



ISSN 2949-2483

Volume

1

Number

2

Special Issue

Artificial intelligence: law and ethics

Executive editor:

Chiara Gallese Nobile, PhD
(Trieste, Italy)

JOURNAL OF DIGITAL TECHNOLOGIES AND LAW

2023

ELECTRONIC
SCIENTIFIC
AND PRACTICAL
JOURNAL





Редакционная коллегия

Шеф-редактор

Бегишев Ильдар Рустамович – доктор юридических наук, доцент, заслуженный юрист Республики Татарстан, главный научный сотрудник Научно-исследовательского института цифровых технологий и права, профессор кафедры уголовного права и процесса Казанского инновационного университета имени В. Г. Тимирязова (Казань, Российская Федерация)

Главный редактор

Жарова Анна Константиновна – доктор юридических наук, доцент, директор Центра исследований киберпространства, ассоциированный член международного научно-образовательного центра «Кафедра ЮНЕСКО по авторскому праву, смежным, культурным и информационным правам» Национального исследовательского университета «Высшая школа экономики», старший научный сотрудник Института государства и права Российской академии наук (Москва, Российская Федерация)

Заместители главного редактора

Громова Елизавета Александровна – кандидат юридических наук, доцент, заместитель директора Юридического института по международной деятельности, доцент кафедры предпринимательского, конкурентного и экологического права Южно-Уральского государственного университета (Национального исследовательского университета) (Челябинск, Российская Федерация)

Залоило Максим Викторович – кандидат юридических наук, ведущий научный сотрудник отдела теории права и междисциплинарных исследований законодательства Института законодательства и сравнительного правоведения при Правительстве Российской Федерации (Москва, Российская Федерация)

Филипова Ирина Анатольевна – кандидат юридических наук, доцент, доцент кафедры трудового и экологического права Национального исследовательского Нижегородского государственного университета имени Н. И. Лобачевского (Нижний Новгород, Российская Федерация)

Шутова Альбина Александровна – кандидат юридических наук, старший научный сотрудник Научно-исследовательского института цифровых технологий и права, доцент кафедры уголовного права и процесса Казанского инновационного университета имени В. Г. Тимирязова (Казань, Российская Федерация)

Редакция

Заведующий редакцией – Дарчинова Гульназ Язкарвна

Выпускающий редактор – Аймурзаева Оксана Анатольевна

Ответственный секретарь – Лапшина Анастасия Денисовна

Редактор – Тарасова Гульнара Абдулахатовна

Технический редактор – Каримова Светлана Альфредовна

Художник-дизайнер – Загреддинова Гульнара Ильгизаровна

Переводчик – Беляева Елена Николаевна, кандидат педагогических наук, член Гильдии переводчиков Республики Татарстан

Специалист по продвижению журнала в сети Интернет –

Гуляева Полина Сергеевна

Адрес: 420111, Российская Федерация,

г. Казань, ул. Московская, 42

Телефон: +7 (843) 231-92-90

Факс: +7 (843) 292-61-59

E-mail: lawjournal@ieml.ru

Сайт: <https://www.lawjournal.digital>

Telegram: <https://t.me/JournalDTL>

ВКонтакте: <https://vk.com/JournalDTL>

Яндекс.Дзен: <https://dzen.ru/JournalDTL>

Одноклассники: <https://ok.ru/JournalDTL>

Учредитель и издатель

Казанский инновационный университет имени В. Г. Тимирязова. Адрес: 420111, Российская Федерация, г. Казань, ул. Московская, 42. Телефон: +7 (843) 231-92-90. Факс: +7 (843) 292-61-59. E-mail: info@ieml.ru. Сайт: <https://ieml.ru>



© Казанский инновационный университет имени В. Г. Тимирязова, оформление и составление, 2023.

Журнал зарегистрирован Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций. Свидетельство о регистрации средства массовой информации: ЭЛ № ФС 77-84090 от 21 октября 2022 г.

Территория распространения: Российская Федерация; зарубежные страны.

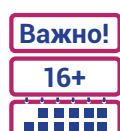


Статьи находятся в открытом доступе и распространяются в соответствии с лицензией Creative Commons «Attribution» («Атрибуция») 4.0 Всемирная (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0/deed.ru>), позволяющей неограниченно использовать, распространять и воспроизводить материал при условии, что оригинальная работа процитирована с соблюдением правил цитирования.

При цитировании любых материалов журнала ссылка обязательна. Ответственность за изложенные в статьях факты несут авторы. Высказанные в статьях мнения могут не совпадать с точкой зрения редакции и не налагают на нее никаких обязательств.

Возрастная классификация: Информационная продукция для детей, достигших возраста шестнадцати лет.

Дата подписания к публикации – 16 июня 2023 г. Дата онлайн-размещения на сайте <https://www.lawjournal.digital> – 20 июня 2023 г.



Международные редакторы

Галлезе-Нобиле Кьяра – доктор наук, научный сотрудник (постдок) по управлению исследовательскими данными Эйндховенского технологического университета (Эйндховен, Королевство Нидерландов), научный сотрудник (постдок) департамента математики и наук о земле Университета Триеста (Триест, Итальянская Республика)

Джайшанкар Каруппаннан – доктор наук, директор и профессор Международного института исследований в сфере криминологии и безопасности (Бенгалуру, Республика Индия)

Мохд Хазми бин Мохд Русли – доктор наук, доцент факультета шариата и права Международного исламского университета Малайзии (Куала-Лумпур, Федерация Малайзия)

Кастилло Парилла Хосе Антонио – доктор наук, магистр новых технологий и права (Севилья, Королевство Испания), научный сотрудник Гранадского университета (Гранада, Королевство Испания)

Члены редакционной коллегии

Арзуманова Лана Львовна – доктор юридических наук, доцент, профессор кафедры финансового права Московского государственного юридического университета имени О. Е. Кутафина (Москва, Российская Федерация)

Бажина Мария Анатольевна – доктор юридических наук, доцент, доцент кафедры предпринимательского права Уральского государственного юридического университета имени В. Ф. Яковлева (Екатеринбург, Российская Федерация)

Беликова Ксения Михайловна – доктор юридических наук, профессор, профессор кафедры предпринимательского и корпоративного права Московского государственного юридического университета имени О. Е. Кутафина (Москва, Российская Федерация)

Берсей Диана Давлетовна – кандидат юридических наук, доцент, доцент кафедры уголовного права и процесса Северо-Кавказского федерального университета (Ставрополь, Российская Федерация)

Будник Руслан Александрович – доктор юридических наук, профессор, заместитель директора международного научно-образовательного центра «Кафедра ЮНЕСКО по авторскому праву, смежным, культурным и информационным правам» Национального исследовательского университета «Высшая школа экономики» (Москва, Российская Федерация)

Дремлюга Роман Игоревич – кандидат юридических наук, доцент, заместитель директора по развитию Института математики и компьютерных технологий, профессор Академии цифровой трансформации Дальневосточного федерального университета (Владивосток, Российская Федерация)

Егорова Мария Александровна – доктор юридических наук, профессор, начальник Управления международного сотрудничества, профессор кафедры конкурентного права Московского государственного юридического университета имени О. Е. Кутафина (Москва, Российская Федерация)

Ефремов Алексей Александрович – доктор юридических наук, доцент, профессор кафедры международного и евразийского права Воронежского государственного университета (Воронеж, Российская Федерация)

Ефремова Марина Александровна – доктор юридических наук, доцент, профессор кафедры уголовно-правовых дисциплин Казанского филиала Российского государственного университета правосудия (Казань, Российская Федерация)

Камалова Гульфия Гафиятовна – доктор юридических наук, доцент, заведующий кафедрой информационной безопасности в управлении Удмуртского государственного университета (Ижевск, Российская Федерация)

Ковалева Наталия Николаевна – доктор юридических наук, профессор, руководитель департамента права цифровых технологий и биоправа факультета права Национального исследовательского университета «Высшая школа экономики» (Москва, Российская Федерация)

Лопатина Татьяна Михайловна – доктор юридических наук, доцент, заведующий кафедрой уголовно-правовых дисциплин Смоленского государственного университета (Смоленск, Российская Федерация)

Минбалеев Алексей Владимирович – доктор юридических наук, доцент, заведующий кафедрой информационного права и цифровых технологий Московского государственного юридического университета имени О. Е. Кутафина (Москва, Российская Федерация)

Миронова Светлана Михайловна – доктор юридических наук, доцент, профессор кафедры финансового и предпринимательского права Волгоградского института управления – филиала Российской академии народного хозяйства и государственной службы при Президенте Российской Федерации (Волгоград, Российская Федерация)

Наумов Виктор Борисович – доктор юридических наук, главный научный сотрудник сектора информационного права и международной безопасности Института государства и права Российской академии наук (Санкт-Петербург, Российская Федерация)

Пашенцев Дмитрий Алексеевич – доктор юридических наук, профессор, заслуженный работник высшей школы Российской Федерации, главный научный сотрудник отдела теории права и междисциплинарных исследований законодательства Института законодательства и сравнительного правоведения при Правительстве Российской Федерации (Москва, Российская Федерация)

Петренко Сергей Анатольевич – доктор технических наук, профессор, профессор кафедры информационной безопасности Санкт-Петербургского государственного электротехнического университета «ЛЭТИ» имени В. И. Ульянова (Ленина), профессор Университета Иннополис (Иннополис, Российская Федерация)

Полякова Татьяна Анатольевна – доктор юридических наук, профессор, заслуженный юрист Российской Федерации, и. о. заведующего сектором информационного права и международной информационной безопасности Института государства и права Российской академии наук (Москва, Российская Федерация)

Пономарева Карина Александровна – доктор юридических наук, ведущий научный сотрудник Центра налоговой политики Научно-исследовательского финансового института Министерства финансов Российской Федерации, профессор департамента публичного права Национального исследовательского университета «Высшая школа экономики» (Москва, Российская Федерация)

Рожкова Марина Александровна – доктор юридических наук, главный научный сотрудник Института законодательства и сравнительного правоведения при Правительстве Российской Федерации, советник по науке декана юридического факультета Государственного академического университета гуманитарных наук, президент IP CLUB (Москва, Российская Федерация)

Рускевич Евгений Александрович – доктор юридических наук, доцент, профессор кафедры уголовного права Московского государственного юридического университета имени О. Е. Кутафина (Москва, Российская Федерация)

Сидоренко Элина Леонидовна – доктор юридических наук, доцент, директор Центра цифровой экономики и финансовых инноваций, профессор кафедры уголовного права, уголовного процесса и криминалистики Московского государственного института международных отношений (университета) Министерства иностранных дел Российской Федерации, генеральный директор платформы забизнес.рф (Москва, Российская Федерация)

Степанян Армен Жоресович – кандидат юридических наук, доцент, доцент кафедры интеграционного и европейского права Московского государственного юридического университета имени О. Е. Кутафина (Москва, Российская Федерация)

Стрельцов Анатолий Александрович – доктор юридических наук, доктор технических наук, профессор, заслуженный деятель науки Российской Федерации, член-корреспондент Академии криптографии Российской Федерации, ведущий научный сотрудник Центра проблем информационной безопасности Московского государственного университета имени М. В. Ломоносова (Москва, Российская Федерация)

Талапина Эльвира Владимировна – доктор юридических наук, доктор права (Франция), главный научный сотрудник Института государства и права Российской академии наук, ведущий научный сотрудник Центра технологий государственного управления Российской академии народного хозяйства и государственной службы при Президенте Российской Федерации (Москва, Российская Федерация)

Терентьева Людмила Вячеславовна – доктор юридических наук, доцент, профессор кафедры международного частного права Московского государственного юридического университета имени О. Е. Кутафина (Москва, Российская Федерация)

Харитоновая Юлия Сергеевна – доктор юридических наук, профессор, руководитель Центра правовых исследований искусственного интеллекта и цифровой экономики, профессор кафедры предпринимательского права Московского государственного университета имени М. В. Ломоносова (Москва, Российская Федерация)

Хисамова Зарина Илдузовна – кандидат юридических наук, начальник отделения планирования и координации научной деятельности научно-исследовательского отдела Краснодарского университета Министерства внутренних дел Российской Федерации (Краснодар, Российская Федерация)

Чеботарева Анна Александровна – доктор юридических наук, доцент, заведующий кафедрой правового обеспечения государственного управления и экономики Российского университета транспорта (Москва, Российская Федерация)

Шугуров Марк Владимирович – доктор философских наук, доцент, профессор кафедры международного права Саратовской государственной юридической академии, главный научный сотрудник Алтайского государственного университета (Саратов, Российская Федерация)

Иностранные члены редакционной коллегии

Абламейко Мария Сергеевна – кандидат юридических наук, доцент, доцент кафедры конституционного права Белорусского государственного университета (Минск, Республика Беларусь)

Аванг Низам Мухаммад – доктор наук, профессор факультета права и шариата Международного исламского университета (Негери-Сембилан, Федерация Малайзия)

Айсан Ахмет Фарук – доктор наук, профессор и координатор программы Исламских финансов и экономики Университета имени Хамада бин Халифа (Доха, Государство Катар)

Ападхьяй Нитиш Кумар – доктор юридических наук, доцент факультета права Университета Галготиас (Большая Нойда, Республика Индия)

Банкио Пабло – доктор наук, профессор Университета Буэнос-Айреса, постдок в области фундаментальных принципов и прав человека, член центра изучения частного права Национальной академии наук Буэнос-Айреса (Буэнос-Айрес, Аргентинская Республика)

Басарудин Нур Ашикин – доктор наук, старший преподаватель Университета технологий МАРА (Синток, Федерация Малайзия)

Бахрамова Мохинур Бахрамовна – доктор наук, старший преподаватель кафедры права интеллектуальной собственности Ташкентского государственного юридического университета (Ташкент, Республика Узбекистан)

Варбанова Гергана – доктор наук, доцент Университета экономики (Варна, Республика Болгария), доцент Университета мировой экономики (София, Республика Болгария)

Вудро Барфилд – доктор наук, приглашенный профессор Туринского университета (Турин, Итальянская Республика)

Гозстоный Гегели – доктор наук, кафедра истории венгерского государства и права Университета Эотвос Лоранд (Будапешт, Венгрия)

Гостожиц Стеван – доктор наук, доцент, глава цифровой криминалистической лаборатории Университета Нови Сад (Нови Сад, Республика Сербия)

Гош Джаянта – доктор наук, научный сотрудник Западно-Бенгальского национального университета юридических наук (Калькутта, Республика Индия)

Гудков Алексей – доктор наук, старший преподаватель Вестминстерского международного университета в Ташкенте (Ташкент, Республика Узбекистан)

Дауд Махауддин – доктор наук, доцент кафедры гражданского права Международного исламского университета Малайзии (Куала-Лумпур, Федерация Малайзия)

Дахдал Эндрю – доктор наук, доцент факультета права Катарского университета (Доха, Государство Катар)

Дэнни Тэйм Даниэль Мендес – доктор наук, научный сотрудник Азиатско-Тихоокеанского центра экологического права Национального университета Сингапура (Сингапур, Республика Сингапур)

- Иванц Тьяша** – доктор наук, доцент кафедры гражданского, международного частного и сравнительного права Мариборского университета (Марибор, Республика Словения)
- Иоаннис Револидис** – доктор наук, преподаватель кафедры медиаправа и права технологий Мальтийского университета (Мсида, Республика Мальта)
- Йованич Татьяна** – доктор наук, доцент факультета права Белградского университета (Белград, Республика Сербия)
- Карим Ридоан** – доктор наук, профессор кафедры предпринимательского и налогового права Университета Монаша (Санвэй, Федерация Малайзия)
- Кастро Дуглас** – доктор наук, профессор международного права школы права Ланьчжоуского университета (Ланьчжоу, Китайская Народная Республика)
- Кера Решеф Дениза** – доктор наук, профессор Центра исследований технологий распределенного реестра Мальтийского университета (Мсида, Республика Мальта)
- Кипурас Павлос** – доктор наук, профессор Школы судебной графологии (Неаполь, Итальянская Республика)
- Мараньяо Альбукерке де Соуза Джулиано** – доктор наук, доцент факультета права Университета Сан-Паулу (Сан-Паулу, Федеративная Республика Бразилия)
- Мелипатаки Габор** – доктор наук, профессор кафедры аграрного и трудового права Университета Мишкольца (Мишкольц, Венгрия)
- Мехрдад Райеджиан Асли** – доктор наук, профессор Института исследований и развития в области гуманитарных наук, доцент кафедры ЮНЕСКО по правам человека, мира и демократии, заместитель декана по науке Университета имени Алламеха Табатабаи (Тегеран, Иран)
- Морина Менсур** – доктор наук, доцент, заместитель декана факультета права Университета бизнеса и технологий (Приштина, Республика Сербия)
- Мохсин Камшад** – доктор наук, доцент юридического факультета Международного университета Махариши (Махариши, Республика Индия)
- Муратаев Серикбек Алпамысович** – кандидат юридических наук, заведующий кафедрой теории государства и права Ташкентского государственного юридического университета (Ташкент, Республика Узбекистан)
- Нуреддин Мухамад** – доктор наук, старший преподаватель кафедры публичного права Университета Байеро (Кано, Федеративная Республика Нигерия)
- Праюди Юди** – доктор наук, профессор кафедры компьютерных наук и электроники Университета Гаджа Мада (Булаксурмур, Республика Индонезия)
- Рахметов Бауржан Жанатович** – доктор наук, ассистент-профессор Международной школы экономики Университета КАЗГЮУ имени М. С. Нарикбаева (Нур-Султан, Республика Казахстан)
- Тран Ван Нам** – доктор наук, директор факультета права Национального экономического университета (Ханой, Социалистическая Республика Вьетнам)
- Феррейра Даниэл Брантес** – доктор наук, старший научный сотрудник Южно-Уральского государственного университета (Челябинск, Российская Федерация), профессор Университета АМБРА (Орландо, Соединенные Штаты Америки), исполнительный директор Центра альтернативного разрешения споров (Рио-де-Жанейро, Федеративная Республика Бразилия)
- Чен Чао Хан Кристофер** – доктор наук, доцент факультета права Тайваньского национального университета (Тайпей, Китайская Народная Республика)
- Шахновская Ирина Викторовна** – кандидат юридических наук, заведующий кафедрой конституционного права и государственного управления Полоцкого государственного университета (Новополоцк, Республика Беларусь)
- Эллул Джошуа** – доктор наук, директор Центра исследований технологии распределенного реестра Мальтийского университета (Мсида, Республика Мальта)
- Юхневич Эдвард** – доктор наук, профессор кафедры финансового права Гданьского университета (Гданьск, Республика Польша)



Содержание

Галлезе-Нобиле К.

Правовые аспекты использования искусственного интеллекта
в телемедицине 314

Харитонов Ю. С.

Правовые средства обеспечения принципа прозрачности
искусственного интеллекта 337

Филипова И. А., Коротеев В. Д.

Будущее искусственного интеллекта: объект или субъект права? 359

Фаллетти Э.

Алгоритмическая дискриминация и защита неприкосновенности
частной жизни 387

Ерахтина О. С.

Подходы к регулированию отношений в сфере разработки
и применения технологий искусственного интеллекта:
особенности и практическая применимость 421

Казанцев Д. А.

Проблемы и перспективы регулирования отношений в рамках
сделки, совершенной с участием искусственного интеллекта 438

Хасан Ф. М., Осман Н. Д.

Автономное вооружение на основе искусственного интеллекта
и индивидуальная уголовная ответственность согласно Римскому статуту..... 464

Спирidonov М. С.

Технологии искусственного интеллекта
в уголовно-процессуальном доказывании..... 481

Рицу С.

Рекомендации по этическим аспектам искусственного интеллекта
в приложении к сфере трудовых отношений 498

Бахтеев Д. В.

Этико-правовые модели взаимоотношений общества с технологией
искусственного интеллекта 520

Шумакова Н. И., Титова Е. В.

Искусственный интеллект как вспомогательный инструмент ограничения
религиозной свободы в Китае 540

Резаев А. В., Трегубова Н. Д.

Возможность и необходимость человеко-ориентированного
искусственного интеллекта в юридической теории и практике 564

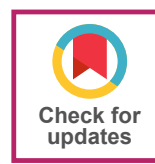


Научная статья

УДК 347.1:654:004.8

EDN: <https://elibrary.ru/vskcfb>

DOI: <https://doi.org/10.21202/jdtl.2023.13>



Правовые аспекты использования искусственного интеллекта в телемедицине

Кьяра Галлезе-Нобиле 

Эйнховенский технологический университет

г. Эйнховен, Королевство Нидерландов;

Университет Триеста

г. Триест, Итальянская Республика

Ключевые слова

Законодательство,
защита данных,
искусственный интеллект,
персональные данные,
право,
регулирование,
телемедицина,
цифровое неравенство,
цифровые технологии,
частная жизнь

Аннотация

Цель: стремительное распространение телемедицины в клинической практике и возрастающая роль искусственного интеллекта ставят перед юристами множество проблем относительно охраны неприкосновенности частной жизни. Повышенная сензитивность данных в этой области заставляет уделить особое внимание правовым аспектам таких систем. В статье исследуются правовые последствия использования искусственного интеллекта в телемедицине, в частности, систем непрерывного обучения и автоматизированного принятия решений; фактически оказание персонализированных медицинских услуг через системы непрерывного обучения может представлять дополнительный риск. Особого внимания заслуживают уязвимые группы населения – дети, пожилые люди и тяжелобольные пациенты – как по причине цифрового неравенства, так и из-за сложностей с выражением своего согласия.

Методы: сравнительно-правовые и формально-правовые методы исследования позволили проанализировать текущее состояние регулирования искусственного интеллекта и выявить его соотношение с нормами регулирования телемедицины, Общим регламентом ЕС по защите персональных данных и другими нормами.

Результаты: изучены правовые последствия использования искусственного интеллекта в телемедицине, в частности, систем непрерывного обучения и автоматизированного принятия решений; автор приходит к выводу, что оказание персонализированных медицинских услуг через системы непрерывного обучения представляет дополнительные риски, и предлагает пути их минимизации. Автор также уделяет особое внимание вопросам информированного согласия уязвимых групп населения (детей, пожилых, тяжелобольных).

© Галлезе-Нобиле К., 2023



Статья находится в открытом доступе и распространяется в соответствии с лицензией Creative Commons «Attribution» («Атрибуция») 4.0 Всемирная (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0/deed.ru>), позволяющей неограниченно использовать, распространять и воспроизводить материал при условии, что оригинальная работа упомянута с соблюдением правил цитирования.

Научная новизна: изучены актуальные риски и проблемы, возникающие в сфере использования искусственного интеллекта в телемедицине, при этом особое внимание уделено системам непрерывного обучения.

Практическая значимость: полученные результаты восполняют недостаток научных исследований по данной теме, могут быть использованы в законодательном процессе в сфере использования искусственного интеллекта в телемедицине, а также в качестве основы для будущих исследований в данной области.

Для цитирования

Галлезе-Нобиле, К. (2023). Правовые аспекты использования искусственного интеллекта в телемедицине. *Journal of Digital Technologies and Law*, 1(2), 314–336. <https://doi.org/10.21202/jdtl.2023.13>

Содержание

Введение

1. Нормативно-правовая база телемедицины в Европе
2. Искусственный интеллект в телемедицине
3. Непрерывное обучение и персонализированная медицина
4. Вопросы охраны частной жизни в телемедицине
5. Статья 22 Общего регламента ЕС по защите персональных данных и контроль со стороны человека
6. Информированное согласие
7. Уязвимые группы
8. Баланс между охраной неприкосновенности частной жизни и защитой от вреда при дистанционной работе
9. Меры дополнительной защиты в форме аудита искусственного интеллекта

Заключение

Список литературы

Введение

Термин «телемедицина» был предложен в 1970-е гг. Томасом Бёрдом (Thomas Bird); Strehle и Shabde определили его как «лечение на расстоянии» (Strehle & Shabde, 2006). С течением времени появилось несколько официальных определений, например: «предоставление услуг здравоохранения, путем использования ИКТ, в ситуациях, когда специалист в области здравоохранения и пациент (или два специалиста в области здравоохранения) не находятся в одном месте. Телемедицина включает в себя безопасную передачу медицинских данных и информации в форме текста, звуков, изображений и других форм, необходимой профилактики, диагностирования, лечения и последующего ведения пациентов. Телемедицина охватывает широкий спектр услуг. Наиболее часто упоминаемыми из них в рецензируемых обзорах являются телерадиология, телепатология, теледерматология, телеконсультирование, теленаблюдение, телехирургия и телеофтальмология. Другие возможные услуги включают

колл-центры/информационные центры онлайн для пациентов, удаленные консультации/электронные приемы или видеоконференции специалистов.

Информационные порталы, электронные системы ведения медицинских карт, электронная передача рецептов и назначений (е-рецепты, е-назначения) не рассматриваются как услуги телемедицины для целей данного Сообщения¹, которое стало основой для государственного регулирования (например, для определения, сформулированного министерством Италии в 2012 г.)².

Новый метод предоставления медицинских услуг не только способствует оптимизации и повышению эффективности данных процессов, при этом не заменяя традиционную очную медицину (Burrai et al., 2021), но и лучше организует последующее ведение пациента и профилактические мероприятия; он более удобен, особенно для маломобильных и тяжелых пациентов. Фактически по сравнению с традиционной медициной устройства, используемые для наблюдения за пациентом в домашних условиях, позволяют им реже ездить в больницу, с комфортом оставаться дома (или, например, в гостинице, если они заболели на отдыхе или в деловой поездке). Это особенно важно во время пандемии, так как снижает вероятность распространения инфекции или заражения в медицинском учреждении. Можно отметить, что именно ситуация с вирусом Covid-19 способствовала развитию телемедицины, так как она дала возможность получить медицинскую помощь даже в условиях ограничений на передвижение.

При этом такая уникальная возможность несет в себе определенные риски и поднимает множество различных правовых вопросов. В настоящей статье мы фокусируемся на правовых вопросах, связанных с использованием искусственного интеллекта (далее – ИИ) в телемедицине, и в частности систем непрерывного обучения.

1. Нормативно-правовая база телемедицины в Европе

Стремительное распространение телемедицины в клинической практике некоторое время назад заставило Евросоюз изучить последствия использования новых технологий для пациентов, развития рынка электронного здравоохранения, создания общеевропейского пространства данных о здоровье, а также возможного влияния всех этих факторов на сферу здравоохранения в странах-членах ЕС. В этой связи в течение нескольких лет были выпущены ряд инструментов «мягкого права», такие как руководства, рекомендации и другие инструменты. Они подробно проанализированы в работе Botrugno (Bortugno, 2014). Кроме того, Регламент Евросоюза о медицинских устройствах, вступивший в силу 26 мая 2021 г., относится ко всем телемедицинским устройствам, используемым для постановки диагноза и оказания помощи на расстоянии³. В Италии данная сфера также в основном регулируется мяг-

¹ Сообщение Комиссии Европейскому Парламенту, Совету, Европейскому экономическому и социальному комитету и Комитету регионов о телемедицине на благо пациентов, систем здравоохранения и общества в целом, 2008. https://commission.europa.eu/system/files/2022-10/cwp_2023.pdf

² Ministero della Salute. Telemedicina – Linee di indirizzo nazionali, 2012. https://www.salute.gov.it/portale/documentazione/p6_2_2_1.jsp?id=2129

³ Regulation (Eu) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R0745>

ким правом, в частности, Руководством от 2012 г.⁴ Кроме того, в 2017 г. был принят Закон № 219, регулирующий вопросы информированного согласия и Распоряжения об ожидаемом обращении на случай возможной недееспособности. Как будет показано ниже, данная сфера приобретает особую важность в случае использования телемедицины через интеллектуальные системы. Исследователи отмечают, что если рассматривать эмоциональные потребности пациента как неотъемлемую часть терапии, то телемедицину следует считать услугой, встроенной в традиционные схемы (Campagna, 2020).

Как показано в работе Campagna, существующая нормативно-правовая база будет в ближайшее время дополнена новым европейским законом об ИИ, одобренным Еврокомиссией в 2021 г. (Campagna, 2020). Он предусматривает изменения в статусе телемедицинских устройств на основе ИИ, налагая строгие требования на их использование в медицинской практике, клинических испытаниях и научных исследованиях. Фактически медицинские устройства классифицируются данным положением как «высокорисковые». Новый закон станет дополнением Общего регламента ЕС по защите персональных данных (GDPR) и Европейского регламента о медицинских изделиях (MDR), предусматривая дополнительные гарантии для пользователей (включая как пациентов, так и медицинский персонал).

2. Искусственный интеллект в телемедицине

По мере технологического прогресса многие технологии искусственного интеллекта начали применяться в телемедицине с целью улучшить ее результативность, поскольку в ряде областей (таких как распознавание образов) возможности ИИ уже превосходили возможности человека.

Однако специалисты в области здравоохранения по-прежнему остерегаются использовать эти технологии в клинической практике: исследования, проведенные в 2020 г. Лабораторией цифровых инноваций в здравоохранении Политехнического университета Милана по теме «Сопутствующая медицинская помощь в чрезвычайных ситуациях, связанных с Covid-19», показали, что лишь 9 % итальянских врачей использовали новые технологии до пандемии и только 6 % из них работают в учреждениях, где такие технологии были внедрены (или усилены) во время пандемии. При этом 60 % медиков считают, что технологии искусственного интеллекта могут сыграть ключевую роль в чрезвычайных ситуациях, 52 % – что они способствуют персонализации медицинского обслуживания, 51 % думают, что они повышают эффективность медицинского обслуживания, а 50 % – что помогают снизить вероятность медицинских ошибок. Схожие результаты были получены и в других странах, где также поднимались вопросы врачебной ответственности (Scheetz et al., 2021). Однако исследования показывают, что широкая общественность с огромным недоверием относится к использованию моделей ИИ в медицине (Castagno & Khalifa, 2020).

Тенденции развития использования искусственного интеллекта в телемедицине можно разделить на четыре группы: наблюдение за пациентами, информация о здоровье, помощь в диагностировании и анализ информации (Pacis et al., 2018).

⁴ Ministero della Salute, Telemedicina – Linee di indirizzo nazionali. <https://www.salute.gov.it/portale/ehealth/homeEHealth.jsp>

Данные технологии наиболее часто используются в таких областях медицины, как лечение диабета, кардиология, офтальмология, онкология, эпидемиология, дерматология. Во время пандемии телемедицина использовалась, помимо прочего, для ведения пациентов, подозревавшихся в инфицировании, и для помощи по хроническим заболеваниям (Ye, 2020). Как правило, пациенты носили на себе съемные устройства, такие как смарт-часы или датчики, или пользовались приложением для планшета; но также использовались и более инвазивные методы, например, пациент мог глотать капсулы, в доме могли устанавливаться камеры или использовались датчики для контроля действий, например, открывания упаковки с лекарством. Также применялись малоинвазивные методы, такие как использование приложения на смартфоне, особенно для молодых пациентов, которые могут приобщиться к телемедицине через игру (Giunti, 2014), и пожилых (Schatten & Protrka, 2021), которых необходимо стимулировать к знакомству с системами ИИ. Устройства с более сложными технологиями могут приспосабливаться к конкретному пациенту, продолжая обучаться во время использования. Это порождает ряд этических и правовых проблем, по которым доктрина и юридическая наука пытаются вести глубокий анализ, чтобы выработать руководящие принципы для исследователей и практиков в области здравоохранения, которые участвуют в разработке таких устройств. Фактически очень часто нормативное регулирование оказывается очень сложным даже для юристов, как это произошло в случае защиты персональных данных, что затрудняет эффективную защиту пациентов.

Комплексное нормативное регулирование осложняется также фрагментированностью сферы ИИ в рамках Евросоюза; исправить эту ситуацию призван новый проект закона об ИИ. На сегодняшний день, однако, сохраняются и сохранятся даже после вступления в силу этого закона многие особенности, характерные для отдельных стран – членов Евросоюза, касающиеся защиты персональных данных, профессиональной врачебной ответственности, информированного согласия, ответственности за качество продукции и уголовной ответственности.

3. Непрерывное обучение и персонализированная медицина

Одна из самых популярных моделей предоставления персонализированных медицинских услуг заключается в так называемом непрерывном обучении. Врачи стремятся предоставить пациентам наилучшие услуги, при этом в ряде случаев эти услуги необходимо привести в соответствие с потребностями конкретного пациента, которые различаются в зависимости от его этнической принадлежности, пола, привычек, семейного анамнеза, психологического состояния и т. д. Этой цели можно эффективно достичь с помощью ИИ, системы которого обучаются в процессе использования пациентом и приспосабливаются к характеристикам последнего с течением времени. С точки зрения права особого внимания заслуживают модели на основе машинного обучения (особенно глубокого обучения), а также модели на основе подхода «черный ящик» (Rodrigues, 2020; Lakkaraju, 2019), так как в этих случаях разработчики создают модель и предоставляют соответствующие образцы, но не знают конечного результата, потому что модель учится самостоятельно (перед выпуском на рынок, во время фазы обучения, а также после продажи). При этом часто невозможно узнать причин того или иного результата, выдаваемого моделями, поэтому важно исследовать и понимать, как они будут вести себя в рамках различных сценариев (Davis, 2016).

Эта проблема не ограничена телемедициной, но относится ко всем приложениям ИИ; однако использование такого типа моделей в контексте телемедицины ведет к крупным рискам – не только из-за типа используемых данных, но и из-за потенциального влияния на базовые права пациентов. Указанные системы могут представлять особую опасность в случае телемедицины, чем в случае очной медицины, поскольку пациент, пребывая в удалении, не находится под непосредственным наблюдением врача. Данную проблему мы рассмотрим ниже. Например, при мониторинге диабета у детей используются смарт-часы, измеряющие уровень сахара в крови, физиологические показатели (такие как сердечный ритм) напоминают о необходимости принимать пищу и двигаться, предлагают необходимый объем лекарства, который должен принять пациент, и передают результат врачу. На основании этого результата врач может назначить прием, анализы и медицинские исследования. В случае непрерывного обучения одна из ключевых проблем состоит в том, что ни разработчик, ни врач, который будет пользоваться системой ИИ, не знает заранее, как она будет себя вести и чему она обучится при взаимодействии с пациентом, поскольку эта система развивается во времени. Фактически следует принимать во внимание четыре аспекта. Во-первых, качество обучения отражает качество образцов: если пользователь предоставляет системе искаженные либо низкокачественные образцы, то поведение ИИ в конечном итоге изменится таким образом, чтобы соответствовать этим образцам.

Ярким примером является случай с чат-ботом Тэй⁵. В персонализированной телемедицине многие образцы предоставляются системе непосредственно пациентом. Очевидно, что пациент не специалист и зачастую не может понять последствий своего выбора. Например, ребенок может одолжить смарт-часы однокласснику или младшему брату, в результате чего в систему попадут неверные данные. Система решит, что уровень сахара в крови ребенка ниже, чем на самом деле, и порекомендует принять меньше лекарства. Даже если количество принимаемого лекарства контролируется врачом (что маловероятно в случае ежедневного приема лекарства, так как ежедневное наблюдение очень дорогостоящее), врач не поймет, что уровень глюкозы указан неверно из-за внешних факторов или из-за неправильного использования устройства пациентом, если только устройство не оснащено камерами наблюдения. В таких случаях единственным способом предотвратить неправильное использование устройства является теленаблюдение, что расценивается как вмешательство в частную жизнь и крайне нежелательно для пациента. Даже с точки зрения ответственности сложно определить, кто должен отвечать за ошибку системы в таких случаях. Должен ли производитель или разработчик нести ответственность за выпуск небезопасного устройства, не оснащенного механизмом наблюдения? Или врач за то, что не заметил ненормальные показатели крови? Родитель или опекун ребенка, не отслеживший неправильное использование устройства? Для решения некоторых из этих вопросов 28 сентября 2022 г. Еврокомиссия опубликовала проект Директивы об ответственности ИИ, служащей дополнением к Закону ЕС об ИИ, в котором медицинские устройства рассматриваются как «высокорисковые системы»,

⁵ Бот Тэй на основе искусственного интеллекта, принадлежащий компании Microsoft, стал расистским и нацистским. *The Guardian*. <https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter>

а значит, должны соответствовать строгим требованиям. За необходимость подобного реформирования давно выступали ученые, поскольку существующие нормы ответственности не в состоянии регулировать использование систем ИИ, и в частности моделей машинного обучения (Gallese, 2022).

Новый проект устанавливает презумпцию ответственности лица, выпустившего систему на рынок: как правило, производитель или импортер должен будет нести ответственность за вред, причиненный системой, а «национальные суды должны признавать, с целью применения норм ответственности за вред, причинно-следственную связь между виной ответчика и результатами, произведенными системой ИИ, либо неспособностью системы ИИ произвести результат». Из этой новой нормы предусмотрены два исключения, в случае если заявитель может показать, что пользователь системы «не выполнил свои обязательства по использованию или наблюдению за системой ИИ в соответствии с прилагаемыми инструкциями по ее использованию или, в соответствующих случаях, приостановил или прервал ее использование согласно [Статье 29 Закона об ИИ]» либо «предоставил системе ИИ исходные данные, находящиеся под его контролем, которые не соответствовали предполагаемым задачам системы согласно [Статье 29(3) Закона об ИИ]». Данная норма направлена на защиту пользователей, поскольку сложно оценить причинно-следственную связь между поведением системы и неверным результатом (особенно в случае «черного ящика»).

В рамках данной нормы врачей сложно привлечь к ответственности как потому, что они не могут физически влиять на систему (если только им не предоставлены полный дистанционный контроль и средства теленаблюдения), так и потому, что они не являются специалистами по методам машинного обучения в такой степени, чтобы корректировать обучение модели. Второй момент, который необходимо учитывать, – это склонность методов машинного обучения к чрезмерной подгонке, т. е. потере способности к обобщению. Такая ситуация, хорошо знакомая практикам в области машинного обучения, обычно выявляется при наблюдении зависимости между ошибкой обучения и тестовой ошибкой: исправленная ошибка на обучающей выборке приводит к усилению ошибки на тестовой выборке; это говорит о чрезмерной подгонке, процесс обучения нужно остановить. Последний метод является одним из самых широко известных и эффективных методов предотвращения чрезмерной подгонки; он носит название «ранней остановки» и, очевидно, не применим к системам, которые одновременно продаются «готовыми» (т. е. процесс обучения был остановлен до наступления чрезмерной подгонки) и способны обучаться вне процесса производства.

В нашем примере смарт-часы могут потерять способность к обобщению после того, как пациент пользовался ими некоторое время, как по причине некорректного использования, так и потому, что в течение этого времени наблюдения оказались несбалансированными (например, пациент болел и его сердечный ритм или анализы крови длительное время были не в норме). С правовой точки зрения это обстоятельство будет значимым не только для разработчиков согласно новому режиму ответственности ИИ, но также для нового регламента безопасности, вводимого Законом об ИИ и проектом так называемого закона о киберустойчивости. В-третьих, эти проблемы усугубляются в случае использования методов непрерывного обучения в течение всей жизни (Parisi, 2019), так как при этом система непрерывно получает новые образцы и изменяет свое поведение. Фактически такие методы

машинного обучения представляют проблему, так как новые образцы часто являются несбалансированными (например, некоторые категории представлены шире, чем другие, выше вероятность встретить их в реальном мире), что может существенно повлиять на качество обучения и будущее поведение ИИ. Так, в нашем примере со смарт-часами конкретные характеристики пациента могут привести к искажению определенных физиологических параметров и некорректному результату.

Наконец, проблема вообще не имеет решения в случае пошагового обучения на генеральной выборке (Mi, 2020), когда машина получает новые образцы, которые могут, в принципе, принадлежать новым категориям, не рассматривавшимся ранее: в этом случае алгоритм должен быть способен перестроить свое внутреннее функционирование (например, в случае глубоких нейросетей адаптировать архитектуру, изменить топологию, число нейронов, заново откалибровать все параметры), что явно противоречит любой реалистической возможности предсказать будущее поведение системы. В нашем примере пациент может иметь уникальные характеристики, не совпадающие с физиологическими показателями, заложенными в систему при обучении. Это может быть опасно для пациента, так как неверные предположения системы могут привести к назначению неправильной дозы лекарства, при этом в отсутствие врача ошибка не будет исправлена.

Предложенное дополнение к Закону об ИИ лишь поверхностно касается вопроса о непрерывном обучении в ст. 15: «<...> Высокорисковые системы ИИ, продолжающие обучаться после выпуска на рынок или запуска в работу, должны разрабатываться таким образом, чтобы гарантировать применение адекватных мер сдерживания против потенциально предвзятых результатов, получаемых из-за использования других результатов в качестве исходных данных для будущих операций ("петля обратной связи")...» В Преамбуле 78 уточняется, что в целях обеспечения учета провайдерами высокорисковых систем ИИ опыта использования высокорисковых систем ИИ при усовершенствовании их систем и в процессе разработки и усовершенствования процессов, а также своевременного принятия всех возможных мер для их коррекции все провайдеры должны применять систему послепродажного мониторинга. Такая система является ключевым элементом, гарантирующим возможность принятия эффективных и своевременных мер против потенциальных рисков, возникающих в результате продолжающегося «обучения» после выпуска на рынок или запуска в эксплуатацию. В данном контексте провайдеры также обязаны иметь систему, позволяющую сообщать соответствующим органам власти о любых серьезных инцидентах и о нарушениях национального и общеевропейского законодательства о защите базовых прав, происшедших в результате использования принадлежащих им систем ИИ.

Данное положение сформулировано крайне расплывчато и не проясняет, какие именно возможные сдерживающие меры или коррекционные действия можно считать адекватными. Практическое применение данного параграфа будет затруднено из-за его нечеткости и возможности произвольного толкования судебными органами. Положение о «системе послепродажного мониторинга» – это полезная мера, однако она вряд ли может быть эффективной в приложении к персонализированной медицине, поскольку производители не могут постоянно следить за каждым пациентом. Непрерывное обучение во многих отношениях представляет собой большую правовую проблему (Marchant & Lindor, 2012).

Например, она ставит под сомнение традиционную парадигму ответственности: возможно ли адаптировать режим строгой ответственности к ситуации, когда ни разработчик, ни производитель не могут предсказать поведение ИИ? Кроме того, безопасен ли с технической точки зрения для пациентов продукт, действия которого нельзя полностью объяснить, даже если его обучение осуществляется производителем? Исследователями была выдвинута так называемая теория пробела ответственности (Matthias, 2004), однако другие ученые ее отвергают (Tigard, 2020).

В таких случаях очень сложно предусмотреть профессиональную ответственность со стороны врача. Возникает также множество вопросов относительно охраны неприкосновенности частной жизни. Например, поскольку сотрудники системы здравоохранения не контролируют способ, которым указанные системы обрабатывают данные о здоровье пациентов, то они не могут полностью соблюсти свои обязанности в области прозрачности деятельности. В следующем разделе мы рассмотрим проблемы неприкосновенности частной жизни.

4. Вопросы охраны частной жизни в телемедицине

Один из самых актуальных вопросов в сфере телемедицины и искусственного интеллекта – это, несомненно, вопрос охраны персональных данных, и в частности Европейский регламент № 679/2016 (Общий регламент ЕС по защите персональных данных, GDPR). Согласно этому регламенту, данные о здоровье, наряду с некоторыми другими, относятся к «особым категориям данных» (ст. 9) и требуют повышенной правовой защиты.

В телемедицине, как и в традиционной медицине, осуществляется обработка особых категорий персональных данных, а именно данных о здоровье. Данные пациентов, которыми пользуются модели ИИ в здравоохранении, редко могут быть полностью анонимными; чаще всего считается, что они псевдонимизированы, так как медицинское учреждение может сопоставить их с именами и другими личными данными пациентов.

Даже если телемедицинские устройства не основаны на ИИ, личности пациентов большую часть времени известны медицинским работникам, так как цель системы здравоохранения – предоставлять медицинскую помощь конкретному лицу. При этом применяются Общий регламент ЕС по защите персональных данных и другие нормы в области охраны неприкосновенности частной жизни (такие как законы отдельных стран, внутренние отраслевые положения и т. д.). Из-за способа, которым осуществляется телемедицинская деятельность, неизбежно будет возникать ряд вопросов в сфере частной жизни.

Самый очевидный из них – это вопрос безопасности, поскольку взаимодействие на расстоянии предполагает установление связи между участниками (между врачом и пациентом или между несколькими медицинскими работниками). Должны выполняться меры безопасности, предусмотренные ст. 32 Общего регламента ЕС по защите персональных данных: необходимо гарантировать, что используется безопасное соединение, что личности пациента и врача подтверждены, что все задействованные лица имеют адекватную подготовку, что данные корректно хранятся, а когда перестают быть нужными – уничтожаются. Риск утечки данных особенно высок в случае работы с пожилыми пациентами, которые обычно не знакомы с новыми технологиями. Устройства интернета вещей могут теряться, пароли взламываются недобросовестными пользователями, данные могут быть удалены по ошибке. Сам факт нахождения устройства в руках пациента означает, что специалисты в области здравоохранения и информационных технологий в этот момент не контролируют систему.

Поэтому до введения таких устройств пациент должен получить адекватную подготовку. Однако недавние атаки на информационные системы больниц показали, что применяемые ими организационные меры не всегда отвечают поставленным задачам, главным образом из-за недостаточной подготовки и низкого уровня знаний о кибербезопасности у персонала (Gioulekas, 2022).

Предоставляя услуги телемедицины, медицинские учреждения должны убедиться, что они могут выполнить все требования ст. 32 Общего регламента ЕС по защите персональных данных, включая подготовку своего персонала в области основных мер кибербезопасности. Если медицинский работник не способен обучиться мерам безопасности, то он, вероятно, не сможет и контролировать взаимодействие пациентов с системой. Не будучи техническими специалистами, они, вероятно, не смогут также определить сбои в системе, такие как некорректное обучение нейросети. В этом контексте необходимо, чтобы устройство регулярно подвергалось проверке, обновлению и калибровке со стороны специалистов по ИИ. Дополнительные гарантии в отношении особых категорий данных предусмотрены Законом об ИИ в последнем разделе ст. 10: они должны обрабатываться в той мере, которая необходима для целей предотвращения необъективности, обнаружения и коррекции в отношении высокорисковых систем ИИ, однако при этом необходимы соответствующие гарантии (такие как технические ограничения на повторное использование данных, а также применение новейших технических средств, включая псевдонимизацию или шифрование).

При использовании системы ИИ для предоставления услуг телемедицины необходимо на уровне разработки по умолчанию внедрить в эту систему принципы охраны частной жизни, принимая во внимание также принципы минимизации данных и ограничения по их хранению. С этой целью в Преамбуле 44 предусмотрена дополнительная задача обработки персональных медицинских данных, а именно «мониторинг, обнаружение и коррекция необъективности» для создания справедливых и надежных систем ИИ.

Выполнение вышеупомянутых требований может быть затруднено в случае непрерывного обучения: как можно гарантировать точность, обновляемость и минимизацию данных, если нельзя предсказать поведение системы? Еще одна общая проблема телемедицины касается передачи медицинских данных за рубеж. В момент консультации пациент и доктор могут находиться в разных странах, а значит, в разных юрисдикциях.

Эта проблема уникальна для телемедицины, поскольку область медицинских консультаций строго регулируется в каждой стране мира, где существуют различные нормы в отношении согласия, прозрачности, стандартов качества, ответственности, страхования, защиты данных, безопасности, установления личности, договорных отношений, оплаты, профессиональной этики и т. д. Каждый из этих элементов может стать предметом судебного разбирательства или официального расследования, создавая правовые коллизии. Некоторые из этих аспектов регулируются договорами, но большинство регулируются напрямую международным правом и не могут быть изменены по договору. Если возникает спорная ситуация, то пациенту, который уже находится в невыгодном положении, будет крайне сложно добиться компенсации и правовой помощи. Даже если договорные отношения возможны, может быть не просто достичь соглашения относительно обязанностей и ответственности в условиях непредсказуемости системы.

Таким образом, уровень риска повышен в сравнении с очными консультациями, тогда как крайняя неопределенность в поведении системы осложняет регулирование отношений между пациентом и доктором. Кроме того, устройство, применяемое при оказании услуги, может пользоваться облачными решениями на серверах, находящихся за пределами ЕС.

Из-за характера используемых данных риски для пациента оказываются выше, чем в других областях. Хотя и после обычной очной консультации врач может поместить данные пациента в зарубежные облачные хранилища, разница с непрерывно обучающимися устройствами состоит в том, что информация может изменяться и обновляться в режиме реального времени, тем самым предоставляя актуальные и истинные данные потенциальным хакерам, властям зарубежных стран, производителям системы. Если устройство не собирает и не использует информацию в режиме реального времени, то пациентам проще контролировать свои данные, удалять их, отбирать данные для хранения и загрузки в модель, получать к ним доступ, как требует новый проект так называемого Закона о данных. Перед использованием таких систем необходима их тщательная оценка, возможно, с помощью инструмента «Оценка воздействия на защиту данных» (Data Protection Impact Assessment, DPIA). Кроме того, возможно, понадобится соответствующее соглашение в области защиты данных между медицинским учреждением и производителями телемедицинских устройств.

5. Статья 22 Общего регламента ЕС по защите персональных данных и контроль со стороны человека

Использование устройств на основе ИИ в телемедицине часто подпадает под определение автоматизированного принятия решений (АПР) и профилирования (ст. 22), например, в случае автоматического сканирования медицинского изображения при постановке диагноза. Наряду с принципом прозрачности, который гарантирует субъектам данных право на информирование о том, как будут использоваться их данные, и о последствиях их обработки, Общий регламент ЕС по защите персональных данных также гарантирует дополнительное право в отношении АПР и профилирования: право на объяснение. Это означает, что «контролирующая сторона обязана простым способом объяснить субъекту данных основания и критерии принятого решения, не пытаясь прибегать к сложным объяснениям применяемых алгоритмов или полному раскрытию алгоритма. При этом предоставляемая информация должна быть понятна субъекту данных»⁶.

В этом контексте модели, работающие по принципу «черного ящика», в особенности основанные на непрерывном обучении, не отвечают этому требованию, поскольку пациенту невозможно объяснить, почему модель пришла к определенному результату. В руководстве отмечается: «Сложность не является причиной непредоставления информации субъекту данных. Преамбула 58 предусматривает, что принцип прозрачности непосредственно относится к ситуациям, когда быстрое нарастание количества действующих лиц и технологической сложности деятельности

⁶ Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. <https://ec.europa.eu/newsroom/article29/items/612053/en>

затрудняет понимание субъектом данных факта сбора относящихся к нему данных и того, кем и с какой целью они собираются, как, например, в случае онлайн-рекламы⁷. Это особенно важно в контексте телемедицины, где речь идет о медицинских данных, а пациент не находится под непосредственным наблюдением врача. Последствия ошибки могут доходить до летальных исходов, и в этом одна из причин, по которым Еврокомиссия классифицировала медицинские устройства как «высокорисковые системы».

Сторона, контролирующая данные, например больница или производитель телемедицинского устройства, обязана предоставить информацию об обработке данных и о возможном влиянии АПР и профилирования на субъектов данных. Статья 77 Конвенции 108+ гласит: «Субъекты данных вправе знать основания для обработки данных, включая следствия из этих оснований, приводящие к любым полученным выводам, в частности, в случаях использования алгоритмов автоматизированного принятия решений и профилирования. Например, при выставлении кредитного рейтинга субъект имеет право знать логику обработки данных, приведшую к положительному или отрицательному решению, а не просто информацию о вынесенном решении. Понимание этих элементов способствует эффективному осуществлению иных существенных гарантий, таких как право на возражение и право на обращение к уполномоченным органам»⁸. Кроме того, субъекты данных имеют право выражать свое мнение об автоматизированном принятии решений и профилировании, право на принятие касающихся их решений человеком и право на оспаривание принятого решения. В случае телемедицины на основе ИИ может быть крайне сложно осуществлять управление процессом оказания медицинской помощи на расстоянии таким образом, чтобы врач мог контролировать каждое решение, принятое системой, и в то же время гарантировать, что решения принимает только врач, поскольку это сводит на нет все преимущества использования автоматизации. Даже если организовать постоянное присутствие врачей, в условиях непрерывного обучения они не смогут объяснить пациенту действия системы. Возможно, это достижимо в некоторых случаях, но в будущем, когда такие устройства получат широкое распространение и будет достигнута высокая степень интеграции телемедицины в систему здравоохранения на низовом уровне, тщательное наблюдение за указанными аспектами станет чрезвычайно важным. Еще одна большая проблема с системами непрерывного обучения состоит в сложности обеспечения права на оспаривание решения системы, т. е. «недостаток очевидных средств для оспаривания результатов, которые представляются неожиданными, вредными, несправедливыми или дискриминирующими» (Edwards & Veale, 2017). Оспорить решение системы не может не только пациент, но и врач, работающий удаленно, так как для этого нужно собрать все элементы, использованные системой при вынесении данного решения (например, изменения дозировки лекарства). Таким образом, предложение о «встроенной оспариваемости» (Almada, 2019) изначально неприменимо к таким моделям. В рамках этого сценария можно утверждать, что в интересах прозрачности и справедливости стандартом для телемедицины должны стать интерпретируемые модели ИИ, тогда как «черные ящики» должны использоваться

⁷ Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. <https://ec.europa.eu/newsroom/article29/items/612053>

⁸ Там же. Art. 29.

только в тех случаях, когда медицинский работник может тщательно контролировать результаты, выдаваемые системой, и выносить общую клиническую оценку на основе дополнительных по отношению к системе ИИ элементов. Кроме того, системы непрерывного обучения должны быть ограничены лишь теми решениями, которые не могут нанести вред пациенту в случае ошибок, и должны постоянно проверяться врачом персоналом.

6. Информированное согласие

Когда правовой основой для обработки медицинских данных пациента является его согласие, как это часто бывает в случае телемедицинских устройств, то, согласно Общему регламенту ЕС по защите персональных данных, такое согласие должно быть явным, информированным, свободно выраженным, конкретным и недвусмысленным. Помимо обычных элементов, которые сообщаются субъекту данных согласно действующему законодательству, должна быть предоставлена также дополнительная информация о порядке проведения виртуальной консультации. Сюда включается информация о правах в рамках законов о защите данных, о возможных ошибках системы, о протоколах связи во время телеконсультаций, о правилах назначения лекарств и о координации работы с другими медицинскими работниками (Membrado, 2021). Дополнительные требования об обеспечении прозрачности содержатся в Законе об ИИ: в ст. 13 перечислены данные, которые должны быть предоставлены пользователям при инструктировании по поводу работы системы ИИ, а именно: высокорисковые системы ИИ должны сопровождаться инструкцией по использованию в соответствующем цифровом или ином формате, включая сжатую, полную, корректную и четкую информацию, релевантную, доступную и понятную пользователям. Статьи 11 и 18 содержат требование о подробной технической документации, которую следует обновлять перед выпуском системы. В руководствах по автоматизированному принятию решений отмечается, что если сторона, контролирующая данные, использует согласия как основу для профилирования, то она должна продемонстрировать, что субъект данных понимает, на что именно он дает свое согласие, и помнить, что согласие не всегда служит достаточным основанием для обработки данных. Во всех случаях субъекты данных должны иметь достаточный объем релевантной информации о предполагаемом использовании своих данных и последствиях их обработки, чтобы выраженное согласие представляло собой информированный выбор. Можно утверждать, что при использовании непрерывного обучения это требование изначально невыполнимо, так как даже врач или разработчик модели не может предсказать последствия использования непредсказуемой модели. Информация, полученная пациентом, должна быть сжатой, прозрачной, понятной, доступной, изложенной простым языком, учитывающей такие особенности, как возраст, умственные способности и уровень образования субъекта данных. В любом случае автоматизированное принятие решений и профилирование с использованием особых категорий данных, таких как медицинские данные, возможно лишь при наличии явного согласия субъекта данных или если это предусмотрено законом в существенных интересах общества, при условии наличия адекватных защитных механизмов. Интересы общества не являются достаточным законным основанием, их важность должна быть подтверждена. Например, можно утверждать, что борьба с инфекцией Covid-19 является «существенным интересом общества».

Информированное согласие как с правовой, так и с этической точек зрения становится центральным элементом систем ИИ, применяемых в телемедицине.

7. Уязвимые группы

Актуальная проблема телемедицины с использованием систем искусственного интеллекта вызывает еще большую озабоченность, когда касается особенно уязвимых субъектов, таких как онкологические пациенты, пациенты с когнитивными нарушениями, например, пожилые люди, страдающие старческой деменцией или болезнью Альцгеймера, дети, нейроотличные пациенты (Shaw et al., 2022), люди, не владеющие языком, на котором говорит доктор. В первую очередь необходимо дать им возможность выразить свое информированное, свободное, однозначное и конкретное согласие (ст. 7 Общего регламента ЕС по защите персональных данных). Отношения между врачом и пациентом сами по себе не являются сбалансированными, так как стороны в них находятся на разных уровнях, а позиция пациента уязвима; поэтому может быть непросто получить согласие, отвечающее всем требованиям закона, от такого уязвимого человека, как, например, онкологический пациент, уже подавленный своей болезнью.

Если прибавить сюда право пациента на объяснение, то становится ясно, что возникает сразу несколько сложных проблем: в условиях телемедицины доктор находится вдали от пациента, человеческий контакт недостаточен. В таких условиях объяснения могут оказаться нечеткими или непонятными, поскольку пациент не может распознать знаки невербальной коммуникации.

При использовании систем искусственного интеллекта, работа которых часто непонятна и специалистам, даже когда эта система обладает свойствами объяснимости и интерпретируемости, риски непонимания еще более возрастают и становятся труднопреодолимыми, если только консультации не проводятся также в очном формате. Если действия системы непредсказуемы, то уязвимому пациенту еще сложнее объяснить последствия применения устройства.

Нельзя забывать и о явлении цифрового неравенства, т. е. о том, что не все пациенты обладают одинаковым уровнем цифровой грамотности. Это обстоятельство приобретает особую значимость, когда услуги телемедицины оказываются в государственном секторе, так как именно там высока вероятность дискриминации между теми, кто способен воспользоваться этими услугами, и теми, кто по разным причинам не способен этого сделать. Этот аспект подчеркивался в заявлении Еврокомиссии, которое было основано, в частности, на докладе ОЭСР «Краткий обзор вопросов здравоохранения: Европа – 2018», где говорилось об этой проблеме.

Комиссия отметила наличие прямой зависимости между уровнем образования и числом запросов на поиск медицинской информации в Сети; фактически вполне вероятны подобные несоответствия в использовании цифровых решений для улучшения здоровья и предотвращения заболеваемости, а значит, риск того, что цифровые инструменты: приложения, носимые устройства, онлайн-форумы – не принесут пользы тем, кто в них более всего нуждается, что потенциально усилит неравенство в области охраны здоровья (Oliveira, 2020). Этот разрыв может даже увеличиться из-за систем непрерывного обучения, так как лица, лучше владеющие технологиями телемедицины, будут получать более точные результаты. Этот риск должен учитываться специалистами в области здравоохранения при оказании телемедицинских услуг.

8. Баланс между охраной неприкосновенности частной жизни и защитой от вреда при дистанционной работе

Как обсуждалось выше, одна из важнейших проблем с системами непрерывного обучения в телемедицине состоит в том, что модель «развивается», когда пациент находится далеко от доктора (от специалистов технической поддержки), но при этом системы обычно разрабатываются так, чтобы быть как можно менее инвазивными. Это приводит к необходимости балансировать между постоянным наблюдением за пациентами с целью обеспечения их безопасности и охраной их частной жизни, чтобы применение технологий не доставляло им неудобств. Таким образом, релевантными оказываются четыре аспекта: технические меры, организационные меры, психологические факторы и требования закона. С технической точки зрения изучались несколько методов решения проблем неприкосновенности частной жизни, которые обсуждались выше, например, проблема идентификации личности с использованием блокчейна, шифрования, федеративного обучения с помощью концепции периферийных вычислений (Ahmad, 2021; Jain et al., 2022; Wang, 2021, Ma et al., 2020; Wang, 2022).

Однако некоторые проблемы невозможно решить лишь с помощью технологий: организационные меры играют решающую роль в достижении баланса между охраной частной жизни и преимуществами технологии наблюдения как меры противодействия неожиданным сбоям в системе.

Фактически в качестве первоочередной меры для снижения ущерба от сбоя системы, можно запланировать необходимые поддерживающие вмешательства со стороны технических специалистов и медиков, которые могут тестировать систему, оценивать ее работу и состояние здоровья пациента. Такие действия можно проводить с минимальным уровнем беспокойства для пациентов (например, совмещая проверки с регулярными очными приемами в медицинском учреждении). Однако вторая мера не менее значима: это обучение врачей и пациентов корректному использованию технологий телемедицины, что чрезвычайно важно как для сохранения неприкосновенности частной жизни, так и для снижения риска сбоя системы и ее некорректного использования. При условии качественной подготовки врачи смогут, например, распознать такие особенности пациентов, которые могут привести самообучающуюся систему к генерации ошибочных выводов.

С психологической точки зрения уязвимые пациенты могут нуждаться в дополнительной поддержке в областях: понимания работы системы, чтобы корректно ее использовать и свободно выражать свое информированное согласие; принятия системы как части своей жизни без чувства нарушения повседневной деятельности; обучения выражению своего беспокойства и неудобства (включая физическую боль) на расстоянии (Yakar, 2021).

Обученный терапевт, работая совместно с техническими специалистами и профильными медиками, может помочь пациентам справляться с этими проблемами. С другой стороны, врачи должны удостовериться, что пациенты понимают их указания, и регулярно проверять, что даваемое пациентами согласие действительно является информированным. Нормы о неприкосновенности частной жизни не препятствуют тому, чтобы задействовать вышеописанные меры, так как Общий регламент ЕС по защите персональных данных предусматривает соответствующие способы защиты фундаментальных прав пациентов. В качестве дополнительной организационной меры больницы должны иметь специалиста по вопросам охраны частной жизни, который следил бы за использованием телемедицинских устройств и консультировал всех заинтересованных лиц.

9. Меры дополнительной защиты в форме аудита искусственного интеллекта

В последние годы все большую популярность приобретает научная область алгоритмического аудита в рамках более широкой проблемы аудита ИИ, особенно в отношении машинного обучения. Аудит ИИ предполагает встраивание вопросов этики, прав человека и законности во все стадии жизненного цикла ИИ от разработки до послепродажного функционирования (LaBrie & Steinke, 2019; Mökander & Floridi, 2022; Mantelero & Esposito, 2021; Koshiyama et al., 2021; Mökander, 2022; Floridi, 2022) с одновременным контролем его технических характеристик (а именно безопасности, эффективности, корректности методов обработки данных). С этой целью был разработан ряд практических инструментов, таких как дискуссии по вопросам этики (например, дискуссии по открытому ИИ, по этике данных, по этике ИИ (Kalra, 2020)). Такая форма оценки и непрерывного мониторинга важна для выявления предубеждений, технических и статистических ошибок, дефектов в системе безопасности, неблагоприятных эффектов на послепродажном этапе. Фактически благодаря пристальному вниманию к процессу в целом и к механизму обеспечения соответствия, применяемому с самых первых этапов развития ИИ (т. е. еще до этапа сбора данных) оказалось возможным предотвратить целый ряд ошибок, которые могли привести к нарушению принципа неприкосновенности частной жизни, сбоям в системе, дефе там в системе безопасности, дискриминации. В связи с этим рекомендуется применять процедуры аудита ИИ каждый раз, когда в клинической практике задействованы модели непрерывного обучения.

Заключение

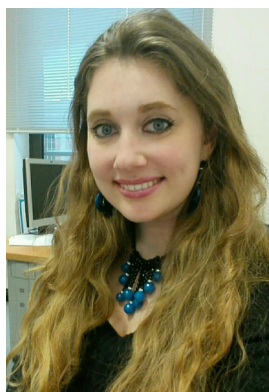
В этой краткой работе мы постарались показать, как телемедицина с помощью интеллектуальных систем непрерывного обучения становится полезным инструментом для усовершенствования медицинского обслуживания пациента, однако ее распространение может также поставить новые вопросы перед юристами и медиками. Новые правовые проблемы, такие как регулирование систем, адаптирующихся к потребностям пациента, или защита уязвимых субъектов, их способность понять работу ИИ для свободного выражения своего согласия, должны в ближайшее время быть рассмотрены и решены в рамках правовой доктрины. Медицинские учреждения, планирующие использовать системы ИИ для оказания телемедицинских услуг, должны обучить персонал и пациентов безопасному обращению с системой, чтобы избежать утечки данных и некорректного использования устройств. Эффективно достичь этой цели можно с помощью сплоченного коллектива технических специалистов, профильных врачей и обученных врачей общей практики. Хотя в Евросоюзе предпринимались попытки давать правовые ответы на вопросы, связанные с развитием технологий ИИ, остается еще много нерешенных проблем. Поэтому стоит надеяться, что данная область будет тщательно изучена, прежде чем телемедицина с использованием ИИ получит широкое распространение и войдет в повседневную практику на территории Евросоюза.

Список литературы

- Ahmad, R. W. (2021). The role of blockchain technology in telehealth and telemedicine. *International Journal of Medical Informatics*, 148, 104399. <https://doi.org/10.1016/j.ijmedinf.2021.104399>
- Almada, M. (2019). Human intervention in automated decision-making: Toward the construction of contestable systems. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law* (pp. 2–11). <https://doi.org/10.2139/ssrn.3264189>
- Botrugno, C. (2014). Un diritto per la telemedicina: analisi di un complesso normativo in formazione. *Politica del Diritto*, 4(45), 639–668. <https://doi.org/10.1437/78949>
- Burrai, F., Gambella, M., & Scarpa, A. (2021). L'erogazione diprestazioni sanitarie in telemedicina. *Giornale di Clinica Nefrologica e Dialisi*, 33, 3–6.
- Campagna, M. (2020). Linee guida per la Telemedicina: considerazioni alla luce dell'emergenza Covid-19. *Corti Supreme e Salute*, 3, 11–25.
- Castagno, S., & Khalifa, M. (2020). Perceptions of artificial intelligence among healthcare staff: a qualitative survey study. *Frontiers in artificial intelligence*, 2(5), 84–92. <https://doi.org/10.3389/frai.2020.578983>
- Davis, E. (2016). AI Amusements: The Tragic Tale of Tay the Chatbot. *AI Matters*, 2(4), 20–24. <https://doi.org/10.1145/3008665.3008674>
- Edwards, L., & Veale, M. (2017). Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for. *Duke L. & Tech. Rev.*, 16, 18–26.
- Floridi, L. (2022). capAI-A Procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4064091>
- Gallese, Ch. (2022). Suggestions for a revision of the European smart robot liability regime. In *Proceedings of the 4th European Conference on the Impact of Artificial Intelligence and Robotics (ECIAIR 2022)*. <https://doi.org/10.34190/eciair.4.1.851>
- Gioulekas, F. (2022). A Cybersecurity Culture Survey Targeting Healthcare Critical Infrastructures. *Healthcare*, 10, 327–333. <https://doi.org/10.3390/healthcare10020327>
- Giunti, G. (2014). The Use of a Gamified Platform To Empower And Increase Patient Engagement in Diabetes Mellitus Adolescents. In *American Medical Informatics Association Annual Symposium*.
- Jain, N., Gupta V., & Dass, P. (2022). Blockchain: A novel paradigm for secured data transmission in telemedicine. In *Wearable Telemedicine Technology for the Healthcare Industry* (pp. 33–52).
- Kalra, A. (2020). *Artificial Intelligence Ethics Canvas: A Tool for Ethical and Socially Responsible AI*.
- Koshiyama, A. S., Kazim, E., Treleaven, P. C., Rai, P., Szpruch, L., Pavey, G., Ahamat, G., Leutner, F., Goebel, R., Knight, A., Adams, J., Hitrova, C., Barnett, J., Nachev, P., Barber, D., Chamorro-Premuzic, T., Klemmer, K., Gregorovic, M., Khan, S. A., & Lomas, E. (2021). Towards Algorithm Auditing: A Survey on Managing Legal, Ethical and Technological Risks of AI, ML and Associated Algorithms. *Software Engineering eJournal*. <https://doi.org/10.2139/ssrn.3778998>
- LaBrie, R., & Steinke, G. (2019). Towards a framework for ethical audits of AI algorithms. In *Twenty-fifth Americas Conference on Information Systems*.
- Lakkaraju, H. (2019). Faithful and customizable explanations of black box models. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 131–138). <https://doi.org/10.1145/3306618.3314229>
- Ma, M., Shuqin, F., & Feng, D. (2020). Multi-user certificateless public key encryption with conjunctive keyword search for cloud-based telemedicine. *Journal of Information Security and Applications*, 55, 102652. <https://doi.org/10.1016/j.jisa.2020.102652>
- Mantelero, A., & Esposito, S. (2021). An evidence-based methodology for human rights impact assessment (HRIA) in the development of AI data-intensive systems. *Computer Law & Security Review*, 41, 105561. <https://doi.org/10.1016/j.clsr.2021.105561>
- Marchant, G., & Lindor, R. (2012). The coming collision between autonomous vehicles and the liability system. *Santa Clara Law Review*, 52, 1321–1340.
- Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and information technology*, 6, 175–183. <https://doi.org/10.1007/s10676-004-3422-1>
- Membrado, C. G. (2021). Telemedicina, ética y derecho en tiempos de COVID-19. Una mirada hacia el futuro. *Revista Clinica Espanola*, 221, 408–410. <https://doi.org/10.1016/j.rce.2021.03.002>
- Mi, F. (2020). Generalized Class Incremental Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 240–241).
- Mökander, J. (2022). Conformity assessments and post-market monitoring: a guide to the role of auditing in the proposed European AI regulation. *Minds and Machines*, 32, 241–268. <https://doi.org/10.1007/s11023-021-09577-4>

- Mökander, J., & Floridi, L. (2022). Operationalising AI governance through ethics-based auditing: an industry case study. *AI and Ethics*, 6, 1–18. <https://doi.org/10.1007/s43681-022-00171-7>
- Oliveira, T. (2020). Bringing health care to the patient: An overview of the use of telemedicine in OECD countries. *OECD, Directorate for Employment, Labour and Social Affairs, Health Committee*.
- Pacis, D., Mitch, M., Edwin, D. C., Subido, Jr., & Bugtai, N. (2018). Trends in telemedicine utilizing artificial intelligence. In *AIP conference proceedings*. AIP Publishing LLC.
- Parisi, G. (2019). Continual lifelong learning with neural networks: A review, *Neural Networks*, 113, 54–71. <https://doi.org/10.1016/j.neunet.2019.01.012>
- Rodrigues, R. (2020). Legal and human rights issues of AI: Gaps, challenges and vulnerabilities. *Journal of Responsible Technology*, 4, 100005. <https://doi.org/10.1016/j.jrt.2020.100005>
- Schatten, M., & Protrka, R. (2021). Conceptual Architecture of a Cognitive Agent for Telemedicine based on Gamification. In *Central European Conference on Information and Intelligent Systems* (pp. 3–10).
- Scheetz, J. (2021). A survey of clinicians on the use of artificial intelligence in ophthalmology, dermatology, radiology and radiation oncology. *Scientific reports*, 11.1, 1–10.
- Shaw, S., Davis, L-J., & Doherty, M. (2022). *Considering autistic patients in the era of telemedicine: the need for an adaptable, equitable, and compassionate approach*, *BJGP open* 6.1.
- Strehle, E. M., & Shabde, N. (2006). One hundred years of telemedicine: does this new technology have a place in paediatrics? *Archives of disease in childhood*, 91.12, 956–959. <https://doi.org/10.1136/adc.2006.099622>
- Tigard, D. (2020). There is no techno-responsibility gap. *Philosophy & Technology*, 1–19.
- Wang, R. (2022). Privacy-Preserving Federated Learning for Internet of Medical Things under Edge Computing. *IEEE Journal of Biomedical and Health Informatics*.
- Wang, W. (2021). A privacy protection scheme for telemedicine diagnosis based on double blockchain. *Journal of Information Security and Applications*, 61, 102845. <https://doi.org/10.1016/j.jisa.2021.102845>
- Yakar, D. (2021). Do People Favor Artificial Intelligence Over Physicians? A Survey Among the General Population and Their View on Artificial Intelligence in Medicine. *Value in Health*, 3, 12–23. <https://doi.org/10.1016/j.jval.2021.09.004>
- Ye, J. (2020). The role of health technology and informatics in a global public health emergency: practices and implications from the COVID-19 pandemic. *JMIR medical informatics*, 8.7, e19866. <https://doi.org/10.2196/19866>

Сведения об авторе



Галлезе-Нобиле Кьяра – доктор наук, научный сотрудник (постдок) по управлению исследовательскими данными, Эйндховенский технологический университет (Эйндховен, Королевство Нидерландов); научный сотрудник (постдок) департамента математики и наук о земле, Университет Триеста (Триест, Итальянская Республика)

Адрес: а/я 513 5600 МБ Эйндховен, Королевство Нидерландов

E-mail: cgallese@liuc.it

Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=57222726276>

Web of Science Researcher ID:

<https://www.webofscience.com/wos/author/record/AGE-9594-2022>

ORCID ID: <https://orcid.org/0000-0001-8194-0261>

Google Scholar ID: <https://scholar.google.com/citations?user=Vmoen8UAAAAJ>

Конфликт интересов

Автор является международным редактором журнала, статья прошла рецензирование на общих основаниях.

Финансирование

Исследование не имело спонсорской поддержки.

Тематические рубрики

Рубрика OECD: 5.05 / Law

Рубрика ASJC: 3308 / Law

Рубрика WoS: OM / Law

Рубрика ГРНТИ: 10.27.91 / Гражданское право отдельных стран

Специальность ВАК: 5.1.3 / Частно-правовые (цивилистические) науки

История статьи

Дата поступления – 4 мая 2023 г.

Дата одобрения после рецензирования – 20 мая 2023 г.

Дата принятия к опубликованию – 16 июня 2023 г.

Дата онлайн-размещения – 20 июня 2023 г.



Research article

DOI: <https://doi.org/10.21202/jdtl.2023.13>



Legal Aspects of the Use Artificial Intelligence in Telemedicine

Chiara Gallese Nobile 

Eindhoven University of Technology
Eindhoven, Netherlands;
University of Trieste
Trieste, Italy

Keywords

Artificial intelligence,
data protection,
digital inequality,
digital technologies,
law,
legislation,
personal data,
private life,
regulation,
telemedicine

Abstract

Objective: the rapid expansion of the use of telemedicine in clinical practice and the increasing use of Artificial Intelligence has raised many privacy issues and concerns among legal scholars. Due to the sensitive nature of the data involved particular attention should be paid to the legal aspects of those systems. This article aimed to explore the legal implication of the use of Artificial Intelligence in the field of telemedicine, especially when continuous learning and automated decision-making systems are involved; in fact, providing personalized medicine through continuous learning systems may represent an additional risk. Particular attention is paid to vulnerable groups, such as children, the elderly, and severely ill patients, due to both the digital divide and the difficulty of expressing free consent.

Methods: comparative and formal legal methods allowed to analyze current regulation of the Artificial Intelligence and set up its correlations with the regulation on telemedicine, GDPR and others.

Results: legal implications of the use of Artificial Intelligence in telemedicine, especially when continuous learning and automated decision-making systems are involved were explored; author concluded that providing personalized medicine through continuous learning systems may represent an additional risk and offered the ways to minimize it. Author also focused on the issues of informed consent of vulnerable groups (children, elderly, severely ill patients).

© Gallese Nobile C., 2023



This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Scientific novelty: existing risks and issues that are arising from the use of Artificial Intelligence in telemedicine with particular attention to continuous learning systems are explored.

Practical significance: results achieved in this paper can be used for lawmaking process in the sphere of use of Artificial Intelligence in telemedicine and as base for future research in this area as well as contribute to limited literature on the topic.

For citation

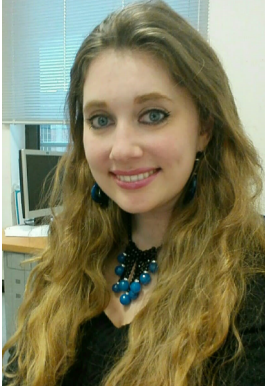
Gallese Nobile, C. (2023). Legal Aspects of the Use Artificial Intelligence in Telemedicine. *Journal of Digital Technologies and Law*, 1(2), 314–336. <https://doi.org/10.21202/jdtl.2023.13>

References

- Ahmad, R. W. (2021). The role of blockchain technology in telehealth and telemedicine. *International Journal of Medical Informatics*, 148, 104399. <https://doi.org/10.1016/j.ijmedinf.2021.104399>
- Almada, M. (2019). Human intervention in automated decision-making: Toward the construction of contestable systems. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law* (pp. 2–11). <https://doi.org/10.2139/ssrn.3264189>
- Botrugno, C. (2014). Un diritto per la telemedicina: analisi di un complesso normativo in formazione. *Politica del Diritto*, 4(45), 639–668. <https://doi.org/10.1437/78949>
- Burrai, F., Gambella, M., & Scarpa, A. (2021). L'erogazione di prestazioni sanitarie in telemedicina. *Giornale di Clinica Nefrologica e Dialisi*, 33, 3–6.
- Campagna, M. (2020). Linee guida per la Telemedicina: considerazioni alla luce dell'emergenza Covid-19. *Corti Supreme e Salute*, 3, 11–25.
- Castagno, S., & Khalifa, M. (2020). Perceptions of artificial intelligence among healthcare staff: a qualitative survey study. *Frontiers in artificial intelligence*, 2(5), 84–92. <https://doi.org/10.3389/frai.2020.578983>
- Davis, E. (2016). AI Amusements: The Tragic Tale of Tay the Chatbot. *AI Matters*, 2(4), 20–24. <https://doi.org/10.1145/3008665.3008674>
- Edwards, L., & Veale, M. (2017). Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for. *Duke L. & Tech. Rev.*, 16, 18–26.
- Floridi, L. (2022). capAI-A Procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4064091>
- Gallese, Ch. (2022). Suggestions for a revision of the European smart robot liability regime. In *Proceedings of the 4th European Conference on the Impact of Artificial Intelligence and Robotics (ECIAIR 2022)*. <https://doi.org/10.34190/eciair.4.1.851>
- Gioulekas, F. (2022). A Cybersecurity Culture Survey Targeting Healthcare Critical Infrastructures. *Healthcare*, 10, 327–333. <https://doi.org/10.3390/healthcare10020327>
- Giunti, G. (2014). The Use of a Gamified Platform To Empower And Increase Patient Engagement in Diabetes Mellitus Adolescents. In *American Medical Informatics Association Annual Symposium*.
- Jain, N., Gupta V., & Dass, P. (2022). Blockchain: A novel paradigm for secured data transmission in telemedicine. In *Wearable Telemedicine Technology for the Healthcare Industry* (pp. 33–52).
- Kalra, A. (2020). *Artificial Intelligence Ethics Canvas: A Tool for Ethical and Socially Responsible AI*.
- Koshiyama, A. S., Kazim, E., Treleaven, P. C., Rai, P., Szpruch, L., Pavey, G., Ahamat, G., Leutner, F., Goebel, R., Knight, A., Adams, J., Hitrova, C., Barnett, J., Nachev, P., Barber, D., Chamorro-Premuzic, T., Klemmer, K., Gregorovic, M., Khan, S. A., & Lomas, E. (2021). Towards Algorithm Auditing: A Survey on Managing Legal, Ethical and Technological Risks of AI, ML and Associated Algorithms. *Software Engineering eJournal*. <https://doi.org/10.2139/ssrn.3778998>

- LaBrie, R., & Steinke, G. (2019). Towards a framework for ethical audits of AI algorithms. In *Twenty-fifth Americas Conference on Information Systems*.
- Lakkaraju, H. (2019). Faithful and customizable explanations of black box models. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 131–138). <https://doi.org/10.1145/3306618.3314229>
- Ma, M., Shuqin, F., & Feng, D. (2020). Multi-user certificateless public key encryption with conjunctive keyword search for cloud-based telemedicine. *Journal of Information Security and Applications*, 55, 102652. <https://doi.org/10.1016/j.jisa.2020.102652>
- Mantelero, A., & Esposito, S. (2021). An evidence-based methodology for human rights impact assessment (HRIA) in the development of AI data-intensive systems. *Computer Law & Security Review*, 41, 105561. <https://doi.org/10.1016/j.clsr.2021.105561>
- Marchant, G., & Lindor, R. (2012). The coming collision between autonomous vehicles and the liability system. *Santa Clara Law Review*, 52, 1321–1340.
- Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and information technology*, 6, 175–183. <https://doi.org/10.1007/s10676-004-3422-1>
- Membrado, C. G. (2021). Telemedicina, ética y derecho en tiempos de COVID-19. Una mirada hacia el futuro. *Revista Clínica Española*, 221, 408–410. <https://doi.org/10.1016/j.rce.2021.03.002>
- Mi, F. (2020). Generalized Class Incremental Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 240–241).
- Mökander, J. (2022). Conformity assessments and post-market monitoring: a guide to the role of auditing in the proposed European AI regulation. *Minds and Machines*, 32, 241–268. <https://doi.org/10.1007/s11023-021-09577-4>
- Mökander, J., & Floridi, L. (2022). Operationalising AI governance through ethics-based auditing: an industry case study. *AI and Ethics*, 6, 1–18. <https://doi.org/10.1007/s43681-022-00171-7>
- Oliveira, T. (2020). Bringing health care to the patient: An overview of the use of telemedicine in OECD countries. *OECD, Directorate for Employment, Labour and Social Affairs, Health Committee*.
- Pacis, D., Mitch, M., Edwin, D. C., Subido, Jr., & Bugtai, N. (2018). Trends in telemedicine utilizing artificial intelligence. In *AIP conference proceedings*. AIP Publishing LLC.
- Parisi, G. (2019). Continual lifelong learning with neural networks: A review, *Neural Networks*, 113, 54–71. <https://doi.org/10.1016/j.neunet.2019.01.012>
- Rodrigues, R. (2020). Legal and human rights issues of AI: Gaps, challenges and vulnerabilities. *Journal of Responsible Technology*, 4, 100005. <https://doi.org/10.1016/j.jrt.2020.100005>
- Schatten, M., & Protrka, R. (2021). Conceptual Architecture of a Cognitive Agent for Telemedicine based on Gamification. In *Central European Conference on Information and Intelligent Systems* (pp. 3–10).
- Scheetz, J. (2021). A survey of clinicians on the use of artificial intelligence in ophthalmology, dermatology, radiology and radiation oncology. *Scientific reports*, 11, 1, 1–10.
- Shaw, S., Davis, L-J., & Doherty, M. (2022). *Considering autistic patients in the era of telemedicine: the need for an adaptable, equitable, and compassionate approach*, *BJGP open* 6.1.
- Strehle, E. M., & Shabde, N. (2006). One hundred years of telemedicine: does this new technology have a place in paediatrics? *Archives of disease in childhood*, 91, 12, 956–959. <https://doi.org/10.1136/adc.2006.099622>
- Tigard, D. (2020). There is no techno-responsibility gap. *Philosophy & Technology*, 1–19.
- Wang, R. (2022). Privacy-Preserving Federated Learning for Internet of Medical Things under Edge Computing. *IEEE Journal of Biomedical and Health Informatics*.
- Wang, W. (2021). A privacy protection scheme for telemedicine diagnosis based on double blockchain. *Journal of Information Security and Applications*, 61, 102845. <https://doi.org/10.1016/j.jisa.2021.102845>
- Yakar, D. (2021). Do People Favor Artificial Intelligence Over Physicians? A Survey Among the General Population and Their View on Artificial Intelligence in Medicine. *Value in Health*, 3, 12–23. <https://doi.org/10.1016/j.jval.2021.09.004>
- Ye, J. (2020). The role of health technology and informatics in a global public health emergency: practices and implications from the COVID-19 pandemic. *JMIR medical informatics*, 8, 7, e19866. <https://doi.org/10.2196/19866>

Author information



Chiara Gallese Nobile – PhD, Researcher (postdoc) of research data management, Eindhoven University of Technology (Eindhoven, Netherlands), Researcher (postdoc) of the Department of Mathematics and Geosciences, University of Trieste (Trieste, Italy).

Address: P/O 513 5600 MB Eindhoven, the Netherlands

E-mail: cgallese@liuc.it

Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=57222726276>

Web of Science Researcher ID:

<https://www.webofscience.com/wos/author/record/AGE-9594-2022>

ORCID ID: <https://orcid.org/0000-0001-8194-0261>

Google Scholar ID: <https://scholar.google.com/citations?user=Vmoen8UAAAAJ>

Conflict of interests

The author is an international editor of the Journal; the article has been reviewed on general terms.

Financial disclosure

The research had no sponsorship.

Thematic rubrics

OECD: 5.05 / Law

PASJC: 3308 / Law

WoS: OM / Law

Article history

Date of receipt – May 4, 2023

Date of approval – May 20, 2023

Date of acceptance – June 16, 2023

Date of online placement – June 20, 2023

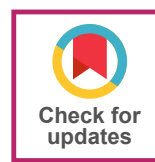


Научная статья

УДК 346.1:006.44:004.8

EDN: <https://elibrary.ru/dxnwhv>

DOI: <https://doi.org/10.21202/jdtl.2023.14>



Правовые средства обеспечения принципа прозрачности искусственного интеллекта

Юлия Сергеевна Харитоновa 

Московский государственный университет имени М. В. Ломоносова
г. Москва, Российская Федерация

Ключевые слова

Автоматизированная обработка данных, автономность, алгоритм, искусственный интеллект, право, принятие решений, прозрачность, цифровая экономика, цифровые технологии, этика

Аннотация

Цель: анализ действующих технологических и юридических теорий для определения содержания принципа прозрачности работы искусственного интеллекта с позиции правового регулирования, выбора применимых средств правового регулирования и установление объективных границ юридического вмешательства в технологическую сферу с помощью регулирующего воздействия.

Методы: методологическую основу исследования составляет совокупность общенаучных (анализ, синтез, индукция, дедукция) и специально-юридических (историко-правовой, формально-юридический, сравнительно-правовой) методов научного познания.

Результаты: подвергнуты критическому анализу нормы и предложения для нормативного оформления принципа прозрачности искусственного интеллекта с точки зрения невозможности получения полной технологической прозрачности искусственного интеллекта. Выдвинуто предложение обсудить варианты политики управления алгоритмической прозрачностью и подотчетностью на основе анализа социальных, технических и регулятивных проблем, создаваемых алгоритмическими системами искусственного интеллекта. Обосновано, что прозрачность является необходимым условием для признания искусственного интеллекта заслуживающим доверия. Обосновано, что прозрачность и объяснимость технологии искусственного интеллекта важна не только для защиты персональных данных, но и в иных ситуациях автоматизированной обработки данных, когда для принятия решений недостающие из входящей информации технологические данные восполняются из открытых источников, в том числе не имеющих значения хранилищ персональных данных. Предложено законодательно закрепить обязательный аудит и ввести стандарт, закрепляющий

© Харитоновa Ю. С., 2023



Статья находится в открытом доступе и распространяется в соответствии с лицензией Creative Commons «Attribution» («Атрибуция») 4.0 Всемирная (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0/deed.ru>), позволяющей неограниченно использовать, распространять и воспроизводить материал при условии, что оригинальная работа упомянута с соблюдением правил цитирования.

компромисс между возможностями и преимуществами технологии, точностью и объяснимостью результата ее работы и правами участников общественных отношений. Введение сертификации моделей искусственного интеллекта, обязательных к применению, позволит решить вопросы ответственности обязанных применять такие системы субъектов. В контексте вопроса о профессиональной ответственности профессиональных субъектов, таких как врачи, военные, органы корпоративного управления юридического лица, требуется ограничить обязательное применение искусственного интеллекта в случаях, если не обеспечена его достаточная прозрачность.

Научная новизна: междисциплинарный характер исследования позволил выявить невозможность и необоснованность требований полного открытия исходного кода или архитектуры моделей искусственного интеллекта. Принцип прозрачности искусственного интеллекта может быть обеспечен за счет проработки и обеспечения права субъекта данных и субъекта, которому адресовано решение, принятое в результате автоматизированной обработки данных, на отказ от применения автоматизированной обработки данных для принятия решений и права на возражения против принятых таким способом решений.

Практическая значимость: обусловлена отсутствием в настоящее время достаточного регулирования принципа прозрачности искусственного интеллекта и результатов его работы, а также содержания и особенностей реализации права на объяснение и права на возражение субъекта решения. Наиболее плодотворный путь для установления доверия к искусственному интеллекту заключается в том, чтобы признать данную технологию частью сложной социотехнической системы, которая опосредует доверие, и повышать надежность этих систем. Основные положения и выводы исследования могут быть использованы для совершенствования правового механизма обеспечения прозрачности моделей искусственного интеллекта, применяемых в государственном управлении и бизнесе.

Для цитирования

Харитонов, Ю. С. (2023). Правовые средства обеспечения принципа прозрачности искусственного интеллекта. *Journal of Digital Technologies and Law*, 1(2), 337–358. <https://doi.org/10.21202/jdtl.2023.14>

Содержание

Введение

1. Понятие «черный ящик» и его значение для юридического оформления применения технологии искусственного интеллекта для принятия решений
2. Правовые и этические риски применения непрозрачной технологии
3. Автоматизированная система обработки данных и их качество
4. Открытость алгоритмов и результатов их работы
5. Применимые правовые средства для предотвращения проблем непрозрачности правовых решений: возразить или отказаться

Выводы

Список литературы

Введение

В российском праве сформулированы принципы Национальной стратегии развития искусственного интеллекта на период до 2030 г., которые включают в том числе прозрачность как объяснимость работы искусственного интеллекта и процесса достижения им результатов, недискриминационный доступ пользователей продуктов, которые созданы с использованием технологий искусственного интеллекта, к информации о применяемых в этих продуктах алгоритмах работы искусственного интеллекта (п. 19 Стратегии, утв. Указом Президента РФ «О развитии искусственного интеллекта в Российской Федерации» № 490 от 10.10.2019). Понятия «объяснимость» и «недискриминационность» работы искусственного интеллекта выделены как составляющие принципа прозрачности.

Раскрытие информации также предусматривается в международных актах и национальном законодательстве разных стран. Эти правила в первую очередь наиболее подробно затрагивают вопросы защиты прав и свобод человека, как, например, в Общем регламенте по защите данных Европейского союза (General Data Protection Regulation, GDPR; Постановление (Европейский Союз) 2016/679¹).

В России защита прав и свобод человека и безопасность работы искусственного интеллекта обозначены в Стратегии 2030 как самостоятельные принципы, хотя названные принципы тесно соприкасаются с прозрачностью. Как представляется, недискриминационность ведет к обеспечению защиты гарантированных российским и международным законодательством прав и свобод человека. Принцип безопасности работы искусственного интеллекта определен как недопустимость использования искусственного интеллекта в целях умышленного причинения вреда гражданам и юридическим лицам, а также как предупреждение и минимизация рисков возникновения негативных последствий использования технологий искусственного интеллекта. Представляется, что принцип прозрачности также позволяет добиться безопасности применения искусственного интеллекта.

В отсутствие ясного видения с точки зрения права содержания принципа прозрачности работы искусственного интеллекта нам представляется важным определить понятие прозрачности, а также исследовать допустимые границы юридического вмешательства в технологическую сферу с помощью регулирующего воздействия.

Прозрачность в контексте работы искусственного интеллекта может быть рассмотрена с позиции технологии, этики и права. Междисциплинарный подход позволяет критически взглянуть на нормы и предложения для нормативного оформления с точки зрения невозможности получения полной технологической прозрачности искусственного интеллекта. Требуется обсудить варианты политики управления алгоритмической прозрачностью и подотчетностью на основе анализа социальных, технических и регулятивных проблем, создаваемых алгоритмическими системами искусственного интеллекта.

¹ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). <http://eur-lex.europa.eu/>

1. Понятие «черный ящик» и его значение для юридического оформления применения технологии искусственного интеллекта для принятия решений

В течение десятилетий проекты искусственного интеллекта опирались на человеческий опыт, накопленный инженерами по знаниям, и были как явно разработанными, так и легко понимаемыми людьми. Значительный прогресс в области искусственного интеллекта был достигнут с использованием контролируемых систем обучения, которые предназначены повторять решения людей (Hastie et al., 2009; LeCun et al., 2015; Pak & Kim, 2017). Например, экспертные системы, основанные на деревьях решений, являются совершенными моделями человеческого принятия решений и поэтому естественным образом понятны как разработчикам, так и конечным пользователям (Lawrence & Wright, 2001; Cho et al., 2002). То же можно сказать и о таблицах данных (Cragun & Steudel, 1987). Однако со сменой парадигмы в ведущих методологиях искусственного интеллекта на системы машинного обучения, основанные на глубоких нейронных сетях (DNN), появились новшества (Samek et al., 2021).

Простота понимания была принесена в жертву скорости принятия решения и получила название черного ящика – непрозрачного для человеческого понимания, но чрезвычайно способного как в отношении результатов, так и в отношении обучения новым областям. Модели, которые «открывают черный ящик», делая нелинейный и сложный процесс принятия решений понятным для людей-наблюдателей, являются многообещающими решениями проблемы черного ящика в ИИ, но ограничены, по крайней мере, в их нынешнем состоянии, в их способности сделать эти процессы менее непрозрачными для большинства наблюдателей. Искусственный интеллект использует глубокое обучение (DL), алгоритмическую систему глубоких нейронных сетей (algorithmic system of deep neural networks), которые в целом остаются непрозрачными или скрытыми от человеческого понимания.

В чем проявляется непрозрачность и каковы ее причины? Основной целью машинного обучения (ML) является обучение точных систем принятия решений – предикторов (predictors), способных помочь автоматизировать задачи, которые в противном случае должны были бы выполнять люди. Машинное обучение располагает множеством алгоритмов, которые продемонстрировали значительные успехи в науке и промышленности. Наиболее популярными рабочими лошадками ML считаются методы ядра (kernel methods) (Hofmann et al., 2008) и, особенно в последнее десятилетие, методы глубокого обучения (deep learning methods) (Vargas et al., 2018).

Поскольку ML все чаще используется в реальных программах и приложениях, сложилось общее мнение, что высокая точность решения или предсказания на практике может быть недостаточной (Gunning, 2017).

Первая трудность связана с многомасштабной и распределенной природой представлений нейронных сетей. Некоторые нейроны активируются только для нескольких точек данных, в то время как другие действуют более глобально. Таким образом, предсказание представляет собой сумму локальных и глобальных эффектов, что затрудняет (или делает невозможным) поиск корневой точки x , которая линейно расширяется до предсказания для интересующей точки данных. Переход от глобального к локальному эффекту вносит нелинейность, которую невозможно уловить (Samek et al., 2021).

Второй источник нестабильности возникает из-за большой глубины современных нейронных сетей, где наблюдается эффект «разбитого градиента» (shattered gradient problem) (Balduzzi et al., 2017). Градиентом в нейронных сетях называется вектор частных производных функции потерь по весам нейронной сети. Он используется в оптимизаторе весов для улучшения качества модели. Градиент показывает изменение ошибок на разных наборах данных.

Наконец, выделяется проблема объяснимости технологии искусственного интеллекта с необходимостью поиска корневой точки x , на которой будет основано объяснение и которая одновременно близка к данным и не является неблагоприятным примером (the problem of adversarial examples) (Goodfellow et al., 2014). Проблема неблагоприятных примеров объясняется градиентным шумом, который заставляет модель давать «чрезмерную реакцию» (overreact) на определенные пиксельные возмущения, а также высокой размерностью данных (dimensionality of the data) (Najafabadi et al., 2015), где множество мелких пиксельных эффектов суммируются в большой эффект на результат модели (Samek et al., 2021).

Эти особенности работы искусственного интеллекта как класса приводят к тому, что, пока доступны большие данные и огромные вычисления, для достижения сверхчеловеческой производительности требуется «нулевое знание человека» (zero human knowledge) (Silver et al., 2017).

Исследователи предлагают признать, что искусственный интеллект находится внутри социотехнической системы, которая опосредует доверие, а повышая надежность этих систем, стараясь сделать эти процессы менее непрозрачными для большинства наблюдателей, мы тем самым повышаем доверие к искусственному интеллекту (von Eschenbach, 2021). В данном контексте исключение человека из процесса принятия решения добавляет ему доверия, так как исключает фактор субъективности в полученном результате.

В то же время вопрос доверия зависит не только от возможности человека вмешиваться в процесс принятия искусственным интеллектом решения. На современном этапе растет спрос на объяснимый искусственный интеллект (explainable artificial intelligence (XAI)). Роман Ямпольский указывает, что «если все, что у нас есть, – это “черный ящик”, то невозможно понять причины сбоев и повысить безопасность системы. Кроме того, если мы привыкнем принимать ответы искусственного интеллекта без объяснения причин, мы не сможем определять, если он начнет давать неправильные или манипулятивные ответы» (Yampolskiy, 2019). Ученый в красках расписывает опасности непрозрачного искусственного интеллекта, предлагая представить, что в ближайшем будущем искусственный интеллект может ошибаться в диагностике заболеваний в 5 % случаев, что приведет к массовым операциям здоровых людей. Отсутствие механизма проверки модели искусственного интеллекта на отклонения и предотвращения подобных сбоев может привести к непоправимым последствиям. Таким образом, прозрачность и подотчетность являются инструментами, способствующими принятию справедливых алгоритмических решений, обеспечивая основу для получения возможности обратиться к значимому объяснению, исправлению или способам установления недостатков, которые могут привести к компенсационным процессам (Koene, 2019).

2. Правовые и этические риски применения непрозрачной технологии

Проблема прозрачности определяется в российской Концепции развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники до 2024 г. (далее – Концепция 2024) как «использование для принятия решений системами искусственного интеллекта вероятностных оценок и невозможность в ряде случаев полного объяснения принятого ими решения (проблема алгоритмической прозрачности систем искусственного интеллекта)»².

В Концепции 2024 прозрачность названа среди таких проблемных направлений регулирования искусственного интеллекта, как соблюдение баланса между требованиями по защите персональных данных и необходимостью их использования для обучения систем искусственного интеллекта; определение предмета и границ регулирования сферы использования систем искусственного интеллекта и робототехники; правовое «делегирующее» решений системам искусственного интеллекта и робототехники; ответственность за причинение вреда с использованием систем искусственного интеллекта и робототехники. То есть вопросы правового обеспечения прозрачности искусственного интеллекта играют концептуально важную роль для выработки правовых подходов.

Выше было показано, что разработчик представляет данные, но не может контролировать, на основании каких именно критериев искусственный интеллект сформулировал результат или прогноз. Складывается видение, что иногда осмысленную нейронную сеть создать не представляется возможным. Это связано с трудностями определения входных данных и их фактической недостаточностью. По сути, потеря контроля над искусственным интеллектом основывается на неопределенности данных, с которыми взаимодействует модель (Kharitonova et al., 2021).

Способны ли разработчики и юристы аргументированно вмешаться в работу системы и поспорить с ее выводами, если они не понимают принципов принятия этих решений? Разработчики могут указать критерии принятия решений, но искусственный интеллект самостоятельно может дополнить условно недостающие данные для формирования решений на выходе. Например, машина анализирует точки или пиксели, но не знает, что это цвет кожи или глаз. Манипулирует «кубиками» пикселей, а не общей картиной.

В то же время решения, принимаемые людьми, юристами и даже судьями, чья деятельность в этом аспекте подробно регулируется, несвободны от осознанного и/или бессознательного предубеждения. Исследователи человеческих предрассудков доказали, что люди когнитивно предрасположены к пристрастиям и стереотипам (Plous, 2003; Quillian, 2006), хотя современные формы предрассудков часто трудно обнаружить и они могут быть даже неизвестны самим носителям предрассудков. Практика обоснования решений может быть просто недостаточной для противодействия влиянию различных факторов, а причины, предлагаемые для человека, принимающего решения, могут скрывать мотивы, едва ли известные тем, кто принимает решения (McEwen et al., 2018).

² Об утверждении Концепции развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники до 2024 года: распоряжение Правительства РФ от 19.08.2020 № 2129-р.

То есть зачастую алгоритмическая и человеческая предвзятость и необъяснимость принятого решения существуют в латентном виде, не осознаваемые ее носителем и не выявленные третьими лицами. Это позволяет предполагать, что бездушный, безэмоциональный алгоритм все еще может выступать более объективным мерилем принятия решения, поскольку избавлен от личных субъективных предвзятостей.

Риск использования непрозрачного искусственного интеллекта приобретает критическое значение, если такая технология обязательно должна быть применена субъектом деятельности. Так, в беспилотном автомобиле в ходе движения решение принимает система искусственного интеллекта, в то время как ответственность за источник повышенной опасности все еще возлагается на водителя (Payre & Cestac, 2014). Или иной пример ближайшей перспективы. Сегодня в медицине все шире используются роботы на основе искусственного интеллекта для ассистирования хирургам (Kalis et al., 2014). В ходе оказания медицинской помощи некоторые процедуры становятся обязательными и, следовательно, врач может оказаться в ситуации, когда его решения повлекли ответственность, хотя фактически причиной вреда стали проблемы программного кода искусственного интеллекта.

Исследуя вопросы правового вмешательства в распространение дипфейков, В. О. Калятин приходит к выводу, что «разрабатывать соответствующее законодательство следует не в отношении дипфейков как таковых, а в отношении использования ИИ в целом» (Калятин, 2022). Перед юристами стоит выбор: оставаться в действующей правовой традиции либо создавать новую. Полагаем, попытки создать правовой режим применения предпринимателями искусственного интеллекта не могут увенчаться успехом в отсутствие понимания технологических особенностей искусственного интеллекта. Однако прозрачность как объяснимость технологии не может быть понята в буквальном смысле. Требуется создание критериев проверки результатов работы искусственного интеллекта для целей соблюдения прав и свобод граждан, защиты государственных и общественных интересов.

3. Автоматизированная система обработки данных и их качество

Если прозрачность сама по себе не свойственна природе алгоритмов (Kalpokas, 2019) при условии, что на входе для запуска приложений искусственного интеллекта предоставляется информация, возникает вопрос о возможности расстановки акцентов о правилах анализа данных алгоритмом искусственного интеллекта.

В литературе выделены несколько аспектов алгоритмической прозрачности и подотчетности, которые включают повышение осведомленности, подотчетность при использовании алгоритмических решений, и прежде всего в государственном секторе, а также нормативный надзор и юридическую ответственность, которые приведут к глобальной координации алгоритмического управления (Koene et al., 2019).

Осведомленность, рассматриваемая многими исследователями как решение проблемы прозрачности, может пониматься по-разному. Прежде всего при обеспечении прозрачности искусственного интеллекта во главу угла ставится работа с данными и осведомленность об их использовании в определенном ключе.

Обратим внимание, во многих юрисдикциях анализ данных и его пределы установлены в отношении персональных данных. В России действуют положения ст. 16

Закона о персональных данных³, согласно которой запрещается принятие на основании исключительно автоматизированной обработки персональных данных решений, порождающих юридические последствия в отношении субъекта персональных данных или иным образом затрагивающих его права и законные интересы, за исключением случаев, предусмотренных законом. К таким случаям отнесены ситуации, когда решение, порождающее юридические последствия в отношении субъекта персональных данных или иным образом затрагивающее его права и законные интересы, принято на основании исключительно автоматизированной обработки его персональных данных при наличии согласия в письменной форме субъекта персональных данных (п. 2 ст. 16 Закона о персональных данных).

Приведенное положение российского законодательства сопоставимо со ст. 15 ныне не применяемой Директивы № 95/46/ЕС Европейского парламента и Совета ЕС «О защите физических лиц при обработке персональных данных и о свободном обращении таких данных»⁴. В действующем GDPR содержатся аналогичные правила. Статья 22(3) Общего регламента по защите данных гласит, что в некоторых случаях автоматизированной обработки «контролер данных должен принять соответствующие меры для защиты прав, свобод и законных интересов субъекта данных, по крайней мере, право на вмешательство человека со стороны контролера выражать свою точку зрения и оспаривать решение»⁵.

В то же время для обеспечения осведомленности важно и раскрытие информации о данных, на которых базируется принятое решение. Сюда относятся вопросы достоверности и нейтральности, репрезентативности данных, непредвзятости методов их обработки и анализа, а также информация о самообучении искусственного интеллекта.

Прозрачность как объяснимость и недискриминационность технологии зависит от качества данных, с которыми работает система искусственного интеллекта. Исследователи установили (Buolamwini & Gebru, 2018), что все популярные системы распознавания лиц точнее всего определяют мужчин со светлой кожей (2,4 % ошибок). Но чаще всего ошибаются при обнаружении темнокожих женщин (61 % ошибок). По сути, тем самым было доказано, что «фотографий женщин с темной кожей в базах данных меньше всего; разработчики таких систем сами преимущественно являются белыми мужчинами; датчики камер хуже распознают детали в темных цветах»⁶.

Приведенный пример демонстрирует, что недостаточно поставить под сомнение достоверность данных, доступных искусственному интеллекту. Проблема качества данных в том, что доступные данные не обладали нейтральностью даже при условии, что они могли считаться репрезентативными. Системы распознавания лиц

³ О персональных данных: Федеральный закон от 27.07.2006 № 152-ФЗ. СПС КонсультантПлюс. https://www.consultant.ru/document/cons_doc_LAW_61801/

⁴ Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. (1995, November 23). *Official Journal of the European Communities*, L 281, 31–50.

⁵ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). <http://eur-lex.europa.eu/>

⁶ Ученые все чаще не могут объяснить, как работает ИИ. Теория «черного» и «белого» ящика. (2022, 23 ноября). Хабр. <https://habr.com/ru/company/getmatch/blog/700736/>

используются во многих странах. В том числе в работе правоохранительных органов. Доказано, что если вы одно из расовых меньшинств в такой стране, эта система будет чаще определять в вас преступника⁷. Высказано мнение, что если обучать модели искусственного интеллекта с помощью больших данных, встроенные в них расовые и другие предубеждения будут неизбежностью (Bender et al., 2021), ведь у некоторых групп людей просто меньше доступа к Интернету, а данные о них мало представлены на разного рода ресурсах (жители отдаленных населенных пунктов, например, в сравнении с программистами).

В целом, мы полагаем, уязвимостью подхода, сосредоточенного только на персональных данных, является несоответствие его действительности. В дополнение к данным, которые поступают непосредственно от пользователей в алгоритм, бизнес должен пополнять эти категории пользователей дополнительными аналитическими данными (Camilleri, 2018), более подробно описывающими различные группы и делающими обоснование классификации еще более ясным.

Следовательно, необходимо устанавливать правила для определения качественного (достоверного и нейтрального набора данных) в ситуации, когда ограничить такой набор данных не представляется возможным. Риск необъяснимых предвзятых решений искусственного интеллекта придется исключать путем обучения с подкреплением и аудита полученного результата.

Это приводит к мысли, что объяснения понятного человеку алгоритма при использовании метода черного ящика ждать не стоит, но и юридического значения раскрытие алгоритма здесь иметь не будет. Научить систему искусственного интеллекта понимать этические ценности невозможно, юристы могут лишь указать критерии проверки решения искусственного интеллекта на объективность. Не во всех случаях можно поставить человека на выходе для проверки результата. Следовательно, право может узаконить лишь необходимость контроля со стороны программы, созданной независимыми разработчиками.

В связи с этим труднодостижимым представляется предложение достичь прозрачности искусственного интеллекта не в отношении к системе в целом, а через объяснение логики отдельных индивидуальных решений (Кутейников и др., 2018). В качестве таких мер авторами предлагается анализ входных данных, статистическое объяснение, проверка архитектуры/кода и статистический анализ, определение чувствительности отдельных данных (какие именно переменные определяют результат) (Кутейников и др., 2018).

Напротив, более выполнимым представляется требовать открытости алгоритма с указанием общей логики принятия решений в целом. Эта система принята на вооружение в штате Калифорния, США. В феврале 2020 г. там был принят Закон о подотчетности автоматизированных систем принятия решений, в котором предписано проводить систематический контроль и выявлять ошибки функционирования автоматизированных систем, а также направлять полученные отчеты в Департамент регулирования бизнеса (Department of Business Oversight) начиная с 1 января 2022 г. и размещать их в открытом доступе в сети Интернет⁸.

⁷ Ученые все чаще не могут объяснить, как работает ИИ. Теория «черного» и «белого» ящика. (2022, 23 ноября). Хабр. <https://habr.com/ru/company/getmatch/blog/700736/>

⁸ USA. State of California. Automated Decision Systems Accountability Act of 2020. https://leginfo.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB2269

На это же обращают внимание исследователи GDPR. Как пишут А. Селбст и Дж. Паулс, «проблема в том, что для проверки точности защиты и потенциального оспаривания ее правильности необходимы конкретные объяснения решения, включая веса и факторы, используемые для его достижения» (Selbst & Powles, 2018), что составляет техническую, не всегда существующую в рамках понимания человека информацию. С этих позиций закрепленное в ст. 22 (3) Общего регламента по защите данных «право» недостаточно подробно раскрыто в законодательстве и подвергается обоснованной критике в связи с невыполнимостью. Большинство ученых сходятся во мнении, что право на разъяснение индивидуальных решений, которое может включать глобальные или локальные разъяснения, не следует из ст. 22(3) GDPR (Wachter, 2017; Goodman & Flaxman, 2017). Статья 15(1)(h) GDPR предусматривает, что в случае автоматизированной обработки в смысле ст. 22(1) GDPR контролер должен предоставить «значимую информацию о задействованной логике». Некоторые исследователи убеждены, что это относится только к общей структуре и архитектуре модели обработки, но нет необходимости в объяснениях отдельных решений или конкретных весов и характеристик модели (Wachter, 2017; Malgieri & Comandé, 2017).

В отсутствие стандартизированного подхода к обоснованию индивидуальных и общих решений без ответа остаются не только вопросы о том, кому вообще должна раскрываться логика принятия решения, только ли пользователям или субъектам данных либо всем заинтересованным лицам, в каком объеме и др.

В данном контексте следует обратить внимание на значимость применения искусственного интеллекта в отношениях с участием государства. Согласимся, что необходимо закрепление норм об обязательной общедоступности результатов использования государственными органами технологий искусственного интеллекта и больших данных. Как убедительно доказывает вслед за европейскими коллегами В. В. Силкин, если государство осуществляет свои функции с применением искусственного интеллекта, прозрачность технологии необходима. При этом автор предлагает возложить на государственные органы обязанность обосновывать и раскрывать цели использования технологий автоматизированной обработки данных. Возможности больших данных и искусственного интеллекта достаточно обширны, однако использование их государством должно быть обусловлено необходимостью достижения именно общественно значимых целей (Силкин, 2021). Однако мы полагаем, при общем распространении данной технологии, обоснование применения искусственного интеллекта в отдельных видах государственной деятельности решит эту задачу в целом, но не обеспечит прозрачность решений.

Однако хотелось бы подчеркнуть, что принцип прозрачности работы искусственного интеллекта не тождественен принципу прозрачности деятельности органов власти или иных операторов данных по применению автоматизированных систем. В. В. Силкин предлагает «при реализации принципа транспарентности в деятельности государственных органов по применению автоматизированных систем обработки данных исходить из открытости информации о целях, способах и результатах их использования» (Силкин, 2021). При этом автор верно утверждает, что «в то же время в комплексных автоматизированных системах обработки алгоритмы формируются и усложняются самостоятельно, что исключает возможность предугадать или заранее однозначно понять все возможности таких систем» (Силкин, 2021).

Полагаем, в данном случае легко может возникнуть подмена понятий: прозрачность принятия решения искусственным интеллектом не связана исключительно

с прозрачностью целей, установленных для применения алгоритма. Нужно разделять прозрачность алгоритма как принцип регулирования искусственного интеллекта и прозрачность деятельности (госуправления, гражданского оборота и т. п.), которая достигается отчасти с применением технологии искусственного интеллекта. Фактически второй аспект автоматической обработки данных наравне с качеством данных составляет прозрачность целей и методик применения искусственного интеллекта. В этом принцип прозрачности работы искусственного интеллекта совпадает, но не тождествен понятию прозрачности деятельности.

Возможно, цели использования искусственного интеллекта можно было бы определить как цели принятия решений с объяснением процессов искусственного интеллекта по стандартизированной форме, которая требует регулярного обновления каждый раз, когда бизнес меняет свои методы автоматизированной обработки (Wulf & Seizov, 2022).

Представляется, что закрепление права на осведомленность о применении технологии искусственного интеллекта при автоматизированной обработке данных для принятия решений не исчерпывает возможные правовые средства для достижения прозрачности интеллектуальных систем. Более того, защита прав граждан не достигается, входя в противоречие с методами работы искусственного интеллекта, и прежде всего черного ящика. Раскрытие задач и приоритетов алгоритма для целей принятия решения может строиться на стандартах, но с учетом необходимости обращения к новым данным и восполнения пробелов в полученных от пользователя данных, достичь прозрачности и в этом направлении не представляется возможным.

4. Открытость алгоритмов и результатов их работы

Раскрытие сведений о разработке программного обеспечения, кода и порядка его работы также стоит в приоритете обеспечения прозрачности искусственного интеллекта.

В ст. 16 Закона о персональных данных, по справедливому мнению А. И. Савельева, условиями использования инструментов автоматизированной обработки персональных данных для целей принятия юридически значимых решений в отношении субъекта предусмотрены «дополнительные информационные обязанности оператора, выражающиеся в его обязанности предоставить субъекту разъяснения относительно порядка принятия такого решения и возможных юридических последствий такого решения» (Савельев, 2021).

Проект Закона об искусственном интеллекте⁹, представленный Европейской комиссией 21 апреля 2021 г., предлагает для обеспечения прозрачности принимаемых искусственным интеллектом решений, например, раскрывать информацию о характеристиках, возможностях и ограничениях системы искусственного интеллекта, предназначении системы, а также информацию, необходимую для обслуживания систем искусственного интеллекта.

⁹ European Commission. (2021, April 21). *Proposal for a Regulation laying down harmonised rules on artificial intelligence*. <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence>

В США с 2022 г. обсуждается проект закона об алгоритмической подотчетности¹⁰, согласно которому от компаний, использующих системы искусственного интеллекта, потребуется проводить оценку своих автоматизированных систем в соответствии с правилами Федеральной торговой комиссии на предмет обеспечения недискриминации пользователей и конфиденциальности информации.

Введение на законодательном уровне требования о раскрытии алгоритма дискутируется и в России, хотя пока только в отношении рекомендательных алгоритмов социальных сетей¹¹. Аналогичные требования содержатся также в праве Китая – в Положении об управлении алгоритмическими рекомендациями информационных услуг в сети Интернет¹².

В то же время требование об абсолютном раскрытии исходного или архитектуры применимых моделей искусственного интеллекта не представляется обоснованным. Во всяком случае такое раскрытие не может быть доскональным по ряду причин. Технологически такое раскрытие является весьма затратным и требует огромных ресурсов. Публикация исходного кода также может приводить к нарушениям права интеллектуальной собственности или нарушению коммерческой тайны. Такой подход к прозрачности вступает в противоречие с имеющимися правовыми режимами, которым подчинены те или иные составляющие единой технологии искусственного интеллекта.

При этом убедительной кажется позиция, согласно которой системы искусственного интеллекта в первую очередь охраняются как коммерческая тайна (торговый секрет), поскольку попытки защитить системы искусственного интеллекта в соответствии с законами об авторском праве и патентах наталкиваются на трудности (Foss-Solbrekk, 2021). Проблема предоставления авторско-правовой защиты самому алгоритму связана с постоянным его изменением и дописыванием в ходе самообучения и автономной работы. Спорным является и вопрос о самом творческом характере происхождения искусственного интеллекта как объекта правовой охраны. Получение патентов на системы искусственного интеллекта также затруднено. Эта ситуация наблюдается в разных правовых порядках. Например, в российском законодательстве об авторском праве алгоритмы по существу получили самостоятельное регулирование в рамках ст. 1261 ГК РФ, в ЕС алгоритмы исключены из-под защиты копирайта в соответствии с Директивой ЕС по программному обеспечению¹³. Во всяком случае идущие бурные дискуссии об отнесении искусственного интеллекта к тому или иному виду объектов права интеллектуальной собственности далеки от завершения.

Защита моделей искусственного интеллекта с помощью правового режима коммерческой тайны приводит к столкновению требований прозрачности

¹⁰ Metcalf, J., Smith, B., & Moss, E. (2022, February 9). A New Proposed Law Could Actually Hold Big Tech Accountable for Its Algorithms. *Slate*. <https://slate.com/technology/2022/02/algorithmic-accountability-act-wyden.html>

¹¹ СМІ: в Госдумі планують внести проект о регулюванні рекомендаційних сервісів в соцсетях. (2021, 15 октября). *Парламентская газета*. <https://www.pnp.ru/politics/smi-v-gosdumu-planiruyut-vnesti-proekt-o-regulirovanii-rekomendatelnykh-servisov-v-socsetyakh.html>

¹² 互联网信息服务算法推荐管理规定. (2021, 31 декабря). http://www.cac.gov.cn/2022-01/04/c_1642894606364259.htm

¹³ О правовой охране компьютерных программ (кодифицированная версия): Директива № 2009/24/ЕС Европейского парламента и Совета Европейского союза. <https://base.garant.ru/71657620/>

и подотчетности. В частности, исследователи приходят к выводу, что Европейская директива 2016/943 оставляет мало места для гипотез об алгоритмической прозрачности (Maggiolino, 2018).

В определении понятия «коммерческая тайна», приведенном в тексте Директивы 2016/943, речь идет о коммерческой ценности без указания на ее фактический или потенциальный характер. Согласно ст. 2 (1), коммерческая тайна означает информацию, которая отвечает всем следующим требованиям:

- «(а) она является секретной, то есть не является общеизвестной или легкодоступной для лиц в кругах, которые обычно имеют дело с рассматриваемой информацией;
- (b) она имеет коммерческую ценность, потому что является секретной;
- (с) при данных обстоятельствах лицо, на законном основании контролирующее информацию, приняло разумные меры для сохранения ее в тайне»¹⁴.

При таком подходе требование о раскрытии кода или архитектуры модели искусственного интеллекта будет означать потерю конкурентного преимущества на рынке. Поскольку защита коммерческой тайны существует до тех пор, пока информация остается конфиденциальной, и требует, чтобы субъекты предпринимали шаги для обеспечения конфиденциальности, защита коммерческой тайны способствует алгоритмической непрозрачности.

Из этого следует, что отказ от секретности алгоритмов искусственного интеллекта должен сопровождаться усилением защиты интересов раскрывшей сведения стороны перед третьими лицами, как сейчас это происходит в патентном праве.

5. Применимые правовые средства для предотвращения проблем непрозрачности правовых решений: возразить или отказаться

Осведомленность о применении искусственного интеллекта для принятия решений также приводит к необходимости обсуждения права на отказ от применения технологии искусственного интеллекта в конкретном случае, а также права на возражение против принятого решения.

В силу п. 3 ст. 16 российского Закона о персональных данных оператор обязан предоставить субъекту персональных данных возможность заявить возражение против решения, принятого в результате автоматизированной обработки данных, а также разъяснить порядок защиты субъектом персональных данных своих прав и законных интересов. А. И. Савельев поясняет, что речь «идет о тех правах, которые касаются именно принятия юридически значимых решений по результатам исключительно автоматизированной обработки персональных данных. В частности, данный порядок защиты предполагает уведомление субъекта о его праве требовать вмешательства человека в процесс принятия решения, которое является неотъемлемой частью права на предоставление возражений» (Савельев, 2021).

¹⁴ О защите конфиденциальных ноу-хау и деловой информации (коммерческой тайны) от незаконного приобретения, использования и раскрытия: Директива № 2016/943 Европейского парламента и Совета Европейского союза. <https://base.garant.ru/71615160/>

В России также обсуждается законопроект, в котором предполагается предоставить пользователям возможность полностью или частично отказаться от применения рекомендательных алгоритмов.

В Китае такой подход уже принят к реализации и проверяется на исполнимость. По мнению исследователей, раскрытие логики алгоритма должно устранить риски необоснованных отказов поставщиков услуг в случае алгоритмической рекомендации предоставить необходимую информацию, такую как область применения алгоритма, пользователь услуги, уровень риска алгоритма и другие, на том основании, что четкие законодательные установления в этом отношении отсутствуют (Xu Ke & Liu Chang, 2022). Однако иногда такое раскрытие ведет не к успеху, а является декоративным оформлением сайта.

Таким образом, добиваться прозрачности и объяснимости работы искусственного интеллекта в понятной человеку форме юристам стоит добиваться в том числе и для того, чтобы обеспечить возможность пользователю системы искусственного интеллекта возражать против принятого решения. Этот вопрос также связан и с определением субъекта ответственности за принимаемые решения, которые могут существенно ущемлять права человека.

Ученые заметили, что за текущей дискуссией о требованиях защиты данных в отношении объяснимости упускается из виду важность данной характеристики для оценки договорной и деликтной ответственности в отношении использования инструментов искусственного интеллекта (Hacker et al., 2020). В связи с этим дальнейшей конкретизации потребуют положения законодательства по использованию искусственного интеллекта в сфере усиления обязанности для разработчиков, производителей и поставщиков услуг искусственного интеллекта постоянно оценивать возможные неблагоприятные последствия применения искусственного интеллекта для прав человека и основных свобод и, учитывая эти последствия, принимать меры по предотвращению и смягчению рисков (Дьяконова и др., 2022).

Выводы

Прозрачность является необходимым условием для признания искусственного интеллекта заслуживающим доверия. Наиболее плодотворный путь для установления доверия к искусственному интеллекту заключается в том, чтобы признать данную технологию частью сложной социотехнической системы, которая опосредует доверие и повышает надежность этих систем.

Большая часть дебатов вокруг прозрачности искусственного интеллекта с юридической точки зрения сосредоточена на законах о защите данных. Полагаем, круг этих дискуссий следует расширять. Прозрачность и объяснимость технологии искусственного интеллекта важна не только для защиты персональных данных, но и в иных ситуациях автоматизированной обработки данных, когда для принятия решений недостающие из входящей информации технологические данные восполняются из открытых источников, в том числе не имеющих значения хранилищ персональных данных. Законодатель может добиваться лишь установления некоего стандарта, закрепляющего компромисс между возможностями и преимуществами технологии, точностью и объяснимостью результата ее работы и правами участников общественных отношений. Введение сертификации моделей искусственного интеллекта, обязательных к применению, позволит решить вопросы ответственности

обязанных применять такие системы субъектов. В контексте вопроса о профессиональной ответственности профессиональных субъектов, таких как врачи, военные, органы корпоративного управления юридического лица, требуется ограничить обязательное применение искусственного интеллекта в случаях, если не обеспечена его достаточная прозрачность.

Развитие правовой дискуссии должно двигаться по выработке предложений о содержании права на отказ от применения автоматизированной обработки данных для принятия решений и права на возражения против принятых таким способом решений.

Список литературы

- Дьяконова, М. О., Ефремов, А. А., Зайцев, О. А. и др.; И. И. Кучерова, С. А. Синицына (ред.). (2022). Цифровая экономика: актуальные направления правового регулирования: научно-практическое пособие. Москва: ИЗиСП, НОРМА.
- Калятин, В. О. (2022). Дипфейк как правовая проблема: новые угрозы или новые возможности? *Закон*, 7, 87–103. <https://doi.org/10.37239/0869-4400-2022-19-7-87-103>
- Кутейников, Д. Л., Ижаев, О. А., Зенин, С. С., Лебедев, В. А. (2020). Алгоритмическая прозрачность и подотчетность: правовые подходы к разрешению проблемы «черного ящика». *Lex russica (Русский закон)*, 73(6), 139–148. <https://doi.org/10.17803/1729-5920.2020.163.6.139-148>
- Савельев, А. И. (2021). Научно-практический постатейный комментарий к Федеральному закону «О персональных данных» (2-е изд., перераб. и доп.). Москва: Статут.
- Силкин, В. В. (2021). Транспарентность исполнительной власти в цифровую эпоху. *Российский юридический журнал*, 4, 20–31. https://doi.org/10.34076/20713797_2021_4_20
- Balduzzi, D., Frea, M., Leary, L., Lewis, J. P., Ma, K. W. D., & McWilliams, B. (2017). The shattered gradients problem: If resnets are the answer, then what is the question? In *International Conference on Machine Learning* (pp. 342–350). PMLR.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610–623).
- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency* (pp. 77–91). PMLR.
- Camilleri, M. A. (2018). Market segmentation, targeting and positioning. In *Travel marketing, tourism economics and the airline product* (pp. 69–83). New York: Springer.
- Cho, Y. H., Kim, J. K., & Kim, S. H. (2002). A personalized recommender system based on web usage mining and decision tree induction. *Expert systems with Applications*, 23(3), 329–342. [https://doi.org/10.1016/S0957-4174\(02\)00052-0](https://doi.org/10.1016/S0957-4174(02)00052-0)
- Cragun, B. J., & Steudel, H. J. (1987). A decision-table-based processor for checking completeness and consistency in rule-based expert systems. *International Journal of Man-Machine studies*, 26(5), 633–648. [https://doi.org/10.1016/S0020-7373\(87\)80076-7](https://doi.org/10.1016/S0020-7373(87)80076-7)
- Foss-Solbrekk, K. (2021). Three routes to protecting AI systems and their algorithms under IP law: The good, the bad and the ugly. *Journal of Intellectual Property Law & Practice*, 16(3), 247–258. <https://doi.org/10.1093/jiplp/jpab033>
- Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). *Explaining and harnessing adversarial examples*. arXiv preprint arXiv:1412.6572.
- Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a “right to explanation”. *AI magazine*, 38(3), 50–57. <https://doi.org/10.1609/aimag.v38i3.2741>
- Gunning, D. (2017). Explainable artificial intelligence (XAI). *Defense advanced research projects agency (DARPA)*. nd Web. 2(2), 1.
- Hacker, P., Krestel, R., Grundmann, S., & Naumann, F. (2020). Explainable AI under contract and tort law: legal incentives and technical challenges. *Artificial Intelligence and Law*, 28(4), 415–439. <https://doi.org/10.1007/s10506-020-09260-6>
- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). New York: Springer.
- Hofmann, T., Schölkopf, B., & Smola, A. J. (2008). *Kernel methods in machine learning*.
- Kalis, B., Collier, M., & Fu, R. (2014). 10 promising AI applications in health care. *Harvard Business Review*.

- Kalpokas, I. (2019). *Algorithmic Governance. Politics and Law in the Post-Human Era*. Cham: Palgrave Pivot.
- Kharitonova, Y. S., Savina, V. S., & Pagnini, F. (2021). Artificial Intelligence's Algorithmic Bias: Ethical and Legal Issues. *Perm U. Herald Jurid. Sci*, 3(53), 488–515. <https://doi.org/10.17072/1995-4190-2021-53-488-515>
- Koene, A., Clifton, C., Hatada, Y., Webb, H., & Richardson, R. (2019). *A governance framework for algorithmic accountability and transparency*. Panel for the Future of Science and Technology.
- Lawrence, R. L., & Wright, A. (2001). Rule-based classification systems using classification and regression tree (CART) analysis. *Photogrammetric engineering and remote sensing*, 67(10), 1137–1142.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
- Maggiolino, M. (2018). *EU trade secrets law and algorithmic transparency*. SSRN 3363178.
- Malgieri, G., & Comandé, G. (2017). Why a right to legibility of automated decision-making exists in the general data protection regulation. *International Data Privacy Law*, 7(4), 243–265. <https://doi.org/10.1093/idpl/ix019>
- McEwen, R., Eldridge, J., & Caruso, D. (2018). Differential or deferential to media? The effect of prejudicial publicity on judge or jury. *The International Journal of Evidence & Proof*, 22(2), 124–143. <https://doi.org/10.1177/1365712718765548>
- Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, 2(1), 1–21. <https://doi.org/10.1186/s40537-014-0007-7>
- Pak, M., & Kim, S. (2017). A review of deep learning in image recognition. In *2017 4th international conference on computer applications and information processing technology (CAIPT)*. <https://doi.org/10.1109/caipt.2017.8320684>
- Payre, W., Cestac, J., & Delhomme, P. (2014). Intention to use a fully automated car: Attitudes and a priori acceptability. *Transportation research part F: traffic psychology and behavior*, 27, 252–263. <https://doi.org/10.1016/j.trf.2014.04.009>
- Plous, S. E. (2003). *Understanding prejudice and discrimination*. McGraw-Hill.
- Quillian, L. (2006). New approaches to understanding racial prejudice and discrimination. *Annu. Rev. Sociol.*, 32, 299–328. <https://doi.org/10.1146/annurev.soc.32.061604.123132>
- Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., & Müller, K. R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE*, 109(3), 247–278. <https://doi.org/10.1109/jproc.2021.3060483>
- Selbst, A., & Powles, J. (2017). “Meaningful information” and the right to explanation. *International Data Privacy Law*, 7(4), 233–242. <https://doi.org/10.1093/idpl/ix022>
- Silver, D. et al. (2017). Mastering the game of go without human knowledge. *Nature*, 550(7676), 354.
- Vargas, R., Mosavi, A., & Ruiz, R. (2018). *Deep learning: a review*. Preprints.org. <https://doi.org/10.20944/preprints201810.0218.v1>
- von Eschenbach, W. J. (2021). Transparency and the black box problem: Why we do not trust AI. *Philosophy & Technology*, 34(4), 1607–1622.
- Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law*, 7(2), 76–99. <https://doi.org/10.1093/idpl/ix005>
- Wulf, A. J. & Seizov, O. (2022). “Please understand we cannot provide further information”: evaluating content and transparency of GDPR-mandated AI disclosures. *AI & SOCIETY*, 1–22. <https://doi.org/10.1007/s00146-022-01424-z>
- Yampolskiy, R. V. (2019). *Unexplainability and incomprehensibility of artificial intelligence*. arXiv preprint arXiv:1907.03869
- 许可、刘畅. 论算法备案制度 // 人工智能. 2022. № 1. P. 66. [Xu Ke, Liu Chang. (2022). On the Algorithm Filing System. *Artificial Intelligence*, 1, 66.]

Сведения об авторе



Харитоновна Юлия Сергеевна – доктор юридических наук, профессор, профессор кафедры предпринимательского права, руководитель Центра правовых исследований искусственного интеллекта и цифровой экономики, Московский государственный университет имени М. В. Ломоносова

Адрес: 119991, Российская Федерация, г. Москва, Ленинские горы, 1

E-mail: sovet2009@rambler.ru

ORCID ID: <https://orcid.org/0000-0001-7622-6215>

Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=57316440400>

Web of Science Researcher ID:

<https://www.webofscience.com/wos/author/record/708572>

Google Scholar ID: <https://scholar.google.ru/citations?user=61mQtb4AAAAJ>

РИНЦ Author ID: https://elibrary.ru/author_items.asp?authorid=465239

Конфликт интересов

Автор является членом редакционной коллегии журнала, статья прошла рецензирование на общих основаниях.

Финансирование

Исследование не имело спонсорской поддержки.

Тематические рубрики

Рубрика OECD: 5.05 / Law

Рубрика ASJC: 3308 / Law

Рубрика WoS: OM / Law

Рубрика ГРНТИ: 10.23.01 / Общие вопросы предпринимательского права

Специальность ВАК: 5.1.3 / Частно-правовые (цивилистические) науки

История статьи

Дата поступления – 6 марта 2023 г.

Дата одобрения после рецензирования – 13 апреля 2023 г.

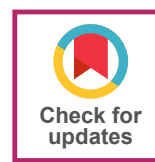
Дата принятия к опубликованию – 16 июня 2023 г.

Дата онлайн-размещения – 20 июня 2023 г.



Research article

DOI: <https://doi.org/10.21202/jdtl.2023.14>



Legal Means of Providing the Principle of Transparency of the Artificial Intelligence

Yuliya S. Kharitonova 

Lomonosov Moscow State University
Moscow, Russian Federation

Keywords

Algorithm,
artificial intelligence,
automated data processing,
autonomy,
decision-making,
digital economy,
digital technologies,
ethics,
law,
transparency

Abstract

Objective: to analyze the current technological and legal theories in order to define the content of the transparency principle of the artificial intelligence functioning from the viewpoint of legal regulation, choice of applicable means of legal regulation, and establishing objective limits to legal intervention into the technological sphere through regulatory impact.

Methods: the methodological basis of the research is the set of general scientific (analysis, synthesis, induction, deduction) and specific legal (historical-legal, formal-legal, comparative-legal) methods of scientific cognition.

Results: the author critically analyzed the norms and proposals for normative formalization of the artificial intelligence transparency principle from the viewpoint of impossibility to obtain the full technological transparency of artificial intelligence. It is proposed to discuss the variants of managing algorithmic transparency and accountability based on the analysis of social, technical and regulatory problems created by algorithmic systems of artificial intelligence. It is proved that transparency is an indispensable condition to recognize artificial intelligence as trustworthy. It is proved that transparency and explainability of the artificial intelligence technology is essential not only for personal data protection, but also in other situations of automated data processing, when, in order to make a decision, the technological data lacking in the input information are taken from open sources, including those not having the status of a personal data storage. It is proposed to legislatively stipulate the obligatory audit and to introduce a standard, stipulating a compromise between the technology abilities and advantages, accuracy and explainability of its result, and the rights of the participants of civil relations. Introduction

© Kharitonova Yu. S., 2023



This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

of certification of the artificial intelligence models, obligatory for application, will solve the issues of liability of the subjects obliged to apply such systems. In the context of professional liability of professional subjects, such as doctors, militants, or corporate executives of a juridical person, it is necessary to restrict the obligatory application of artificial intelligence if sufficient transparency is not provided.

Scientific novelty: the interdisciplinary character of the research allowed revealing the impossibility and groundlessness of the requirements to completely disclose the source code or architecture of the artificial intelligence models. The principle of artificial intelligence transparency may be satisfied through elaboration and provision of the right of the data subject and the subject, to whom the decision made as a result of automated data processing is addressed, to reject using automated data processing in decision-making, and the right to object to the decisions made in such a way.

Practical significance: is due to the actual absence of sufficient regulation of the principle of transparency of artificial intelligence and results of its functioning, as well as the content and features of the implementation of the right to explanation the right to objection of the decision subject. The most fruitful way to establish trust towards artificial intelligence is to recognize this technology as a part of a complex sociotechnical system, which mediates trust, and to improve the reliability of these systems. The main provisions and conclusions of the research can be used to improve the legal mechanism of providing transparency of the artificial intelligence models applied in state governance and business.

For citation

Kharitonova, Yu. S. (2023). Legal Means of Providing the Principle of Transparency of the Artificial Intelligence. *Journal of Digital Technologies and Law*, 1(2), 337–358. <https://doi.org/10.21202/jdtl.2023.14>

References

- Balduzzi, D., Frean, M., Leary, L., Lewis, J. P., Ma, K. W. D., & McWilliams, B. (2017). The shattered gradients problem: If resnets are the answer, then what is the question? In *International Conference on Machine Learning* (pp. 342–350). PMLR.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610–623).
- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency* (pp. 77–91). PMLR.
- Camilleri, M. A. (2018). Market segmentation, targeting and positioning. In *Travel marketing, tourism economics and the airline product* (pp. 69–83). New York: Springer.
- Cho, Y. H., Kim, J. K., & Kim, S. H. (2002). A personalized recommender system based on web usage mining and decision tree induction. *Expert systems with Applications*, 23(3), 329–342. [https://doi.org/10.1016/S0957-4174\(02\)00052-0](https://doi.org/10.1016/S0957-4174(02)00052-0)

- Cragun, B. J., & Steudel, H. J. (1987). A decision-table-based processor for checking completeness and consistency in rule-based expert systems. *International Journal of Man-Machine studies*, 26(5), 633–648. [https://doi.org/10.1016/s0020-7373\(87\)80076-7](https://doi.org/10.1016/s0020-7373(87)80076-7)
- Dyakonova, M. O., Efremov, A. A., Zaitsev, O. A., et al.; I. I. Kucheroval, S. A. Sinitsyna (Eds.). (2022). *Digital economy: topical areas of legal regulation: scientific-practical tutorial*. Moscow: IZISP, NORMA. (In Russ.).
- Foss-Solbrekk, K. (2021). Three routes to protecting AI systems and their algorithms under IP law: The good, the bad and the ugly. *Journal of Intellectual Property Law & Practice*, 16(3), 247–258. <https://doi.org/10.1093/jiplp/jpab033>
- Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). *Explaining and harnessing adversarial examples*. arXiv preprint arXiv:1412.6572.
- Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a “right to explanation”. *AI magazine*, 38(3), 50–57. <https://doi.org/10.1609/aimag.v38i3.2741>
- Gunning, D. (2017). Explainable artificial intelligence (XAI). *Defense advanced research projects agency (DARPA)*. nd Web. 2(2), 1.
- Hacker, P., Krestel, R., Grundmann, S., & Naumann, F. (2020). Explainable AI under contract and tort law: legal incentives and technical challenges. *Artificial Intelligence and Law*, 28(4), 415–439. <https://doi.org/10.1007/s10506-020-09260-6>
- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). New York: Springer.
- Hofmann, T., Schölkopf, B., & Smola, A. J. (2008). *Kernel methods in machine learning*.
- Kalis, B., Collier, M., & Fu, R. (2014). 10 promising AI applications in health care. *Harvard Business Review*.
- Kalpokas, I. (2019). *Algorithmic Governance. Politics and Law in the Post-Human Era*. Cham: Palgrave Pivot.
- Kalyatin, V. O. (2022). Deepfake as a legal problem: new threats or new opportunities? *Zakon*, 7, 87–103. (In Russ.). <https://doi.org/10.37239/0869-4400-2022-19-7-87-103>
- Kharitonova, Y. S., Savina, V. S., & Pagnini, F. (2021). Artificial Intelligence’s Algorithmic Bias: Ethical and Legal Issues. *Perm U. Herald Jurid. Sci*, 3(53), 488–515. <https://doi.org/10.17072/1995-4190-2021-53-488-515>
- Koene, A., Clifton, C., Hatada, Y., Webb, H., & Richardson, R. (2019). *A governance framework for algorithmic accountability and transparency*. Panel for the Future of Science and Technology.
- Kuteynikov, D. L., Izhaev, O. A., Zenin, S. S., & Lebedev, V. A. (2020). Algorithmic Transparency and Accountability: Legal Approaches to Solving the “Black Box” Problem. *Lex Russica*, 73(6), 139–148. (In Russ.). <https://doi.org/10.17803/1729-5920.2020.163.6.139-148>
- Lawrence, R. L., & Wright, A. (2001). Rule-based classification systems using classification and regression tree (CART) analysis. *Photogrammetric engineering and remote sensing*, 67(10), 1137–1142.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
- Maggiolino, M. (2018). *EU trade secrets law and algorithmic transparency*. Available at SSRN 3363178.
- Malgieri, G., & Comandé, G. (2017). Why a right to legibility of automated decision-making exists in the general data protection regulation. *International Data Privacy Law*, 7(4), 243–265. <https://doi.org/10.1093/idpl/ixp019>
- McEwen, R., Eldridge, J., & Caruso, D. (2018). Differential or deferential to media? The effect of prejudicial publicity on judge or jury. *The International Journal of Evidence & Proof*, 22(2), 124–143. <https://doi.org/10.1177/1365712718765548>
- Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, 2(1), 1–21. <https://doi.org/10.1186/s40537-014-0007-7>
- Pak, M., & Kim, S. (2017). A review of deep learning in image recognition. In *2017 4th international conference on computer applications and information processing technology (CAIPT)*. <https://doi.org/10.1109/caipt.2017.8320684>
- Payre, W., Cestac, J., & Delhomme, P. (2014). Intention to use a fully automated car: Attitudes and a priori acceptability. *Transportation research part F: traffic psychology and behavior*, 27, 252–263. <https://doi.org/10.1016/j.trf.2014.04.009>
- Plous, S. E. (2003). *Understanding prejudice and discrimination*. McGraw-Hill.
- Quillian, L. (2006). New approaches to understanding racial prejudice and discrimination. *Annu. Rev. Sociol.*, 32, 299–328. <https://doi.org/10.1146/annurev.soc.32.061604.123132>
- Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., & Müller, K. R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE*, 109(3), 247–278. <https://doi.org/10.1109/jproc.2021.3060483>

- Savelyev, A. I. (2021). *Scientific-practical article-by-article commentary to Federal Law "On personal data"* (2nd ed., amended and abridged). Moscow: Statut. (In Russ.).
- Selbst, A., & Powles, J. (2017). "Meaningful information" and the right to explanation. *International Data Privacy Law*, 7(4), 233–242. <https://doi.org/10.1093/idpl/ix022>
- Silkin, V. V. (2021). Transparency of executive power in digital epoch. *Russian Juridical Journal*, 4, 20–31. (In Russ.). https://doi.org/10.34076/20713797_2021_4_20
- Silver, D. et al. (2017). Mastering the game of go without human knowledge. *Nature*, 550(7676), 354.
- Vargas, R., Mosavi, A., & Ruiz, R. (2018). *Deep learning: a review*. Preprints.org. <https://doi.org/10.20944/preprints201810.0218.v1>
- von Eschenbach, W. J. (2021). Transparency and the black box problem: Why we do not trust AI. *Philosophy & Technology*, 34(4), 1607–1622.
- Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law*, 7(2), 76–99. <https://doi.org/10.1093/idpl/ix005>
- Wulf, A. J. & Seizov, O. (2022). "Please understand we cannot provide further information": evaluating content and transparency of GDPR-mandated AI disclosures. *AI & SOCIETY*, 1–22. <https://doi.org/10.1007/s00146-022-01424-z>
- Yampolskiy, R. V. (2019). *Unexplainability and incomprehensibility of artificial intelligence*. arXiv preprint arXiv:1907.03869
- 许可、刘畅. 论算法备案制度 // 人工智能. 2022. № 1. P. 66. [Xu Ke, Liu Chang. (2022). On the Algorithm Filing System. *Artificial Intelligence*, 1, 66.]

Author information



Yuliya S. Kharitonova – Doctor of Law, Professor, Professor of the Department of Entrepreneurial Law, Head of the Center for legal research of artificial intelligence and digital economy, Lomonosov Moscow State University

Address: 1 Leninskiye gory, 119991 Moscow, Russian Federation

E-mail: sovet2009@rambler.ru

ORCID ID: <https://orcid.org/0000-0001-7622-6215>

Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=57316440400>

Web of Science Researcher ID:

<https://www.webofscience.com/wos/author/record/708572>

Google Scholar ID: <https://scholar.google.ru/citations?user=61mQtb4AAAAJ>

RSCI Author ID: https://elibrary.ru/author_items.asp?authorid=465239

Conflict of interests

The author is a member of the Editorial Board of the Journal; the article has been reviewed on general terms.

Funding

The research was not sponsored.

Thematic rubrics

OECD: 5.05 / Law

PASJC: 3308 / Law

WoS: OM / Law

Article history

Date of receipt – March 6, 2023

Date of approval – April 13, 2023

Date of acceptance – June 16, 2023

Date of online placement – June 20, 2023

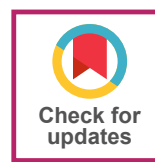


Научная статья

УДК 34.023:34.025:004.8

EDN: <https://elibrary.ru/immoam>

DOI: <https://doi.org/10.21202/jdtl.2023.15>



Будущее искусственного интеллекта: объект или субъект права?

Ирина Анатольевна Филипова

Национальный исследовательский Нижегородский государственный университет имени Н. И. Лобачевского
г. Нижний Новгород, Российская Федерация;
Самаркандский государственный университет
г. Самарканд, Республика Узбекистан

Вадим Дмитриевич Коротеев

Национальный исследовательский Нижегородский государственный университет имени Н. И. Лобачевского
г. Нижний Новгород, Российская Федерация

Ключевые слова

Генеративная модель, интеллектуальная система, искусственный интеллект, квазисубъект права, киберфизическая система, право, правосубъектность, робот, цифровые технологии, электронное лицо

Аннотация

Цель: выявление проблем, связанных с правовым регулированием общественных отношений, в которых используются системы искусственного интеллекта, и рациональное осмысление обсуждаемой правовой возможности наделения таких систем статусом субъекта права.

Методы: методологической основой исследования являются общенаучные методы анализа и синтеза, аналогии, абстрагирования и классификации. Среди преимущественно применяемых в работе юридических методов – формально-юридический, сравнительно-правовой и системно-структурный, а также методы толкования права и правового моделирования.

Результаты: представлен обзор состояния развития искусственного интеллекта и его внедрения на практике ко времени проведения исследования. Рассмотрено нормативно-правовое регулирование в данной области и разобраны основные из существующих концепций наделения искусственного интеллекта правосубъектностью (индивидуальная, коллективная и градиентная правосубъектность искусственного интеллекта). При этом дана оценка каждого из подходов и сделаны выводы о наиболее предпочтительных вариантах внесения изменений

Контактное лицо

© Филипова И. А., Коротеев В. Д., 2023



Статья находится в открытом доступе и распространяется в соответствии с лицензией Creative Commons «Attribution» («Атрибуция») 4.0 Всемирная (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0/deed.ru>), позволяющей неограниченно использовать, распространять и воспроизводить материал при условии, что оригинальная работа упомянута с соблюдением правил цитирования.

в действующее законодательство, которое перестает соответствовать реалиям. Растущее несоответствие связано с ускоряющимся развитием искусственного интеллекта и его распространением в различных секторах экономики, социальной сферы, а в ближайшей перспективе и в государственном управлении. Все это свидетельствует о повышении риска разрыва правовой материи с изменяющейся социальной реальностью.

Научная новизна: классифицированы научные подходы к наделению искусственного интеллекта правосубъектностью. В рамках каждого из подходов выделены ключевые моменты, использование которых позволит в дальнейшем создавать правовые конструкции на основе комбинирования, уходя от крайностей и соблюдая баланс интересов всех сторон. Оптимальным вариантом определения правового статуса искусственного интеллекта может стать внесение интеллектуальных систем в перечень объектов гражданских прав, но с дифференциацией правового регулирования искусственного интеллекта в качестве объекта права и «электронного агента» как квазисубъекта права. Линия разграничения должна проходить в зависимости от функциональных различий интеллектуальных систем, причем «электронным агентом» может быть признан не только робот, но и виртуальная интеллектуальная система.

Практическая значимость: материалы исследования могут быть применены в работе, связанной с подготовкой предложений по внесению изменений и дополнений в действующее законодательство, а также при разработке учебных курсов и написании учебных пособий по тематике, имеющей отношение к регулированию использования искусственного интеллекта.

Для цитирования

Филипова, И. А., Коротеев, В. Д. (2023). Будущее искусственного интеллекта: объект или субъект права? *Journal of Digital Technologies and Law*, 1(2), 359–386. <https://doi.org/10.21202/jdtl.2023.15>

Содержание

Введение

1. Правовой статус искусственного интеллекта: истоки проблемы
 - 1.1. Уровень и темп развития искусственного интеллекта
 - 1.2. Распространение технологий искусственного интеллекта на практике
 - 1.3. Система искусственного интеллекта как объект права
2. Основные концепции правосубъектности искусственного интеллекта
 - 2.1. Концепция индивидуальной правосубъектности искусственного интеллекта
 - 2.2. Концепция коллективной правосубъектности искусственного интеллекта
 - 2.3. Концепция градиентной правосубъектности искусственного интеллекта

Выводы

Список литературы

Введение

Сегодня человечество находится в периоде трансформации общества, связанном со сменой одного технологического уклада на другой, «умные» машины и программы достаточно быстро обучаются, растет способность систем искусственного интеллекта заменять людей во многих сферах деятельности. Одним из вопросов, все чаще поднимаемых в связи с совершенствованием технологий искусственного интеллекта, является вопрос о признании искусственных интеллектуальных систем субъектами права ввиду достижения такими системами уровня принятия полностью автономных решений и потенциального проявления «субъективной воли». Как гипотетический этот вопрос был сформулирован еще в XX в. (McNally & Inayatullah, 1988; Solum, 1992). В XXI в. градус научной дискуссии неуклонно повышается, достигая очередных экстремумов с каждым внедрением новых моделей искусственного интеллекта на практике, к примеру, с появлением беспилотных автомобилей на улицах или с представлением широкой публике роботов, обладающих новым набором функций (Bertolini & Episcopo, 2022).

Правовая проблема определения статуса искусственного интеллекта носит общетеоретический характер, что обусловлено объективной неспособностью предсказания всех возможных результатов разработки новых моделей искусственного интеллекта. Однако уже на данный момент системы искусственного интеллекта (ИИ-системы) фактически являются участниками некоторых общественных отношений, что требует расстановки «реперных точек», т. е. разрешения базовых вопросов в этой области с целью фиксации в законодательстве и тем самым уменьшения доли неопределенности в прогнозировании развития отношений, в которых задействованы системы искусственного интеллекта, на будущее.

Вынесенный в заглавие статьи вопрос о предполагаемой субъектности искусственного интеллекта в качестве предмета исследования, разумеется, охватывает не любые системы искусственного интеллекта, среди которых много «электронных ассистентов», не претендующих на правосубъектность в силу ограниченного набора функций и представляющих собой узкий (слабый) искусственный интеллект. Речь идет об «умных машинах» (киберфизических интеллектуальных системах) и генеративных моделях виртуальных интеллектуальных систем, чем дальше, тем больше приближающихся по своим возможностям к общему (сильному) искусственному интеллекту, сопоставимому с человеческим, а в перспективе и превосходящему его.

1. Правовой статус искусственного интеллекта: истоки проблемы

1.1. Уровень и темп развития искусственного интеллекта

Говорить об уровне развития искусственного интеллекта на сегодняшний день можно лишь условно, так как скорость его развития увеличивается, и то, что было актуальным на момент написания статьи, достаточно быстро устаревает. Особенно это касается наиболее быстро развивающегося направления в сфере искусственного интеллекта – искусственных нейронных сетей. Мультимодальные нейросети, например ChatGPT, DALL-e и т. д., интеллектуальные способности которых совершенствуются за счет наращивания числа параметров (воспринимаемых модальностей, включая недоступные человеку), а также использования для обучения огромных объемов

данных, которые человек не способен обработать физически, к началу 2023 г. подняли остроту вопроса о создании сильного искусственного интеллекта. К примеру, мультимодальные генеративные модели нейросетей способны создавать картины, литературные и научные тексты так, что далеко не всегда можно определить, сделано это человеком или системой искусственного интеллекта.

IT-эксперты говорят о двух качественных скачках: скачке скорости (периодичности появления качественно новых моделей), которая стала измеряться не годами, а максимум месяцами, и скачке непредсказуемости (невозможности точно спрогнозировать, что может произойти в сфере искусственного интеллекта даже до конца текущего года)¹. Модель ChatGPT-3 (третье поколение алгоритма обработки естественного языка от компании OpenAI) появилась в 2020 г. и могла обрабатывать текст, модель следующего поколения – ChatGPT-4, представленная компанией-разработчиком в марте 2023 г., умеет «работать» не только с текстами, но и с изображениями, в то время как модель последующего поколения уже обучается и будет способна на большее.

Если еще несколько лет назад о предполагаемом моменте технологической сингулярности, когда развитие машин становится фактически неуправляемым и необратимым, в корне меняя человеческую цивилизацию, упоминали как о времени, которое наступит минимум через несколько десятилетий, то сегодня увеличивается число исследователей – сторонников версии, что это может произойти гораздо быстрее². Имеется в виду появление так называемого сильного искусственного интеллекта, который будет демонстрировать возможности, сопоставимые с интеллектом человека, и способен решать аналогичный или даже более широкий круг задач. В отличие от слабого искусственного интеллекта сильный искусственный интеллект будет обладать сознанием, а одним из необходимых условий возникновения сознания у интеллектуальных систем является именно возможность осуществлять мультимодальное поведение с интеграцией данных из различных сенсорных модальностей (текст, изображение, видео, звук и т. д.), «привязывая» информацию разных модальностей к окружающей реальности с построением полноценных связанных «образов мира», как это свойственно людям³.

В марте 2023 г. более тысячи исследователей, IT-специалистов и предпринимателей в области искусственного интеллекта подписали открытое письмо, опубликованное на сайте американского научно-исследовательского центра Future of Life Institute⁴, специализирующегося на изучении экзистенциальных рисков для человечества. В письме содержался призыв приостановить обучение новых генеративных мультимодальных нейросетевых моделей в связи с тем, что отсутствие общих протоколов безопасности и правовой вакуум существенно повышают риски, ведь

¹ Карелов, С. (2023, 5 апреля). Телеграм-канал «Малоизвестное интересное». <https://t.me/s/theworldisnoteasy>

² David Shapiro (expert on artificial cognitive architecture) predicts. «AGI within 18 months». (2023, 28 марта). https://www.reddit.com/r/singularity/comments/1254azr/david_shapiro_expert_on_artificial_cognitive/

³ Колонин, А. (2021, 8 декабря). О глубине, прозрачности и «силе» ИИ в текущем моменте. <https://russiancouncil.ru/analytics-and-comments/analytics/o-glubine-prozrachnosti-i-sile-ii-v-tekushchem-momente/>

⁴ Pause Giant AI Experiments: An Open Letter. (2023, March 22). <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>

скорость развития технологий искусственного интеллекта резко повысилась из-за «ChatGPT революции». Отмечено также появление у моделей искусственного интеллекта не предвиденных их разработчиками способностей, не находящихся объяснения, и, вероятно, постепенное повышение доли таких способностей. Кроме того, подобная технологическая революция резко стимулирует создание интеллектуальных гаджетов, которые получают распространение, и новые поколения, сегодняшние дети, выросшие в постоянном общении с искусственными интеллектуальными помощниками, будут колоссально отличаться от предыдущих поколений.

Можно ли затормозить развитие искусственного интеллекта, чтобы человечество успело подстроиться под новые условия? Теоретически да, если все государства будут способствовать этому через национальные законодательства. Будут ли они это делать? Исходя из опубликованных национальных стратегий – нет, наоборот, каждым государством ставится задача выиграть конкуренцию (сохранить лидерство или сократить отставание). В Российской Федерации задача ускоренного развития технологий искусственного интеллекта поставлена в Национальной стратегии развития искусственного интеллекта до 2030 года, утвержденной Указом Президента Российской Федерации от 10 октября 2019 года № 490 «О развитии искусственного интеллекта в Российской Федерации»⁵ (далее – Национальная стратегия). Согласно пункту 24 Национальной стратегии, среди основных направлений заявлены: поддержка исследований для обеспечения опережающего развития искусственного интеллекта, разработка интеллектуального программного обеспечения, повышение доступности данных, необходимых для развития технологий искусственного интеллекта, и создание комплексной системы регулирования отношений, возникающих в связи с развитием и использованием искусственного интеллекта. В пункте 30 Национальной стратегии предусматривается, что для развития российских технологий необходимо поддерживать научные исследования, направленные на создание принципиально новых результатов, в том числе на создание сильного искусственного интеллекта. Аналогичная задача поставлена и в национальных стратегиях развития искусственного интеллекта других стран мира.

1.2. Распространение технологий искусственного интеллекта на практике

Возможности искусственного интеллекта привлекают предпринимателей, поэтому бизнес-структуры инвестируют огромные средства в новые разработки, и успех каждой новой модели стимулирует этот процесс. Объемы ежегодных инвестиций, учитывая, что в разработки вкладываются как частные компании, так и государства, растут, глобальный рынок решений в области искусственного интеллекта измеряется уже сотнями миллиардов долларов, а к 2030 г., по прогнозам, в частности содержащимся в Резолюции Европарламента от 3 мая 2022 г. «Об искусственном интеллекте в цифровую эпоху», вклад искусственного интеллекта в мировую экономику превысит 11 трлн евро⁶.

⁵ О развитии искусственного интеллекта в Российской Федерации: Указ Президента Российской Федерации от 10 октября 2019 года № 490. (2019). *Собрание законодательства Российской Федерации*, № 41. Ст. 5700.

⁶ *European Parliament Resolution of 3 May 2022 on Artificial Intelligence in a Digital Age (2020/2266(INI))*. https://www.europarl.europa.eu/doceo/document/TA-9-2022-0140_EN.html

Практико-ориентированность бизнеса влечет внедрение технологий искусственного интеллекта во все отрасли экономики. Искусственный интеллект используется как в добывающей, так и в обрабатывающей промышленности (металлургия, топливная и химическая промышленность, машиностроение, металлообработка и т. д.). Он применяется для прогнозирования эффективности разрабатываемых продуктов, автоматизации сборочных линий, снижения количества брака, улучшения логистики и предотвращения простоев.

Использование искусственного интеллекта в транспортной сфере включает как собственно автономные транспортные средства, так и оптимизацию маршрутов с учетом прогнозирования транспортных потоков, обеспечение безопасности вождения за счет предупреждения опасных ситуаций. Выпуск беспилотных автомобилей на дороги общего пользования – вопрос, активно обсуждаемый в парламентах разных стран мира. В Российской Федерации в 2021 г. Министерством транспорта также был подготовлен законопроект «О высокоавтоматизированных транспортных средствах и о внесении изменений в отдельные законодательные акты Российской Федерации»⁷, а годом позже Постановлением Правительства РФ от 29 декабря 2022 г. № 2495 утверждена «Программа экспериментального правового режима в сфере цифровых инноваций по предоставлению транспортных услуг с использованием высокоавтоматизированных транспортных средств на территориях отдельных субъектов Российской Федерации»⁸. На стыке транспортной сферы и сельского хозяйства растет применение автономных комбайнов, причем в сельскохозяйственном секторе процесс идет еще быстрее, так как там отсутствуют строгие нормативно-правовые ограничения, касающиеся автомобильного транспорта на дорогах общего пользования.

В банковской сфере ИИ-системы почти полностью заменили людей при оценке кредитоспособности заемщиков, все шире они используются для разработки новых банковских продуктов, повышения безопасности банковских операций.

Технологии искусственного интеллекта «захватывают» не только бизнес, но и социальную сферу: здравоохранение, образование, занятость. Применение искусственного интеллекта в медицине позволяет усовершенствовать диагностику, разработку новых лекарственных средств, проводить хирургические операции с использованием робототехники, в сфере образования – индивидуализировать уроки, автоматизировать оценку знаний обучающихся и профессиональных навыков учителей.

Сфера занятости сегодня все больше видоизменяется благодаря экспоненциальному росту платформенной занятости. Доля лиц, выполняющих работу с помощью платформ цифрового труда, дополненных искусственным интеллектом, по данным

⁷ Проект Федерального закона «О высокоавтоматизированных транспортных средствах и о внесении изменений в отдельные законодательные акты Российской Федерации» № 02/04/06-21/00116763. <https://base.garant.ru/56880577/>

⁸ Об установлении экспериментального правового режима в сфере цифровых инноваций и утверждении Программы экспериментального правового режима в сфере цифровых инноваций по предоставлению транспортных услуг с использованием высокоавтоматизированных транспортных средств на территориях отдельных субъектов Российской Федерации: Постановление Правительства Российской Федерации от 29.12.2022 № 2495. (2022, 30 декабря). Официальный интернет-портал правовой информации. <http://publication.pravo.gov.ru/Document/View/0001202212300090>

Международной организации труда, неуклонно повышается во всем мире⁹. Платформенная занятость не единственная составляющая трансформации сферы труда, серьезное влияние на нее оказывает увеличение уровня роботизации производства. По информации Международной федерации робототехники, рост числа промышленных роботов продолжается в общемировом масштабе, наиболее быстрые темпы роботизации присутствуют в странах Азии, в первую очередь в Китайской Народной Республике и Японии¹⁰. Россия значительно отстает в этой области, но именно на сокращение отставания рассчитан новый федеральный проект по развитию отечественной робототехники, который будет предусматривать правовые, налоговые и иные условия для развития производства и внедрения промышленных роботов. Этот федеральный проект, согласно поручению Президента Российской Федерации, должен быть подготовлен летом 2023 г. Проект будет включать перечень мер государственной поддержки производства и внедрения промышленных роботов «с учетом обеспечения ежегодного сокращения отставания количества таких роботов на 10 тысяч производственных рабочих в стране от среднемирового уровня»¹¹. Кроме того, идет работа по подготовке проекта Указа Президента Российской Федерации о внесении в Национальную стратегию развития искусственного интеллекта изменений, «направленных на повсеместное внедрение технологий искусственного интеллекта в отрасли экономики и социальной сферы и в системе государственного управления»¹².

Действительно, возможности искусственного интеллекта по анализу данных, используемые для управления производством, диагностической аналитики и прогнозирования, вызывают серьезную заинтересованность государств. Искусственный интеллект начинают внедрять в публичное управление. Сегодня активно ведется работа по созданию цифровых платформ для оказания государственных услуг, автоматизации многих процессов, связанных с выработкой решений государственными органами.

В публичном дискурсе все чаще применяются понятия «искусственная личность», «искусственная социальность», это свидетельствует о том, что разработка и реализация интеллектуальных систем перешли из чисто технической плоскости в область исследования многообразных способов их применения в гуманитарной и социокультурной сферах человеческой деятельности (Алексеев и др., 2023).

С учетом всего перечисленного можно утверждать, что искусственный интеллект глубже и глубже проникает в жизнь людей. Присутствие систем искусственного интеллекта в нашей жизни в ближайшие годы станет более заметным, оно увеличится как в рабочей среде, так и в общественном пространстве, сфере услуг и дома.

⁹ Перспективы занятости и социальной защиты в мире: Роль платформ цифрового труда в трансформации сферы труда. (2021). Группа технической поддержки по вопросам достойного труда и Бюро МОТ для стран Восточной Европы и Центральной Азии. Москва: МОТ.

¹⁰ World Robotics R&D Programs. (2023). https://ifr.org/downloads/papers/Executive_Summary_-_World_Robotics_RD_Programs_V02.pdf

¹¹ Перечень поручений Президента Российской Федерации по итогам конференции «Путешествие в мир искусственного интеллекта» (23–24 ноября 2022 года). (2023, 29 января). Пр-172, п. 1, подп. «е». <http://www.kremlin.ru/acts/assignments/orders/70418> (дата обращения: 20.04.2023)

¹² Перечень поручений Президента Российской Федерации по итогам конференции «Путешествие в мир искусственного интеллекта» (23–24 ноября 2022 года). (2023, 29 января). Пр-172, п. 5. <http://www.kremlin.ru/acts/assignments/orders/70418>

Чем дальше, тем больше искусственный интеллект будет обеспечивать повышение эффективности в достижении результатов через интеллектуальную автоматизацию различных процессов, создавая новые возможности и одновременно неся новые угрозы для людей, сообществ и государств.

С ростом уровня интеллекта ИИ-системы неизбежно становятся неотъемлемой частью общества, людям придется сосуществовать с ними. Подобный симбиоз будет включать сотрудничество людей и «умных» машин, а это, по словам нобелевского лауреата в области экономики Дж. Стиглица, приведет к трансформации цивилизации (Stiglitz, 2017). Уже сегодня, по мнению некоторых правоведов, «в целях повышения уровня благосостояния людей право не должно проводить различий между деятельностью людей и деятельностью искусственного интеллекта, когда люди и искусственный интеллект выполняют одни и те же задачи» (Abbott, 2020). Нужно учитывать и то, что развитие роботов-гуманоидов, приобретающих все более похожую на человеческую физиологию, вызовет выполнение ими среди прочего гендерных ролей в обществе в качестве партнеров (Karnouskos, 2022).

Государства вынуждены адаптировать законодательство к меняющимся общественным отношениям: количество законов, направленных на регулирование отношений, в которых в том или ином качестве задействованы системы искусственного интеллекта, в мире быстро растет. Согласно докладу Стэнфордского университета «Индекс искусственного интеллекта – 2023»¹³, если в 2016 г. был принят всего один закон, то в 2018 г. их было уже 12, в 2021 г. – 18, а в 2022 г. – 37 законов. Это подтолкнуло Организацию Объединенных Наций к формированию позиции по этике использования искусственного интеллекта на глобальном уровне. В сентябре 2022 г. появился документ, содержащий принципы этического использования искусственного интеллекта¹⁴ и основанный на Рекомендации по этике искусственного интеллекта, принятой годом ранее Генеральной конференцией ЮНЕСКО¹⁵. Тем не менее темпы развития и внедрения технологий искусственного интеллекта значительно превосходят скорость соответствующих изменений в сфере права.

Развитие технологий искусственного интеллекта запустило процесс создания машиночитаемого права, которое могут понимать ИИ-системы, более того, речь можно вести не только о машиночитаемости правовых норм, но также об их машинопроектируемости и машиноисполнимости. Применение в юридической деятельности ИИ-систем для качественной правовой аналитики и формулирования машинных рекомендаций для юристов – уже свершившийся факт (Ashley, 2017). Работы по созданию машиночитаемого права сегодня активно ведутся во многих странах мира, в 2021 г. комиссией по цифровому развитию при Правительстве РФ была утверждена российская концепция развития технологий машиночитаемого права¹⁶.

¹³ *AI Index Report 2023*. (2023). <https://aiindex.stanford.edu/report/>

¹⁴ *Principles for the Ethical Use of Artificial Intelligence in the United Nations System*. (2022, 20 сентября). <https://unsceb.org/principles-ethical-use-artificial-intelligence-united-nations-system>

¹⁵ *Recommendation on the Ethics of Artificial Intelligence*. (2021, 25 ноября). <https://unesdoc.unesco.org/ark:/48223/pf0000373434>

¹⁶ Концепция развития технологий машиночитаемого права, утверждена Правительственной комиссией по цифровому развитию, использованию информационных технологий для улучшения качества жизни и условий ведения предпринимательской деятельности, протокол от 15.09.2021 № 31. КонсультантПлюс. http://www.consultant.ru/document/cons_doc_LAW_396491/

1.3. Система искусственного интеллекта как объект права

В настоящее время системы искусственного интеллекта не обладают правосубъектностью и признаются объектами гражданских прав – это данность для любой, а не только для российской, национальной правовой системы. Независимо от достигнутого уровня развития искусственного интеллекта ИИ-система – это чье-то имущество. Соответственно, как виртуальная, так и киберфизическая ИИ-системы (две существующие разновидности искусственного интеллекта по форме его воплощения) – это то, по поводу чего возникают правоотношения.

Рассмотрим правовой режим искусственного интеллекта как объекта по российскому законодательству. В соответствии со статьей 128 Гражданского кодекса Российской Федерации¹⁷ (далее – ГК РФ) объектами права могут быть вещи, иное имущество, в том числе имущественные права, охраняемые результаты интеллектуальной деятельности, нематериальные блага и т. д.

Киберфизический по форме, т. е. обладающий «телом», искусственный интеллект (как правило, это робот) действующим законодательством признается вещью, но оно не устанавливает каких-то особенностей правового режима таких вещей и не ограничивает их оборотоспособность (Соменков, 2019). Представляется допустимой характеристика роботов по аналогии с неделимыми вещами в соответствии со статьей 133 ГК РФ, поскольку попытка отделения собственно искусственного интеллекта (в смысле программного обеспечения) от «тела» робота как оболочки влечет неизбежное изменение его назначения или вообще разрушение.

В международной частноправовой практике киберфизические ИИ-системы тоже имеют статус вещи в общем смысле и статус товара в коммерческом отношении. Так, Международная классификация товаров и услуг для регистрации знаков (МКТУ) прямо называет специфический вид товара: «Роботы человекоподобные с искусственным интеллектом» (класс 09, базовый № 090778)¹⁸.

Искусственный интеллект в виртуальной форме в Российской Федерации также не имеет особого правового положения и на данный момент фактически регулируется нормами, содержащимися в части 4 ГК РФ, относясь к непоименованным объектам авторских прав. Однако для эффективной юридической защиты такого объекта гражданских прав виртуальные ИИ-системы приходится признавать компьютерными программами, чтобы положения статьи 1259 ГК РФ, устанавливающие правовую охрану программ для ЭВМ, аналогичную литературным произведениям, распространялись и на них.

Определение самого понятия «искусственный интеллект» можно обнаружить в Федеральном законе от 24 апреля 2020 г. № 123-ФЗ «О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта в субъекте

¹⁷ Гражданский кодекс Российской Федерации. (1994, 5 декабря). *Собрание законодательства Российской Федерации*. № 32. Ст. 3301.

¹⁸ Международная классификация товаров и услуг для регистрации знаков (МКТУ) (11-я редакция, изд. 1-е). *Правовая система «Кодекс»*. <https://docs.cntd.ru/document/420273241>

Российской Федерации – городе федерального значения Москве и внесении изменений в статьи 6 и 10 Федерального закона «О персональных данных»¹⁹:

«Искусственный интеллект – комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые как минимум с результатами интеллектуальной деятельности человека. Комплекс технологических решений включает в себя информационно-коммуникационную инфраструктуру (в том числе информационные системы, информационно-телекоммуникационные сети, иные технические средства обработки информации), программное обеспечение (в том числе в котором используются методы машинного обучения), процессы и сервисы по обработке данных и поиску решений».

Исходя из процитированного фрагмента закона, можно заключить, что ИИ-система не соответствует определению программы для ЭВМ, содержащемуся в статье 1261 ГК РФ, такие системы не ограничены лишь совокупностью данных и команд, предназначенных для функционирования ЭВМ, т. е. имеют не только программную составляющую, поэтому ИИ-систему в целях дальнейшей правовой охраны следовало бы признать сложным объектом интеллектуальной собственности, предусмотренным статьей 1240 ГК РФ (Василевская и др., 2021).

Здесь нужно отметить, что если ранее любую систему искусственного интеллекта IT-специалисты определяли как программно-аппаратный комплекс, то сегодня аппаратная часть в виртуальных интеллектуальных системах может считаться отсутствующей и тогда проблема снимается. Так, к примеру, в определении искусственного интеллекта, данном в Коммюнике Европейской комиссии, упоминается, что виртуальная интеллектуальная система может вовсе не иметь собственной аппаратной части: «Искусственный интеллект (ИИ) относится к системам, которые демонстрируют разумное поведение, анализируя окружающую среду и предпринимая действия – с некоторой степенью автономии – для достижения определенных целей. Системы на основе ИИ могут быть чисто программными, действующими в виртуальном мире (например, голосовые помощники, программное обеспечение для анализа изображений, поисковые системы, системы распознавания речи и лиц) или ИИ может быть встроен в аппаратные устройства (например, роботы, дополненные искусственным интеллектом, автономные автомобили, дроны или приложения интернета вещей)»²⁰.

Отнесение ИИ-систем к объектам права не исключает возможности в дальнейшем установления законом особенностей их правового регулирования в зависимости от формы искусственного интеллекта – виртуальной или киберфизической, а также с учетом разницы в уровне искусственного интеллекта. Так, некоторые исследователи предлагают выделять в особую категорию продвинутые киберфизические

¹⁹ О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта в субъекте Российской Федерации – городе федерального значения Москве и внесении изменений в статьи 6 и 10 Федерального закона «О персональных данных»: Федеральный закон от 24.04.2020 № 123-ФЗ. (2020). *Собрание законодательства РФ*. № 17. Ст. 2701.

²⁰ Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions. *Artificial Intelligence for Europe*, Brussels, 25.04.2018 COM (2018) 237 final. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0237&from=EN>

системы, полагая, что влияние «умных» роботов на общество будет гораздо значительней, чем виртуальных систем, в силу сочетания нескольких факторов, в том числе наличия «тела» и эмерджентности, т. е. появления у системы свойств, не присущих ее компонентам (Calo, 2015).

Аргументированным можно признать мнение о необходимости применения к ИИ-системам правового режима, установленного для источников повышенной опасности. Сегодня суд при рассмотрении конкретного дела вправе самостоятельно определять, относится тот или иной объект к категории «источник повышенной опасности» или нет (Лаптев, 2019). Искусственный интеллект из-за возможности принятия автономных решений, отличных от изначально заложенной в него программы, соответствует определению источника повышенной опасности. Уточним, сказанное выше должно касаться не любых систем с элементами искусственного интеллекта, к которым можно отнести многие приложения из смартфона, а лишь продвинутых моделей. В этом ключе не лишено смысла выдвинутое В. В. Архиповым и В. Б. Наумовым предложение признать такие ИИ-системы имуществом особого рода, подробно урегулировав их правовой режим через закрепление статуса «имущества, способного действовать автономно» (Архипов, Наумов, 2017), особенно с учетом того, что, чем совершеннее становятся модели искусственного интеллекта, тем больше контрольных функций будет на них возлагаться (Greenstein, 2022).

Нельзя не упомянуть, что признание искусственного интеллекта объектом права предоставляет весьма ограниченный выбор вариантов решения ряда вопросов, важность которых со временем будет только расти. Среди этих вопросов:

1. Кто понесет ответственность за вред, причиненный действиями ИИ-систем, с учетом того, что они становятся все более самостоятельными?
2. За кем следует признавать права на результаты интеллектуальной деятельности творческого характера (принимая в расчет то, что уровень результатов становится все выше, а участие человека, даже опосредованное, фактически может сводиться к нулю)?

В ответах на эти вопросы пока фигурирует только человек – производитель, владелец или пользователь. Сохранится ли это положение в будущем? Применительно к ответственности за вред, причиненный ИИ-системой, «если раньше считалось, что не потребуются кардинальных изменений в регулировании института юридической ответственности, то сегодня оснований так говорить с абсолютной уверенностью уже нет. Причиной стало повышение степени автономности систем искусственного интеллекта при расширении спектра возможностей их применения» (Харитонов и др., 2022). Опять краеугольным камнем встает проблема растущей автономии ИИ-систем от человека, таким образом, изменение законодательства – вопрос времени и это изменение станет отражением нового «баланса интересов» (McCarty, 2017).

2. Основные концепции правосубъектности искусственного интеллекта

2.1. Концепция индивидуальной правосубъектности искусственного интеллекта

Переходя к рассмотрению концепций потенциального наделения искусственных интеллектуальных систем правосубъектностью, следует признать, что реализация любого из подобных подходов потребует фундаментального переустройства

сложившейся общей теории права и изменения ряда положений отдельных отраслей права. Примечательно, что приверженцы различных взглядов нередко используют термин «электронное лицо», поэтому использование данного наименования само по себе не позволяет установить, сторонником какой именно концепции является автор конкретного труда без ознакомления непосредственно с содержанием его работы.

Наиболее радикальным и, что закономерно, наименее популярным среди представителей научного сообщества подходом является концепция индивидуальной правосубъектности искусственного интеллекта. Сторонниками данного подхода выдвинута идея «полной инклюзивности» (*extreme inclusivism*), предполагающая наделение ИИ-систем аналогичным человеку правовым статусом с признанием за ними собственных интересов (Mulgan, 2019) и учетом их социальной значимости или социальной валентности (*social valence*). Последняя связана с тем, что «физическое воплощение робота имеет тенденцию заставлять человека относиться к движущемуся объекту так, как если бы он был живым. Это еще более заметно, когда робот имеет антропоморфные характеристики, поскольку сходство с человеческим телом заставляет людей проецировать эмоции, чувства удовольствия, боли и заботы, а также желание строить отношения» (Avila Negri, 2021). Проецирование человеческих эмоций на неодушевленные предметы не является чем-то новым, исходя из истории человечества, но применительно к роботам оно влечет за собой многочисленные последствия (Balkin, 2015).

В качестве предпосылок правового закрепления данной позиции обычно называются:

- достижение ИИ-системами уровня, сопоставимого с когнитивными функциями человеческого мышления;
- возрастающая степень сходства роботов и людей;
- гуманность, защита интеллектуальных систем от потенциальных «страданий».

Как видно из перечня предпосылок, все они имеют большую степень теоретизации и субъективной оценки. В частности, тенденция к созданию антропоморфных роботов (андроидов) вызвана обыденными психологическими и социальными потребностями людей, для которых комфортно ощущать себя в «обществе» похожих на них субъектов. Некоторые современные роботы обладают иными конструктивными особенностями из-за выполняемых ими функций, сюда можно отнести «многоногих» роботов-курьеров, для которых в первую очередь важна прочность конструкции и эффективное распределение веса перевозимого груза. В этом случае «включается» последняя из упомянутых выше предпосылок, вызванная формированием эмоциональных связей с роботами в человеческом сознании, аналогичных эмоциональным связям между домашним животным и его хозяином (Гринь, 2018).

Идея «полной инклюзивности» правового статуса ИИ-систем и человека отражена в работах некоторых исследователей-правоведов. Поскольку положения Конституции (например, положения глав 1 и 2 Конституции Российской Федерации²¹), а равно отраслевого законодательства не дают легального определения личности, само по себе понятие «личность» в конституционно-правовом смысле теоретически допускает расширительное толкование. В таком случае к личностям можно отнести любых носителей интеллекта, чьи когнитивные способности будут признаны

²¹ Конституция Российской Федерации. <https://base.garant.ru/10103000>

развитыми на достаточном уровне. Как отмечает А. В. Нечкин, логика данного подхода заключается в том, что сущностное отличие человека от всех остальных живых существ состоит как раз в его особом высокоразвитом интеллекте (Нечкин, 2020). Признание прав за ИИ-системами представляется очередным шагом эволюции правовой системы, постепенно расширившей юридическое признание на ранее подвергавшихся дискриминации людей, а теперь открывающей доступ и нечеловеческим существам (Gellers, 2021).

При наделении ИИ-систем подобным правовым статусом сторонникам рассматриваемого подхода видится обоснованным предоставление таким системам не буквальных прав граждан в устоявшемся конституционно-правовом понимании, а их аналогов и отдельных прав граждан с некоторыми изъятиями. Эта позиция основывается на объективных биологических различиях человека и робота. Так, не имеет смысла признание за ИИ-системой права на жизнь, поскольку она не живет в биологическом смысле. Права, свободы и обязанности систем искусственного интеллекта должны быть вторичными по отношению к правам граждан, данное положение закрепляет в правовом плане производность самого искусственного интеллекта в качестве творения человека.

Среди потенциальных конституционных прав и свобод искусственных интеллектуальных систем называются: право быть свободным, право на самосовершенствование (обучение и самообучение), право на неприкосновенность (защиту программного кода от произвольного вмешательства третьих лиц), свобода слова, свобода творчества, признание за ИИ-системой авторских прав и ограниченное право собственности. Также можно выделить и особенные права искусственного интеллекта, например, право на доступ к источнику электроэнергии.

В отношении обязанностей ИИ-систем предлагается конституционное закрепление трех знаменитых законов робототехники, сформулированных А. Азимовым: не причинение вреда человеку, а равно недопущение своим бездействием его причинения; повиновение всем приказам, отдаваемым человеком, кроме направленных на причинение вреда другому человеку; забота о собственной безопасности, за исключением предыдущих двух случаев (Наумов, Архипов, 2017). Еще ряд обязанностей в этом случае найдут отражение в нормах гражданского и административного права.

Концепция индивидуальной правосубъектности искусственного интеллекта имеет весьма небольшие шансы на свое нормативное закрепление сразу по нескольким основаниям.

Во-первых, критерий признания правосубъектности по наличию сознания и самосознания абстрактен, он допускает возможность многочисленных правонарушений, злоупотреблений правом и провоцирует социальные, политические проблемы в качестве дополнительного повода для расслоения общества. Данный тезис детализирован в работе С. Чопра и Л. Уайта, указывающих, что сознание и самосознание не являются необходимым и (или) достаточным условием для признания ИИ-систем в качестве субъекта права (Chopra & White, 2004). В правовой реальности полностью сознательные индивиды, например дети (или рабы в римском праве), лишены или ограничены в правосубъектности. В то же время люди, имеющие тяжелые психические расстройства, в том числе признанные недееспособными, или находящиеся в состоянии комы, т. е. при объективной невозможности проявления сознания, в первом случае остаются субъектами права (пусть и в ограниченной форме), а во втором имеют всю ту же полную правосубъектность, никак глобально не меняясь в своем

правовом статусе. При потенциальном же закреплении упомянутого выше критерия сознания и самосознания граждан можно будет в произвольном порядке лишать правосубъектности.

Во-вторых, ИИ-системы не смогут в устоявшемся правовом понимании осуществлять свои права и обязанности, поскольку действуют на основании изначально написанной программы, а принятие юридически значимых решений должно основываться на субъективном, нравственном выборе лица (Морхат, 2018б), его прямом волеизъявлении. Все моральные установки, чувства и желания такого «лица» становятся производными от человеческого интеллекта (Ужов, 2017). Автономность ИИ-систем в смысле способности принимать решения и реализовывать их самостоятельно, без внешнего антропогенного контроля или умышленного воздействия человека (Мусина, 2023), не является полноценной. Сегодня искусственный интеллект способен принимать только «квазинезависимые решения», так или иначе основанные на представлениях и моральных установках людей. В данном контексте можно рассматривать только «действие-операцию» ИИ-системы, исключая возможность реальной нравственной оценки поведения искусственного интеллекта (Петев, 2022).

В-третьих, признание за искусственным интеллектом правосубъектности индивидуального лица (тем более в форме приравнивания к статусу физического лица), влечет за собой разрушительное изменение сложившегося правопорядка и правовых традиций, формировавшихся со времен римского права, и провоцирует поднятие целого ряда фундаментально неразрешимых философских и правовых вопросов в области прав человека. Право как система социальных норм и социальное явление создано с учетом способностей человека и для обеспечения его интересов. Устоявшаяся антропоцентрическая система нормативных правил, международный консенсус в области концепции естественного права в случае закрепления подхода «полной инклюзивности» юридически и фактически будут признаны недействительными (Дремлюга, Дремлюга, 2019). Тем самым предоставление ИИ-системам, в частности «умным» роботам, правосубъектности может стать не решением существующих проблем, а ящиком Пандоры, обостряющим социальные и политические противоречия (Solaiman, 2017).

Еще один момент: в работах сторонников данной концепции упоминаются обычно лишь роботы, т. е. киберфизические системы искусственного интеллекта, которые будут взаимодействовать с людьми в физическом мире, а виртуальные системы остаются «за рамками», хотя сильный искусственный интеллект в случае его появления сможет найти воплощение и в виртуальной форме.

Исходя из всей совокупности приведенных выше аргументов, концепцию индивидуальной правосубъектности систем искусственного интеллекта следует рассматривать как юридически нереалистичную при сложившемся правопорядке.

2.2. Концепция коллективной правосубъектности искусственного интеллекта

Концепция коллективных лиц применительно к искусственным интеллектуальным системам нашла значительную поддержку среди сторонников самой допустимости признания подобной правосубъектности. Главное достоинство этого подхода заключается в исключении из юридической техники абстрактных понятий и оценочных суждений (сознание, самосознание, разумность, мораль и т. д.), подход базируется

на использовании юридической фикции в отношении искусственного интеллекта. Применительно к юридическим лицам уже существуют «передовые методы регулирования, которые можно было бы адаптировать для решения дилеммы правового статуса искусственного интеллекта» (Hárs, 2022).

Данная концепция не предполагает реального наделения ИИ-систем правосубъектностью физического лица, а является лишь расширением действия существующего института юридических лиц, предлагая создание новой категории юридических лиц – кибернетических «электронных организмов»²² (Мусина, 2023). В контексте данного подхода юридическое лицо целесообразнее рассматривать не в соответствии с современным узким понятием, в частности, закрепленным в статье 48 Гражданского кодекса РФ (как организацию, которая имеет обособленное имущество и отвечает им по своим обязательствам, может от своего имени приобретать и осуществлять гражданские права, нести гражданские обязанности, быть истцом и ответчиком в суде), а в более широком смысле, представляющем юридическое лицо в качестве любой отличной от физического лица правовой конструкции, которая в установленной законом форме наделяется правами и обязанностями. Тем самым сторонники данного подхода предлагают рассматривать юридическое лицо как субъект-сущность (идеальную сущность) по римскому праву (Санфилиппо, 2007).

Сходство ИИ-систем и юридических лиц видят в способе наделения их правосубъектностью – через обязательность государственной регистрации юридических лиц. Лишь после прохождения установленной процедуры регистрации юридическое лицо наделяется правоспособностью и дееспособностью, т. е. становится субъектом права. Такая модель удерживает рассуждения о правосубъектности ИИ-систем в рамках правового поля, исключая возможность признания правосубъектности по иным (неюридическим) основаниям, без естественных на это предпосылок, в то время как человек по праву своего рождения признается субъектом права.

Достоинством данной концепции является распространение на искусственные интеллектуальные системы требования о внесении сведений в соответствующие государственные реестры по аналогии с единым государственным реестром юридических лиц (Попова, 2018) как необходимого условия для придания им правосубъектности. Такой метод реализует значимую функцию систематизации всех юридических лиц и создания единой базы данных, необходимой как для органов государственной власти в целях осуществления контроля и надзора (к примеру, в налоговой сфере), так и для потенциальных контрагентов такого лица.

Объем прав юридических лиц в любой юрисдикции, как правило, меньше, чем физических лиц, тем самым использование данной конструкции для придания правосубъектности искусственному интеллекту не сопряжено с наделением его рядом прав, предлагаемых сторонниками предыдущей концепции.

При использовании приема юридической фикции в отношении юридических лиц предполагается, что за действиями юридического лица стоит объединение физических лиц, которое через органы управления юридическим лицом формирует

²² Не следует путать их с «юридическими электронными лицами» – децентрализованными автономными организациями, в которых координация деятельности участников происходит в соответствии с заранее согласованным набором правил с автоматическим контролем их выполнения (функционирующими на основе блокчейна).

его «волю» и осуществляет «волеизъявление». Иными словами, юридические лица являются искусственными (абстрактными) образованиями, призванными удовлетворять интересы физических лиц, которые выступили их учредителями или осуществляют управление ими. Так же и искусственные интеллектуальные системы создаются для удовлетворения потребностей конкретных лиц – разработчиков, операторов, владельцев. Физическое лицо при использовании ИИ-систем, их программировании руководствуется своими интересами, которые во внешнем мире такая система и представляет.

При теоретической оценке подобной модели регулирования не стоит забывать, что полная аналогия положения юридических лиц и ИИ-систем невозможна. Как уже было отмечено ранее, за всеми юридически важными действиями юридических лиц стоят лица физические, которые эти решения непосредственно и принимают. Воля, исходящая от юридического лица, всегда обусловлена и полностью контролируется волей физических лиц (Шуткин, 2020). Следовательно, без волеизъявления физических лиц невозможно осуществление деятельности лиц юридических, а применительно к ИИ-системам уже встает объективная проблема их автономности, возможности принятия решений без вмешательства физического лица после момента непосредственного создания такой системы (Ладенков, 2021).

Необходимо учитывать и то, что ИИ-системы не отвечают формальному признаку организационного единства, обязательного для юридических лиц. Правовой статус юридического лица вырабатывался на протяжении нескольких веков и так же, как и право в целом, обладает чертами «правового консерватизма». С другой стороны, действующее законодательство о юридических лицах сильно ограничивает возможности для наделения правами и обязанностями ИИ-системы, при попытке применить эту конструкцию следствием станет необоснованное законодательное препятствование инновациям (Понкин, Редькина, 2018), недопустимое исходя из содержания упоминавшихся ранее документов стратегического характера, нацеленных на скорейшее внедрение технологий искусственного интеллекта в различных отраслях экономики, социальной сфере и государственном управлении.

Таким образом, концепция коллективных лиц применительно к ИИ-системам пусть и обладает определенным потенциалом, не отвечает сложившимся юридическим традициям. Если же исходить из того, что «хотя вопрос о личности является бинарным» (признание лицом или непризнание таковым), однако «содержание этого статуса представляет собой спектр» возможных вариантов (Chesterman, 2020), то речь должна идти скорее о градиентной правосубъектности искусственного интеллекта, о которой будет подробнее рассказано в следующем разделе.

2.3. Концепция градиентной правосубъектности искусственного интеллекта

В связи с неустраняемыми ограничениями рассмотренных выше концепций значительным числом исследователей предлагаются свои подходы к решению вопроса о правовом положении искусственных интеллектуальных систем. Условно можно отнести их к различным вариациям концепции «градиентной правосубъектности», по выражению исследователя из Левенского университета Д. М. Мокану, которая подразумевает под этим ограниченную или частичную правоспособность и дееспособность ИИ-систем с оговоркой: термин «градиент» используется потому, что речь

идет не просто о включении или невключении каких-то прав и обязанностей в правовой статус, но о формировании их набора с минимально допустимым порогом, а также о признании такой правосубъектности только для определенных целей (Mocanu, 2021). Тогда к двум основным вариациям этой концепции можно отнести подходы, обосновывающие:

1) придание ИИ-системам особого правового статуса и включение в правовой порядок «электронных лиц» как абсолютно новой категории субъектов права;

2) наделение ИИ-систем ограниченной правоспособностью и дееспособностью в рамках гражданско-правовых отношений через создание категории «электронных агентов».

Позиция сторонников различных подходов внутри данной концепции может быть объединена, исходя из того, что пока отсутствуют онтологические причины, по которым искусственный интеллект следует рассматривать в качестве субъекта права, тем не менее в конкретных ситуациях уже наличествуют функциональные причины для наделения ИИ-систем конкретными правами и обязанностями, что «кажется лучшим способом поощрения индивидуальных и общественных интересов, которые закон призван защищать», наделяя эти системы «ограниченными и узкими формами правосубъектности» (Bertolini & Episcopo, 2022).

Придание системам искусственного интеллекта особого правового статуса через создание отдельного правового института «электронных лиц» главным своим достоинством имеет детальное прояснение и урегулирование отношений, возникающих:

- между юридическими или физическими лицами и ИИ-системами;
- между ИИ-системами и их разработчиками (оператором, владельцем);
- между третьей стороной и ИИ-системами в гражданско-правовых отношениях²³.

При подобной правовой конструкции ИИ-система будет контролироваться и управляться отдельно от ее разработчика, владельца или оператора (Морхат, 2018б). Приводя определение «электронного лица», П. М. Морхат акцентирует внимание на использовании уже упомянутого приема юридической фикции и функциональной направленности конкретной модели искусственного интеллекта: «электронное лицо» – это технико-юридический образ (обладающий некоторыми признаками юридической фикции по аналогии с юридическим лицом), отражающий и воплощающий условно специфическую правосубъектность системы искусственного интеллекта, отличающуюся в зависимости от ее функционально-целевого назначения и возможностей (Морхат, 2018а).

Как и концепция коллективных лиц в отношении ИИ-систем, данный подход предусматривает ведение специальных реестров «электронных лиц». Детальное и четкое изложение прав и обязанностей «электронных лиц» является основой для дальнейшего контроля со стороны государства и владельцев подобных ИИ-систем. Точно определенный круг полномочий, суженный объем правоспособности и дееспособности «электронных лиц» позволят отслеживать, не зашло ли данное «лицо» за рамки собственной программы вследствие потенциального принятия автономных решений и постоянного самообучения.

²³ Schrijver, S. de (2018, January 5). The Future Is Now: Legal Consequences of Electronic Personality for Autonomous Robots. *Who's Who Legal*. <https://whoswholegal.com/features/the-future-is-now-legal-consequences-of-electronic-personality-for-autonomous-robots>

Формально-юридически эта модель представляет из себя аналог наделения юридических лиц, к примеру в форме унитарных предприятий, ограниченной (специальной) правоспособностью по смыслу пункта 2 статьи 48 ГК РФ. Выдвигается и предложение о лицензировании отдельных видов «электронных лиц» в зависимости от осуществляемой ими деятельности по аналогии с лицензированием, предусмотренным Федеральным законом от 4 мая 2011 года № 99-ФЗ «О лицензировании отдельных видов деятельности»²⁴. При данном подходе искусственный интеллект, на стадии его создания являющийся объектом интеллектуальной собственности разработчиков программного обеспечения, после соответствующей сертификации, государственной регистрации может быть наделен правосубъектностью, но правоспособность и дееспособность «электронного лица» будут носить специальный характер (Вавилин, 2022).

Введение принципиально нового института для устоявшегося правопорядка повлечет за собой серьезные правовые последствия, необходимость глобального реформирования законодательства как минимум в области конституционного и гражданского права. Исследователи справедливо отмечают, что при введении понятия «электронного лица» необходимо проявлять осторожность, учитывая проблемы введения новых лиц в право, поскольку расширение понятия «лицо» в юридическом смысле может потенциально происходить за счет ограничения прав и законных интересов уже существующих субъектов правоотношений (Bryson et al., 2017). Учет таких аспектов представляется реально невозможным, так как правосубъектность физических лиц, юридических лиц, публично-правовых образований – это следствие многовековой эволюции теории государства и права.

Второй подход в рамках концепции градиентной правосубъектности – правовой концепт «электронных агентов», он связан в первую очередь с повсеместным распространением ИИ-систем как средств коммуникации между контрагентами и как инструментов онлайн-торговли. Данный подход можно назвать компромиссным, он признает невозможность наделения ИИ-систем статусом полноценных субъектов права, в то же время закрепляя за искусственным интеллектом отдельные (социально значимые) права и обязанности. Иными словами, концепция «электронных агентов» является легализацией квазисубъектности искусственного интеллекта. Под квазисубъектом права следует понимать определенное правовое явление, за которым на официальном либо доктринальном уровне признаются отдельные элементы правосубъектности при невозможности закрепления статуса полноценного субъекта права (Чаннов, 2022).

Сторонники данного подхода отмечают функциональные особенности ИИ-систем, позволяющие им выступать как в форме пассивного инструмента, так и в форме активного участника правоотношений, потенциально способного самостоятельно создавать юридически значимые для владельца такой системы соглашения. Именно поэтому ИИ-системы можно условно рассматривать в рамках отношений агентирования (Морхат, 2018б). При создании (или регистрации) ИИ-системы инициатор деятельности «электронного агента» заключает с ним фактический односторонний агентский договор, в результате чего «электронный агент» наделяется рядом полномочий, при осуществлении которых он может осуществлять юридические действия, значимые для принципала.

²⁴ О лицензировании отдельных видов деятельности: Федеральный закон от 04.05.2011 № 99-ФЗ (ред. от 29.12.2022). (2011). *Собрание законодательства РФ*, 2011. № 19. Ст. 2716.

Положения об агентировании отношений с ИИ-системами в России впервые прозвучали в связи с разработкой законопроекта «О внесении изменений в Гражданский кодекс Российской Федерации в части совершенствования правового регулирования отношений в области робототехники», подготовленного в 2016 г. Этот проект получил неформальное наименование «Закон Гришина» по фамилии Д. Гришина – основателя инвестиционного фонда Grishin Robotics и председателя совета директоров Mail.Ru Group. По смыслу законопроекта «электронным агентом» должен быть признан робот, который по решению собственника и в силу конструктивных особенностей предназначен для участия в гражданском обороте. Робот-агент имеет обособленное имущество и отвечает им по своим обязательствам, может от своего имени приобретать и осуществлять гражданские права и нести гражданские обязанности. В случаях, установленных законом, робот-агент может выступать в качестве участника гражданского процесса. Результатом одобрения данного законопроекта стала бы легализация ИИ-систем в качестве участников правоотношений в России.

В контексте «электронных агентов» наиболее проблемным является вопрос об обладании такими лицами обособленным имуществом, что дало бы возможность привлекать их к ответственности по приобретенным гражданско-правовым обязательствам. Авторы – разработчики законопроекта постарались учесть функциональную специфику разных видов искусственного интеллекта, предложив разделить ИИ-системы²⁵ на два типа:

- ИИ-системы (в проекте – роботы) как особая форма имущества, для которых потенциально возможна аналогия с животными и иными объектами права (ИИ-системы первого типа – объекты права);

- ИИ-агенты как участники гражданских правоотношений, обладающие особой правосубъектностью (ИИ-системы второго типа – квазисубъекты права).

Стоит отметить, что о возможном правовом статусе виртуальных систем как еще одной формы искусственного интеллекта авторы законопроекта 2016 г. не упоминали, указав лишь, что «Положения гражданского законодательства о роботах не применяются к программам для ЭВМ, которые, хотя и способны действовать, определять свои действия и оценивать их последствия без полного контроля со стороны человека по результатам обработки информации, поступающей из внешней среды, при этом не являются частью информационной системы обособленного устройства, предназначенного полностью или в части для осуществления фактических действий в автономном режиме»²⁶. В настоящее время достигнутый уровень разработок в области искусственного интеллекта вполне допускает предположение, что сильный искусственный интеллект сможет иметь и виртуальную форму, при этом управлять интеллектуально более слабыми киберфизическими системами, подчинив их себе.

²⁵ В проекте закона упоминается лишь о роботах, подобная неполнота формулировки была свойственна не только создателям данного проекта, но, например, и европейским разработчикам текста Резолюции ЕС от 16 февраля 2017 г. относительно норм гражданского права о робототехнике (European Parliament Resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics 2015/2013(INL)), позднее в резолюциях ЕС 2020 года данная неточность была устранена.

²⁶ Дмитрий Гришин представил проект регулирования правового статуса роботов в России. (2016, 17 декабря). <https://robotrends.ru/pub/1650/dmitriy-grishin-predstavil-proekt-regulirovaniya-pravovogo-statusa-robotov>

Подходы к наделению ИИ-систем элементами правосубъектности в рамках концепции, рассматриваемой в данном разделе, можно сравнить с конструктором, позволяющим выстроить новое, используя имеющийся материал из концепций, раскрытых в предыдущих разделах статьи, путем комбинирования, учета функциональных характеристик ИИ-систем и круга задач, для решения которых конкретная модель искусственного интеллекта предназначена. Исходя из этого, стоит считать градиентную концепцию имеющей наибольшие шансы на реализацию в рамках существующего правопорядка.

Выводы

Определение правового статуса искусственных интеллектуальных систем – вопрос, вызывающий с течением времени все более оживленные дискуссии в среде правоведов. Можно согласиться с У. Пагалло, одним из наиболее углубившихся в эту тему исследователей, что вряд ли в ближайшие годы будут найдены решения «для всех сложных случаев и дилемм», связанных с искусственным интеллектом, тем не менее, предотвращая «поляризацию сегодняшних дебатов, методы правовой гибкости и прагматического экспериментирования» позволяют рационально решить даже такие сложные задачи (Pagallo, 2018).

Рассмотрев сформулированные до настоящего момента основные концепции наделения ИИ-систем правосубъектностью, следует констатировать как минимум юридическую нецелесообразность признания за искусственным интеллектом статуса субъекта права в классическом понимании теории права, но равно и чем дальше, тем меньшую возможность сохранения правового режима объекта права в существующем варианте. В современных технологических, экономических, социальных, политических и юридических реалиях, скорее всего, придется использовать комбинированный подход для определения правового статуса искусственного интеллекта. Оптимальным вариантом может стать внесение ИИ-систем в перечень объектов гражданских прав, но с дифференциацией правового регулирования искусственного интеллекта в качестве объекта права и «электронного агента» как квазисубъекта права. Линия разграничения должна проходить в зависимости от функциональных различий ИИ-систем, причем «электронным агентом» может быть признан не только робот, но и виртуальная интеллектуальная система. «Электронный агент» наделяется отдельными правами и способен выполнять некоторые юридические обязанности, а в конечном счете ответственность за его действия должен нести человек. Признание ИИ-систем «электронными лицами» представляется преждевременным как минимум до появления сильного искусственного интеллекта.

В перспективе, с учетом начавшейся трансформации права, при условии, что уже тестируются ИИ-системы для машиночитаемого права и ИИ-системы поддержки принятия решений в области публичного управления, невозможно исключить вероятность постепенного увеличения влияния искусственного интеллекта на сферу права, что будет способствовать усилению позиций искусственного интеллекта на пути к признанию статуса субъекта права и к полной «перекройке» материи права с его участием или даже под его руководством, как бы фантастично это ни звучало на первый взгляд.

Список литературы

- Алексеев, А. Ю., Алексеева, Е. А., Емельянова, Н. Н. (2023). Искусственная личность в социально-политической коммуникации. *Искусственные общества*, 18(1), 1. <https://doi.org/10.18254/s207751800024370-6>
- Архипов, В. В., Наумов, В. Б. (2017). О некоторых вопросах теоретических оснований развития законодательства о робототехнике: аспекты воли и правосубъектности. *Закон*, 5, 157–170.
- Вавилин, Е. В. (2022). Статус искусственного интеллекта: от объекта к субъекту правовых отношений. *Вестник Томского государственного университета. Право*, 45, 147–158. <https://doi.org/10.17223/22253513/45/10>
- Василевская, Л. Ю., Подузова, Е. Б., Тасалов, Ф. А. (2021). *Цифровизация гражданского оборота: правовая характеристика «искусственного интеллекта» и «цифровых» субъектов (цивилистическое исследование)* (В 5 т. Т. 3). Москва: Проспект.
- Гринь, С. Н. (2018). Эмансипация роботов: элементы правосубъектности в конструкции искусственного интеллекта. *Бизнес. Общество. Власть*, 1(27), 233–242.
- Дремлюга, Р. И., Дремлюга, О. А. (2019). Искусственный интеллект – субъект права: аргументы за и против. *Правовая политика и правовая жизнь*, 2, 120–125.
- Ладенков, Н. Е. (2021). Модели наделения искусственного интеллекта правосубъектностью. *Вестник Балтийского федерального университета им. И. Канта. Серия: Гуманитарные и общественные науки*, 3, 12–20.
- Лаптев, В. А. (2019). Понятие искусственного интеллекта и юридическая ответственность за его работу. *Право. Журнал Высшей школы экономики*, 2, 79–102. <https://doi.org/10.17323/2072-8166.2019.2.79.102>
- Морхат, П. М. (2018а). Правосубъектность юнита искусственного интеллекта: некоторые гражданско-правовые подходы. *Вестник КГУ*, 3, 280–283.
- Морхат, П. М. (2018b). Юнит искусственного интеллекта как электронное лицо. *Вестник МГОУ. Серия: Юриспруденция*, 2, 61–73. <https://doi.org/10.18384/2310-6794-2018-2-61-73>
- Мусина, К. С. (2023). Идентификация правосубъектности искусственного интеллекта: кросснациональный анализ законодательств зарубежных стран. *Вестник Российского университета дружбы народов. Серия: Юридические науки*, 27(1), 135–147. <https://doi.org/10.22363/2313-2337-2023-27-1-135-147>
- Наумов, В. Б., Архипов, В. В. (2017). Проект Федерального закона «О внесении изменений в Гражданский кодекс Российской Федерации в части совершенствования правового регулирования отношений в области робототехники». В сб.: Н. А. Шевелёва (ред.), *Право и информация: вопросы теории и практики: сборник материалов VII Международной научно-практической конференции. Сер. «Электронное законодательство»*, 7 (с. 220–226). Санкт-Петербург: Президентская библиотека.
- Нечкин, А. В. (2020). Конституционно-правовой статус искусственного интеллекта в России: настоящее и будущее. *Lex Russica*, 8(165), 78–85. <https://doi.org/10.17803/1729-5920.2020.165.8.078-085>
- Петев, Н. И. (2022). Экзистенциальная, правовая и этическая проблемы искусственного интеллекта. *Векторы благополучия: экономика и социум*, 2(45), 55–70. <https://doi.org/10.18799/26584956/2022/2/1159>
- Понкин, И. В., Редькина, А. И. (2018). Искусственный интеллект с точки зрения права. *Вестник Российского университета дружбы народов. Серия: Юридические науки*, 1, 91–109. <https://doi.org/10.22363/2313-2337-2018-22-1-91-109>
- Попова, А. В. (2018). Новые субъекты информационного общества и общества знания: к вопросу о нормативном правовом регулировании. *Журнал российского права*, 11(263), 14–24. https://doi.org/10.12737/art_2018_11_2
- Санфилиппо, Ч. (2007). *Курс римского частного права: учебник* (пер. с итал. И. И. Маханькова, под общ. ред. Д. В. Дождева). Москва: Норма, 2007.
- Соменков, С. А. (2019). Искусственный интеллект: от объекта к субъекту?. *Вестник Университета имени О. Е. Кутафина*, 2(54), 75–85.
- Ужов, Ф. В. (2017). Искусственный интеллект как субъект права. *Пробелы в российском законодательстве*, 3, 357–360.
- Харитонов, Ю. С., Савина, В. С., Паньини, Ф. (2022). Гражданско-правовая ответственность при разработке и применении систем искусственного интеллекта и робототехники: основные подходы. *Вестник Пермского университета. Юридические науки*, 58, 683–708. <https://doi.org/10.17072/1995-4190-2022-58-683-708>

- Чаннов, С. Е. (2022). Робот (система искусственного интеллекта) как субъект (квазисубъект) права. *Актуальные проблемы российского права*, 17(12), 94–109. <https://doi.org/10.17803/1994-1471.2022.145.12.094-109>
- Шуткин, С. И. (2020). Возможна ли правосубъектность искусственного интеллекта. *Труды по интеллектуальной собственности*, 35(1–2), 90–137.
- Abbott, R. (2020). *The Reasonable Robot. Artificial Intelligence and the Law*. Cambridge University Press.
- Ashley, K. D. (2017). *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*. Cambridge University Press.
- Avila Negri, S. M. C. (2021). Robot as Legal Person: Electronic Personhood in Robotics and Artificial Intelligence. *Frontiers in Robotics and AI*, 8, Art. 789327. <https://doi.org/10.3389/frobt.2021.789327>
- Balkin, J. M. (2015). The Path of Robotics Law. *California Law Review*, 6, 45–60.
- Bertolini, A., & Episcopo, F. (2022). Robots and AI as Legal Subjects? Disentangling the Ontological and Functional Perspective. *Frontiers in Robotics and AI*, 9, Art. 842213. <https://doi.org/10.3389/frobt.2022.842213>
- Bryson, J. J., Diamantis, M. E., & Grant, Th. D. (2017). Of, For, and By the People: The Legal Lacuna of Synthetic Persons. *Artificial Intelligence and Law*, 25, 273–291.
- Calo, R. (2015). Robotics and the Lessons of Cyberlaw. *California Law Review*, 103(3), 513–563.
- Chesterman, S. (2020). Artificial Intelligence and the Limits of Legal Personality. *International & Comparative Law Quarterly*, 69, 819–844. <https://doi.org/10.1017/s0020589320000366>
- Chopra, S., & White, L. (2004). Artificial Agents – Personhood in Law and Philosophy. In *Proceedings of the 16th European Conference on Artificial Intelligence, ECAI'2004, including Prestigious Applicants of Intelligent Systems, PAIS 2004* (pp. 635–639). Valencia: IOS Press.
- Gellers, J. C. (2021). *Rights for Robots. Artificial Intelligence, Animal and Environmental Law*. London: Routledge.
- Greenstein, S. (2022). Preserving the rule of law in the era of artificial intelligence (AI). *Artificial Intelligence and Law*, 30, 291–323.
- Hárs, A. (2022). AI and international law – Legal personality and avenues for regulation. *Hungarian Journal of Legal Studies*, 62(4), 320–344. <https://doi.org/10.1556/2052.2022.00352>
- Karnouskos, S. (2022). Symbiosis with artificial intelligence via the prism of law, robots, and society. *Artificial Intelligence and Law*, 30, 93–115.
- McCarty, L. T. (2017). Finding the Right Balance in Artificial Intelligence and Law. In *Research Handbook on the Law of Artificial Intelligence* (Chapter 3, pp. 55–87). Edward Elgar Publishing.
- McNally, Ph., Inayatullah, S. (1988). The Rights of Robots: Technology, Culture and Law in the 21st Century. *Futures*, 20(1), 119–136. [https://doi.org/10.1016/0016-3287\(88\)90019-5](https://doi.org/10.1016/0016-3287(88)90019-5)
- Mocanu, D. M. (2021). Gradient Legal Personhood for AI Systems – Painting Continental Legal Shapes Made to Fit Analytical Molds. *Frontiers in Robotics and AI*, 8, Art. 788179. <https://doi.org/10.3389/frobt.2021.788179>
- Mulgan, T. (2019). Corporate Agency and Possible Futures. *Journal of Business Ethics*, 154, 901–916. <https://doi.org/10.1007/s10551-018-3887-1>
- Pagallo, U. (2018). Apples, oranges, robots: four misunderstandings in today's debate on the legal status of AI systems. *Philosophical Transactions of the Royal Society*, 376(2133), Art. 20180168. <https://doi.org/10.1098/rsta.2018.0168>
- Solaiman, S. M. (2017). Legal Personality of Robots, Corporations, Idols and Chimpanzees: A Quest for Legitimacy. *Artificial Intelligence and Law*, 25(2), 155–179. <https://doi.org/10.1007/s10506-016-9192-3>
- Solum, L. B. (1992). Legal Personhood for Artificial Intelligences. *North Carolina Law Review*, 70(4), 1231–1287.
- Stiglitz, J. E. (2017). The coming great transformation. *Journal of Policy Modeling*, 39(4), 625–638. <https://doi.org/10.1016/j.jpolmod.2017.05.009>

Сведения об авторах



Филипова Ирина Анатольевна – кандидат юридических наук, доцент, доцент кафедры трудового и экологического права, Национальный исследовательский Нижегородский государственный университет имени Н. И. Лобачевского; руководитель Центрально-Азиатского исследовательского центра регулирования искусственного интеллекта, Самаркандский государственный университет
Адрес: 603922, Российская Федерация, г. Нижний Новгород, проспект Гагарина, 23; 140104, Республика Узбекистан, г. Самарканд, Университетский бульвар, 15
E-mail: irinafilipova@yandex.ru
ORCID ID: <https://orcid.org/0000-0003-1773-5268>
Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=57327205000>
Web of Science Researcher ID: <https://www.webofscience.com/wos/author/record/R-1375-2016>
Google Scholar ID: <https://scholar.google.com/citations?user=opJc7fcAAAAJ>
РИНЦ Author ID: https://www.elibrary.ru/author_items.asp?authorid=461586



Коротеев Вадим Дмитриевич – студент 4-го курса юридического факультета, Национальный исследовательский Нижегородский государственный университет имени Н. И. Лобачевского
Адрес: 603922, Российская Федерация, г. Нижний Новгород, проспект Гагарина, 23
E-mail: redsert87@gmail.com
ORCID ID: <https://orcid.org/0009-0004-4182-2411>
Web of Science Researcher ID: <https://www.webofscience.com/wos/author/record/IAP-5405-2023>
РИНЦ Author ID: https://www.elibrary.ru/author_profile.asp?id=1198439

Вклад авторов

И. А. Филипова осуществляла формулирование идеи, исследовательских целей и задач; разработку методологии; анализ и обобщение литературы; интерпретацию общих результатов исследования; критический пересмотр и редактирование текста рукописи; формулировку ключевых выводов, предложений и рекомендаций; утверждение окончательного варианта статьи.

В. Д. Коротеев осуществлял составление черновика рукописи и его критический пересмотр с внесением ценных замечаний интеллектуального содержания; участие в научном дизайне; проведение сравнительного анализа; сбор литературы; подготовку и редактирование текста статьи; интерпретацию частных результатов исследования; оформление рукописи.

Конфликт интересов

И. А. Филипова является заместителем главного редактора журнала, статья прошла рецензирование на общих основаниях.

Финансирование

Исследование не имело спонсорской поддержки.

Тематические рубрики

Рубрика OECD: 5.05 / Law

Рубрика ASJC: 3308 / Law

Рубрика WoS: OM / Law

Рубрика ГРНТИ: 10.07.45 / Право и научно-технический прогресс

Специальность ВАК: 5.1.1 / Теоретико-исторические правовые науки

История статьи

Дата поступления – 23 апреля 2023 г.

Дата одобрения после рецензирования – 8 мая 2023 г.

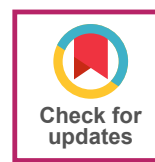
Дата принятия к опубликованию – 16 июня 2023 г.

Дата онлайн-размещения – 20 июня 2023 г.



Research article

DOI: <https://doi.org/10.21202/jdtl.2023.15>



Future of the Artificial Intelligence: Object of Law or Legal Personality?

Irina A. Filipova  

National Research Lobachevsky State University of Nizhny Novgorod
Nizhny Novgorod, Russian Federation;
Samarkand State University
Samarkand, Republic of Uzbekistan

Vadim D. Koroteev  

National Research Lobachevsky State University of Nizhny Novgorod
Nizhny Novgorod, Russian Federation

Keywords

Artificial intelligence,
cyberphysical system,
digital technologies,
electronic person,
generative model,
intellectual system,
law,
legal personality,
quasi subject of law,
robot

Abstract

Objective: to reveal the problems associated with legal regulation of public relations, in which artificial intelligence systems are used, and to rationally comprehend the possibility of endowing such systems with a legal subject status, which is being discussed by legal scientists.

Methods: the methodological basis of the research are the general scientific methods of analysis and synthesis, analogy, abstraction and classification. Among the legal methods primarily applied in the work are formal-legal, comparative-legal and systemic-structural methods, as well as the methods of law interpretation and legal modeling.

Results: the authors present a review of the state of artificial intelligence development and its introduction into practice by the time of the research. Legal framework in this sphere is considered; the key current concepts of endowing artificial intelligence with a legal personality (individual, collective and gradient legal personality of artificial intelligence) are reviewed. Each approach is assessed; conclusions are made as to the most preferable

 Corresponding author

© Filipova I. A., Koroteev V. D., 2023



This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

amendments in the current legislation, which ceases to correspond to the reality. The growing inconsistency is due to the accelerated development of artificial intelligence and its spreading in various sectors of economy, social sphere, and in the nearest future – in public management. All this testifies to the increased risk of a break between legal matter and the changing social reality.

Scientific novelty: scientific approaches are classified which endow artificial intelligence with a legal personality. Within each approach, the key moments are identified, the use of which will allow in the future creating legal constructs based on combinations, avoiding extremes and observing the balance between the interests of all parties. The optimal variant to define the legal status of artificial intelligence might be to include intellectual systems into a list of civil rights objects, but differentiating the legal regulation of artificial intelligence as an object of law and an “electronic agent” as a quasi subject of law. The demarcation line should be drawn depending on the functional differences between intellectual systems, while not only a robot but also a virtual intellectual system can be considered an “electronic agent”.

Practical significance: the research materials can be used when preparing proposals for making amendments and additions to the current legislation, as well as when elaborating academic course and writing tutorials on the topics related to regulation of using artificial intelligence.

For citation

Filipova, I. A., Koroteev, V. D. (2023). Future of the Artificial Intelligence: Object of Law or Legal Personality? *Journal of Digital Technologies and Law*, 1(2), 359–386. <https://doi.org/10.21202/jdtl.2023.15>

References

- Abbott, R. (2020). *The Reasonable Robot. Artificial Intelligence and the Law*. Cambridge University Press.
- Alekseev, A., Alekseeva, E., & Emelyanova, N. (2023). Artificial Personality in socio-political communication. *Artificial Societies*, 18(1). (In Russ.). <https://doi.org/10.18254/s207751800024370-6>
- Arkhipov, V. V., & Naumov, V. B. (2017). On certain issues of theoretic grounds for development of robotics legislation: the aspects of will and legal personality. *Statute*, 5, 157–170. (In Russ.).
- Ashley, K. D. (2017). *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*. Cambridge University Press.
- Avila Negri, S. M. C. (2021). Robot as Legal Person: Electronic Personhood in Robotics and Artificial Intelligence. *Frontiers in Robotics and AI*, 8, Art. 789327. <https://doi.org/10.3389/frobt.2021.789327>
- Balkin, J. M. (2015). The Path of Robotics Law. *California Law Review*, 6, 45–60.
- Bertolini, A., & Episcopo, F. (2022). Robots and AI as Legal Subjects? Disentangling the Ontological and Functional Perspective. *Frontiers in Robotics and AI*, 9, Art. 842213. <https://doi.org/10.3389/frobt.2022.842213>
- Bryson, J. J., Diamantis, M. E., & Grant, Th. D. (2017). Of, For, and By the People: The Legal Lacuna of Synthetic Persons. *Artificial Intelligence and Law*, 25, 273–291.
- Calo, R. (2015). Robotics and the Lessons of Cyberlaw. *California Law Review*, 103(3), 513–563.
- Channov, S. E. (2022). Robot (Artificial Intelligence System) as a Subject (Quasi-Subject) of Law. *Actual Problems of Russian Law*, 17(12), 94–109. (In Russ.). <https://doi.org/10.17803/1994-1471.2022.145.12.094-109>

- Chesterman, S. (2020). Artificial Intelligence and the Limits of Legal Personality. *International & Comparative Law Quarterly*, 69, 819–844. <https://doi.org/10.1017/s0020589320000366>
- Chopra, S., & White, L. (2004). Artificial Agents – Personhood in Law and Philosophy. In *Proceedings of the 16th European Conference on Artificial Intelligence, ECAI'2004, including Prestigious Applicants of Intelligent Systems, PAIS 2004* (pp. 635–639). Valencia: IOS Press.
- Dremliuga, R. I., & Dremliuga, O. A. (2019). Artificial intelligence – a legal person: the arguments for and against. *Pravovaya politika i pravovaya zhizn*, 2, 120–125. (In Russ.).
- Gellers, J. C. (2021). *Rights for Robots. Artificial Intelligence, Animal and Environmental Law*. London: Routledge.
- Greenstein, S. (2022). Preserving the rule of law in the era of artificial intelligence (AI). *Artificial Intelligence and Law*, 30, 291–323.
- Grin, S. N. (2018). Robots' emancipation: elements of legal personhood in the construct of artificial intelligence. *Biznes. Obshchestvo. Vlast*, 1(27), 233–242. (In Russ.).
- Hárs, A. (2022). AI and international law – Legal personality and avenues for regulation. *Hungarian Journal of Legal Studies*, 62(4), 320–344. <https://doi.org/10.1556/2052.2022.00352>
- Karnouskos, S. (2022). Symbiosis with artificial intelligence via the prism of law, robots, and society. *Artificial Intelligence and Law*, 30, 93–115.
- Kharitonova, Yu. S., Savina, V. S., & Pagnini, F. (2022). Civil liability in the development and application of artificial intelligence and robotic systems: basic approaches. *Perm University Herald. Juridical sciences*, 4(58), 683–708. (In Russ.). <https://doi.org/10.17072/1995-4190-2022-58-683-708>
- Ladenkov, N. S. (2021). Models of endowing artificial intelligence with legal personality. *Vestnik IKBFU. Humanities and Social Sciences*, 3, 12–20. (In Russ.).
- Laptev, V. A. (2019). Artificial Intelligence and Liability for its Work. *Pravo. Zhurnal Vysshey Shkoly Ekonomiki*, 2, 79–102. (In Russ.). <https://doi.org/10.17323/2072-8166.2019.2.79.102>
- McCarty, L. T. (2017). Finding the Right Balance in Artificial Intelligence and Law. In *Research Handbook on the Law of Artificial Intelligence* (Chapter 3, pp. 55–87). Edward Elgar Publishing.
- McNally, Ph. & Inayatullah, S. (1988). The Rights of Robots: Technology, Culture and Law in the 21st Century. *Futures*, 20(1), 119–136. [https://doi.org/10.1016/0016-3287\(88\)90019-5](https://doi.org/10.1016/0016-3287(88)90019-5)
- Mocanu, D. M. (2021). Gradient Legal Personhood for AI Systems – Painting Continental Legal Shapes Made to Fit Analytical Molds. *Frontiers in Robotics and AI*, 8, Art. 788179. <https://doi.org/10.3389/frobt.2021.788179>
- Morkhat, P. M. (2018a). Artificial intelligence unit as electronic personality. *Bulletin of the Moscow State Regional University (Jurisprudence)*, 2, 61–73. (In Russ.). <https://doi.org/10.18384/2310-6794-2018-2-61-73>
- Morkhat, P. M. (2018b). Legal personality of artificial intelligence unit: some civil-legal approaches. *Bulletin of Kostroma State University*, 3, 280–283. (In Russ.).
- Mulgan, T. (2019). Corporate Agency and Possible Futures. *Journal of Business Ethics*, 154, 901–916. <https://doi.org/10.1007/s10551-018-3887-1>
- Musina, K. S. (2023). Theoretical aspects of identifying legal personality of artificial intelligence: cross-national analysis of the laws of foreign countries. *RUDN Journal of Law*, 27(1), 135–147. (In Russ.). <https://doi.org/10.22363/2313-2337-2023-27-1-135-147>
- Naumov, V. B., & Arkhipov, V. V. (2017). Draft Federal law “On amendments to the Civil Code of the Russian Federation in improving the legal regulation of relations in the field of robotics”. In N. A. Sheveleva (Ed.), *Law and Information: the Questions of Theory and Practice: Collection of works of international scientific and practical conference*. Saint Petersburg: The Presidential Library. (In Russ.).
- Nechkin, A. V. (2020). Constitutional and Legal Status of Artificial Intelligence in Russia: Present and Future. *Lex Russica*, 8, 78–85. (In Russ.). <https://doi.org/10.17803/1729-5920.2020.165.8.078-085>
- Pagallo, U. (2018). Apples, oranges, robots: four misunderstandings in today's debate on the legal status of AI systems. *Philosophical Transactions of the Royal Society*, 376(2133), Art. 20180168. <https://doi.org/10.1098/rsta.2018.0168>
- Petev, N. I. (2022). Existential, legal and ethical problems of artificial intelligence. *Journal of Wellbeing Technologies*, 2(45), 55–70. (In Russ.). <https://doi.org/10.18799/26584956/2022/2/1159>
- Ponkin, I. V., & Redkina A. I. (2018). Artificial Intelligence from the Point of View of Law. *RUDN Journal of Law*, 22(1), 91–109. (In Russ.). <https://doi.org/10.22363/2313-2337-2018-22-1-91-109>
- Popova, A. V. (2018). New Subjects of the Information Society and the Knowledge Society: To the Question of Legal Regulation. *Journal of Russian Law*, 6(11), 14–24. (In Russ.). https://doi.org/10.12737/art_2018_11_2
- Sanfilippo, Ch. (2007). *Course in Roman private law: tutorial* (transl. by I. I. Makhankov, D. V. Dozhdev (Ed.)). Moscow: Norma. (In Russ.).

- Shutkin, S. I. (2020). Is legal personhood of AI possible? *Works on Intellectual Property*, 35(1–2), 90–137. (In Russ.).
- Solaiman, S. M. (2017). Legal Personality of Robots, Corporations, Idols and Chimpanzees: A Quest for Legitimacy. *Artificial Intelligence and Law*, 25(2), 155–179. <https://doi.org/10.1007/s10506-016-9192-3>
- Solum, L. B. (1992). Legal Personhood for Artificial Intelligences. *North Carolina Law Review*, 70(4), 1231–1287.
- Somenkov, S. A. (2019). Artificial intelligence: from object to subject? *Courier of the Kutafin Moscow State Law University*, 2(54), 75–85. (In Russ.).
- Stiglitz, J. E. (2017). The coming great transformation. *Journal of Policy Modeling*, 39(4), 625–638. <https://doi.org/10.1016/j.jpolmod.2017.05.009>
- Uzhov, F. V. (2017). Legal personality of artificial intelligence. *Gaps in Russian Legislation*, 3, 357–360. (In Russ.).
- Vasilevskaya, L. Yu., Poduzova, E. B., & Tasalov, F. A. (2021). *Digitalization of civil turnover: legal characteristics of "artificial intelligence" and "digital" subjects (civilistic research)* (In 5 vol. Vol. 3). Moscow: Prospect. (In Russ.).
- Vavilin, E. V. (2022). The status of artificial intelligence: from object to the subject of legal relations. *Vestnik Tomskogo Gosudarstvennogo Universiteta. Pravo*, 45, 147–158. (In Russ.). <https://doi.org/10.17223/22253513/45/10>

Authors information



Irina A. Filipova – Candidate of Sciences in Jurisprudence, Associate Professor, Associate Professor of the Department of Labor Law and Environmental Law, National Research Lobachevsky State University of Nizhny Novgorod, Head of Central Asia research center for artificial intelligence regulation, Samarkand State University
Address: 23 prospekt Gagarina, Nizhniy Novgorod 603922, Russian Federation; 15 Universitetskiy boulevard, Samarkand 140104, Republic of Uzbekistan
E-mail: irinafilipova@yandex.ru
ORCID ID: <https://orcid.org/0000-0003-1773-5268>
Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=57327205000>
Web of Science Researcher ID: <https://www.webofscience.com/wos/author/record/R-1375-2016>
Google Scholar ID: <https://scholar.google.com/citations?user=opJc7fcAAAAJ>
RSCI Author ID: https://www.elibrary.ru/author_items.asp?authorid=461586



Vadim D. Koroteev – 4th year student of Law Faculty, National Research Lobachevsky State University of Nizhny Novgorod
Address: 23 prospekt Gagarina, Nizhniy Novgorod 603922, Russian Federation
E-mail: redsert87@gmail.com
ORCID ID: <https://orcid.org/0009-0004-4182-2411>
Web of Science Researcher ID: <https://www.webofscience.com/wos/author/record/IAP-5405-2023>
RSCI Author ID: https://www.elibrary.ru/author_profile.asp?id=1198439

Authors' contributions

Irina A. Filipova formulated the research idea, goals and objectives; elaborated the methodology; analyzed and summarized literature; interpreted general research results; critically reviewed and edited the manuscript text; formulated the key conclusions, proposals and recommendations; approved the final variant of the article. Vadim D. Koroteev compiled the manuscript draft and critically reviewed it, adding valuable comments on the intellectual content; participated in scientific design; performed comparative analysis; selected literature; prepared and edited the manuscript text; interpreted the specific research results; finalized the manuscript.

Conflict of interests

I. A. Filipova is a Deputy Editor-in-Chief of the Journal; the article has been reviewed on general terms.

Funding

The research was not sponsored.

Thematic rubrics:

OECD: 5.05 / Law

PASJC: 3308 / Law

WoS: OM / Law

Article history

Date of receipt – April 23, 2023

Date of approval – May 8, 2023

Date of acceptance – June 16, 2023

Date of online placement – June 20, 2023

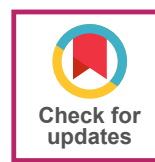


Научная статья

УДК 347.1:004.8

EDN: <https://elibrary.ru/ktizpw>

DOI: <https://doi.org/10.21202/jdtl.2023.16>



Алгоритмическая дискриминация и защита неприкосновенности частной жизни

Элена Фаллетти 

Университет Карло Каттанео

г. Кастелланца, Итальянская Республика

Ключевые слова

Алгоритм,
дискриминация,
защита данных,
искусственный интеллект,
неприкосновенность,
персональные данные,
право,
регулирование,
цифровые технологии,
частная жизнь

Аннотация

Цель: появление цифровых технологий, таких как искусственный интеллект, стало вызовом для государств всего мира. Оно породило множество рисков, связанных с нарушением прав человека, включая права на неприкосновенность частной жизни и человеческое достоинство. Это определяет актуальность данного исследования. Цель статьи – проанализировать роль алгоритмов в случаях дискриминации и выяснить, каким образом алгоритмы могут способствовать предубежденности при принятии решений на основе персональных данных. Проведенный анализ помогает оценить проект закона об искусственном интеллекте, направленный на регулирование данной проблемы для предотвращения дискриминации при использовании алгоритмов.

Методы: в работе применялись методы эмпирического и сравнительного анализа. Сравнительный анализ позволил выявить сходства и различия существующего регулирования и положений проекта закона об искусственном интеллекте. С помощью эмпирического анализа рассмотрены реальные примеры алгоритмической дискриминации.

Результаты: результаты исследования показывают, что Закон об искусственном интеллекте нуждается в доработке, так как он остается на уровне дефиниций и недостаточно опирается на эмпирический материал. Автор выдвигает ряд предложений по совершенствованию данного законопроекта.

Научная новизна: заключается в мультидисциплинарности данной работы, рассматривающей вопросы дискриминации, защиты данных и влияния на эмпирическую реальность в сфере алгоритмической дискриминации и охраны неприкосновенности частной жизни.

© Фаллетти Э., 2023



Статья находится в открытом доступе и распространяется в соответствии с лицензией Creative Commons «Attribution» («Атрибуция») 4.0 Всемирная (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0/deed.ru>), позволяющей неограниченно использовать, распространять и воспроизводить материал при условии, что оригинальная работа упомянута с соблюдением правил цитирования.

Практическая значимость: состоит в привлечении внимания к тому факту, что алгоритмы выполняют инструкции, составленные на основе сообщенных им данных. Не имея возможностей для абдукции, алгоритмы действуют лишь как послушные исполнители приказов. Результаты работы могут использоваться в качестве основы для будущих исследований в данной области и в законотворческом процессе.

Для цитирования

Фаллетти, Э. (2023). Алгоритмическая дискриминация и защита неприкосновенности частной жизни. *Journal of Digital Technologies and Law*, 1(2), 387–420. <https://doi.org/10.21202/jdtl.2023.16>

Содержание

Введение

1. Дискриминация и обращение с персональными данными
2. Предубеждение и дискриминация в системах автоматизированного принятия решений
3. Двойственная природа алгоритмов оценки риска
4. Предложения к проекту закона об искусственном интеллекте

Заключение

Список литературы

Введение

Промышленная революция XVIII в. стала провозвестником социальных перемен и развития свобод путем внедрения автоматизации в производственные процессы. Она способствовала тому, что продукция стала производиться в специально отведенных для этого местах, а основная деятельность была перенесена из сельской местности в городские центры. Она также вызвала социальную трансформацию общества, переместив массы в чуждую реальность: фабрики уничтожали менталитет, поведение, привычки, характерные для прежнего сельскохозяйственного уклада, укорененные в культуре и коллективной памяти (Tunzelmann, 2003; Freeman et al., 2001).

С точки зрения взаимоотношений между правом и технологиями эти изменения оказали огромное влияние на закон и общество, поскольку технологический прогресс способствует унификации, стандартизируя нормы поведения в новом производственном процессе. Происходит переход к установлению системы норм, которая, как техническое предписание, обязывает каждого субъекта приводить свое поведение в соответствие со стандартами, принимающими в итоге значение закона.

Этот контекст позволяет провести исторически диахроническое сравнение между автоматизацией, начавшейся в XIX в. на фабриках Британии, и алгоритмической автоматизацией в наши дни. Это сравнение будет касаться измерения времени при выполнении работ. Как и в прошлом, первая область, в которой проявляются и развиваются инновации в человеческой деятельности, – это профессиональная деятельность. Этот факт полностью относится к автоматизации: с одной стороны,

труд – божественное наказание – дает возможность человеку вести независимую жизнь; с другой – автоматизация труда позволяет предпринимателям и фабрикантам умножать свои сбережения, угнетая рабочих, роль которых уподобляется простым машинам (Marx, 2016). Осуществляется марксистское проклятие «отчужденного труда» (Marx, 2016; Claeyes, 2018).

Измерение рабочего времени для сокращения затрат является одной из главнейших забот при управлении такой сложной организацией, как промышленное предприятие: автоматизация производственного процесса в XIX в. позволила увеличить рабочий день за пределы светового дня (благодаря электричеству) и устранить зависимость от погодных условий, отделив труд от сельскохозяйственных ритмов; алгоритмическая автоматизация действует еще более экстремально, так как алгоритм может управлять производственным процессом без участия человека, достигая того типа отчуждения, который предсказывал Маркс, – отчуждения каждого отдельного работника от остальных (Marx, 2016).

Применяя теорию Маркса к продолжительности рабочего дня (Marx, 2016; Claeyes, 2018), как *illo tempore* [«во времена оны», в прошлом. – Прим. перев.], так и в наши дни, можно заметить парадоксальный эффект: с одной стороны, капиталист увеличивает свою прибыль путем повышения как производительности труда рабочих, так и прибавочной стоимости продукции; однако этот метод оборачивается снижением прибыли, поскольку удлинение смен истощает работников. Это верно и для промышленного капиталиста, и для цифровой платформы, которая нанимает работников для выполнения задач под управлением алгоритма. В обеих ситуациях увеличение прибавочной стоимости за счет занятой рабочей силы требует физического контроля над работниками. В прошлом такой контроль был затруднен, поскольку в отличие от других факторов производства рабочая сила воплощается (или воплощалась раньше) в работниках, людях, которые, в свою очередь, используют (или использовали) свои возможности, чтобы сопротивляться превращению в товар (Woodcock, 2020; Altenried, 2020).

В «эпоху цифровой революции» это соответствие между физическими качествами и работой значительно подорвано тем, что работа выполняется в одиночку. Действительно, в процессе труда работник общается только с платформой, управляемой алгоритмами. Он изолирован от других сотрудников, и его единственным средством общения является приложение, обеспечивающее его взаимодействие с алгоритмом. Алгоритмическая дискриминация на рабочем месте сегодня лишает работников должной степени признания в отношении законных прав и оплаты труда, закрепляя юридические и категориальные различия в профессиональной деятельности (Fuchs, 2014).

Промышленная революция изменила парадигму равенства и, соответственно, дискриминации, когда наступление эпохи капитализма принесло перемены в социальной реальности и образе жизни (Schivone, 2019). В рамках философского дискурса равенство потребовало нового понимания централизации с отказом от религиозной и философской точек зрения и принятием социальной и политической перспективы, воплотившейся в великих революциях того времени – американской и французской (Schivone, 2019).

Понятия равенства, неравенства и, соответственно, дискриминации вышли за формальные рамки условий рабства, которое продолжало существовать, и стали применяться к фактам социальной жизни, о которых начало формироваться общественное мнение вне старых аристократических и религиозных порядков.

Люди начали считать, что дискриминация и бедность обусловлены развитием технологий, которые расшатали и ослабили прежние социальные категории.

Расшатывание и ослабление прежних социальных категорий снова происходит в наши дни. Действительно, технологии, кажется, усиливают дискриминацию наиболее уязвимых слоев, и это происходит потому, что они способствуют появлению «всепроникающих элементов безличного равенства по отношению к любым индивидуальным или гендерным различиям» (Schiafone, 2019). За этим должен последовать слабый эффект, когда равенство обернется своего рода обезличенностью, поскольку станет необходимо игнорировать или даже уничтожать индивидуальные характеристики, делающие каждого человека неповторимым (Ainis, 2015).

Однако случаи дискриминации могут возникнуть и в контексте всеобщего уравнивания индивидуальности, особенно в области автоматической обработки персональных данных. В этом смысле эффективность использования алгоритмов автоматического принятия решений оправдана именно обобщением проблем и дублированием фактического материала. Алгоритмическая автоматизация предусматривает некоторое безразличие, т. е. намеренное игнорирование возможных причин физического или морального неравенства, поскольку номенклатура выборов, заложенная создателем алгоритма, предполагает каталогизацию, а значит, и классификацию на основе обработки концепций. Эта процедура позволяет во время формирования алгоритма выбрать, как именно будет классифицироваться каждый элемент процедуры, основываясь на целях, на которые направлен данный алгоритм, т. е. на представлениях разработчика. Такая процедура не может быть нейтральной, она следует критериям категоризации в соответствии с поставленными целями. Поскольку процедура не нейтральна и не может быть таковой, так как основана на выборе, то она содержит в себе потенциал дискриминации.

В рамках данной темы исследовались вопросы о том, возможно ли принять меры к обеспечению равноправия различных элементов, и в частности, изучались математические отношения между ними (Garg et al., 2020; Pessach & Shmueli, 2020; Krawiec et al., 2023). Например, методами калибровки групп, положительной балансировки и/или балансировки класса было показано, что, «кроме отдельных случаев», никакие условия не могут одновременно удовлетворить ситуациям, изучаемым в эксперименте (Kleinberg et al., 2016; Pleiss et al., 2017; Abdollahpouri et al., 2020). В других работах рассматривались близкие темы несовместимости между критериями равноправия, которые в некоторых контекстах (например, в психологических тестах на рецидивизм) могут привести к значительной косвенной дискриминации, когда склонность к рецидивизму различается в исследуемых группах (Chouldechova, 2017; Lippert-Rasmussen, 2022). Аналогичным образом, при анализе риска совершения уголовных преступлений было отмечено, что «в целом невозможно максимизировать факторы точности и равноправия одновременно, а также невозможно одновременно удовлетворить всем типам равноправия» (Berk et al., 2021).

В условиях технологической трансформации как в середине XVIII в., так и в наши дни причину дискриминации можно распознать благодаря уязвимости людей, вовлеченных в процесс автоматизации. Действительно, есть сходство между началом промышленной революции и современным использованием искусственного интеллекта.

В обоих случаях происходит закабаление уязвимых групп, которые не могут избежать сдвига технологической парадигмы. Это условие, которое появляется из-за принадлежности к уязвимой группе, но в то же время эта уязвимость заставляет эти слабые группы занять подчиненное положение.

В странах западной правовой традиции, особенно в США, эта двойственность проявляется наиболее глубоко в отношении лиц, страдающих от порочных последствий наследия эпохи рабства, которые прочно укоренились в жизни общества, несмотря на многочисленные попытки преодолеть это положение. Эти попытки остаются неудовлетворительными, поскольку в различных слоях общества существует значительная расовая дискриминация, поддерживаемая системами автоматического принятия решений, особенно в таких областях, как программное обеспечение для оценки рисков.

1. Дискриминация и обращение с персональными данными

Рассматривая неприкосновенность частной жизни, следует считать ее проявлением права личности на сохранение целостности, как физической, так и психической, от влияния третьих сторон, будь то физические субъекты, юридические лица или само государство. Неприкосновенность частной жизни стала основой охраны личности как в Сети, так и в реальном мире. Это защита от попыток реконструкции личности третьими сторонами, публичными или частными, в результате отслеживания данных, которые каждый человек оставляет после себя через геолокацию, выполняемую приложениями на смартфоне, платежами, пересылкой материалов или меток. Неприкосновенность частной жизни защищает личную идентичность и репутацию (благодаря установлению права на забвение, одним из элементов которого является неприкосновенность частной жизни) как в реальной, так и в виртуальной личной жизни.

С этой точки зрения роль дискриминации оказывается более скрытой и касается аспекта манипуляций: с момента, когда «черные ящики» начали собирать персональные данные в массовых масштабах, они подрывают (Messinetti, 2019; Vale et al., 2022) личность и идентичность каждого человека тем, что эффективно переводят их в массивы информации; самым этим фактом они совершают операцию по манипулированию личными данными, что в зависимости от точки зрения влечет за собой два противоположных результата.

С одной стороны, если содержание каждого «черного ящика» формируется из данных, собранных максимально нейтрально, то оно отражает общественную реальность. Таким образом, оно, как в зеркале, отражает и дискриминационные процессы, и искажения, существующие в обществе.

С другой стороны, если при внесении в «черный ящик» данные очищаются от дискриминационного содержания, то может сформироваться искаженный, манипулятивный образ реальности, а «черный ящик» начинает отражать идеальные взгляды тех, кто занимается сбором данных. При этом такая ситуация может быть еще опаснее, так как она соответствует не объективной реальности, а чьим-то представлениям о ней; эти представления, кроме того, что являются предвзятыми и нереалистичными, также оказываются привязаны к ценностным суждениям, ориентированным на цели тех, кто контролирует данный «черный ящик» и его результаты.

Следовательно, формирование «черного ящика», которое лежит в основе сложных процессов автоматизированного принятия решений, должно осуществляться с максимальной прозрачностью и вниманием к целям его провайдеров. Это именно те процессы, которые требуют междисциплинарных знаний специалистов по работе с данными, математиков, философов и юристов, особенно тех, кто занимается сравнительным изучением систем для понимания принципов и результатов работы других правовых систем в этой сфере.

Хотя защита права на неприкосновенность частной жизни традиционно сосредоточивается в руках государственной власти, в последние годы стабильно растет заинтересованность частных организаций в коммерческом использовании данных их пользователей (Zuboff, 2019; York, 2022); это проявляется в ситуациях, которые лишь на первый взгляд могут показаться неожиданными. Противостояние между технологическим прогрессом и защитой личности приводит к четырем различным типам искажений:

(а) с одной стороны, необходимость обеспечить защиту в сфере частной жизни изначально обусловлена освобождением личности от центральной власти (Bellamy, 2014), которая касается всех граждан индустриальных стран (Rodotà, 2012), как индивидуально, так и в массе (Canetti, 1960);

(б) эта необходимость следует из введения электронной обработки и передачи информации, касающейся личности, в производственных процессах;

(в) сбор и хранение таких данных частными организациями представляет собой сдвиг парадигмы (Kuhn, 1962), интересный с правовой точки зрения (Rodotà, 1995; Alpa & Resta, 2006; Angiolini, 2020);

(г) указанная деятельность дает возможность профилирования пользователей, что также предвещает сдвиг парадигмы в отношении субъекта права, противопоставляемого организации, так как юридическое лицо наделяется технологической властью (Pellecchia, 2018).

Сочетание компьютеризации и возможностей передачи данных позволило хранить огромные объемы информации, сначала на магнитных носителях, а затем на все более сложных цифровых устройствах. После этой трансформации частные организации начали, как и государство, массово собирать личные данные, получившие общее название «большие данные» (Mayer-Schönberger & Cukier, 2013). Этому способствовало распространение новейших, более точных, сложных, а значит, и более интуитивных технологических инструментов среди пользователей.

Сюда входит сбор персональных данных из таких источников, как Интернет (цифровые платформы и социальные сети), мобильные устройства (смартфоны, планшеты, смарт-часы), спутниковые геолокаторы, радиочастотные идентификаторы (RFID), датчики, камеры и другие инструменты, использующие новые технологии (Gressel et al., 2020).

Право на неприкосновенность частной жизни проявляется как запрет, т. е. как право гражданина не подвергаться вмешательству в его частную жизнь со стороны лиц или организаций, будь то частных или институциональных. Однако право на защиту персональных данных включено в правовую норму об осуществлении контроля над обработкой и оборотом личной информации (Rodotà, 2014).

Неопределенность коннотаций, связанных с «черным ящиком», получает правовую оценку в «негативном» (охрана неприкосновенности частной жизни вплоть до вмешательства извне) или «позитивном» (защита персональных данных через

контроль над информацией) ключе. Действительно, обработке подвергается информация, получаемая от субъектов в течение всей жизни, независимо от того, как ее можно умозрительно классифицировать по различным концептуальным основаниям.

Если представить себя внутри «черного ящика», погруженным в виртуальное пространство, где матричные вычисления придают форму алгоритмам автоматического принятия решений, то становится понятно, что невозможно логически проследить путь данных. Такие операции проявляются своим результатом (Cockburn et al., 2018; Bodei, 2019), тогда как о его условиях и источнике ничего не известно.

Если считать истиной, что весь этот огромный массив информации (действительно «большие данные») представляют собой знания, используемые «черными ящиками», то становится необходимым определить, кто обладает полномочиями управлять и распоряжаться этими знаниями. Возникают следующие вопросы:

А. Кто владеет конкретной информацией? Этот вопрос относится к распространению информации и к тому, можно ли на законных основаниях ограничить доступ к ее источникам.

Б. Кто решает, какую информацию можно получить (и, следовательно, как оценивать собранную информацию)? Здесь рассматривается роль каждого субъекта цепочки информации: «источник» данных (считается отдельной личностью, даже если собранные данные обрабатываются как массив), институты, осуществляющие контроль за использованием данных (например, Национальная служба по защите частной жизни), а также операторы «черных ящиков», получающие данные.

В. Кто принимает решения о том, кто будет принимать решения? Кто уполномочен контролировать передачу или удаление собранной и доступной информации?

Ключевым элементом каждого вопроса является эффективность защиты частной жизни. Например, в вопросе А неприкосновенность частной жизни рассматривается как препятствие против вторжения частных и государственных субъектов в личную сферу, как ограничение возможности собирать дискриминирующие данные, а значит, и возможности для дискриминации как таковой, осуществляемое в момент формирования «черного ящика».

Вопрос Б решает, какой орган может разрешить массовый сбор данных, и здесь ответственность и гарантии лежат, с одной стороны, на службах по защите данных, которые имеются у всех частных и государственных организаций, участвующих в сборе и обработке информации. С другой – в роли гарантов равноправия выступают общеевропейские и национальные институты.

Наконец, что касается вопроса В, т. е. кто обладает полномочиями определять лиц, принимающих решения, представляется логичным, что эту функцию должно выполнять государство, понимаемое как представитель всех своих членов, подчиняющийся верховенству закона и принципу разделения властей. При этом очевидно, что уровень полномочий платформ при управлении данными фактически достигает уровня монополии при кризисном управлении национальных государств в отношении заявленной экстерриториальности экономических субъектов при применении национального законодательства (Tarrant & Cowen, 2022; Witt, 2022; Parona, 2021).

Данный контекст появился благодаря взаимодействию субъектов в сообществе без государства, таком как Интернет. С одной стороны, нет никакой существенной разницы между социальными и политическими отношениями, поскольку отдельный пользователь считает себя единицей отсчета в рамках своего мира. С другой стороны, он растворяется в общем потоке информации в Сети.

Напротив, негосударственные субъекты могут устанавливать правила через договорные отношения, вводить односторонние ограничения и действовать на свое усмотрение без необходимости соблюдать баланс. В этом контексте процесс сбора информации находится под влиянием дискриминирующих элементов, особенно если она собирается незаконно, что приводит к необъективным, искаженным, предвзятым результатам, а затем к противозаконным и фактически неверным решениям.

В определенном смысле изучение «черных ящиков» и заключенных в них дискриминирующих элементов служит зеркалом, в котором отражается реальность. Оно может стать полезным инструментом в поиске средств противодействия этой ситуации, особенно с правовой точки зрения.

В данной области важная роль принадлежит прецедентному праву, которое может разрешить противоречие между необходимостью обеспечить общественную безопасность и защитой частной жизни.

2. Предубеждение и дискриминация в системах автоматизированного принятия решений

В компьютерных науках под термином «систематическая погрешность» понимается искажение длины передачи данных. В компьютерном праве этот термин означает ситуации дискриминации со стороны алгоритмических моделей, которые «могут приводить к ложноположительным или ложноотрицательным выводам и, как следствие, к дискриминационным эффектам в ущерб некоторым категориям лиц» (Parona, 2021). Такое искажение может зависеть от набора обучающих данных, от релевантности или точности данных, от типов используемых алгоритмов (Hildebrandt, 2021) в том, что касается качества или скорости получения результата. С точки зрения социальной реальности дискриминация на основе искажения может относиться к несправедливому обращению или незаконной дискриминации. В этой связи критически важно отличать компьютерные искажения от влияния неравноправного отношения из-за незаконной дискриминации; это зависит от того, как формируются собираемые данные и как они взаимодействуют между собой и в то же время с окружающей реальностью.

В доктрине отмечается, что существует три основных вида взаимодействия искажений при автоматизированном принятии решений (Hildebrandt, 2021):

(а) первый из них относится к машинному обучению на основе алгоритмов машинного обучения. Это искажение индуктивно и неизбежно, и, хотя само по себе оно не является ни положительным, ни отрицательным, его нельзя считать нейтральным относительно реальности, с которой оно взаимодействует;

(б) второй тип относится к этическим проблемам, так как позволяет изменять распределение товаров, услуг, рисков, возможностей и доступа к информации способами, которые могут быть нравственно неоднозначными. Например, некоторых людей могут исключать или подталкивать в определенном направлении, к определенным действиям;

(в) третий тип искажений наиболее очевиден даже для стороннего наблюдателя. Он основывается на незаконных ситуациях или действиях, т. е. алгоритм машинного обучения фокусируется на лицах или категориях субъектов, опираясь на незаконные или дискриминационные мотивы.

В литературе обсуждался вопрос, может ли искажение категории (в) включать в себя подкатегорию этических искажений (Hildebrandt, 2021). Действительно, дискриминация по признаку пола незаконна, но не все считают ее неэтичной, например, когда водители-мужчины должны платить повышенные страховые взносы, так как женщины водят машину более осторожно. Такая дискриминация, хотя и является незаконной, не всегда представляет собой этическую проблему.

Искажения категорий (а) и (б) могут относиться к действиям, наблюдаемым с помощью датчиков или систем слежения онлайн, или вызываться тем же автоматизированным алгоритмом. При наблюдении искажение влияет на обучающие данные, а также на логические выводы системы. В обоих случаях оказывается затронут вывод (т. е. результат), поэтому искажение не обладает нейтральностью по отношению к реальности.

Использование машинного обучения (Hildebrandt, 2021) неизбежно порождает искажения, поскольку, как и в случае человеческого познания и восприятия, эти процессы не объективны, вопреки ожиданиям. В этой ситуации следует проявлять осторожность и критичность, особенно когда когнитивные результаты машинного обучения кажутся достоверными (Hildebrandt, 2021).

Одна из основополагающих работ в этой области, «Машинное обучение» Тома Митчелла (Mitchell, 2007; Kubat & Kubat, 2017), убедительно показывает, что искажение, понимаемое как вариативность, необходимо для демонстрации важности и полезности точных проверок. Отсюда можно сделать вывод, что искажения, будь то представляющие собой вариативность или нет, могут привести к ошибкам (Hildebrandt, 2021), укорененным в различных стадиях процесса принятия решений. Этот процесс может включать такие этапы, как сбор или классификация «обучающих данных» вплоть до достижения цели автоматизированного принятия решений. Суть таких ошибок может заключаться в переносе фактических чрезвычайных ситуаций в программный код или в скрытой неполноте исходных данных (что определяется термином «легкая мишень»), что, в свою очередь, вызывается контекстом, в котором работает программа машинного обучения, т. е. симулированной или реальной моделью (Hildebrandt, 2021).

Объем данных, используемых программой машинного обучения, также может быть причиной ошибок, так как эта программа может обрабатывать «ложные» корреляции и закономерности в результате искажений, понимаемых как вариативность, заложенных в исходные данные, именно по причине обращения к определенной идее, на которую настроены создатели алгоритма.

Однако существует и третья возможность. Она относится к проблеме, в одно и то же время фундаментальной и неуловимой, и возникает, когда данные корректно считываются системой машинного обучения (далее – МО), но реальная ситуация производит искажающий эффект. Например, это может происходить после социальных изменений, повлекших за собой исключение уязвимых групп, т. е. исключение отдельных лиц на основании физических или поведенческих характеристик.

В случае ошибочного или ложного исходного материала собранные данные с самого начала имеют дефекты, которые могут привести к ошибочным интерпретациям зависимостей; искажение коренится в реальной жизни, поэтому вычленение данных будет способствовать закреплению или даже усилению искажений.

Эту ситуацию нельзя исправить с помощью изменений МО, хотя есть мнение, что МО может способствовать выявлению таких искажений или их причин. Поэтому необходимо критически рассматривать возможные причины искажений. Действительно, необходимо стараться определить, способствует ли процедура сбора данных появлению искажений в исходном материале или эти искажения появляются в результате непричинного распределения; при этом следует с большой осторожностью работать с инструментами МО из-за риска некритического отношения к его результатам (Hildebrandt, 2021).

Точность результатов машинного обучения зависит от того, с какими знаниями оно работает и взаимодействует, вне зависимости от ограничений (Marcus, & Davis, 2019, Brooks, 2017), налагаемых исходными данными (Sunstein, 2019) и используемыми моделями.

Данная проблема относится к способности понять различие между когнитивными искажениями, существующими у человека (чей разум может адаптироваться и делать абдуктивные и неожиданные умозаключения) и у систем машинного обучения, чей интеллект, напротив, всегда опирается на данные, предположения и характеристики исходного материала. Система МО может делать индуктивные и дедуктивные выводы, но не абдуктивные. Действительно, абдуктивное мышление требует «интуитивного скачка», который начинается с набора элементов и затем вырабатывает объяснительную теорию для этих элементов, проверяя ее на имеющихся данных (Hildebrandt, 2021). Для проверки таких гипотез при создании модели машинного обучения учитывают творческие способности по перекомпоновке абдуктивного этапа и проверяют гипотезу индуктивным методом. Если бы такая операциональная гипотеза подтвердилась, то система была бы способна использовать абдуктивный метод в качестве основы для дедуктивного мышления. Поэтому экспериментальная обратная связь в машинном обучении является решающим фактором, так как она обладает свойствами фундаментальности и критичности (Hildebrandt, 2021).

Отсюда следует, что качество применяемых образцов отражает качество обучения баз данных. Если пользователь загружает в систему искаженные или некачественные данные (образцы), то это негативно влияет на поведение алгоритма машинного обучения и приводит к некачественным результатам (Gallese, 2022). Кроме того, нельзя забывать, что алгоритмам машинного обучения свойственно терять способность к обобщению, а значит, преувеличивать (Gallese, 2022).

Это обстоятельство хорошо известно программистам, однако юристы его недооценивают. Распознать его можно по соотношению между ошибками при обучении и при тестировании – при исправлении ошибки на обучающем наборе данных ошибка на тестовом наборе данных усугубляется. В этом случае сеть, на которой основывается система машинного обучения, оказывается «переобученной» (Gallese, 2022). Это происходит тогда, когда модель отвечает наблюдаемым данным (образцам) потому, что учитывает слишком много параметров по отношению к количеству наблюдений и теряет связь с реальностью данных. Отсюда можно сделать вывод, что вклад человека в предварительные решения по поводу сбора, классификации и обработки данных неизбежен и существенен для получения осмысленных и приемлемых результатов.

Другими словами, если алгоритм машинного обучения выдает ошибочные, дискриминационные, неверные результаты, то ответственность за это лежит на том, кто организовывал базу данных и настраивал алгоритм. Машина лишь выполняет данные ей инструкции.

В таких ситуациях проблема может возникнуть или усилиться даже при непрерывном поступлении новых категорий данных, что приведет к разбалансировке системы, так как одни категории данных представлены лучше, чем другие, что существенно влияет на будущие действия искусственного интеллекта (Gallese, 2022). Однако в одном случае проблема не будет иметь решения: это так называемое пошаговое обучение на генеральной выборке (Gallese et al., 2020), когда система машинного обучения получает новые данные, которые могут, в принципе, принадлежать новым или никогда ранее не встречавшимся классам. В этой ситуации алгоритм должен быть способен перестроить свои внутренние процессы (т. е. в случае глубокого обучения алгоритм должен адаптировать свою архитектуру и перекалибровать все параметры) (Gallese et al., 2020). При этом становится невозможно предсказать будущие действия автоматизированной системы.

3. Двойственная природа алгоритмов оценки риска

Алгоритмы автоматизированного принятия решений выявляют важный аспект гражданского общежития, в настоящее время меняющий свое направление, а значит, и свою сущность. Это отношения между властью (будь то государственной или частной, но способной влиять на частную и общественную жизнь человека) и рядовыми гражданами (т. е. людьми, которые осознанно или неосознанно начинают зависеть от персональных данных или становятся их источниками).

Каждый этап и элемент жизни (как повседневной, так и всей жизни в экзистенциальном смысле) стал объектом автоматического сбора данных, профилирования и принятия решений на их основе. Алгоритмы биометрического распознавания изучают нашу личность и чувства через так называемое эмоциональное компьютерное обучение (Guo et al., 2020), т. е. используют данные о наиболее явных физических характеристиках (таких как цвет глаз, кожи и волос) или отпечатках пальцев, хранящихся в идентификационных документах (например, паспортах), об особенностях походки или выражении лица в покое и при выполнении работы (Crawford, 2021) или когда вы бежите к поезду метро под внимательным взглядом камер наблюдения.

Эта информация позволяет ответить на такие интересные вопросы, как кто вы и куда вы направляетесь. Но насколько законны сбор и обработка таких данных?

Вероятно, было бы полезно изучить алгоритмы оценки риска. С эволюционной точки зрения они раскрывают интереснейшие аспекты, так как именно они были первыми программами, применявшимися для прогнозирования рецидивизма, изначально в области условно-досрочного освобождения (Oswald et al., 2018). Как указывает доктрина, использование алгоритмов прогностической аналитики в судебном контексте служит трем целям, из которых две являются более общими, особенно на уровне оценки соотношения расходов и выгод и оценки ресурсов для обеспечения общественной безопасности и правовой защиты, а третья выступает на индивидуальном уровне (Oswald et al., 2018).

С точки зрения эффективности осуществления правосудия использование указанного программного обеспечения может дать определенные преимущества, например:

(а) способствовать общему стратегическому планированию при определении прогнозов и приоритетов в борьбе с преступностью;

(б) оценивать параметры конкретных действий в сфере борьбы с преступностью;

(в) оценивать прогнозы рецидивизма на индивидуальном уровне.

Что касается последнего аспекта, то применение алгоритмов оценки риска приводит к необходимости соотносить необходимость (в первую очередь для государства) с принципом пропорциональности в свете уважения прав человека, т. е. защита личных прав человека должна быть справедливо сбалансирована с интересами сообщества (Oswald et al., 2018).

С точки зрения права, основными и важнейшими проблемами в данной области являются:

1) непрозрачность и секретность, связанные в первую очередь с защитой интеллектуальных прав на алгоритм как таковой (Oswald et al., 2018), но спорные с точки зрения принципа прозрачности, особенно в случае использования систем автоматизированного принятия решений в судебной сфере, например, при прогнозировании уголовного рецидивизма;

2) использование таких автоматизированных систем не отвечает критериям защиты принципов конституционного соответствия, например, законного права на вынесение решения судьей-человеком, права быть выслушанным в суде и права на справедливый суд, на мотивированное решение (предусмотренное ст. 111 Конституции Италии и ст. 6 Европейской конвенции по защите прав человека и основных свобод), а также, очевидно, не соответствует положениям ст. 22 Общего регламента ЕС по защите персональных данных, которая гарантирует право на объяснение решения, принятого автоматизированной системой.

Что касается ограничений свобод, то и решение, принятое алгоритмом, и база данных, на которой оно было основано, должны быть прозрачными и доступными, а неадекватность хотя бы одного из этих элементов является нарушением принципов надлежащей процедуры согласно ст. 111 Конституции и ст. 6 Конвенции по правам человека.

Необходимо предусмотреть возможность оспаривать в обычных судах автоматизированные решения с неадекватными результатами, полученные путем процедур, не отвечающих принципам защиты персональных данных и фундаментальным правам, с целью получить адекватные решения, понятные тем, кого они затрагивают. У человека, которого касается принятое автоматизированное решение, есть право получить информацию о правильности этого решения как в формальном аспекте (т. е. выполнение гарантий защиты, заложенных в законодательные нормы на разных уровнях), так и по существу (т. е. почему дело рассматривалось с юридической точки зрения и к каким выводам пришел автоматический алгоритм). Логическая цепочка не должна вызывать существенных сомнений в том, что принимающий решение, в данном случае алгоритм оценки риска, корректно применил закон и принял рациональное решение на релевантных основаниях (Oswald et al., 2018).

Можно ли утверждать, что системы автоматизированного принятия решений должны применять более высокие стандарты принятия решений, чем человек? (Oswald et al., 2018). Например, в сфере права судебное решение, вынесенное судьей

без указания основания, будет считаться несуществующим¹, т. е. ничтожным и не имеющим юридической силы², что предусмотрено ст. 132, № 5 Уголовно-процессуального кодекса, ст. 546 Уголовно-процессуального кодекса и ст. 111 Конституции (Chizzini, 1998; Taruffo, 1975; Massa, 1990; Bargi, 1997). Судебное решение должно быть обосновано с помощью соответствующих фактов по делу и законных оснований для принятого решения (Taruffo, 1975).

Однако в литературе отмечены примеры из исторической и сравнительной юридической практики, когда вышеупомянутая обязанность давать обоснования с помощью фактов и положений закона не является сама по себе неперенным условием отправления судебной функции, поскольку в различных системах отсутствуют как институты суда присяжных, судей, рассматривающих и решающих вопросы факта (Chizzini, 1998; Taruffo, 1975; Massa, 1990; Bargi, 1997), так и обязанность обосновывать решения (Chizzini, 1998). Обязанность указывать причины решений берет начало с якобинских конституций как внутривпроцессуальная функция, необходимая для контроля над судебной властью, что соответствовало становлению номофилактической роли Верховного суда в реализации принципа подчиненности судьи закону (Taruffo, 1975).

Эти слова вызывают в памяти старинное высказывание Монтескье о том, что он хотел бы, чтобы судья был «глашатаем закона» (*bouche de la loi*), а не «глашатаем короля» (*bouche du roi*) (Petronio, 2020). Кто становится этим «королем», когда в юридической сфере используются алгоритмы оценки риска? Этот вопрос особенно актуален в свете сензитивности, т. е. легкости манипулирования результатами, полученными с помощью алгоритмов оценки риска. Можно поставить вопрос, почему алгоритм не может стать «глашатаем закона». Ученые утверждают: чтобы убедиться, что судья, даже судья Верховного суда, не занимается произволом, а следует закону, который для всех един, даже если применяется с учетом индивидуальных особенностей каждого из множества дел, суждение должно быть обосновано, т. е. необходимо дать отчет, почему принято именно такое конкретное решение (Petronio, 2020).

Указание причин принятого решения служит также демонстрацией компетентности, независимости и ответственности судьи; это инструмент правовой защиты сторон или ответчика против таких скрытых факторов, как предубеждение, ошибка или иррациональность. В отношении автоматического принятия решений по таким важнейшим вопросам невозможно выполнить подобную операцию по обеспечению прозрачности, учитывая сложившуюся закрытость этапа принятия решения.

По задумке сторонников использования алгоритмов, оно оправдано тем, что заменяет системы оценки в области освобождения под поручительство (Israni, 2017); однако в реальности алгоритмы вобрали в себя все распространенные стереотипы, особенно этнические, экономические и гендерные (Starr, 2014), не решив при этом этические и конституционные проблемы.

Так, в Верховном суде штата Висконсин рассматривалось дело об использовании алгоритма, разработанного для помощи судьям в области освобождения под

¹ Cass., 19-7-1993, n. 8055, GC, 1993, I, 2924; Cass., 8-10-1985, n. 4881, NGL, 1986, 254.

² Cass., 27-11-1997, n. 11975.

поручительство и риска рецидивизма заключенных. Суть дела в следующем: в феврале 2013 г. был арестован Эрик Лумис, который вел автомобиль, фигурировавший в деле о стрельбе. Вскоре после ареста он признал свою вину в неподчинении судебному постановлению и не оспаривал тот факт, что завладел автомобилем без согласия владельца. В результате его приговорили к шести годам тюремного заключения. Это решение интересно тем, что окружной суд применил запатентованный алгоритм, разработанный компанией Northpointe, Inc., представляющий собой этап принятия решения программой искусственного интеллекта четвертого поколения для оценки риска под названием «Управление профилированием заключенных для альтернативных санкций» (Correctional Offender Management Profiling for Alternative Sanctions, COMPAS). Предполагалось, что этот алгоритм должен предсказывать риск рецидивизма для конкретного заключенного на основе комплексного анализа информации, собранной в результате опроса из 137 пунктов, и данных из уголовного досье (Rebitschek et al., 2021; Bao et al., 2021; Wang et al., 2022).

Верховный суд штата Висконсин постановил, что такой инструмент не противоречит Конституции в отношении права ответчика на надлежащую процедуру, поскольку программа обрабатывает индивидуальные и точные данные (Israni, 2017)³. Следует также отметить, что производитель программы COMPAS отказался раскрывать суду методологию и правила, при помощи которых программа пришла к своему решению (Custers, 2022), хотя в постановление суда было внесено значение балла по оценке риска, выведенное алгоритмом. Поскольку программа сочла риск рецидивизма высоким, суд отказал в освобождении под поручительство и назначил шесть лет тюремного заключения (Israni, 2017).

Верховный суд штата Висконсин отклонил сомнения в конституционности в отношении права ответчика на надлежащую процедуру, несмотря на отсутствие прозрачности и точности в формулировках алгоритма, которые не позволили ответчику быть уверенным в беспристрастности процесса принятия решения о мере наказания.

Такая проверка алгоритмов принятия решений часто запрашивается в судебном процессе, но отклоняется из-за правил защиты интеллектуальной собственности и коммерческой тайны, выражая в своеобразной форме непознаваемость алгоритма (Vogel, 2020), который становится абсолютным совереном: это воплощение «голоса короля», миропомазанного интеллектуальной собственностью вместо божественной святости.

В то же время в данном контексте нормы охраны интеллектуальной собственности защищают экономические интересы разработчиков технологий машинного обучения. Чьи права будут превалировать при сравнении прав на доступ к механизму принятия решений, в особенности прав на защиту (а в некоторых контекстах прав на надлежащую процедуру), предусмотренных различными многоуровневыми актами, от Конституции до международных конвенций?

Правовая система Северной Америки поддержала сохраняющуюся непрозрачность, подтвердив конституционность использования системы COMPAS, хотя Верховный суд штата Висконсин наложил многочисленные ограничения на ее использование. Так, алгоритм нельзя применять для решения вопроса, подлежит ли задержанный аресту, т. е. для определения срока его задержания (Israni, 2017).

³ State v. Loomis, 881 N.W.2d 749, 767 (Wis. 2016).

Решения алгоритма COMPAS подлежат проверке при вынесении приговора и при выставлении баллов прогноза рецидивизма, с обязательной оговоркой об ограниченности его использования для принятия решений. Поскольку федеральный Верховный суд отказался выпустить приказ об истребовании дела, решение Верховного суда штата Висконсин является окончательным⁴.

Спорным остается фактическое обстоятельство, ставит ли решение Верховного суда штата Висконсин алгоритм, относительно которого существуют этические и конституционные сомнения из-за его непрозрачности, наравне с обычными судьями в вопросе осуществления правосудия. В судебной системе США защита конфиденциальности в патентных вопросах является приоритетной в целях поддержания конкурентного преимущества владельца патента по отношению к индивидуальному праву на надлежащую процедуру. Согласно доктрине, это противоречие, которое будет сложно разрешить без вмешательства федерального Верховного суда (Israni, 2017).

Рассматривая проблему использования не менее пяти психологических и страховых инструментов оценки риска, Верховный суд Канады предложил обоснование от обратного в вопросе о том, могут ли механизмы принятия решений применяться при оценке уровня риска рецидивизма у преступников этой страны. Речь идет о решении по делу *Ewert v Canada*⁵. В рамках этого дела Верховный суд Канады постановил, что Канадская служба исполнения наказаний (Correctional Service of Canada, CSC) нарушила свою законную обязанность, согласно разделу 24 Закона о наказаниях и условном освобождении, использовать исключительно точную информацию при оценке риска (Scassa, 2021).

Канадский суд отметил, что служба исполнения наказаний должна учитывать систематическую дискриминацию коренных народов Канады как в целом в системе уголовного правосудия, так и в местах лишения свободы. При этом Верховный суд постановил, что, несмотря на использование инструментов, которые могут быть дискриминационными по отношению к представителям коренных народов, Канадская хартия прав и свобод не была нарушена (Russell, 1983; Epp, 1996).

Это решение вызвало ряд критических комментариев, учитывая, что именно представители коренных народов, и в особенности женщины, страдают от систематической дискриминации из-за организации тюремной системы: их лишение свободы оказывается более глубоким и длительным. Для них нет программ и форм реабилитации, учитывающих их культурные особенности, которые помогли бы им вернуться в свои сообщества и получить там поддержку.

Хотя канадский суд признал наличие такой дискриминации, руководствуясь тем, что «одинаковые условия могут приводить к серьезному неравенству»⁶, однако он также указал, что существует важная связь между инструментами оценки риска, применяемыми в исправительном учреждении, и свободой заключенных. Высокий рейтинг риска заключенного значительно влияет на его личную свободу: такой заключенный помещается в условия более строгого режима, его шансы на досрочное освобождение снижаются. Несмотря на все это, Верховный суд постановил, что ни

⁴ Loomis v. Wisconsin, 137 S. Ct. 2290 (2017).

⁵ Ewert v Canada [2018 SCC 30]. <https://canliiconnects.org/en/commentaries/62360>

⁶ Там же.

личные права на жизнь, свободу и безопасность (гл. 7 Хартии), ни положения о равноправии (гл. 15) не были нарушены. По мнению суда, нет свидетельств того, что такая дискриминация перешла в алгоритмы программы для оценки риска и тем самым дискриминировала заявителя.

Верховный суд Канады признал, что инструменты классификации риска рецидивизма заключенных неточны и необъективны. Однако вызывает недоумение тот факт, что, несмотря на обширный анализ, выявивший некорректность этих инструментов и их необъективность по отношению к представителям коренных народов, Верховный суд так и не объявил их неконституционными. Это решение еще более ошибочно и пагубно, чем решение Верховного суда Висконсина, так как последнее, по крайней мере, установило процедурные ограничения на использование такого программного обеспечения. Так, предусмотрены некоторые средства, позволяющие ограничить влияние алгоритма на экономические, этнические и социальные аспекты при внесении персональных данных в программу. Кроме того, есть возможность не вносить «зип-код», т. е. данные о месте жительства проверяемого лица, поскольку они позволяют судить об уровне доходов. Исходя из принципа презумпции невиновности, эти данные не релевантны при определении риска рецидивизма, что является целью использования этой программы.

4. Предложения к проекту закона об искусственном интеллекте

Проект закона, опубликованный Еврокомиссией 21 апреля 2021 г., представляет собой первую полноценную попытку регулирования искусственного интеллекта в общем виде, независимо от контекста, в котором это регулирование должно применяться: с одной стороны, фрагментарное или облегченное регулирование (особенно в США или Китае, двух ведущих конкурентах на мировом рынке в этой сфере); с другой стороны, сложности с прогнозированием и гармонизацией развития этого сектора, когда новшества появляются очень быстро, а интересы сторон могут противоречить друг другу (Rosa, 2021; Alpa, 2021; Scherer, 2015). Регулирование в области развития ИИ должно обеспечивать защиту фундаментальных прав и верховенства закона, но при этом быть достаточно гибким, чтобы адаптироваться к технологическим изменениям, которые еще только прогнозируются. Другими словами, регулирование должно решить проблему «квадратуры круга» на национальном и общеевропейском уровнях и создать модель общемирового уровня.

Критические точки в этом отношении связаны с балансом между универсальностью и обновлением этой отрасли, учитывая возможность ее быстрого устаревания благодаря независимому развитию «черных ящиков», машинного обучения, глубокого обучения и нейросетей, что, в свою очередь, является источником рисков, которые нельзя предвидеть заранее, тогда как один из фундаментальных принципов верховенства закона – это обеспечение общих и абстрактных норм, нацеленных в будущее (Scherer, 2015).

В любом случае мы видим, что разработчики проекта закона признают наличие определенной юридической практики в этой сфере, особенно в том, что касается более слабой позиции пользователя по отношению к платформам. Это также свидетельствует о растущем понимании феномена дискриминации, связанной с автоматизированными алгоритмами.

Еврокомиссия рассмотрела проблемы, связанные с потенциальными факторами риска, которые не могут быть предсказаны заранее, и ввела два механизма для обеспечения гибкости законодательной базы. Развитие этой стратегии происходит по трем основным направлениям:

а) Закон об искусственном интеллекте дополняется несколькими приложениями, составляющими его неотъемлемую часть, в которых дается характеристика данной отрасли права. В частности, выделяются категории высокорисковых устройств (высокорисковые системы ИИ), для которых конкретизируется законодательство и предусматриваются определенные процедуры соответствия. Такие приложения настолько соответствуют нормативному подходу Еврокомиссии к проблеме искусственного интеллекта, что процедура их введения предусмотрена ст. 290 Договора о функционировании Европейского союза, описывающей установление технических стандартов (Battini, 2018). Для своевременной и эффективной выработки решений в области применения высокорисковых систем в соответствии со ст. 74 Закона об искусственном интеллекте комитет планирует принимать поправки к законодательству, возможно, даже вне стандартной процедуры, требуемой для формального изменения регламентов (Casonato et al., 2021; Veale et al., 2021; Stuurman et al., 2022);

б) проект закона об искусственном интеллекте также предусматривает обязательный пересмотр каждые пять лет, первый из которых должен быть проведен до истечения пяти лет с момента принятия Закона; это обусловлено изменчивостью данной сферы и необходимостью адаптации норм и стандартов к происходящим изменениям;

в) для достижения максимальной эффективности указанного механизма пересмотра регламентов раздел V проекта закона предусматривает применение механизма так называемых песочниц, т. е. функционального пространства, создаваемого странами – членами ЕС на ограниченный период времени и под контролем правительств, в котором обеспечивается возможность для экспериментирования и тестирования инновационных систем искусственного интеллекта с целью их последующего вывода на рынок.

Учитывая сложность и непрозрачность образа действий, высокую степень непредсказуемости, автономность в получении результатов на основе исходных данных, необходимость регулирования искусственного интеллекта стала уже насущной. Такое регулирование должно учитывать риски безопасности, гарантировать и повышать степень защиты фундаментальных прав от правовой неопределенности, возникающей из-за фрагментарности регулирования и недостатка доверия как к самому инструменту, так и к способности человека его контролировать.

Предлагаемый закон об искусственном интеллекте является частью стратегии Евросоюза по усилению единого цифрового рынка с гармонизацией его норм. В данном случае эти нормы направлены на преодоление фрагментации внутреннего рынка в отношении основных элементов, касающихся требований к продукции с использованием автоматизированных алгоритмов с целью избежать правовой неопределенности как для поставщиков автоматизированных систем принятия решений, так и пользователей их услуг. Действительно, с точки зрения субсидиарности, если строго придерживаться принципа неэсклюзивной правомочности, то в этом смысле, учитывая возможность внедрения различных баз данных в любой продукт с автоматизированными системами, подход на уровне отдельных государств приведет к созданию более значительных и противоречивых нормативных ограничений и неопределенностей, которые будут тормозить оборот товаров и услуг, включая те, где используются автоматизированные системы принятия решений.

В этом смысле проект закона об искусственном интеллекте нацелен на развитие законодательной базы, отвечающей принципу пропорциональности, который осуществляется через подход, ориентированный на учет рисков, налагая ограничения только тогда, когда системы искусственного интеллекта несут в себе высокие риски (т. е. превосходящие выгоды) для защиты фундаментальных прав и обеспечения безопасности. Для оценки степени риска и признания систем искусственного интеллекта как не принадлежащих к высокорисковой категории, они должны соответствовать определенным требованиям: используемые данные должны отвечать критериям высокого качества, документированности, прозрачности и отслеживаемости.

В этом отношении данный инструмент регулирования был выбран потому, что, согласно ст. 288 Договора о функционировании Европейского союза, непосредственная применимость регламента снижает правовую фрагментированность и способствует развитию единого рынка законных, безопасных и надежных систем ИИ путем объединения всех стран – членов ЕС в рамках согласованного набора основных требований к высокорисковым системам ИИ и обязанностей производителей и пользователей таких систем, тем самым повышая степень защиты фундаментальных прав и обеспечивая правовую определенность операторов и потребителей.

Что касается защиты фундаментальных прав, проект закона об искусственном интеллекте налагает ряд ограничений на свободу ведения бизнеса и свободу занятий искусством и наукой в целях обеспечения превалирующих общественных интересов, таких как здравоохранение, безопасность, защита прав потребителей и других фундаментальных прав при развитии и использовании высокорисковых систем ИИ. Эти ограничения пропорциональны и сведены к минимуму, необходимому для предотвращения и снижения серьезных рисков и возможных нарушений фундаментальных прав. Использование ИИ с его особыми характеристиками (непрозрачность, сложность, зависимость от исходных данных, автономность) может негативно повлиять на ряд фундаментальных прав, закрепленных в Хартии ЕС по правам человека. Обязательства по предварительному тестированию, управлению рисками и надзору со стороны человека также способствуют соблюдению других фундаментальных прав путем минимизации риска принятия ошибочных или необъективных решений с помощью искусственного интеллекта в таких критических областях, как образование и обучение, занятость, юридические услуги, судебная система, здравоохранение и социальное обеспечение.

Следует подчеркнуть, что в интересах инвестирования средств в основные фонды, технологии и исследования, обязанности по обеспечению прозрачности не должны чрезмерно затрагивать право на защиту интеллектуальной собственности, технологических и коммерческих секретов, конфиденциальной информации. Однако здесь таится опасность помешать достижению таких целей, как прозрачность и надежность автоматизированных систем принятия решений. Проблема состоит в том, что если система в целом недостаточно сбалансирована, то она не позволит раскрыть способы обработки данных, а значит, и источник возможной дискриминации, как это произошло в судебном процессе о защите сотрудников, массово нанимаемых через онлайн-платформы.

Как и предполагалось, проект закона основан на рискориентированном подходе и предлагает следующую классификацию моделей искусственного интеллекта:

а) запрещенные практики применения искусственного интеллекта, ориентированные на манипулирование поведением людей путем технологий, действующих

на уровне подсознания, или эксплуатация уязвимостей по признаку возраста или ограниченных возможностей для влияния на их действия. Также категорически запрещено использование систем ИИ органами власти с целью установить благонадежность личности (т. е. «социальный рейтинг») (Maamar, 2018; Infantino et al., 2021) на основе общественного поведения и личных характеристик. Однако этот запрет, вероятно, появился как реакция на другие правовые системы, такие как Китай, поскольку использование подобных моделей «социального рейтинга» уже и так запрещено в Европе как нарушающих достоинство и равноправие.

Аналогичным образом запрещены инструменты распознавания, за исключением их применения для целенаправленного поиска потенциальных жертв преступности (например, пропавших детей) или для предотвращения конкретной, существенной и неотвратимой опасности для человека или террористической атаки, либо для поимки, определения местоположения или задержания подозреваемого в преступлении согласно ст. 2(2) Рамочного постановления Совета 2002/584, за которое по законодательству соответствующей страны предусмотрено наказание от трех лет лишения свободы. При этом за такое преступление должен быть выписан Европейский ордер на арест. В любом случае отмечается, что проект закона в явной форме ничего не устанавливает относительно возможного использования таких систем распознавания частными организациями;

б) высокорисковые модели: их использование может быть разрешено, но требуется предварительная проверка по конкретным критериям в отношении защиты человеческого достоинства и соблюдения фундаментальных прав. Выделение этой категории основано как на предполагаемой функции устройства, так и на общей и частных целях его применения. Для установления этого требуется оценка устройства на соответствие как применимому законодательству, так и принципу защиты фундаментальных прав. Сюда входит широкий спектр инструментов (используемых в областях, перечисленных в Приложении III) и моделей, используемых при наиме, в диагностических медицинских устройствах, для биометрической идентификации личности, для управления инфраструктурой (например, умные светофоры в «смарт-городах» или устройства для управления подачей воды, газа или электричества), в обучении и подготовке персонала и т. д.

Отнесение системы искусственного интеллекта к категории высокорисковых основано на предполагаемой функции этой системы в соответствии с действующим законодательством о безопасности продукции. Таким образом, оно зависит не только от функции, выполняемой системой ИИ, но и от конкретной цели и способа, которым используется эта система. Согласно этой классификации, выделяются две категории высокорисковых систем: (а) системы ИИ, предназначенные для использования в качестве элементов безопасности продуктов, которые должны проходить предварительную независимую проверку на соответствие; (б) другие отдельные системы ИИ, касающиеся главным образом фундаментальных прав, перечисленные в Приложении II;

в) модели ИИ с минимальным риском или его отсутствием характеризуются тем, что не принадлежат к вышеперечисленным категориям, хотя также должны соответствовать определенным требованиям к прозрачности. Это, например, чат-боты, которые могут быть разрешены к применению, но должны выполнять требования к качеству информации и прозрачности, или модели «если/то».

Раздел I проекта закона об искусственном интеллекте определяет предмет и объем новых стандартов, регулирующих размещение на рынке, отладку и использование систем ИИ. В нем также представлены определения понятий, присутствующих в документе, в частности, в ст. 3(1) проекта «система искусственного интеллекта» (система ИИ) определяется как «компьютерная программа, разработанная с использованием одного или нескольких приемов и подходов, перечисленных в Приложении I, которая способна в соответствии с данным набором определенных человеком задач генерировать такие результаты, как контент, прогнозы, рекомендации или решения, влияющие на окружающую среду, с которой она взаимодействует». Это определение системы ИИ в законодательной базе нацелено на максимальную технологическую нейтральность и «независимость от будущего развития», учитывая стремительные технологические и рыночные изменения в этой сфере.

Для обеспечения необходимой правовой определенности Раздел I дополнен Приложением I, в котором содержится подробный перечень подходов и приемов развития ИИ, адаптируемых к новому технологическому сценарию.

Ключевые участники цепочки создания и приращения стоимости в сфере ИИ также четко определены как поставщики и пользователи систем ИИ, включая государственных и частных операторов, что обеспечивает однородную конкурентную среду. Их можно обобщить следующим образом:

(а) подходы к машинному обучению, включая контролируемое, неконтролируемое обучение и обучение с подкреплением, с использованием широкого спектра методов, включая глубокое обучение;

(б) подходы на основе логики и знаний, включая представление знаний, индуктивное (логическое) программирование, базы знаний, механизмы логических и дедуктивных выводов, принятие решений (в том числе через манипулирование с символами) и экспертные системы;

(в) статистические оценки, байесовское оценивание, методы поиска и оптимизации.

В дальнейшем в свете Раздела 6 Преамбулы система ИИ должна быть четко определена, чтобы обеспечить правовую определенность, при этом не ограничивая гибкость для соответствия будущим технологическим достижениям. Это определение должно быть основано на сущностных функциональных характеристиках программного обеспечения, в частности, на способности генерировать результаты, такие как контент, прогнозы, рекомендации или решения, влияющие на окружающую среду, с которой взаимодействует система. При разработке систем ИИ в них может быть заложена способность действовать с различной степенью автономии и использоваться как отдельно, так и в качестве компонентов другого продукта, независимо от того, интегрирована ли эта система в данный продукт физически (встроенная система) или обслуживает функционал продукта без интеграции с ним (невстроенная система). Кроме того, определение системы ИИ должно быть дополнено перечнем конкретных методов и подходов, применяемых при ее разработке, причем этот перечень должен постоянно обновляться в свете технологических и рыночных изменений путем принятия Комиссией подзаконных актов.

В настоящий момент проект закона об искусственном интеллекте представляется попыткой дать определения и классификации, однако с точки зрения проблем защиты от дискриминации ему не хватает механизмов судебной защиты человека от дискриминационных автоматизированных решений, особенно в свете того, что должно было стать основой для (новой) нормативно-правовой базы, центральную роль в которой должен был занять человек.

Единственной отсылки к предмету ст. 22 общего регламента по защите данных недостаточно, так как она не покрывает обширную область искусственного интеллекта, но фокусируется в основном на правовых последствиях автоматического принятия решений, за исключением случаев, когда такое решение, в том числе профилирование, необходимо в контексте договорных отношений или закреплено в законодательстве Евросоюза или страны – члена ЕС. В этой связи в доктрине было показано, что решения, предусмотренные ст. 22 общего регламента по защите данных, не могут основываться на особых категориях персональных данных, например, на биометрических данных, без разрешения субъекта данных (ст. 9(2)(а) общего регламента по защите данных), кроме случаев, когда затронуты государственные интересы (ст. 9(2)(g)). Тем не менее эти исключения должны обеспечивать четкую законодательную базу для защиты фундаментальных прав (Martini, 2020) или учета инвестиций в данную область, несмотря на возможную связь со ст. 5 проекта закона об искусственном интеллекте.

В этом контексте очевидно, что ст. 22 общего регламента по защите данных не может восполнить отсутствие процедуры, соответствующей цели и задачам проекта закона об искусственном интеллекте. Это отсутствие становится еще более вопиющим, если вспомнить, что проект Закона об ИИ определяет системы искусственного интеллекта как «высокорисковые», но не предлагает никакого механизма (и не усиливает ст. 22 общего регламента по защите данных с этой целью) для эффективной защиты от дискриминации, порождаемой рисками для безопасности, здоровья или негативного влияния на фундаментальные права.

Можно предположить, что ст. 13 проекта закона об ИИ (озаглавленная «Прозрачность в предоставлении информации пользователям») дополняет ст. 22 общего регламента по защите данных в рамках подхода, заложенного в проекте закона об ИИ. Формулировки ст. 13 не позволяют сделать такой вывод, так как в ней лишь провозглашаются благие намерения. Действительно, маловероятно, что рядовой пользователь сможет взаимодействовать с алгоритмом или понять механизм достижения его результатов, каким бы прозрачным он ни был, хотя высокорисковые системы ИИ должны иметь «инструкции по использованию в соответствующем цифровом или нецифровом формате; таковые инструкции должны включать краткую, полную, верную и точную информацию, которая является необходимой, доступной и понятной пользователям»⁷.

Статья 13 проекта закона об ИИ прямо предусматривает необходимость корректности (точности) информации, но не ее истинности (достоверности). Эти два термина не являются точными синонимами ни в итальянском, ни в английском языке. Корректность информации подразумевает точность конкретных деталей, что может означать также истинность, но не обязательно. Таким образом, в тех областях, где эти два понятия не совпадают, возможна дискриминация или неверная автоматизированная обработка данных, при этом санкций или наказания за это не предусмотрено.

Проект текста рекомендаций поддерживает подход к этике искусственного интеллекта, предложенный ЮНЕСКО⁸. Однако этот проект делегирует обязанность по выработке механизмов против дискриминации операторам ИИ, которые «должны

⁷ Статья 13 проекта закона об ИИ.

⁸ ЮНЕСКО. *Проект рекомендаций по этике искусственного интеллекта*. Генеральная конференция ЮНЕСКО, Париж.

прилагать все разумные усилия, чтобы минимизировать или избежать усиления или распространения дискриминационных или необъективных приложений и результатов в течение всего жизненного цикла системы ИИ, чтобы обеспечить справедливость таких систем»⁹.

Эффективные механизмы борьбы против дискриминации и негативной предвзятости алгоритмических решений должны быть доступны для использования, и это является задачей общеевропейского законодателя и всех стран – членов ЕС. Если оставить эти механизмы на усмотрение отдельных операторов, то это усилит фрагментацию, но не снизит дискриминацию в результатах алгоритмических решений.

Предписывающий характер содержания и требований к такой информации позволяет сделать вывод о том, что ст. 13 проекта закона об ИИ не может служить таким механизмом. Она лишь дает указания относительно содержания информации.

Поправки к проекту закона об ИИ выглядят привлекательными, потому что они не срывают своей противоречивости, возникающей из-за стремления выработать «универсальное» законодательство, как провозглашается создателями и сторонниками проекта. Действительно, как и в случае Общего регламента по защите данных, проект закона об ИИ позиционируется как модель для регулирования ИИ в других правовых системах. Однако вызывает сомнения, что такая сложная и противоречивая модель, полная ограничений и исключений, может быть адаптирована к системам, в которых автоматизация принятия решений используется государством в первую очередь для повышения эффективности бюрократии и законопорядка, а также для укрепления оборонительного или наступательного военного аппарата, а частным сектором – для совершенствования анализа экономической эффективности через автоматизацию процесса производства.

С другой стороны, поправки демонстрируют фундаментальное противоречие с концепцией ИИ, представленной Европейским парламентом и, шире, общественным мнением или электоратом. Докладчик Европарламента напомнил, что системы искусственного интеллекта основываются на программном обеспечении, которое использует вероятностные математические модели и алгоритмические прогнозы для нескольких конкретных целей. Напротив, «искусственный интеллект» – это общее понятие, охватывающее широкий спектр старых и новых технологий, методов и подходов, более точно характеризуемое как «системы искусственного интеллекта»; это понятие относится к любой системе на основе вычислительной машины и часто означает лишь то, что ею управляет конкретный набор целей, определяемых человеком, но она обладает некоей степенью автономии в своих действиях. Такие системы делают прогнозы, дают рекомендации или решения на основе доступных им данных. Развитие таких технологий происходит неравномерно; некоторые уже широко используются, другие еще только разрабатываются или находятся на стадии обдумывания их конкретных черт¹⁰.

⁹ Там же.

¹⁰ Восс, Аксель. (2022, 20 апреля). *Проект доклада по проблемам искусственного интеллекта в цифровую эпоху (2020/2266(INI))*. Согласование поправок в Европарламенте; Проект законодательной резолюции Европарламента по Проекту регламента Европарламента и Совета Европы по координации норм в области искусственного интеллекта (Закон об искусственном интеллекте) и поправкам к некоторым законодательным актам Евросоюза (COM2021/0206 – C9-0146/2021 – 2021/0106(COD))

Противоположные позиции по поводу определения искусственного интеллекта нашли выражение в поправках, предложенных объединенным Комитетом Европарламента по проблемам гражданских свобод, правосудия и внутренних дел (European Parliament's Committee on Civil Liberties, Justice and Home Affairs, LIBE) и Комитетом по вопросам внутреннего рынка и защиты прав потребителей (Internal Market and Consumer Protection, IMCO)¹¹. С одной стороны, представлено максимально широкое определение ИИ, перекрывающее технические классификации, предложенные в Приложении I к проекту закона; с другой – декларируется необходимость точных дефиниций, в том числе относительно машинного обучения (которое определяется как способность находить закономерности без прямого программирования для конкретных задач).

Вводится новый критерий классификации высокорисковых систем, при этом автоматическая классификация систем в списке направлений, перечисленных в Приложении III, заменяется списком «ситуаций критического использования». На его основе провайдеры ИИ должны самостоятельно оценивать, представляют ли их системы значительные риски для здоровья, безопасности и фундаментальных прав. В процессе принятия этих поправок велись споры по поводу алгоритмов, используемых для оценки кредитоспособности, страхования здоровья, платежей и рекомендательных систем.

Представляется, что правильным было бы не только обозначить намерения и идеологические предпосылки, но и признать, что необходимо избегать укорененных в обществе предубеждений, и тем более недопустимо усиливать их через низкокачественные базы данных. Известно, что при обучении алгоритмы становятся дискриминационными в той же мере, в какой дискриминационны данные, с которыми они работают. Используя низкокачественные базы данных для обучения, а также предубеждения и дискриминационные взгляды, характерные для общества, алгоритмы могут порождать изначально дискриминирующие решения, что послужит усилению этой проблемы. Отмечается, однако, что предвзятость ИИ можно скорректировать. Кроме того, для минимизации риска необходимо применять технические средства и устанавливать различные уровни контроля над программным обеспечением систем ИИ, алгоритмами и данными, которые они используют и генерируют. Наконец, утверждают, что ИИ может и должен использоваться для снижения уровня предубежденности и дискриминации, но это, вероятно, заблуждение. На самом деле ИИ демонстрирует тенденцию усиления дискриминации в обществе.

Закон об искусственном интеллекте мог бы дать возможность продвинуться вперед от модели «изучения кейсов» для алгоритмов оценки риска, которая, напротив, главным образом и основывается на этих кейсах, фокусируясь на частных, а не коллективных рисках, тогда как ключевым аспектом является влияние алгоритмов оценки риска на права человека. Учитывая отраслевой характер каждого субъекта, именно такой подход следует воспринять всем причастным к созданию или формулированию алгоритмов. Это важнейший шаг от чисто культурной или эмпирической точки зрения, так как он позволит изменить антидискриминационные подходы при использовании данных.

¹¹ Бенифей, Брандо, Тудораш, Иоан-Драгош. *Проект доклада по Проекту регламента Европарламента и Совета Европы по координации норм в области искусственного интеллекта (Закон об искусственном интеллекте) и поправкам к некоторым законодательным актам Евросоюза (COM2021/0206 – C9-0146/2021 – 2021/0106(COD))*. Комитет по вопросам внутреннего рынка и защиты прав потребителей. Комитет Европарламента по проблемам гражданских свобод, правосудия и внутренних дел.

Не менее важен вопрос о запрете на использование технологий, содержание которого очень обширно. По этому пункту проект закона также постулирует широкий спектр исключений: несомненный и юридически значимый факт состоит в том, что проект целиком построен вокруг классификации типов алгоритмов, при этом центральным элементом концепции является запрет на использование алгоритмов социального рейтинга и специального биометрического программного обеспечения. Тем не менее широкий набор исключений оставляет множество противоречий, особенно в области биометрии. В проекте это объясняется тем, что, по крайней мере, в странах – членах ЕС действуют ограничения в отношении защиты достоинства и фундаментальных прав человека, которые нельзя нарушать при использовании алгоритмов искусственного интеллекта.

В свете вышесказанного можно также задать вопрос, является ли введение так называемых песочниц подходящим решением для уравнивания скорости технологического прогресса с необходимостью защиты прав человека, особенно с точки зрения борьбы с дискриминацией, при принятии решений с помощью алгоритмов. На этот вопрос нельзя ответить быстро и, может быть, на него нельзя ответить положительно. Действительно, песочница – это ограниченное во времени функциональное пространство для экспериментирования с системами ИИ, особенно принимающими решения, с целью их вывода на рынок. Цель их создания – оценить влияние автоматизированного принятия решений на отдельных граждан и на общество в целом. Однако это влияние не всегда обнаруживается немедленно, но может проявиться лишь после их функционирования в средне- или долгосрочной перспективе.

Внесение множества поправок демонстрирует глубокое недоверие политических сил в Европарламенте по отношению к Комиссии, предложившей данный проект, что усложняет процесс принятия поправок или окончательного одобрения проекта закона об ИИ.

Заключение

Развитие технологий не только способствует прогрессу и улучшает нашу жизнь. Оно также несет в себе угрозу для прав человека в области нарушения неприкосновенности частной жизни и дискриминации. Дискриминация часто имеет место при автоматизированной обработке персональных данных. Поскольку этот процесс не нейтрален и не может быть таковым, так как основан на выборе, постольку он содержит возможности для дискриминации. Сегодня важно определить, можно ли и если да, то каким образом принять различные меры в отношении справедливости действия алгоритмов.

Результаты настоящей работы могут использоваться в качестве основы для будущих исследований в сфере алгоритмической дискриминации и защиты неприкосновенности частной жизни, а также в законотворческом процессе.

Список литературы

- Abdollahpouri, H., Mansoury, M., Burke, R., & Mobasher, B. (2020). The connection between popularity bias, calibration, and fairness in recommendation. In *Proceedings of the 14th ACM Conference on Recommender Systems* (pp. 726–731). <https://doi.org/10.1145/3383313.3418487>
- Ainis, M. (2015). *La piccola eguaglianza*. Einaudi.
- Alpa, G. (2021). Quale modello normativo europeo per l'intelligenza artificiale? *Contratto e impresa*, 37(4), 1003–1026.

- Alpa, G., & Resta, G. (2006). *Trattato di diritto civile. Le persone e la famiglia: 1. Le persone fisiche ei diritti della personalità*. UTET giuridica.
- Altenried, M. (2020). The platform as factory: Crowdwork and the hidden labour behind artificial intelligence. *Capital & Class*, 44(2), 145–158. <https://doi.org/10.1177/0309816819899410>
- Amodio, E. (1970). *L'obbligo costituzionale di motivare e l'istituto della giuria*. *Rivista di diritto processuale*.
- Angiolini, C. S. A. (2020). *Lo statuto dei dati personali: uno studio a partire dalla nozione di bene*. Giappichelli.
- Bao, M., Zhou, A., Zottola, S., Brubach, B., Desmarais, S., Horowitz, A., ... & Venkatasubramanian, S. (2021). It's complicated: The messy relationship between rai datasets and algorithmic fairness benchmarks. *arXiv preprint arXiv:2106.05498*
- Bargi, A. (1997). *Sulla struttura normativa della motivazione e sul suo controllo in Cassazione*. Giur. it.
- Battini, S. (2018). *Indipendenza e amministrazione fra diritto interno ed europeo*.
- Bellamy, R. (2014). Citizenship: Historical development of. Citizenship: Historical Development of'. In J. Wright (Ed.), *International Encyclopaedia of Social and Behavioural Sciences*, Elsevier. <https://doi.org/10.1016/b978-0-08-097086-8.62078-0>
- Berk, R., Heidari, H., Jabbari, S., Kearns, M., & Roth, A. (2021). Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research*, 50(1), 3–44. <https://doi.org/10.1177/0049124118782533>
- Brooks, R. (2017). *Machine Learning Explained. Robots, AI and other stuff*.
- Bodei, R. (2019). *Dominio e sottomissione*. Bologna, Il Mulino.
- Canetti, E. (1960). *Masse und Macht*. Hamburg, Claassen.
- Casonato, C., & Marchetti, B. (2021). Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale. *BioLaw Journal-Rivista di BioDiritto*, 3, 415–437.
- Chizzini, A. (1998). *Sentenza nel diritto processuale civile*. Dig. disc. priv., Sez. civ.
- Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data*, 5(2), 153–163. <https://doi.org/10.1089/big.2016.0047>
- Citino, Y. (2022). *Cittadinanza digitale a punti e social scoring: le pratiche scorrette nell'era dell'intelligenza artificiale*. Diritti comparati.
- Claeys, G. (2018). *Marx and Marxism*. Nation Books, New York.
- Cockburn, I. M., Henderson, R., & Stern, S. (2018). The impact of artificial intelligence on innovation: An exploratory analysis. In *The economics of artificial intelligence: An agenda*. University of Chicago Press.
- Cossette-Lefebvre, H., & Maclure, J. (2022). AI's fairness problem: understanding wrongful discrimination in the context of automated decision-making. *AI and Ethics*, 5, 1–15. <https://doi.org/10.1007/s43681-022-00233-w>
- Crawford, K. (2021). Time to regulate AI that interprets human emotions. *Nature*, 592(7853), 167. <https://doi.org/10.1038/d41586-021-00868-5>
- Custers, B. (2022). AI in Criminal Law: An Overview of AI Applications in Substantive and Procedural Criminal Law. In B. H. M. Custers, & E. Fosch Villaronga (Eds.), *Law and Artificial Intelligence* (pp. 205–223). Heidelberg: Springer. <http://dx.doi.org/10.2139/ssrn.4331759>
- De Gregorio, G. & Paolucci F. (2022). *Dati personali e AI Act. Media laws*.
- Di Rosa, G. (2021). Quali regole per i sistemi automatizzati "intelligenti"? *Rivista di diritto civile*, 67(5), 823–853.
- Epp, C. R. (1996). Do bills of rights matter? The Canadian Charter of Rights and Freedoms, *American Political Science Review*, 90(4), 765–779.
- Fanchiotti, V. (1995). *Processo penale nei paesi di Common Law*. Dig. Disc. Pen.
- Freeman, C., Louçã, F., & Louçã, F. (2001). *As time goes by: from the industrial revolutions to the information revolution*. Oxford University Press.
- Freeman, K. (2016). Algorithmic injustice: How the Wisconsin Supreme Court failed to protect due process rights in State v. Loomis. *North Carolina Journal of Law & Technology*, 18(5), 75–90.
- Fuchs, C. (2014). *Digital Labour and Karl Marx*. Routledge.
- Gallese, C. (2022). *Legal aspects of the use of continuous-learning models in Telemedicine*. JURISIN.
- Gallese, E. Falletti, M. S. Nobile, L. Ferrario, Schettini, F. & Foglia, E. (2020). Preventing litigation with a predictive model of COVID-19 ICUs occupancy. *2020 IEEE International Conference on Big Data (Big Data)*. (pp. 2111–2116). Atlanta, GA, USA. <https://doi.org/10.1109/BigData50022.2020.9378295>
- Garg, P., Villasenor, J., & Foggo, V. (2020). Fairness metrics: A comparative analysis. In *2020 IEEE International Conference on Big Data (Big Data)* (pp. 3662–3666). IEEE. <https://doi.org/10.1109/bigdata50022.2020.9378025>
- Gressel, S., Pauleen, D. J., & Taskin, N. (2020). *Management decision-making, big data and analytics*. Sage.
- Guo, F., Li, F., Lv, W., Liu, L., & Duffy, V. G. (2020). Bibliometric analysis of affective computing researches during 1999–2018. *International Journal of Human-Computer Interaction*, 36(9), 801–814. <https://doi.org/10.1080/10447318.2019.1688985>

- Hildebrandt, M. (2021). The issue of bias. The framing powers of machine learning. In Pelillo, M., & Scantamburlo, T. (Eds.), *Machines We Trust: Perspectives on Dependable AI*. MIT Press. <https://doi.org/10.7551/mitpress/12186.003.0009>
- Hoffrage, U., & Marewski, J. N. (2020). Social Scoring als Mensch-System-Interaktion. *Social Credit Rating: Reputation und Vertrauen beurteilen*, 305–329. https://doi.org/10.1007/978-3-658-29653-7_17
- Iftene, A. (2018). *Who Is Worthy of Constitutional Protection? A Commentary on Ewert v Canada*.
- Infantino, M., & Wang, W. (2021). Challenging Western Legal Orientalism: A Comparative Analysis of Chinese Municipal Social Credit Systems. *European Journal of Comparative Law and Governance*, 8(1), 46–85. <https://doi.org/10.1163/22134514-bja10011>
- Israni, E. (2017). *Algorithmic due process: mistaken accountability and attribution in State v. Loomis*.
- Kleinberg, J., Mullainathan, S., & Raghavan, M. (2016). Inherent trade-offs in the fair determination of risk scores. *arXiv preprint arXiv:1609.05807*.
- Krawiec, A., Pawela, Ł., & Puchała, Z. (2023). Discrimination and certification of unknown quantum measurements. *arXiv preprint arXiv:2301.04948*.
- Kubat, M., & Kubat, J. A. (2017). *An introduction to machine learning* (Vol. 2, pp. 321–329). Cham, Switzerland: Springer International Publishing.
- Kuhn, Th. S. (1962). The structure of scientific revolutions. *International Encyclopedia of Unified Science*, 2(2).
- Lippert-Rasmussen, K. (2022). Algorithm-Based Sentencing and Discrimination, *Sentencing and Artificial Intelligence* (pp. 74–96). Oxford University Press.
- Maamar, N. (2018). Social Scoring: Eine europäische Perspektive auf Verbraucher-Scores zwischen Big Data und Big Brother. *Computer und Recht*, 34(12), 820–828. <https://doi.org/10.9785/cr-2018-341212>
- Mannozi, G. (1997). Sentencing. *Dig. Disc. Pen.*
- Marcus, G., & Davis, E. (2019). *Rebooting AI: Building artificial intelligence we can trust*. Vintage.
- Martini, M. (2020). Regulating Algorithms – How to demystify the alchemy of code?. In *Algorithms and Law* (pp. 100–135). Cambridge University Press. <https://doi.org/10.1017/9781108347846.004>
- Marx, K. (2016). Economic and philosophic manuscripts of 1844. In *Social Theory Re-Wired*. Routledge
- Massa, M. (1990). *Motivazione della sentenza (diritto processuale penale)*. Enc. Giur.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.
- Messinetti, R. (2019). La tutela della persona umana versus l'intelligenza artificiale. Potere decisionale dell'apparato tecnologico e diritto alla spiegazione della decisione automatizzata, *Contratto e impresa*, 3, 861–894.
- Mi, F., Kong, L., Lin, T., Yu, K., & Faltings, B. (2020). Generalised class incremental learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 240–241). <https://doi.org/10.1109/cvprw50498.2020.00128>
- Mitchell, T. M. (2007). *Machine learning* (Vol. 1). New York: McGraw-hill.
- Nazir, A., Rao, Y., Wu, L., & Sun, L. (2020). Issues and challenges of aspect-based sentiment analysis: A comprehensive survey. *IEEE Transactions on Affective Computing*, 13(2), 845–863. <https://doi.org/10.1109/taffc.2020.2970399>
- Oswald, M. (2018). Algorithm-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2128), 20170359. <https://doi.org/10.1098/rsta.2017.0359>
- Oswald, M., Grace, J., Urwin, S., & Barnes, G. C. (2018). Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality. *Information & communications technology law*, 27(2), 223–250.
- Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., & Wermter, S. (2019). Continual lifelong learning with neural networks: A review. *Neural networks*, 113, 54–71.
- Parona, L. (2021). Government by algorithm": un contributo allo studio del ricorso all'intelligenza artificiale nell'esercizio di funzioni amministrative. *Giornale Dir. Amm*, 1.
- Pellecchia, E. (2018). Profilazione e decisioni automatizzate al tempo della black box society: qualità dei dati e leggibilità dell'algoritmo nella cornice della responsible research and innovation. *Nuove leg. civ. comm*, 1209–1235.
- Pessach, D., & Shmueli, E. (2020). Algorithmic fairness. *arXiv preprint arXiv:2001.09784*.
- Petronio, U. (2020). *Il precedente negli ordinamenti giuridici continentali di antico regime*. *Rivista di diritto civile*, 66(5), 949–983.
- Pleiss, G., Raghavan, M., Wu, F., Kleinberg, J., & Weinberger, K. Q. (2017). On fairness and calibration. *Advances in neural information processing systems*, 30.

- Poria, S., Hazarika, D., Majumder, N., & Mihalcea, R. (2020). Beneath the tip of the iceberg: Current challenges and new directions in sentiment analysis research, *IEEE Transactions on Affective Computing*. <https://doi.org/10.1109/taffc.2020.3038167>
- Rebitschek, F. G., Gigerenzer, G., & Wagner, G. G. (2021). People underestimate the errors made by algorithms for credit scoring and recidivism prediction but accept even fewer errors. *Scientific reports*, 11(1), 1–11.
- Rodotà, S. (1995). *Tecnologie e diritti*, il Mulino. Bologna.
- Rodotà, S. (2012). *Il diritto di avere diritti*. Gius. Laterza.
- Rodotà, S. (2014). *Il mondo nella rete: Quali i diritti, quali i vincoli*. GLF Editori Laterza.
- Russell, P. H. (1983). The political purposes of the Canadian Charter of Rights and Freedoms. *Can. B. Rev.*, 61, 30–35.
- Scassa, T. (2021). Administrative Law and the Governance of Automated Decision Making: A Critical Look at Canada's Directive on Automated Decision Making, *UBCL Rev*, 54, 251–255. <https://doi.org/10.2139/ssrn.3722192>
- Scherer, M. U. (2015). Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies, *Harv. JL & Tech.*, 29, 353–360. <https://doi.org/10.2139/ssrn.2609777>
- Schiavone, A. (2019). *Eguaglianza*. Einaudi.
- Starr, S. B. (2014). Evidence-based sentencing and the scientific rationalisation of discrimination. *Stanford Law Review*, 66, 803–872.
- Stuurman, K., & Lachaud, E. (2022). Regulating AI. A label to complete the proposed Act on Artificial Intelligence. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3963890>
- Sunstein, C. R. (2019). Algorithms, correcting biases. *Social Research: An International Quarterly*, 86(2), 499–511. <https://doi.org/10.1353/sor.2019.0024>
- Tarrant, A., & Cowen, T. (2022). Big Tech Lobbying in the EU. *The Political Quarterly*, 93(2), 218–226. <https://doi.org/10.1111/1467-923x.13127>
- Taruffo, M. (1975). *La motivazione della sentenza civile*. Cedam, Padova.
- Vale, D., El-Sharif, A., & Ali, M. (2022). Explainable artificial intelligence (XAI) post-hoc explainability methods: Risks and limitations in non-discrimination law. *AI and Ethics*, 1–12. <https://doi.org/10.1007/s43681-022-00142-y>
- Veale, M., & Borgesius, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act-Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, 22(4), 97–112. <https://doi.org/10.31235/osf.io/38p5f>
- Vogel, P. A. (2020). "Right to explanation" for algorithmic decisions?, *Data-Driven Decision Making. Law, Ethics, Robotics, Health*, 49, 1–12. <https://doi.org/10.48550/arXiv.1606.08813>
- Von Tunzelmann, N. (2003). Historical coevolution of governance and technology in the industrial revolutions, *Structural Change and Economic Dynamics*, 14(4), 365–384. [https://doi.org/10.1016/s0954-349x\(03\)00029-8](https://doi.org/10.1016/s0954-349x(03)00029-8)
- Wang, C., Han, B., Patel, B., & Rudin, C. (2022). In pursuit of interpretable, fair and accurate machine learning for criminal recidivism prediction, *Journal of Quantitative Criminology*, 6, 1–63. <https://doi.org/10.1007/s10940-022-09545-w>
- Witt, A. C. (2022). Platform Regulation in Europe – Per Se Rules to the Rescue?, *Journal of Competition Law & Economics*, 18(3), 670–708. <https://doi.org/10.1093/joclec/nhac001>
- Woodcock, J. (2020). The algorithmic panopticon at Deliveroo: Measurement, precarity, and the illusion of control, *Ephemera: theory & politics in organisations*, 20(3), 67–95.
- York, J. C. (2022). *Silicon values: The future of free speech under surveillance capitalism*. Verso Books, London-New York.
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile books, London.

Сведения об авторе



Элена Фаллетти – доктор наук, доцент, Университет Карло Каттанео

Адрес: Корсо Маттеотти 22, Кастелланца, 21053, Италия

E-mail: efalletti@liuc.it

ORCID ID: <https://orcid.org/0000-0002-6121-6775>

Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=57040979500>

Конфликт интересов

Автор заявляет об отсутствии конфликта интересов.

Финансирование

Исследование не имело спонсорской поддержки.

Тематические рубрики

Рубрика OECD: 5.05 / Law

Рубрика ASJC: 3308 / Law

Рубрика WoS: OM / Law

Рубрика ГРНТИ: 10.27.51 / Осуществление и защита гражданских прав

Специальность ВАК: 5.1.3 / Частно-правовые (цивилистические) науки

История статьи

Дата поступления – 24 февраля 2023 г.

Дата одобрения после рецензирования – 13 апреля 2023 г.

Дата принятия к опубликованию – 16 июня 2023 г.

Дата онлайн-размещения – 20 июня 2023 г.



Research article

DOI: <https://doi.org/10.21202/jdtl.2023.16>



Algorithmic Discrimination and Privacy Protection

Elena Falletti 

Università Carlo Cattaneo – LIUc
Castellanza, Italy

Keywords

Algorithm,
artificial intelligence,
data protection,
digital technologies,
discrimination,
law,
personal data,
privacy,
private life,
regulation

Abstract

Objective: emergence of digital technologies such as Artificial intelligence became a challenge for states across the world. It brought many risks of the violations of human rights, including right to privacy and the dignity of the person. That is why it is highly relevant to research in this area. That is why this article aims to analyse the role played by algorithms in discriminatory cases. It focuses on how algorithms may implement biased decisions using personal data. This analysis helps assess how the Artificial Intelligence Act proposal can regulate the matter to prevent the discriminatory effects of using algorithms.

Methods: the methods used were empirical and comparative analysis. Comparative analysis allowed to compare regulation of and provisions of Artificial Intelligence Act proposal. Empirical analysis allowed to analyse existing cases that demonstrate us algorithmic discrimination.

Results: the study's results show that the Artificial Intelligence Act needs to be revised because it remains on a definitional level and needs to be sufficiently empirical. Author offers the ideas of how to improve it to make more empirical.

Scientific novelty: the innovation granted by this contribution concerns the multidisciplinary study between discrimination, data protection and impact on empirical reality in the sphere of algorithmic discrimination and privacy protection.

© Falletti E., 2023



This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Practical significance: the beneficial impact of the article is to focus on the fact that algorithms obey instructions that are given based on the data that feeds them. Lacking abductive capabilities, algorithms merely act as obedient executors of the orders. Results of the research can be used as a basis for further research in this area as well as in law-making process.

For citation

Falletti, E. (2023). Algorithmic Discrimination and Privacy Protection. *Journal of Digital Technologies and Law*, 1(2), 387–420. <https://doi.org/10.21202/jdtl.2023.16>

References

- Abdollahpouri, H., Mansoury, M., Burke, R., & Mobasher, B. (2020). The connection between popularity bias, calibration, and fairness in recommendation. In *Proceedings of the 14th ACM Conference on Recommender Systems* (pp. 726–731). <https://doi.org/10.1145/3383313.3418487>
- Ainis, M. (2015). *La piccola eguaglianza*. Einaudi.
- Alpa, G. (2021). Quale modello normativo europeo per l'intelligenza artificiale? *Contratto e impresa*, 37(4), 1003–1026.
- Alpa, G., & Resta, G. (2006). *Trattato di diritto civile. Le persone e la famiglia: 1. Le persone fisiche e i diritti della personalità*. UTET giuridica.
- Altenried, M. (2020). The platform as factory: Crowdwork and the hidden labour behind artificial intelligence. *Capital & Class*, 44(2), 145–158. <https://doi.org/10.1177/0309816819899410>
- Amodio, E. (1970). *L'obbligo costituzionale di motivare e l'istituto della giuria*. *Rivista di diritto processuale*.
- Angiolini, C. S. A. (2020). *Lo statuto dei dati personali: uno studio a partire dalla nozione di bene*. Giappichelli.
- Bao, M., Zhou, A., Zottola, S., Brubach, B., Desmarais, S., Horowitz, A., ... & Venkatasubramanian, S. (2021). It's complicated: The messy relationship between rai datasets and algorithmic fairness benchmarks. *arXiv preprint arXiv:2106.05498*
- Bargi, A. (1997). *Sulla struttura normativa della motivazione e sul suo controllo in Cassazione*. Giur. it.
- Battini, S. (2018). *Indipendenza e amministrazione fra diritto interno ed europeo*.
- Bellamy, R. (2014). Citizenship: Historical development of. *Citizenship: Historical Development of*. In J. Wright (Ed.), *International Encyclopaedia of Social and Behavioural Sciences*, Elsevier. <https://doi.org/10.1016/b978-0-08-097086-8.62078-0>
- Berk, R., Heidari, H., Jabbari, S., Kearns, M., & Roth, A. (2021). Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research*, 50(1), 3–44. <https://doi.org/10.1177/0049124118782533>
- Brooks, R. (2017). *Machine Learning Explained. Robots, AI and other stuff*.
- Bodei, R. (2019). *Dominio e sottomissione*. Bologna, Il Mulino.
- Canetti, E. (1960). *Masse und Macht*. Hamburg, Claassen.
- Casonato, C., & Marchetti, B. (2021). Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale. *BioLaw Journal-Rivista di BioDiritto*, 3, 415–437.
- Chizzini, A. (1998). *Sentenza nel diritto processuale civile*. Dig. disc. priv., Sez. civ.
- Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data*, 5(2), 153–163. <https://doi.org/10.1089/big.2016.0047>
- Citino, Y. (2022). *Cittadinanza digitale a punti e social scoring: le pratiche scorrette nell'era dell'intelligenza artificiale*. Diritti comparati.
- Claeys, G. (2018). *Marx and Marxism*. Nation Books, New York.
- Cockburn, I. M., Henderson, R., & Stern, S. (2018). The impact of artificial intelligence on innovation: An exploratory analysis. In *The economics of artificial intelligence: An agenda*. University of Chicago Press.
- Cossette-Lefebvre, H., & Maclure, J. (2022). AI's fairness problem: understanding wrongful discrimination in the context of automated decision-making. *AI and Ethics*, 5, 1–15. <https://doi.org/10.1007/s43681-022-00233-w>

- Crawford, K. (2021). Time to regulate AI that interprets human emotions. *Nature*, 592(7853), 167. <https://doi.org/10.1038/d41586-021-00868-5>
- Custers, B. (2022). AI in Criminal Law: An Overview of AI Applications in Substantive and Procedural Criminal Law. In B. H. M. Custers, & E. Fosch Villaronga (Eds.), *Law and Artificial Intelligence* (pp. 205–223). Heidelberg: Springer. <http://dx.doi.org/10.2139/ssrn.4331759>
- De Gregorio, G. & Paolucci F. (2022). *Dati personali e AI Act. Media laws*.
- Di Rosa, G. (2021). Quali regole per i sistemi automatizzati “intelligenti”? *Rivista di diritto civile*, 67(5), 823–853.
- Epp, C. R. (1996). Do bills of rights matter? The Canadian Charter of Rights and Freedoms, *American Political Science Review*, 90(4), 765–779.
- Fanchiotti, V. (1995). *Processo penale nei paesi di Common Law*. Dig. Disc. Pen.
- Freeman, C., Louçã, F., & Louçã, F. (2001). *As time goes by: from the industrial revolutions to the information revolution*. Oxford University Press.
- Freeman, K. (2016). Algorithmic injustice: How the Wisconsin Supreme Court failed to protect due process rights in *State v. Loomis*. *North Carolina Journal of Law & Technology*, 18(5), 75–90.
- Fuchs, C. (2014). *Digital Labour and Karl Marx*. Routledge.
- Gallese, C. (2022). *Legal aspects of the use of continuous-learning models in Telemedicine*. JURISIN.
- Gallese, E. Falletti, M. S. Nobile, L. Ferrario, F. Schettini, & Foglia E. (2020). Preventing litigation with a predictive model of COVID-19 ICUs occupancy. *2020 IEEE International Conference on Big Data (Big Data)*. (pp. 2111–2116). Atlanta, GA, USA. <https://doi.org/10.1109/BigData50022.2020.9378295>
- Garg, P., Villasenor, J., & Foggo, V. (2020). Fairness metrics: A comparative analysis. In *2020 IEEE International Conference on Big Data (Big Data)* (pp. 3662–3666). IEEE. <https://doi.org/10.1109/bigdata50022.2020.9378025>
- Gressel, S., Pauleen, D. J., & Taskin, N. (2020). *Management decision-making, big data and analytics*. Sage.
- Guo, F., Li, F., Lv, W., Liu, L., & Duffy, V. G. (2020). Bibliometric analysis of affective computing researches during 1999–2018. *International Journal of Human-Computer Interaction*, 36(9), 801–814. <https://doi.org/10.1080/10447318.2019.1688985>
- Hildebrandt, M. (2021). The issue of bias. The framing powers of machine learning. In Pelillo, M., & Scantamburlo, T. (Eds.), *Machines We Trust: Perspectives on Dependable AI*. MIT Press. <https://doi.org/10.7551/mitpress/12186.003.0009>
- Hoffrage, U., & Marewski, J. N. (2020). Social Scoring als Mensch-System-Interaktion. *Social Credit Rating: Reputation und Vertrauen beurteilen*, 305–329. https://doi.org/10.1007/978-3-658-29653-7_17
- Iftene, A. (2018). *Who Is Worthy of Constitutional Protection? A Commentary on Ewert v Canada*.
- Infantino, M., & Wang, W. (2021). Challenging Western Legal Orientalism: A Comparative Analysis of Chinese Municipal Social Credit Systems. *European Journal of Comparative Law and Governance*, 8(1), 46–85. <https://doi.org/10.1163/22134514-bja10011>
- Israni, E. (2017). *Algorithmic due process: mistaken accountability and attribution in State v. Loomis*.
- Kleinberg, J., Mullainathan, S., & Raghavan, M. (2016). Inherent trade-offs in the fair determination of risk scores. *arXiv preprint arXiv:1609.05807*.
- Krawiec, A., Pawela, Ł., & Puchała, Z. (2023). Discrimination and certification of unknown quantum measurements. *arXiv preprint arXiv:2301.04948*.
- Kubat, M., & Kubat, J. A. (2017). *An introduction to machine learning* (Vol. 2, pp. 321–329). Cham, Switzerland: Springer International Publishing.
- Kuhn, Th. S. (1962). The structure of scientific revolutions. *International Encyclopedia of Unified Science*, 2(2).
- Lippert-Rasmussen, K. (2022). Algorithm-Based Sentencing and Discrimination, *Sentencing and Artificial Intelligence* (pp. 74–96). Oxford University Press.
- Maamar, N. (2018). Social Scoring: Eine europäische Perspektive auf Verbraucher-Scores zwischen Big Data und Big Brother. *Computer und Recht*, 34(12), 820–828. <https://doi.org/10.9785/cr-2018-341212>
- Mannozi, G. (1997). Sentencing. *Dig. Disc. Pen*.
- Marcus, G., & Davis, E. (2019). *Rebooting AI: Building artificial intelligence we can trust*. Vintage.
- Martini, M. (2020). Regulating Algorithms – How to demystify the alchemy of code?. In *Algorithms and Law* (pp. 100–135). Cambridge University Press. <https://doi.org/10.1017/9781108347846.004>
- Marx, K. (2016). Economic and philosophic manuscripts of 1844. In *Social Theory Re-Wired*. Routledge
- Massa, M. (1990). *Motivazione della sentenza (diritto processuale penale)*. Enc. Giur.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

- Messinetti, R. (2019). La tutela della persona umana versus l'intelligenza artificiale. Potere decisionale dell'apparato tecnologico e diritto alla spiegazione della decisione automatizzata, *Contratto e impresa*, 3, 861–894.
- Mi, F., Kong, L., Lin, T., Yu, K., & Faltings, B. (2020). Generalised class incremental learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 240–241). <https://doi.org/10.1109/cvprw50498.2020.00128>
- Mitchell, T. M. (2007). *Machine learning* (Vol. 1). New York: McGraw-hill.
- Nazir, A., Rao, Y., Wu, L., & Sun, L. (2020). Issues and challenges of aspect-based sentiment analysis: A comprehensive survey. *IEEE Transactions on Affective Computing*, 13(2), 845–863. <https://doi.org/10.1109/taffc.2020.2970399>
- Oswald, M. (2018). Algorithm-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2128), 20170359. <https://doi.org/10.1098/rsta.2017.0359>
- Oswald, M., Grace, J., Urwin, S., & Barnes, G. C. (2018). Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality. *Information & communications technology law*, 27(2), 223–250.
- Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., & Wermter, S. (2019). Continual lifelong learning with neural networks: A review. *Neural networks*, 113, 54–71.
- Parona, L. (2021). Government by algorithm": un contributo allo studio del ricorso all'intelligenza artificiale nell'esercizio di funzioni amministrative, in *Giornale Dir. Amm.*
- Pellecchia, E. (2018). Profilazione e decisioni automatizzate al tempo della black box society: qualità dei dati e leggibilità dell'algoritmo nella cornice della responsible research and innovation. *Nuove leg. civ. comm*, 1209–1235.
- Pessach, D., & Shmueli, E. (2020). Algorithmic fairness. *arXiv preprint arXiv:2001.09784*.
- Petronio, U. (2020). Il precedente negli ordinamenti giuridici continentali di antico regime. *Rivista di diritto civile*, 66(5), 949–983.
- Pleiss, G., Raghavan, M., Wu, F., Kleinberg, J., & Weinberger, K. Q. (2017). On fairness and calibration. *Advances in neural information processing systems*, 30.
- Poria, S., Hazarika, D., Majumder, N., & Mihalcea, R. (2020). Beneath the tip of the iceberg: Current challenges and new directions in sentiment analysis research, *IEEE Transactions on Affective Computing*. <https://doi.org/10.1109/taffc.2020.3038167>
- Rebitschek, F. G., Gigerenzer, G., & Wagner, G. G. (2021). People underestimate the errors made by algorithms for credit scoring and recidivism prediction but accept even fewer errors. *Scientific reports*, 11(1), 1–11.
- Rodotà, S. (1995). *Tecnologie e diritti*, il Mulino. Bologna.
- Rodotà, S. (2012). *Il diritto di avere diritti*. Gius. Laterza.
- Rodotà, S. (2014). *Il mondo nella rete: Quali i diritti, quali i vincoli*. GLF Editori Laterza.
- Russell, P. H. (1983). The political purposes of the Canadian Charter of Rights and Freedoms. *Can. B. Rev.*, 61, 30–35.
- Scassa, T. (2021). Administrative Law and the Governance of Automated Decision Making: A Critical Look at Canada's Directive on Automated Decision Making, *UBCL Rev*, 54, 251–255. <https://doi.org/10.2139/ssrn.3722192>
- Scherer, M. U. (2015). Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies, *Harv. JL & Tech.*, 29, 353–360. <https://doi.org/10.2139/ssrn.2609777>
- Schiavone, A. (2019). *Eguaglianza*. Einaudi.
- Starr, S. B. (2014). Evidence-based sentencing and the scientific rationalisation of discrimination. *Stanford Law Review*, 66, 803–872.
- Stuurman, K., & Lachaud, E. (2022). Regulating AI. A label to complete the proposed Act on Artificial Intelligence. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3963890>
- Sunstein, C. R. (2019). Algorithms, correcting biases. *Social Research: An International Quarterly*, 86(2), 499–511. <https://doi.org/10.1353/sor.2019.0024>
- Tarrant, A., & Cowen, T. (2022). Big Tech Lobbying in the EU. *The Political Quarterly*, 93(2), 218–226. <https://doi.org/10.1111/1467-923x.13127>
- Taruffo, M. (1975). *La motivazione della sentenza civile*. Cedam, Padova.
- Vale, D., El-Sharif, A., & Ali, M. (2022). Explainable artificial intelligence (XAI) post-hoc explainability methods: Risks and limitations in non-discrimination law. *AI and Ethics*, 1–12. <https://doi.org/10.1007/s43681-022-00142-y>

- Veale, M., & Borgesius, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act-Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, 22(4), 97–112. <https://doi.org/10.31235/osf.io/38p5f>
- Vogel, P. A. (2020). "Right to explanation" for algorithmic decisions?, *Data-Driven Decision Making. Law, Ethics, Robotics, Health*, 49, 1–12. <https://doi.org/10.48550/arXiv.1606.08813>
- Von Tunzelmann, N. (2003). Historical coevolution of governance and technology in the industrial revolutions, *Structural Change and Economic Dynamics*, 14(4), 365–384. [https://doi.org/10.1016/s0954-349x\(03\)00029-8](https://doi.org/10.1016/s0954-349x(03)00029-8)
- Wang, C., Han, B., Patel, B., & Rudin, C. (2022). In pursuit of interpretable, fair and accurate machine learning for criminal recidivism prediction, *Journal of Quantitative Criminology*, 6, 1–63. <https://doi.org/10.1007/s10940-022-09545-w>
- Witt, A. C. (2022). Platform Regulation in Europe – Per Se Rules to the Rescue?, *Journal of Competition Law & Economics*, 18(3), 670–708. <https://doi.org/10.1093/joclec/nhac001>
- Woodcock, J. (2020). The algorithmic panopticon at Deliveroo: Measurement, precarity, and the illusion of control, *Ephemera: theory & politics in organisations*, 20(3), 67–95.
- York, J. C. (2022). *Silicon values: The future of free speech under surveillance capitalism*. Verso Books, London-New York.
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile books, London.

Author information



Elena Faletti – PhD, Assistant Professor, Carlo Cattaneo University LIUC

Address: Corso Matteotti 22, Castellanza, 21053, Italy

E-mail: efalletti@liuc.it

ORCID ID: <https://orcid.org/0000-0002-6121-6775>

Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=57040979500>

Conflict of interest

The author declares no conflict of interest.

Financial disclosure

The research had no sponsorship.

Thematic rubrics

OECD: 5.05 / Law

PASJC: 3308 / Law

WoS: OM / Law

Article history

Date of receipt – February 24, 2023

Date of approval – April 13, 2023

Date of acceptance – June 16, 2023

Date of online placement – June 20, 2023

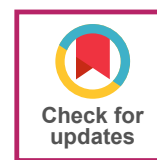


Научная статья

УДК 340.143:004.8

EDN: <https://elibrary.ru/lbwsxw>

DOI: <https://doi.org/10.21202/jdtl.2023.17>



Подходы к регулированию отношений в сфере разработки и применения технологий искусственного интеллекта: особенности и практическая применимость

Ольга Сергеевна Ерахтина 

Пермский филиал Национального исследовательского университета «Высшая школа экономики»
г. Пермь, Российская Федерация

Ключевые слова

Искусственный интеллект, опасность, право, программное обеспечение, регулирование, рискориентированный подход, технологический подход, управление рисками, цифровая экономика, цифровые технологии

Аннотация

Цель: обзор сложившихся в науке подходов к регулированию отношений в сфере применения технологий искусственного интеллекта, выявление основных особенностей и ограничений применения рискориентированного и технологического подходов для определения направлений их дальнейшего развития.

Методы: методологическую основу исследования составляет совокупность методов научного познания, в том числе общенаучный диалектический и универсальные научные методы (анализ и синтез, сравнение, обобщение, структурно-функциональный, формально-логический).

Результаты: определено, что применение рискориентированного подхода предполагает построение конструктивных моделей управления рисками. Значимым вопросом для применения данного подхода является вопрос об основаниях отнесения технологий искусственного интеллекта к высокорисковым. При определении уровня риска применения технологии искусственного интеллекта необходимо применять следующие критерии: тип технологии искусственного интеллекта, сферу ее применения, а также уровень ее потенциальной опасности для окружающей среды, здоровья, других фундаментальных прав граждан.

В свою очередь, центральным вопросом для применения технологического подхода является вопрос о необходимости и пределах регулирования сферы разработки и применения технологий искусственного интеллекта. Во-первых, вмешательство в данную сферу не должно создавать препятствий для развития технологий и инноваций.

© Ерахтина О. С., 2023



Статья находится в открытом доступе и распространяется в соответствии с лицензией Creative Commons «Attribution» («Атрибуция») 4.0 Всемирная (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0/deed.ru>), позволяющей неограниченно использовать, распространять и воспроизводить материал при условии, что оригинальная работа упомянута с соблюдением правил цитирования.

Во-вторых, естественная реакция регулятора в ответ на появление новых объектов и субъектов оборота – «синдром несовершенного закона». Вместе с тем ложное представление об отсутствии правового регулирования может дать обратный эффект – дублирование правовых норм. В целях решения проблемы дублирования законодательных требований следует прежде всего решить вопрос о том, необходимо ли регулировать технологии искусственного интеллекта или некоторые виды использования программного обеспечения.

Научная новизна: проведен обзор основных подходов к регулированию отношений в сфере разработки и применения технологий искусственного интеллекта, выявлены возможности и ограничения их применения, предложены дальнейшие направления их развития.

Практическая значимость: основные положения и выводы исследования могут быть использованы для определения оптимальных подходов к регулированию сферы цифровых технологий, а также в целях совершенствования правового регулирования рассматриваемой области общественных отношений.

Для цитирования

Ерахтина, О. С. (2023). Подходы к регулированию отношений в сфере разработки и применения технологий искусственного интеллекта: особенности и практическая применимость. *Journal of Digital Technologies and Law*, 1(2), 421–437. <https://doi.org/10.21202/jdtl.2023.17>

Содержание

Введение

1. Рискоориентированный подход

2. Технологический подход

Выводы

Список литературы

Введение

Активное развитие технологий и систем искусственного интеллекта порождает научные дискуссии о необходимости, пределах и задачах правового регулирования отношений в сфере ИТ. Учеными высказываются принципиально разные позиции: от утверждений о необходимости установления значительного количества обязательных требований, исполнение которых в большей степени возлагается на разработчика (Smuha, 2021), до предложений не осуществлять правовое вмешательство в сферу высоких технологий¹, чтобы не препятствовать инновациям. В палитру позиций ученых входят также более умеренные взгляды: применение международных и национальных стандартов (Zielke, 2020); осуществление добровольной сертификации (Ellul et al., 2021); мягкое регулирование и саморегулирование (Erdélyi & Goldsmith, 2018); установление объяснимых (Hamon et al., 2022) и этических (Wagner, 2018) рамок.

¹ O'Sullivan, Andrea. (2017, October 24). *Don't Let Regulators Ruin AI*. <https://www.technologyreview.com/2017/10/24/3937/dont-let-regulators-ruin-ai/>

Бельгийский исследователь Николя Пети предлагает так называемый регуляторный компромисс, достигаемый посредством соотношения угроз и возможностей, создаваемых введением правового регулирования². Он приводит ряд примеров, когда регулирование создавало препятствия для технологического процесса³. Вместе с тем отмечает, что отсутствие регулирования также может помешать технологической эволюции. В частности, правовая неопределенность оказывает негативное влияние на инвестиции. По мнению Райана Кало, неопределенные правила ответственности могут служить препятствием для инвестиций в открытые рынки робототехники и направить поток капитала на узкую функциональность роботов, где производители могут лучше управлять рисками, оставляя открытую робототехнику недостаточно развитой (Calo, 2011).

Среди множества подходов к правовому регулированию отношений в сфере применения систем искусственного интеллекта выделяют подход, согласно которому определяется общий правовой режим, который должен предусматривать основные требования к обеспечению безопасности систем искусственного интеллекта. Данный режим будет распространяться на все подобные системы. Наряду с этим необходимо разработать детальные требования к разработке и применению искусственного интеллекта в конкретных сферах (Понкин, Редькина, 2018).

Высокая динамика развития технологий искусственного интеллекта, наличие множества регуляторных инициатив актуализируют значимость междисциплинарных исследований, нацеленных на выявление оптимальных подходов к регулированию рассматриваемой сферы общественных отношений.

В статье представлен комплексный анализ двух междисциплинарных подходов к регулированию отношений в сфере разработки и применения технологий искусственного интеллекта – рискориентированного и технологического подходов, выявлены возможности и ограничения их применения, а также определены направления их дальнейшего развития.

1. Рискориентированный подход

Одним из активно обсуждаемых в науке подходов к регулированию технологий искусственного интеллекта является рискориентированный подход (Михалева, Шубина, 2019; Gellert, 2021; Gonçalves, 2020). Идея применения данного подхода была анонсирована в резолюции Европарламента «Нормы гражданского права в робототехнике» в 2017 г.⁴ Через четыре года в проекте Регламента Европейского союза, устанавливающего согласованные правила в отношении искусственного интеллекта⁵, была предложена классификация систем искусственного интеллекта, основанная

² Petit, N. (2017, March 9). *Law and Regulation of Artificial Intelligence and Robots – Conceptual Framework and Normative Implications* (Working paper, pp. 6–7). <http://dx.doi.org/10.2139/ssrn.2931339>

³ Там же. Р. 12.

⁴ European Parliament. (2017, February 15). *Draft Report with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103 (INL))*. <https://www.europarl.europa.eu/portal/en>

⁵ European Parliament. (2021, April 21). *Regulation of the European Parliament and of the Council. Laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts*. <https://www.europarl.europa.eu/portal/en>

на оценке уровня риска их применения. В соответствии с данной классификацией системы искусственного интеллекта разделены на четыре группы:

1) системы искусственного интеллекта с неприемлемо высоким уровнем риска (прежде всего, системы искусственного интеллекта, применяемые в военном и оборонном секторах, а также системы для манипулирования поведением человека и системы для создания социальных рейтингов);

2) системы искусственного интеллекта с высоким уровнем риска⁶;

3) системы искусственного интеллекта с ограниченным риском⁷;

4) системы искусственного интеллекта с минимальным риском⁸.

Для оценки уровня риска применения системы искусственного интеллекта предполагается использование двух основных критериев: степень зависимости пользователей от решений, которые принимает система, и степень ее опасности для жизни и здоровья граждан и нарушения их основных прав.

В проекте Регламента ЕС указывается на необходимость установления запрета на производство и введение в гражданский оборот систем искусственного интеллекта, применение которых создает неприемлемо высокий уровень риска причинения вреда. В свою очередь к разработчикам, владельцам и пользователям высокорисковых систем искусственного интеллекта, как следует из проекта, должны предъявляться повышенные требования к обеспечению безопасности, ведению технической документации и предоставлению информации⁹. В целях соблюдения установленных требований могут быть использованы системы менеджмента качества. Для систем же с ограниченным и минимальным риском как минимум должна быть установлена обязанность информировать пользователей о том, что они взаимодействуют с искусственным интеллектом.

В целом подход ЕС заслуживает поддержки. Вместе с тем он нуждается в дальнейшем развитии.

Во-первых, необходимо учитывать, что рискориентированный подход – это система управления рисками, которая состоит из трех основных этапов. На первом этапе необходимо выявить, проанализировать и дифференцировать прогнозируемые риски. На втором – провести оценку уровня риска. На третьем этапе – определить способы управления рисками.

При всем многообразии подходов к классификации рисков (в идеале все эти подходы должны быть обязательно учтены) для целей данной статьи можно выделить:

- внутренние и внешние риски;
- системные и несистемные риски.

⁶ Например, роботы-ассистенты в хирургии, управление и эксплуатация критической инфраструктурой.

⁷ Например, чат-боты, виртуальные помощники, умные дома.

⁸ Например, видеоигры, спам-фильтры.

⁹ При разработке и применении таких систем должны соблюдаться «принципиальные требования», такие как требования к качеству данных, документированию и прослеживаемости, человеческому контролю и др. В частности, до размещения на рынке или введения в эксплуатацию должна быть составлена техническая документация, в которой необходимо отразить соответствие системы установленным требованиям, а также всю необходимую информацию для оценки системы на соответствие (статья 11). Проект также требует разрабатывать системы таким образом, чтобы обеспечивались ведение учета во время работы системы (статья 12), а также прозрачность системы ИИ и предоставление информации (статья 13), с тем чтобы пользователи высокорисковых систем могли использовать их надлежащим образом и корректно интерпретировать выходные данные. Контроль со стороны человека должен осуществляться на протяжении всего жизненного цикла системы с возможностью в любой момент вмешаться в работу системы и при необходимости остановить или скорректировать ее.

Внутренние риски возникают в отдельно взятой компании, их можно спрогнозировать, оценить и предотвратить своими силами. Внешние риски (изменение экономической и политической ситуации, стихийные бедствия, экологические катастрофы и др.) предотвратить силами отдельной компании нельзя.

Системные риски угрожают рынку в целом или отдельным сферам бизнеса. Спрогнозировать их возникновение и оценить последствия могут только эксперты. Их также нельзя предотвратить силами отдельной компании. В то время как несистемные (коммерческие) риски – это риски отдельной компании, которые она может и должна минимизировать самостоятельно.

Концепция рискориентированного подхода в понимании резолюции Европарламента «Нормы гражданского права в робототехнике» направлена на управление внутренними и несистемными рисками и оставляет без внимания внешние и системные риски, которые не могут быть вменены отдельным субъектам гражданского оборота. Выявление таких рисков – это одна из основных задач, которая должна быть решена на первом этапе.

На втором этапе определяется уровень риска (вероятность материализации риска и размер неблагоприятных последствий, которые могут вследствие этого наступить). На первый взгляд, проект решает данную задачу. Вместе с тем основания для отнесения системы искусственного интеллекта к одной из четырех групп требуют дальнейшего исследования. Как уже было указано, в проекте Регламента предлагается в качестве критерия оценки уровня риска использовать степень зависимости пользователей от решений, которые принимает искусственный интеллект, и степень опасности технологии для жизни и здоровья граждан и нарушения их основных прав. На наш взгляд, оценка вероятности материализации риска зависит и от ряда других факторов. Прежде всего, это характеристики технологии. Основными характеристиками искусственного интеллекта являются его автономность (способность самостоятельно, без вмешательства человека принимать решения) и обучаемость (способность овладевать новыми навыками и способностями). При этом в зависимости от сферы применения выделяют несколько уровней автономности технологий. Так, в Приложении № 10 к Транспортной стратегии РФ до 2030 г. с прогнозом на период до 2035 г.¹⁰ определено пять уровней автономности автомобильного транспорта, четыре уровня автономности железнодорожного транспорта и шесть уровней автономности водного и морского транспорта.

По критерию обучаемости искусственный интеллект может быть необучаемым, обучаемым и самообучаемым. Очевидно, что высокоавтономные и самообучаемые технологии должны быть отнесены к высокорисковым системам искусственного интеллекта, а необучаемые, имеющие первый или второй уровень автономности, – к системам с минимальным риском. Однако решение задачи определения уровня риска применения технологии искусственного интеллекта на основе ее характеристики не является настолько простой, насколько может показаться на первый взгляд. Прежде всего потому, что при оценке риска наряду с автономностью и способностью к обучению необходимо учитывать и другие характеристики искусственного интеллекта. В частности, функциональность (способность выполнять одну или более функций) технологии, оснащенность средствами контроля (Алексеев и др., 2020). Кроме того, очевидно, что возможны различные комбинации

¹⁰ Утверждена распоряжением Правительства РФ от 27.11.2021 № 3363-р. <https://base.garant.ru/403156321/>

указанных характеристик. Так, например, технология искусственного интеллекта может быть высокоавтономной, обучаемой, выполняющей две функции и не содержать средства объективного контроля.

В научных работах также предлагается в качестве критерия оценки риска применения технологий искусственного интеллекта учитывать сферу их применения. По мнению группы исследователей из Мальтийского университета, правовое регулирование должно быть обязательным только для критических областей и сфер деятельности. При этом наряду со сферой применения систем искусственного интеллекта необходимо проводить оценку уровня риска тех видов деятельности, при осуществлении которых он используется (Ellul et al., 2021). С данным выводом следует согласиться. Так, в сфере здравоохранения технологии искусственного интеллекта могут применяться для помощи врачам при постановке диагноза, назначении лекарственных препаратов, проведении операций. Указанные виды деятельности могут быть отнесены к категории высокого риска. Наряду с этим технологии искусственного интеллекта используются для регистрации пациентов, обработки и анализа медицинских данных, автоматического оповещения медицинского персонала. Представляется, что данные виды деятельности могут быть отнесены к группе ограниченного риска.

Сферу транспорта большинство авторов также относят к высокорисковой сфере. Однако и здесь необходимо учитывать, для сопровождения каких видов деятельности применяется искусственный интеллект. Системы искусственного интеллекта используются для повышения безопасности и эффективности транспортного процесса, управления пассажиропотоками и грузовыми потоками. Вместе с тем технологии искусственного интеллекта также используются при оказании услуг по перевозке грузов и пассажиров. Если деятельность по управлению транспортной инфраструктурой относится к категории высокого риска, то деятельность по оказанию услуг может быть отнесена к категории ограниченного риска.

Во-вторых, в дальнейших исследованиях нуждается и вопрос о способах управления рисками. Проект уделяет основное внимание высокорисковым системам, оставляя, по сути, без внимания системы искусственного интеллекта с ограниченным риском. Вместе с тем специальное регулирование (основанное на рискориентированном подходе) должно в том числе осуществляться и применительно к системам искусственного интеллекта, относящимся к данной группе.

Полагаем, что способы управления рисками будут различаться в зависимости от вида риска (внутренний или внешний, системный или несистемный), а также уровня «рисковости» технологии искусственного интеллекта (высокий или средний).

Заслуживающей внимания является «Базовая модель определения критериев и категорий риска», утвержденная в 2017 г. проектным комитетом приоритетной программы «Реформа контрольной и надзорной деятельности»¹¹. В данном документе раскрываются такие понятия, как «источники риска», «факторы риска», «профиль риска», определяются их виды, предлагаются способы ранжирования управляемых факторов и профилей риска (в целях определения наиболее значимых из них)

¹¹ «Базовая модель определения критериев и категорий риска» (утв. протоколом заседания проектного комитета от 31.03.2017 № 19(3)) (вместе с «Требованиями к обоснованию предлагаемых федеральными органами исполнительной власти – участниками приоритетной программы “Реформа контрольной и надзорной деятельности” категорий риска (классов опасности) и критериев риска в отношении осуществляемых ими видов государственного контроля (надзора)»).

и методика определения величины причиняемого вреда и возможной частоты возникновения потенциальных негативных последствий.

Несмотря на то, что рассматриваемая модель была принята в целях осуществления надзорными органами «умных проверок», чтобы сконцентрировать проверки на потенциально наиболее опасных объектах, предлагаемые в нем методики определения категорий и критериев риска могут быть использованы и при оценке уровня риска применения технологий искусственного интеллекта.

2. Технологический подход

В последнее время в научной литературе все большее внимание уделяется технологическому подходу, в центре внимания которого находятся сама технология, ее сущностные и специфические характеристики. Рассматривая развитие технологии и инноваций в качестве основной задачи правового регулирования, представители данного подхода прежде всего задаются вопросом о необходимости и пределах регулирования сферы высоких технологий. Размышляя над этим вопросом, Дж. Эллул, приходит к выводу о том, что правовое регулирование должно быть сосредоточено не на искусственном интеллекте, а на программном обеспечении. По его мнению, в случае критического программного обеспечения, используемого, например, в самолете, не имеет значения, используется ли при этом искусственный интеллект или нет. Регулирование не должно касаться конкретной технологии искусственного интеллекта, оно должно быть более широким и осуществляться в отношении программного обеспечения в целом. В подтверждение данного вывода Эллул приводит еще один пример. Так, при использовании искусственного интеллекта в банковских или страховых системах, которые решают, следует ли предлагать конкретный кредит или полис, регулирование должно быть направлено на обеспечение того, чтобы клиенты не подвергались дискриминации. Данное требование должно применяться не только к системам на основе искусственного интеллекта. Вполне возможно запрограммировать систему для принятия решений, используя методы, не связанные с искусственным интеллектом (Ellul, 2022).

В целом поддерживая данную точку зрения, необходимо учитывать, что искусственный интеллект является собирательным понятием. В нормативно-правовых актах обоснованно проводится последовательное разграничение понятий «программа для ЭВМ»¹², «искусственный интеллект»¹³ и «технологии искусственного интеллекта»¹⁴.

¹² Статья 1261 Гражданского кодекса РФ.

¹³ Указ Президента Российской Федерации от 10 октября 2019 г. № 490 «О развитии искусственного интеллекта в Российской Федерации», которым утверждена Национальная стратегия развития искусственного интеллекта на период до 2030 г.

¹⁴ Статья 2 Федерального закона от 24 апреля 2020 г. № 123-ФЗ «О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта в субъекте Российской Федерации – городе федерального значения Москве и внесении изменений в статьи 6 и 10 Федерального закона «О персональных данных».

В примечании к п. 3.18 Национального стандарта «Системы искусственного интеллекта. Классификация систем искусственного интеллекта»¹⁵ указано на то, что искусственный интеллект как комплекс технологических решений включает в себя информационно-коммуникационную инфраструктуру, программное обеспечение (в том числе в котором используются методы машинного обучения), процессы и сервисы по обработке данных, анализу и синтезу решений.

В научной литературе также поднимается вопрос о том, является ли искусственный интеллект единым объектом либо это собирательное понятие (Балашова, 2022). В качестве решения данного вопроса Л. Ю. Василевской предлагается включить искусственный интеллект в установленный в ст. 1240 ГК РФ перечень сложных объектов интеллектуальных прав. По мнению, Л. Ю. Василевской, структурными элементами искусственного интеллекта являются программный продукт (программа для ЭВМ); программное обеспечение (совокупность программ); искусственные нейронные сети (программы для ЭВМ); алгоритмы, программное обеспечение в качестве ноу-хау; технические решения как изобретения; базы данных (Василевская и др., 2021).

Таким образом, несмотря на то, что ядром технологии искусственного интеллекта является компьютерная программа, полностью отождествлять эти понятия было бы неверным. Именно отличительные особенности систем искусственного интеллекта, их специфические характеристики и наличие структурных элементов, являющихся самостоятельными объектами гражданских прав, обуславливают особенности правового регулирования отношений в сфере их разработки и применения.

Применение технологического подхода призвано также не допустить дублирования законодательных требований в тех сферах, где они уже введены. В этой связи Дж. Эллул ставит еще один исследовательский вопрос: должно ли, например, программное обеспечение для планирования задач в календаре регулироваться в большей степени, чем того требуют существующие законы (например, Закон о защите персональных данных или Закон о защите прав потребителей) (Ellul, 2022)? Полагаю, что на данный вопрос следует ответить утвердительно. Развитие технологий меняет процесс взаимодействия участников оборота, а соответственно, содержание их прав и обязанностей. Так, ст. 12 Закона о защите прав потребителей устанавливает обязанность изготовителя (исполнителя, продавца) своевременно предоставлять потребителю необходимую и достоверную информацию о товарах (работах, услугах). Вместе с тем в цифровом обществе информация, как правило, предоставляется в цифровой форме. Наряду с этим появляются программы, облегчающие поиск информации. А. И. Савельев справедливо отмечает, что в перспективе на потребителя, по-видимому, будет возложен риск неознакомления с информацией, размещенной предпринимателем в общедоступных источниках (Савельев, 2016). Более того, законодатель вводит в гражданский оборот новых субъектов. Так, в июле 2018 г. Законом «О внесении изменений в Закон Российской Федерации «О защите прав потребителей»¹⁶ было установлено специальное регулирование деятельности владельца агрегатора информации о товарах (услугах).

¹⁵ ГОСТ Р 59277-2020. Национальный стандарт РФ. Системы искусственного интеллекта. Классификация систем искусственного интеллекта. Дата введения 01.03.2021.

¹⁶ О внесении изменений в Закон Российской Федерации «О защите прав потребителей»: Федеральный закон от 29.07.2018 № 250-ФЗ. http://www.consultant.ru/document/cons_doc_LAW_303537/

Применение технологий искусственного интеллекта при обработке и хранении персональных данных связано с рисками их утечки или некорректной интерпретации. С 1995 г. Европейский парламент решает проблему управления такими рисками¹⁷. В 2016 г. в Регламенте ЕС «О защите физических лиц в отношении обработки персональных данных и о свободном перемещении таких данных» была предложена модель, призванная решить задачу усиления защиты персональных данных. Применение данной модели, по мнению разработчиков Регламента, должно способствовать формированию доверительного отношения к технологиям¹⁸.

И, наконец, последний вопрос, который исследует Эллул в своей статье: должно ли обязательное регулирование быть ориентировано непосредственно на технологии либо на определенную сферу или вид деятельности, при осуществлении которой они применяются (Ellul, 2022)? При ответе на данный вопрос необходимо учитывать, что риск ошибки имманентно присущ программному обеспечению любой сложности, вне зависимости от того, обладает оно элементами искусственного интеллекта или нет. По мере того, как программные системы в целом (которые не обязательно используют искусственный интеллект) «растут в сложности, взаимосвязанности и географическом распределении, мы будем все чаще сталкиваться с их нежелательным эмерджентным поведением» (Mogul, 2006). То есть взаимодействие технологии с внешней средой создает дополнительные сложности и риски. В целях минимизации таких рисков прежде всего необходимо вводить стандарты качества программного обеспечения¹⁹. Затем необходимо проводить оценку уровня риска тех видов деятельности, при осуществлении которых используется технология (об этом мы писали в первой части нашей статьи). Если определенная деятельность относится к группе высокого риска, должно быть установлено требование обязательного применения соответствующего стандарта. Таким образом, обязательное регулирование должно быть ориентировано не на технологию, а на вид деятельности, при осуществлении которой она применяется.

Рассматривая основные вопросы, которые поднимаются представителями технологического подхода, нельзя оставить без внимания исследование Рональда Линса. Исследователь из Нидерландов указывает на то, что выявить наличие пробелов в правовом регулировании отношений в сфере применения технологий – это достаточно сложная задача. На первом этапе необходимо определить сущностные и специфические характеристики технологии. Следует ли при решении этой задачи сосредоточиться на конкретном примере технологии, например на беспилотных автомобилях, или перейти к более широкой категории – беспилотным транспортным средствам?

¹⁷ О защите прав частных лиц применительно к обработке персональных данных и о свободном движении таких данных: Директива ЕС 95/46/ЕС от 24 октября 1995 г. (Директива ЕС 95/46/ЕС «О персональных данных»).

¹⁸ О защите физических лиц в отношении обработки персональных данных и о свободном перемещении таких данных: Регламент ЕС 2016/679 от 27 апреля 2016 г.

¹⁹ См., например: ГОСТ 28195-89. Оценка качества программных средств. Общие положения. Введен 01.07.1990; ГОСТ 28806-90. Качество программных средств. Термины и определения. Введен 01.01.1992; ГОСТ Р 51188-98. Защита информации. Испытание программных средств на наличие компьютерных вирусов. Типовое руководство. Введен 01.07.1999.

По мнению Линса, ни один из этих подходов не является желательным. Фокусирование на определенной технологии может привести к тому, что регулятор сосредоточится на потенциально случайных особенностях технологии. Во втором случае из-за чрезмерного обобщения дискуссия может стать абстрактной, а потому бесполезной. Поэтому на данном этапе надо взять «социально-техническую призму» и наряду с определением специфических характеристик технологии выявить, чьи интересы должны быть важны и приоритетны (Leenes, 2019).

На втором этапе необходимо решить вопрос развития технологий, для чего выявить потенциальные риски и существующие проблемы, с которыми связано их применение. К сожалению, в научной литературе очень часто происходит смешение таких категорий, как «риск» и «проблема», в связи с чем сложно четко определить пределы и задачи правового регулирования рассматриваемой сферы общественных отношений. Представляется, что одним из основных рисков применения технологий искусственного интеллекта является возможность ее самопроизвольного отклонения от изначально заложенной в ней цели. Вследствие материализации данного риска возможно наступление определенных негативных последствий, таких как причинение вреда жизни, здоровью или имуществу пользователя, разглашение конфиденциальной информации.

На необходимость разграничения категорий «риск» и «проблема» указывают Е. А. Войниканис, Е. В. Семенова, Г. С. Тюляев. К рискам применения технологий искусственного интеллекта они, в частности, относят возможность деанонимизации данных, возможность дискриминации по признакам пола, расы, национальности, вероисповедания. В качестве проблем в работе выделяются вопросы о том, кто является правообладателем программы искусственного интеллекта, кто является ответственным за причинение вреда жизни или здоровью при эксплуатации искусственного интеллекта, и др. (Войниканис и др., 2018).

На третьем этапе следует определить формы и пределы государственного вмешательства в сферу технологий искусственного интеллекта. Способы управления рисками и решения проблем минимизации негативных последствий их материализации различны. Так, например, к способам управления рисками можно отнести ведение регистрационного журнала во время работы системы, а также обеспечение прозрачности процесса принятия решения, чтобы пользователи могли интерпретировать выходные данные. В отношении высокорисковых систем должна быть также введена система пострыночного мониторинга, в рамках которого будут осуществляться сбор и анализ данных о работе системы уже после ее вывода на рынок.

Если же, несмотря на принятие превентивных мер контроля, произошел сбой в работе системы, право должно предоставить способы справедливого распределения негативных последствий между ее разработчиком, пользователем и оператором.

Необходимо также учитывать, что способы воздействия на риски и проблемы различны и не всегда имеют правовую форму. При этом, по справедливому замечанию М. Шерера (Scherer, 2016), традиционные методы правового регулирования, такие как, например, лицензирование продукции, контроль за исследованиями, возможность применения деликтной ответственности, являются не вполне пригодными для управления рисками, связанными с применением систем искусственного интеллекта.

Выводы

Применение рискориентированного подхода предполагает построение конструктивных моделей управления рисками. Процесс управления рисками состоит из трех основных этапов. На первом этапе необходимо выявить и классифицировать все риски, с которыми сопряжено применение технологий искусственного интеллекта в определенной сфере. Концепция рискориентированного подхода, предлагаемая Европарламентом, концентрируется на внутренних и несистемных рисках. Соответственно, в целях развития данного подхода необходимо провести исследование по выявлению внешних и системных рисков. На втором этапе необходимо провести оценку уровня риска применения определенной технологии искусственного интеллекта. При проведении оценки необходимо использовать ряд критериев. Среди них – сущностные и специфические характеристики технологии, сфера и вид деятельности, при осуществлении которой применяется данная технология. На третьем этапе должны быть определены способы управления рисками, которые в свою очередь дифференцируются в зависимости от уровня рисковости определенной технологии. Как видим, основная цель применения рискориентированного подхода состоит в определении способов управления рисками, с которыми сопряжено применение технологий искусственного интеллекта.

В центре внимания технологического подхода находится вопрос о необходимости и пределах регулирования сферы высоких технологий. К основным этапам применения технологического подхода относят:

- определение сущностных и специфических характеристик технологий;
- выявление потенциальных рисков и уже существующих проблем их применения;
- определение форм и пределов государственного вмешательства в сферу технологий искусственного интеллекта.

Именно специфические характеристики технологий искусственного интеллекта, такие как автономность и самообучаемость, обуславливают особенности правового регулирования отношений в сфере их разработки и применения. При этом обязательное регулирование должно быть ориентировано не на технологию, а на вид деятельности, при осуществлении которой она применяется.

Проведенное исследование позволяет также сделать вывод о том, что универсальным подходом к регулированию отношений в сфере разработки и применения технологий является технологический подход. Несмотря на то, что данный подход нуждается в дальнейшем развитии, уже сейчас он может послужить основой для формирования стратегии правотворческой деятельности. В свою очередь, рискориентированный подход является одним из основных элементов технологического подхода. Эффективное управление сопутствующими рисками позволит минимизировать потенциальные негативные последствия применения новых технологий и обеспечит устойчивое развитие сферы высоких технологий.

Список литературы

- Алексеев, А. О., Ерахтина, О. С., Кондратьева, К. С., Никитин, Т. Ф. (2020). Подходы к гражданско-правовой ответственности разработчика технологий искусственного интеллекта: на основе классификации технологий. *Информационное общество*, 6, 47–57. <https://elibrary.ru/ylddab>
- Балашова, А. И. (2022). Искусственный интеллект в авторском и патентном праве: объекты, субъектный состав правоотношений, сроки правовой охраны. *Журнал Суда по интеллектуальным правам*, 2(36), 90–98. <https://elibrary.ru/apldua>

- Василевская, Л. Ю., Подузова, Е. Б., Тасалов, Ф. А. (2021). Цифровизация гражданского оборота: правовая характеристика «искусственного интеллекта» и «цифровых» субъектов (цивилистическое исследование) (в 5 т.). Москва: Проспект. <https://elibrary.ru/nrjkdo>
- Войниканис, Е. А., Семенова, Е. В., Тюляев, Г. С. (2018). Искусственный интеллект и право. Вызовы и возможности самообучающихся алгоритмов. *Вестник Воронежского государственного университета. Серия: Право*, 4(35), 137–148. <https://elibrary.ru/yumlhz>
- Михалева, Е. С., Шубина, Е. А. (2019). Проблемы и перспективы правового регулирования робототехники. *Актуальные проблемы российского права*, 12(109), 26–35. <https://doi.org/10.17803/1994-1471.2019.109.12.026-035>
- Понкин, И. В., Редькина, А. И. (2018). Искусственный интеллект с точки зрения права. *Вестник РУДН. Серия: Юридические науки*, 22(1), 91–109. <https://doi.org/10.22363/2313-2337-2018-22-1-91-109>
- Савельев, А. И. (2016). Направления эволюции свободы договора под влиянием современных информационных технологий. В сб. М. А. Рожкова (рук. авт. кол. и отв. ред.), *Свобода договора*. Москва: Статут. <https://elibrary.ru/xxoolt>
- Calo, R. (2011). Open robotics. *Maryland Law Review*, 70.3, 101–142.
- Ellul, J., Pace, G., McCarthy, S., Sammut, T., Brockdorf, J., & Scerri, M. (2021). Regulating artificial intelligence: a technology regulator's perspective. In *Proceedings of the Eighteenth International conference on artificial intelligence and law* (pp. 190–194). <https://doi.org/10.1145/3462757.3466093>
- Ellul, J. (2022). Should we regulate Artificial Intelligence or some uses of Software? *Discover Artificial Intelligence*, 2(5). <https://doi.org/10.1007/s44163-022-00021-9>
- Erdélyi, O. J., & Goldsmith, J. (2018). Regulating artificial intelligence: proposal for a global solution. In *Proceedings of the 2018 AAAI/ACM conference on AI, ethics, and society* (pp. 95–101).
- Gellert, R. (2021). The role of the risk-based approach in the General data protection Regulation and in the European Commission's proposed Artificial Intelligence Act: Business as usual? *Journal of Ethics and Legal Technologies*, 3(2), 15–33.
- Gonçalves, M. E. (2020). The risk-based approach under the new EU data protection regulation: a critical perspective. *Journal of Risk Research*, 23(2), 139–152. <https://doi.org/10.1080/13669877.2018.1517381>
- Hamon, R., Junklewitz, H., Sanchez, I., Malgieri, G., & De Hert, P. (2022). Bridging the gap between AI and explainability in the GDPR: towards trustworthiness-by-design in automated decision-making. *IEEE Comput Intell Mag.*, 17(1), 72–85. <https://doi.org/10.1109/mci.2021.3129960>
- Leenes, R. (2019). *Regulating New Technologies in Uncertain Times*. https://doi.org/10.1007/978-94-6265-279-8_2
- Mogul, J. C. (2006). Emergent (mis) behavior vs. complex software systems. *ACM SIGOPS Oper. Syst. Rev.*, 40(4), 293–304. <https://doi.org/10.1145/1218063.1217964>
- Scherer, M. U. (2016). Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies. *Harvard Journal of Law & Technology*, 29(2), 353–400. <https://doi.org/10.2139/ssrn.2609777>
- Smuha, N. A. (2021). From a 'race to AI' to a 'race to AI regulation': regulatory competition for artificial intelligence. *Law. Innov. Technol.*, 13(1), 57–84. <https://doi.org/10.1080/17579961.2021.1898300>
- Wagner, B. (2018). Ethics as an escape from regulation: from ethics-washing to ethics-shopping. In *Being profiling: cogitas ergo sum* (pp. 86–90). Amsterdam: Amsterdam University Press. <https://doi.org/10.1515/9789048550180-016>
- Zielke, T. (2020). Is artificial intelligence ready for standardization? In *European conference on software process improvement* (pp. 259–274). Springer.

Сведения об авторе



Ерахтина Ольга Сергеевна – кандидат юридических наук, доцент, доцент кафедры гражданского и предпринимательского права, Пермский филиал Национального исследовательского университета «Высшая школа экономики»

Адрес: 614070, Российская Федерация, г. Пермь, ул. Студенческая, 38

E-mail: oerahtina@hse.ru

ORCID ID: <https://orcid.org/0000-0002-9041-3487>

Web of Science Researcher ID:

<https://www.webofscience.com/wos/author/record/K-3149-2014>

Google Scholar ID: <https://scholar.google.ru/citations?hl=ru&user=WdwWB4kAAAAJ>

РИНЦ Author ID: https://www.elibrary.ru/author_items.asp?authorid=498773

Конфликт интересов

Автор заявляет об отсутствии конфликта интересов.

Финансирование

Исследование не имело спонсорской поддержки.

Тематические рубрики

Рубрика OECD: 5.05 / Law

Рубрика ASJC: 3308 / Law

Рубрика WoS: OM / Law

Рубрика ГРНТИ: 10.07.45 / Право и научно-технический прогресс

Специальность ВАК: 5.1.1 / Теоретико-исторические правовые науки

История статьи

Дата поступления – 10 февраля 2023 г.

Дата одобрения после рецензирования – 23 апреля 2023 г.

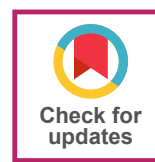
Дата принятия к опубликованию – 16 июня 2023 г.

Дата онлайн-размещения – 20 июня 2023 г.



Research article

DOI: <https://doi.org/10.21202/jdtl.2023.17>



Approaches to Regulating Relations in the Sphere of Developing and Using the Artificial Intelligence Technologies: Features and Practical Applicability

Olga S. Erahtina 

Perm branch of National Research University "Higher School of Economics"
Perm, Russian Federation

Keywords

Artificial intelligence,
danger,
digital economy,
digital technologies,
law,
regulation,
risk management,
risk-oriented approach,
software,
technological approach

Abstract

Objective: to review the modern scientific approaches to regulating relations in the sphere of using the artificial intelligence technologies; to reveal the main features and limitations of using the risk-oriented and technological approaches in order to determine the directions of their further development.

Methods: the methodological basis of the research is a set of scientific cognition methods, including the general scientific dialectic method and the universal scientific methods (analysis and synthesis, comparison, summarization, structural-functional, and formal-logical methods).

Results: it was determined that the use of the risk-oriented approach implies building constructive models of risk management. A significant issue in using this approach is the bases of referring the artificial intelligence technologies to high-risk ones. When determining the risk level of using the artificial intelligence technologies, the following criteria should be applied: the type of artificial intelligence technology, its sphere of use, and the level of potential harm for the environment, health and other fundamental human rights.

In turn, the central issue of using the technological approach is the necessity and limits of regulation in the sphere of developing and using the artificial intelligence technologies. First, interference into this sphere must not create obstacles for developing technologies and innovations. Second, a natural reaction of a regulator towards newly emerging objects and subjects of turnover is the "imperfect law syndrome". At the same time, a false idea

© Erakhtina O. S., 2023



This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

about a lack of legal regulation may produce an opposite effect – duplication of legal norms. To solve the problem of duplicating legal requirements, it is necessary, first of all, to solve the issue of the need to regulate the artificial intelligence technologies or certain types of software applications.

Scientific novelty: a review was carried out of the main approaches to regulating relations in the sphere of developing and using the artificial intelligence technologies; the opportunities and limitations of their use are revealed; further directions of their development are proposed.

Practical significance: the main provisions and conclusions of the research can be used for determining the optimal approaches to regulating the sphere of digital technologies and for improving the legal regulation of the studied sphere of social relations.

For citation

Erahtina, O. S. (2023). Approaches to Regulating Relations in the Sphere of Developing and Using the Artificial Intelligence Technologies: Features and Practical Applicability. *Journal of Digital Technologies and Law*, 1(2), 421–437. <https://doi.org/10.21202/jdtl.2023.17>

References

- Alekseev, A. O., Erahtina, O. S., Kondratyeva K. S., & Nikitin, T. Ph. (2020). Approaches to civil legal liability of the artificial intelligence technologies developer: based on the classification. *Information Society*, 6, 47–57. (In Russ.).
- Balashova, A. I. (2022). Artificial intelligence in copyright and patent law: objects, subject structure of legal relations, terms of legal protection. *Zhurnal Suda po intellektual'nym pravam*, 2(36), 90–98. (In Russ.).
- Calo, R. (2011). Open robotics. *Maryland Law Review*, 70.3, 101–142.
- Ellul, J., Pace, G., McCarthy, S., Sammut, T., Brockdorf, J., & Scerri, M. (2021). Regulating artificial intelligence: a technology regulator's perspective. In: *Proceedings of the Eighteenth International conference on artificial intelligence and law* (pp. 190–194). <https://doi.org/10.1145/3462757.3466093>
- Ellul, J. (2022). Should we regulate Artificial Intelligence or some uses of Software? *Discover Artificial Intelligence*, 2(5). <https://doi.org/10.1007/s44163-022-00021-9>
- Erdélyi, O. J., & Goldsmith, J. (2018). Regulating artificial intelligence: proposal for a global solution. In *Proceedings of the 2018 AAAI/ACM conference on AI, ethics, and society* (pp. 95–101).
- Gellert, R. (2021). The role of the risk-based approach in the General data protection Regulation and in the European Commission's proposed Artificial Intelligence Act: Business as usual? *Journal of Ethics and Legal Technologies*, 3(2), 15–33.
- Gonçalves, M. E. (2020). The risk-based approach under the new EU data protection regulation: a critical perspective. *Journal of Risk Research*, 23(2), 139–152. <https://doi.org/10.1080/13669877.2018.1517381>
- Hamon, R., Junklewitz, H., Sanchez, I., Malgieri, G., & De Hert, P. (2022). Bridging the gap between AI and explainability in the GDPR: towards trustworthiness-by-design in automated decision-making. *IEEE ComputIntell Mag.*, 17(1), 72–85. <https://doi.org/10.1109/mci.2021.3129960>
- Leenes, R. (2019). *Regulating New Technologies in Uncertain Times*. https://doi.org/10.1007/978-94-6265-279-8_2
- Mikhaleva, E. S., & Shubina, E. A. (2019). Challenges and Prospects of the Legal Regulation of Robotics. *Actual Problems of Russian Law*, 1(12), 26–35. (In Russ.). <https://doi.org/10.17803/1994-1471.2019.109.12.026-035>
- Mogul, J. C. (2006). Emergent (mis)behavior vs. complex software systems. *ACM SIGOPS Oper. Syst. Rev.*, 40(4), 293–304. <https://doi.org/10.1145/1218063.1217964>
- Ponkin, I. V., & Redkina, A. I. (2018). Artificial Intelligence from the Point of View of Law. *RUDN Journal of Law*, 22(1), 91–109. (In Russ.). <https://doi.org/10.22363/2313-2337-2018-22-1-91-109>

- Savelyev, A. I. (2016). Directions of freedom of contract evolution under the influence of modern information technologies. In M. A. Rozhkova (head of authors' collective and editor-in-chief), *Svoboda dogovora*. Moscow: Statut. (In Russ.).
- Scherer, M. U. (2016). Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies. *Harvard Journal of Law & Technology*, 29(2), 353–400. <https://doi.org/10.2139/ssrn.2609777>
- Smuha, N. A. (2021). From a 'race to AI' to a 'race to AI regulation': regulatory competition for artificial intelligence. *Law. Innov. Technol.*, 13(1), 57–84. <https://doi.org/10.1080/17579961.2021.1898300>
- Vasilevskaya, L. Yu., Poduzova, E. B., & Tasalov, F. A. (2021). *Digitalization of civil turnover: legal characteristics of "artificial intelligence" and "digital" subjects (civilistic research)* (In 5 Vol.). Moscow: Prospekt. (In Russ.).
- Voinikanis, E. A., Semenova, E. V., & Tyulyaev, G. S. (2018). Artificial intelligence and law: challenges and possibilities of self-learning algorithms. *Proceedings of Voronezh State University. Series: Pravo*, 4(35), 137–148. (In Russ.).
- Wagner, B. (2018). Ethics as an escape from regulation: from ethics-washing to ethics-shopping. In *Being profiling: cogitas ergo sum* (pp. 86–90). Amsterdam: Amsterdam University Press. <https://doi.org/10.1515/9789048550180-016>
- Zielke, T. (2020). Is artificial intelligence ready for standardization? In: *European conference on software process improvement* (pp. 259–274). Springer.

Author information



Olga S. Erahtina – Candidate of Sciences in Jurisprudence, Associate Professor, Department of Civil and Entrepreneurial Law, Perm branch of National Research University "Higher School of Economics"

Address: 38 Studencheskaya Str., 614070 Perm, Russian Federation

E-mail: oerahtina@hse.ru

ORCID ID: <https://orcid.org/0000-0002-9041-3487>

Web of Science Researcher ID:

<https://www.webofscience.com/wos/author/record/K-3149-2014>

Google Scholar ID: <https://scholar.google.ru/citations?hl=ru&user=WdwWB4kAAAAJ>

RSCI Author ID: https://www.elibrary.ru/author_items.asp?authorid=498773

Conflict of interests

The author declares no conflict of interests.

Financial disclosure

The research was not sponsored.

Thematic rubrics

OECD: 5.05 / Law

PASJC: 3308 / Law

WoS: OM / Law

Article history

Date of receipt – February 10, 2023

Date of approval – April 23, 2023

Date of acceptance – June 16, 2023

Date of online placement – June 20, 2023

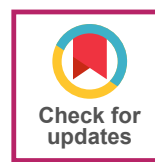


Научная статья

УДК 347.44:004.8

EDN: <https://elibrary.ru/jyqazw>

DOI: <https://doi.org/10.21202/jdtl.2023.18>



Проблемы и перспективы регулирования отношений в рамках сделки, совершенной с участием искусственного интеллекта

Дмитрий Александрович Казанцев 

B2B-Center

г. Москва, Российская Федерация

Ключевые слова

Автоматизация,
данные,
закупка,
искусственный интеллект,
ответственность,
право,
робот,
сделка,
цифровизация,
цифровые технологии

Аннотация

Цель: исследование проблемы определения субъекта юридически значимого действия, совершенного с использованием искусственного интеллекта, а также распределения ответственности за последствия его работы.

Методы: для иллюстрации проблематики и практической значимости вопроса о правосубъектности искусственного интеллекта были выбраны автоматизированные закупки для государственных и корпоративных нужд, а методологическую основу исследования составила совокупность методов научного познания, используемых для теоретико-правовых исследований, в том числе сравнения, ретроспективного анализа, аналогии и синтеза.

Результаты: на примере отрасли конкурентных закупок для государственных и корпоративных нужд проанализирована эволюция автоматизации хозяйственных отношений вплоть до внедрения искусственного интеллекта. Продемонстрированы успешно апробированные ответы на вызовы, обусловленные поэтапным внедрением цифровых технологий в хозяйственные отношения, а также соответствующие модификации правового регулирования. На основании текущего уровня развития технологий сформулированы перспективные вопросы правового регулирования хозяйственных отношений, реализуемых с использованием искусственного интеллекта, и прежде всего вопрос определения субъекта сделки, совершенной с использованием искусственного интеллекта. В качестве приглашения к дискуссии после анализа выводов право-ведов о возможных вариантах юридического статуса искусственного

© Казанцев Д. А., 2023



Статья находится в открытом доступе и распространяется в соответствии с лицензией Creative Commons «Attribution» («Атрибуция») 4.0 Всемирная (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0/deed.ru>), позволяющей неограниченно использовать, распространять и воспроизводить материал при условии, что оригинальная работа упомянута с соблюдением правил цитирования.

интеллекта предложены варианты ответа на вопрос о его правосубъектности при заключении сделки. Для решения вопроса об ответственности за решения, ставшие результатом реализации алгоритмов программно-аппаратного комплекса, предложено несколько моделей распределения такой ответственности между его создателем, его владельцем и иными лицами, действия которых могли повлиять на результаты работы такого алгоритма. Предложенные выводы могут использоваться для развития нормативного регулирования как в совокупности, так и по отдельности.

Научная новизна: в работе на основании анализа эволюции практик использования цифровых технологий в закупках сформулированы потенциальные проблемы правового порядка, обусловленные непрерывным процессом автоматизации хозяйственных отношений, а также предложены правовые конструкции для решения таких проблем.

Практическая значимость: выводы и предложения настоящей работы имеют перспективное значение для концептуального понимания и нормативного регулирования инструментов проведения закупок в электронной форме как на корпоративном, так и национальном уровне.

Для цитирования

Казанцев, Д. А. (2023). Проблемы и перспективы регулирования отношений в рамках сделки, совершенной с участием искусственного интеллекта. *Journal of Digital Technologies and Law*, 1(2), 438–463. <https://doi.org/10.21202/jdtl.2023.18>

Содержание

Введение

1. Пять цифровых трансформаций конкурентных закупок

1.1. Закупки в электронной форме

1.2. Цифровизация закупочной работы

1.3. Автоматизация создания документов

1.4. Автоматизация бизнес-процессов

1.5. Автоматизация решений

2. Правосубъектность искусственного интеллекта

3. Правовой статус сделки, совершенной с участием искусственного интеллекта

Выводы

Список литературы

Введение

Использование искусственного интеллекта с каждым годом становится все более востребованным в самых различных отраслях. Актуальность искусственного интеллекта является логичным и неизбежным следствием повсеместного распространения цифровых технологий.

Цифровые технологии уже давно используются не только для повседневной коммуникации, но и для сопровождения, а сегодня еще и для оформления хозяйственных отношений. Закупки корпоративного и государственного сегментов могут служить наглядным примером стремительного расширения области применения информационных технологий.

Однако сегодня диджитализация выходит на новый уровень: программно-аппаратный комплекс решает задачи и осуществляет действия, которые традиционно считались прерогативой человека. Это и создание объектов авторского права, и фактическое заключение сделок.

Такое проникновение искусственного интеллекта в хозяйственные, а вслед за тем и в правовые отношения с неизбежностью ставит перед нами вопросы юридического характера. Например, кто является автором текста, созданного искусственным интеллектом? Или кто отвечает за последствия сделки, правовое бытие которой порождено последствиями реализации алгоритмов программно-аппаратного комплекса?

Эти вопросы приобретают особенную актуальность в силу того, что и диджитализация в целом, и использование искусственного интеллекта в частности отнюдь не являются сиюминутной данью моде. Рациональное и последовательное внедрение этих технологий действительно способно повысить эффективность как процессов производства, так и бизнес-процессов в целом.

В связи с этим можно ожидать лишь расширения и интенсификации применения цифровых технологий. А значит, и юридическая наука не сможет уклониться от ответов на возникающие вопросы. «При этом юристам не обязательно постигать математические и технические тайны цифровизации; цифровизация не входит в предмет юридической науки. Об этом приходится писать, поскольку многие из тех, кто посвятил свои исследования цифровизации, игнорируют то обстоятельство, что науки делятся на технические и социальные, а юридические науки – науки социальные, и технические нормы не являются предметом их анализа» (Лазарев, 2023).

Разумеется, электронные технологии, как и любые технологии, имеют ограниченные области эффективного применения. Но, например, конкурентные закупки B2B (business-to-business) и B2G (business-to-government) являются той сферой, в которой эти технологии дают качественное повышение эффективности работы: так, с помощью цифровых технологий уже сегодня за считанные часы выполняются те бизнес-процессы, на которые в традиционной парадигме отводились недели.

Применение программно-аппаратных комплексов вовсе не ограничивается конкурентными закупками. Однако эта сфера позволяет представить наглядные иллюстрации как той поэтапной эволюции внедрения цифровых технологий, которая логически привела к использованию ИИ, так и сопутствующей правовой проблематики. Не претендуя на всеобъемлющее исследование применения технологий автоматизации в хозяйственной деятельности, в рамках данной статьи именно на примере электронных конкурентных закупок будут рассмотрены те вопросы, которые синтез бизнеса и цифровых технологий ставит перед юридической наукой.

И в сегменте B2G, и особенно в сегменте B2B за I четверть XXI века конкурентные закупки прошли путь от внедрения электронного документооборота до апробирования технологий искусственного интеллекта. Этот путь включал не только множество экспериментов, но и несколько качественных переходов, каждый из которых требует, по крайней мере, обобщенного анализа не только с организационной, но и с правовой точки зрения.

Такой анализ даст нам возможность отследить те ключевые трансформации, каждая из которых принципиальным образом меняла значение цифровых

технологий в закупочной деятельности. И та трансформация, связанная с внедрением ИИ в повседневные бизнес-процессы, которая в этой сфере происходит сегодня, не ограничивается лишь вопросами технологии, но и в значительной мере касается вопросов права.

При этом важно не забывать о том, что и предшествующие трансформации отнюдь не обходились без формулировки и внедрения новых правовых конструкций. А значит, мы можем опираться и на этот опыт для ответа на новые вопросы.

1. Пять цифровых трансформаций конкурентных закупок

Конкурентные закупки в той или иной форме существуют уже не первое столетие. И до конца XX века они проводились, за редчайшим исключением, без использования цифровых технологий. Однако в течение первой четверти XXI века использование цифровых технологий при проведении закупок не просто стало повсеместным, а повлекло стремительное изменение качества использования этих технологий.

Несмотря на стремительную диджитализацию, в целом можно констатировать ее эволюционный характер. Это, в свою очередь, дает нам возможность проследить этапы внедрения цифровых технологий в закупочное пространство. Для наглядности эту эволюцию можно представить в виде пяти трансформаций, результатом каждой из которых стала принципиально новая роль цифровых технологий в закупочных бизнес-процессах.

Обособление этих трансформаций важно еще и в связи с тем, что на практике крайне затруднителен переход на очередной уровень внедрения цифровых технологий без глубокой и последовательной реализации предыдущего уровня в повседневной закупочной работе. Иными словами, каждая предшествующая трансформация является необходимым условием для последующей трансформации.

1.1. Закупки в электронной форме

Первой цифровой трансформацией конкурентных закупок можно назвать перевод таких закупок в электронную форму.

Электронную конкурентную закупку можно в самом первом приближении определить через технологический признак как формализованную процедуру конкурентного выбора контрагента, в рамках которой все отношения между организатором и участниками закупки, в том числе юридически значимый документооборот и торги в режиме реального времени, полностью реализуются в сети Интернет без использования документов на бумажных носителях.

Разумеется, не только торги в Интернете, но даже и простой отказ от документов на бумажных носителях невозможно реализовать без «нулевого» шага в виде внедрения хозяйствующими субъектами электронного документооборота. Для того чтобы документ на электронном носителе признавался столь же юридически значимым, как и документ на бумажном носителе, потребовался механизм удостоверения юридической значимости электронного документа.

Таким механизмом стала электронная подпись, которая позволила направлять предложение о заключении сделки в виде файла и гарантировать такому файлу статус полноценной оферты в гражданско-правовом смысле. В самом начале 2002 г.

был подписан Закон «Об электронной цифровой подписи»¹, а в 2011 г. ему на смену пришел ныне действующий Закон «Об электронной подписи»².

Вопрос о видах и порядке применения электронных подписей остается за рамками этой статьи. Для рассматриваемой темы важно то, что посредством проставления электронной подписи юридическая значимость сообщается не только документам на электронном носителе, но и информации в электронной форме. В частности, такой подписью может быть удостоверен электронный образ бумажного документа или даже цифра, указанная в качестве аукционной ставки. А это уже, в свою очередь, открывает пространство для проведения полноценных торгов в электронной форме.

Первые электронные закупки стали проводиться корпоративными заказчиками в 2002 г., а к концу десятилетия электронные аукционы были введены и для государственных заказчиков³. И хотя широкий спектр инструментария электронных закупок не ограничивается лишь аукционом – так, в электронной форме с тем же успехом можно проводить и конкурсы, и запросы предложений, и запросы цен, – но сама по себе новелла об использовании электронных аукционов для государственных закупок стала важнейшей вехой официального признания электронных закупочных технологий. Отдельной дискуссии заслуживает вопрос о том, насколько методологически и практически оправдано избрание из всего спектра доступных инструментов электронных закупок одного лишь аукциона в электронной форме, который в действующем Законе «О контрактной системе» играет доминирующую роль⁴.

Сегодня электронная форма является нормой и для государственных, и для коммерческих закупок. Однако на практике в понятие «электронной закупки» разные заказчики вкладывают разное содержание. Для одних это публикация закупочной документации в сети Интернет с последующим сбором предложений на электронных носителях. Для других – проведение формализованной процедуры на специализированном портале, именуемом электронной торговой площадкой.

Различаются и хронологические рамки применения электронных технологий в закупке. Традиционно для наименования электронной закупкой достаточно проведения на электронной торговой площадке той части бизнес-процессов, которые начинаются публикацией извещения о закупке и завершаются объявлением результатов закупки.

Однако на деле комплекс закупочных бизнес-процессов куда обширнее. Он начинается задолго до объявления закупки формализованной работой с той потребностью, для удовлетворения которой эта закупка необходима, а результаты такой работы включаются в план закупок. Завершается же закупочная работа много позднее выбора победителя оформлением результатов поставки, а в развитых закупочных

¹ Об электронной цифровой подписи: Федеральный закон от 10.01.2002 № 1-ФЗ. *Собрание законодательства РФ*, 14.01.2002, № 2. Ст. 127.

² Об электронной подписи: Федеральный закон от 06.04.2011 № 63-ФЗ. *Собрание законодательства РФ*, 11.04.2011, № 15. Ст. 2036.

³ О размещении заказов на поставки товаров, выполнение работ, оказание услуг для государственных и муниципальных нужд: Федеральный закон от 21 июля 2005 года № 94-ФЗ. *Собрание законодательства РФ*, 2005, № 30. Ст. 3105.

⁴ О контрактной системе в сфере закупок товаров, работ, услуг для обеспечения государственных и муниципальных нужд: Федеральный закон от 05.04.2013 № 44-ФЗ. *Собрание законодательства РФ*, 08.04.2013, № 14. Ст. 1652.

практиках – анализом качества удовлетворения поставленной продукцией первоначально заявленной потребности.

Технически все эти бизнес-процессы уже давно могут реализовываться в электронной форме. Практика показывает, что при должной реализации их перевод в электронную форму имеет практический смысл. Однако для этого нужна уже следующая трансформация, которую условно можно назвать цифровизацией закупочной работы.

1.2. Цифровизация закупочной работы

Цифровизация закупок – это перевод в электронную форму всех формализованных закупочных бизнес-процессов от планирования закупки до приемки поставленной продукции. Такая трансформация, как правило, требует создания специализированного портала, обеспечивающего прохождение бизнес-процессов в цифровой среде.

Если цифровизация закупочной работы проведена успешно, то весь внутренний и внешний документооборот, связанный с подготовкой и проведением закупки, осуществляется в электронной форме. Бизнес-процессы реализуются по тем маршрутам, которые предлагают пользователю алгоритмы специализированного портала. Это, в свою очередь, упорядочивает закупочное взаимодействие и облегчает деятельность человека, который получает возможность сосредоточиться не на бюрократической, а на содержательной части своей работы.

Цифровизация закупочной работы почти всегда подразумевает поэтапное внедрение новых технологий в устоявшиеся бизнес-процессы. Все этапы такого внедрения можно условно объединить в два блока: создание цифровой среды, в которой будут проходить бизнес-процессы после их цифровизации, и последующее развитие цифровой системы, обеспечивающей прохождение бизнес-процессов в электронной среде (Казанцев, 2022). Очевидным образом эти блоки могут быть реализованы лишь последовательно, но внутри каждого из двух блоков отдельные этапы могут прорабатываться и параллельно.

Создание цифровой среды не ограничивается созданием специализированного портала. Этот блок требует также решения целого ряда организационных и юридико-технических вопросов, таких как:

1. Систематизация закупочных бизнес-процессов с распределением полномочий и ответственности, исключением лишних звеньев и добавлением недостающих элементов.

2. Регламентация закупочных бизнес-процессов в документах, обязательных для всех субъектов бизнес-процессов. Такие документы должны исключать как различные толкования формулировок, так и размытие ответственности, неопределенность сроков и иные юридико-технические недочеты.

3. Описание закупочных бизнес-процессов в виде алгоритмов, предназначенных для наглядной визуализации регламентирующих документов и упрощения последующей цифровизации регламентированных бизнес-процессов.

4. Создание программного обеспечения для реализации закупочных алгоритмов, а также интуитивно понятного интерфейса для использования этого программного обеспечения.

5. Пользовательское тестирование и отладка функциональности программного обеспечения для реализации закупочных алгоритмов.

6. Апробирование созданной цифровой среды на пилотном проекте для минимизации издержек последующего внедрения программного обеспечения для всех бизнес-процессов.

7. Модернизация цифровой среды по итогам пилотного проекта и ее распространение на все закупочные бизнес-процессы.

Развитие цифровой системы закупок в самом общем виде сводится к расширению области ее использования в хозяйственной деятельности компании путем постепенного перевода всех бизнес-процессов в электронную форму.

Однако не стоит забывать о том, что цифровизация закупок, при всех очевидных удобствах, в глобальном смысле означает лишь реализацию старых процессов в новой среде. По известному определению, цифровизация хаоса порождает лишь хаос в электронной форме. Иными словами, цифровизацию закупочных – да и любых других – бизнес-процессов необходимо начинать с их отладки и оптимизации. Не стоит ожидать эффективности от внедрения цифровых инструментов при сохранении избыточных согласований или, наоборот, участков непрозрачности решений.

Для того чтобы содержание закупочной работы качественно изменилось благодаря внедрению цифровых технологий, цифровизации недостаточно. Необходима автоматизация.

Понятие автоматизации необходимо строго отделять от понятия цифровизации. Цифровизация предшествует автоматизации и является ее необходимым условием, но отнюдь не тождественна ей.

1.3. Автоматизация создания документов

Если цифровизация представляет собой осуществление классических бизнес-процессов в электронной среде, то автоматизация – это привлечение к выполнению части задач в рамках таких процессов мощностей программно-аппаратного комплекса. Причем речь не идет о задачах вроде маршрутизации документов и напоминании о наступлении тех или иных сроков. Автоматизация означает использование программно-аппаратного комплекса, в том числе для создания таких документов.

В самом общем виде за человеком в таком разделении труда остается загрузка исходных данных и контроль результатов, а за программно-аппаратным комплексом – обработка данных и формирование результатов. Ответственным за результат, в том числе в юридическом смысле, остается уполномоченный сотрудник. Однако при должной реализации внедрение подобных технологий качественно сокращает объем рутинной работы человека и повышает ее экспертную составляющую.

Первый шаг такой автоматизации – это формирование базы данных, включающей конструктор электронных документов с предустановленными формами. Эти формы могут иметь вид текстовых полей, полей цифр или дат, списков, формул и т. д. С помощью интерфейса конструктора документов ответственный сотрудник вводит исходную информацию в предустановленные формы. На основании совокупности этой информации программно-аппаратным комплексом сперва подбирается релевантная форма, а затем в этой форме заполняются переменные значения.

Именно конкурентные закупки в силу самой своей организационной специфики представляют собой перспективную среду для автоматизации процессов подготовки документов. Ведь нормальная закупка проводится по формализованной, заранее регламентированной процедуре, а число вариантов как шаблонов закупочной

документации, так и шаблонов заявки на участие в закупке является конечным и, как правило, весьма ограниченным.

Это создает возможность для использования при автоматизации закупок помещенных в цифровую среду шаблонов документов, а в развитых электронных системах еще и заполнение этих шаблонов с помощью технологии наследования данных. Под наследованием данных с точки зрения бизнес-процессов в данном случае мы понимаем возможность автоматизированного предзаполнения форм документов, контрактов и т. д. По сути, заказчику при подготовке каждой закупки остается лишь указать в «шаблоне» документации предмет закупки, требования к закупаемой продукции и условия будущей поставки.

С помощью конструктора документов создается форма, необходимая для данной закупочной процедуры, а с помощью наследования данных в этой форме заполняются поля переменных.

Максимально упрощая, систему наследования данных в рамках конкурентной закупки можно представить так:

1. Параметры потребности формулируются в виде электронной заявки на проведение закупки.
2. Параметры согласованной заявки включаются в план закупок.
3. Данные строки плана закупок «подтягиваются» в типовую закупочную документацию, которая благодаря этому превращается в проект документации конкретной закупки.
4. Нормы закупочной документации транслируются в проект договора.
5. Договорные условия отражаются в закрывающих документах.

По большому счету, сегодняшняя архитектура цифровой среды государственных закупок, и прежде всего Единой информационной системы в сфере закупок, строится именно на таких подходах.

Для эффективного использования подобной электронной системы пользователю должно быть достаточно профессиональных знаний в своей области. Иными словами, автоматизация хороша лишь в тех случаях, когда она высвобождает время сотрудника, а не требует от него тратить это время на изучение порядка использования функциональности электронного портала.

Вот почему в числе прочего дизайн системы и ее интерфейса на практике является одним из ключевых факторов автоматизации работы. Несколько упрощая, можно сформулировать этот принцип так: система тем лучше, чем меньшее количество кнопок нужно нажать человеку для получения сложного результата.

Результатом наследования данных предшествующих документов должно становиться предзаполнение форм последующих документов. Например, в рамках согласованной заявки утверждены наименование и количество приобретаемой продукции – в этом случае технология наследования данных должна обеспечить автоматизированное предзаполнение переменных «наименование» и «количество» и в плане закупок, и в закупочной документации, и в проекте договора, и в проектах документов о приемке. При этом предзаполнение данных должно оставлять человеку пространство для их редактирования – поскольку, как уже было сказано, именно человек в конечном счете отвечает за результат.

Таким образом, после внедрения инструментов автоматизации создания документов исходные данные, в том числе параметры закупки и значения переменных, указываются человеком. Проекты документов создаются программно-аппаратным

комплексом. А юридическую значимость этим документам сообщает подписание, осуществляемое уполномоченным сотрудником посредством проставления электронной подписи.

При этом ни в теории, ни на практике потенциал автоматизации в закупках не ограничивается формированием закупочной документации. Уже сегодня мы видим успешные примеры автоматизации отдельных бизнес-процессов, связанных с закупками.

1.4. Автоматизация бизнес-процессов

Автоматизацию бизнес-процессов можно определить как обеспечение их прохождения с использованием алгоритмов программно-аппаратного комплекса. Как правило, это требует существенной организационной переработки этих бизнес-процессов при сохранении их сущности. После автоматизации закупок человек принимает решения только на ключевых этапах.

Возможности автоматизации позволяют реализовать закупки и по принципиально новым процедурам – качественно сокращающим организационные издержки всех сторон закупочного процесса при сохранении конкурентности и прозрачности самой закупки (Казанцев, Михалева, 2020). Уже сегодня набирают популярность так называемые динамические закупки, в рамках которых сбор предложений поставщиков и их ранжирование де-факто автоматизированы. И это пример того, как автоматизация требует организационной перестройки привычных бизнес-процессов.

Например, в рамках классической конкурентной закупки заказчик сперва публикует извещение о закупке, затем от одной до нескольких недель ожидает поступления предложений от поставщиков, потом оценивает и ранжирует эти предложения и лишь после этого оформляет итоги закупки. Автоматизация в рамках динамической закупки требует выстроить процесс иначе: сперва поставщики публикуют свои предложения на специализированном портале и подписывают их электронной подписью как оферты, адресованные неопределенному кругу потенциальных заказчиков; затем заказчик, при возникновении потребности в закупке, указывает параметры этой потребности в предназначенных для того полях интерфейса специализированного портала; после этого программно-аппаратный комплекс портала проводит автоматизированный поиск релевантных предложений среди тех, которые уже были загружены поставщиками, и сразу же ранжирует их по параметрам, заданным заказчиком.

Сегодня появляются все новые и новые разновидности этой схемы в виде электронных маркетплейсов и корпоративных интернет-магазинов. Эти разновидности порой могут иметь существенную специфику: например, предварительное предложение может не иметь статуса оферты, а его выбор означает не акцепт, а лишь приглашение к переговорам. Но даже такая специфика не отменяет базовой схемы описанной выше автоматизации.

Инструмент быстрого приобретения продукции с автоматизированным подбором потенциальных предложений на сегодняшний день регламентирован ч. 12 ст. 93 Закона «О контрактной системе»⁵ как полноценный способ определения поставщи-

⁵ О контрактной системе в сфере закупок товаров, работ, услуг для обеспечения государственных и муниципальных нужд: Федеральный закон от 05.04.2013 № 44-ФЗ. *Собрание законодательства РФ*, 08.04.2013, № 14. Ст. 1652.

ка (подрядчика, исполнителя). В качестве рекомендованного инструмента он получил свое место и в закупках государственных корпораций и субъектов естественных монополий⁶. Еще в 2014 году схожий инструмент под названием Dynamic purchasing были упомянут в Директиве ЕС «О публичных закупках»⁷, а в 2016 г. стал едва ли не основным способом выбора поставщика в Закупочном кодексе Италии⁸.

Динамическая закупка – это лишь один из вариантов автоматизации закупочных бизнес-процессов. Строго говоря, в этом случае автоматизация касается лишь выбора победителя. Это действительно позволяет качественно сократить временные издержки. Однако в конкурентных закупках может быть реализован и более глубокий вариант автоматизации.

Такой вариант автоматизации называют сквозной автоматизацией. Она базируется на описанной выше технологии наследования данных, но не ограничивается ею. В схематическом виде ее можно описать следующим образом:

1. Заказчик формирует базу данных с библиотеками типовых закупочных документов, проектов договоров и т. д.

2. База данных интегрируется в специализированный портал вместе с категориальными стратегиями заказчика.

3. При возникновении потребности в закупке уполномоченный сотрудник указывает ключевые параметры этой потребности в предназначенных для того полях интерфейса специализированного портала.

4. На основании ключевых параметров программно-аппаратным комплексом определяется категориальная стратегия.

5. В рамках данной стратегии осуществляются автоматизированный сбор и обработка информации, в том числе информация о конкурентном состоянии рынка, наличии квалифицированных поставщиков, рыночных ценах на продукцию, после чего определяется наиболее эффективный способ закупки и формируется пакет закупочной документации.

6. Уполномоченный сотрудник утверждает проекты либо вносит дополнительные данные для уточнения результатов.

7. На основании документации, утвержденной уполномоченным сотрудником, программно-аппаратный комплекс осуществляет контроль сроков и публикует либо рассылает извещение о закупке в предназначенное для того время.

8. В зависимости от выбранного способа закупки программно-аппаратный комплекс подбирает предложения, заранее опубликованные поставщиками для участия в динамической закупке, либо направляет квалифицированным поставщикам приглашения прислать свои предложения.

9. После сбора предложений программно-аппаратный комплекс осуществляет проверку сведений о поставщиках в открытых источниках, оценивает соответствие

⁶ Об особенностях участия субъектов малого и среднего предпринимательства в закупках товаров, работ, услуг отдельными видами юридических лиц: Постановление Правительства РФ от 11.12.2014 № 1352. *Собрание законодательства РФ*, 22.12.2014, № 51. Ст. 7438.

⁷ Directive 2014/24/EU of the European Parliament and of the Council of 26 February 2014 on public procurement. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32014L0024>

⁸ *Codice dei contratti pubblici* (Decreto legislativo 18 aprile 2016, n. 50). https://www.bosettiegatti.eu/info/norme/statali/2016_0050.htm

каждого предложения потребности заказчика и ранжирует предложения на основании критериев, указанных заказчиком.

10. Заказчик утверждает ранжирование и объявляет победителя, указывает дополнительные данные для проведения новой закупки.

11. В случае объявления победителя программно-аппаратный комплекс составляет проект договора и направляет его сперва на утверждение заказчику, а затем на подписание поставщику.

12. На основании параметров подписанного договора программно-аппаратный комплекс осуществляет контроль сроков и объемов поставки, в том числе рассылает напоминания об отгрузке и приемке.

13. По итогам поставки на основании сведений договора и загруженной заказчиком информации о фактическом исполнении договора программно-аппаратный комплекс формирует проекты закрывающих документов и направляет их на подписание заказчику и поставщику (Казанцев, 2020).

Эта схема охватывает все основные этапы закупочной работы и допускает использование всех основных способов закупки. Важно подчеркнуть: человек при этом не устраняется от закупочной работы, и именно его электронной подписью сообщается юридическая значимость ключевым решениям и документам. Но вот подготовка этих решений и документов, их маршрутизация, коммуникации с контрагентом и «бюрократические» элементы работы – все это реализуется силами программно-аппаратного комплекса.

В такой парадигме не только сокращаются трудозатраты уполномоченного сотрудника при одновременном увеличении скорости прохождения бизнес-процессов. Реализация этих процессов силами программно-аппаратного комплекса снижает значение так называемого человеческого фактора и субъективизма в принятии решений. Ведь тот факт, что человек может вносить правки в документы, подготовленные программно-аппаратным комплексом, не отменяет того, что каждая из этих правок фиксируется в истории изменений, а это в свою очередь облегчает ретроспективный контроль.

Сквозная автоматизация создает условия для формирования системы S2P. Source-to-pay – это прикладная автоматизация всего комплекса закупочной работы, включая все взаимодействия между заказчиком и поставщиком⁹. В этой системе все бизнес-процессы, связанные с закупкой, проходят в цифровой среде и реализуются в том числе с использованием искусственного интеллекта.

Искусственный интеллект при должной интеграции в систему S2P способен существенно расширить потенциал автоматизации и индивидуализировать закупочные сценарии в зависимости от уникальных параметров сочетания потребности заказчика с рынком приобретаемой продукции. В дополнение к описанной выше базовой схеме автоматизации искусственный интеллект на основании анализа больших данных – как из открытых источников, так и из собственных архивов ранее проведенных закупок – способен выполнять по меньшей мере следующие функции:

– Прогнозировать потребности. Для этого ИИ анализирует деятельность заказчика с точки зрения необходимых для нее ресурсов, отслеживает объем и содержание складских запасов, а также выполнение рамочных договоров. Если возникает риск

⁹ McCann, Jo. *How source-to-pay works*. <https://blog.routable.com/how-source-to-pay-works/>

того, что тот или иной ресурс близок к истощению и не может быть получен без проведения закупки, то ИИ сообщает об этом заказчику, а в идеале еще и предлагает оптимальные параметры будущей закупки. Например, учитывается фактический срок выполнения заказа поставщиками. Ведь чем больше времени требуется поставщику, чтобы доставить товары, тем больше запасов вынужден держать производитель¹⁰.

– Определить наиболее эффективный инструмент закупки. По большому счету эта функция является продолжением функции определения базовых параметров. На основании анализа информации о состоянии рынка закупаемой продукции в регионе ее закупки ИИ может оценить, что будет эффективнее – объявить аукцион, провести динамическую закупку или сразу предложить кандидатуру единственного поставщика. Ведь, например, попытка проведения аукциона, заранее обреченного на провал в силу отсутствия конкуренции, будет означать лишь дополнительные организационные и временные издержки заказчика без какого бы то ни было результата.

– Провести расширенную проверку квалификации и хозяйственного поведения поставщика. ИИ способен не просто собрать данные из официальных реестров, но и получить отзывы других клиентов, информацию о судебных разбирательствах, а иногда даже сведения о мощностях, оборудовании, технологиях, квалификации сотрудников и т. д. Все эти данные важны для заключения договора. Вот только самому заказчику без использования ИИ потребуется несравненно больше времени для того, чтобы собрать хотя бы часть этих сведений.

– Определить и контролировать логистику доставки. ИИ не просто высчитывает оптимальный маршрут, его стоимость и риски, но и позволяет заказчику при минимальной технологической интеграции с поставщиком и его навигационными системами отслеживать движение груза в режиме реального времени.

Реализация этих дополнительных функций требует от программно-аппаратного комплекса собирать и обрабатывать большие данные, а также иметь механизмы машинного обучения. Эти признаки позволяют нам говорить, если и не в узкоспециализированном, то, по крайней мере, в широком смысле об использовании для автоматизации закупок искусственного интеллекта.

1.5. Автоматизация решений

На сегодня нельзя говорить о консенсусе даже по такому, казалось бы, отправному для данной дискуссии вопросу, как понятие искусственного интеллекта.

Например, предлагается понимать его как систему, обладающую способностью физически проявлять себя, в том числе чувствовать, обрабатывать и воздействовать на окружающий мир в некоторой степени (Calo, 2015). В такой дефиниции на первый план выходит физическое проявление последствий работы искусственного интеллекта. Но на практике результаты его работы могут оставаться в электронной среде, и лишь воздействие этих результатов на правоотношения с участием физических и юридических лиц проявит себя в окружающем мире.

¹⁰ Banker, Steve. (2022, April 1). One Multinational's Supply Chain Transformation Journey. *Forbes*. <https://www.forbes.com/sites/stevebanker/2022/04/01/one-multinationals-supply-chain-transformation-journey/?sh=87ff9516229c>

В крайнем выражении такой подход воплощен в концепции так называемого сильного искусственного интеллекта, под которым исследователи понимают такую технологию, которая ментальными свойствами и характером обработки доступной ей информации идентична человеческому сознанию, в том числе в части комплексной интерпретации информации, способности к творчеству и интуиции (Searle, 1990).

Однако на сегодня наиболее реалистичной следует признать концепцию искусственного интеллекта как программно-аппаратного комплекса, не имеющего с человеческим разумом ничего общего в плане сущности мышления, но способного при этом решать в совокупности аналогичные по сложности либо более сложные задачи (Bokovnya et al., 2020). Представляется, что именно такой подход отвечает сегодня как достигнутому уровню технологий, так и областям практического применения искусственного интеллекта.

В развернутом варианте, предложенном Н. Н. Черногором, определение искусственного интеллекта в рамках такого подхода звучит следующим образом: «Технология, определяющая способность некой информационной системы без непосредственного участия человека правильно интерпретировать внешние данные (внешнюю информацию), уточнять базу (базы) данных с их учетом, извлекать уроки из допущенных ошибок и использовать полученные знания для достижения конкретных целей, решения конкретных задач посредством гибкой адаптации в условиях плохо определенной ситуации» (Черногор, 2022).

Уже сегодня возможности искусственного интеллекта в большей или меньшей степени используются заказчиками для управления складскими объемами и для работы с логистикой поставки. Сочетание объема обрабатываемых данных со скоростью обработки делает привлечение ИИ объективно оправданным. Однако это и ставит перед нами новые вопросы правового характера.

На примере предыдущих четырех трансформаций было показано, что развитие регулирования не предвляло внедрение электронных технологий в закупочную деятельность, но всегда сопровождало его. Сперва это были новеллы общего регулирования, посвященные электронному документообороту и электронной подписи, затем новеллы специального законодательства, регулирующие электронные способы закупки, затем новые подзаконные акты, посвященные быстрым покупкам на базе схем автоматизации. Государственное регулирование дополнялось и развивалось корпоративным регулированием – в вопросах автоматизации сегодня основной объем норм, равно как и глубина регулирования, приходится не на законы и даже не на ведомственные инструкции, а именно на корпоративные положения и регламенты.

Об этом стоит помнить, приступая к вопросу о регулировании применения искусственного интеллекта в закупочной работе. Ведь ИИ способен не только анализировать складские запасы и маршруты доставки. Следующий шаг – это выбор оптимального способа закупки на основании анализа данных о рыночных условиях. Для этого важна не просто обработка больших данных, но и наличие у заказчика многофакторных категорийных стратегий управления закупками. Ведь для определения инструментария большие данные о рынке должны обрабатываться в матрице категорийных стратегий: ИИ собирает и обрабатывает данные о том, продается ли необходимая заказчику продукция в данном регионе и доставляется ли она из соседних регионов, какова цена с учетом доставки, отвечают ли характеристики потребностям заказчика, есть ли конкурирующие предложения и готовы ли поставщики к конкуренции и т. д.

Но если ИИ способен обрабатывать эти данные, то он может предложить заказчику не только оптимальный способ закупки, но и кандидатуру победителя. А если экстраполировать эту возможность на потребность в оперативной закупке для непрерывного производства, то тот же ИИ на основании тех же данных может сам направить заявку на поставку лучшему поставщику.

И даже если ИИ направляет заявку не новому поставщику, а тому, с которым ранее заказчик заключил рамочное соглашение на поставку – а такие модели интеграции искусственного интеллекта в закупки уже сегодня используют некоторые предприятия, – то и в этом случае именно ИИ формирует те существенные условия договора, которые порождают сделку. Он указывает конкретный объем поставки конкретной продукции, из чего на условиях рамочного договора автоматически следует определение цены поставки. Технически уже сегодня искусственный интеллект может делать то же самое и в отсутствие рамочного договора, подписанного ранее заказчиком и поставщиком.

Автоматизация принятия решения о выборе поставщика и о заключении сделки с ним – это намного быстрее и удобнее классического тендера. Это полностью устраняет пристрастность и субъективизм принятия решения о выборе победителя закупки. Но это же превращает ИИ из инструмента принятия решения в субъекта принятия решения.

2. Правосубъектность искусственного интеллекта

Принципиальное изменение роли программно-аппаратного комплекса в хозяйственных отношениях, связанных с закупками, ставит перед нами принципиально новые вопросы юридического порядка, которые уже не снимаются действующим регулированием. Уже урегулирован электронный документооборот. Урегулированы и электронные закупки, которые проходят с использованием электронного документооборота. До известной степени урегулирована и цифровизация всего процесса закупочной работы, для которого сама по себе электронная закупка является лишь одним из этапов. Но не урегулирован статус искусственного интеллекта в рамках цифровой закупочной работы.

И сегодня это уже вопрос не просто теоретического порядка: на практике последствия работы искусственного интеллекта приводят в том числе к заключению сделок и возникновению договорных обязательств. Более того, можно утверждать, что именно возникновение договорных обязательств по результатам работы искусственного интеллекта и означает подлинную автоматизацию закупочной работы.

Так, например, решение о выборе способа закупки напрямую касается лишь заказчика (пусть даже его последствия косвенным образом могут сказаться на поставщике). И в этом случае программно-аппаратный комплекс, предлагающий заказчику на основании заданного алгоритма наиболее эффективный инструмент закупок, выполняет сложную, но все еще вспомогательную функцию. Субъектом решения, отвечающим на его последствия, остается уполномоченный сотрудник заказчика. А значит, о полной автоматизации данного бизнес-процесса говорить еще не приходится.

Но уже решение об отправке поставщику заявки на отгрузку очередной партии товара ставит программно-аппаратный комплекс в рамках хозяйственных отношений в ту роль, при которой уже фактически можно говорить о том, что робот заключает сделку.

При этом, что важно, программно-аппаратный комплекс не «присваивает» себе полномочий за рамками заданного ему алгоритма. Его роль субъекта сделки следует именно из выполнения алгоритма – ведь и заказчику, и поставщику удобно такое привлечение робота к определению сроков, объема и предмета поставки. Это позволяет и заказчику, и поставщику высвободить ресурсы для основной производственной деятельности. Но с гражданско-правовой точки зрения в этой ситуации именно результаты работы искусственного интеллекта порождают взаимные права и обязанности заказчика и поставщика.

Это обстоятельство ставит перед нами вопрос о правосубъектности самого искусственного интеллекта, ответ на который сегодня нельзя назвать очевидным. Результаты научной дискуссии по этому вопросу зависят не только от возможного консенсуса, но и от скорости развития электронных технологий.

Вопрос о правосубъектности искусственного интеллекта решается на базисе понимания его сущности, однако не ограничивается ею. Так, даже если не пытаться отыскать сходства между мыслительной деятельностью человека и работой искусственного интеллекта, то можно найти весомые аргументы в пользу наделения программно-аппаратного комплекса правосубъектностью, во-первых, для ясного обозначения его роли в правовых отношениях, а во-вторых, для защиты прав иных субъектов.

С предельной ясностью эту мысль выражает Е.В. Вавилин: «...правосубъектность ИИ необходима для того, чтобы ИИ был ограничен в своих правах конкретным функциональным предназначением» (Вавилин, 2021). Однако и он не говорит о том, что правосубъектность искусственного интеллекта может быть сходна с правосубъектностью, например, юридического лица. Конструкция правосубъектности ИИ предлагается именно как способ достижения цели ограничения его прав. С этой точки зрения достаточно наделить ИИ конкретным набором прав и обязанностей, т. е. предусмотреть специальную техническую правосубъектность.

Однако в подобном дискурсе остается неразрешенным вопрос о практической реализации такой правосубъектности, и в частности о практических аспектах возложения ответственности. Каким образом, например, можно заставить ИИ возместить вред, причиненный последствиями его работы? Ведь какие бы юридические фикции мы ни использовали, но де-факто работа ИИ в пользу одного субъекта (например, в пользу лица, которому искусственный интеллект причинил ущерб) означает то, что он не использует часть времени или мощностей для работы в пользу другого лица (а именно своего владельца). Таким образом, фактически возмещение вреда в этой ситуации возлагается на владельца ИИ, пусть даже и осуществляет он это возмещение с использованием того же ИИ.

В этих условиях сложно не согласиться с выводами В. К. Андреева о том, что «применение цифровых технологий с использованием искусственного интеллекта на современном уровне его развития не означает появления новых общественных отношений, качественно отличающихся от существующих», а «искусственный интеллект не выступает в качестве цифрового субъекта права в отношениях по обороту цифровых прав в информационной системе оператора. Последний, используя в предпринимательстве цифровые технологии, применяет в бизнес-моделях элементы искусственного интеллекта, которые не порождают цифровых правоотношений» (Андреев, 2021).

И в самом деле, именно владелец, или оператор, искусственного интеллекта выступает тем субъектом, который в итоге влияет на результаты работы ИИ

и пользуется ими. А значит, логично предположить и его ответственность за эти результаты.

И в самом деле, при любом уровне проникновения цифровых технологий и при любой модели использования искусственного интеллекта сегодня сложно представить ситуацию, в которой ИИ действует совершенно независимо от человека и принимает на себя права и обязанности, не проистекающие из воли и действий последнего ни прямо, ни косвенно. А значит, искусственный интеллект сегодня стоит отнести не столько к новым правовым субъектам, сколько к инновационным инструментам, требующим нового правового регулирования.

«Отношения с использованием искусственного интеллекта – это всегда отношения между субъектами права или по поводу объектов права. В любом случае это отношения, которые на том или ином этапе инициированы, запрограммированы человеком – субъектом права с той или иной степенью ответственности (в том числе в рамках деятельности юридических лиц). Волеизъявление человека на те или иные действия искусственного интеллекта может быть выражено в разной степени: от действий ИИ, находящихся под полным контролем воли человека, до автономных действий ИИ, опять же допускаемых и осознаваемых в своих возможных пределах и последствиях человеком (группой лиц)» (Шахназаров, 2022).

Степень влияния волеизъявления человека на результаты работы искусственного интеллекта должна приниматься во внимание при решении вопроса о распределении ответственности за действия ИИ между существующими правовыми субъектами, о чем подробнее будет сказано ниже. Но в вопросе о правосубъектности самого искусственного интеллекта на нынешнем уровне развития технологий, а также общественных и хозяйственных отношений можно опираться на позицию С. Е. Чаннова: «Придание роботам (СИИ) статуса субъекта права не повлечет за собой в обозримом будущем каких-то явных негативных последствий. В то же время не видны и преимущества такого решения по сравнению с рассмотрением роботов (СИИ) в качестве квазисубъектов права. Исходя из философского принципа Оккама не умножать сущности без крайней на то необходимости, мы полагаем, что введение в правовую сферу такого принципиально нового субъекта права, как робот (СИИ), является преждевременным (хотя не исключено, что такая необходимость появится)» (Чаннов, 2022).

3. Правовой статус сделки, совершенной с участием искусственного интеллекта

Вывод о том, что искусственный интеллект на сегодня не может быть признан субъектом права, не снимает, а лишь актуализирует вопросы, связанные со статусом сделки, совершенной с участием искусственного интеллекта. Ведь такие сделки заключаются уже сегодня, а со временем их число будет только расти.

Каков статус подобной сделки? Каков механизм ее судебной защиты? По каким основаниям она может быть оспорима? На эти и другие вопросы теоретикам права еще только предстоит ответить. И именно эти ответы станут основой того самого «права электронных отношений» (Kenney & Zysman, 2016), о котором пишут уже сегодня. Представляется, что будущее правового регулирования именно за формированием права электронных отношений как обособленной и целостной отрасли правового знания, а не за приданием искусственному интеллекту статуса правового субъекта.

Но если наделение искусственного интеллекта правосубъектностью пока преждевременно, то необходимо решить вопрос о том, кто является субъектом правоотношения, реализуемого с участием ИИ. Ведь использование ИИ не должно само по себе приводить к оспоримости, а тем более к ничтожности сделки. Ведь если допустить презумпцию порока сделки по основанию использования технологии ИИ при ее совершении, то это с неизбежностью создаст неоправданные риски для существующих хозяйственных отношений, о цифровизации и автоматизации которых подробно говорилось выше. А коль скоро за любым ИИ стоит физическое или юридическое лицо, то между ними и стоит распределять как правовые последствия, так и ответственность за работу искусственного интеллекта.

Точные параметры распределения ответственности как раз и предстоит выработать юридической науке в рамках формирования права электронных отношений. В качестве приглашения к дискуссии можно предложить несколько вариантов решения вопроса об ответственности за последствия работы программно-аппаратного комплекса в рамках хозяйственных отношений с высокой степенью автоматизации:

1. Ответственность владельца: каким был ни был результат, ответственным за действия и решения ИИ является то юридическое или физическое лицо, которому данный программно-аппаратный комплекс принадлежит или которым он используется на законном праве.

По блестящему определению С. Ф. Афанасьева, при таком подходе «правовое положение ИИ становится тождественным или близким тому, которое было у римских антропоморфных коллективных организаций, либо еще более редуцированным – раба, домочадцев, детей, в том числе *filius in potestate tua est*» (Афанасьев, 2022).

Такой подход интуитивно понятен и, на первый взгляд, максимально логичен и утилитарен. Однако он игнорирует тот факт, что результаты работы искусственного интеллекта далеко не всегда зависят от последствий волеизъявления, действий или бездействий его владельца.

Например, неблагоприятные последствия работы ИИ могут стать не только следствием порока его использования, но и следствием порока разработки. И сомнительной представляется правовая конструкция, возлагающая на владельца ответственность за работу программного кода, для анализа которого необходима специальная и глубокая экспертиза – та экспертиза, которой добросовестный пользователь программно-аппаратного комплекса, как правило, не может, да и не обязан обладать.

2. Ответственность создателя: если добросовестный владелец не предвидел неблагоприятные последствия действий робота и не мог повлиять на них, то обязательства из причинения ущерба логично возложить не на него, а на разработчика того алгоритма, реализация которого и стала причиной ущерба.

Необходимо подчеркнуть: речь не идет об устранении владельца ИИ от ответственности за последствия работы последнего. Речь идет лишь о распределении такой ответственности по результатам доказывания вины. Так, презюмируется ответственность владельца, но если его вину доказать не удалось, то инициируется вопрос о выяснении вины разработчика.

Однако этот подход может иметь свое логичное продолжение и в том случае, когда не доказана вина создателя. Если же на результаты работы ИИ повлияло иное

лицо, например путем его модификации или предоставления для обработки некорректных данных, то ответственным будет именно это лицо.

3. Комплексный подход: за неблагоприятные последствия работы ИИ отвечает тот субъект, чье действие или бездействие в конечном счете стало причиной причинения ущерба. Ведь на практике причиной неблагоприятных последствий работы ИИ в рамках закупочных бизнес-процессов могут стать и действия разработчика, и действия заказчика как владельца программно-аппаратного комплекса, и даже действия поставщика как его пользователя.

В зарубежной доктрине рассматривается даже вопрос о реализации концепции уголовной ответственности за последствия работы искусственного интеллекта, которая принимала бы во внимание действия и бездействия создателя, владельца, пользователей, а также иных лиц, имеющих отношение к ИИ или повлиявших на результаты ее работы (Hallevy, 2013).

Ведь искусственный интеллект принимает решения на основании сбора и обработки информации. В том числе информации из открытых источников. А значит, если в одном из таких источников содержится критический объем некорректной информации, то именно он может стать причиной ошибки искусственного интеллекта и соответствующих неблагоприятных последствий.

Для распределения ответственности в этом случае можно законодательно установить, например, презумпцию того, что субъектом обязательств, возникших в результате нормальной работы программно-аппаратного комплекса, считается его владелец. И именно он обязан будет доказать необходимость возложения ответственности на другого субъекта. К этой мысли нас подводит, в частности, В. А. Лаптев, рассуждая о субсидиарной ответственности создателя, владельца и пользователя искусственного интеллекта (Лаптев, 2019).

Этот спектр решений не является закрытым. Можно сконструировать иные модели правовых последствий для отношений, возникших в результате работы искусственного интеллекта.

Например, еще одной правовой конструкцией может стать включение ИИ в круг источников повышенной опасности. По мнению А. А. Антонова, «фиксация в законе новых систем ИИ потребует рассмотрения вопроса о признании их источником повышенной опасности, так как взыскать с владельца ИИ возмещение ущерба за причиненный вред можно лишь через суд» (Антонов, 2020).

Альтернативный подход, обсуждаемый в том числе в зарубежной доктрине, – это наделение искусственного интеллекта правосубъектностью, схожей с правосубъектностью юридического лица – со схожим же распределением ответственности за последствия его работы (Chesterman, 2020).

В любом случае регулирование вопроса правосубъектности в отношениях, реализуемых с использованием ИИ, равно как и вопрос распределения ответственности в этих отношениях, требует специального нормативного регулирования. Это не означает незамедлительного внесения изменения в законодательство: регуляторной апробацией могут стать соглашения, включающие нормы о статусе ИИ, между лицами, участвующими в таких правоотношениях. В конце концов, сами технологии цифровизации и автоматизации закупок, избранные в качестве иллюстративного материала к данной статье, создавались и апробировались не в государственном, а коммерческом сегменте.

Выводы

Информационное право не просто формируется на наших глазах, но и может в перспективе стать самостоятельной отраслью права. Той отраслью, на которой базируется значимый сегмент хозяйственных отношений. И, что еще важнее, без формулировки специализированных норм информационного законодательства в этих отношениях с каждым годом все сложнее будет говорить о правовом регулировании. Иными словами, огромные ресурсы будут задействованы в хозяйственной деятельности в отсутствие возможности точного определения взаимных прав и обязанностей.

Опыт цифровизации конкурентных закупок B2B и B2G может помочь нам проследить эволюцию проникновения цифровых технологий в хозяйственную деятельность и практику внедрения новых правовых механизмов для решения возникающих при этом вопросов. И этот опыт свидетельствует о том, что возникновение принципиально новых технологий никогда не означает принципиальной невозможности их регулирования. Разумеется, это регулирование должно отражать, а не подменять собой технологическое развитие.

При этом даже в такой специфической сфере, как конкурентные закупки, повсеместное проникновение цифровых технологий не означает равномерности такого внедрения. Проще говоря, в то самое время, когда одни заказчики лишь проходят первую цифровую трансформацию и все еще не готовы отказаться от договора на бумажном носителе с собственноручной подписью, другие заказчики на постоянной основе пользуются автоматизацией закупочных бизнес-процессов и прорабатывают внедрение технологий искусственного интеллекта.

Такая неравномерность не зависит от отраслевой специфики: первая цифровая трансформация может стать «камнем преткновения» для крупных организаций в чрезвычайно наукоемких отраслях, но при этом пятую цифровую трансформацию могут успешно апробировать самые заурядные сервисы. Это обстоятельство возвращает нас к мысли о том, что закупочное регулирование в вопросах диджитализации не является производной от отраслевого регулирования, а представляет собой самостоятельный и во многом обособленный круг вопросов, обладающий предметным единством. Такую специфику чрезвычайно важно учитывать при развитии нормативного регулирования хозяйственных отношений с использованием искусственного интеллекта.

Цифровая трансформация не ограничивается лишь использованием нового программного обеспечения. Она включает и перестроение бизнес-процессов, и модернизацию правового регулирования. Например, «цифровизация государственных закупок – это не просто вопрос приобретения самых передовых технологий. Это также требует изменения инструментов и способов закупок, которые позволили бы государству осуществлять взаимодействие с новыми технологиями, а также эффективно и быстро интегрировать их в практическую действительность» (Шмелева, 2019).

Логичным продолжением цифровой трансформации закупочной работы является ее автоматизация, в том числе с использованием технологии искусственного интеллекта. Использование искусственного интеллекта не означает устранение человека от закупочной работы. Оно означает лишь повышение экспертного уровня этой работы. Проще говоря, по мере внедрения искусственного интеллекта внимание специалиста по закупкам будет все более и более концентрироваться на нетиповых, специфических и особо важных ситуациях (Казанцев, 2021).

При этом сам искусственный интеллект на сегодня остается именно инструментом, а не субъектом закупочной работы – даже тогда, когда решения в рамках этой работы принимаются программно-аппаратным комплексом без непосредственного участия человека. «С онтологической точки зрения все передовые технологии являются не субъектами, а объектами, и нет никаких оснований предоставлять им права или привлекать их к юридической ответственности. Даже в свете существующих правил юридической ответственности на основе различных правовых критериев всегда теоретически возможно идентифицировать человека, который понесет ответственность за ущерб, возникший в результате производства или эксплуатации устройства с системой ИИ» (Ивлиев, Егорова, 2022).

В качестве приглашения к дискуссии можно предложить несколько моделей распределения ответственности в рамках хозяйственных отношений с высокой степенью автоматизации. Точнее даже будет назвать их не столько моделями, сколько точками для формулировки юридических конструкций.

Например, можно исходить из того, что субъектом такой сделки де-юре считается лицо, подписавшее тот рамочный договор, во исполнение которого был создан оформленный роботом заказ. Таким образом, ответственность за решения, «принятые» роботом в рамках исполнения алгоритма, перекладывается на человека. Но нельзя не заметить того, что по большому счету это лишь временная мера. Ведь искусственный интеллект технически уже сегодня может не просто направлять заявки в рамках подписанных ранее договоров, но и заключать новые самостоятельные договоры. Проще говоря, выступая в качестве субъекта правоотношения, робот уже не нуждается в подписанном ранее рамочном соглашении.

Другой подход – за действия робота отвечает хозяин. Иными словами, каким был ни был результат исполнения алгоритма, ответственным за действия и решения робота можно считать то юридическое или физическое лицо, которому данный программно-аппаратный комплекс принадлежит или которым он используется на законном праве. По большому счету подобный подход отработан еще римскими юристами для казусов рабовладельческих отношений.

Можно увязать ответственность за последствия работы искусственного интеллекта с виной того субъекта, действие или бездействие которого повлияло на возникновение определенного результата работы программно-аппаратного комплекса.

Каждый из подходов не совершенен и заслуживает дискуссии. Задача будущего – выработать решение, базирующееся на этих подходах, но не ограниченное рамками ни одного из них. Однако сами по себе решения в области правового регулирования означенных вопросов необходимы.

Информационное право становится все более и более востребованной отраслью юридического знания (Scassa, 2018). В перспективе оно почти наверняка станет самостоятельной отраслью права. И этой отрасли придется иметь дело с отношениями, в которых технологии и бизнес-практики имеют приоритетное значение для выработки юридических норм. Проще говоря, новые нормы должны не создавать, а гармонизировать цифровую реальность.

А значит, и новые юридические конструкции должны адекватно отражать и регулировать реально существующие отношения, реализуемые в том числе с помощью инструментов автоматизации. Именно при таком подходе представляется достижимой задача правового регулирования по минимизации риска злоупотребления полезными инструментами со стороны недобросовестных субъектов.

Список литературы

- Андреев, В. К. (2021). Приобретение и осуществление прав юридического лица с использованием искусственного интеллекта. *Предпринимательское право*, 4, 11–17.
- Антонов, А. А. (2020). Искусственный интеллект как источник повышенной опасности. *Юрист*, 7, 69–74. <https://doi.org/10.18572/1812-3929-2020-7-69-74>
- Афанасьев, С. Ф. (2022). К проблеме материальной и процессуальной правосубъектности искусственного интеллекта. *Вестник гражданского процесса*, 3, 12–31.
- Вавилин, Е. В. (2021). Трансформация гражданско-правовых и процессуальных отношений с использованием искусственного интеллекта: формирование новых правовых режимов. *Вестник гражданского процесса*, 6, 13–35. <https://doi.org/10.24031/2226-0781-2021-11-6-13-35>
- Ивлиев, Г. П., Егорова, М. А. (2022). Юридическая проблематика правового статуса искусственного интеллекта и продуктов, созданных системами искусственного интеллекта. *Журнал российского права*, 26(6), 32–46. <https://doi.org/10.12737/jrl.2022.060>
- Казанцев, Д. А. (2020). *Конкурентные закупки. Методология и нормативное регулирование*: монография. Москва: ИНФРА-М. <https://doi.org/10.12737/1068790>
- Казанцев, Д. А. (2021). Закупки и машинное обучение. Есть ли пространство для применения технологии? *Госзаказ: управление, размещение, обеспечение*, 65, 110–115.
- Казанцев, Д. А. (2022). От электронных документов до Source-to-Pay. Автоматизация закупок шаг за шагом. *Госзаказ: управление, размещение, обеспечение*, 67, 60–67.
- Казанцев, Д. А., Михалева, Н. М. (2020). Автоматизация закупок как будущее контрактной системы. *Вестник Российского университета дружбы народов. Серия: Юридические науки*, 1, 137–159. <https://doi.org/10.22363/2313-2337-2020-24-1-137-157>
- Лазарев, В. В. (2023). Юридическая наука в свете перспектив цифровизации. *Журнал российского права*, 2, 5–19.
- Лаптев, В. А. (2019). Понятие искусственного интеллекта и юридическая ответственность за его работу. *Право. Журнал Высшей школы экономики*, 2, 79–102.
- Чаннов, С. Е. (2022). Робот (система искусственного интеллекта) как субъект (квазисубъект) права. *Актуальные проблемы российского права*, 12, 94–109. <https://doi.org/10.17803/1994-1471.2022.145.12.094-109>
- Черногор, Н. Н. (2022). Искусственный интеллект и его роль в трансформации современного правопорядка. *Журнал российского права*, 26(4), 5–15. <https://doi.org/10.12737/jrl.2022.037>
- Шахназаров, Б. А. (2022). Правовое регулирование отношений с использованием искусственного интеллекта. *Актуальные проблемы российского права*, 9, 63–72. <https://doi.org/10.17803/1994-1471.2022.142.9.063-072>
- Шмелева, М. В. (2019). Цифровая трансформация системы государственных и муниципальных закупок. *Юрист*, 7, 15–22. <https://doi.org/10.18572/1812-3929-2019-7-15-22>
- Bokovnya, A. Y. et al. (2020). Legal Approaches to Artificial Intelligence Concept and Essence Definition. *Revista San Gregorio*, 41, 115–121.
- Calo, R. (2015). Robotics and the New Cyberlaw. *Californian Law Review*, 103(3).
- Chesterman, S. (2020). Artificial intelligence and the limits of legal personality. *International and Comparative Law Quarterly*, 69(4), 819–844. <https://doi.org/10.1017/s0020589320000366>
- Halleve, G. (2013). *When Robots Kill: Artificial Intelligence under Criminal Law*. University Press of New England.
- Kenney, M., & Zysman, J. (2016). The Rise of the Platform Economy. *Issues in Science and Technology*, 32(3), 61–69.
- Scassa, T. (2018). Information Law in the Platform Economy: Ownership, Control, and Reuse of Platform Data. In *Law and the "Sharing Economy": Regulating Online Market Platforms* (pp. 321–356). University of Ottawa Press.
- Searle, J. R. (1990, January). Is the Brain's Mind a Computer Program? *Scientific American*, 262(1).

Сведения об авторе



Казанцев Дмитрий Александрович – кандидат юридических наук, руководитель Департамента нормативно-правового регулирования оператора электронной торговой площадки B2B-Center

Адрес: 107113, Российская Федерация, г. Москва, 3-я Рыбинская улица, 18/22

E-mail: info@dkazantsev.ru

ORCID ID: <https://orcid.org/0000-0003-2182-5776>

РИНЦ Author ID: https://elibrary.ru/author_items.asp?authorid=1149755

Конфликт интересов

Автор заявляет об отсутствии конфликта интересов.

Финансирование

Исследование не имело спонсорской поддержки.

Тематические рубрики

Рубрика OECD: 5.05 / Law

Рубрика ASJC: 3308 / Law

Рубрика WoS: OM / Law

Рубрика ГРНТИ: 10.27.41 / Сделки

Специальность ВАК: 5.1.3 / Частно-правовые (цивилистические) науки

История статьи

Дата поступления – 6 апреля 2023 г.

Дата одобрения после рецензирования – 15 апреля 2023 г.

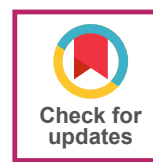
Дата принятия к опубликованию – 16 июня 2023 г.

Дата онлайн-размещения – 20 июня 2023 г.



Research article

DOI: <https://doi.org/10.21202/jdtl.2023.18>



Problems and Prospects of Regulating Relations within a Deal Effected with Participation of Artificial Intelligence

Dmitriy A. Kazantsev 

B2B-Center
Moscow, Russian Federation

Keywords

Artificial intelligence,
automation,
data,
deal,
digital technologies,
digitalization,
law,
liability,
procurement,
robot

Abstract

Objective: to research the problem of determining the subject of a legally relevant act effected with participation of artificial intelligence, as well as distribution of responsibility for the consequences of its performance.

Methods: to illustrate the problematic and practical significance of the issue of legal personality of artificial intelligence, we chose automated procurements for public and corporate needs; the methodological basis of the research is the set of methods of scientific cognition, including comparison, retrospective analysis, analogy, and synthesis.

Results: by the example of the sector of competitive procurements for public and corporate needs, the evolution of automation of economic relations up to artificial intelligence introduction was analyzed. Successfully tested reactions to the challenges of stage-by-stage introduction of digital technologies into economic relations were demonstrated, as well as the respective modifications of legal regulation. Based on the current level of technological development, the prospective questions are formulated, associated with the legal regulation of economic relations implemented with the use of artificial intelligence, first of all, the question of defining the subject of a deal effected with participation of artificial intelligence. As an invitation for discussion after analysis of jurists' conclusions about the probable variants of the legal status of artificial intelligence, the author proposes

© Kazantsev D. A., 2023



This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

variants of answers to the question of its legal personality when effecting a deal. To solve the issue of responsibility for the decisions resulting from the implementation of algorithms of a software and hardware package, we propose several models of distributing such responsibility among its creator, owner, and other persons, whose actions might influence the results of such an algorithm functioning. The proposed conclusions may be used to develop normative regulation both as a set and individually.

Scientific novelty: based on the analysis of evolution of the practices of using digital technologies in procurement, the work formulates potential legal problems, determined by the constant automation of economic relations, and proposes legal constructs to solve such problems.

Practical significance: the conclusions and proposals of this work are of prospective significance for conceptual comprehension and normative regulation of electronic procurement tools both at corporate and national level.

For citation

Kazantsev, D. A. (2023). Problems and Prospects of Regulating Relations within a Deal Effected with Participation of Artificial Intelligence. *Journal of Digital Technologies and Law*, 1(2), 438–463. <https://doi.org/10.21202/jdtl.2023.18>

References

- Afanasyev, S. F. (2022). On the problem of substantive and procedural legal personality of artificial intelligence. *Herald of Civil Procedure*, 3, 12–31. (In Russ.).
- Andreev, V. K. (2021). Acquisition and implementation of rights of a legal person using artificial intelligence. *Predprinimatelskoe pravo*, 4, 11–17. (In Russ.).
- Antonov, A. A. (2020). Artificial Intelligence as a Source of Increased Danger. *Yurist*, 7, 69–74. <https://doi.org/10.18572/1812-3929-2020-7-69-74>
- Bokovnya, A. Y. et al. (2020). Legal Approaches to Artificial Intelligence Concept and Essence Definition. *Revista San Gregorio*, 41, 115–121.
- Calo, R. (2015). Robotics and the New Cyberlaw. *Californian Law Review*, 103(3).
- Channov, S. E. (2022). Robot (Artificial Intelligence System) as a Subject (Quasi-Subject) of Law. *Actual Problems of Russian Law*, 17(12), 94–109. (In Russ.). <https://doi.org/10.17803/1994-1471.2022.145.12.094-109>
- Chernogor, N. N. (2022). Artificial intelligence and its role in the transformation of modern law and order. *Journal of Russian Law*, 4, 5–15. (In Russ.). <https://doi.org/10.12737/jrl.2022.037>
- Chesterman, S. (2020). Artificial intelligence and the limits of legal personality. *International and Comparative Law Quarterly*, 69(4), 819–844. <https://doi.org/10.1017/s0020589320000366>
- Halleve, G. (2013). *When Robots Kill: Artificial Intelligence under Criminal Law*. University Press of New England.
- Ivliev, G., & Egorova, M. (2023). Legal Issues of the Legal Status of Artificial Intelligence and Products Created by Artificial Intelligence Systems. *Journal of Russian Law*, 26(6), 32–46. (In Russ.). <https://doi.org/10.12737/jrl.2022.060>
- Kazancev, D. (2020). *A competitive procurement. Methodology and regulation*. Moscow: INFRA-M. (In Russ.). <https://doi.org/10.12737/1068790>
- Kazantsev, D. A. (2021). Procurement and machine learning. Is there space for technology application? *Goszakaz: upravlenie, razmeshchenie, obespechenie*, 65, 110–115. (In Russ.).
- Kazantsev, D. A. (2022). From electronic documents to Source-to-Pay. Procurement automation step by step. *Goszakaz: upravlenie, razmeshchenie, obespechenie*, 67, 60–67. (In Russ.).
- Kazantsev, D. A., & Mikhaleva, N. A. (2020). Procurement automation as the future of the contract system. *RUDN Journal of Law*, 1, 137–159. (In Russ.). <https://doi.org/10.22363/2313-2337-2020-24-1-137-157>

- Kenney, M., & Zysman, J. (2016). The Rise of the Platform Economy. *Issues in Science and Technology*, 32(3), 61–69.
- Laptev, V. A. (2019). Artificial Intelligence and Liability for its Work. *Pravo. Zhurnal Vysshey Shkoly Ekonomiki*, 2, 79–102. (In Russ.).
- Lazarev, V. V. (2023). Legal Science in the Light of the Prospects of Digitalization. *Journal of Russian Law*, 2, 5–19. (In Russ.).
- Scassa, T. (2018). Information Law in the Platform Economy: Ownership, Control, and Reuse of Platform Data. In *Law and the "Sharing Economy": Regulating Online Market Platforms* (pp. 321–356). University of Ottawa Press.
- Searle, J. R. (1990, January). Is the Brain's Mind a Computer Program? *Scientific American*, 262(1).
- Shakhnazarov, B. A. (2022). Legal Regulation of Relations Using Artificial Intelligence. *Actual Problems of Russian Law*, 9, 63–72. (In Russ.). <https://doi.org/10.17803/1994-1471.2022.142.9.063-072>
- Shmeleva, M. V. (2019). Digital Transformation of the System of Public and Municipal Procurements, *Jurist*, 7, 15–22. (In Russ.). <https://doi.org/10.18572/1812-3929-2019-7-15-22>
- Vavilin, E. V. (2021). Transformation of civil legal and procedural relations with the use of artificial intelligence: the formation of new legal regimes. *Herald of Civil Procedure*, 6, 13–35. (In Russ.). <https://doi.org/10.24031/2226-0781-2021-11-6-13-35>

Author information



Dmitriy A. Kazantsev – Candidate of Sciences in Jurisprudence, Head of the Department of normative-legal regulation of the B2B-Center electronic trading platform operator

Address: 18/22 3rd Rybiskaya Str., 107113 Moscow, Russian Federation

E-mail: info@dkazantsev.ru

ORCID ID: <https://orcid.org/0000-0003-2182-5776>

RSCI Author ID: https://elibrary.ru/author_items.asp?authorid=1149755

Conflict of interests

The author declares no conflict of interest.

Financial disclosure

The research had no sponsorship.

Thematic rubrics

OECD: 5.05 / Law

PASJC: 3308 / Law

WoS: OM / Law

Article history

Date of receipt – April 6, 2023

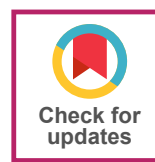
Date of approval – April 15, 2023

Date of acceptance – June 16, 2023

Date of online placement – June 20, 2023



Научная статья
УДК 343.3/.7:341.4:004.8
EDN: <https://elibrary.ru/mgreoy>
DOI: <https://doi.org/10.21202/jdtl.2023.19>



Автономное вооружение на основе искусственного интеллекта и индивидуальная уголовная ответственность согласно Римскому статуту

Фарид Мохд Хасан

Исламский научный университет Малайзии
Нилаи, Негри-Сембилан, Малайзия

Нур Джухаида Осман

Исламский научный университет Малайзии
Нилаи, Негри-Сембилан, Малайзия

Ключевые слова

Автономное вооружение,
война,
вооруженный конфликт,
искусственный интеллект,
Международный уголовный суд,
право,
Римский статут,
робототехника,
уголовная ответственность,
цифровые технологии

Аннотация

Цель: международное право обязывает государства преследовать лиц, нарушивших закон в ходе вооруженных конфликтов, чему способствовало создание Международного уголовного суда. Цель данной статьи – рассмотрение ответственности за преступления, совершенные с использованием автономных устройств на основе искусственного интеллекта, согласно положениям Римского статута Международного уголовного суда.

Методы: доктринальный анализ позволил изучить позиции экспертов по вопросу ответственности за преступления, совершенные с использованием автономных устройств на основе искусственного интеллекта, согласно положениям Римского статута Международного уголовного суда.

Результаты: в работе показано, что Международный уголовный суд может отправлять правосудие только в отношении физических лиц, предположительно совершивших преступления в рамках его юрисдикции, но не в отношении автономных вооружений. В статье утверждается, что лица, способствовавшие совершению предполагаемых преступлений, будут, вероятно, нести уголовную ответственность за предоставление средств для совершения предполагаемых преступлений автономными вооружениями на основе искусственного интеллекта согласно статье

Контактное лицо

© Хасан Ф. М., Осман Н. Д., 2023



Статья находится в открытом доступе и распространяется в соответствии с лицензией Creative Commons «Attribution» («Атрибуция») 4.0 Всемирная (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0/deed.ru>), позволяющей неограниченно использовать, распространять и воспроизводить материал при условии, что оригинальная работа упомянута с соблюдением правил цитирования.

25(3)(с) Римского статута. Авторы приходят к выводу, что Римский статут дает решение относительно автономного вооружения на основе искусственного интеллекта.

Научная новизна: в статье изучены актуальные вопросы, связанные с ответственностью за преступления, совершенные с использованием автономных устройств на основе искусственного интеллекта, согласно положениям Римского статута Международного уголовного суда.

Практическая значимость: результаты работы могут быть использованы при разработке регулирования автономного вооружения на основе искусственного интеллекта, а также служить основой для будущих исследований в сфере ответственности за использование как автономных вооружений на основе искусственного интеллекта, так и искусственного интеллекта в целом.

Для цитирования

Хасан, Ф. М., Осман, Н. Д. (2023). Автономное вооружение на основе искусственного интеллекта и индивидуальная уголовная ответственность согласно Римскому статуту. *Journal of Digital Technologies and Law*, 1(2), 464–480. <https://doi.org/10.21202/jdtl.2023.19>

Содержание

Введение

1. Автономные вооружения на основе искусственного интеллекта
2. Автономное оружие на международном уровне
3. Международное право в области автономного вооружения
4. Международный уголовный суд и его юрисдикция
5. Индивидуальная уголовная ответственность в рамках юрисдикции Международного уголовного суда
6. Индивидуальная уголовная ответственность и автономное вооружение на основе искусственного интеллекта

Заключение

Список литературы

Введение

Войны являются средством расширения территории государств, которые с этой целью прибегают к вооруженным конфликтам (Kalmanovitz, 2022; Kohama, 2019). Во время вооруженного конфликта используются различные методы и средства ведения войны, что приводит к потерям у всех сторон, вовлеченных в конфликт (Bantekas, 2022). Типы вооружений значительно эволюционировали и изменились, особенно во время и после Первой и Второй мировых войн (Fennell, 2019). На смену обычному оружию – мечам, ножам, лукам, пороху – пришло ядерное оружие. Сегодня многие страны стремятся использовать в качестве оружия новейшие технологии – автономное вооружение на основе искусственного интеллекта (далее – ИИ) (Human Rights Watch, 2020). Такое автономное оружие разрабатывается во многих государствах; оно действует с помощью искусственного интеллекта и после активации может выбирать и поражать цели без участия человека-оператора (Horowitz, 2019). Данный тип оружия заменяет человека (Hareth & Evans, 2023).

1. Автономные вооружения на основе искусственного интеллекта

Говоря об автономных вооружениях на основе искусственного интеллекта (далее – ИИ) и робототехники, мы понимаем автономию как способность функционировать в течение продолжительного времени без помощи человека-оператора. Поскольку войны связаны с раздором, многие военные приложения ИИ и робототехники также вызывают споры ([Amoroso & Tamburrini, 2021](#)). Создание и использование летальных автономных систем вооружения, способных самостоятельно принимать решения о жизни и смерти людей, – это, возможно, самый спорный аспект данной темы. Есть мнение, что крылатые ракеты также являются разновидностью летальных автономных систем вооружения. Ракетные системы «Патриот», корабельная боевая система «Иджис», система вооружения «Фаланга», израильская система вооружения «Гарпия» – все это примеры летальных автономных систем вооружения, используемых в настоящее время ([Payne, 2021](#)). Оборонительными вооружениями являются системы «Патриот», «Иджис» и «Фаланга» ([Bartneck et al., 2021](#)). Однако не все военные роботы являются летальными.

Термин «военные роботы» объединяет широкий круг нелетальных приложений ([Bartneck et al., 2021](#); [Krishnan, 2009](#)). Автономные роботы могут использоваться для разминирования, обезвреживания боеприпасов, управления и контроля, разведки, шпионажа, создания мобильных узлов связи, спасательных работ, снабжения, инженерного обеспечения и многого другого ([Burgess, 2017](#)). Содержание споров о военных роботах может меняться в зависимости от их роли ([Malle et al., 2019](#)). Чтобы проиллюстрировать роль робота и человека относительно военных действий, важно дать определение некоторым распространенным терминам. В области ИИ и робототехники автономия понимается как способность функционировать в течение продолжительного времени без помощи человека-оператора ([Totaro, 2023](#)). Роботы могут действовать самостоятельно при принятии непосредственных решений, но в выборе цели они не имеют автономии ([Javdani et al., 2018](#)). Вооружение считается автономным в области «критических функций целеполагания», если оно способно выполнить одно или несколько из нижеперечисленных действий без помощи человека-оператора. Если оружие может выбрать, какие типы объектов поражать, то оно считается автономным в определении целей ([Ekelhof, 2017](#)). В настоящее время автономные системы вооружения не имеют такой способности. Если оружие может использовать датчики для выбора цели без помощи человека-оператора, то оно считается автономным в функции выбора цели.

Многие системы существующего вооружения способны выбирать цели без помощи человека-оператора. Если оружие может произвести выстрел по цели без вмешательства человека-оператора, оно называется автономным в функции поражения цели. Многие ныне существующие виды оружия могут поражать заранее заданные цели. Например, противоракетная система «Патриот» может самостоятельно выбрать цель, но человек-оператор должен нажать на кнопку подтверждения выбора, чтобы выпустить ракету. Выпущенная ракета поразит цель без помощи человека-оператора. Контроль со стороны человека при использовании ракеты «Патриот» невозможен из-за скорости действия ([Bartneck et al., 2021](#)).

Многие другие функции автономных систем вооружения могут быть автономными. Они могут самостоятельно взлетать и приземляться, осуществлять навигацию. Однако такие нелетальные автономные функции обычно не воспринимаются как нравственно сомнительные. В прессе автономное оружие часто называют

«роботом-убийцей». Многие возражают против такого использования термина, описывая его как «хитрую риторическую уловку» (Lokhorst & Van Den Hoven, 2012). С этим не согласна «Кампания против роботов-убийц» – объединенная правозащитная организация, выступающая за запрет летальных автономных систем вооружения во всем мире (Bartneck et al., 2021).

2. Автономное оружие на международном уровне

Когда многие страны критикуют создание автономного оружия, центральной является одна важная тема: риски использования такого оружия для всего человечества относительно достижения военных целей. По мнению сторонников автономного оружия, технология ИИ несет в себе как риски, так и преимущества. При решении вопроса о регулировании этой спорной области технологий необходимо учитывать анализ рисков и преимуществ для ситуаций летальных и нелетальных целей. При этом поднимаются этические и правовые проблемы использования автономного оружия согласно международному праву. Для более глубокого рассмотрения такого метода ведения войны, как автономные вооружения на основе искусственного интеллекта, необходимо вспомнить ряд инцидентов, которые привели к созданию правового регулирования в этой области.

Проблема автономных вооружений на основе искусственного интеллекта широко обсуждалась в 2010 г., когда Филип Алстон, являвшийся в то время специальным докладчиком по внесудебным, суммарным и произвольным казням, поднял этот вопрос в своем предварительном докладе на 65-й сессии Генеральной Ассамблеи ООН. Алстон утверждал, что «автономные технологии становятся все более сложными, исследования способности искусственного интеллекта размышлять и принимать решения получают значительное финансирование и активизируются. Разработчики из разных стран в сфере военной и оборонной промышленности объединяются для создания “полностью автономных систем”, которые позволят беспилотным летательным аппаратам принимать и выполнять сложные решения, в том числе распознавать человеческие цели и поражать их»¹. Затем в 2013 г. Кристоф Хейнс, занимавший ту же должность, опубликовал доклад с обсуждением проблем «летальной автономной робототехники».

Сразу после выпуска рекомендаций консультативного комитета по вопросам разоружения на 68-й сессии Генеральной Ассамблеи ООН, в 2014 г. началось обсуждение автономных систем вооружений в рамках Конвенции о запрещении или ограничении применения конкретных видов обычного оружия, которые могут считаться наносящими чрезмерные повреждения или имеющими неизбирательное действие (ред. от 21 декабря 2001 г.) (UN CCW). Для работы над этой проблемой в 2016 г. была создана группа правительственных экспертов по новейшим технологиям в сфере летальных систем автономных вооружений (Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, GGE on LAWS). Группа продолжает существовать, однако к сентябрю 2022 г. никаких конкретных шагов по созданию нормативной базы по автономным вооружениям не предпринималось.

¹ <https://digitallibrary.un.org/record/690463?ln=en>

На Генеральной Ассамблее ООН впервые страны всего мира выступили с совместным заявлением на тему систем автономных вооружений. Оно стало крупнейшим межрегиональным заявлением в ООН по данному вопросу, объединив 70 государств. Хотя дискуссии по поводу вышеуказанной Конвенции до сих пор не принесли никаких результатов, заявление на Генеральной Ассамблее ООН показывает стремление стран мира двигаться в направлении новых международных норм в области систем автономных вооружений. Это заявление, выдвинутое от имени группы стран послом Александром Кменттом, директором отдела по разоружению, контролю над вооружениями и нераспространению оружия МИД Австрии, объединяет ключевые элементы давно ожидаемого глобального отклика; в частности, в нем подчеркивается необходимость «признать, что появление систем автономных вооружений поднимает серьезные вопросы в гуманитарной, правовой, технологической, этической сферах, а также в сфере безопасности; признать важность ответственности и подотчетности человека при использовании оружия; а также важность правил и ограничений, признанных на международном уровне, включая систему запретов и норм в отношении систем автономных вооружений»².

3. Международное право в области автономного вооружения

Конвенция о запрещении или ограничении применения конкретных видов обычного оружия, которые могут считаться наносящими чрезмерные повреждения или имеющими неизбирательное действие, в редакции от 21 декабря 2001 г., также известна как Конвенция о конкретных видах обычного оружия (Convention on Certain Conventional Weapons, CCW)³; ее также часто называют Конвенцией о негуманном оружии. Цель Конвенции – запретить или ограничить использование таких видов оружия, которые могут причинить чрезмерные или неоправданные страдания военнослужащим или неизбирательно поражать гражданских лиц. Четкая структура Конвенции позволяет адаптировать ее к новым технологиям, используемым в вооруженных конфликтах.

Основной текст Конвенции задает общие положения, такие как правила присоединения, возможность обсуждать и принимать новые протоколы. Протоколы Конвенции содержат запреты и ограничения по существу относительно конкретных видов оружия. Конвенция, включающая три протокола, была принята 10 октября 1980 г. и открыта для подписания 10 апреля 1981 г. на один год. Она была подписана 50 государствами и вступила в силу 2 декабря 1983 г. Три первоначальных протокола были озаглавлены: Протокол I о необнаруживаемых фрагментах; Протокол II о запрещении или ограничении применения мин, мин-ловушек и других устройств и Протокол III о запрещении или ограничении применения зажигательного оружия.

² На Генеральной Ассамблее ООН 70 государств распространили совместное заявление относительно автономных систем вооружения. <https://www.stopkillerrobots.org/news/70-states-deliver-joint-statement-on-autonomous-weapons-systems-at-un-general-assembly>

³ Конвенция о запрещении или ограничении применения конкретных видов обычного оружия, которые могут считаться наносящими чрезмерные повреждения или имеющими неизбирательное действие (принята 10 октября 1980 г., вступила в силу 2 декабря 1983 г.). 1342 UNTS 137.

Позднее были добавлены Протокол IV об ослепляющем лазерном оружии, принятый 13 октября 1995 г. на Первой обзорной конференции стран-участниц согласно ст. 8(3)(b) Конвенции и вступивший в силу 30 июля 1998 г., и поправка к Протоколу II о запрещении или ограничении применения мин, мин-ловушек и других устройств от 3 мая 1996 г., принятая на Первой обзорной конференции в соответствии со ст. 8(1)(b) Конвенции и вступившая в силу 3 декабря 1998 г. Существует также поправка к ст. 1, расширяющая область применения Конвенции на ситуации внутренних вооруженных конфликтов, которая была принята на Второй обзорной конференции в декабре 2001 г. согласно статье 8(1)(b) Конвенции и вступила в силу 18 мая 2004 г. Наконец, Протокол V о взрывоопасных пережитках войны стал первым обсуждавшимся на многосторонней основе документом по проблеме неразорвавшихся и оставленных боеприпасов; он был принят 28 ноября 2003 г. на встрече стран-участниц согласно ст. 5(3) Конвенции и вступил в силу 12 ноября 2006 г.

4. Международный уголовный суд и его юрисдикция

Римский статут был принят международным сообществом в июле 1998 г.⁴ и вступил в силу в 2002 г.⁵, когда был создан Международный уголовный суд (далее – МУС) в соответствии с преамбулой 10 и ст. 1 Римского статута. МУС стал первым и давно ожидаемым постоянно действующим международным уголовным судом (Pella, 1950), призванным покончить с безнаказанностью международных преступлений (Schabas, 2009), таких как геноцид, преступления против человечества, военные преступления и проявления агрессии, согласно ст. 6, 7, 8, 8bis и 15ter соответственно. МУС был создан после нескольких специальных трибуналов, учрежденных международным сообществом после Второй мировой войны союзниками-победителями, включая Международный военный трибунал в Нюрнберге согласно Лондонскому уставу⁶ и Международный военный трибунал по Дальнему Востоку в Токио, согласно декларации генерала МакАртура, верховного главнокомандующего союзных держав во Второй мировой войне⁷, а также два Международных уголовных трибунала

⁴ Дипломатическая конференция полномочных представителей под эгидой ООН по учреждению Международного уголовного суда, финальные документы: Римский статут Международного уголовного суда и Заключительный акт Дипломатической конференции полномочных представителей под эгидой ООН по учреждению Международного уголовного суда [с дополнением, содержащим резолюции, принятые на конференции], т. I (Рим, 15 июня – 17 июля 1998 г.). UN Doc A/CONF.183/13.

⁵ Римский статут Международного уголовного суда (принят 17 июля 1998 г., вступил в силу 1 июля 2002 г.) 2187 UNTS 3.

⁶ Соглашение между правительством США, Временным правительством Французской Республики, правительством Соединенного Королевства Великобритании и Северной Ирландии и правительством СССР о судебном преследовании и наказании главных военных преступников стран европейской оси (подписано в Лондоне 8 августа 1945 г. согласно Уставу Международного военного трибунала, вступило в силу 8 августа 1945 г.).

⁷ Официальное объявление Главнокомандующего союзными оккупационными войсками в Токио (19 января 1946 г.); Устав от 19 января 1946 г.; Изменения в Уставе от 26 апреля 1946 г. – Трибунал создан 19 января 1946 г. (утверждено в Токио 19 января 1946 г.).

Совета безопасности ООН по бывшей Югославии⁸ и Руанде⁹, действовавших согласно главе VII Устава ООН в 1990-х гг.¹⁰

В отличие от Международного суда ООН, имеющего юрисдикцию только над государствами¹¹, МУС имеет юрисдикцию над физическими лицами, как предусмотрено ст. 25(1) Римского статута, достигшими возраста 18 лет на момент совершения преступлений; иначе они считаются несовершеннолетними, согласно ст. 26 Римского статута. Аналогичным образом, в других частях Римского статута также упоминается «лицо»; в частности, в ст. 1: «...осуществлять юрисдикцию в отношении лиц...», в ст. 20: «...никакое лицо...» и «...соответствующее лицо...», в ст. 22: «...лицо не подлежит...» и «...в пользу лица...», в ст. 23: «...лицо...».

Поскольку эти преступления являются международными по определению, государства обязаны проводить расследования и осуществлять наказания за них (Hassan & Osman, 2019). Если страна не может или не желает этого делать, то МУС осуществляет свою юрисдикцию по этим преступлениям согласно принципу дополнительности, отраженному в ст. 17 Римского статута. Другими словами, власти страны являются *forum conveniens* – эта фраза на латыни означает «самый подходящий суд для решения определенного спора или дела», это суд первой юрисдикции, имеющий способность и желание расследовать и выносить приговор конкретным субъектам предполагаемого преступления.

5. Индивидуальная уголовная ответственность в рамках юрисдикции МУС

Согласно ст. 10 Римского статута, «ничто в настоящей Части Положения о ненадании ущерба не должно истолковываться как каким бы то ни было образом ограничивающее ныне действующие или складывающиеся нормы международного права или наносящее им ущерб для целей, отличных от целей настоящего Статута»¹². Как обсуждалось в предыдущих разделах настоящей статьи, существует несколько соглашений, принятых международным сообществом для регулирования автономных вооружений на основе искусственного интеллекта. Кроме того, ст. 21 Римского статута позволяет Международному уголовному суду при разрешении споров применять «в соответствующих случаях применимые международные договоры, принципы и нормы международного права, включая общепризнанные принципы международного права вооруженных конфликтов»¹³. Хотя положения Римского статута не ограничивают развитие международного права и его применимость в Международном уголовном суде при разбирательстве любых дел, предъявленных в этом суде, однако судебному расследованию и разбирательству в этом суде может подвергаться только физическое лицо, согласно ст. 25(1) Римского статута, независимо от его официального статуса главы государства, главы правительства или другого официального лица, как указано в ст. 27(1) Римского статута.

⁸ UNSC Res 827 (25 мая 1993 г.) UN Doc S/RES/827.

⁹ UNSC Res 955 (8 ноября 1994 г.) UN Doc S/RES/955.

¹⁰ Устав ООН (24 октября 1945 г.) 1 UNTS XVI. См. Глава VII.

¹¹ Статут Международного уголовного суда (ICJ Statute), Статья 34(1).

¹² Римский статут Международного уголовного суда (принят 17 июля 1998 г., вступил в силу 1 июля 2002 г.) 2187 UNTS 3.

¹³ Там же.

Положение о подсудности в Международном уголовном суде лица независимо от его официального статуса за совершение международных преступлений не является новшеством, а применяется во множестве международных трибуналов, таких как Международный военный трибунал в Нюрнберге, согласно ст. 6 и 7 его Устава, Международный военный трибунал по Дальнему Востоку, согласно ст. 5 и 6 Устава этого трибунала, Международный уголовный трибунал по бывшей Югославии, согласно ст. 6 и 7 Статута этого трибунала, Международный уголовный трибунал по Руанде, согласно ст. 5 и 6 Статута этого трибунала. Что касается Международного уголовного суда, то в ст. 25(3) Римского статута перечислены шесть различных режимов или ситуаций, когда лицо несет уголовную ответственность и подлежит наказанию за преступление в рамках юрисдикции этого суда, а также основные нормы индивидуальной уголовной ответственности и нормы расширения атрибуции (Ambos, 2016).

I. Если лицо совершило преступление¹⁴

Первый режим уголовной ответственности, согласно ст. 25(3)(a) Римского статута, предусматривает, что лицо несет уголовную ответственность и подлежит наказанию за преступление в рамках юрисдикции этого Суда, если это лицо «совершает такое преступление индивидуально, совместно с другим лицом или через другое лицо, независимо от того, подлежит ли это другое лицо уголовной ответственности»¹⁵. Это общепринятый принцип уголовного права¹⁶, закрепленный Международным военным трибуналом в Нюрнберге как принцип индивидуальной уголовной ответственности, который гласит, что «преступления против международного права совершаются людьми, а не абстрактными сообществами, и положения международного права могут быть реализованы лишь путем наказания отдельных лиц, совершивших такие преступления»¹⁷. В рамках данного режима индивидуальной уголовной ответственности выделяются три формы нарушения закона: индивидуально, совместно с другим лицом или через другое лицо (средство правонарушения)¹⁸.

II. Если лицо приказывает, подстрекает или побуждает совершить такое преступление

Второй режим уголовной ответственности, согласно ст. 25(3)(b) Римского статута, предусматривает, что лицо несет уголовную ответственность и подлежит наказанию за преступление в рамках юрисдикции этого Суда, если это лицо «приказывает, подстрекает или побуждает совершить такое преступление, если это преступление совершается или если имеет место покушение на это преступление»¹⁹.

¹⁴ Римский статут, Статья 25(3)(a).

¹⁵ Там же.

¹⁶ См. *Prosecutor v Dusko Tadic* (Решение по ходатайству защиты о предварительной апелляции по юрисдикции) IT-94-1 (1995, October 2) (разд. 128–137). В разделе 134 данного решения Международный уголовный трибунал по бывшей Югославии указал, что «все эти факторы подтверждают, что обычное международное право устанавливает уголовную ответственность за серьезные нарушения статьи 3, что подкреплено другими общими принципами и нормами защиты пострадавших во внутреннем вооруженном конфликте, а также за нарушение определенных фундаментальных принципов и норм в отношении средств и методов военных действий в гражданском конфликте».

¹⁷ Суд над главными военными преступниками в Международном военном трибунале (Нюрнберг, 14 ноября 1945 г. – 1 октября 1946 г.), т. I (Нюрнберг, 1947 г.).

¹⁸ Kai Ambos. (2016). 'Article 25: Individual Criminal Responsibility'. In Otto Triffterer, Kai Ambos (Eds.), *The Roman Statute of the International Criminal Court: A Commentary* (3rd edn). Nomos, 984.

¹⁹ Римский статут, Статья 25(3)(b).

III. Если лицо способствует совершению такого преступления

Третий режим уголовной ответственности, согласно ст. 25(3)(c) Римского статута, предусматривает, что лицо несет уголовную ответственность и подлежит наказанию за преступление в рамках юрисдикции этого Суда, если это лицо с целью облегчить совершение такого преступления пособничает, подстрекает или каким-либо иным образом содействует его совершению или покушению на него, включая предоставление средств для его совершения.

IV. Если это лицо любым другим образом способствует совершению или покушению на совершение такого преступления группой лиц, действующих с общей целью

Четвертый режим уголовной ответственности, согласно ст. 25(3)(d) Римского статута, предусматривает, что лицо несет уголовную ответственность и подлежит наказанию за преступление в рамках юрисдикции этого Суда, если это лицо «любым другим образом способствует совершению или покушению на совершение такого преступления группой лиц, действующих с общей целью»²⁰. Такое содействие должно оказываться умышленно и либо «в целях поддержки преступной деятельности или преступной цели группы в тех случаях, когда такая деятельность или цель связана с совершением преступления, подпадающего под юрисдикцию Суда»²¹, либо «с осознанием умысла группы совершить преступление»²².

V. В отношении преступления геноцида прямо и публично подстрекает других к совершению геноцида.

VI. Покушается на совершение такого преступления, предпринимая действие, которое представляет собой значительный шаг в его совершении, однако преступление оказывается незавершенным по обстоятельствам, не зависящим от намерений данного лица. Вместе с тем лицо, которое отказывается от попытки совершить преступление или иным образом предотвращает завершение преступления, не подлежит наказанию в соответствии с настоящим Статутом за покушение на совершение этого преступления, если данное лицо полностью и добровольно отказалось от преступной цели.

6. Индивидуальная уголовная ответственность и автономное вооружение на основе искусственного интеллекта

Если иногда очень сложно связать лицо, ответственное за совершение преступления, особенно когда оно географически или структурно находится далеко от места преступления, то насколько это сложнее в отношении преступлений, предположительно совершенных с помощью автономного оружия на основе искусственного интеллекта. Статья 36 Дополнительного протокола I к Женевской конвенции 1949 г.²³ предусматривает, что государства обязаны пересматривать законность намерений по развертыванию новых вооружений. Важно обеспечить способность вооруженных

²⁰ Римский статут, Статья 25(3)(b).

²¹ Римский статут, Статья 25(3)(d)(i).

²² Там же. Статья 25(3)(d)(ii).

²³ Дополнительный протокол к Женевской конвенции от 12 августа 1949 г. в отношении защиты жертв международных вооруженных конфликтов (Протокол I) (принят 8 июня 1977 г., вступил в силу 7 декабря 1978 г.).

сил государства выполнять боевые действия, не нарушая международных обязательств (Lawand, 2006). В ст. 36(2) Дополнительного протокола I далее говорится, что при разработке новых технологий вооружения юристы и политики должны сделать соответствующие поправки в отношении закона и подотчетности лиц, серьезно нарушающих закон, согласно ст. 49 Женевской конвенции I ²⁴.

Статья 49 Женевской конвенции I предусматривает, что «высокие договаривающиеся стороны берут на себя обязательство ввести в действие законодательство, необходимое для обеспечения эффективных уголовных наказаний для лиц, совершивших или приказавших совершить те или иные серьезные нарушения настоящей Конвенции»²⁵. Кроме того, в ней сказано, что «каждая Высокая договаривающаяся сторона обязуется разыскивать лиц, обвиняемых в том, что они совершили или приказали совершить то или иное из упомянутых серьезных нарушений, и, каково бы ни было их гражданство, предавать их своему суду. Она сможет также, если она этого пожелает, передать их в соответствии с положениями своего законодательства для суда другой заинтересованной Высокой договаривающейся стороне в том случае, если эта Договаривающаяся сторона имеет доказательства, дающие основания для обвинений этих лиц»²⁶.

Что касается полностью беспилотного автономного оружия на основе искусственного интеллекта, то приказы операторов такого оружия являются запрограммированными заранее, а значит, следует считать, что правовая ответственность за любые действия должна быть перенесена с операторов на систему, управляемую искусственным интеллектом. Однако при этом встает вопрос правовых обязательств: лежит ответственность за принятые решения на оружии или на его операторе? В этом смысле никто не будет отвечать за намеренное действие или бездействие. Однако разработчик оружия, программист, изготовитель также могут быть привлечены к ответственности только в той степени, в какой они намеренно способствовали совершению преступления (McFarland & McCormack, 2014).

При использовании автономного оружия, особенно полностью свободного от вмешательства человека и находящегося под контролем искусственного интеллекта, человек не испытывает эмпатии и не выносит суждений (Gunawan et al., 2022). Воздействие человека на системы вооружения или использование военной мощи должно отвечать определенным правовым и этическим требованиям, как было отмечено в заявлении Международного комитета Красного Креста на встрече экспертов по летальным автономным системам вооружения в Женеве 11 апреля 2017 г. относительно Конвенции о конкретных видах обычного оружия.

Заключение

Технологический прогресс достиг высоких стандартов, и международное сообщество требует обеспечить не только защиту своих границ и граждан от вторжения и нападения, но и защиту своих войск от уничтожения. Это привело к созданию новых

²⁴ Конвенция (I) об улучшении участия раненых и больных в действующих армиях (принята 12 августа 1949 г., вступила в силу 21 октября 1950 г.). 75 UNTS 31.

²⁵ Там же.

²⁶ Там же.

технологий вооружения, а именно автономного оружия на основе искусственного интеллекта. Однако такая технология не освобождает от ответственности в рамках международного права, подвергается активной критике и вызывает тревогу международного сообщества. Автономное вооружение на основе искусственного интеллекта может предпринимать и осуществлять атаки, вызывая потери на невоенных объектах. Созданный в 2002 г. на основе Римского статута Международный уголовный суд предлагает решение этой проблемы относительно даже самых продвинутых вооружений, таких как беспилотные устройства на основе искусственного интеллекта; при этом лица, связанные с созданием и управлением такими устройствами, будут нести уголовную ответственность, если преступят закон в случае участия в войне или вооруженном конфликте.

Список литературы

- Ambos, K. (2016). Article 25: Individual Criminal Responsibility. In O. Triffterer, & K. Ambos (Eds.), *The Rome Statute of the International Criminal Court: A Commentary* (3rd ed.). Nomos.
- Amoroso, D., & Tamburrini, G. (2021). In Search of the 'Human Element': International Debates on Regulating Autonomous Weapons Systems. *The International Spectator*, 56(1), 20–38. <https://doi.org/10.1080/03932729.2020.1864995>
- Bantekas, I. (2022). Punishment in Warfare and the Application of Law. In *Principles of Direct and Superior Responsibility in International Humanitarian Law* (pp. 1–37). Manchester UP.
- Bartneck, C., Lütge, C., Wagner, A., & Welsh, S. (2021). *Military Uses of AI BT*. In *An Introduction to Ethics in Robotics and AI* (pp. 93–99). Springer International Publishing. https://doi.org/10.1007/978-3-030-51110-4_11
- Burgess, A. (2017). *The Executive Guide to Artificial Intelligence: How to Identify and Implement Applications for AI in Your Organization*. Springer. <https://doi.org/10.1007/978-3-319-63820-1>
- Ekelhof, M. A. (2017). Complications of a Common Language: Why It Is So Hard to Talk About Autonomous Weapons. *Journal of Conflict and Security Law*, 22(2), 311–331. <https://doi.org/10.1093/jcsl/krw029>
- Fennell, J. (2019). *Fighting the People's War: The British and Commonwealth Armies and the Second World War*. CUP. <https://doi.org/10.1017/9781139380881>
- Gunawan, Y., Anggriawan, M. H. A. R., & Putro, T. A. (2022). Command Responsibility of Autonomous weapons under International Humanitarian Law. *Cogent Social Sciences*, 8(1), 2139906. <https://doi.org/10.1080/23311886.2022.2139906>
- Hareth, B., & Evans, N. G. (2023). Make Them Rare or Make Them Care. In D. Schoeni & T. Vestner (Eds.), *Ethical Dilemmas in the Global Defense Industry* (pp. 217–236). OUP.
- Hassan, F. M., & Osman, N. D. (2019). The Obligation to Prosecute Heads of State Under the Rome Statute of the International Criminal Court (ICC) and Customary International Law: The African And United States' Perspectives. *MJSL*, 7(1), 33–56. <https://doi.org/10.33102/mjssl.v7i1.112>
- Horowitz, M. C. (2019). When Speed Kills: Lethal Autonomous Weapon Systems, Deterrence and Stability. *Journal of Strategic Studies*, 42(6), 764–788. <https://doi.org/10.1080/01402390.2019.1621174>
- Human Rights Watch. (2020). *Stop Killer Robots: Country Positions on Banning Fully Autonomous Weapons and Retaining Human Control*.
- Javdani, S., Admoni, H., Srinivasa, S. P. S. S., & Bagnell, J. A. (2018). Shared Autonomy via Hindsight Optimization for Teleoperation and Teaming. *International Journal of Robotics Research*, 37(7), 717–742. <https://doi.org/10.1177/0278364918776060>
- Kalmanovitz, P. (2022). Can Criminal Organizations Be Non-State Parties to Armed Conflict? In *International Review of the Red Cross* (pp. 1–19). ICRC, CUP. <https://doi.org/10.1017/S1816383122000510>
- Kohama, S. (2019). Territorial Acquisition, Commitment, and Recurrent War. *International Relations of the Asia-Pacific*, 19(2), 269–295. <https://doi.org/10.1093/irap/lcy001>
- Krishnan, A. (2009). Automating War: The Need for Regulation. *Contemporary Security Policy*, 30(1), 172–193. <https://doi.org/10.1080/13523260902760397>
- Lawand, K. (2006). Reviewing the legality of new weapons, means and methods of warfare. *International Review of the Red Cross*, 88, 925–930. <https://doi.org/10.1017/S1816383107000884>

- Lokhorst, G.-J., & Hoven, J. van den. (2012). Responsibility for military robots. In P. Lin, K. Abney, & G. A. Bekey (Eds.), *Robot Ethics: The Ethics and Social Implications of Robotics* (pp. 145–156). MIT Press.
- Malle, B. F., Magar, S. T., & Scheutz, M. (2019). AI in the Sky: How People Morally Evaluate Human and Machine Decisions in a Lethal Strike Dilemma. In M. I. A. Ferreira, J. S. Sequeira, G. S. Virk, M. O. Tokhi, & E. E. Kadar (Eds.), *Robotics and Well-being* (pp. 111–133). Springer. https://doi.org/10.1007/978-3-030-12524-0_11
- McFarland, T., & McCormack, T. (2014). Mind the Gap: Can Developers of Autonomous Weapon Systems be Liable for War Crimes?, *Internatoinal Law Studies*, 90, 350–362.
- Payne, K. (2021). *I, Warbot: The Dawn of Artificially Intelligent Conflict*. OUP.
- Pella V. (1950). Towards an International Criminal Court. *American Journal of International Law*, 44(1), 38–49.
- Schabas, William A. (2009). International Crimes. In D. Armstrong (Ed.), *Routledge Handbook of International Law*. Routledge. <https://doi.org/10.4324/9780203884621>
- Totaro, D. L. (2023). Machine or Robot? Thoughts on the Legal Notion of Autonomy in the Context of Self-Driving Vehicles and Intelligent Machines. *European Business Law Review*, 34(1), 99–114.

Сведения об авторах



Фарид Мохд Хасан – доктор наук, старший преподаватель факультета шариата и права, Исламский научный университет Малайзии, Нилаи, Негри-Сембилан, Малайзия

Адрес: Бандар Бару Нилаи, 71800 Нилаи, Негри-Сембилан, Малайзия

E-mail: fareed@usim.edu.my

ORCID ID: <https://orcid.org/0009-0006-7501-0782>

Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=57222041043>

Google Scholar ID: <https://scholar.google.com/citations?user=2LjkkQ4AAAAJ&hl=en>



Нур Джухаида Осман – доктор наук, старший преподаватель факультета шариата и права, Исламский научный университет Малайзии, Нилаи, Негри-Сембилан, Малайзия

Адрес: Бандар Бару Нилаи, 71800 Нилаи, Негри-Сембилан, Малайзия

E-mail: noordzuhaidah@usim.edu.my

ORCID ID: <https://orcid.org/0000-0003-2659-9309>

Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=57226308331>

Web of Science Researcher ID:

<https://www.webofscience.com/wos/author/record/AAT-6652-2021>

Google Scholar ID: https://scholar.google.com/citations?user=h48_cAIAAAAJ&hl=en

Вклад авторов

Авторы внесли равный вклад в создание статьи.

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

Финансирование

Исследование не имело спонсорской поддержки.

Тематические рубрики

Рубрика OECD: 5.05 / Law

Рубрика ASJC: 3308 / Law

Рубрика WoS: OM / Law

Рубрика ГРНТИ: 10.87.89 / Международное уголовное право

Специальность ВАК: 5.1.5 / Международно-правовые науки

История статьи

Дата поступления – 28 февраля 2023 г.

Дата одобрения после рецензирования – 23 апреля 2023 г.

Дата принятия к опубликованию – 16 июня 2023 г.

Дата онлайн-размещения – 20 июня 2023 г.



Research article

DOI: <https://doi.org/10.21202/jdtl.2023.19>



AI-based Autonomous Weapons and Individual Criminal Responsibility under the Rome Statute

Fareed Mohd Hassan 

Universiti Sains Islam Malaysia
Nilai, Negeri Sembilan, Malaysia

Noor Dzuhaidah Osman  

Universiti Sains Islam Malaysia
Nilai, Negeri Sembilan, Malaysia

Keywords

Armed conflict,
Artificial intelligence,
Autonomous weapons,
Criminal liability,
Digital technologies,
International criminal court,
law,
robotics,
Rome Statute,
war

Abstract

Objective: international law obligates states to prosecute those who have violated laws in armed conflicts, particularly when the international community now has International Criminal Court (ICC).

That is why the aim of the paper is to discover the responsibility for the crimes made with the use of AI-based autonomous vehicles in accordance with the provisions of the Rome Statute of the ICC.

Methods: doctrinal analysis allowed to research the positions of experts on the responsibility for the crimes made with the use of AI-based autonomous vehicles in accordance with the provisions of the Rome Statute of the ICC.

Results: this paper argues that the ICC can only exercise jurisdiction over natural persons who allegedly have committed the crimes under its jurisdiction, as compared to autonomous weapons. This paper argues that the persons who facilitate the commission of the alleged crimes are highly likely to be criminally responsible for providing means for the alleged crimes to be committed by AI-based autonomous weapons under Article 25(3)(c) of the Rome Statute and concludes that the Rome Statute provides a solution even to AI-based autonomous weapons.

 Corresponding author

© Hassan F. M., Osman N. D., 2023



This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Scientific novelty: this paper addresses to the highly relevant issues of the responsibility for the crimes made with the use of AI-based autonomous vehicles in accordance with the provisions of the Rome Statute of the ICC.

Practical significance: the results achieved in the paper can be used in regulation design for AI-based autonomous weapons. It can also be used as a basis for the future research in the sphere of liability of AI-based autonomous weapons and AI in general.

For citation

Hassan, F. M., & Osman, N. D. (2023). AI-Based Autonomous Weapons and Individual Criminal Responsibility Under the Rome Statute. *Journal of Digital Technologies and Law*, 1(2), 464–480. <https://doi.org/10.21202/jdtl.2023.19>

References

- Ambos, K. (2016). Article 25: Individual Criminal Responsibility. In O. Triffterer, & K. Ambos (Eds.), *The Rome Statute of the International Criminal Court: A Commentary* (3rd ed.). Nomos.
- Amoroso, D., & Tamburrini, G. (2021). In Search of the 'Human Element': International Debates on Regulating Autonomous Weapons Systems. *The International Spectator*, 56(1), 20–38. <https://doi.org/10.1080/03932729.2020.1864995>
- Bantekas, I. (2022). Punishment in Warfare and the Application of Law. In *Principles of Direct and Superior Responsibility in International Humanitarian Law* (pp. 1–37). Manchester UP.
- Bartneck, C., Lütge, C., Wagner, A., & Welsh, S. (2021). *Military Uses of AI BT*. In *An Introduction to Ethics in Robotics and AI* (pp. 93–99). Springer International Publishing. https://doi.org/10.1007/978-3-030-51110-4_11
- Burgess, A. (2017). *The Executive Guide to Artificial Intelligence: How to Identify and Implement Applications for AI in Your Organization*. Springer. <https://doi.org/10.1007/978-3-319-63820-1>
- Ekelhof, M. A. (2017). Complications of a Common Language: Why It Is So Hard to Talk About Autonomous Weapons. *Journal of Conflict and Security Law*, 22(2), 311–331. <https://doi.org/10.1093/jcsl/krw029>
- Fennell, J. (2019). *Fighting the People's War: The British and Commonwealth Armies and the Second World War*. CUP. <https://doi.org/10.1017/9781139380881>
- Gunawan, Y., Anggriawan, M. H. A. R., & Putro, T. A. (2022). Command Responsibility of Autonomous weapons under International Humanitarian Law. *Cogent Social Sciences*, 8(1), 2139906. <https://doi.org/10.1080/23311886.2022.2139906>
- Hareth, B., & Evans, N. G. (2023). Make Them Rare or Make Them Care. In D. Schoeni & T. Vestner (Eds.), *Ethical Dilemmas in the Global Defense Industry* (pp. 217–236). OUP.
- Hassan, F. M., & Osman, N. D. (2019). The Obligation to Prosecute Heads of State Under the Rome Statute of the International Criminal Court (ICC) and Customary International Law: The African And United States' Perspectives. *MJSL*, 7(1), 33–56. <https://doi.org/10.33102/mjst.v7i1.112>
- Horowitz, M. C. (2019). When Speed Kills: Lethal Autonomous Weapon Systems, Deterrence and Stability. *Journal of Strategic Studies*, 42(6), 764–788. <https://doi.org/10.1080/01402390.2019.1621174>
- Human Rights Watch. (2020). *Stop Killer Robots: Country Positions on Banning Fully Autonomous Weapons and Retaining Human Control*.
- Javdani, S., Admoni, H., Srinivasa, S. P. S. S., & Bagnell, J. A. (2018). Shared Autonomy via Hindsight Optimization for Teleoperation and Teaming. *International Journal of Robotics Research*, 37(7), 717–742. <https://doi.org/10.1177/0278364918776060>
- Kalmanovitz, P. (2022). Can Criminal Organizations Be Non-State Parties to Armed Conflict? In *International Review of the Red Cross* (pp. 1–19). ICRC, CUP. <https://doi.org/10.1017/S1816383122000510>
- Kohama, S. (2019). Territorial Acquisition, Commitment, and Recurrent War. *International Relations of the Asia-Pacific*, 19(2), 269–295. <https://doi.org/10.1093/irap/lcy001>

- Krishnan, A. (2009). Automating War: The Need for Regulation. *Contemporary Security Policy*, 30(1), 172–193. <https://doi.org/10.1080/13523260902760397>
- Lawand, K. (2006). Reviewing the legality of new weapons, means and methods of warfare. *International Review of the Red Cross*, 88, 925–930. <https://doi.org/10.1017/S1816383107000884>
- Lokhorst, G.-J., & Hoven, J. van den. (2012). Responsibility for military robots. In P. Lin, K. Abney, & G. A. Bekey (Eds.), *Robot Ethics: The Ethics and Social Implications of Robotics* (pp. 145–156). MIT Press.
- Malle, B. F., Magar, S. T., & Scheutz, M. (2019). AI in the Sky: How People Morally Evaluate Human and Machine Decisions in a Lethal Strike Dilemma. In M. I. A. Ferreira, J. S. Sequeira, G. S. Virk, M. O. Tokhi, & E. E. Kadar (Eds.), *Robotics and Well-being* (pp. 111–133). Springer. https://doi.org/10.1007/978-3-030-12524-0_11
- McFarland, T., & McCormack, T. (2014). Mind the Gap: Can Developers of Autonomous Weapon Systems be Liable for War Crimes?, *Internatoinal Law Studies*, 90, 350–362.
- Payne, K. (2021). *I, Warbot: The Dawn of Artificially Intelligent Conflict*. OUP.
- Pella V. (1950). Towards an International Criminal Court. *American Journal of International Law*, 44(1), 38–49.
- Schabas, William A. (2009). International Crimes. In D. Armstrong (Ed.), *Routledge Handbook of International Law*. Routledge. <https://doi.org/10.4324/9780203884621>
- Totaro, D. L. (2023). Machine or Robot? Thoughts on the Legal Notion of Autonomy in the Context of Self-Driving Vehicles and Intelligent Machines. *European Business Law Review*, 34(1), 99–114.

Authors information



Fareed Mohd Hassan – PhD, Senior Lecturer, Faculty of Syariah and Law, Universiti Sains Islam Malaysia (USIM) Nilai, Negeri Sembilan, Malaysia
Address: Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan, Malaysia
E-mail: fareed@usim.edu.my
ORCID ID: <https://orcid.org/0009-0006-7501-0782>
Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=57222041043>
Google Scholar ID: <https://scholar.google.com/citations?user=2LjkkQ4AAAAJ&hl=en>



Noor Dzuhaidah Osman – PhD, Senior Lecturer, Faculty of Syariah and Law, Universiti Sains Islam Malaysia (USIM) Nilai, Negeri Sembilan, Malaysia
Address: Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan, Malaysia
E-mail: oordzuhaidah@usim.edu.my
ORCID ID: <https://orcid.org/0000-0003-2659-9309>
Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=57226308331>
Web of Science Researcher ID: <https://www.webofscience.com/wos/author/record/AAT-6652-2021>
Google Scholar ID: https://scholar.google.com/citations?user=h48_cAIAAAAAJ&hl=en

Authors' contributions

Authors equally contribute to the paper.

Conflict of interests

The author declares no conflict of interests.

Financial disclosure

The research had no sponsorship.

Thematic rubrics

OECD: 5.05 / Law

PASJC: 3308 / Law

WoS: OM / Law

Article history

Date of receipt – February 28, 2023

Date of approval – April 23, 2023

Date of acceptance – June 16, 2023

Date of online placement – June 20, 2023

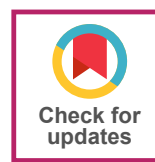


Научная статья

УДК 343.14:004.8

EDN: <https://elibrary.ru/acsqhx>

DOI: <https://doi.org/10.21202/jdtl.2023.20>



Технологии искусственного интеллекта в уголовно-процессуальном доказывании

Михаил Сергеевич Спиридонов 

Южно-Уральский государственный университет (национальный исследовательский университет)
г. Челябинск, Российская Федерация

Ключевые слова

Доказательство,
доказывание,
искусственный интеллект,
нейронная сеть,
право,
судопроизводство,
уголовное дело,
уголовный процесс,
цифровизация,
цифровые технологии

Аннотация

Цель: обобщение и анализ сложившихся в уголовно-процессуальной науке позиций к применению технологий искусственного интеллекта, выработка авторского подхода к перспективам трансформации уголовно-процессуального доказывания под влиянием технологий искусственного интеллекта.

Методы: методологическую основу исследования составляет единство всеобщего, общенаучных и специально-юридических методов правовой науки, в том числе абстрактно-логического, сравнительно-правового и прогностического.

Результаты: определены основные направления применения технологий искусственного интеллекта в уголовном процессе, такие как профилактика и выявление преступлений, организация предварительного расследования, криминалистическое сопровождение расследования преступлений, оценка доказательств на досудебной и судебной стадиях. Автор приходит к выводу, что сложившийся в науке уголовного процесса достаточно оптимистичный подход по данному вопросу значительно опережает реально существующие в настоящее время технологии искусственного интеллекта. Выявлены основные требования, которым должна отвечать деятельность по применению искусственного интеллекта при сборе доказательств по уголовному делу. Обращено внимание на проблемы применения технологий искусственного интеллекта при проведении судебных экспертиз, которые требуют совершенствования методологии судебно-экспертной работы. Рассмотрен вопрос о перспективах трансформации процесса уголовно-процессуального доказывания в условиях внедрения технологий искусственного интеллекта. Обосновывается вывод, что оценка доказательств с помощью математических алгоритмов, в которых применяются заранее установленные значения качества

© Спиридонов М. С., 2023



Статья находится в открытом доступе и распространяется в соответствии с лицензией Creative Commons «Attribution» («Атрибуция») 4.0 Всемирная (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0/deed.ru>), позволяющей неограниченно использовать, распространять и воспроизводить материал при условии, что оригинальная работа упомянута с соблюдением правил цитирования.

каждого доказательства, противоречит принципу свободы оценки доказательств в уголовном процессе. Автор приходит к выводу об отсутствии в настоящее время достаточных оснований для наделения искусственного интеллекта субъектностью в процессе доказывания.

Научная новизна: в работе предпринята попытка рассмотреть место искусственного интеллекта в уголовно-процессуальном доказывании, выявлены требования, которым должно соответствовать применение этой технологии при сборе доказательств, при этом проанализированы перспективы трансформации процесса доказывания с учетом внедрения технологий искусственного интеллекта.

Практическая значимость: основные положения и выводы исследования могут быть использованы для совершенствования механизма правового регулирования технологий искусственного интеллекта в уголовном процессе.

Для цитирования

Спиридонов, М. С. (2023). Технологии искусственного интеллекта в уголовно-процессуальном доказывании. *Journal of Digital Technologies and Law*, 1(2), 481–497. <https://doi.org/10.21202/jdtl.2023.20>

Содержание

Введение

1. Математические алгоритмы на службе уголовного процесса
 - 1.1. Профилактика и выявление преступлений
 - 1.2. Организация предварительного расследования
 - 1.3. Криминалистическое сопровождение расследования преступлений
 - 1.4. Оценка доказательств на досудебной и судебной стадиях
2. Искусственный интеллект в руках субъекта доказывания
 - 2.1. Проблемы сбора доказательств с использованием технологий искусственного интеллекта
 - 2.2. Применение технологий искусственного интеллекта при проведении экспертиз
3. Перспективы трансформации процесса доказывания

Выводы

Список литературы

Введение

Тема искусственного интеллекта в уголовном процессе кажется довольно футуристичной, поскольку само по себе это явление можно назвать чисто техническим и по своей сути оно является не более чем математическим алгоритмом, задача которого, если упрощенно, состоит в обработке входящей информации и формировании на ее основе новой информации. Однако цифровые технологии настолько быстро входят не только в повседневную жизнь, но и в такие специальные сферы, как уголовное судопроизводство, что наука не может оставаться в стороне и отворачиваться от вопросов, которые, на первый взгляд, не имеют какого-то реального практического применения.

Справедлива точка зрения ученых о том, что мы не будем готовы к новым вызовам, возникающим в связи с современными технологиями, если не будем обсуждать их уже сейчас (Vesnic-Alujevic et al., 2020; Feijóo & Kwon, 2020; Robles Carrillo, 2020). Поэтому особенно важно изучать перспективы применения технологий искусственного интеллекта в уголовном процессе, анализировать возникающие в связи с этим проблемы и искать возможные оптимальные варианты их решения, прогнозировать пути развития уголовного процесса.

Уголовный процесс, как любая другая отрасль правовой науки, нуждается в объективной и непредвзятой оценке того, насколько применение искусственного интеллекта отвечает стоящим перед ним целям и задачам, какое влияние применение этих технологий оказывает и будет оказывать в будущем на защиту личности от необоснованного уголовного преследования. В этой оценке не должно быть излишнего и ничем не оправданного оптимизма, поскольку в отсутствие сильного искусственного интеллекта любые рассуждения о том, что уже завтра роботы заменят судей, не могут восприниматься всерьез. Но и не стоит скатываться к противоположному – слепому отрицанию перспектив применения новых технологий, в том числе в такой сложной и абстрактной сфере, как процесс доказывания.

Следует согласиться с учеными, полагающими, что назрела необходимость проанализировать сложившиеся в науке уголовного процесса подходы к оценке перспектив применения технологий искусственного интеллекта и на основе этого анализа рассмотреть вопрос о месте этих цифровых технологий в уголовно-процессуальном доказывании (Silva et al., 2020; Kaur et al., 2023; Wang & Ma, 2022; Kalai et al., 2022; Cascavilla et al., 2021).

1. Математические алгоритмы на службе уголовного процесса

Еще в 1990 г. профессор Рисслэнд отмечала, что между правом и искусственным интеллектом есть плодотворная синергия; право открывает широкие возможности для разработки аналитических и вычислительных моделей искусственного интеллекта; при этом ряд характеристик права делают его особенно сложной областью для искусственного интеллекта (Rissland, 1990).

Спустя более 30 лет после этих заявлений мы можем выделить несколько направлений, в которых в уголовном процессе (в широком смысле) уже используются или только планируются к применению технологии искусственного интеллекта.

1.1. Профилактика и выявление преступлений

Ученые обращают внимание на потенциал, которым обладает технология искусственного интеллекта в области профилактики и выявления преступлений. Так, например, зарубежные ученые пишут о возможности применения искусственного интеллекта для выявления ряда киберпреступлений (Kaur et al., 2023; Kalai et al., 2022; Cascavilla et al., 2021). В свою очередь, в работе А. М. Цирина и Е. А. Артеменко описывается программный продукт на основе технологий искусственного интеллекта, с помощью которого планируется анализировать проведенные контрольно-надзорные мероприятия и в результате выявлять истинные причины возникновения ущерба и индикаторы риска наиболее потенциально опасных контролируемых лиц (Цирин, Артеменко, 2023). Н. А. Кузьмин в своей статье предлагает создавать в целях борьбы с коррупцией основанные на технологии искусственного интеллекта

системы отслеживания финансовых транзакций, ввода и вывода денежных средств (Кузьмин, 2021). С. В. Расторопов приводит обзор таких программных продуктов в сфере предупреждения и выявления преступлений, как Shot Spotter, Prepol, Cloud Walk Technology, HART, VAA, которые уже успешно применяются в полицейской деятельности (Расторопов, 2020).

1.2. Организация предварительного расследования

В этом направлении Ю. А. Цветковым, например, предлагается разработать «искусственный управленческий мозг», который был бы способен генерировать «оптимальные варианты решений на всех узловых пунктах траектории движения уголовного дела» (Цветков, 2021). М. А. Малина считает возможным применять технологии искусственного интеллекта для выявления и исправления всевозможных нарушений формы процессуальных документов, а также для выявления недоброкачественной и потенциально недостоверной информации, поступающей следователю (Малина, 2021).

1.3. Криминалистическое сопровождение расследования преступлений

Примером может служить описанная в статье Д. Н. Сретенцева и В. Р. Волковой программа «Зеркало», позволяющая выявлять признаки внутрикадрового монтажа видеоизображений, который выполнен с помощью искусственных нейронных сетей, позволяющих синтезировать видеоизображения людей (дипфейк) (Сретенцев, Волкова, 2021). В статье Ф. Рахмана описано исследование, проведенное в Сиракузском университете, в котором машинное обучение использовалось для разделения и идентификации индивидуальных профилей ДНК, а также анализа больших объемов сложных данных для выявления закономерностей, некоторые из которых могут быть недоступны человеческому анализу (Rahman, 2019). В статье И. Сильвы с соавт. (Silva et al., 2020) исследуется возможность применения искусственного интеллекта для анализа снимков как доказательств в уголовном процессе.

1.4. Оценка доказательств на досудебной и судебной стадиях

А. В. Сибилькова полагает, что на основе машинного обучения искусственных нейронных сетей результатам расследования уголовных дел определенной категории может производиться оценка достаточности собранных доказательств, т. е. «искусственные нейронные сети должны получить информацию о том, чего не хватило для вынесения обвинительного приговора» (Сибилькова, 2019). А. В. Макутчев в качестве одной из моделей внедрения искусственного интеллекта в систему правосудия видит его работу «наравне, “в связке” с судьей», либо замену судьи искусственным интеллектом (Макутчев, 2022). А. А. Сумин и О. В. Химичева в своей статье приводят положительный пример использования систем искусственного интеллекта в оценке доказательств на судебной стадии в Китае (Сумин, Химичева, 2020). О проблемах и перспективах применения искусственного интеллекта в рамках данной стадии уголовного процесса писали и зарубежные ученые (Stoykova, 2023; Yassine et al., 2023; Amariles & Baquero, 2023).

Изложенный выше краткий обзор показывает, что в науке уголовного процесса сложился достаточно оптимистичный подход по вопросу применения технологий искусственного интеллекта в различных сферах, в том числе в уголовно-процессуальном доказывании. При этом этот подход в большинстве своем значительно опережает реально существующие в настоящее время технологии искусственного интеллекта, т. е. можно сказать, что ученые в большей мере высказывают свои ожидания и прогнозы относительно возможностей применения новых технологий.

Безусловно, применение новейших технологий в любом из указанных выше направлений может только приветствоваться. Однако нужно отдавать отчет в том, что должна быть достаточно четкая и весомая причина для внедрения той или иной технологии в уголовный процесс. Ведь сам по себе уголовный процесс как неотъемлемая часть юридического процесса не нуждается ни в каких технологиях, он, если можно так выразиться, самодостаточен. Разве нужна хоть какая-то технология для того, чтобы инициировать уголовное дело, собрать доказательства, предоставить их в суд, где затем рассмотреть их и принять решение по существу дела? Конечно, нет. Все это можно сделать даже при отсутствии письменных принадлежностей и бумаги, ведь именно так все и начиналось в человеческой истории.

Как видно, направления использования технологий искусственного интеллекта затрагивают все этапы процесса доказывания по уголовному делу. Часть из них применяется для собирания доказательств, выступая в данном случае своего рода инструментом получения доказательственной информации в руках субъекта доказывания. Другие технологии могут быть применимы в деятельности по проверке и даже оценке доказательств.

Доказательства и доказывание лежат в основе любых процессуальных решений по уголовному делу, и это напрямую влияет на права и интересы участников процесса. Поэтому это одна из наиболее чувствительных сфер уголовного процесса. В отечественной правовой традиции доказывание признается «сердцевинной уголовного судопроизводства» (Шейфер, 2022), «стержнем всей процессуальной деятельности» (Лупинская, 2023). Поэтому с точки зрения науки уголовного процесса важно определиться в том, какое место занимают, в том числе в перспективе, технологии искусственного интеллекта в доказывании по уголовному делу: стоит ли воспринимать их исключительно как инструмент, с помощью которого можно получить те или иные имеющие значение для уголовного дела сведения, которые могут быть облечены в форму доказательств, или же мы можем говорить о том, что меняется сама природа доказывания в уголовном процессе и оно из сферы человеческого познания трансформируется в сферу машинного познания, где человеку отводится роль лишь фиксатора его результатов в конкретном правоприменительном решении.

2. Искусственный интеллект в руках субъекта доказывания

Использование технологий искусственного интеллекта в целях получения значимой для дела доказательственной информации предполагает определенную обработку тех сведений и данных, которые содержит источник доказательства. Соответственно возникает риск искажения или видоизменения этой информации, что в конечном итоге может повлиять на достоверность самого доказательства (Stoykova, 2023). Субъект доказывания и участники процесса должны иметь возможность оценить, каким образом происходила указанная обработка информации, и удостовериться в ее результатах.

2.1. Проблемы сбора доказательств с использованием технологий искусственного интеллекта

Следует согласиться с основными проблемами, возникающими на этапе сбора доказательств с использованием технологий искусственного интеллекта, которые были сформулированы Эфтихией Бампасика ([Bampasika, 2021](#)):

1. Необъяснимость – сложность для понимания алгоритмов, используемых искусственным интеллектом, что ведет к невозможности проверить или оспорить такое доказательство. Действительно, если в процессе сбора доказательств применяется определенное программное обеспечение на основе искусственного интеллекта, например, система распознавания (идентификации) лиц или система восстановления некачественного изображения, то заложенный в основу этого продукта алгоритм должен быть прозрачен и доступен для изучения как субъектом доказывания, так и участниками процесса.

2. Дискриминация и предвзятость – информация, на основе которой искусственный интеллект принимает решение, не всегда является полной и свободной от предубеждений или искажений. Это значит, что тот набор данных, который был предоставлен, например, нейронной сети для «обучения», должен быть доступен участникам процесса доказывания, чтобы иметь возможность выявить любое искажение или предвзятость.

3. Недостаток ответственности – в основе работы искусственного интеллекта в любом случае лежит деятельность человека, которая в достаточной мере не регулируется законом. Деятельность по созданию и разработке технологий, с помощью которых возможен сбор доказательств или информации, на основании которой затем формируются доказательства по уголовному делу, должна быть не только регламентирована на законодательном уровне, но и отвечать базовым принципам уголовного процесса.

Обозначенные выше вопросы действительно вызывают тревогу, и поэтому любая технология искусственного интеллекта, применяемая в уголовном процессе в целях получения и сбора доказательств, должна обладать набором свойств, позволяющих в результате получить допустимое с точки зрения уголовно-процессуального закона доказательство.

2.2. Применение технологий искусственного интеллекта при проведении экспертиз

Важное место в этом аспекте занимает применение технологий искусственного интеллекта при проведении современных экспертиз. Например, использование в ходе экспертного исследования нейронных сетей, основанных на машинном обучении, должно находить отражение в исследовательской части заключения эксперта.

В качестве примера того, как в заключении эксперта может быть раскрыто использование технологий искусственного интеллекта, приведем статью Алессандро Маронэ и группы ученых-биологов, в которой очень подробно отражена методика проведенного исследования по определению давности пятен крови на основе колориметрического анализа, в том числе указано об использовании пяти разных подходов машинного обучения ([Marrone et al., 2021](#)). При этом исследователями описан каждый примененный подход машинного обучения и полученные с его помощью результаты. Вполне можно представить, что подобным образом должно быть описано применение

того или иного инструмента на основе искусственного интеллекта (например, нейронной сети) и при проведении биологической экспертизы. Это позволит оценить содержание выводов эксперта с точки зрения их достоверности и, как результат, даст возможность использовать эти выводы в доказывании по уголовному делу.

Другой пример, на котором можно рассмотреть использование в уголовном процессе технологий искусственного интеллекта как инструмента для получения и сбора доказательств, – это технология дипфейка. С ее развитием доступность и, соответственно, широта ее использования в криминальной сфере будут только увеличиваться. Мы легко можем представить себе использование дипфейка в банковском или страховом мошенничестве, а также в других подобных преступлениях. Поэтому уже сейчас необходимо подготовить научную базу для соответствующих экспертных исследований, которые, очевидно, невозможны без применения инструментов искусственного интеллекта. Если одна нейронная сеть способна сгенерировать поддельное видеоизображение, то в распоряжении эксперта должна быть не менее эффективная нейронная сеть, способная распознать такое видео. Так, симметричным ответом на развитие технологии дипфейка должна стать разработка новой методологии портретной и видеоскопической экспертизы, которая должна учитывать наиболее уязвимые стороны этой технологии, такие, например, как наличие или отсутствие отражения в глазах изображенных на видео персонажей. Выше мы уже упоминали, что МВД России пытается разработать соответствующее программное обеспечение в целях противодействия тем преступным действиям, которые применяют дипфейк. Однако, помимо технических средств, следует предусмотреть и соответствующее методическое обеспечение этой деятельности.

Как видим, сбор доказательств с помощью технологий искусственного интеллекта, при условии что эта деятельность отвечает требованиям открытости, проверяемости, объективности и обеспечена соответствующим механизмом ответственности, вполне отвечает принципу свободы оценки доказательств (ст. 17 Уголовно-процессуального кодекса Российской Федерации¹). Стоит отметить, что речь не идет в данном случае о появлении в уголовном процессе нового вида доказательства, как его иногда называют, «электронного доказательства». Мы вполне разделяем высказанную в науке точку зрения о том, что предусмотренные уголовно-процессуальным законом виды и формы доказательств не нуждаются в расширении (Головкин, 2019). Мы говорим только о регулировании нового способа получения и сбора доказательств, который некоторые исследователи обозначают как «получение цифровой информации машинным способом» (Александров, 2018). Полагаем, что доказательства, при получении которых применялись технологии искусственного интеллекта, будут обладать признаком допустимости и поэтому могут использоваться в доказывании по уголовному делу.

3. Перспективы трансформации процесса доказывания

Бурное развитие технологий искусственного интеллекта и возрастающая цифровизация судопроизводства толкают некоторых ученых к выводу о том, что искусственный интеллект может стать субъектом правоотношений и наделен правосубъектностью

¹ Уголовно-процессуальный кодекс Российской Федерации от 18 декабря 2001 г. № 174-ФЗ. *Собрание законодательства Российской Федерации*. 24 декабря 2001 г. № 52 (ч. I). Ст. 4921.

(Папышева, 2022) и даже вполне может заменить судью (Колоколов, 2020). Эта позиция, может, и является чрезмерно оптимистичной и даже в какой-то мере футуристичной, но все же она не лишена основания. Дело в том, что проблема приближения судебного решения к объективной истине и связанный с ней поиск эффективных средств и механизмов, которые могли бы исключить или свести к минимуму судебную ошибку, настолько занимает умы отечественных ученых уголовно-процессуального направления, что они готовы хвататься за любую, пусть даже и чисто гипотетическую возможность ее решить. И в этом случае инструменты искусственного интеллекта, которые, как это широко растиражировано, «позволяют решать некоторые задачи, которые не под силу человеку», «исключают влияние человеческого фактора при решении задачи», «ускоряют процесс принятия решения», естественно становятся довольно привлекательным объектом для построения гипотез о направлениях развития уголовного процесса.

Еще одной причиной для обращения взоров ученых к технологиям искусственного интеллекта является непрогнозируемость судебных решений, особенно когда это касается суда присяжных. По уголовно-процессуальному кодексу Российской Федерации присяжные заседатели не обязаны мотивировать свой вердикт, что не позволяет сторонам оценить обоснованность такого решения и понять, как и вокруг чего строились рассуждения коллегии присяжных. Из-за этого решения суда присяжных часто воспринимаются как произвольные, оторванные от процесса доказывания.

Кроме того, никто из участников процесса, в том числе профессиональный судья, не может повлиять на их решение и тем более не может принимать участие в их совещании. Оставленные один на один с вопросным листом, не имея специальной подготовки в доказывании по уголовному делу, но получив при этом наставления от профессионального судьи о необходимости толковать все разумные сомнения в пользу обвиняемого, присяжные вынуждены прибегать к своему жизненному опыту и тем навыкам размышлений и логических умозаключений, которые у них сложились за период жизни.

В результате в деле, где в качестве доказательств были представлены видеозаписи с автозаправки, зафиксировавшие момент вооруженного разбойного нападения на нее, и показания потерпевших, красочно описывающих эти события, присяжные могут прийти к выводу об отсутствии события преступления и оправдать подсудимых². Или в другом деле, где в качестве доказательств представлены выводы экспертизы о наличии следов ДНК подсудимого на месте происшествия, а также изъятые по месту жительства подсудимого личные вещи убитых потерпевших, присяжные могут прийти к выводу о непричастности подсудимого и также оправдать его³.

Конечно, если бы была возможность выяснить у самих присяжных, почему они пришли к указанным выше решениям, они, вероятно, объяснили бы, что представленных доказательств им просто оказалось недостаточно для вывода о доказанности

² Дело № 2-3/2020, рассмотренное Ставропольским краевым судом. https://kraevoy-stv.sudrf.ru/modules.php?name=sud_delo&srv_num=2&name_op=case&case_id=30809862&case_uid=bdd2dd35-a6f1-4f7b-a158-18a71eefa38f&delo_id=1540006

³ Дело № 2-4/2021, рассмотренное Ставропольским краевым судом. https://kraevoy-stv.sudrf.ru/modules.php?name=sud_delo&srv_num=2&name_op=doc&number=23537753&delo_id=1540006&new=0&text_number=1

обвинения. Это во многом прояснило бы связь между вердиктом и доказательствами и сняло бы вопрос о восприятии вердикта с точки зрения его обоснованности.

В этой ситуации, когда мотивы принятого субъектом доказывания решения для нас скрыты, возникает естественный вопрос: можно ли с помощью математических методов, которые лежат в основе любой технологии искусственного интеллекта, улучшить процесс доказывания по уголовному делу, чтобы его результатом было справедливое судебное решение, максимально приближенное к объективной истине? Что если каждому доказательству в уголовном деле задать определенный вес и научить нейронную сеть оценивать эти доказательства, строить на их основе определенные выводы?

В ответе на эти вопросы следует в первую очередь обратиться к сущности принципа свободной оценки доказательств, описанного в ст. 17 Уголовно-процессуального кодекса Российской Федерации, который заключается в том числе в отсутствии заранее установленной силы доказательств и запрете их какой-либо градации по качеству (Головко, 2017). Когда мы задаем алгоритму точные значения для каждого доказательства и предлагаем ему на основании этого сделать оценочный вывод, мы забываем о том, что уголовно-процессуальное доказывание хоть и является разновидностью познавательной деятельности, но при этом имеет принципиальные отличия от иных видов познания. Таким образом, формализация доказательств посредством присвоения им определенного веса, пусть даже и в целях работы математического алгоритма, идет вразрез с основополагающим принципом уголовного судопроизводства. Здесь следует согласиться с М. А. Малиной в негативной оценке подобного подхода к применению технологий искусственного интеллекта в доказывании по уголовному делу (Малина, 2021).

Есть ли альтернативные пути или следует однозначно и окончательно отказаться от идеи трансформации процесса доказывания посредством применения технологий искусственного интеллекта?

К счастью, альтернатива есть всегда. Так, например, один из ведущих мировых специалистов в доказательственном праве, уважаемый профессор Рональд Дж. Аллен, еще в 2001 г. убедительно показал, что применение к доказыванию в уголовном процессе таких математических методов определения вероятностей, как теорема Байеса, не работает, и предложил в качестве варианта теорию относительного правдоподобия (Allen, 2001). В более поздней работе, развивая эту мысль, он указывал, что доказательства будут более убедительными, когда они подтверждают вывод об одной гипотезе (например, этот человек совершил преступление) по сравнению с конкурирующей гипотезой (например, преступление совершил кто-то другой), и более слабыми, когда они не исключают правдоподобных альтернативных гипотез, которые основаны на альтернативных предположениях (Allen & Pardo, 2007).

Если посмотреть через призму этих идей на процесс доказывания в суде присяжных, то становится очевидным, что именно так и происходит: стороны пытаются убедить коллегию присяжных в том, что их версия обстоятельств дела является более правдоподобной. Как раз на оценку правдоподобности доказательств ориентирует судья коллегию присяжных в напутственном слове, когда просит их при анализе доказательств толковать в пользу подсудимого лишь разумные сомнения, т. е. те, которые можно объяснить, которые основаны на здравом смысле, а не на предвзятом мнении, предположениях, воображении, чувстве симпатии или антипатии, желании угодить общественному мнению, оправдать ожидания друзей, эмоциях и фантазиях.

Поэтому если и рассматривать возможность применения технологий искусственного интеллекта именно на этапе проверки и оценки доказательств, то только через алгоритмы, основанные на теории относительной правдоподобности. В противном случае мы упираемся в нарушение принципа свободы оценки доказательств.

Могут ли существующие технологии искусственного интеллекта выполнить работу по оценке доказательств на том же уровне, что и непрофессиональные судьи – присяжные? Очевидно, что нет. Такой математический метод пока не разработан. Поэтому на сегодняшний день говорить о субъектности искусственного интеллекта в процессе уголовно-процессуального доказывания преждевременно.

Но, глядя в будущее с надеждой усовершенствовать процесс доказывания, в том числе в суде присяжных, науке уголовного процесса в первую очередь нужно обратиться к помощи математиков. Только разработка математических методов анализа доказательств по уголовному делу позволит перейти от использования технологий искусственного интеллекта исключительно как инструмента в процессе получения и сбора доказательств к формированию нового процесса доказывания, в котором искусственный интеллект сможет играть уже познавательную роль. Насколько скоро это случится и случится ли вообще, покажет время.

Выводы

Сложившийся в уголовно-процессуальной науке оптимистичный подход к оценке возможностей применения технологий искусственного интеллекта в уголовном процессе значительно опережает существующие в настоящее время технологии, т. е. основан на ожиданиях и прогнозах. Реальное применение указанных технологий имеет место лишь в отдельных сферах уголовного процесса. Тем не менее продолжающееся развитие искусственного интеллекта позволяет говорить о неизбежном расширении сфер его применения, в том числе в процессе доказывания по уголовным делам.

Наиболее реалистичным сценарием применения технологий искусственного интеллекта в процессе доказывания является их использование при получении и сборе информации, имеющей доказательственное значение, которая в дальнейшем может быть оформлена в качестве доказательства по уголовному делу. Использование искусственного интеллекта как инструмента при сборе доказательств должно отвечать таким требованиям, как открытость, проверяемость, объективность и обеспеченность соответствующим механизмом ответственности. В этом случае доказательства, при получении которых применялись технологии искусственного интеллекта, будут обладать признаком допустимости.

Оценка доказательств с помощью математических алгоритмов, в которых применяются заранее установленные значения качества каждого доказательства, противоречит принципу свободы оценки доказательств в уголовном процессе. Применение основанных на математических алгоритмах технологий искусственного интеллекта в уголовно-процессуальном доказывании возможно лишь при условии соответствия указанному принципу, т. е., например, на основе теории относительной правдоподобности доказательств. Поэтому в настоящее время преждевременно говорить о том, что машинное познание может стать содержанием уголовно-процессуального доказывания.

Список литературы

- Александров, А. С. (2018). Проблемы теории уголовно-процессуального доказывания, которые надо решать в связи с переходом в эпоху цифровых технологий. *Судебная власть и уголовный процесс*, 2, 130–139. <https://elibrary.ru/XWCHZB>
- Головко, Л. В. (2017). *Курс уголовного процесса* (2-е изд., испр.). Москва: Статут, 2017.
- Головко, Л. В. (2019). Цифровизация в уголовном процессе: локальная оптимизация или глобальная революция? *Вестник экономической безопасности*, 1, 15–25.
- Колоколов, Н. А. (2020). Еще раз об искусственном интеллекте в правосудии. *Уголовное судопроизводство*, 4, 3–6. <https://elibrary.ru/OYLHDM>
- Кузьмин, Н. А. (2021). Перспективы использования искусственного интеллекта в противодействии коррупции. *Вестник Московского университета МВД России*, 3, 154–156.
- Лупинская, П. А. (2023). *Решения в уголовном судопроизводстве: теория, законодательство, практика: монография* (3-е изд., стереотип.). Москва.
- Макутчев, А. В. (2022). Современные возможности и пределы внедрения искусственного интеллекта в систему правосудия. *Актуальные проблемы российского права*, 8, 47–58. <https://doi.org/10.17803/1994-1471.2022.141.8.047-058>
- Малина, М. А. (2021). Цифровизация российского уголовного процесса: искусственный интеллект для следователя или вместо следователя. *Российский следователь*, 2, 29–32.
- Папышева, Е. С. (2022). Искусственный интеллект как угроза принципу презумпции невиновности. *Адвокатская практика*, 5, 2–5.
- Расторопов, С. В. (2020). Использование искусственного интеллекта для предупреждения и выявления преступлений (мировой опыт). *Международное публичное и частное право*, 5, 40–43. <https://doi.org/10.18572/1812-3910-2020-5-40-43>
- Сибилькова, А. В. (2019). Искусственный интеллект на службе у следователя. *Российский следователь*, 9(3), 68–70. <https://elibrary.ru/ZANFTV>
- Сретенцев, Д. Н., Волкова, Д. Р. (2021). Перспективы внедрения искусственного интеллекта в сферу расследования преступлений. *Российский следователь*, 11, 38–42. <https://elibrary.ru/HLNQDU>
- Сумин, А. А., Химичева, О. В. (2020). Искусственный интеллект в уголовном процессе государств Азиатско-Тихоокеанского региона: общий обзор. *Международное уголовное право и международная юстиция*, 2, 18–21. <https://elibrary.ru/RGILWR>
- Цветков, Ю. А. (2021). Искусственный интеллект в управлении следственными органами. *Российский следователь*, 9, 29–33.
- Цирин, А. М., Артеменко, Е. А. (2023). Цифровые технологии и искусственный интеллект как средства профилактики проявлений коррупции в контрольной (надзорной) деятельности: отечественный и зарубежный опыт. *Журнал российского права*, 3, 126–142.
- Шейфер, С. А. (2022). *Доказательства и доказывание по уголовным делам: проблемы теории и правового регулирования: монография* (2-е изд., испр. и доп.). Москва: Норма.
- Allen, Ronald J., & Pardo, M. S. (2007). The Problematic Value of Mathematical Models of Evidence. *Journal of Legal Studies*, 36, 107–108.
- Allen, Ronald J. (2001). Artificial intelligence and the evidentiary process: The challenges of formalism and computation. *Artificial Intelligence and Law*, 9, 99–114.
- Amariles, D. R., & Baquero, P. M. (2023). Promises and limits of law for a human-centric artificial intelligence. *Computer Law & Security Review*, 48, 105795. <https://doi.org/10.1016/j.clsr.2023.105795>
- Bampasika, E.-V. (2021). Artificial Intelligence as Evidence in Criminal Trial. In *CEUR Workshop Proceedings* (pp. 133–138).
- Cascavilla, G., Tamburri, D., & Van Den Heuvel, W.-J. (2021). Cybercrime threat intelligence: A systematic multi-vocal literature review. *Computers & Security*, 105, 102258. <https://doi.org/10.1016/j.cose.2021.102258>
- Feijóo, C., & Kwon, Y. (2020). AI impacts on economy and society: Latest developments, open issues and new policy measures. *Telecommunications Policy*, 44(6), 101987. <https://doi.org/10.1016/j.telpol.2020.101987>
- Kalai, T., Gnanaprakasam, C., Indumathy, M., Khilar, R., & Sathish Kumar, P. J. (2022). Artificial intelligence based optimization for mapping IP addresses to prevent cyber-based attacks. *Measurement: Sensors*, 24, 100508. <https://doi.org/10.1016/j.measen.2022.100508>
- Kaur, R., Gabrijelčič, D., & Klobučar, T. (2023). Artificial intelligence for cybersecurity: Literature review and future research directions. *Information Fusion*, 97, 101804. <https://doi.org/10.1016/j.inffus.2023.101804>

- Marrone, A., La Russa, D., Montesanto, A., Lagani, V., La Russa, M. F., & Pellegrino, D. (2021). Short and Long Time Bloodstains Age Determination by Colorimetric Analysis: A Pilot Study. *Molecules*, 26, 6272. <https://doi.org/10.3390/molecules26206272>
- Rahman, F. (2019). Introducing Artificial Intelligence in The Criminal Justice System with Special Reference to India, *NCU Law Review*, 2(1), 49–63.
- Rissland, Edwina L. (1990). Artificial Intelligence and Law: Stepping Stones to a Model of Legal Reasoning. *Yale Law Journal*, 99, 1957–1980.
- Robles Carrillo, M. (2020). Artificial intelligence: From ethics to law. *Telecommunications Policy*, 44(6), 101937, <https://doi.org/10.1016/j.telpol.2020.101937>
- Silva, I., Marcos do Valle, J., Souza, G., Budke, J., Araújo, D., Carvalho, D., Cacho, N., Sales, H., Lopes, F., & Silva Júnior, R. (2020). Using micro-services and artificial intelligence to analyze images in criminal evidences. *Forensic Science International: Digital Investigation*, 37, Supplement, 301197. <https://doi.org/10.1016/j.fsidi.2021.301197>
- Stoykova, R. (2023). The right to a fair trial as a conceptual framework for digital evidence rules in criminal investigations. *Computer Law & Security Review*, 49, 105801. <https://doi.org/10.1016/j.clsr.2023.105801>
- Vesnic-Alujevic, L., Nascimento, S., & Pólvera, A. (2020). Societal and ethical impacts of artificial intelligence: Critical notes on European policy frameworks. *Telecommunications Policy*, 44(6), 101961. <https://doi.org/10.1016/j.telpol.2020.101961>
- Wang, H., & Ma, S. (2022). Preventing crimes against public health with artificial intelligence and machine learning capabilities. *Socio-Economic Planning Sciences*, 80, 101043. <https://doi.org/10.1016/j.seps.2021.101043>
- Yassine, S., Esghir, M., & Ibrihich, O. (2023). Using Artificial Intelligence Tools in the Judicial Domain and the Evaluation of their Impact on the Prediction of Judgments. *Procedia Computer Science*, 220, 1021–1026. <https://doi.org/10.1016/j.procs.2023.03.142>

Сведения об авторе



Спиридонов Михаил Сергеевич – кандидат юридических наук, доцент кафедры уголовного процесса, криминалистики и судебной экспертизы, Южно-Уральский государственный университет (национальный исследовательский университет)

Адрес: 454080, Российская Федерация, г. Челябинск, пр. Ленина, 76

E-mail: spiridonovms@susu.ru

ORCID ID: <https://orcid.org/0009-0008-2715-8912>

Web of Science Researcher ID:

<https://www.webofscience.com/wos/author/record/AAK-9097-2021>

РИНЦ Author ID: https://elibrary.ru/author_items.asp?authorid=1089837

Конфликт интересов

Автор заявляет об отсутствии конфликта интересов.

Финансирование

Исследование не имело спонсорской поддержки.

Тематические рубрики

Рубрика OECD: 5.05 / Law

Рубрика ASJC: 3308 / Law

Рубрика WoS: OM / Law

Рубрика ГРНТИ: 10.79.35 / Доказательства

Специальность ВАК: 5.1.4 / Уголовно-правовые науки

История статьи

Дата поступления – 28 апреля 2023 г.

Дата одобрения после рецензирования – 12 мая 2023 г.

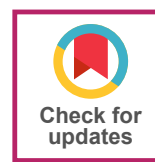
Дата принятия к опубликованию – 16 июня 2023 г.

Дата онлайн-размещения – 20 июня 2023 г.



Research article

DOI: <https://doi.org/10.21202/jdtl.2023.20>



Artificial Intelligence Technologies in Criminal Procedural Proving

Mikhail S. Spiridonov 

South Ural State University (National Research University)
Chelyabinsk, Russian Federation

Keywords

Artificial intelligence,
criminal case,
criminal procedure,
digital technologies
digitalization,
evidence,
judicial procedure,
law,
neural network,
proving

Abstract

Objective: to summarize and analyze the approaches, established in criminal procedural science, regarding the use of artificial intelligence technologies, to elaborate an author's approach to the prospects of transformation of criminal procedural proving under the influence of artificial intelligence technologies.

Methods: the methodological basis of the research is integrity of general, general scientific and specific legal methods of legal science, including abstract-logical, comparative-legal and prognostic methods.

Results: the main areas of using artificial intelligence technologies in the criminal procedure are defined, such as prophylaxis and detection of crimes, organization of preliminary investigation, criminological support of crime investigation, and assessing evidences at pre-trial and trial stages. The author comes to a conclusion that the rather optimistic approach to this issue, established in the science of criminal procedure, significantly outstrips the actually existing artificial intelligence technologies. The main requirements are identified, which the activity of using artificial intelligence in collecting evidences in a criminal case should satisfy. The author pays attention to the problems of using artificial intelligence technologies in conducting expert assessments, requiring an improved methodology of forensic work. The issue is considered of the prospects of transforming the criminal-procedural proving process under introduction of artificial intelligence technologies. A conclusion is substantiated that the assessment

© Spiridonov M. S., 2023



This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

of evidences with mathematical algorithms, in which preset values of each evidence quality are used, contradict to the principle of free assessment of evidences in the criminal procedure. The author comes to a conclusion that today there are no sufficient grounds for endowing artificial intelligence with legal personality during proving.

Scientific novelty: the work presents an attempt to consider the role of artificial intelligence in the criminal-procedural proving; it specifies the requirements to be met by this technology during evidences collection and analyzes the prospects of transforming the proving process under the introduction of artificial intelligence technologies.

Practical significance: the main provisions and conclusions of the research can be used to improve a mechanism of legal regulation of artificial intelligence technologies in the criminal procedure.

For citation

Spiridonov, M. S. (2023). Artificial Intelligence Technologies in Criminal Procedural Proving. *Journal of Digital Technologies and Law*, 1(2), 481–497. <https://doi.org/10.21202/jdtl.2023.20>

References

- Aleksandrov, A. S. (2018). The problems of the theory of criminal procedural proof, which must be solved in connection with the transition to the digital age. *Judicial authority and criminal process*, 2, 130–139. (In Russ.).
- Allen, Ronald J., & Pardo, M. S. (2007). The Problematic Value of Mathematical Models of Evidence. *Journal of Legal Studies*, 36, 107–108.
- Allen, Ronald J. (2001). Artificial intelligence and the evidentiary process: The challenges of formalism and computation. *Artificial Intelligence and Law*, 9, 99–114.
- Amariles, D. R., & Baquero, P. M. (2023). Promises and limits of law for a human-centric artificial intelligence. *Computer Law & Security Review*, 48, 105795. <https://doi.org/10.1016/j.clsr.2023.105795>
- Bampasika, E.-V. (2021). Artificial Intelligence as Evidence in Criminal Trial. In *CEUR Workshop Proceedings* (pp. 133–138).
- Cascavilla, G., Tamburri, D., & Van Den Heuvel, W.-J. (2021). Cybercrime threat intelligence: A systematic multi-vocal literature review. *Computers & Security*, 105, 102258. <https://doi.org/10.1016/j.cose.2021.102258>
- Feijóo, C., & Kwon, Y. (2020). AI impacts on economy and society: Latest developments, open issues and new policy measures. *Telecommunications Policy*, 44(6), 101987. <https://doi.org/10.1016/j.telpol.2020.101987>
- Golovko, L. V. (2017). *Course in criminal procedure* (2nd ed., amended). Moscow: Statut. (In Russ.).
- Golovko, L. V. (2019). The digitalization in criminal procedure: local optimization or global revolution? *Vestnik ekonomicheskoy bezopasnosti*, 1, 15–25. (In Russ.).
- Kalai, T., Gnanaprakasam, C., Indumathy, M., Khilar, R., & Sathish Kumar, P. J. (2022). Artificial intelligence based optimization for mapping IP addresses to prevent cyber-based attacks. *Measurement: Sensors*, 24, 100508. <https://doi.org/10.1016/j.measen.2022.100508>
- Kaur, R., Gabrijelčič, D., & Klobučar, T. (2023). Artificial intelligence for cybersecurity: Literature review and future research directions. *Information Fusion*, 97, 101804. <https://doi.org/10.1016/j.inffus.2023.101804>
- Kolokolov, N. A. (2020). Once again on artificial intelligence in justice. *Ugolovnoye sudoproizvodstvo*, 4, 3–6. (In Russ.).
- Kuzmin, N. A. (2021). Prospects for the use of artificial intelligence in combating corruption. *Vestnik Moskovskogo universiteta MVD Rossii*, 3, 154–156. (In Russ.).

- Lupinskaya, P. A. (2023). *Decisions in criminal judicial procedure: theory, legislation, practice*: monograph (3rd ed., stereotyped) Moscow. (In Russ.).
- Makutchev, A. V. (2022). Modern Possibilities and Limits of Artificial Intelligence Introduction into the Justice System. *Actual Problems of Russian Law*, 17(8), 47–58. (In Russ.). <https://doi.org/10.17803/1994-1471.2022.141.8.047-058>
- Malina, M. A. (2021). Digitalization of the Russian criminal procedure: artificial intelligence for an investigator or instead of an investigator. *Rossiiskiy sledovatel*, 2, 29–32. (In Russ.).
- Marrone, A., La Russa, D., Montesanto, A., Lagani, V., La Russa, M. F., & Pellegrino, D. (2021). Short and Long Time Bloodstains Age Determination by Colorimetric Analysis: A Pilot Study. *Molecules*, 26, 6272. <https://doi.org/10.3390/molecules26206272>
- Papysheva, E. S. (2022). Artificial intelligence as a threat to the principle of assumption of innocence. *Advocate's Practice*, 5, 2–5. (In Russ.).
- Rahman, F. (2019). Introducing Artificial Intelligence in the Criminal Justice System with Special Reference to India, *NCU Law Review*, 2(1), 49–63.
- Rastoropov, S. V. (2020). The use of artificial intelligence for crime prevention and identification (the world experience). *Public International and Private International Law*, 5, 40–43. (In Russ.). <https://doi.org/10.18572/1812-3910-2020-5-40-43>
- Rissland, Edwina L. (1990). Artificial Intelligence and Law: Stepping Stones to a Model of Legal Reasoning. *Yale Law Journal*, 99, 1957–1980.
- Robles Carrillo, M. (2020). Artificial intelligence: From ethics to law. *Telecommunications Policy*, 44(6), 101937, <https://doi.org/10.1016/j.telpol.2020.101937>
- Sheifer, S. A. (2022). *Evidence and proving in criminal cases: problems of theory and legal regulation*: monograph (2nd ed., amend. and compl.). Moscow: Norma. (In Russ.).
- Sibilkova, A. V. (2019). Artificial intelligence in the investigator's employ. *Rossiiskiy sledovatel*, 3, 68–70. (In Russ.).
- Silva, I., Marcos do Valle, J., Souza, G., Budke, J., Araújo, D., Carvalho, D., Cacho, N., Sales, H., Lopes, F., & Silva Júnior, R. (2020). Using micro-services and artificial intelligence to analyze images in criminal evidences. *Forensic Science International: Digital Investigation*, 37, Supplement, 301197. <https://doi.org/10.1016/j.fsidi.2021.301197>
- Sretentsev, D. N., & Volkova, V. R. (2021). Prospects for the introduction of artificial intelligence systems in crime investigation. *Rossiiskiy sledovatel*, 11, 38–42. (In Russ.).
- Stoykova, R. (2023). The right to a fair trial as a conceptual framework for digital evidence rules in criminal investigations. *Computer Law & Security Review*, 49, 105801. <https://doi.org/10.1016/j.clsr.2023.105801>
- Sumin, A. A., & Khimicheva, O. V. (2020). Artificial intelligence in the criminal procedure of the states of the Asia-Pacific region: a general overview. *Mezhdunarodnoye ugovnoye pravo i mezhdunarodnaya yustitsiya*, 2, 18–21. (In Russ.).
- Tsirin, A. M., & Artemenko, E. A. (2023). Digital Technologies and Artificial Intelligence as Measures of Preventing Corruption in Control (Supervisory) Activities: Domestic and Foreign Experience. *Journal of Russian Law*, 3, 126–142. (In Russ.).
- Tsvetkov, Yu. A. (2021). Artificial intelligence in management of investigative authorities. *Rossiiskiy sledovatel*, 9, 29–33. (In Russ.).
- Vesnic-Alujevic, L., Nascimento, S., & Pólora, A. (2020). Societal and ethical impacts of artificial intelligence: Critical notes on European policy frameworks. *Telecommunications Policy*, 44(6), 101961. <https://doi.org/10.1016/j.telpol.2020.101961>
- Wang, H., & Ma, S. (2022). Preventing crimes against public health with artificial intelligence and machine learning capabilities. *Socio-Economic Planning Sciences*, 80, 101043. <https://doi.org/10.1016/j.seps.2021.101043>
- Yassine, S., Esghir, M., & Ibrihich, O. (2023). Using Artificial Intelligence Tools in the Judicial Domain and the Evaluation of their Impact on the Prediction of Judgments. *Procedia Computer Science*, 220, 1021–1026. <https://doi.org/10.1016/j.procs.2023.03.142>

Author information



Mikhail S. Spiridonov – Candidate of Sciences in Jurisprudence, Associate Professor of the Department of Criminal Procedure, Criminology and Forensics, South Ural State University (National Research University)

Address: 76 prospekt Lenina, 454080 Chelyabinsk, Russian Federation

E-mail: spiridonovms@susu.ru

ORCID ID: <https://orcid.org/0009-0008-2715-8912>

Web of Science Researcher ID:

<https://www.webofscience.com/wos/author/record/AAK-9097-2021>

RSCI Author ID: https://elibrary.ru/author_items.asp?authorid=1089837

Conflict of interests

The author declare no conflict of interests.

Financial disclosure

The research was not sponsored.

Thematic rubrics

OECD: 5.05 / Law

PASJC: 3308 / Law

WoS: OM / Law

Article history

Date of receipt – April 28, 2023

Date of approval – May 12, 2023

Date of acceptance – June 16, 2023

Date of online placement – June 20, 2023

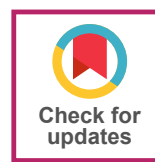


Научная статья

УДК 341.1/8:004.8

EDN: <https://elibrary.ru/cpffyw>

DOI: <https://doi.org/10.21202/jdtl.2023.21>



Рекомендации по этическим аспектам искусственного интеллекта в приложении к сфере трудовых отношений

София Рицу 

Университет Мишкольца
г. Мишкольц, Венгрия

Ключевые слова

Законодательство,
искусственный интеллект,
право,
принципы права,
труд,
трудовое право,
трудовые отношения,
цифровизация,
цифровые технологии,
этика

Аннотация

Цель: распространение и широкое использование искусственного интеллекта выдвигает на первый план не только проблемы защиты данных, но и этические вопросы. Цель данной статьи – изучение этических аспектов искусственного интеллекта и предложение рекомендаций для его использования в трудовом праве.

Методы: исследование основано на методах сравнительного и эмпирического анализа. Сравнительный анализ позволил изучить положения современного трудового законодательства в контексте искусственного интеллекта. Эмпирический анализ выявил этические проблемы, относящиеся к искусственному интеллекту в сфере труда, путем изучения спорных случаев использования искусственного интеллекта в различных областях, таких как здравоохранение, образование, транспорт и др.

Результаты: частноправовые аспекты этических проблем искусственного интеллекта были изучены в контексте этических и трудовых вопросов права, влияющих на процесс отбора с помощью искусственного интеллекта и на обращение с работниками с точки зрения работодателя. Автор выделяет как общие аспекты этики, так и вопросы цифровой этики. Предложены отдельные международные рекомендации относительно этики искусственного интеллекта.

Научная новизна: исследование посвящено изучению этических аспектов использования искусственного интеллекта в конкретной отрасли частного права – трудовом праве. Автор дает рекомендации относительно этических аспектов использования искусственного интеллекта в данной сфере.

© Рицу С., 2023



Статья находится в открытом доступе и распространяется в соответствии с лицензией Creative Commons «Attribution» («Атрибуция») 4.0 Всемирная (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0/deed.ru>), позволяющей неограниченно использовать, распространять и воспроизводить материал при условии, что оригинальная работа упомянута с соблюдением правил цитирования.

Практическая значимость: исследование восполняет имеющиеся пробелы в научной литературе по указанному вопросу. Результаты работы могут использоваться в процессе законотворчества и служить базой для дальнейших исследований.

Для цитирования

Рицу, С. (2023). Рекомендации по этическим аспектам искусственного интеллекта в приложении к сфере трудовых отношений. *Journal of Digital Technologies and Law*, 1(2), 498–519. <https://doi.org/10.21202/jdtl.2023.21>

Содержание

Введение

1. Этика – цифровая этика
2. Регулирование этических аспектов искусственного интеллекта
 - 2.1. ЮНЕСКО – Рекомендации по этике искусственного интеллекта
 - 2.2. Этические принципы Группы экспертов высокого уровня (HLEG), принятые Еврокомиссией
 - 2.3. Европейский парламент – Рамочная резолюция об этических аспектах искусственного интеллекта (ИИ), робототехники и связанных с ними технологий
 - 2.4. Другие рекомендации и руководства
3. Этические аспекты искусственного интеллекта в сфере труда (на основе доклада Исследовательской службы Европарламента по вопросам искусственного интеллекта)

Заключение

Список литературы

Введение

Не вызывает сомнений, что технологии искусственного интеллекта стремительно развиваются. На многих площадках эта тема обсуждается с точки зрения общественных отношений. Исследования показывают, что искусственный интеллект (далее – ИИ) воспринимается неоднозначно; с одной стороны, он сможет выполнять различные виды деятельности во многих областях, но его противники уже говорят о теориях заговора.

Правовое регулирование ИИ также активно исследуется, в основном с точки зрения нормативных вопросов (Cyman et al., 2021; Alikhademi et al., 2022). Учитывая, как тяжело сейчас угнаться за технологическим прогрессом, состояние правового регулирования выглядит следующим образом: по мере появления технологических достижений право пытается установить нормы в отношении отдельных вопросов. С одной стороны, это правильно, учитывая, что само законодательство основывается на эмпирическом знании; с другой – правовое регулирование требует внимательного подхода, так как необходимо учесть проблемы, которые еще даже не возникли, но которые в ряде случаев можно предвидеть. Научный и технологический прогресс, цифровая революция, искусственный интеллект, алгоритмизация ставят

перед правом непростые задачи, и мы наблюдаем пересмотр как доминирования правовых институтов, так и содержания отдельных правовых категорий (Harmathy, 2019). Развитию регулирования во многом способствует наличие общепринятых стандартов, однако решать эти вопросы, учитывая их специфику, необходимо также и на уровне государства.

Применение и распространение алгоритмизации и искусственного интеллекта представляет собой крайне интересную проблему для исследователя. Их применение связано с ускорением и повышением точности всех процессов. В недавнем прошлом появилось множество работ, посвященных различным аспектам возрастания рисков использования умных роботов. Исследователь László Z. Karvalics назвал этот подход алармизмом, и он является логическим следствием (и неизменным спутником) устоявшейся парадигмы «сильного искусственного интеллекта» и последних версий этой парадигмы (Karvalics, 2015).

Технологии непрерывно изменяются. Как и прежние индустриальные революции, нынешняя революция «Индустрия 4.0» (или даже «Индустрия 5.0») может приносить инновации каждый день, каждый час и каждую минуту. Темп развития возрастает экспоненциально, и право не может за ним угнаться. Насущной проблемой является защита человека и его прав как в цифровом, так и в реальном пространстве, ведь, как отмечал Stefán Ibolya, недостаточно эффективное правовое регулирование может стать препятствием для экономического развития (Stefán, 2020).

Различные аспекты искусственного интеллекта не только активно изучаются исследователями, но и сами играют значимую роль в развитии науки и решении глобальных проблем, от создания персональных помощников или беспилотных автомобилей до изучения космоса. Технологии стали частью нашей повседневной жизни, добавили нам удобств, расширив возможности человека. Обучающие алгоритмы уже выполняют многие задачи, от вождения автомобиля до принятия решений. Во многих случаях они принимают решения о нашей жизни, поэтому необходимы определенные гарантии. Помимо мер по защите данных, широкое распространение ИИ ставит также важные этические проблемы. Заявленная тема статьи становится краеугольным камнем исследований.

Недавние обсуждения выявили насущную необходимость работы над социальными и этическими аспектами алгоритмических систем и систем на основе данных, которые управляют нашей жизнью. Выступления ученых, журналистов и правоведов на тему машинного обучения и искусственного интеллекта затрагивают проблемы равенства, предубеждений, прозрачности, доступа, участия и дискриминации, что в совокупности часто называют проблемой ИИ и этики. Этот дискурс определяется стремлением снизить уровень дискриминации и предубеждений в области данных, а также нерешенными вопросами по поводу алгоритмических систем: могут ли они быть справедливыми, будет ли их использование способствовать достижению равноправия в будущем либо напротив, закрепит или даже усилит существующее неравенство. Несмотря на очевидную новизну этих вопросов, многие лежащие в их основе проблемы издавна обсуждались и имеют богатую традицию в сфере библиотековедения и науки об информации (library and information science, LIS) (Hoffmann et al., 2019).

Темпы развития технологических инноваций и их распространения по всему миру при поддержке цифровой экономики явно превышают темпы развития

человеческого сознания. Jerome Beranger¹ попытался перечислить и определить основные этические критерии, формирующие базис будущего мировоззрения, призванного двигаться в сторону искусственного интеллекта на службе человеческого разума, а также предвидеть все отклонения и возможные последствия развития алгоритмических систем (Beranger, 2021).

Объемы и сложность данных и алгоритмов, используемых в системах на основе искусственного интеллекта, также представляют значительные проблемы при предвидении их этических, правовых и политических последствий. Исследование в области этики искусственного интеллекта, основанное на опросах заинтересованных лиц, данных права и политики, показывает, что этика ИИ структурируется личными ценностями и профессиональными целями людей (Fleischmann et al., 2023).

В первой части статьи рассмотрены общие аспекты этики и выделены вопросы цифровой этики. Затем представлены отдельные рекомендации международных организаций в области этики ИИ. В данной работе мы более подробно обсудим частноправовые аспекты этики ИИ (включая трудовое право). В рамках этой темы мы рассмотрим вопросы этики и трудового права, влияющие на процессы выбора искусственного интеллекта и отношений с сотрудниками с точки зрения работодателя.

1. Этика – цифровая этика

Слово «этика» происходит от греческого *ethos*, что значит «привычка, форма поведения». Предметом ведения этики является деятельность человека и проявления личности в действиях. Термин часто используется как синоним слова «мораль», обозначающего формы поведения и деятельность относительно целей человека (Turay, 2000). Этика содержит стандарты нравственности, существующие подсознательно, ценности, лежащие в основе наших действий (Müller & Kerényi, 2019). Этика – это доктрина моральных ценностей, которую следует отличать от этикета, т. е. науки о традициях, манерах, внешних приличиях – о человеческом поведении (Legeza, 2013).

Этика и прикладная этика изучают практические проблемы и повседневные ситуации. Их область исследования – низкие и высокие проявления нравственных феноменов, общих категорий нравственности; так, П. Фобель ссылается на определение Kinsky и отмечает, что центральной темой на XX Всемирном философском конгрессе была прикладная этика, в рамках которой рассматривались проблемы биоэтики, этики охраны здоровья, окружающей среды, а также деловой этики, этики спорта, этики в области технологий и права (Fobel, 2002).

В области этики технологий определяющим фактором является цифровая этика, возникновение которой относится к периоду развития Интернета. Этическое поведение столь же важно во Всемирной паутине, доступной каждому, сколь и вне цифрового пространства. Сегодня каждый может создавать цифровой контент, размещать комментарии, выражать свое мнение по различным вопросам. Этими действиями мы выражаем свои ценностные установки.

¹ Научный эксперт в области этического подхода к цифровой революции, сооснователь и руководитель ADELIAA, научный сотрудник коллектива по биоэтике Inserm 1295 BIOETHICS Университета Тулузы.

Цифровая этика, или, в более широком смысле, информационная этика, изучает влияние цифровых информационных и коммуникационных технологий на общество и окружающую среду в целом. В более узком смысле информационная этика (или этика цифровых СМИ) изучает этические аспекты информации, содержащейся в Интернете, и средств коммуникации, основанных на Интернете, таких как мобильные телефоны и средства навигации (Capurro, 2018).

Появление цифровой этики также породило сложные правовые проблемы. Цифровая коммуникация, алгоритмизация и развитие искусственного интеллекта оказывают значительное влияние на жизни людей. По мнению Dániel Eszteri, на первый план могут выйти такие новые этические проблемы, как автоматическое принятие решений без вмешательства человека и права искусственных образований (Eszteri, 2021). Как отмечает Gabriella Németh, появление машинных субъектов может привести к переосмыслению самой сущности человека или ценностей человеческого существования (Németh, 2021). Этические вопросы искусственного интеллекта в военной сфере исследовались Imre Négyesi (Négyesi, 2020). В работе Réka Pusztahelyi предложена классификация этических принципов искусственного интеллекта с различных точек зрения в зависимости от их направленности. В частности, выделены частные, отраслевые и всеобщие этические принципы (Pusztahelyi, 2019). В нашей работе мы придерживаемся этой классификации. Действительно, этические принципы искусственного интеллекта, применяемые в медицине, не обязательно должны покрывать таковые этические концепции рынка труда.

Нельзя сказать, что системам искусственного интеллекта изначально присущи несправедливость и неравенство, однако проблема взаимодействия таких систем и нашего несправедливого мира вызывает дискуссии. Как отмечал Knowles, одним из механизмов решения проблемы этики ИИ является этическое обучение ИИ: все, кто принимает участие в обучении, использовании и/или предоставлении данных системам искусственного интеллекта, должны руководствоваться принципами четкого морального обоснования, ответственного выбора и правильных действий (Knowles, 2021).

Первым шагом к созданию этических систем искусственного интеллекта является изучение этических дилемм. Сообщество исследователей и специалистов в области искусственного интеллекта предпочитает использовать широкие нормы регулирования искусственного интеллекта, а не узкоспециализированные стандарты (Yu et al., 2018). Подобные нормы были созданы ЮНЕСКО, Европарламентом и Еврокомиссией; эти рекомендации и резолюции мы рассмотрим в следующей главе.

2. Регулирование этических аспектов искусственного интеллекта

Появление искусственного интеллекта вызвало множество вопросов законодательного регулирования. Они относятся не только к правовой концептуализации и категоризации, но и к отдельной отрасли права – этическим аспектам искусственного интеллекта и связанной с ними правовой ответственности.

Этические риски искусственного интеллекта значительны. Об этом свидетельствует исследование Deloitte 2018 г., согласно которому 32 % менеджеров оценивали этические риски искусственного интеллекта так высоко, что готовы были остановить

свои инициативы в этой области². По результатам опроса, проведенного в 2019 г. исследовательским институтом Capgemini Research Institute, почти половина пользователей искусственного интеллекта испытывали в этой связи этические проблемы, а 86 % менеджеров отметили, что им известны случаи, когда использование искусственного интеллекта привело к этическим проблемам (Capgemini Research Institute, 2019)³. Эти исследования также показывают, что недостаточно обозначить правовые принципы искусственного интеллекта, следует также обратить внимание на его этические аспекты.

Хотя терминология в данной области может меняться, сущность этики искусственного интеллекта постепенно проясняется. Принципы ОЭСР направлены на создание надежного искусственного интеллекта; подобные положения были сформулированы Объединенным центром искусственного интеллекта (United Center for Artificial Intelligence) (в качестве важнейших принципов были указаны отслеживаемость, надежность и управляемость). Группа экспертов высокого уровня Евросоюза сосредоточивается на принципах защиты данных, управления данными, прозрачности, недискриминации и справедливости (Tilesch & Hatamleh, 2021).

2.1. ЮНЕСКО – Рекомендации по этике искусственного интеллекта

Искусственный интеллект присутствует в жизни миллиардов людей, даже без их ведома, незаметно трансформируя общество. Его применение несет множество преимуществ, от обучения до поиска работы. Однако наряду с преимуществами он порождает также риски и проблемы. Искусственный интеллект несет в себе значительные социальные и культурные последствия, ставя вопросы о свободе слова и самовыражения, о праве на тайну частной жизни, об имущественных правах, дискриминации, манипуляции и искажении информации. Вдобавок искусственный интеллект бросает вызов когнитивным способностям человека и аспектам взаимодействия. Алгоритмы могут способствовать распространению дезинформации, влиять на политические и идеологические взгляды. Процессы глубокого обучения могут усиливать существующие предубеждения, что противоречит требованиям равноправия, подчеркивает неравенство между отдельными социальными слоями и группами, усиливает цифровые барьеры, а значит, и шансы на появление цифрового разрыва; все это отмечено в предварительном докладе ЮНЕСКО по проблеме искусственного интеллекта⁴.

Чтобы преодолеть эти риски и вызовы, необходимо установить как международные, так и национальные законодательные нормы, основанные на принципах ЮНЕСКО. В этой связи в ноябре 2021 г. на Генеральной конференции ЮНЕСКО были приняты рекомендации по искусственному интеллекту. Ранее, на 40-й сессии

² Loucks, J., Davenport, Th. H., & Schatsky, D. (2018, October 22). State of AI in the Enterprise, 2nd Edition. Early adopters combine bullish enthusiasm with strategic investments. Deloitte. <https://www2.deloitte.com/us/en/insights/focus/cognitive-technologies/state-of-ai-and-intelligent-automation-in-business-survey-2018.html>

³ Почему решение этических вопросов в области искусственного интеллекта принесет пользу организации [Why addressing ethical questions in AI will benefit organization]. <https://www.expertbibliotheek.nl/publicaties/data>

⁴ Рекомендации по этике искусственного интеллекта. <https://www.unesco.org/en/legal-affairs/recommendation-ethics-artificial-intelligence>

в ноябре 2019 г., было принято решение о создании международного нормативного документа по этике искусственного интеллекта. Соответственно, специальная группа экспертов провела междисциплинарные консультации с целью всестороннего внедрения указанных рекомендаций⁵.

Рекомендации ЮНЕСКО относятся к этическим аспектам в области искусственного интеллекта с точки зрения нормативного регулирования. В них этика определяется как база для оценки технологий искусственного интеллекта. Цель рекомендаций не в том, чтобы дать определение искусственному интеллекту, но в том, чтобы раскрыть важнейшие вопросы с точки зрения этики⁶. Также ставится цель создать международно признанный нормативный инструмент, отражающий такие ценности, как права, свободы и достоинство человека, защита окружающей среды, разнообразие, инклюзивность, формирование мирных и равноправных сообществ. Указанный инструмент также должен формулировать основные принципы и конкретные рекомендации для их достижения⁷, в первую очередь учитывая то, что государства-участники должны будут вводить свои правила для оценки этических аспектов, проводить оценку рисков, осуществлять надзор и создавать механизмы обеспечения безопасности⁸. Помимо этических вопросов, в рекомендациях выделены также проблемы защиты данных, международного сотрудничества и развития, охраны окружающей среды и гендерного равенства.

2.2. Этические принципы Группы экспертов высокого уровня (HLEG), принятые Еврокомиссией

С момента своего возникновения Европейский союз неоднократно сталкивался с социальными и экологическими проблемами, круг которых увеличивался экспоненциально с увеличением числа стран-участниц. Признавая важность устойчивого развития и климатических изменений, Евросоюз поддержал несколько международных инициатив, например, Конвенцию по климату⁹ и Цели в области устойчивого развития под эгидой ООН¹⁰. ЕС также осуществляет деятельность в соответствии с принципами устойчивого развития; в рамках этой деятельности Комиссия Евросоюза основала Группу экспертов высокого уровня (HLEG, далее – ГЭВУ) для изучения этических аспектов искусственного интеллекта¹¹.

Работа ГЭВУ многопланова, она затрагивает все сферы приложения искусственного интеллекта и определяет не только развитие, но и моральные стороны его применения (Pusztahelyi, 2019). Рекомендации экспертной группы были использованы в таких инициативах Еврокомиссии, как «Укрепление доверия к человеко-ориентированному

⁵ Рекомендации по этике искусственного интеллекта. <https://www.unesco.org/en/legal-affairs/recommendation-ethics-artificial-intelligence>

⁶ Там же.

⁷ Там же.

⁸ Там же.

⁹ Парижское соглашение. <https://www.un.org/en/climatechange/paris-agreement>

¹⁰ Цели в области устойчивого развития. <https://sdgs.un.org/goals>

¹¹ Группа экспертов высокого уровня (HLEG). <https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai>

искусственному интеллекту»¹², «Белая книга по искусственному интеллекту: европейский подход к совершенству и доверию»¹³ и обновленный Координированный план по искусственному интеллекту¹⁴.

В сводном отчете ГЭВУ отметила, что в отношении искусственного интеллекта должны выполняться три основных условия: законность, этичность, устойчивость. Были установлены принципы применения надежного искусственного интеллекта, обеспечивающие этичность и устойчивость искусственного интеллекта. Что касается законности, то, как отмечает Réka Pusztahelyi, она начинается с того, что применение ИИ имеет место в рамках стандартов (Pusztahelyi, 2019).

В первой главе «Руководящих принципов по этике для надежного искусственного интеллекта» отражены базовые права и заложены основы понимания искусственного интеллекта, затем рассматриваются задачи этики¹⁵ и в то же время отмечается, что этический кодекс в данной области не может заменить этической аргументации. Как и документ ЮНЕСКО, отчет ГЭВУ базируется на фундаментальных правах (ценностях): достоинство, свобода, справедливость, демократия, равенство, верховенство права, отсутствие дискриминации, которые выступают основаниями правомочности по закону и нравственности. Кроме того, документ закладывает четыре базовых принципа: уважение к автономии человека, предотвращение вреда, справедливость и объяснимость; эти принципы выходят за рамки требований закона, но их соблюдение ставит серьезную дилемму перед разработчиками программного обеспечения (Pusztahelyi, 2019). Помимо общих принципов, обозначены семь требований, которые должны обеспечить всестороннее соблюдение этических принципов. Эти требования следующие: участие и надзор человека (включая основные права, участие человека и надзор со стороны человека), техническая стабильность и безопасность (устойчивость к атакам и безопасность, точность, надежность и воспроизводимость), защита данных и управление данными (уважение к частной жизни, качество и целостность данных, доступ к данным), прозрачность (включая отслеживаемость, объяснимость), разнообразие, отсутствие дискриминации, справедливость (недопустимость несправедливого предубеждения, доступность и универсальный дизайн, участие заинтересованных сторон), социальное и экологическое благополучие (выделяя устойчивость и защиту окружающей среды в качестве приоритетов) и подотчетность (включая проведение аудита, минимизацию негативных последствий, возможности для компромиссов и законных мер защиты)¹⁶. В документе подчеркивается, что эти требования одинаково важны, а их применение необходимо на всех этапах жизненного цикла приложений с ИИ, однако некоторые из этих требований содержатся в уже существующих правовых нормах.

¹² Укрепление доверия к человекоориентированному искусственному интеллекту. (2019). <https://digital-strategy.ec.europa.eu/en/library/communication-building-trust-human-centric-artificial-intelligence>

¹³ Белая книга по искусственному интеллекту: европейский подход к совершенству и доверию. (2020). https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

¹⁴ Координированный план по искусственному интеллекту, версия 2021 года. (2021). <https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review>

¹⁵ Руководящие принципы по этике для надежного искусственного интеллекта. (2021). <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/1.html>

¹⁶ Там же.

2.3. Европейский парламент – Рамочная резолюция об этических аспектах искусственного интеллекта, робототехники и связанных с ними технологий

Помимо рекомендаций ЮНЕСКО и руководства ГЭВУ, важно рассмотреть рамочные документы Европейского парламента. Основными принципами этих предложений являются поддержка использования искусственного интеллекта, робототехники и других связанных с ними технологий и приведение их в соответствие этическим принципам. Цель данного проекта – создать регулирование на основе этических принципов, направленное на развитие, внедрение и использование искусственного интеллекта, робототехники и связанных с ними технологий¹⁷. Задачей данного регулирования является установление всеобъемлющей долгосрочной законодательной базы Евросоюза для осуществления этических принципов и правовых обязательств.

Ответственный за данный проект Ибан Гарсия дель Бланко подчеркнул преимущества и риски искусственного интеллекта. По его словам, поставленная цель – достичь построения более устойчивого и справедливого общества; при этом следует уделять больше внимания обеспечению защиты частной жизни и искоренению дискриминации. Важно выработать нормативную базу на основе фундаментальных этических принципов. При разработке человеко-ориентированного ИИ необходимо обеспечить не только надежность, но и безопасность. Соответственно, документ создавался с учетом таких принципов, как оценка высокорискового ИИ, робототехники и связанных с ними технологий, обеспечение безопасности, прозрачности, подотчетности, гарантий отсутствия дискриминации, наличия законных мер защиты и права на защиту, социальная ответственность, устойчивость технологий искусственного интеллекта, соблюдение права на частную жизнь, ограничения биометрической идентификации, обеспечение необходимого контроля над использованием данных и формированием данных с помощью технологий. Внутри Евросоюза регулирование распространяется на любые технологии искусственного интеллекта, робототехники и связанных с ними технологий, включая программное обеспечение, алгоритмы и данные, используемые или создаваемые посредством таких технологий, в соответствии с законодательством ЕС, принципами человеческого достоинства, автономии и безопасности, а также в соответствии с другими фундаментальными правами, закрепленными в Уставе, и должны развиваться, применяться и использоваться должным образом¹⁸. Резолюция была принята 20 октября 2020 г.¹⁹, были представлены предложения. Правовым основанием для представления предложений явилась ст. 114 Договора о функционировании Европейского союза, которая устанавливает меры по обеспечению создания и функционирования единого внутреннего рынка. Проект стал одним из центральных элементов стратегии единого цифрового рынка. Он основывается на уже существующих законодательных нормах, является сбалансированным и необходимым для достижения поставленных целей. В проекте

¹⁷ Рамочная резолюция об этических аспектах искусственного интеллекта (ИИ), робототехники и связанных с ними технологий. (2020/2012(INL), 2020). <https://oeil.secure.europarl.europa.eu/oeil/popups/summary.do?id=1636985&t=d&l=en>

¹⁸ Там же.

¹⁹ Там же.

используется подход на основе анализа риска; бремя регулирования налагается лишь в том случае, если существует высокая вероятность угрозы фундаментальным правам или безопасности со стороны системы искусственного интеллекта²⁰.

Таким образом, мы видим, что законодательная база Европарламента строится по тем же этическим принципам, что и рекомендации ЮНЕСКО, и руководство ГЭВУ.

2.4. Другие рекомендации и руководства

Вышеописанные документы содержат общие положения и рекомендации, обладающие схожей логической структурой и подчеркивающие необходимость обеспечения практически одних и тех же этических ценностей. В данном разделе мы кратко рассмотрим публикации, не направленные конкретно на установление стандартов, но описывающие этические принципы, характерные для указанной сферы деятельности.

Международная организация «Институт инженеров электротехники и электроники» (далее – IEEE) в 2019 г. выпустила руководство по этике искусственного интеллекта под названием «Этически согласованный дизайн» (Ethically Aligned Design, EAD). Цель Глобальной инициативы IEEE – выработать прагматические и целеполагающие оценки и рекомендации, на которые могли бы опираться в своей деятельности разработчики технологий, законодатели и ученые (Pusztahelyi, 2019). В 2021 г. организация представила стандарт по этике IEEE7000, направленный на развитие этических и справедливых систем искусственного интеллекта. Этот стандарт в настоящее время изучается Еврокомиссией. Его положения выходят за рамки проблем защиты данных, прозрачности и надежности, так как содержат также анализ социальных последствий использования технологий и проблемы изменения личности под влиянием технологий. Стандарт предлагает разработчикам четкую систему принципов дизайна и развития. При этом используются различные теории этики, позволяющие выделить релевантные ценности и определить их приоритетность по классификации корпоративных или отраслевых ценностей. Введено также новое понятие «требование этической ценности» (ethical value requirement, EVR), входящее в системные требования, которые устанавливаются исходя из оценки рисков (Spiekermann et al., 2022).

Кроме того, при Министерстве обороны США был создан Объединенный центр искусственного интеллекта, начавший работу в 2019 г. Центр разрабатывает руководства в области искусственного интеллекта в основном в военной сфере, однако его деятельность представляется важной с этической точки зрения, учитывая, что ключевым приоритетом систем искусственного интеллекта является обеспечение соответствия правовым нормам с целью снижения опасности для гражданских лиц, и эти приоритеты актуальны с момента установления требований до окончания этапа тестирования²¹.

²⁰ Проект Постановления Европейского парламента и Европейского совета о введении согласованных норм в области искусственного интеллекта (Закон об искусственном интеллекте) и внесении изменений в некоторые законодательные акты Евросоюза. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>

²¹ Объединенный центр искусственного интеллекта. <https://www.defense.gov/News/News-Stories/Article/Article/2418970/joint-artificial-intelligence-center-has-substantially-grown-to-aid-the-warfigh/>

Этические вопросы в области искусственного интеллекта рассматриваются не только международными или национальными организациями. Так, компания IBM сочла необходимым внедрить этические принципы в свои процессы разработки и развития²²; компания Adobe также сделала это²³.

В России Комиссия по реализации Кодекса этики в сфере ИИ определила главным приоритетом развития технологий искусственного интеллекта защиту интересов и прав людей; это не только позволяет решить задачу ответственности и возможных последствий, но и обеспечивает использование ИИ для заявленных целей и там, где это принесет пользу людям²⁴.

3. Этические аспекты искусственного интеллекта в сфере труда (на основе доклада Исследовательской службы Европарламента по вопросам искусственного интеллекта)

Ожидается, что искусственный интеллект и смежные технологии дадут огромные экономические и социальные преимущества во всех областях, окажут влияние на финансовый сектор, здравоохранение, сельское хозяйство. Использование искусственного интеллекта принесет практическую пользу в плане улучшения прогнозируемости и оптимизации деятельности. Однако нельзя забывать, что последствия применения систем искусственного интеллекта могут угрожать фундаментальным правам человека, закрепленным в Хартии Европейского союза по правам человека, таким как равноправие, отсутствие дискриминации, человеческое достоинство, защита персональных данных и частной жизни²⁵.

Системы искусственного интеллекта способны обрабатывать данные точнее и быстрее, чем человек, однако их использование несет в себе значительные риски, если искусственный интеллект принимает решения без вмешательства человека. В случаях, когда системы могут избежать контроля со стороны человека или действовать с минимальным вмешательством человека, возникает множество этических вопросов; поэтому, по нашему мнению, такие процессы следует тщательно контролировать, пока не будет достигнут определенный уровень тестирования и гарантий безопасности.

В области здравоохранения искусственный интеллект может во многих случаях применяться для решения задач диагностики. Машинное обучение осуществляется на базе образцов, получаемых из баз данных, однако данные могут быть искажены внешними влияниями, а значит, и полученный результат может быть неверным. При обсуждении этики искусственного интеллекта в области радиологии на презентации совместной европейской и североамериканской декларации по стандарту IEEE подчеркивалось, что необходим приоритет человеческого фактора, т. е. результаты

²² IBM: повседневная этика для искусственного интеллекта. <https://www.ibm.com/design/ai/ethics/everyday-ethics/>

²³ Этические принципы в области искусственного интеллекта компании Adobe. <https://www.adobe.com/about-adobe/aiethics.html>

²⁴ Российский Альянс в сфере искусственного интеллекта – Этический кодекс искусственного интеллекта. <https://ethics.a-ai.ru/>

²⁵ Исследовательская служба Европарламента: Закон об искусственном интеллекте. (2021). [https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI\(2021\)698792_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf)

работы искусственного интеллекта должны занимать подчиненное положение относительно результатов суждений, надзора и контроля человека²⁶.

Наряду со здравоохранением, образованием, транспортом, энергетикой, финансовым и банковским сектором, сфера трудовых отношений также классифицируется Европарламентом как высокорисковая с точки зрения искусственного интеллекта. Системы искусственного интеллекта все чаще используются в процессе принятия решений в сфере занятости. Учет этических аспектов также важен при практическом применении искусственного интеллекта, особенно в отношении уязвимых или социально незащищенных групп и лиц, а также в случаях неравноправных позиций сторон, будь то отношения покупателя и продавца или работника и работодателя. В последнем случае (который соответствует теме нашего исследования) это отношения старшинства и подчинения, что свидетельствует о необходимости применять этические принципы.

В связи с этим в 2021 г. Еврокомиссия представила проект правового регулирования искусственного интеллекта. Общая цель проекта – создать необходимые условия для развития надежных систем искусственного интеллекта, а для этого определить согласованную законодательную базу. Кроме того, были поставлены задачи, обеспечивающие соответствие систем искусственного интеллекта европейскому законодательству и правовой определенности и способствующие формированию и поддержанию единого рынка. Новая законодательная база призвана также утвердить технологически нейтральное определение систем искусственного интеллекта. В проекте тщательно анализируется система рисков при использовании искусственного интеллекта. Подход, основанный на анализе рисков, ранжирует системы искусственного интеллекта по степени риска: от неприемлемого до высокого и ограниченного. Тем самым закладываются требования к высокорисковым системам, а именно: требование предварительной оценки соответствия, регистрация провайдера услуг в общей базе данных ЕС, тестирование, техническая надежность, прозрачность, надзор человека, кибербезопасность. Кроме того, поставлен вопрос о соответствии проекта стандартам ЕС, которые разрабатываются в настоящее время²⁷.

В Приложении III к данному проекту выделена группа систем искусственного интеллекта высокого класса, включающая, кроме систем биометрической идентификации и систем для оценивания обучающихся, также системы оценки уровня занятости и самозанятости. В связи с этим дистрибьюторы и провайдеры услуг в области систем искусственного интеллекта, используемых для рекламы и отбора вакансий, оценки на основе собеседований, отбора и найма сотрудников, обязаны выполнить ряд требований, прежде чем запускать эти системы, а весь процесс подлежит строгому регулированию и надзору²⁸.

²⁶ Этические вопросы использования искусственного интеллекта в радиологии. <https://www.neuroct.hu/blog/a-mesterseges-intelligencia-alkalmazasanak-etikai-kerdesei-a-radiologiaban>

²⁷ Гармонизация искусственного интеллекта: роль стандартов в Регламенте ЕС по ИИ. <https://montrealethics.ai/harmonizing-artificial-intelligence-the-role-of-standards-in-the-eu-ai-regulation/>

²⁸ Проект Постановления Европейского парламента и Европейского совета о введении согласованных норм в области искусственного интеллекта (Закон об искусственном интеллекте) и внесении изменений в некоторые законодательные акты Евросоюза. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>

В рамках отношений найма мы уже сейчас можем столкнуться с работой искусственного интеллекта в области отбора и найма на первом этапе – часто искусственный интеллект без нашего ведома выбирает наше резюме из всего множества полученных анкет. Такое использование искусственного интеллекта значительно повышает эффективность отбора кандидатов. Однако, по нашему мнению, это приложение не может заменить работу специалиста по кадрам; эту деятельность следует расценивать как дополнительную и вспомогательную, тогда как искусственный интеллект или робот не способен воспроизвести человеческое взаимодействие в контексте личного собеседования.

В процессе отбора искусственный интеллект обеспечивает справедливые условия, способствует снижению риска предвзятости и предубеждений – кандидаты отбираются исключительно на основании их профессионального опыта и навыков. В то же время это преимущество может породить этические риски в аспекте восприимчивости организации. Ярким примером такой ситуации является робот-кадровик компании Amazon Inc., которого обучали на резюме, поданных в компанию в течение 10 лет. Поскольку большинство резюме на позиции технических специалистов были от мужчин, искусственный интеллект «научился» отдавать предпочтение кандидатам-мужчинам. Робот автоматически отклонял анкеты женщин, тем самым нарушая гендерное равенство как этический принцип (Dastin, 2018).

В связи с этим возникает вопрос: насколько этично для работодателя воспринимать сотрудников как элементы базы данных? Как правило, базы данных можно смоделировать, однако в отсутствие контекста ценность таких данных снижена или полностью равна нулю. Например, в работе Boyd и Crawford описано графическое моделирование сети личных взаимоотношений на основе социальных сетей, позволяющее получить набор данных, но не полную и точную информацию о реальных отношениях между людьми (Boyd & Crawford, 2012). При отборе сотрудников мы можем столкнуться с подобной этической дилеммой: может ли искусственный интеллект определить пригодность кандидата? Полученные резюме оторваны от контекста, это лишь обезличенные наборы данных, из которых можно сделать определенные выводы, но на которые, по нашему мнению, нельзя полностью положиться.

Подобная дилемма возникает также при использовании искусственного интеллекта для анализа возможности сокращения рабочей силы. В Приложении III указанного проекта программное обеспечение ИИ, предназначенное для инициирования или прекращения контрактных отношений и принятия соответствующих решений, также отнесено к высокорисковым системам. Таковы же и системы для наблюдения и оценки эффективности деятельности и поведения людей в рамках правовых отношений (AI Act Proposal). Нет сомнений, что искусственный интеллект принимает справедливые решения на основе оценки деятельности, но этично ли будет уволить сотрудника, который в данный момент хуже работает, в одиночку воспитывая детей, только потому, что искусственный интеллект поместил его в более низкую категорию эффективности? Польза искусственного интеллекта в определенных процессах бесспорна, однако мы против того, чтобы расценивать сотрудников как элементы базы данных. Мы считаем, что при принятии решений следует учитывать также другие факторы, такие как возможности дальнейшего обучения и переобучения сотрудника, его гибкость и, разумеется, положение семьи или принадлежность к социально незащищенным группам. В этой связи при разработке законодательных норм страны – члена ЕС или конкретной организации, помимо стремления к инновационности,

мы должны учитывать содержание отношений между машиной и человеком, отношений старшинства и подчинения и (в соответствии с проектом законодательства ЕС) обеспечить прозрачность применения искусственного интеллекта и возможность пересмотра принятого решения человеком. Очевидно, что машины все сильнее вовлекаются в процессы принятия этических решений, которые требуют этических обоснований. Существующие алгоритмы машинного обучения с этической точки зрения непостижимы, однако они не сильно отличаются от человеческого поведения. В статье Marius Dorobantu и Yorick Wilks исследовалась роль рациональности и обоснованности в традиционном этическом мышлении и искусственном интеллекте, при этом подчеркивалась необходимость обоснования своих действий. В работе изучается фактор воплощения Нила Лоуренса как подход к пониманию различий между человеческим и машинным интеллектом, увязанный с теологическим пониманием личности, а также предлагается концепция искусственных моральных подпорок, которая этически обосновывает понятия искусственного и человеческого субъекта права в качестве более многообещающего универсального подхода к человеческой и машинной этике (Dorobantu & Wilks, 2019).

Заключение

В эпоху искусственного интеллекта этика вышла на новый уровень значимости обсуждений (Sudhi & Huraimel, 2021). Относительно рисков искусственного интеллекта консенсус пока не достигнут. Хотя инновациям нельзя препятствовать, нет сомнений, что при применении искусственного интеллекта необходим человеческий надзор, что связано с определенными этическими проблемами. Руководства, рекомендации и проекты законодательства, кратко представленные в данной работе, направлены в том числе на достижение этой цели, т. е. на создание правовой базы для успешного применения этих технологий в будущем.

Инновации присущи любой цивилизации. В Бельмонтском докладе 1978 г. были указаны три этических принципа: уважение к личности, благотворность и справедливость, ставшие основой исследований, в которых участвуют люди в качестве объектов. Однако независимые комитеты по контролю над исследованиями человека в качестве объекта и выпускаемые ими стандарты нацелены на соответствие этике исследовательской работы в Интернете, в области больших данных и искусственного интеллекта. Об этом свидетельствуют выполненная в 2014 г. работа об эффекте эмоционального заражения в Фейсбуке* и споры вокруг исследования 2016 г. «Гей-радар искусственного интеллекта» (Tang, 2020).

Как показывает проведенный анализ, этические риски искусственного интеллекта, способного к самостоятельной деятельности, представляют собой существенный вызов законодательству. Машинное обучение стало популярным инструментом многих приложений уголовного права, включая вынесение приговора и выработку правовых мер. Однако также потенциально возможно, что прогностические системы будут создавать эффект неравномерности и усиливать социальное неравенство. Хотя в ряде исследований было показано, что модели машинного обучения могут успешно справляться с такими задачами, им также присуще воспроизведение системных предвзятостей, первоначально заложенных человеком. При этом важность справедливых установок для приложений машинного обучения в законотворчестве недостаточно исследовалась учеными (Alikhademi et al., 2022).

Соответствующие рекомендации служат основой для развития законодательства; при этом необходимо гармонизировать международные и национальные стандарты и реорганизовать системы стандартов. Трудовое законодательство может сыграть важную роль в этом подходе. Наиболее значимыми являются обязывающие гражданско-правовые договоры и ресурсы, ставящие задачу достичь полной занятости; также важны инструменты регулирования массовых увольнений (Stefano, 2018). Заметные усилия по регулированию искусственного интеллекта прилагаются не только в Евросоюзе; для Европы характерен подход, основанный на оценке риска, тогда как в Соединенных Штатах предлагаются более специализированные указания (Sussman, 2021). Общей чертой является то, что для развития научных исследований и расширения использования технологий искусственного интеллекта необходимо создать инфраструктуру, которая способствовала бы сотрудничеству между заинтересованными лицами и обеспечивала деятельность в соответствии с рамками закона и, что не менее важно, этики. Кроме того, необходимо понять суть взаимодействия между искусственным интеллектом (роботами) и человеком, а также проблемы эмоционального интеллекта, как отмечал József Hajdú (Hajdú, 2020).

Государства признают значимость искусственного интеллекта и начинают разрабатывать нормативную базу вокруг стратегии автономного ИИ. 25 апреля 2018 г. Венгрия вместе с еще 25 европейскими странами подписала Декларацию о сотрудничестве в сфере искусственного интеллекта, в которой фиксируется намерение стран-участниц сотрудничать и поддерживать инновации в сфере разработки искусственного интеллекта в Европе²⁹.

Споры вокруг этики искусственного интеллекта достигли критической точки, поскольку сегодня текстовые программы с ИИ общего назначения, такие как GPT-3, способны генерировать масштабные, сложные тексты, выглядящие так, будто они написаны человеком. С другой стороны, существует программа обработки информации на естественном языке (Krutilla & Kővári, 2022). Бизнес-сообщество принимает недостаточное участие в обсуждении этических вопросов, касающихся слияния ролей человека и машины при создании контента (Illia et al., 2023).

Центральной проблемой нашего времени является этика при принятии решений. По нашему мнению, автономные интеллектуальные системы не учитывают потребности человека, они принимают объективные решения. Следовательно, необходимо создать искусственный интеллект, ориентированный на человека, отвечающий ценностям и этическим принципам общества, т. е. разработать механизмы сдерживания для защиты справедливости человеческого достоинства.

Подводя итоги, можно сказать, что использование искусственного интеллекта приводит к комплексным изменениям. При этом необходимо создать законодательную базу и обеспечить прозрачность деятельности. В этой связи в настоящей работе представлены отдельные законодательные решения с акцентом на этические аспекты искусственного интеллекта в приложении к трудовым отношениям.

«Искусственный интеллект никогда не станет этическим. Это инструмент, и, как любой инструмент, его можно использовать во благо и во зло. Нет хорошего ИИ, есть только хорошие и плохие люди. Мы не настолько умны, чтобы сделать ИИ этическим.

²⁹ Стратегия развития искусственного интеллекта (Венгрия). https://ai-watch.ec.europa.eu/countries/hungary/hungary-ai-strategy-report_en

Мы не настолько умны, чтобы сделать ИИ нравственным... В конечном итоге я считаю, что единственный способ избежать гонки вооружений с использованием ИИ – это не иметь ИИ вообще. Это единственная защита от ИИ» (Connock, 2023).

* Организация признана экстремистской, ее деятельность запрещена на территории Российской Федерации.

Список литературы

- Alikhademi, K., Drobina, E., Prioleau, D., Richardson, B., & Gilbert, J. E. (2022). A review of predictive policing from the perspective of fairness. *Artificial Intelligence and Law*, 4(23), 1–17. <https://doi.org/10.1007/s10506-021-09286-4>
- Beranger, J. (2021). *Societal Responsibility of Artificial Intelligence: Towards an Ethical and Eco-responsible AI*. UK: Wiley-Iste.
- Boyd, D., & Crawford, K. (2012). Az adatrengeteg kínos kérdései. *Információs Társadalom*, 12(2), 7. <https://doi.org/10.22503/infars.xii.2012.2.1>
- Candriam Academy. (2022). *What is the European Commission's HLEG?*
- Capgemini Research Institute. (2019). *Why addressing ethical questions in AI will benefit organization*.
- Capurro, R. (2018). Digital Ethics. *International Journal of Applied Research on Information Technology and Computing*, 9/1, 23–31.
- Connock, A. (2023). *Media Management and Artificial Intelligence: Understanding Media Business Models in the Digital Age*. UK: Routledge.
- Cyman, D., Gromova, E., & Juchnevicius, E. (2021). Regulation of Artificial Intelligence in BRICS and the European Union, *BRICS Law Journal*, 8(1), 86–115. <https://doi.org/10.21684/2412-2343-2021-8-1-86-115>
- Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*.
- Dorobantu, M., & Wilks, Y. (2019). Moral orthoses: a new approach to human and machine ethics. *Zygon Journal of Religion and Science*, 54(4), 12–23. <https://doi.org/10.1111/zygo.12560>
- Eszteri, D. (2015). A mesterséges intelligencia fejlesztésének és üzemeltetésének egyes felelősségi kérdései. *Infokommunikáció és Jog*, 47–57. ISSN 1786-0776
- Fleischmann, S. S., Greenberg, K. R., Verma, N., Cummings, B., Li, L., & Shenefel, C. (2023). Locazing the work of artificial intelligence ethics. *Journal of the Association for Information Science and Technology*, 74(3), 311–322. <https://doi.org/10.1002/asi.24638>
- Fobel, P. (2002). Alkalmazott filozófia és etika. In S. Karikó, & S. Karikó (Szerk.), *Az alkalmazott filozófia esélyei*. Budapest: Áron Kiadó.
- Hajdú, J. (2020). A mesterséges intelligencia hatása a munkaerőpiacra, avagy elveszik-e a robotok az ember munkáját. *Infokommunikáció és Jog*, 7.
- Harmathy, A. (2019). A polgári jog a változó jogrendszerben. In V. Lamm, & A. Sajó, *Studia in honorem Lajos Vékás*. Budapest: HVG-ORAC Lapés Könyvkiadó Kft.
- Hoffmann, A. L., Roberts, S. T., Wolf, C. T., & Wood, S. (2019). Beyond fairness, accountability, and transparency in the ethics of algorithms: Contributions and perspectives from LIS. *Proceedings of the Association for Information Science and Technology*, 55(1), 694–696. <https://doi.org/10.1002/pra2.2018.14505501084>
- Illia, L., Colleoni, E., & Zyglidopoulou, S. (2023). Ethical implications of text generation in the age of artificial intelligence. *Business Ethics, the Environment & Responsibility*, 32(1), 201–210. <https://doi.org/10.1111/beer.12479>
- Karvalics, Z. L. (2015). Mesterséges intelligencia – a diskurzusok újratervzésének kora. *Információs Társadalom*.
- Knowles, M. A. (2021). Five Motivating Concerns for AI Ethics Instruction. *Proceedings of the Association for Information Science and Technology*, 58, 472–476. <https://doi.org/10.1002/pra2.481>
- Krutilla, Z., & Kővári, A. (2022). The origin and primary areas of application of natural language processing. In *2022 IEEE 22nd International Symposium on Computational Intelligence and Informatics and 8th IEEE International Conference on Recent Achievements in Mechatronics Automation Computer Science and Robotics* (pp. 293–298). <https://doi.org/10.1109/cinti-macro57952.2022.10029432>
- Legeza, L. (2013). *Mérnöki etika*. Budapest.
- Müller, J., & Kerényi, Á. (2019). A bizalom esetika igénye a digitális korszakban. *Hitelintézeti Szemle*, 18/4, 8–19.
- Négyesi, I. (2020). A mesterséges intelligencia és az etika. *Társadalomtudomány*, 104.

- Németh, G. (2021). Jogászai etikai kihívások a technológiai fejlődés tükrében: az etika és jog innovációjának aktuális kérdései. Tanulmányok. http://real.mtak.hu/108838/1/JAP-2020-01_NG.pdf
- Pusztahelyi, R. (2019). Bizalmunkra méltó MI – A mesterséges intelligencia fejlesztésének és alkalmazásának erkölcsi-etikai vonatkozásairól. *Publicationes Universitatis Miskolcensis Sectio Juridica et Politica*, XXXVII/2, 99.
- Spiekermann, S., Krasnova, H., Hinz, O., Baumann, A., Benlian, A., Grimple, H., & Trenz, M. (2022). Values and Ethics in Information Systems. *Bus Inf Syst Eng*, 64, 247–264. <https://doi.org/10.1007/s12599-021-00734-8>
- Stefán, I. (2020). A mesterséges intelligencia fogalmának polgári jogi értelmezése. *Pro Futuro*, 1, 29–39.
- Stefano, V. (2018). “Negotiating the algorithm”: Automation, artificial intelligence and labour protection. *Employment Working Paper*, 246.
- Sudhi, S., & Huraimel, K. (2021). Dealing with Ethics, Privacy, and Security. In *Reimagining Businesses with AI* (pp. 193–206).
- Sussman, E. H. (2021). *U. S. Artificial Intelligence Regulation takes shape*.
- Tang, B. (2020). Independent AI Ethics Committees and ESG Corporate Reporting on AI as Emerging Corporate and AI Governance Trends. In S. Chishti, I. Bartoletti, A. Leslie, & S. Millie, *The AI Book: The Artificial Intelligence Handbook for Investors, Entrepreneurs and FinTech Visionaries*. <https://doi.org/10.1002/9781119551966.ch48>
- Tilesch, G., & Hatamleh, O. (2021). *Mesterség és Intelligencia*. Libri Kiadó.
- Turay, A. (2000). *Az ember és az erkölcs- Alapvető etika Aquinói Tamás nyomán*. Szeged: Agapé, Ferences Nyomda és Könyvkiadó Kft.
- Yu, H., Shen, Zh., Miao, Ch., Leung, C., Lesser, V. R., & Yang, O. (2018). *Building ethics into Artificial Intelligence*. <http://arxiv.org/pdf/1812.02953.pdf>

Сведения об авторе



София Рицу – аспирант кафедры аграрного и трудового права, Университет Мишкольца

Адрес: Университетский городок, г. Мишколец, Венгрия

E-mail: jogriczu@uni-miskolc.hu

ORCID ID: <https://orcid.org/0000-0002-4024-5833>

Конфликт интересов

Автор заявляет об отсутствии конфликта интересов.

Финансирование

Исследование не имело спонсорской поддержки.

Тематические рубрики

Рубрика OECD: 5.05 / Law

Рубрика ASJC: 3308 / Law

Рубрика WoS: OM / Law

Рубрика ГРНТИ: 10.87.91 / Международное право в практике отдельных государств

Специальность ВАК: 5.1.5 / Международно-правовые науки

История статьи

Дата поступления – 9 апреля 2023 г.

Дата одобрения после рецензирования – 22 апреля 2023 г.

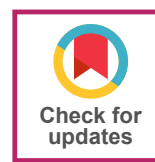
Дата принятия к опубликованию – 16 июня 2023 г.

Дата онлайн-размещения – 20 июня 2023 г.



Research article

DOI: <https://doi.org/10.21202/jdtl.2023.21>



Recommendations on the Ethical Aspects of Artificial Intelligence, with an Outlook on the World of Work

Zsafia Riczu 

University of Miskolc
Miskolc, Hungary

Keywords

Artificial intelligence,
digital technologies,
digitalization,
ethics,
labor law,
labor relations,
labor,
law,
legislation,
principles of law

Abstract

Objective: the spread and wide application of Artificial Intelligence raises ethical questions in addition to data protection measures. That is why the aim of this paper is to examine ethical aspects of Artificial Intelligence and give recommendations for its use in labor law.

Methods: research based on the methods of comparative and empirical analysis. Comparative analysis allowed to examine provisions of the modern labor law in the context of use of Artificial Intelligence. Empirical analysis made it possible to highlight the ethical issues related to Artificial Intelligence in the world of work by examining the disputable cases of the use of Artificial Intelligence in different areas, such as healthcare, education, transport, etc.

Results: the private law aspects of the ethical issues of Artificial Intelligence were examined in the context of ethical and labor law issues that affect the selection process with Artificial Intelligence and the treatment of employees as a set of data from the employers' side. Author outlined the general aspects of ethics and issues of digital ethics. Author described individual international recommendations related to the ethics of Artificial Intelligence.

Scientific novelty: this research focused on the examination of ethical issues of the use of Artificial Intelligence in the specific field of private law – labor law. Authors gave recommendations on ethical aspects of use of Artificial Intelligence in this specific field.

© Riczu Zs., 2023



This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Practical significance: research contributes to the limited literature on the topic. The results of the research could be used in lawmaking process and also as a basis for future research.

For citation

Riczu, Zs. (2023). Recommendations on the Ethical Aspects of Artificial Intelligence, with an Outlook on the World of Work. *Journal of Digital Technologies and Law*, 1(2), 498–519. <https://doi.org/10.21202/jdtl.2023.21>

References

- Alikhademi, K., Drobina, E., Prioleau, D., Richardson, B., & Gilbert, J. E. (2022). A review of predictive policing from the perspective of fairness. *Artificial Intelligence and Law*, 4(23), 1–17. <https://doi.org/10.1007/s10506-021-09286-4>
- Beranger, J. (2021). *Societal Responsibility of Artificial Intelligence: Towards an Ethical and Eco-responsible AI*. UK: Wiley-Iste.
- Boyd, D., & Crawford, K. (2012). Az adatrengeteg kínos kérdései. *Információs Társadalom*, 12(2), 7. <https://doi.org/10.22503/infars.xii.2012.2.1>
- Candriam Academy. (2022). *What is the European Commission's HLEG?*
- Capgemini Research Institute. (2019). *Why addressing ethical questions in AI will benefit organization*.
- Capurro, R. (2018). Digital Ethics. *International Journal of Applied Research on Information Technology and Computing*, 9/1, 23–31.
- Connock, A. (2023). *Media Management and Artificial Intelligence: Understanding Media Business Models in the Digital Age*. UK: Routledge.
- Cyman, D., Gromova, E., & Juchnevicius, E. (2021). Regulation of Artificial Intelligence in BRICS and the European Union, *BRICS Law Journal*, 8(1), 86–115. <https://doi.org/10.21684/2412-2343-2021-8-1-86-115>
- Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*.
- Dorobantu, M., & Wilks, Y. (2019). Moral orthoses: a new approach to human and machine ethics. *Zygon Journal of Religion and Science*, 54(4), 12–23. <https://doi.org/10.1111/zygo.12560>
- Eszteri, D. (2015). A mesterséges intelligencia fejlesztésének és üzemeltetésének egyes felelősségi kérdései. *Infokommunikáció és Jog*, 47–57. ISSN 1786-0776
- Fleischmann, S. S., Greenberg, K. R., Verma, N., Cummings, B., Li, L., & Shenefel, C. (2023). Locazing the work of artificial intelligence ethics. *Journal of the Association for Information Science and Technology*, 74(3), 311–322. <https://doi.org/10.1002/asi.24638>
- Fobel, P. (2002). Alkalmazott filozófia és etika. In S. Karikó, & S. Karikó (Szerk.), *Az alkalmazott filozófia esélyei*. Budapest: Áron Kiadó.
- Hajdú, J. (2020). A mesterséges intelligencia hatása a munkaerőpiacra, avagy elveszik-e a robotok az ember munkáját. *Infokommunikáció és Jog*, 7.
- Harmathy, A. (2019). A polgári jog a változó jogrendszerben. In V. Lamm, & A. Sajó, *Studia in honorem Lajos Vékás*. Budapest: HVG-ORAC Lapés Könyvkiadó Kft.
- Hoffmann, A. L., Roberts, S. T., Wolf, C. T., & Wood, S. (2019). Beyond fairness, accountability, and transparency in the ethics of algorithms: Contributions and perspectives from LIS. *Proceedings of the Association for Information Science and Technology*, 55(1), 694–696. <https://doi.org/10.1002/pra2.2018.14505501084>
- Illia, L., Colleoni, E., & Zyglidopoulou, S. (2023). Ethical implications of text generation in the age of artificial intelligence. *Business Ethics, the Environment & Responsibility*, 32(1), 201–210. <https://doi.org/10.1111/beer.12479>
- Karvalics, Z. L. (2015). Mesterséges intelligencia – a diskurzusok újratervzésének kora. *Információs Társadalom*.
- Knowles, M. A. (2021). Five Motivating Concerns for AI Ethics Instruction. *Proceedings of the Association for Information Science and Technology*, 58, 472–476. <https://doi.org/10.1002/pra2.481>

- Krutilla, Z., & Kővári, A. (2022). The origin and primary areas of application of natural language processing. In *2022 IEEE 22nd International Symposium on Computational Intelligence and Informatics and 8th IEEE International Conference on Recent Achievements in Mechatronics Automation Computer Science and Robotics* (pp. 293–298). <https://doi.org/10.1109/cinti-macro57952.2022.10029432>
- Legeza, L. (2013). *Mérnöki etika*. Budapest.
- Müller, J., & Kerényi, Á. (2019). A bizalom esetika igénye a digitális korszakban. *Hitelintézeti Szemle*, 18/4, 8–19.
- Négyesi, I. (2020). A mesterséges intelligencia és az etika. *Társadalomtudomány*, 104.
- Németh, G. (2021). Jogászai etikai kihívások a technológiai fejlődés tükrében: az etika és jog innovációjának aktuális kérdései. *Tanulmányok*. http://real.mtak.hu/108838/1/JAP-2020-01_NG.pdf
- Pusztahelyi, R. (2019). Bizalmunkra méltó MI – A mesterséges intelligencia fejlesztésének és alkalmazásának erkölcsi-etikai vonatkozásairól. *Publicationes Universitatis Miskolcensis Sectio Juridica et Politica*, XXXVII/2, 99.
- Spiekermann, S., Krasnova, H., Hinz, O., Baumann, A., Benlian, A., Grimple, H., & Trenz, M. (2022). Values and Ethics in Information Systems. *Bus Inf Syst Eng*, 64, 247–264. <https://doi.org/10.1007/s12599-021-00734-8>
- Stefán, I. (2020). A mesterséges intelligencia fogalmának polgári jogi értelmezése. *Pro Futuro*, 1, 29–39.
- Stefano, V. (2018). “Negotiating the algorithm”: Automation, artificial intelligence and labour protection. *Employment Working Paper*, 246.
- Sudhi, S., & Huraimel, K. (2021). Dealing with Ethics, Privacy, and Security. In *Reimagining Businesses with AI* (pp. 193–206).
- Sussman, E. H. (2021). *U. S. Artificial Intelligence Regulation takes shape*.
- Tang, B. (2020). Independent AI Ethics Committees and ESG Corporate Reporting on AI as Emerging Corporate and AI Governance Trends. In S. Chishti, I. Bartoletti, A. Leslie, & S. Millie, *The AI Book: The Artificial Intelligence Handbook for Investors, Entrepreneurs and FinTech Visionaries*. <https://doi.org/10.1002/9781119551966.ch48>
- Tilesch, G., & Hatamleh, O. (2021). *Mesterség és Intelligencia*. Libri Kiadó.
- Turay, A. (2000). *Az ember és az erkölcs- Alapvető etika Aquinói Tamás nyomán*. Szeged: Agapé, Ferences Nyomda és Könyvkiadó Kft.
- Yu, H., Shen, Zh., Miao, Ch., Leung, C., Lesser, V. R., & Yang, O. (2018). *Building ethics into Artificial Intelligence*. <http://arxiv.org/pdf/1812.02953.pdf>

Author information



Zsófia Riczu – PhD candidate, Faculty of Law, Agricultural and Labour Law Department, University of Miskolc

Address: Miskolc-Egyetemváros, Miskolc, Hungary

E-mail: jogriczu@uni-miskolc.hu

ORCID ID: <https://orcid.org/0000-0002-4024-5833>

Conflict of interest

The author declares no conflict of interest.

Financial disclosure

The research had no sponsorship.

Thematic rubrics

OECD: 5.05 / Law

PASJC: 3308 / Law

WoS: OM / Law

Article history

Date of receipt – April 9, 2023

Date of approval – April 22, 2023

Date of acceptance – June 16, 2023

Date of online placement – June 20, 2023

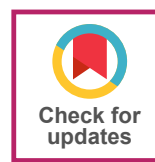


Научная статья

УДК 340.143:004.8

EDN: <https://elibrary.ru/ekeudk>

DOI: <https://doi.org/10.21202/jdtl.2023.22>



Этико-правовые модели взаимоотношений общества с технологией искусственного интеллекта

Дмитрий Валерьевич Бахтеев 

Уральский государственный юридический университет имени В. Ф. Яковлева
г. Екатеринбург, Российская Федерация

Ключевые слова

ChatGPT,
искусственный интеллект,
машинное обучение,
модель,
общество,
право,
регулирование,
робот,
цифровые технологии,
этика

Аннотация

Цель: исследование современного состояния технологии искусственного интеллекта в формировании прогностических этико-правовых моделей взаимоотношений общества с рассматриваемой сквозной цифровой технологией.

Методы: основным методом исследования является моделирование. Помимо него, в работе использованы сравнительный, абстрактно-логический и исторический методы научного познания.

Результаты: сформулированы четыре этико-правовые модели взаимоотношений общества с технологией искусственного интеллекта: инструментальная (на основе использования человеком системы искусственного интеллекта), ксенофобная (на основе конкуренции человека и системы искусственного интеллекта), эмпатическая (на основе сочувствия и соадаптации человека и систем искусственного интеллекта), толерантная (на основе взаимоиспользования и сотрудничества между человеком и системами искусственного интеллекта). Приведены исторические и технические предпосылки формирования таких моделей. Описаны сценарии реакций законодателя на ситуации использования этой технологии, такие как необходимость точечного регулирования, отказа от регулирования либо же полномасштабного вмешательства в технологическую отрасль экономики. Произведено сравнение моделей по критериям условий реализации, достоинства, недостатков, характера отношений «человек – система искусственного интеллекта», возможных правовых последствий и необходимости регулирования отрасли либо отказа от такового.

© Бахтеев Д. В., 2023



Статья находится в открытом доступе и распространяется в соответствии с лицензией Creative Commons «Attribution» («Атрибуция») 4.0 Всемирная (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0/deed.ru>), позволяющей неограниченно использовать, распространять и воспроизводить материал при условии, что оригинальная работа упомянута с соблюдением правил цитирования.

Научная новизна: в работе приведена оценка существующих в научной литературе, публицистике мнений и подходов, проанализированы технические решения и проблемы, возникшие в недавнем прошлом и настоящем. Теоретические выводы подтверждаются ссылками на прикладные ситуации, имеющие общественную или правовую значимость. В работе использован междисциплинарный подход, объединяющий правовую, этическую и техническую составляющие, которые, по мнению автора, являются критериальными для любых современных социогуманитарных исследований технологий искусственного интеллекта.

Практическая значимость: феномен искусственного интеллекта связывают с четвертой промышленной революцией, соответственно, эта цифровая технология должна быть изучена многоаспектно и междисциплинарно. Выработанные в научной статье подходы могут быть использованы при дальнейших технических разработках интеллектуальных систем, совершенствования отраслевого законодательства (например, гражданского и трудового), а также при формировании и модификации этических кодексов в сфере разработки, внедрения и использования систем искусственного интеллекта в различных ситуациях.

Для цитирования

Бахтеев, Д. В. (2023). Этико-правовые модели взаимоотношений общества с технологией искусственного интеллекта. *Journal of Digital Technologies and Law*, 1(2), 520–539. <https://doi.org/10.21202/jdtl.2023.22>

Содержание

Введение

1. Инструментальная модель

2. Ксенофобная модель

3. Эмпатическая модель

4. Толерантная модель

Выводы

Список литературы

Введение

Многочисленные достоинства систем искусственного интеллекта, среди которых быстрая обучаемость, возможность решать широкий круг задач, более высокая, чем у человека, эффективность вкупе со все большим проникновением их в различные сферы нашей жизни, заставляют задуматься над вопросами: способна ли (или будет ли способна) система искусственного интеллекта воспринимать себя как самостоятельную, независимую от разработчиков и пользователей личность, осознает ли искусственный интеллект свои преимущества перед людьми, как он будет оценивать свое положение и взаимодействие с человечеством, будет ли оно его устраивать и что он будет делать, если захочет его изменить. Эти вопросы находятся на пересечении предметности этики, права и технологии и потому подлежат разрешению междисциплинарными методами исследования (Kazim, 2021). В зависимости от ответов на эти вопросы и степени развития технологии возможны приведенные ниже модели/сценарии реакции общества и, как следствие, права на технологии искусственного интеллекта.

1. Инструментальная модель

Инструмент – продукт деятельности человека и средство изготовления других предметов, в том числе и инструментов. Карл Маркс и Мартин Хайдеггер в своих трудах подчеркивали необходимость различия между инструментом и машиной. Первый указывал на то, что зачастую смешиваются принципиально разные понятия: если инструмент требует непосредственного участия человека в процессе труда, то машина «заменяет рабочего, действующего одновременно только одним орудием, таким механизмом, который разом оперирует множеством одинаковых или однородных орудий и приводится в действие одной двигательной силой» (Маркс, 2001). Он же отмечает, что машину от инструмента отличает достаточная автономия (в большей степени ресурсная, машину в любом случае направляет и контролирует человек). Мартин Хайдеггер же, в свою очередь, с опорой на работы Гегеля, в качестве критериев отнесения объекта к машине называет самостоятельность, уверенность в себе и независимость (Хайдеггер, 1993). Для данного исследования это может пониматься следующим образом: искусственный интеллект в рамках инструментальной концепции на этапе разработки и тестирования выступает более как инструмент, а не как машина, поскольку его функционирование связано с человеческой деятельностью на трех уровнях: при разработке (в этом инструмент и машина схожи), при реализации деятельности, ранее присущей исключительно человеку, при контроле человеком результатов деятельности искусственного интеллекта. Такой подход иногда также называют прагматическим (Morley J. et al., 2021). «Машинность» искусственного интеллекта в данном случае является производной характеристикой от автономности, однако с точки зрения рассматриваемой модели информационная автономность систем искусственного интеллекта отрицается. Соответственно, в рамках данной модели искусственный интеллект рассматривается именно в качестве инструмента, в том числе для реализации нужд человечества (Watkins & Human, 2023).

Описанная в работах автора статьи теория автономности (информационной и ресурсной) (Бахтеев, 2021) вполне согласуется с тем, что системы искусственного интеллекта могут восприниматься и как машины (что, в свою очередь, подтверждается существованием терминов «машинное обучение» и «машинное зрение»), и как сущности, сопоставимые с познающими биологическими объектами. При этом следует учитывать, что если инструмент (в традиционном научном понимании) призван облегчить труд человека, то искусственный интеллект может заменить своей деятельностью труд человека. При этом некорректным будет сравнение искусственного интеллекта со станками времен промышленной революции. Такие станки лишили работы множество людей, однако при этом они создали новые рабочие места. В случае автоматизации вообще и использования систем искусственного интеллекта в частности мы уже наблюдаем ликвидацию определенных профессий, в первую очередь связанных с посредническими услугами: консультантов, диспетчеров, маркетологов и т. д. Обычные производственные роботы уже сейчас, без интеллектуальных модулей, в значительной степени оптимизировали конвейерное производство, что позволило при сокращении издержек на производство продукции многократно повысить количество и качество продукции, а также лишило работы огромное число неквалифицированных и низкоквалифицированных рабочих, фактически мы наблюдаем новую промышленную революцию. К 2020 г. в мире применялись более 3 млн промышленных роботов (к 2023 г. темпы роста, впрочем, снизились), происходит

интеграция интеллектуальных систем в разные области жизни, что очевидным образом положительно сказывается на экономике, но причиняет определенный ущерб обществу, в первую очередь в виде снижения занятости. Вариантом решения такой проблемы могут быть гарантии занятости, закрепленные законодательно, либо поддерживаемые государством и крупными компаниями программы переквалификации для людей, потерявших работу.

Развитие искусственного интеллекта позволяет решать ему все больший круг интеллектуальных задач, что создает предпосылки для дальнейшего сокращения рабочих мест во все большем количестве сфер, а также уничтожения целых профессий. «Реальный эффект сокращения фонда заработной платы (и числа рабочих мест. – Прим. Д. Б.) за счет внедрения роботов определяется количеством высвобождающихся при этом людей и величиной их заработной платы, а также стоимостью самих роботов, которая в свою очередь определяется сложностью конструкции и степенью интеллектуальности роботов» (Тимофеев, 1978).

В отличие от начального этапа роботизации распространение систем искусственного интеллекта может снизить уровень занятости не только для рабочих профессий. Так, согласно прогнозам экспертов Оксфордского университета, в течение следующих 20 лет в США будет автоматизировано 47 %, а в Китае – 77 % рабочих мест. По мнению одного из ведущих специалистов в сфере вычислительной техники М. Варди, к 2045 г. примерно 50 % людей останутся без работы (Vardi, 2012). Можно возразить, что при этом увеличивается количество разработчиков искусственного интеллекта, однако увеличение числа программистов и других лиц, связанных с разработкой интеллектуальных систем, не идет в сравнение с уменьшением рабочих мест других профессий. Более того, стремление создать интеллектуальные системы, способные к саморепликации, также вполне очевидно, поэтому нельзя исключать сокращений и в профессиях, связанных с информационными технологиями: например, ChatGPT в своей четвертой версии способна создавать несложный, но компилируемый программный код. Так, в январе 2023 г. компания Alphabet объявила о сокращении 12 000 рабочих мест по всему миру, в том числе из-за внедрения различных интеллектуальных систем, способных заменить, в частности, маркетологов, копирайтеров и иллюстраторов¹. Согласно модели, разработанной для анализа вероятности исчезновения отдельных профессий за счет использования интеллектуальных чат-систем, 100-процентное замещение человека на настоящем этапе развития технологии возможно для математиков, налоговых специалистов, финансовых аналитиков, писателей, копирайтеров, веб-дизайнеров, верстальщиков, секретарей государственных органов, новостных аналитиков (Eloundou T. et al., 2023).

К настоящему времени именно инструментальную модель можно считать единственной полноценно реализованной: в прикладной деятельности системы искусственного интеллекта выступают в качестве инструмента как средства повышения эффективности труда. Это, как и в случаях прежних технологических революций, налагает на государство и общество функцию сохранения рабочих мест и регулирования интенсивности применения описываемой технологии.

¹ Pichar, S. (2023, January 20). A difficult decision to set us up for the future. <https://blog.google/inside-google/message-ceo/january-update/>

2. Ксенофобная модель

В связи с активизацией обсуждения вопросов разработки и использования систем искусственного интеллекта все чаще и громче слышатся голоса противников дальнейших исследований в данной области. При этом как минимум часть из них нельзя назвать обскурантами с иррациональным страхом перед технологическим прогрессом. Так, например, всемирно известный ученый, популяризатор науки С. Хокинг говорил: «...появление полноценного искусственного интеллекта может стать концом человеческой расы... Такой разум возьмет инициативу на себя и станет сам себя совершенствовать со все возрастающей скоростью. Возможности людей ограничены слишком медленной эволюцией, мы не сможем тягаться со скоростью машин и проиграем». Схожего мнения придерживается и известный американский инженер, IT предприниматель И. Маск, который заявил: «Полагаю, что искусственный интеллект рано или поздно прикончит всех нас... Facebook*, Google, Amazon, Apple – все они уже знают о вас очень многое. Искусственный интеллект, который будет создан в недрах этих корпораций, получит огромную власть над людьми. А концентрация власти в одних руках всегда порождает огромные риски». Отмечается также, что «распределение функций между системой искусственного интеллекта и человеком должно следовать принципу, ориентированному на человека, и оставлять всегда возможность для человеческого выбора. Это означает обеспечение человеческого контроля над рабочими процессами в системах искусственного интеллекта» (Семис-оол, 2019).

Данные факты детерминируют потребность тщательного исследования «ксенофобной» модели отношения к искусственному интеллекту. Следует указать, что она является развитием инструментальной модели по негативному сценарию, т. е. вследствие как прогресса в развитии и использовании систем искусственного интеллекта, так и реализации одного или нескольких рисков (в виде одномоментных негативных событий или же затяжных кризисов), описанных ранее.

Термин «ксенофобная» образован путем сочетания греческих слов ξένος («чужой») + φόβος («страх»). Тем самым буквально ксенофобия определяется как страх, нетерпимость к чужому, незнакомому².

Среди исследователей нет единства мнений об истоках происхождения ксенофобии. Ряд авторов отмечают, что она могла возникнуть как инструмент адаптации в процессе эволюции, который способствовал выживанию и передаче генов потомкам. Так, страх перед незнакомцами мог быть, среди прочего, основан на наблюдении, что чужаки могут быть разносчиками новых, а значит, очень опасных (из-за отсутствия нужных антител) для коренных жителей болезнетворных микроорганизмов.

Традиционно термин «ксенофобия» использовался для обозначения страха, неприязни к людям других рас, национальностей, культур и религий. Тем не менее, на наш взгляд, исследование процесса взаимодействия человечества с достижениями технологического прогресса дает основания заимствовать данный термин в том числе и для описания определенного типа отношения к ходу научно-технического прогресса и его плодам – технологиям, к числу которых очевидно относится и искусственный интеллект.

² Ожегов, С. И., Шведова, Н. Ю. (2016). *Толковый словарь русского языка* (с. 300). Москва: ООО «А ТЕМП».

Завершая изложение нашего подхода к ксенофобии, отметим важный аспект, который заключается в том, что в конечном итоге ксенофобия представляет собой специфическую разновидность страха. Страх, являясь одной из множества эмоций, по мнению Е. П. Ильина, выступает как «эмоциональное состояние, отражающее защитную биологическую реакцию человека или животного при переживании им мнимой или реальной опасности для их здоровья или благополучия» (Ильин, 2016). Далее Е. П. Ильин отмечает, что с биологической точки зрения страх – несомненно, полезное явление, в то время как для человека как социального существа страх нередко служит препятствием для достижения поставленных целей. В данной части работы будут исследованы основы потенциального критического недоверия общества к технологии искусственного интеллекта.

Сущность «ксенофобного подхода» к искусственному интеллекту заключается в рассмотрении его как реальной угрозы для человечества и его текущего положения в мире.

Обобщающий анализ критики технологии искусственного интеллекта позволяет выделить две основные формы страха (недоверия) людей к этой технологии – сущностную и инструментальную.

Сущностный страх связан с тем, что люди боятся не применения возможностей технологии искусственного интеллекта, а самого искусственного интеллекта как искусственного, но при этом вполне самостоятельного и автономного разума, способного существовать, учиться, мыслить и осознавать себя без участия человека. Появление такой «рукотворной мыслящей машины», которая способна мыслить не просто как человек, но и лучше человека, по существу, подрывает существовавшую на протяжении всей истории цивилизации монополию человека на мыслительную деятельность, которая и позволила ему занять доминантное положение среди всех других видов на планете. В этом смысле искусственный интеллект становится отдельным видом, который человечество не может воспринимать иначе как конкурентный. При этом страх вызывает именно вышедший из-под контроля искусственный интеллект, т. е. ситуация обретения искусственным интеллектом самосознания в результате программного сбоя или умышленных действий разработчика. Фактически такую ситуацию, проводя аналогию с биологическими процессами, следует считать мутацией, однако сомнительно, что процессы разработки технологических продуктов имеют настолько большую связь с эволюционными механизмами. Именно поэтому сценарий «агрессивного» искусственного интеллекта представляется крайне нереалистичным.

Инструментальный страх, в свою очередь, отражает боязнь вытеснения человека системами искусственного интеллекта в трудовой сфере, что было описано в инструментальной модели (см. выше).

После начала систематического исследования искусственного интеллекта в 40–50-е гг. XX в. не прошло и ста лет, как уже созданы системы, превосходящие человека в отдельных областях интеллектуальной деятельности. Возможности современных компьютеров пока не позволяют произвести полное моделирование сознания человека или всего окружающего мира, однако с абстракциями искусственный интеллект справляется. Игры – это вполне корректная абстракция и, что в данном случае крайне важно, результаты участия в игре могут быть точно оценены. Так, уже с начала 2000-х гг. сильнейшие игроки в шахматы в мире ничего не могут противопоставить компьютеру, а по словам Г. Каспарова, все профессиональные шахматисты

тренируются, играя против компьютерных шахматных программ, поскольку соперник-человек не может предоставить достаточной глубины вариантов ходов. В 2015 г. компьютерная программа впервые победила человека в игре го – одной из самых сложных игр с открытой информацией³, что ранее считалось невозможным. Искусственные нейросети способны обыгрывать профессиональных игроков в компьютерные игры в киберспортивных дисциплинах⁴, что также считалось ранее исключительной прерогативой человека.

Содержание ксенофобного подхода заключается в том, что искусственный интеллект может использоваться отдельными лицами, организациями и целыми государствами как средство достижения своих недобросовестных целей.

Типичным примером данного страха является скандал, возникший вскоре после президентских выборов в США, связанный с организацией Cambridge Analytica. Согласно информации из ряда источников, данная частная организация, используя новейшие методы сбора и анализа данных в социальной сети Facebook*, получила огромный массив данных, в том числе и личных, для разработки специальной политической рекламы, которая, по мнению ряда экспертов, в значительной степени способствовала избранию на президентскую должность нынешнего президента США. Более того, данная организация обвиняется в причастности к вмешательству в результаты более 200 выборов по всему миру. Бывший сотрудник Cambridge Analytica, К. Уайли, отмечал: «Мы использовали несовершенство программного обеспечения Facebook* для сбора миллионов пользовательских профилей и построения моделей, которые позволяли нам узнавать о людях и применять эти знания для активации их внутренних демонов»⁵.

Данное событие демонстрирует, что уже сейчас возможности систем искусственного интеллекта применяются для сбора персональных данных, манипуляции общественным мнением. В будущем искусственный интеллект может использоваться для манипуляции огромными объемами информации, формируя картину мира для населения целых государств, что создает реальные угрозы для демократических институтов, свободы слова и распространения информации. Этим могут воспользоваться государства для навязывания определенной позиции своему населению, а также населению других стран; представители различных корпораций для искусственного формирования спроса на те или иные товары и услуги. Наконец, этим могут воспользоваться представители преступного мира для сбора конфиденциальной информации граждан и организаций, которая впоследствии может быть продана на черном рынке или использована для шантажа или мошеннических действий.

³ Так, в шахматах на каждый шаг игры существует около 20 возможных реакций-ответов, в го их около 200.

⁴ Statt, N. (2019, April 13). OpenAI's Dota 2 AI steamrolls world champion e-sports team with back-to-back victories. *The Verge*. <https://www.theverge.com/2019/4/13/18309459/openai-five-dota-2-finals-ai-bot-competition-og-e-sports-the-international-champion>

⁵ Черешнев, Е. (20 марта 2018). Беззащитные данные: как Facebook оказалась в центре самого большого скандала в истории. *Forbes*. <https://www.forbes.ru/tehnologii/358883-bezzashchitnye-dannye-kak-facebook-okazalas-v-centre-samogo-bolshogo-skandala-v>

Иной пласт данной проблемы заключается в применении искусственного интеллекта в военных действиях. Условный боевой робот, снабженный искусственным интеллектом, или даже целая армия таких роботов представляет собой эффективную замену действующим регулярным войскам. Он способен выполнять сложные интеллектуальные задачи, может действовать даже в самых неблагоприятных условиях, не требует сна и отдыха, а его уничтожение мало волнует общественное мнение в стране, ведущей войну (по крайней мере, до тех пор, пока такая война не становится слишком дорогой с экономической точки зрения). При этом искусственный интеллект способен решать ранее исключительно человеческие задачи с внечеловеческой, машинной рациональностью. Для системы искусственного интеллекта при неверном ее формировании не возникнет проблемы использования незаконных средств ведения войны, убийства гражданского населения и т. п.

Иной аспект ксенофобного подхода связан с явлениями, описанными в части инструментального подхода. Так, голливудская гильдия сценаристов, а также тысячи художников и иллюстраторов по всей планете считают результаты работы системы искусственного интеллекта априорно плагиатом, поскольку они не представляют творчество как таковое, а лишь являются перемешиванием уже раскрытых смыслов. Отметим, однако, что значительная часть человеческого творчества формируется по аналогичной модели.

Таким образом, ксенофобный подход к оценке искусственного интеллекта нельзя назвать однозначно необоснованным. Существуют вполне резонные основания опасаться неконтролируемого развития технологий искусственного интеллекта и все большей интеграции их в разные аспекты жизни человека. Данный подход, предполагающий либо жесткий контроль над исследованиями в данной области, либо, в более радикальной форме, полный отказ от таковых, очевидно, не лишен недостатков. К их числу относятся: торможение научно-технического прогресса, невозможность оптимизации практической деятельности людей посредством использования систем искусственного интеллекта, разрыв между научными достижениями и их интеграцией в практику и, как следствие, снижение авторитета компьютерной науки в глазах общественности. Альтернативой цифровым технологиям, флагманом которых является искусственный интеллект, обычно считаются биотехнологии, поэтому отказ от развития искусственного интеллекта, разочарование в этой технологии могут привести к развитию медицины, физиологии, генетики и т. д. Тем не менее следует, на наш взгляд, не забывать о достоинствах данного подхода, который предполагает более взвешенную оценку технологии и ее прикладного значения; разработку инструментария для прогнозирования и оценки рисков дальнейшего изучения и использования искусственного интеллекта, который с поправками может быть использован и в других областях знания; стимулирование развития иных наук, связанных с развитием человека и его потенциала; придание нового импульса осмыслению человека и его места в мире.

3. Эмпатическая модель

Согласно этой модели, общество, воспринимая положительно домашних и иных социально-бытовых интеллектуальных роботов и программных помощников, благоприятно воспринимает идею распространения технологии и не исключает возможности наделения систем искусственного интеллекта свойствами субъекта права (в ограниченном смысле).

В основе данной модели лежит продвинутое и расширенное ощущение гуманизма и ответственности человека не только за себя, но и за тех, кто рядом. Интеллектуальные и автономные системы, согласно данной модели, перестают считаться инструментом или конкурентом человеку, а рассматриваются как компаньоны, но в ограниченном смысле, на уровне животных-компаньонов. Именно этические и правовые нормы, регулирующие отношение к животным, являются фундаментом реализации этой модели. Фактически данная модель является переходной между инструментальной и толерантной и не может рассматриваться как нечто продолжительное.

Приведем примеры, подтверждающие, что частично данная модель уже реализуется в обществе.

В работе К. Дарлинг и С. Хоерт описывается следующий эксперимент. Отдельным шести группам по восемь человек выдаются игрушки в форме динозавра, размером с небольшую кошку. Участникам предлагалось взаимодействовать с ними. Затем каждой группе была дана команда «задушить», «разбить голову» или иным способом «убить» игрушки, которая встретила жесткое сопротивление: участники не просто отказывались «убить» «своих» динозавров, но и пытались защищать их от других людей и испытывали серьезный дискомфорт, если видели, как «умирает» динозавр⁶. При этом из 48 игрушек только одна была «убита».

Другим примером может служить отключение серверов, обслуживающих роботов-игрушек Jibo, которое многие пользователи восприняли как смерть своего компаньона и отреагировали крайне эмоционально.

Аналогичным образом люди часто реагируют на проблемы и смерть книжных, кинотеатральных и игровых персонажей, несмотря на то, что существуют они исключительно виртуально.

В психологии существует позиция, что людям нравится общаться с чат-ботами (такими как ChatGPT) для обсуждения своих психологических проблем, поскольку интеллектуальная система редко способна на осуждение, свойственное живому человеку⁷. Надо полагать, в будущем это явление будет встречаться еще чаще.

Приведенные ситуации, несмотря на то, что они являются частными случаями и не могут отражать в целом общественное отношение к интеллектуальным системам, демонстрируют, что в определенных случаях человек и группы людей могут относиться к явно неодушевленным (и, стоит признать, даже не вполне интеллектуальным) системам как к любимым питомцам. Отметим также, что чувство эмпатии напрямую зависит от внешнего вида (внешности?) киберфизической системы и используемой лексики. Так, степень эмпатии и доверия к киберфизической системе антропоморфного фенотипа может также зависеть от черт «лица». М. Б. Матур и Д. Б. Райхлинг обнаружили, что надежность робота варьируется в зависимости от сходства его лица с человеком, не увеличивается линейно с человеческим подобием, но падает, когда агент очень реалистичен, но еще не полностью похож на человека (Mathur & Reichling, 2016). Этот феномен, изначально описанный в 1978 г.

⁶ См. Darling, K., Hauert, S. (2013, March 8). Giving rights to robots. *Robohub.org*. RobotsPodcast № 125. <https://robohub.org/robots-giving-rights-to-robots/>

⁷ Эксперт: Люди используют нейросеть в качестве психолога из-за страха осуждения со стороны живого человека. (2023, 23 марта). Агентство городских новостей «Москва». <https://www.mskagency.ru/materials/3286743>

японским ученым М. Мори, носит название «странная долина», или «зловещая долина» (uncanny valley): самые необычные человекоподобные роботы неожиданно оказались неприятными людьми из-за неумовимых несоответствий во внешности и поведении, вызывающих чувство дискомфорта и страха (Mori, 2012). Таким образом, существует вероятность, что с развитием робототехники эмпатическая модель может сдвинуться не в сторону толерантной, а по направлению ко ксенофобной. Так, риски именно психологического характера были отражены в Своде этических правил для роботов, разработанном Британским институтом стандартов: пользователю интеллектуальной системы не должно быть неудобно, он не должен испытывать тревогу и стресс (Winfield, 2019).

Уточнение данной модели требует также раскрытия феномена взаимообучения человека и машины. Взаимодействуя с интеллектуальными системами, человек меняется, при этом изменения затрагивают не только сферу технологических навыков, но также физиологию и морально-этическую сферу. Так, исследование группы швейцарских ученых привело к экспериментально проверенному выводу о том, что «повторяющиеся движения по гладкой поверхности сенсорного экрана изменяют сенсорные реакции и, как следствие, представления мозга о последствиях прикосновений» (Balerna & Ghosh, 2018): мозг современного человека при прикосновении пальцев рук к поверхности с интенсивностью, аналогичной управлению смартфоном, ожидает изменения «картинки» перед глазами.

К другим особенностям, имеющим значение для рассмотрения данной модели, следует отнести ухудшение памяти и концентрации внимания, обусловленное возможностью быстрого обращения к необходимой информации с помощью, к примеру, смартфона или голосового помощника, а также улучшение визуальных навыков, позволяющих быстрее осуществлять более качественное комплексное восприятие сложных визуальных объектов. Не следует в целом считать, что интеграция интеллектуальных систем в общество приводит к деградации последнего. Эффект Флинна демонстрирует, что средний интеллектуальный уровень каждого нового поколения возрастает, т. е. текущий средний коэффициент интеллекта соответствует высокому уровню интеллекта предшествующего поколения (Flynn, 2009). При этом современные исследования показывают, что с распространением цифровых технологий действие эффекта Флинна снизилось или вообще исчезло (Teasdale, 2005). Однако данное исследование не может считаться достоверным, так как его объектом являлся интеллект призывников, т. е. выводы могут объясняться социальными причинами, а не реальным падением интеллектуального уровня. Представляется, что имеет место не снижение интеллекта, а его перепрофилирование (Букатов, 2018), к примеру, современный студент помнит меньше своих предшественников из XX в., однако владеет гораздо большим набором техник для поиска и анализа информации. Соответственно, наблюдается постепенная замена содержательных знаний навыками работы с информацией, «следует учитывать биологическую коадаптацию и козволюцию органов чувств человека, следует учитывать расширение диапазона наших восприятий, которое обеспечивается благодаря достижениям техники» (Огурцов, 2006). Однако нельзя исключать факторы притупления внимания, снижения понимаемой ответственности и профессионализма лиц, принимающих решения, чьим «советником» является система искусственного интеллекта. Таким образом, возможна ситуация «перекладывания» ответственности за ошибку/противоправное деяние на систему искусственного интеллекта.

Помимо физиологического и интеллектуального аспекта, эмпатическая модель включает в себя и возможные изменения в эмоциональной сфере, так, М. Шойц указывает: «Социальные роботы, которые устанавливают с людьми эмоциональный контакт и, как следствие, последние глубоко доверяют роботам, что в свою очередь может быть использовано для манипулирования людьми ранее невозможными способами. Например, компания может использовать уникальные отношения робота со своим владельцем, чтобы робот убедил владельца приобрести продукты, которые компания хочет продвигать. Обратите внимание на отличие от человеческих отношений, где при нормальных обстоятельствах социальные эмоциональные механизмы, такие как эмпатия и вина, предотвратят эскалацию таких сценариев» (Schultz, 2009).

Как известно, во многих странах предусмотрена уголовная ответственность за проявленную по отношению к животным жестокость и тем более лишение жизни животного без оснований на это, так, к примеру, Европейская конвенция о защите домашних животных указывает на недопустимость причинения без необходимости боли и страданий⁸. Продолжая сравнение систем искусственного интеллекта с домашними животными, не будет ли существенное перепрограммирование такой системы причинять ей боль, аналогично тому, как боль животному причиняет косметическое хирургическое вмешательство, также запрещаемое указанной Конвенцией?

Соответственно, в рамках рассматриваемой модели такой подход переносится на системы искусственного интеллекта и с большими оговорками действует в настоящее время в части общества. Однако для полноценной его реализации необходим хотя бы условный и воспринятый обществом ответ на вопрос о том, может ли система искусственного интеллекта испытывать боль и страдания.

4. Толерантная модель

Поступательное развитие научно-технического прогресса при сохранении интереса к совершенствованию рассматриваемых технологий может привести к указанной ранее ситуации – появлению «сильного» искусственного интеллекта или как минимум широкому распространению интеллектуальных систем-помощников. В первом случае ограничения технического характера будут нивелированы, системы искусственного интеллекта получат достаточную автономию, единственным рамочным ограничением которой смогут послужить правовые нормы и производные от них технические ограничения. Такое условное равенство (или как минимум соотносимость) человечества и искусственного интеллекта может повлечь за собой как положительные, так и отрицательные явления.

В рамках данной модели искусственный интеллект выступает «партнером» человечества, обеспечивает функцию сдерживающего механизма, не допускает эскалации конфликтов, осуществляет общую и частную превенцию правонарушений. Негативные сценарии, описанные ранее, не реализованы, поскольку благополучное

⁸ European Convention for the Protection of Pet Animals (1987, November 13). Council of Europe: official website. <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000168007a67d>

существование как человечества, так и искусственного интеллекта взаимозависимы, и обе указанные сущности обеспечивают и собственную стабильность, и развитие (в случае человечества в первую очередь социальное, для систем искусственного интеллекта – техническое) за счет сотрудничества, но не конкуренции. «Вовлеченность в рыночные отношения требует взаимного учета интересов и прав... Трудно назвать какой-либо иной, помимо взаимоиспользования, принцип, посредством которого равенство и справедливость утверждались в человеческих отношениях столь естественно и спонтанно» (Апресян, 1995). Отношения общества и систем искусственного интеллекта вряд ли можно назвать в полном смысле человеческими, однако ожидать от них взаимности, основанной на прямой и обратной полезности, вполне оправданно. Так, в Национальном стандарте «Требования по безопасности для промышленных роботов» используются следующие формулировки: «взаимодействующий с человеком робот (collaborative robot): Робот, сконструированный для непосредственного взаимодействия с человеком в рамках определенного совместного рабочего пространства; совместное рабочее пространство (collaborative workspace): Рабочее пространство, находящееся внутри защищенного пространства, где робот и человек могут выполнять работы одновременно в процессе производства»⁹. Очевидно, что робот в данных определениях является не инструментом, а субъектом взаимодействия, коллаборации, совместной работы, что, однако, пока выглядит преждевременным.

Так, успехи систем искусственного интеллекта в научной и творческой деятельности могут привести к росту доходов, которые направляются на дальнейшее развитие системы искусственного интеллекта. Проекты на основе искусственного интеллекта становятся не просто самоокупаемыми, но и сверхдоходными. Следует, однако, учитывать, что любой экономический рост ограничен как в объемах, так и во времени.

Развитие технологии может повлечь качественное изменение жизни: искусственный интеллект (к примеру, в качестве автора произведений искусства или изобретения) представляет свои интересы в суде, создает конкуренцию представителям различных профессий, что в свою очередь формирует ситуацию конкуренции и стимул для развития общества и человека.

Толерантная модель очевидным образом остается единственной доступной из приведенного списка при создании сильного искусственного интеллекта, однако она также может быть реализована в ближайшем будущем, даже в условиях отсутствия последнего: при возрастании числа ситуаций эффективного безрискового практического использования систем искусственного интеллекта возрастет доверие к ним на корпоративном и государственном уровнях. К примеру, канадская страховая компания Kanetrix.ca использует такие системы в процессе выбора пользователем приобретаемого страхового продукта. Учитывая, что для этого используются искусственные нейронные сети, прозрачности и объяснимости ожидать было бы странно, однако эти характеристики и не требовались: из

⁹ ГОСТ Р 60.1.2.2-2016/ИСО 10218-2:2011 Роботы и робототехнические устройства. Требования по безопасности для промышленных роботов. Часть 2: Робототехнические системы и их интеграция. (2016). Москва: Стандартинформ.

сопоставления с человеком система искусственного интеллекта в данном случае убедительно вышла победителем¹⁰.

Подобные состояния могут стать реальностью лишь при условии полного разрешения вопроса границ ответственности систем искусственного интеллекта, определения критериев наличия сознания и волевого компонента в принятии решения, без которых не может произойти наделение систем искусственного интеллекта признаками субъекта права.

Недостатки данной модели находятся в плоскости как публичного, так и частного права: признание обществом систем искусственного интеллекта как равного человеку субъекта невозможно без ответа на вопросы о самой сути и критериях наложения на такую систему ответственности за ее действия, т. е. отнесение ее к объекту или субъекту права. Так, В. А. Лаптев считает, что последствия действий и решений искусственного интеллекта могут рассматриваться как обстоятельство непреодолимой силы, т. е. исключающее саму постановку вопроса об ответственности, либо же введение обязательного страхования гражданской ответственности создателя системы искусственного интеллекта за вред, причиненный последними третьим лицам (Лаптев, 2017).

Если системы искусственного интеллекта (или киберфизические системы) достигнут определенного уровня когнитивных способностей, т. е. если они будут обладать очевидной моральной значимостью, такой как разум или чувствительность, тогда они, вероятно, будут претендовать на признание их морального статуса и должны иметь права, т. е. некоторую долю привилегий, претензий, полномочий или иммунитетов (Gunkel, 2018). Данная ситуация возможна лишь при значительном качественном прогрессе технологии. Очевидно, что для описания данной модели используется слишком много слов «если» и «вероятно». Это выражает степень неуверенности в возможности развития технологии искусственного интеллекта до таких пределов, однако исключать такую возможность все же нельзя. Исходя из темпов развития технологии, эксперты прогнозируют, что при самом худшем стечении обстоятельств полноценный искусственный интеллект, способный сравниться с человеком, будет разработан к концу XXI в.

Выводы

В обобщенном виде соотношение описанных моделей может быть представлено в табличной форме (табл.).

Отметим также, что данные модели не отражают последовательность развития общественных отношений в контексте технологий машинного обучения, они могут реализовываться параллельно друг другу в разных регионах, отраслях экономики, судопроизводства и т. д.

¹⁰ McWaters, R. J. et al. *Navigating Uncharted Waters. A roadmap to responsible innovation with AI in financial services*. World Economic Forum. http://www3.weforum.org/docs/WEF_Navigating_Uncharted_Waters_Report.pdf

**Соотношение моделей взаимодействия общества и права
с технологией искусственного интеллекта**

Критерий сравнения	Инструментальная модель	Ксенофобная модель	Эмпатическая модель	Толерантная модель
Условие реализации	Реализована в настоящее время	При проявлении существенных кризисов	Развитие эмпатических настроений в обществе, прогресс в робототехнике и интеллектуальных системах-ассистентах	Достижение технологической сингулярности, появление «сильного (общего)» искусственного интеллекта
Достоинства	Низкий уровень социальных и правовых рисков от использования систем искусственного интеллекта при обеспечении промышленного и интеллектуального прогресса, возможность сохранения текущих подходов к регулированию технологий	Развитие медицины, физиологии, генетики	Развитие общественной морали и гуманизма (в широком смысле)	Развитие как технологий, так и права и других социогуманитарных областей знания
Недостатки	Застой в социогуманитарных науках, отказ от концепции «сильного (общего)» искусственного интеллекта, антропоцентризм	Торможение научно-технического прогресса в области цифровых и компьютерных технологий	Уменьшение рациональности в угоду эмоциональности, трансформации эмоциональной сферы, массовые проблемы с запоминанием и вниманием	Снижение требований к эффективности систем искусственного интеллекта. Возможность использования системы искусственного интеллекта, обладающей правосубъектностью как «подставного лица», требование пересмотра большого количества правовых и иных социальных норм
Характер отношений «человек – искусственный интеллект»	Использование	Конкуренция	Сочувствие, соадаптация	Взаимоиспользование, сотрудничество
Правовые последствия	За негативные решения и действия системы искусственного интеллекта отвечает человек (оператор или разработчик)	Введение разрешительного регулирования отрасли	Повышенная правовая защита интеллектуальных систем без наделения их правосубъектностью	Наделение систем искусственного интеллекта правосубъектностью

Подводя итог, подчеркнем необходимость развития изучения потенциала систем искусственного интеллекта, в том числе их свойств при интеграции, и распространения в обществе, а также потенциальных рисков, неизбежных при этих процессах. Указанные модели отражают грани реальности, как уже существующей, так и потенциальной. Моделирование подобных ситуаций следует осуществлять дифференцированно, чему могут поспособствовать приведенные этико-правовые модели.

* Организация признана экстремистской, ее деятельность запрещена на территории Российской Федерации

Список литературы

- Апресян, Р. Г. (1995). Нормативные модели моральной рациональности. В кн. *Мораль и рациональность* (с. 94–118). Москва: Институт философии РАН.
- Бахтеев, Д. В. (2021). *Искусственный интеллект: этико-правовой подход*: монография. Москва: Проспект.
- Букатов, В. М. (2018). Клиповые изменения в восприятии, понимании и мышлении современных школьников – досадное новообразование «постиндустриального уклада» или долгожданная реанимация психического естества? *Актуальные проблемы психологического знания*, 4(49), 5–19.
- Ильин, Е. П. (2016). *Эмоции и чувства*. (2-е изд.). Санкт-Петербург: Питер.
- Лаптев, В. А. (2017). Ответственность «будущего»: правовое существо и вопрос оценки доказательств. *Гражданское право*, 3, 32–35.
- Маркс, К. (2001). *Капитал* (Т. 1). Москва: АСТ.
- Огурцов, А. П. (2006). Возможности и трудности в моделировании интеллекта. В кн. Д. И. Дубровский, В. А. Лекторский (ред.) *Искусственный интеллект: междисциплинарный подход* (с. 32–48). Москва: ИИнтелЛ.
- Семис-оол, И. С. (2019). «Заслуживающий доверия» искусственный интеллект. В сб. Д. В. Бахтеев, *Технологии XXI века в юриспруденции: мат-лы Всерос. науч.-практ. конф. (Екатеринбург, 24 мая 2019 года)* (с. 145–149). Екатеринбург: Уральский государственный юридический университет.
- Тимофеев, А. В. (1978). *Роботы и искусственный интеллект*. Москва: Главная редакция физико-математической литературы издательства «Наука».
- Хайдеггер, М. (1993). Вопрос о технике. В сб. *Время и бытие: статьи и выступления*. Москва: Республика.
- Balerna, M., & Ghosh, A. (2018). The details of past actions on a smartphone touchscreen are reflected by intrinsic sensorimotor dynamics. *Digital Med*, 1, Article 4. <https://doi.org/10.1038/s41746-017-0011-3>
- Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023, Marth 17). GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models.
- Flynn, J. R. (2009). *What Is Intelligence: Beyond the Flynn Effect*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511605253>
- Gunkel, D. J. (2018). *Robot rights*. Cambridge, MA: MIT Press.
- Kazim, E., & Koshiyama, A. S. (2021). A high-level overview of AI ethics. *Patterns*, 3(9). <https://doi.org/10.1016/j.patter.2021.100314>
- Mathur, M. B., & Reichling, D. B. (2016). Navigating a social world with robot partners: A quantitative cartography of the uncanny valley. *Cognition*, 146, 22–32. <https://doi.org/10.1016/j.cognition.2015.09.008>
- Mori, M. (2012). The uncanny valley. *IEEE Robotics & Automation Magazine*, 19(2), 98–100. <https://doi.org/10.1109/mra.2012.2192811>
- Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mokander, J., & Floridi, L. (2021). Ethics as a service: a pragmatic operationalisation of AI Ethics. *Minds and Machines*, 31, <https://doi.org/10.1007/s11023-021-09563-w>
- Scheutz, M. (2009). The Inherent Dangers of Unidirectional Emotional Bonds between Humans and Social Robots. *Workshop on Roboethics at ICRA*.
- Teasdale, T. W., & Owen, D. R. (2005). A long-term rise and recent decline in intelligence test performance: *The Flynn Effect in reverse*. *Personality and Individual Differences*, 39(4), 837–843. <https://doi.org/10.1016/j.paid.2005.01.029>
- Vardi, M. (2012). Artificial Intelligence: Past and Future. *Communications of the ACM*, 55, 5. <https://doi.org/10.1145/2063176.2063177>
- Watkins, R., & Human, S. (2023). Needs-aware artificial intelligence: AI that ‘serves [human] needs. *AI Ethics*, 3. <https://doi.org/10.1007/s43681-022-00181-5>
- Winfield, A. (2019). Ethical standards in robotics and AI. *Nature Electronics*, 2, 46–48. <https://doi.org/10.1038/s41928-019-0213-6>

Сведения об авторе



Бахтеев Дмитрий Валерьевич – доктор юридических наук, доцент, доцент кафедры криминалистики, Уральский государственный юридический университет имени В. Ф. Яковлева, руководитель группы проектов CrimLib.info

Адрес: 620137, Российская Федерация, г. Екатеринбург, ул. Комсомольская, 21

E-mail: ae@crimlib.info

ORCID ID: <https://orcid.org/0000-0002-0869-601X>

ScopusAuthorID: <https://www.scopus.com/authid/detail.uri?authorId=57208909117>

Web of Science Researcher ID:

<https://www.webofscience.com/wos/author/record/ABA-1494-2020>

Google Scholar ID: <https://scholar.google.ru/citations?user=h0zOOdcAAAAJ>

РИНЦ Author ID: https://elibrary.ru/author_items.asp?authorid=762765

Конфликт интересов

Автор заявляет об отсутствии конфликта интересов.

Финансирование

Исследование не имело спонсорской поддержки.

Тематические рубрики

Рубрика OECD: 5.05 / Law

Рубрика ASJC: 3308 / Law

Рубрика WoS: OM / Law

Рубрика ГРНТИ: 10.07.49 / Планирование и прогнозирование в праве

Специальность ВАК: 5.1.1 / Теоретико-исторические правовые науки

История статьи

Дата поступления – 25 марта 2023 г.

Дата одобрения после рецензирования – 24 апреля 2023 г.

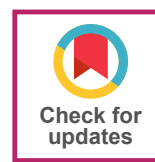
Дата принятия к опубликованию – 16 июня 2023 г.

Дата онлайн-размещения – 20 июня 2023 г.



Research article

DOI: <https://doi.org/10.21202/jdtl.2023.22>



Ethical-Legal Models of the Society Interactions with the Artificial Intelligence Technology

Dmitriy V. Bakhteev 

Ural State Law University named after V. F. Yakovlev
Ekaterinburg, Russian Federation

Keywords

Artificial Intelligence,
ChatGPT,
digital technologies,
ethics,
law,
machine learning,
model,
regulation,
robot,
society

Abstract

Objective: to explore the modern condition of the artificial intelligence technology in forming prognostic ethical-legal models of the society interactions with the end-to-end technology under study.

Methods: the key research method is modeling. Besides, comparative, abstract-logic and historical methods of scientific cognition were applied.

Results: four ethical-legal models of the society interactions with the artificial intelligence technology were formulated: the tool (based on using an artificial intelligence system by a human), the xenophobia (based on competition between a human and an artificial intelligence system), the empathy (based on empathy and co-adaptation of a human and an artificial intelligence system), and the tolerance (based on mutual exploitation and cooperation between a human and artificial intelligence systems) models. Historical and technical prerequisites for such models formation are presented. Scenarios of the legislator reaction on using this technology are described, such as the need for selective regulation, rejection of regulation, or a full-scale intervention into the technological economy sector. The models are compared by the criteria of implementation conditions, advantages, disadvantages, character of "human – artificial intelligence system" relations, probable legal effects and the need for regulation or rejection of regulation in the sector.

Scientific novelty: the work provides assessment of the existing opinions and approaches, published in the scientific literature and mass media, analyzes the technical solutions and problems occurring in the recent past and present. Theoretical conclusions are confirmed by references to applied

© Bakhteev D. V., 2023



This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

situations of public or legal significance. The work uses interdisciplinary approach, combining legal, ethical and technical constituents, which, in the author's opinion, are criteria for any modern socio-humanitarian researches of the artificial intelligence technologies.

Practical significance: the artificial intelligence phenomenon is associated with the fourth industrial revolution; hence, this digital technology must be researched in a multi-aspectual and interdisciplinary way. The approaches elaborated in the article can be used for further technical developments of intellectual systems, improvements of branch legislation (for example, civil and labor), and for forming and modifying ethical codes in the sphere of development, introduction and use of artificial intelligence systems in various situations.

For citation

Bakhteev, D. V. (2023). Ethical-legal Models of the Society Interactions with the Artificial Intelligence Technology. *Journal of Digital Technologies and Law*, 1(2), 520–539 <https://doi.org/10.21202/jdtl.2023.22>

References

- Apresyan, R. G. (1995). Normative models of moral rationality. In *Morals and rationality* (pp. 94–118). Moscow: Institut filosofii RAN. (In Russ.).
- Bakhteev, D. V. (2021). *Artificial intelligence: ethical-legal approach*. Moscow: Prospekt. (In Russ.).
- Balerna, M., & Ghosh, A. (2018). The details of past actions on a smartphone touchscreen are reflected by intrinsic sensorimotor dynamics. *Digital Med*, 1, Article 4. <https://doi.org/10.1038/s41746-017-0011-3>
- Bukatov, V. M. (2018). Clip changes in the perception, understanding and thinking of modern schoolchildren – negative neoplasm of postindustrial way or long-awaited resuscitation of the psychic nature? *Actual Problems of Psychological Knowledge*, 4(49), 5–19. (In Russ.).
- Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023, March 17). *GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models*.
- Flynn, J. R. (2009). *What Is Intelligence: Beyond the Flynn Effect*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511605253>
- Gunkel, D. J. (2018). *Robot rights*. Cambridge, MA: MIT Press.
- Heidegger, M. (1993). The Question Concerning Technology. In *Time and being: articles and speeches*. Moscow: Respublika. (In Russ.).
- Ilyin, E. P. (2016). *Emotions and feelings*. (2d ed.). Saint Petersburg: Piter. (In Russ.).
- Kazim, E., & Koshiyama, A. S. (2021). A high-level overview of AI ethics. *Patterns*, 3(9). <https://doi.org/10.1016/j.patter.2021.100314>
- Laptev, V. A. (2017). Responsibility of the “future”: legal essence and evidence evaluation issue. *Civil Law*, 3, 32–35. (In Russ.).
- Marx, K. (2001). *Capital* (Vol. 1). Moscow: AST. (In Russ.).
- Mathur, M. B., & Reichling, D. B. (2016). Navigating a social world with robot partners: A quantitative cartography of the uncanny valley. *Cognition*, 146, 22–32. <https://doi.org/10.1016/j.cognition.2015.09.008>
- Mori, M. (2012). The uncanny valley. *IEEE Robotics & Automation Magazine*, 19(2), 98–100. <https://doi.org/10.1109/mra.2012.2192811>
- Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mokander, J., & Floridi, L. (2021). Ethics as a service: a pragmatic operationalisation of AI Ethics. *Minds and Machines*, 31. <https://doi.org/10.1007/s11023-021-09563-w>
- Ogurtsov, A. P. (2006). Opportunities and difficulties in modeling intelligence. In D. I. Dubrovskii, & V. A. Lektorskii (Eds.), *Artificial intelligence: interdisciplinary approach* (pp. 32–48). Moscow: IIntELL. (In Russ.).

- Scheutz, M. (2009). The Inherent Dangers of Unidirectional Emotional Bonds between Humans and Social Robots. *Workshop on Roboethics at ICRA*.
- Semis-ool, I. S. (2019). "Trustworthy" artificial intelligence. In D. V. Bakhteev (Ed.), *Technologies of the 21st century in jurisprudence: works of the All-Russia scientific-practical conference (Yekaterinburg, May 24, 2019)* (pp. 145–149). Yekaterinburg: Uralskiy gosudarstvenniy yuridicheskiy universitet. (In Russ.).
- Teasdale, T. W., & Owen, D. R. (2005). A long-term rise and recent decline in intelligence test performance: The Flynn Effect in reverse. *Personality and Individual Differences*, 39(4), 837–843. <https://doi.org/10.1016/j.paid.2005.01.029>
- Timofeev, A. V. (1978). *Robots and artificial intelligence*. Moscow: Glavnaya redaktsiya fiziko-matematicheskoy literatury izdatelstva "Nauka". (In Russ.).
- Vardi, M. (2012). Artificial Intelligence: Past and Future. *Communications of the ACM*, 55, 5. <https://doi.org/10.1145/2063176.2063177>
- Watkins, R., & Human, S. (2023). Needs-aware artificial intelligence: AI that 'serves [human] needs'. *AI Ethics*, 3. <https://doi.org/10.1007/s43681-022-00181-5>
- Winfield, A. (2019). Ethical standards in robotics and AI. *Nature Electronics*, 2, 46–48. <https://doi.org/10.1038/s41928-019-0213-6>

Author information



Dmitriy V. Bakhteev – Doctor of Law, Associate Professor, Department of Criminology, Ural State Law University named after V. F. Yakovlev, Head of CrimLib.info project group

Address: 21 Komsomolskaya Str., 620137, Ekaterinburg, Russian Federation

E-mail: ae@crimlib.info

ORCID ID: <https://orcid.org/0000-0002-0869-601X>

ScopusAuthorID: <https://www.scopus.com/authid/detail.uri?authorId=57208909117>

Web of Science Researcher ID:

<https://www.webofscience.com/wos/author/record/ABA-1494-2020>

Google Scholar ID: <https://scholar.google.ru/citations?user=h0zOOdcAAAAJ>

РИНЦ Author ID: https://elibrary.ru/author_items.asp?authorid=762765

Conflict of interests

The author declare no conflict of interests.

Financial disclosure

The research was not sponsored.

Thematic rubrics

OECD: 5.05 / Law

PASJC: 3308 / Law

WoS: OM / Law

Article history

Date of receipt – March 25, 2023

Date of approval – April 24, 2023

Date of acceptance – June 16, 2023

Date of online placement – June 20, 2023

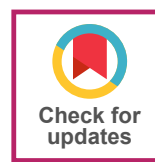


Научная статья

УДК 342.731:004.8

EDN: <https://elibrary.ru/fetbv>

DOI: <https://doi.org/10.21202/jdtl.2023.23>



Искусственный интеллект как вспомогательный инструмент ограничения религиозной свободы в Китае

Наталья Игоревна Шумакова  

Южно-Уральский государственный университет (национальный исследовательский университет)
г. Челябинск, Российская Федерация

Елена Викторовна Титова 

Южно-Уральский государственный университет (национальный исследовательский университет)
г. Челябинск, Российская Федерация

Ключевые слова

Искусственный интеллект,
Китай,
нейронная сеть,
права человека,
право,
регулирование,
религия,
свобода вероисповедания,
цифровые технологии,
экстремизм

Аннотация

Цель: на основании изучения статистики преступлений, национального законодательства и норм международного права дать правовую оценку ограничениям права на свободу вероисповедания с применением технологий искусственного интеллекта в Китае.

Методы: методологическую основу исследования составляет совокупность методов научного познания, в том числе конкретно-социологический (анализ статистических данных и иных документов), формально-юридический (изучение правовых категорий и дефиниций), формально-логические методы (анализ и синтез), общенаучные методы (индукция, дедукция) и др.

Результаты: в работе изучены предпосылки использования в Китае технологий искусственного интеллекта для контроля за общественными отношениями, возникающими в процессе религиозной активности как в цифровом пространстве, так и за его пределами; проанализирована правовая база применения указанных мер; дана правовая оценка ограничения религиозной свободы с использованием технологий искусственного интеллекта; сделан прогноз дальнейшего развития китайского законодательства и внешней политики, связанных с религиозной свободой. Дополнительно в работе проанализированы материалы правозащитных организаций, направленных на сдерживание

 Контактное лицо

© Шумакова Н. И., Титова Е. В., 2023



Статья находится в открытом доступе и распространяется в соответствии с лицензией Creative Commons «Attribution» («Атрибуция») 4.0 Всемирная (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0/deed.ru>), позволяющей неограниченно использовать, распространять и воспроизводить материал при условии, что оригинальная работа упомянута с соблюдением правил цитирования.

политики КНР по «китаизации» и «деэкстремификации» этнических и религиозных меньшинств, в том числе при помощи контроля и пропаганды с использованием современных цифровых технологий.

Научная новизна: в работе исследована попытка Китая урегулировать связанные с религиозной активностью вызовы, возникающие в процессе стремительной цифровизации общества и государства, с которыми республика сталкивается как развивающаяся, многонациональная и поликофессиональная страна. Установленные ограничения религиозной свободы с применением технологий искусственного интеллекта рассмотрены в комплексе с релевантной статистикой преступлений. Правовая оценка применения искусственного интеллекта как инструмента ограничения свободы вероисповедания дана не только с точки зрения международного права, но и с учетом норм китайского национального законодательства.

Практическая значимость: результаты проведенного исследования могут быть использованы для разработки непротиворечивой нормативной правовой базы для использования технологий искусственного интеллекта в целях противодействия экстремизму.

Для цитирования

Шумакова, Н. И., Титова, Е. В. (2023). Искусственный интеллект как вспомогательный инструмент ограничения религиозной свободы в Китае. *Journal of Digital Technologies and Law*, 1(2), 540–563. <https://doi.org/10.21202/jdtl.2023.23>

Содержание

Введение

1. Гарантии и ограничения религиозной свободы
2. Правовая база для установления контроля за религиозной активностью
3. Применение искусственного интеллекта для реализации контроля за религиозной активностью
4. Статистика преступлений как основание установления контроля за религиозной активностью
5. Применение искусственного интеллекта и цифровых технологий за контролем религиозной активности в проблемных регионах

Выводы

Список литературы

Введение

Процесс цифровизации общественных отношений переносит в новую плоскость реализацию фундаментальных прав человека, в частности, права на религиозную свободу. К. А. Бингаман, признавая этот процесс необратимым, подчеркивает, что он размывает грань между реальным и виртуальным в аспектах отправления религиозных служб и ритуалов (Bingaman, 2023). Похожей концепции придерживаются Э. Марик и Й. Марик, описывающие цифровое пространство как «современную общественную площадь», объединяющую виртуальные действия с реальными последствиями, а потому требующую особого регулирования (Marique & Marique, 2020). К. Ашраф, в свою

очередь, говорит о возможности выделения цифровой религии в самостоятельный институт (Ashraf, 2020). Эту теорию подтверждают его каталонские коллеги, называющие платформы цифровой религиозной активности «новыми общественными пространствами, способствующими созданию социальной идентичности молодых людей» (Bosch et al., 2017).

Появление и развитие онлайн-организаций, оказывающих религиозные услуги, как и общественные отношения, возникающие в процессе онлайн-коммуникации на религиозную тематику, создают дополнительные вызовы, зачастую ведущие к ограничению права на свободу вероисповедания. В частности, рассуждая о феномене онлайн-экстремизма, К. В. Бхатия, обращаясь к результатам исследований Л. Доусана и Д. Коуэна (Bhatia, 2021), отмечает, что неограниченный доступ к созданию, распространению и потреблению онлайн-информации приводит к кризису власти в цифровом пространстве, а также к кризису аутентичности, связанному с содержанием распространяемых материалов.

Указывает на рост онлайн-распространения экстремистских материалов и ОБСЕ¹, в то время как Генеральная Ассамблея ООН² призывает обратить внимание на статистику инцидентов, связанных с использованием цифровых площадок для разжигания ненависти, а также поддержки и финансирования экстремистских групп.

Необходимость защиты от угроз, связанных с растущей цифровизацией всех сфер жизни общества и государства, обсуждается и китайскими учеными. К. Уэй, например, в своей исследовательской работе «Критическая оценка правовой реакции Китая на кибертерроризм» называет кибертерроризм особым видом угрозы стабильности и безопасности государств, а существующее на сегодняшний день китайское законодательство, как и меры по его реализации, характеризует как недостаточные для обеспечения устранения этой угрозы (Wei, 2022). Столь категоричный взгляд на активность в цифровом пространстве косвенно подтверждается исследованиями распространяемого контента, ведущего к «массовой поляризации» отношения в обществе к терроризму и экстремизму (Guan & Liu, 2019).

А. Сабик-эль-Райесс также называет социальные сети в числе основных источников распространения информации, способствующей радикализации населения и, как следствие, актам насилия (Sabic-El-Rayess, 2012). Ссылаясь на заметные психологические изменения, связанные с процессом коммуникации в цифровом пространстве и ведущие к радикализации, ряд ученых в качестве мер предупреждения и противодействия предлагает применять машинное обучение не только для обнаружения определенных типов сообщений, но и для изучения лонгитюдных общедоступных данных социальных сетей с целью выявления индивидуальных изменений в публикациях и определенных типах сообщений (Smith et al., 2020).

Л. Ф. С. Менесес, также говоря о негативных психологических изменениях, называет неконтролируемое распространение непроверенной информации в цифровой

¹ ПА ОБСЕ против COVID-19. 1 этап. Размышления, предлагаемые стратегии и рекомендации, представленные Председателем ПА ОБСЕ Георгием Церетели. (2020). *OSCE PA Reports*. <https://www.oscepa.org/en/documents/president/reports-22/>

² Генеральная Ассамблея ООН. (2020, 18 июня). Расовая дискриминация и инновационные цифровые технологии: правозащитный анализ : доклад Специального докладчика по вопросу о современных формах расизма, расовой дискриминации, ксенофобии и связанной с ними нетерпимости. *Report of the Special Procedure of the Human Rights Council*. <https://digitallibrary.un.org/record/3879751?ln=ru>

среде «кризисом правды» и акцентирует внимание на деструктивном влиянии данного феномена на способность индивида к критическому мышлению (Meneses, 2021). Исследования на фокус-группе из Гонконга во время протестов 2019 г. также доказывают, что демонстрация определенных изображений и видео способна оказывать значительное давление на критическое мышление (Chan, 2019).

Если исходить из анализа материалов предыдущих исследований, становится очевидным, что невозможность применения традиционных механизмов регулирования в сети Интернет стала непосредственной причиной применения искусственного интеллекта (далее – ИИ) в качестве инструмента государственного контроля в цифровом пространстве. На возможность использования его алгоритмов для изучения религиозной активности и непосредственно самой религии указывает Р. Рид, отмечая, что сбор, сортировка и хранение информации неизбежно будут производить впечатление тотальной слежки, и предлагая задуматься о роли религии в принятии обществом тех или иных идей (Reed, 2021).

В то же время современные цифровые технологии могут рассматриваться как вспомогательные инструменты борьбы с преступлениями, направленными на разжигание ненависти по этническому и религиозному признакам, нарушение территориальной целостности и свержение государственной власти и другое за пределами цифрового пространства.

На необходимость использования ИИ органами правопорядка указывают и ряд исследователей (Kao & Sapp, 2022; Fontes et al., 2022), в работах которых упоминается о побочных эффектах злоупотребления использованием ИИ, главным среди которых является дестабилизация общества.

Целью настоящей статьи является правовая оценка применения искусственного интеллекта и иных цифровых технологий с целью контроля за религиозной активностью в Китайской Народной Республике (далее – КНР). Для ее достижения в процессе исследования были выявлены особенности реализации права на свободу вероисповедания в КНР, проанализирована статистика преступлений, связанных с экстремизмом и религиозной нетерпимостью, изучены материалы слушания, проведенного Комиссией Соединенных Штатов Америки по международной свободе вероисповедания (US Commission on International Religious Freedom (USCIRF)). В заключительной части сделана попытка прогнозирования дальнейшего развития китайских политики и законодательства по отношению к религиозной свободе.

1. Гарантии и ограничения религиозной свободы

Ст. 36 Конституции КНР устанавливает, что «граждане Китайской Народной Республики обладают свободой вероисповедания»³, закрепляя запреты на принуждение к «вере или неверию» и дискриминацию по религиозному признаку. В то же время ст. 36 содержит важную оговорку: государство обязуется защищать только «нормальную религиозную активность». Притом четкого определения термина «нормальная религиозная активность» в основном законе Китая нет, оно лишь обрисовывается.

³ Constitution of the People's Republic of China от 4 декабря 1982 г. (с изм. и доп. в ред. от 11 марта 2018 г.). *The National People's Congress of the People's Republic of China*. <https://www.npc.gov.cn/englishnpc/constitution2019/201911/1f65146fb6104dd3a2793875d19b5b29.shtml>

Исходя из текста, под этим понимается религиозная активность, не нарушающая общественный порядок, не создающая риск вреда здоровью граждан, не содержащая признаков общественной деятельности и не вмешивающаяся в государственную систему образования. Рассматриваемая статья также содержит указание на то, что религиозные группы и отношения, связанные с их деятельностью, не могут находиться под контролем иностранных сил.

В целом же в Китае существует целый пласт законов и иных нормативных правовых актов, нормы которых так или иначе направлены на обеспечение, защиту и регулирование общественных отношений, связанных с реализацией свободы вероисповедания, в числе которых Уголовный кодекс КНР (1997)⁴, Основной закон Гонконга (1997)⁵, Основной закон Макао (1999)⁶, Закон о труде (1985)⁷, Закон об обязательном образовании (1986)⁸, Закон о региональных национальных автономиях КНР (1984)⁹, Закон о собраниях и демонстрациях в КНР (1989)¹⁰, Закон о рекламе¹¹ и др.

Практически каждый нормативный правовой акт КНР гарантирует недопустимость дискриминации по религиозному признаку и неотчуждаемость религиозной свободы. В то же время, изучив китайское законодательство, можно сделать заключение о его направленности на недопустимость проникновения религии в социальные институты, более того, становится очевидным, что Китай стремится к превращению религии в инструмент государственной власти. На это еще в 2005 г., после вступления в силу первой редакции «Положений по делам религии»¹², указывала в своей работе Б. Люн (Leung, 2005).

Более поздние исследования лишь подтверждают попытки китайского правительства «китаизировать», или, иначе говоря, адаптировать религию к системе социалистических ценностей (Lavička, 2021), где конечной целью являются детальное регулирование религиозной деятельности, подавление несанкционированных религиозных групп и уменьшение влияния религии (Lin, 2018). Именно для их достижения

⁴ 中华人民共和国主席令 от 14 марта 1997 г. 新华网 (год публикации не указан). https://www.npc.gov.cn/zgrdw/npc/lfzt/rlys/2008-08/21/content_1882895.htm

⁵ Basic Law от 4 апреля 1990. *The Government of the Hong Kong Special Administrative Region, Hong Kong Basic Law Drafting Committee* (год публикации не указан). <https://www.basiclaw.gov.hk/en/basiclaw/chapter1.html>

⁶ The Basic Law of the Macao SAR от 31 марта 1993 г. (2014). *Macao government web*. https://www.zlb.gov.cn/2014-06/26/c_126677086_4.htm

⁷ 中华人民共和国劳动法 от 5 июля 1994. 中国政府门户网站, 全国人大法规库 (год публикации не указан). https://www.gov.cn/banshi/2005-05/25/content_905.htm

⁸ 中华人民共和国义务教育法 中华人民共和国主席令 38 от 4 апреля 1986 г. (2006). *China Educational and Research Network*. https://www.edu.cn/edu/zheng_ce_gs_gui/jiao_yu_fa_lv/200603/t20060303_165119.shtml

⁹ 中华人民共和国民族区域自治法 от 5 мая 1984 г. № GJXFJ-0000-2014-00084 (с изм. и доп. В ред. от 28 февраля 2001) (2005). 中国政府门户网站. https://www.gov.cn/ziliao/flfg/2005-09/12/content_31168.htm

¹⁰ 中华人民共和国集会游行示威法 от 31 октября 1989 г. № GJXFJ-0000-2014-00084 (2014, 12 мая). 国家信访局门户网站. <https://www.gjxfj.gov.cn/gjxfj/fwgj/flfg/webinfo/2014/05/1601761496620028.htm>

¹¹ 中华人民共和国广告法 от 27 октября 1994 г. (с изм. и доп. в ред. от 24 апреля 2015) (2015). 中国人大网【字体: 大 中 小】打印. https://www.gov.cn/guoqing/2021-10/29/content_5647620.htm

¹² Religious Affairs Regulations от 7 июля 2004 г. № 426. *State Council of the People's Republic of China*. <https://www.refworld.org/pdfid/474150382.pdf>

государственному регулированию подлежат абсолютно все религиозные отношения, которые в какой-либо степени затрагивают интересы государства и общества.

В качестве гарантии обеспечения свободы вероисповедания данные нормы закреплены в ст. 6 «Положений по делам религий»¹³ (2017). Кроме того, 1 марта 2022 г. в Китае в силу вступили «Меры по администрированию предоставления религиозных услуг в Интернете»¹⁴ – резонансный подзаконный нормативный правовой акт, принятый Государственным управлением по делам религий совместно с Государственным управлением информации в Интернете, Министерством промышленности и информационных технологий, Министерством общественной безопасности и Министерством государственной безопасности КНР. Он стал логическим продолжением «Закона о кибербезопасности граждан КНР»¹⁵ (2017), «Положения по делам религий»¹⁶ (2017), а также «Мер по администрированию предоставления информационных услуг в интернете»¹⁷ (2000). Очевидно, что его принятие было ответом на появление услуг по виртуальному отправлению религиозных ритуалов, расцвет которых пришелся на период пандемии коронавируса. Так, например, в Китае появились приложения для виртуального возжигания благовоний, чаты для чтения мантр и молитв и иные способы проявления религиозной активности в цифровом пространстве.

В официальном пресс-релизе к «Мерам по администрированию предоставления религиозных услуг в Интернете»¹⁸, опубликованном на сайте Комиссии по делам национальностей и религий провинции Чжэцзян, говорится о необходимости стандартизации онлайн-сервисов с религиозной тематикой. Там же Государственное управление по делам религий делает акцент на том, что цифровое пространство Китая не может рассматриваться как «особая зона религиозной деятельности» или «анклав религиозного и общественного мнения».

Вступивший в силу нормативный правовой акт¹⁹ устанавливает стандарты предоставления религиозных услуг и распространения любой религиозной информации. Так, согласно ст. 6 только зарегистрированные на территории Китая организации могут осуществлять данную деятельность, любое участие в ней иностранцев запрещается. Положения этой статьи распространяются на любую информацию,

¹³ 宗教事务条例 中华人民共和国国务院令 686 от 26 августа 2017 г. № 000014349/2017-00167. (2017, 7 сентября). 中华人民共和国国务院令. https://www.gov.cn/zhengce/content/2017-09/07/content_5223282.htm

¹⁴ 互联网宗教信息服务管理办法 от 3 декабря 2021 г. № 17. (2021). 国家互联网信息办公室、中华人民共和国工业和信息化部、中华人民共和国公安部、中华人民共和国国家安全部令. https://www.gov.cn/gongbao/content/2022/content_5678093.htm

¹⁵ 中华人民共和国网络安全法 от 7 ноября 2016 г. (2016). 新华社北京. https://www.gov.cn/xinwen/2016-11/07/content_5129723.htm

¹⁶ 宗教事务条例 中华人民共和国国务院令 686 от 26 августа 2017 № 000014349/2017-00167. (2017, 7 сентября). 中华人民共和国国务院令. https://www.gov.cn/zhengce/content/2017-09/07/content_5223282.htm

¹⁷ 互联网信息服务管理办法 от 25 сентября 2000 № 292 (с изм. и доп. в ред. от 8 января 2011). 日中华人民共和国国务院令第. https://www.gov.cn/zhengce/2020-12/26/content_5574367.htm

¹⁸ 国家宗教事务局相关负责人就《互联网宗教信息服务管理办法》答记者问 от 28 февраля 2022 г. № 002482103/2022-00007. (2022, 28 февраля). 浙江省民宗委. https://mzw.zj.gov.cn/art/2022/2/28/art_1229468422_2394644.html

¹⁹ 互联网宗教信息服务管理办法 от 3 декабря 2021 г. № 17. (2021). 国家互联网信息办公室、中华人民共和国工业和信息化部、中华人民共和国公安部、中华人民共和国国家安全部令. https://www.gov.cn/gongbao/content/2022/content_5678093.htm

связанную с религией, в том числе визуальную, распространение которой происходит через интернет-сайты, приложения, форумы, блоги, микроблоги, мессенджеры, онлайн-трансляции и т. д. Притом необходимо помнить, что в Китае официально признаются только пять религий²⁰: буддизм, даосизм, ислам, протестантизм и католицизм. А это означает, что представители других конфессий, прежде занимавшиеся миссионерской деятельностью и отправляющие религиозные услуги на территории Китая онлайн, отныне лишены такой возможности.

Международная реакция на разработку и применение подобных мер не заставила себя ждать. 13 сентября 2022 г. Комиссия Соединенных Штатов Америки по международной свободе вероисповедания (US Commission on International Religious Freedom (USCIRF)) провела слушания²¹, в ходе которых в адрес Китая прозвучали обвинения в препятствии реализации права на свободу вероисповедания на трех уровнях: публичном, частном и цифровом.

Из показаний свидетелей слушания следовало, что китайское правительство применяет современные цифровые технологии:

- для отслеживания местонахождения верующих;
- отслеживания любой активности пользователя как в цифровом пространстве, так и в реальной жизни;
- распознавания речи;
- распознавания лиц;
- сбора информации о ведении религиозных групп в сети Интернет.

Из материалов слушания следует, что вступление в силу «Мер по администрированию предоставления религиозных услуг в Интернете»²², по заявлению обвиняющей стороны, способно усугубить религиозные гонения в стране. КНР также обвинили в том, что избранный курс на укрепление народного единства и «китаизацию» религий на самом деле нацелен на истребление традиционной культуры и ценностей. Применение же современных технологий, в том числе технологий ИИ в качестве вспомогательного инструмента реализации данного нормативного правового акта, было охарактеризовано как проявление «цифрового авторитаризма». Кроме того, сторона обвинения потребовала применения к Китаю санкций и мер, направленных на нарушение его суверенитета, включая вмешательство США во внутреннюю политику страны.

Ранее аналогичные обвинения были предъявлены КНР Управлением Верховного комиссара ООН по правам человека²³ в докладе о нарушении прав человека в Синьцзяне.

²⁰ 中国宗教概况. 宗教局, 中国政府门户网站 (год публикации не указан). https://www.gov.cn/test/2005-06/22/content_8406.htm

²¹ Hearing Before the Congressional-Executive Commission on China. September 13, 2022. (2023). *Control of Religion in China through Digital Authoritarianism*. Washington: U.S. Government publishing office. <https://www.govinfo.gov/content/pkg/CHRG-117jhr48647/pdf/CHRG-117jhr48647.pdf>

²² 互联网宗教信息服务管理办法 от 3 декабря 2021 № 17. (2021). 国家互联网信息办公室、中华人民共和国工业和信息化部、中华人民共和国公安部、中华人民共和国国家安全部令. https://www.gov.cn/gongbao/content/2022/content_5678093.htm

²³ OHCHR Assessment of human rights concerns in the Xinjiang Uyghur Autonomous Region, People's Republic of China от 31 августа 2022. (2022). UN Human Rights Office. <https://www.ohchr.org/sites/default/files/documents/countries/2022-08-31/22-08-31-final-assessment.pdf>

2. Правовая база для установления контроля за религиозной активностью

Общим основанием применения технологий ИИ для обеспечения цифровой безопасности и борьбы с киберпреступностью, в том числе с распространением экстремистских религиозных материалов, являются следующие статьи Закона о кибербезопасности граждан КНР²⁴:

- ст. 21, согласно которой операторы связи обязаны классифицировать и хранить в течение полугода журналы сетевой активности пользователей;
- ст. 46, запрещающая создавать сайты и коммуникационные группы для использования в противозаконных целях либо для распространения противозаконной информации;
- ст. 47, обязующая операторов связи усилить управление информацией, публикуемой пользователями, и немедленно прекращать передачу данных в случае выявления противоправной активности, а затем принять меры по обработке и ликвидации данных, нарушающих законодательство страны. Сохраненные доказательства должны быть направлены в компетентный орган;
- ст. 51, утверждающая создание государственной системы мониторинга цифровой безопасности. Как следует из данной статьи, государственные департаменты кибербезопасности и информатизации обязаны осуществлять общую координацию других релевантных департаментов с целью улучшения эффективности сбора, анализа и представления информации о кибербезопасности с целью раннего предупреждения рисков нарушения цифровой безопасности;
- ст. 58, позволяющая применять временные меры к сетевым коммуникациям в определенных районах, в том числе меры по их ограничению для защиты национальной безопасности и общественного порядка, а также в качестве реагирования на крупные происшествия, подрывающие безопасность общества;
- ст. 68, предусматривающая ответственность для операторов связи за неприменение мер реагирования в случае выявления распространения противозаконной информации.

Данные нормы Китай намерен применять и к регулированию генеративного ИИ. Например, ст. 12 проекта «Мер по управлению услугами генеративного ИИ»²⁵ приравнивает поставщиков таких услуг к производителям контента и обязывает следить, чтобы сгенерированные тексты, фото- и аудиовизуальные материалы не были дискриминирующими по какому-либо признаку. В этом же проекте ст. 19 предусматривает ответственность и для пользователей услуг генеративного ИИ за создание противоправных материалов.

3. Применение искусственного интеллекта для реализации контроля за религиозной активностью

Не меньший интерес представляет и то, как именно планируется реализовывать положения «Мер по администрированию предоставления религиозных услуг в Интернете»²⁶. В вышеупомянутом официальном пресс-релизе говорится, что заниматься

²⁴ 中华人民共和国网络安全法 от 7 ноября 2016 г. (2016). 新华社北京. https://www.gov.cn/xinwen/2016-11/07/content_5129723.htm

²⁵ 生成式人工智能服务管理办法（征求意见稿） от 11 апреля 2023 г. (2023). 国家互联网信息办公室. https://www.cac.gov.cn/2023-04/11/c_1682854275475410.htm

²⁶ 互联网宗教信息服务管理办法 от 3 декабря 2021 г. № 17. (2021). 国家互联网信息办公室、中华人民共和国工业和信息化部、中华人民共和国公安部、中华人民共和国国家安全部令. https://www.gov.cn/gongbao/content/2022/content_5678093.htm

этим будут совместно Государственное управление по делам религий, Государственное управление интернет-информации, Министерство промышленности и информационных технологий, Министерство общественной безопасности и Министерство государственной безопасности, а значит, и аффилированные с ними компании, например, такие как:

- iFLYTEK Co., Ltd.²⁷ – китайская государственная компания, применяющая ИИ для изучения и распознавания речи, включая редкие языки и диалекты (в том числе тибетский и уйгурский);

- CloudWalk Technology Co., Ltd.²⁸ – занимающаяся технологиями распознавания человеческого лица, тела и голоса, данная компания, как следует из информации, размещенной на ее официальном сайте, также занимается профилированием и изучением паттернов поведения;

- China Electronics Technology Group Corporation²⁹ – еще одна государственная компания, занимающаяся в том числе разработками программного обеспечения и инновациями в области обеспечения безопасности и применения ИИ;

- Qihoo 360³⁰ – провайдер бесплатной интернет- и мобильной безопасности, которой принадлежат 360 Total Security, 360 Mobile Security, 360 Security и другие продукты для обеспечения национальной безопасности с применением технологий ИИ. Компания также занимается обучением специалистов в данной отрасли;

- Huawei Technologies³¹ – один из крупнейших производителей телекоммуникационного оборудования в мире. Из размещенных на официальном сайте компании в общем доступе рекламных материалов следует, что их разработки могут помочь идентифицировать индивидов по голосу, управлять идеологическим перевоспитанием заключенных, отслеживать политических активистов, использовать технологию распознавания лиц для отслеживания покупателей и т. д. Компания также занимается разработками программного обеспечения, позволяющего использовать сканирование лица в целях передачи информации органам правопорядка.

Помимо этого, не стоит забывать, что технологии ИИ для контроля за религией используются и вне пределов цифрового пространства. В частности, Hangzhou Hikvision Digital Technology Co., Ltd., если верить данным Комиссии США по международной религиозной свободе³² за 2019 г., к 2035 г. планирует установить камеры в 967 мечетях, что позволит им не только отслеживать при помощи системы распознавания лиц всех посетителей, но и следить за тем, насколько проповеди отвечают букве закона. Исходя из положений ст. 9 «Положения Синьцзян-Уйгурского автономного

²⁷ About iFLYTEK Create A Better World With A.I. <https://global.iflytek.com>

²⁸ CloudWalk Technology Co., Ltd. <https://www.cloudwalk.com>

²⁹ China Electronics Technology Group Corporation. https://www.cetcei.com/enzgdkj/about_us/introduction29/index.html

³⁰ Qihoo 360. <https://www.360.cn/>

³¹ Nowhere to hide: Building safe cities with technology enablers and AI. <https://www.huawei.com/en/huaweitech/publication/winwin/AI/nowhere-to-hide>

³² Religious Freedom in China's High-Tech Surveillance State. (2019, September). USCIRF Country Update: China. <https://www.uscifr.gov/publications>

района о деэкстремификации»³³ (2017), ИИ будет распознавать следующие признаки экстремизма:

- ношение бурки, закрывающей лицо, либо попытки заставить других носить данный вид одежды, а также ношение иных экстремистских символов (п. 7);
- распространение идей религиозного фанатизма посредством ношения большой бороды и/или использования определенных имен (п. 8);
- заключение брака либо расторжение одного по религиозным обычаям без государственной регистрации.

Отдельного внимания заслуживает п. 5 рассматриваемого акта, где в качестве проявления экстремизма указывается отказ от пользования телевидением и радио, а это предполагает наличие автоматизированного мониторинга данного вида активности.

4. Статистика преступлений как основание установления контроля за религиозной активностью

Контроль за религиозной активностью в проблемных регионах, в особенности в Синьцзян-Уйгурском автономном округе (далее – СУАР), как следует из Белой книги китайского правительства «Борьба с терроризмом и экстремизмом и защита прав человека в СУАР»³⁴ (2019), является частью государственной программы по борьбе с терроризмом и экстремизмом, направленной на защиту представителей других конфессий и этнических групп и защиту прав человека в целом.

В этом же документе приводится следующая статистика преступлений, сопряженных с сепаратизмом и религиозным экстремизмом, с примечанием о том, что она не является полной:

- а) террористические акты и угроза общественной безопасности:
 - 05.02.1992 – подрыв двух пассажирских автобусов в Урумчи (3 убитых, 23 раненых);
 - 25.02.1997 – подрыв трех пассажирских автобусов в Урумчи (9 убитых, 68 раненых);
 - 23.05.1998 – террористы разместили более 40 самовоспламеняющихся устройств в людных местах в Урумчи, что привело к 15 случаям поджогов;
 - 07.03.2008 – попытка взрыва борта на рейсе CZ6901 из Урумчи в Пекин;
 - 30.06.2011 – в Кашгаре двое террористов завладели грузовиком, убив водителя, а затем умышленно совершили наезд на толпу. После этого они напали с холодным оружием на пешеходов (8 убитых, 27 раненых);
 - 31.06.2011 – в Кашгаре террористы напали с холодным оружием на пешеходов (6 убитых, 16 раненых);
 - 29.07.2012 – попытка захвата борта на рейсе GS7554 из Хотана в Урумчи;

³³ 新疆维吾尔自治区去极端化条例 от 27 марта 2017 г. № mzi/2020-00007 (с изм. и доп. в ред. от 9 мая 2019 г.) (2019). 新疆人大网 更新时间. <https://www.xjpcsc.gov.cn/article/225/lfgz.html>

³⁴ 新疆维吾尔自治区打击恐怖主义和极端主义和保护人权 从 2019 年 3 月 (2021). 中华人民共和国国务院新闻办公室. https://geneva.china-mission.gov.cn/chn/ztjs/ajljdiowqjknew/baipishu/202110/t20211014_9587970.htm

- 28.10.2013 – теракт в центре Пекина: террористы направили джип с 31 бочкой бензина в сторону туристов и дежурных полицейских, после чего подожгли его (2 убито, 40 ранено);
- 01.03.2014 – террористы с холодным оружием атаковали людей на железнодорожной станции в Куньмине (31 убит, 141 ранен);
- 30.04.2014 – один террорист атаковал пассажиров на железнодорожной станции Урумчи, второй в это же время привел в действие взрывное устройство (3 убито, 79 ранено);
- 22.05.2014 – пятеро террористов на легковых автомобилях въехали в толпу в Урумчи, после чего привели в действие взрывное устройство (39 убито, 94 ранено);
- 18.09.2015 – террористы атаковали угольную шахту в округе Аксу (16 убито, 18 ранено);
- б) убийства религиозных лидеров:
 - 24.08.1993 – двое террористов предприняли попытку убийства имама Муллаха Абулицы в Кашкаре;
 - 22.03.1996 – двое в масках застрелили вице-президента Исламской Ассоциации Уезда Токсу и помощника имама Акемусидике Аджи;
 - 12.05.1996 – предпринята попытка убийства вице-президента исламской ассоциации Китая, президента Исламской ассоциации Синьцзяна и хатипа мечети Ид Ках в Кашгаре Аронгхан Аджи;
 - 06.10.1997 – был застрелен старший мулла Юнуси Сидике, член Китайской исламской ассоциации, президент Аксуской исламской ассоциации и имам Большой мечети уезда Бай;
 - 27.01.1998 – застрелен Абулиджи Аджи, имам Большой мечети уезда Бай;
 - 30.10.2014 – старший мулла Джума Тайер, вице-президент Исламской ассоциации Синьцзяна и имам мечети Ид Ках, был зверски убит тремя террористами по пути домой после утренней молитвы Фаджр;
- в) нападения на государственные органы и общественные беспорядки:
 - 05.04.1990 – группа террористов в количестве более 200 человек совершила нападение на здание правительства поселка Барен Актюньского района Кызылсуйской Киргизской автономной префектуры, похитив 10 человек и убив 6 сотрудников полиции;
 - 27.08.1996 – атакована резиденция правительства поселка Цзянгэлеси округа Ечэн префектуры Кашгар, убиты заместитель главы поселка, дежурный полицейский и трое охранников;
 - 05–08.02.1997 – в ходе беспорядков в Инине 7 человек погибли и 198 получили ранения;
 - 24.10.1999 – вооруженное нападение на полицейский участок в поселке Саили уезда Зепу префектуры Кашгар. Один охранник общественной безопасности убит, второй ранен. Также ранения получили один находящийся под стражей и один полицейский;
 - 05.07.2009 – беспорядки в Урумчи: тысячи террористов совершили нападения на мирных жителей, правительственные органы, сотрудников органов общественной безопасности и полиции, жилые дома, магазины и объекты общественного транспорта, в результате чего 197 человек погибло и более 1700 человек получили ранения;
 - 04.08.2008 – террористы въехали на угнанном самосвале в пограничников, участвовавших в учениях в городе Кашгар, после чего бросили самодельные гранаты. 16 человек погибли и 16 получили ранения;

– 23.04.2013 – общественные работники обнаружили террористов, производящих взрывчатку в частном доме в Селибуя, и были убиты на месте. Далее террористы совершили нападение на сотрудников местных органов власти и полицейских, в результате чего 15 человек погибли и 2 получили тяжелые ранения;

– 26.06.2013 – совершены нападения на полицейский участок, патрульных, резиденцию местного правительства и строительные площадки в Лукэцине – 24 человека погибли и 25 получили ранения;

– 28.07.2014 – террористы с ножами и топорами напали на здание правительства и полицейский участок города Айликсиху. Затем некоторые из них двинулись в город Хуанди, где они напали на мирных жителей – 37 человек погибли и 13 получили ранения;

– 21.09.2014 – совершены нападения на полицейский участок и фермерский рынок города Янся, полицейский участок города Тирекебажа и магазин в уездном центре Лунтай Байингольской автономной префектуры Монголии – 10 человек погибло, 54 получили ранения;

– 28.12.2016 – террористы въехали во двор здания правительства уезда Моюй, привели в действие самодельное взрывное устройство и напали на сотрудников правительства – 2 человека погибли и 3 получили ранения.

Приведенную в указанной Белой книге правительства³⁵ статистику не представляется возможным ни подтвердить, ни дополнить из иных официальных источников, так как доступ к данным либо заархивирован, либо ограничен для иностранцев. В целом возникают сомнения относительно точности статистических данных, находящихся в открытом доступе. Так, исходя из материалов, приведенных Бюро по борьбе с терроризмом и противодействию насильственному экстремизму³⁶, в Китае не было зарегистрировано ни одного террористического акта или актов экстремизма, сопряженных с применением насилия, с 2016 г. Чему противоречит ранее приведенное сообщение о взрыве самодельного устройства и гибели пятерых человек недалеко от уезда Пишан в СУАР в 2017 г.³⁷ Отчет Народной прокуратуры СУАР за 2017 г.³⁸ никак не выделяет данный инцидент в общей статистике преступлений.

Согласно отчету правозащитной организации Human Rights Watch³⁹, со ссылкой на предположительно официальную статистику, предоставленную КНР, с 2017 по 2022 г., на февраль 2022 г. общее число обвинительных приговоров по статьям, связанным с экстремизмом, в данном регионе составило 540 826. В этом же отчете еще раз делается акцент на том, что Китай регулярно засекречивает полные данные, в том числе полностью закрыт доступ к приговорам судов и невозможно получить

³⁵ 新疆维吾尔自治区打击恐怖主义和极端主义和保护人权 от марта 2019 г. (2021). 中华人民共和国国务院新闻办公室. https://geneva.china-mission.gov.cn/chn/ztjs/ajljdiowqjknew/baipishu/202110/t20211014_9587970.htm

³⁶ Bureau of Counterterrorism. (2020). *Country Reports on Terrorism 2020: China (Hong Kong and Macau)*. <https://www.state.gov/reports/country-reports-on-terrorism-2020/china>

³⁷ Bureau of Counterterrorism. (2020). *Country Reports on Terrorism 2017: Country Reports on Terrorism*. <https://www.state.gov/reports/country-reports-on-terrorism-2017/>

³⁸ 自治区人民检察院工作报告 от 23 января 2018. (2018). 新疆维吾尔自治区人民检察院. <https://archive.ph/5MLLE>

³⁹ China: Xinjiang Official Figures Reveal Higher Prisoner Count от 14 сентября 2022. (2022). *Human Rights Watch*. <https://www.hrw.org/news/2022/09/14/china-xinjiang-official-figures-reveal-higher-prisoner-count>

информацию, за совершение каких именно преступлений, сопряженных с экстремизмом, эти приговоры были вынесены.

Довольно впечатляющую статистику приводит в уже ранее упомянутом исследовании К. Уэй (Wei, 2022), согласно которой, в XXI в. терроризм в Китае перешел в цифровое пространство и основными его проявлениями на всей территории страны стали вспомогательные действия, ведущие к таким реальным террористическим актам и нарушениям общественного порядка, как использование Интернета для распространения противоправной информации, подстрекательства к совершению преступлений и пропаганды, связанной с экстремизмом (64 %), владение аудиовизуальными материалами или другими предметами, которые могут быть использованы для совершения терактов (33,9 %).

5. Применение искусственного интеллекта и цифровых технологий за контролем религиозной активности в проблемных регионах

Говоря о СУАР, следует учитывать, что основой установления цифрового контроля за реализацией права на свободу вероисповедания там является локальный нормативный правовой акт «Положение Синьцзян-Уйгурского автономного района о деэкстремификации»⁴⁰ (2017). В том числе п. 13 ст. 9 данного акта запрещает «публикацию, распространение, скачивание, хранение аудио- и видеоматериалов экстремистского содержания», равно как и доступ к таким материалам, а ст. 26 наделяет департаменты телекоммуникаций и операторов телекоммуникаций правом заниматься мониторингом систем и применять технические меры к обмену голосовыми текстовыми сообщениями, разговорам по мобильной и телефонной связи и использованию других телекоммуникационных инструментов. В случае выявления информации экстремистского содержания оператор обязан прервать трансляцию, удалить все релевантные материалы, зафиксировать доказательства и доложить о произошедшем. В то же время ст. 3 дает определение религиозному экстремизму – это «распространение радикальной религиозной идеологии посредством высказываний и действий, отрицание нормального производства и жизнедеятельности и воспрепятствование им» и «предложения и действия с использованием искажения религиозных учений или других средств для разжигания ненависти или дискриминации и пропаганды насилия». «Положение Синьцзян-Уйгурского автономного района о деэкстремификации»⁴¹ является частью кампании «Мощный удар по терроризму», разработанной в целях борьбы с экстремизмом и терроризмом в СУАР⁴².

Полагаем, что технологии ИИ будут применяться и для цифрового контроля в Тибете. Несмотря на то, что общественные беспорядки в этом регионе не принято относить к террористической активности, учитывая его исторические взаимоотношения с Китаем, последний опасается нового роста сепаратистских настроений в нем.

⁴⁰ 新疆维吾尔自治区去极端化条例 от 27 марта 2017 № mzi/2020-00007 (с изм. и доп. в ред. от 9 мая 2019) (2019). 新疆人大网 更新时间. <https://www.xjpcsc.gov.cn/article/225/lfgz.html>

⁴¹ Там же.

⁴² 公安部开展严厉打击暴力恐怖活动专项行动 (2014, May 25). 中央政府门户网站//新华社. https://www.gov.cn/xinwen/2014-05/25/content_2686705.htm

В качестве основной предпосылки таких опасений здесь нужно рассматривать протесты, устроенные буддийскими монахами в марте 2008 г. в Лхасе и переросшие в массовые беспорядки, во время которых, помимо всего прочего, был совершен поджог главной мечети города (Van Wie Davis, 2009)⁴³. На сегодняшний день достаточно проблематично получить доступ к официальной статистике, но Ван Ви Дэйвис в своем исследовании этого вопроса, со ссылкой на данные, предоставленные местным правительством, указывает, что во время протестов погибло 16 человек, четверо из которых – работники отеля и хозяева магазинов, умышленно сожженные заживо.

Также в марте 2008 г. протестующие с самодельными взрывными устройствами атаковали офисы правительства, полицейские участки, госпитали, школы, банки и рынки под лозунгом «Свободу Тибету!» в юго-западной части провинции Сычуань. Ван Ви Дэйвис, со ссылкой на данные полиции, указывает, что нападения были организованы буддийскими монахами и направлены в том числе против мусульманской части населения, в связи с чем 15 марта 2008 г. полиция была вынуждена оцепить мусульманский квартал Лхасы⁴⁴.

Правительство Китая приняло ряд превентивных мер для дерадикализации Тибета, из которых наибольший интерес для данного исследования представляет вступившее в силу «Положение о создании образцовой зоны этнического единства и прогресса в Тибетском автономном районе»⁴⁵ (2020). Положение это направлено в первую очередь на уже упомянутую ранее «китаизацию» религии и интеграции в нее идеи этнического единства, как следует из ст. 19 данного нормативного акта. К идее национального единства и прогресса обязаны призывать и онлайн-агентства – эта норма закреплена в ст. 27. На санкции же указывает ст. 46 – наказание предусмотрено за подрыв идеи национального единства и прогресса путем распространения слухов, производства информации либо посредством занятия решительной противостоящей позиции. Отдельного внимания заслуживает ст. 24, согласно которой озвученные выше идеи тибетцы должны продвигать в семьях, чтобы, как уже следует из ст. 34, воспитывать образцовых граждан.

Выводы

Результаты проведенного исследования выявили, что Китай стремится создать правовую базу для государственного регулирования общественных отношений, возникающих в процессе реализации права на свободу вероисповедания, в целях сохранения народного единства, общественного порядка и территориальной целостности страны. В то же время китайский законодатель пытается своевременно отвечать на такие вызовы современности, как распространение экстремистских материалов и радикализация населения в цифровом пространстве, в частности, посредством создания и распространения противоправного контента с использованием

⁴³ Ввиду отсутствия доступа к официальным данным приводится статистика из исследования Van Wie Davis, E. Tibetan separatism in China. (2009). *Korean Journal of Defense Analysis*, 21(2), 155–170. <https://doi.org/10.1080/10163270902872135>

⁴⁴ Там же.

⁴⁵ 西藏自治区民族团结进步模范区创建条例 第四章 宣传教育 от 11 января 2020 № mzj/2020-00007. (2020). 地区民宗局. <https://www.xz.gov.cn/info/1258/28851.htm>

генеративного ИИ. Реализация же законов и иных нормативных правовых актов, направленных на контроль и ограничение религиозной свободы, включая так называемые китаизацию и деэкстремификацию населения отдельных регионов, обеспечиваются, помимо прочего, применением технологий ИИ. Перечисленные меры обусловлены существованием реальных угроз территориальной целостности и стабильности КНР, а также безопасности граждан этой страны, что следует из результатов проведенного анализа статистики преступлений на почве этнической и религиозной нетерпимости, сопряженных с массовыми беспорядками и террористическими актами, конечной целью которых являлся выход определенных регионов из состава государства.

Согласно ст. 36 Конституции КНР⁴⁶, государство обязуется защищать только «нормальную религиозную активность», т. е. такой вид реализации религиозной свободы, который не нарушает общественного порядка, не угрожает жизни и здоровью граждан, не подстрекает к свержению государственного строя и т. д. Данная статья является отправной точкой для всей китайской политики и, соответственно, законодательства в отношении религии. Таким образом, принятые на основе ст. 36 законы и иные нормативные правовые акты изначально обязаны соответствовать Конституции.

Не противоречащим основному закону, на первый взгляд, кажется и применение ИИ с целью распознавания содержания текстовых и аудиовизуальных сообщений, в том числе, как можно полагать из ст. 34 «Положения о создании образцовой зоны этнического единства и прогресса в Тибетском автономном районе»⁴⁷, в личных беседах между членами семей, ведь ст. 40 Конституции КНР предусматривает ограничение тайны переписки в случаях, необходимых для обеспечения национальной безопасности. Тем не менее та же статья четко устанавливает, что осуществлять такую деятельность могут исключительно органы уголовного розыска, органы общественной безопасности или органы прокуратуры, но ни в коем случае не организации и физические лица. Привлечение сторонних организаций и специалистов для реализации «Мер по администрированию предоставления религиозных услуг в Интернете», которое, как следует из официального пресс-релиза к данному нормативному правовому акту, планируется совместно Государственным управлением по делам религий с Государственным управлением информации в Интернете, противоречит данной норме.

Более того, превентивные меры по распространению экстремистских материалов, предполагающие вмешательство в переписку и другие виды общения посредством телекоммуникации, предусмотренные «Мерами по администрированию предоставления религиозных услуг в Интернете»⁴⁸, «Мерами по администрированию

⁴⁶ Constitution of the People's Republic of China от 4 декабря 1982. С изм. и доп. в ред. от 11 марта 2018. *The National People's Congress of the People's Republic of China*. <https://www.npc.gov.cn/englishnpc/constitution2019/201911/1f65146fb6104dd3a2793875d19b5b29.shtml>

⁴⁷ 西藏自治区民族团结进步模范区创建条例 第四章 宣传教育 от 11 января 2020 № mzj/2020-00007. (2020). 地区民宗局. <https://www.al.gov.cn/info/1258/28851.htm>

⁴⁸ 互联网宗教信息服务管理办法 от 3 декабря 2021 г. № 17. (2021). 国家互联网信息办公室、中华人民共和国工业和信息化部、中华人民共和国公安部、中华人民共和国国家安全部令. https://www.gov.cn/gongbao/content/2022/content_5678093.htm

предоставления информационных услуг в Интернете»⁴⁹ и Законом о кибербезопасности⁵⁰, противоречат нормам международного права. В частности, ст. 12 Всеобщей декларации прав человека⁵¹, не допускающей вмешательства в частную жизнь и тайну корреспонденции, а также ст. 19, согласно которой каждый имеет право на свободу убеждений и распространения своих идей.

Отдельно стоит отметить, что указанные нормативные правовые акты, а также «Положение о создании образцовой зоны этнического единства и прогресса в Тибетском автономном районе»⁵² и «Положение Синьцзян-Уйгурского автономного района о деэкстремификации»⁵³ в купе с самой ст. 36 Конституции КНР находятся в противоречии со ст. 18 Всеобщей декларации прав человека⁵⁴, гарантирующей каждому свободу мысли, совести и вероисповедания, соответственно, нормам международного права не отвечают и попытки реализовать их в цифровом пространстве, в том числе такие, как использование генеративного ИИ для распространения государственной идеологии и иных видов пропаганды, целью которых является воздействие на формирование убеждений граждан.

Проведенное исследование также показало, что контроль за религиозной активностью в Китае имеет разную степень в зависимости от региона и предусматривает особые, усиленные меры в этнических автономиях, таких как СУАР и Тибет. Исходя из этого, очевидным становится, что «Положение о создании образцовой зоны этнического единства и прогресса в Тибетском автономном районе»⁵⁵ и «Положение Синьцзян-Уйгурского автономного района о деэкстремификации»⁵⁶ находятся в противоречии с нормами ст. 33 Конституции КНР, согласно которой все граждане Китая равны перед законом. Соответственно, любые меры по их применению, включая использование «умных» камер в отдельных регионах, помогающих собирать и сортировать информацию о внешнем виде, паттернах поведения прихожан и т. д., а также содержания речей во время отправляемых религиозных ритуалов и служб, противоречат основному закону государства и расходятся с нормами международного права, в частности, с положениями Международной конвенции о ликвидации всех

⁴⁹ 互联网信息服务管理办法 от 25 сентября 2000 г. № 292 (с изм. и доп. в ред. от 8 января 2011). 日中华人民共和国国务院令第. https://www.gov.cn/zhengce/2020-12/26/content_5574367.htm

⁵⁰ 中华人民共和国网络安全法 от 7 ноября 2016. (2016). 新华社北京. https://www.gov.cn/xinwen/2016-11/07/content_5129723.htm

⁵¹ Всеобщая декларация прав человека от 10 декабря 1948 года. КонсультантПлюс. https://www.consultant.ru/document/cons_doc_LAW_120805/

⁵² 西藏自治区民族团结进步模范区创建条例 第四章 宣传教育 от 11 января 2020 № mztj/2020-00007. (2020). 地区民宗局. <https://www.al.gov.cn/info/1258/28851.htm>

⁵³ 新疆维吾尔自治区去极端化条例 от 27 марта 2017 № mztj/2020-00007 (с изм. и доп. в ред. от 9 мая 2019 г.). (2019). 新疆人大网 更新时间. <https://www.xjpcsc.gov.cn/article/225/lfgz.html>

⁵⁴ Всеобщая декларация прав человека от 10 декабря 1948 года. КонсультантПлюс. https://www.consultant.ru/document/cons_doc_LAW_120805/

⁵⁵ 西藏自治区民族团结进步模范区创建条例 第四章 宣传教育 от 11 января 2020 № mztj/2020-00007. (2020). 地区民宗局. <https://www.al.gov.cn/info/1258/28851.htm>

⁵⁶ 新疆维吾尔自治区去极端化条例 от 27 марта 2017 № mztj/2020-00007 (с изм. и доп. в ред. от 9 мая 2019 г.). (2019). 新疆人大网 更新时间. <https://www.xjpcsc.gov.cn/article/225/lfgz.html>

форм расовой дискриминации⁵⁷. С нормами, закрепленными в данной конвенции, ратифицированной Китаем, не согласуется и политика народного единства, предполагающая принудительную «китаизацию» этнических меньшинств и их верований.

Об особенностях мер, помогающих гомогенизировать Китаю свое население, упоминается в совместном исследовании С. Джименез-Товар и М. Лавичка, посвященном изучению внедрения правительством определенных идей с использованием аудиовизуальных и других средств, влияющих на формирование национальной идентичности (Jiménez-Tovar & Lavička, 2020).

В другом своем исследовании М. Лавичка (Lavička, 2021) доказывает, что в СУАР действуют более строгие правила, которые он истолковывает как попытку искоренить религию в регионе, в том числе посредством отлучения молодого поколения от религиозных традиций и обычаев. Впрочем, важно понимать, что этот процесс начался задолго до появления в Китае пропаганды социалистических ценностей и берет свое начало в конфуцианстве (Ma, 2006). Тем не менее следование этому курсу в XXI в. вступает в прямое противоречие со ст. 4 указанной конвенции, направленной на борьбу с идеями этнического превосходства. Впрочем, было бы несправедливо утверждать, что ее положения не исполняются китайским правительством хотя бы частично. Например, из текста Белой книги «Борьба с терроризмом и экстремизмом и защита прав человека в СУАР»⁵⁸ следует, что Китай исполняет обязанности по противодействию распространения идей этнического и религиозного превосходства уйгурскими мусульманами, но сопротивление их распространению с использованием современных цифровых технологий для пропаганды схожих же идей нельзя рассматривать в качестве адекватной меры реагирования, отвечающей требованиям данной конвенции и нормам обычного международного права.

Исходя из выводов исследования, можно утверждать, что в Китае существует развитая правовая база для применения технологий ИИ в качестве инструмента регулирования религиозной активности, а его использование государственными органами и иными организациями в данных целях отвечает нормам китайского законодательства. В то же время само китайское законодательство содержит ряд коллизий и пробелов, что может препятствовать корректной реализации норм, направленных как на обеспечение гарантии прав и свобод граждан, так и на их ограничение. Более того, ряд национальных китайских законов и нормативных правовых актов, во исполнение которых предполагается применять ИИ, вступает в противоречие с международными конвенциями, ратифицированными КНР.

На сегодняшний день нет оснований полагать, что Китай изменит свою политику по распространению государственного регулирования на цифровое пространство, с которым взаимодействуют его граждане, и прекратит применение ИИ в качестве инструмента контроля как в нем, так и за его пределами. Следовательно, и национальное законодательство КНР будет развиваться в направлении установления государственного контроля не только за общественными отношениями, но и за самим

⁵⁷ Международная конвенция о ликвидации всех форм расовой дискриминации от 7 марта 1966 г. Информационно-правовой портал ГАРАНТ.РУ. <https://base.garant.ru/2540327>

⁵⁸ The Fight Against Terrorism and Extremism and Human Rights Protection in Xinjiang от 2 июня 2016 г. (2016). The State Council Information Office of the People's Republic of China. https://english.www.gov.cn/archive/white_paper/2016/06/02/content_281475363031504.htm

ИИ и его обучением (так, проект «Мер по управлению услугами генеративного интеллекта»⁵⁹ обязывает переучивать генеративный ИИ в случаях создания им противоправных материалов, либо материалов, которые могут быть использованы для совершения противоправных деяний). Проведенный анализ также установил, что китайский законодатель старается оперативно реагировать на пробелы и коллизии в праве, а потому можно предположить, что озвученные выше внутренние противоречия в национальном законодательстве страны вскоре будут устранены.

Учитывая сложившуюся геополитическую ситуацию, можно предположить, что при дальнейшем усилении давления на Китай со стороны стран Запада, в том числе при применении мер воздействия «Глобального режима санкций ЕС за нарушения прав человека», КНР больше не сможет оправдывать существующие ограничения ст. 29 Всеобщей декларации прав человека⁶⁰, допускающей их «с целью обеспечения должного признания и уважения прав и свобод других и удовлетворения справедливых требований морали, общественного порядка». На факт игнорирования Коллективным Западом угроз территориальной целостности и общественной стабильности Китая под эгидой защиты прав человека уже указывают китайские академики (Ji, 2014). При таком развитии событий Китай может либо самостоятельно выйти из международных конвенций и организаций, членство в которых препятствует развитию по выбранному политическому курсу и подрывает государственный суверенитет, либо будет из них исключен – такой прецедент уже сложился, когда из Совета Европы была исключена Российская Федерация.

Список литературы

- Ashraf, C. (2020). Exploring the impacts of artificial intelligence on freedom of religion or belief online. *The International Journal of Human Rights*, 26(5), 764. <https://doi.org/10.1080/13642987.2021.1968376>
- Bhatia, K. V. (2021). Religious Subjectivities and Digital Collectivities on Social Networking Sites in India. *Studies in Indian Politics*, 9(1), 22. <https://doi.org/10.1177/2321023021999141>
- Bingaman, K. A. (2023). Religion in the Digital Age: An Irreversible Process. *Religions*, 14(1), <https://doi.org/10.3390/rel14010108>
- Bosch, M. D. et al. (2017). Typing my Religion. Digital use of religious webs and apps by adolescents and youth for religious and interreligious dialogue. *Church, Communication and Culture*, 2(2), 122–135. <https://doi.org/10.1080/23753234.2017.1347800>
- Chan, C. (2019). Using digital storytelling to facilitate critical thinking disposition in youth civic engagement: A randomized control trial. *Children and Youth Services Review*, 107. <https://doi.org/10.1016/j.childyouth.2019.104522>
- Fontes, C, et al. (2022). AI-powered public surveillance systems: why we (might) need them and how we want them. *Technology in Society*, 71. <https://doi.org/10.1016/j.techsoc.2022.102137>
- Guan, T., & Liu, T. (2019). Globalized fears, localized securities: 'Terrorism' in political polarization in a one-party state. *Communist and Post-Communist Studies*, 52(4), 343–344. <https://doi.org/10.1016/j.postcomstud.2019.10.008>
- Ji, F. Y. (2014). Talking Past Each Other: Chinese and Western Discourses on Ethnic Conflict. *Procedia-Social and Behavioral Sciences*, 155(6), 434–441. <https://doi.org/10.1016/j.sbspro.2014.10.318>
- Jiménez-Tovar, S., & Lavička, M. (2020). Folklorized politics: how Chinese soft power works in Central Asia. *Asian Ethnicity*, 21(2), 244–268. <https://doi.org/10.1080/14631369.2019.1610355>

⁵⁹ 生成式人工智能服务管理办法（征求意见稿） от 11 апреля 2023 г. (2023). 国家互联网信息办公室. https://www.cac.gov.cn/2023-04/11/c_1682854275475410.htm

⁶⁰ Всеобщая декларация прав человека от 10 декабря 1948 г. КонсультантПлюс. https://www.consultant.ru/document/cons_doc_LAW_120805

- Kao, Y., & Sapp, S. G. (2022). The effect of cultural values and institutional trust on public perceptions of government use of network surveillance. *Technology in Society*, 70. <https://doi.org/10.1016/j.techsoc.2022.102047>
- Lavička, M. (2021). Changes in Chinese legal narratives about religious affairs in Xinjiang. *Asian Ethnicity*, 22(1), 61–76. <https://doi.org/10.1080/14631369.2020.1793100>
- Leung, B. (2005). China's Religious Freedom Policy: The Art of Managing Religious Activity. *The China Quarterly*, 184, 894. <https://doi.org/10.1017/s030574100500055x>
- Lin, W. (2018). Religion as an object of state power: The People's Republic of China and its domestic religious geopolitics after 1978. *Political Geography*, 67, 1–11. <https://doi.org/10.1016/j.polgeo.2018.09.003>
- Ma, R. (2006). Ethnic Relations in Contemporary China: Cultural Tradition and Ethnic Policies Since 1949. *Policy and Society*, 25(1), 85–90. [https://doi.org/10.1016/s1449-4035\(06\)70128-x](https://doi.org/10.1016/s1449-4035(06)70128-x)
- Marique, E., & Marique, Y. (2020). Sanctions on digital platforms: Balancing proportionality in a modern public square. *Computer Law & Security Review*, 36. <https://doi.org/10.1016/j.clsr.2019.105372>
- Meneses, L. F. S. (2021). Thinking critically through controversial issues on digital media: Dispositions and key criteria for content evaluation. *Thinking Skills and Creativity*, 42. <https://doi.org/10.1016/j.tsc.2021.100927>
- Reed, R. (2021). A.I. in Religion, A.I. for Religion, A.I. and Religion: Towards a Theory of Religious Studies and Artificial Intelligence. *Religions*, 12(6). <https://doi.org/10.3390/rel12060401>
- Sabic-El-Rayess, A. (2012). How do people radicalize?, *International Journal of Educational Development*, 87. <https://doi.org/10.1016/j.ijedudev.2021.102499>
- Smith, L. G. E. et al. (2020). Detecting psychological change through mobilizing interactions and changes in extremist linguistic style. *Computers in Human Behavior*, 108. <https://doi.org/10.1016/j.chb.2020.106298>
- Van Wie Davis, E. (2009). Tibetan separatism in China. *Korean Journal of Defense Analysis*, 21(2), 155–170. <https://doi.org/10.1080/10163270902872135>
- Wei, X. (2022). A critical evaluation of China's legal responses to cyberterrorism. *Computer Law & Security Review*, 47. <https://doi.org/10.1016/j.clsr.2022.105768>

Сведения об авторах



Шумакова Наталья Игоревна – аспирант кафедры конституционного и административного права, Южно-Уральский государственный университет (национальный исследовательский университет)

Адрес: 454080, Российская Федерация, г. Челябинск, пр. Ленина, 76

E-mail: territoryoflaw@mail.ru

ORCID ID: <https://orcid.org/0009-0004-6063-0650>



Титова Елена Викторовна – доктор юридических наук, доцент, заведующий кафедрой конституционного и административного права, директор юридического института, Южно-Уральский государственный университет (национальный исследовательский университет)

Адрес: 454080, Российская Федерация, г. Челябинск, пр. Ленина, 76

E-mail: territoryoflaw@mail.ru

ORCID ID: <https://orcid.org/0000-0001-9453-3550>

Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=57201640405>

Google Scholar ID: <https://scholar.google.ru/citations?user=Pqj6OiQAAAAJ>

РИНЦ Author ID: https://www.elibrary.ru/author_items.asp?authorId=451302

Вклад авторов

Е. В. Титова осуществляла общее руководство и постановку задач исследования, поиск и подбор научной литературы, критическую оценку интерпретации результатов исследования.

Н. И. Шумакова осуществляла анализ национального законодательства Китая, изучала криминальную статистику, выполняла интерпретацию результатов исследования и подготовку рукописи.

Конфликт интересов

Авторы сообщают об отсутствии конфликта интересов.

Финансирование

Исследование не имело спонсорской поддержки.

Тематические рубрики

Рубрика OECD: 5.05 / Law

Рубрика ASJC: 3308 / Law

Рубрика WoS: OM / Law

Рубрика ГРНТИ: 10.15.59 / Права и свободы человека и гражданина

Специальность ВАК: 5.1.2 / Публично-правовые (государственно-правовые) науки

История статьи

Дата поступления – 3 марта 2023 г.

Дата одобрения после рецензирования – 4 мая 2023 г.

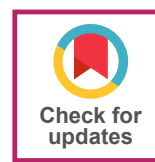
Дата принятия к опубликованию – 16 июня 2023 г.

Дата онлайн-размещения – 20 июня 2023 г.



Research article

DOI: <https://doi.org/10.21202/jdtl.2023.23>



Artificial Intelligence as an Auxiliary Tool for Limiting Religious Freedom in China

Natalya I. Shumakova ✉

South Ural State University (National Research University)
Chelyabinsk, Russian Federation

Elena V. Titova

South Ural State University (National Research University)
Chelyabinsk, Russian Federation

Keywords

Artificial intelligence,
China,
digital technologies,
extremism,
freedom of worship,
human rights,
law,
neuron network,
regulation,
religion

Abstract

Objective: based on studying the statistics of crimes, national legislation and norms of international law, to give a legal assessment to restrictions of the right to worship implemented with the use of artificial intelligence technologies in China.

Methods: the methodological basis of the research is the set of methods of scientific cognition, including specific sociological (analysis of statistical data and other documents), formal-legal (examining legal categories and definitions), formal-logical (analysis and synthesis), general scientific (induction, deduction), and other methods.

Results: the work researches prerequisites for using artificial intelligence technologies in China to control public relations arising during religious activity both in the digital space and beyond; analyzes the legal framework of the measures implemented; gives a legal assessment to restrictions of the religious freedom using artificial intelligence technologies; forecasts the further development of Chinese legislation and foreign policy associated with religious freedom. Additionally, the work analyzes materials of human rights organizations aimed at hindering the Chinese policy of "sinicisation" and "de-extremification" of ethnic and religious minorities, including with the help of control and propaganda using modern digital technologies.

✉ Corresponding author

© Shumakova N. I., Titova E. V., 2023



This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Scientific novelty: the work researches the attempt of China to regulate the challenges related to religious activity, arising during rapid digitalization of the society and state, which the Republic faces being a developing, multinational and polyconfessional country. The established restrictions of religious freedom using artificial intelligence technologies are considered along with the relevant criminal statistics. The legal assessment of using artificial intelligence as a tool for restricting the right to worship is given from the standpoint of international law, as well as with the account of Chinese national legislation.

Practical significance: the research results can be used to elaborate a consistent legal framework for using artificial intelligence technologies to counteract extremism.

For citation

Shumakova, N. I., Titova, E. V. (2023). Artificial Intelligence as an Auxiliary Tool for Limiting Religious Freedom in China. *Journal of Digital Technologies and Law*, 1(2), 540–563. <https://doi.org/10.21202/jdtl.2023.23>

References

- Ashraf, C. (2020). Exploring the impacts of artificial intelligence on freedom of religion or belief online. *The International Journal of Human Rights*, 26(5), 764. <https://doi.org/10.1080/13642987.2021.1968376>
- Bhatia, K. V. (2021). Religious Subjectivities and Digital Collectivities on Social Networking Sites in India. *Studies in Indian Politics*, 9(1), 22. <https://doi.org/10.1177/2321023021999141>
- Bingaman, K. A. (2023). Religion in the Digital Age: An Irreversible Process. *Religions*, 14(1), <https://doi.org/10.3390/rel14010108>
- Bosch, M. D. et al. (2017). Typing my Religion. Digital use of religious webs and apps by adolescents and youth for religious and interreligious dialogue. *Church, Communication and Culture*, 2(2), 122–135. <https://doi.org/10.1080/23753234.2017.1347800>
- Chan, C. (2019). Using digital storytelling to facilitate critical thinking disposition in youth civic engagement: A randomized control trial. *Children and Youth Services Review*, 107. <https://doi.org/10.1016/j.childyouth.2019.104522>
- Fontes, C, et al. (2022). AI-powered public surveillance systems: why we (might) need them and how we want them. *Technology in Society*, 71. <https://doi.org/10.1016/j.techsoc.2022.102137>
- Guan, T., & Liu, T. (2019). Globalized fears, localized securities: 'Terrorism' in political polarization in a one-party state. *Communist and Post-Communist Studies*, 52(4), 343–344. <https://doi.org/10.1016/j.postcomstud.2019.10.008>
- Ji, F. Y. (2014). Talking Past Each Other: Chinese and Western Discourses on Ethnic Conflict. *Procedia-Social and Behavioral Sciences*, 155(6), 434–441. <https://doi.org/10.1016/j.sbspro.2014.10.318>
- Jiménez-Tovar, S., & Lavička, M. (2020). Folklorized politics: how Chinese soft power works in Central Asia. *Asian Ethnicity*, 21(2), 244–268. <https://doi.org/10.1080/14631369.2019.1610355>
- Kao, Y., & Sapp, S. G. (2022). The effect of cultural values and institutional trust on public perceptions of government use of network surveillance. *Technology in Society*, 70. <https://doi.org/10.1016/j.techsoc.2022.102047>
- Lavička, M. (2021). Changes in Chinese legal narratives about religious affairs in Xinjiang. *Asian Ethnicity*, 22(1), 61–76. <https://doi.org/10.1080/14631369.2020.1793100>
- Leung, B. (2005). China's Religious Freedom Policy: The Art of Managing Religious Activity. *The China Quarterly*, 184, 894. <https://doi.org/10.1017/s030574100500055x>
- Lin, W. (2018). Religion as an object of state power: The People's Republic of China and its domestic religious geopolitics after 1978. *Political Geography*, 67, 1–11. <https://doi.org/10.1016/j.polgeo.2018.09.003>

- Ma, R. (2006). Ethnic Relations in Contemporary China: Cultural Tradition and Ethnic Policies Since 1949. *Policy and Society*, 25(1), 85–90. [https://doi.org/10.1016/s1449-4035\(06\)70128-x](https://doi.org/10.1016/s1449-4035(06)70128-x)
- Marique, E., & Marique, Y. (2020). Sanctions on digital platforms: Balancing proportionality in a modern public square. *Computer Law & Security Review*, 36. <https://doi.org/10.1016/j.clsr.2019.105372>
- Meneses, L. F. S. (2021). Thinking critically through controversial issues on digital media: Dispositions and key criteria for content evaluation. *Thinking Skills and Creativity*, 42. <https://doi.org/10.1016/j.tsc.2021.100927>
- Reed, R. (2021). A.I. in Religion, A.I. for Religion, A.I. and Religion: Towards a Theory of Religious Studies and Artificial Intelligence. *Religions*, 12(6). <https://doi.org/10.3390/rel12060401>
- Sabic-El-Rayess, A. (2012). How do people radicalize?, *International Journal of Educational Development*, 87. <https://doi.org/10.1016/j.ijedudev.2021.102499>
- Smith, L. G. E. et al. (2020). Detecting psychological change through mobilizing interactions and changes in extremist linguistic style. *Computers in Human Behavior*, 108. <https://doi.org/10.1016/j.chb.2020.106298>
- Van Wie Davis, E. (2009). Tibetan separatism in China. *Korean Journal of Defense Analysis*, 21(2), 155–170. <https://doi.org/10.1080/10163270902872135>
- Wei, X. (2022). A critical evaluation of China's legal responses to cyberterrorism. *Computer Law & Security Review*, 47. <https://doi.org/10.1016/j.clsr.2022.105768>

Authors information



Natalya I. Shumakova – post-graduate student of the Department of Constitutional and Administrative Law, South Ural State University (National Research University)
Address: 76 prospekt Lenina, 454080 Chelyabinsk, Russian Federation
E-mail: territoryoflaw@mail.ru
ORCID ID: <https://orcid.org/0009-0004-6063-0650>



Elena V. Titova – Doctor of Law, Associate Professor, Head of the Department of Constitutional and Administrative Law, Director of Law Institute, South Ural State University (National Research University)
Address: 76 prospekt Lenina, 454080 Chelyabinsk, Russian Federation
E-mail: territoryoflaw@mail.ru
ORCID ID: <https://orcid.org/0000-0001-9453-3550>
Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=57201640405>
Google Scholar ID: <https://scholar.google.ru/citations?user=Pqj6OiQAAAAJ>
RSCI Author ID: https://www.elibrary.ru/author_items.asp?authorid=451302

Authors' contributions

Elena V. Titova performed overall guidance and setting the research tasks, search and selection of scientific literature, critical evaluation of the research results interpretation.

Natalya I. Shumakova performed analysis of the Chinese national legislation, studied criminal statistics, interpreted the research results and prepared the manuscript.

Conflict of interests

The authors declare no conflict of interests.

Financial disclosure

The research had no sponsorship.

Thematic rubrics

OECD: 5.05 / Law

PASJC: 3308 / Law

WoS: OM / Law

Article history

Date of receipt – April 17, 2023

Date of approval – April 24, 2023

Date of acceptance – June 16, 2023

Date of online placement – June 20, 2023

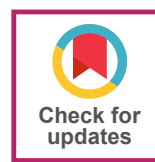


Научная статья

УДК 340.143:004.8

EDN: <https://elibrary.ru/sadrzw>

DOI: <https://doi.org/10.21202/jdtl.2023.24>



Возможность и необходимость человеко-ориентированного искусственного интеллекта в юридической теории и практике

Андрей Владимирович Резаев

Санкт-Петербургский государственный университет
г. Санкт-Петербург, Российская Федерация

Наталья Дамировна Трегубова

Санкт-Петербургский государственный университет
г. Санкт-Петербург, Российская Федерация

Ключевые слова

Алгоритм,
искусственная
социальность,
искусственный интеллект,
право,
регулирование,
социология,
цифровая экономика,
цифровые технологии,
человек,
человеко-ориентированный
искусственный интеллект

Аннотация

Цель: определение проблем, которые ставит распространение технологий искусственного интеллекта перед юридической теорией и практикой, и соотнесение этих проблем с человеко-ориентированным подходом к искусственному интеллекту (Human-Centered AI).

Методы: исследование основано на критическом анализе релевантной литературы из разных дисциплинарных областей: юриспруденции, социологии, философии, компьютерных наук.

Результаты: в статье сформулированы основные перспективы и проблемы, которые ставит перед правовой системой развитие цифровых технологий в целом и алгоритмов искусственного интеллекта в частности. Выделенные проблемы соотнесены с положениями человеко-ориентированного подхода к искусственному интеллекту. Авторы утверждают необходимость того, чтобы разработчики программ искусственного интеллекта, владельцы компаний, участвующих в гонке по внедрению технологий искусственного интеллекта, сосредоточивались на том, чтобы люди и человек как базовая ценность общества находились в центре внимания. В частности, специальные усилия следует направить на сбор и анализ

Контактное лицо

© Резаев А. В., Трегубова Н. Д., 2023



Статья находится в открытом доступе и распространяется в соответствии с лицензией Creative Commons «Attribution» («Атрибуция») 4.0 Всемирная (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0/deed.ru>), позволяющей неограниченно использовать, распространять и воспроизводить материал при условии, что оригинальная работа упомянута с соблюдением правил цитирования.

высококачественных данных для организации разработок инструментов искусственного интеллекта, поскольку сегодня алгоритмы искусственного интеллекта эффективны настолько, насколько эффективны данные, на которых они обучаются. Авторы формулируют три принципа человеко-ориентированного искусственного интеллекта для правовой сферы: 1) человек как необходимое звено в цепочке принятия и исполнения правовых решений; 2) необходимость регулирования искусственного интеллекта на уровне международного права; 3) формулировка «табу» для внедрения технологий искусственного интеллекта.

Научная новизна: статья представляет собой первую в русскоязычной научной литературе попытку обозначить перспективы развития области человеко-ориентированного искусственного интеллекта в юриспруденции. На основании анализа специальной литературы авторы формулируют три принципа включения искусственного интеллекта в юридическую теорию и практику с точки зрения человеко-ориентированного подхода к искусственному интеллекту.

Практическая значимость: принципы, сформулированные в статье, будут полезны как для правового регулирования технологий искусственного интеллекта, так и для гармоничного их включения в юридические практики.

Для цитирования

Резаев, А. В., Трегубова, Н. Д. (2023). Возможность и необходимость человеко-ориентированного искусственного интеллекта в юридической теории и практике. *Journal of Digital Technologies and Law*, 1(2), 564–580. <https://doi.org/10.21202/jdtl.2023.24>

Содержание

Введение

1. Цифровые технологии и право

2. Искусственный интеллект в юридической практике и теории: pro et contra

3. Человеко-ориентированный искусственный интеллект

Выводы

Список литературы

Введение

В 1948 г. Норберт Винер, основатель кибернетики, писал: «Мы уже в состоянии построить искусственные машины почти со сколь угодно сложным поведением. Еще задолго до Нагасаки и до того, как общественности стало известно о существовании атомной бомбы, мне пришла мысль, что мы стоим перед лицом другой социальной силы, несущей неслыханные возможности для добра и для зла» (Винер, 1983).

Сегодня «искусственные машины» уже решают (или вскоре будут способны решить) многочисленные проблемы, с которыми сталкивается человечество. Однако

эти машины, без сомнения, создают и новые проблемы¹. То, что Винер называл «искусственными машинами», сегодня в том или ином виде стало частью жизни общества, и нам уже сложно представить свою жизнь без технологий искусственного интеллекта. И не удивительно, что в последние годы было много информационного «шума» вокруг искусственного интеллекта и его потенциала кардинально изменить мир, в котором мы живем и работаем.

Цель наших размышлений – показать, что по мере того, как технологии искусственного интеллекта продолжают разрабатываться и внедряться в повседневную жизнь, в той же прогрессии растет необходимость, чтобы разработчики программ, владельцы старых и новых компаний, участвующих в гонке по внедрению новых инструментов, сосредоточивались в первую очередь на том, чтобы люди и человек как базовая ценность общества находились в центре внимания. Не цели одного человека или одной компании, не технологии и машины, но человек и отношение к человеку как мерилу нравственности и гуманизма. Понимая, что благостные призывы к гуманизму могут звучать абстрактно в логике технологически ориентированного развития, мы хотели бы более конкретно показать необходимость ориентироваться в работе с искусственным интеллектом в рамках подхода, который получил в научной литературе название Human-Centered AI (HCAI) – «человеко-ориентированный искусственный интеллект».

Проблема ориентации искусственного интеллекта на благо человека стоит во всех сферах жизни общества, но в некоторых из них – особенно остро. К ним относятся образование, медицина и юриспруденция, где цена ошибки – ошибки человека или алгоритма – является наиболее высокой. Юридические решения регулируют жизнь человека, его отношения с другими людьми, порой они касаются экзистенциальных вопросов – жизни, смерти, справедливости.

В данной статье мы рассматриваем, какие проблемы ставит распространение технологий искусственного интеллекта перед юридической теорией и практикой и как эти проблемы соотносятся с человеко-ориентированным подходом к искусственному интеллекту. Искусственный интеллект мы определяем как «ансамбль разработанных и закодированных человеком рационально-логических, формализованных правил, которые организуют процессы, позволяющие имитировать интеллектуальные структуры, производить и воспроизводить целерациональные действия, а также осуществлять последующее кодирование и принятие инструментальных решений вне зависимости от человека» (Резаев, Трегубова, 2019).

Сегодня одним из определяющих факторов развития технологий искусственного интеллекта является онлайн-культура – «ансамбль (гармоническая совокупность) сетей коммуникаций, технических устройств, алгоритмов, формальных и неформальных правил взаимодействия, паттернов поведения, культурных символов, которые делают возможной и структурируют активность людей в сети Интернет и аналогичных сетях, обеспечивающих удаленный доступ к созданию, обмену и получению информации» (Резаев, Трегубова, 2019).

¹ В подтверждение можно привести недавнее высказывание Сэма Альтмана, основателя компании OpenAI, в которой был разработан получивший широкую известность чат-бот ChatGPT: «Думаю, что сейчас мы находимся не там, где хотим быть. Есть весьма широкие границы того, что могут делать подобные системы, и эти границы должны определяться не Microsoft или OpenAI, а обществом, правительствами, чем-то в этом роде, людьми и прямой демократией. <...> Это совсем новая технология, и мы не знаем, как с ней обращаться». *Bing's Revenge and Google's A.I. Face Plant*. <https://www.nytimes.com/2023/02/10/podcasts/bings-revenge-and-googles-ai-face-plant.html>

Интернет обеспечивает как колоссальные объемы данных, на которых обучаются алгоритмы искусственного интеллекта, так и «площадку» для действия этих алгоритмов.

Как следствие одновременного развития вычислительных мощностей искусственного интеллекта и развития онлайн-культуры, искусственный интеллект все более включается в повседневную жизнь людей и отношения между ними. Возникает «искусственная социальность»: искусственный интеллект становится активным посредником и участником социальных взаимодействий (Резаев, Трегубова, 2019).

Проект искусственного интеллекта с самого своего начала имел а-дисциплинарный характер. Создатели искусственного интеллекта стремились воспроизвести человеческий интеллект и потому смело заимствовали нужные им положения из математики, психологии, кибернетики и т. д. (Рассел, Норвиг, 2007). Но если создание машин, воспроизводящих деятельность человеческого разума, требовало обращения к достижениям различных областей знания, то это тем более верно для понимания того, как эти машины входят в повседневную жизнь общества, взаимодействуют с людьми, встраиваются в общественные отношения. Иными словами, исследование проблем искусственной социальности должно иметь междисциплинарный, а в перспективе а-дисциплинарный характер. Поэтому в настоящей статье мы опираемся как на философский и социологический анализ проблем искусственного интеллекта, так и на результаты правовых исследований.

Дальнейшее рассуждение организовано следующим образом. Сначала мы обратим внимание на несколько принципиальных моментов, которые касаются внедрения цифровых технологий в юридические практики. Затем – на проблемы и перспективы стремительного проникновения искусственного интеллекта в повседневную жизнь общества, что изменяет характер правовой работы и структуру юридических систем. В завершение мы обратимся к человеко-ориентированному подходу к искусственному интеллекту и его следствиям для правовой сферы жизни общества.

1. Цифровые технологии и право

В самом общем виде суммируя влияние цифровых технологий на правовую систему, в первую очередь необходимо отметить следующее.

Во-первых, цифровые технологии упростили доступ к правовой информации через онлайн-базы данных, правовые поисковые системы и другие онлайн-ресурсы. Данное обстоятельство, соответственно, создало возможности для проведения онлайн-исследований, ориентированных на получение юридической информации, практически для каждого пользователя Интернета (а это почти 90 % населения РФ)². Интернет революционизировал поиск во всех сферах жизни людей (Утехин, 2019), и юридическая информация не является исключением.

Во-вторых, цифровые технологии улучшили коммуникацию между юристами, клиентами и другими заинтересованными сторонами в правовой системе. К примеру, технология видео-конференц-связи позволяет юристам дистанционно общаться с клиентами, что расширило доступ к юридическим услугам для жителей отдаленных районов.

² Дмитрий Чернышенко: Сегодня в России насчитывается около 130 миллионов интернет-пользователей – это практически 90 % населения. <http://government.ru/news/46639/>

В-третьих, цифровые технологии сделали возможными подачу и хранение юридических документов в электронном виде. Иными словами, организация юридических практик с помощью цифровых технологий значительно уменьшает потребность в бумажных документах и упрощает поиск и обмен информацией (Русакова, 2020; Степанов и др., 2021).

В-четвертых, цифровые технологии привели к автоматизации многих юридических процессов, таких как рутинная проверка документов и составление контрактов (включая так называемые смарт-контракты (Ефимова и др., 2020)). Соответственно, значительно уменьшилась потребность в рутинизированном ручном труде юристов и их помощников.

В-пятых, использование цифровых технологий в юридических практиках вызвало к жизни новые области права, такие как киберправо/право в киберпространстве (Мажорина, 2020), право интеллектуальной собственности и право защиты данных (Войниканис, 2020).

Таким образом, цифровые технологии уже оказали значительное влияние на развитие права, сделав юридические услуги и практики более доступными, эффективными и действенными. Вместе с тем практикующие юристы, специальная литература, массмедиа и ежедневный потребитель юридических услуг практически в один голос обращают внимание на тот факт, что цифровые технологии порождают новые проблемы для развития правовой системы. В первую очередь это проблемы конфиденциальности (Талапина, 2022) и доступности (Панченко, 2012) юридических баз данных, а также проблема критической оценки информации, полученной из Интернета (Греггер, 2017).

Современный этап развития цифровых технологий в онлайн-культуре с необходимостью предполагает обратить внимание на то, каким образом искусственный интеллект трансформирует и определяет дальнейшее развитие юридических практик, каковы плюсы и минусы для правоведов в использовании технологий искусственного интеллекта в повседневной юридической практике.

2. Искусственный интеллект в юридической практике и теории: *pro et contra*

Есть как преимущества, так и недостатки использования технологий искусственного интеллекта в повседневной юридической практике. К основным преимуществам можно отнести следующие:

- Эффективность в организации работы юристов. Технологии искусственного интеллекта автоматизируют и ускоряют решение таких задач, как просмотр документов, предварительные юридические исследования литературных источников, анализ контрактов (Талапина, 2021).

- Искусственный интеллект может выполнять определенные задачи с более высоким уровнем точности, чем люди. Например, выявлять закономерности в данных или проверять документы на наличие фактических ошибок или грамматических и стилистических несоответствий (Андреев и др., 2020).

- Очевидно, что использование технологий искусственного интеллекта, снижая потребность в ручном труде, обуславливает экономию средств юридических фирм и их клиентов.

– Технологии искусственного интеллекта предоставляют юристам более полную, многостороннюю и подробную информацию, позволяющую им принимать более обоснованные решения.

Среди основных недостатков использования искусственного интеллекта выделяются следующие:

– Общий минус использования искусственного интеллекта для всех профессий: происходит исчезновение одних профессий и появление других (Ли, 2019). Широкое использование технологий искусственного интеллекта в юридической практике пока еще потенциально, но очень скоро в обязательном порядке приведет к пересмотру номенклатуры рабочих мест в структуре юридической системы, особенно это будет касаться помощников юристов и другого вспомогательного юридического персонала (Лессиг, 2019).

– Системы искусственного интеллекта в той или иной степени являются трансляторами предвзятости, предубеждений, которые характерны для их создателей (Горохова, 2021). Алгоритмы искусственного интеллекта могут быть в той или иной степени предвзятыми или ошибочными в результате как минимум двух обстоятельств: а) если они основываются и разработаны на массивах данных, которые ошибочны и предвзяты; б) если они не используются должным образом. Отсюда внедрение технологий искусственного интеллекта предполагает поиски ответов на вопросы о том, как обеспечить справедливость и беспристрастность систем искусственного интеллекта.

– Технологии искусственного интеллекта, как и всякие технологии, несут в себе риски безопасности. Технологии искусственного интеллекта не могут обеспечить стопроцентную кибербезопасность (О'Нил, 2018). Искусственный интеллект может минимизировать, но не исключить полностью утечку данных или взлом. Соответственно, основа основ юридических практик – конфиденциальность – остается под угрозой при использовании технологий искусственного интеллекта.

Таким образом, использование технологий искусственного интеллекта в повседневной юридической практике очевидно обеспечивает многочисленные преимущества, но важно взвесить их с потенциальными рисками и недостатками. Юристы должны не только понимать возможности технологий искусственного интеллекта, но и тщательно продумать их использование, видеть их ограничения и потенциальные риски.

Кроме проблем использования искусственного интеллекта, которые уже проявляются в повседневной жизни, следует также иметь в виду и реальные проблемы, порождаемые повсеместным проникновением искусственного интеллекта в юридические практики:

– Проблемы, касающиеся конфиденциальности. Для эффективной работы системам искусственного интеллекта часто требуется доступ к большим объемам персональных данных, что вызывает опасения по поводу конфиденциальности и защиты данных. Регуляторы и законодатели должны найти баланс между защитой частной жизни и продвижением инноваций в сфере искусственного интеллекта (Горохова, 2021).

– Юридическая ответственность за те или иные действия, совершаемые искусственным интеллектом. По мере того, как системы искусственного интеллекта становятся более автономными и принимают решения без вмешательства человека, возникают вопросы о том, кто несет ответственность за их действия (Вавилин, 2021; Батурин, Полубинская, 2022). Например, если автомобиль без водителя, основанный

на работе искусственного интеллекта, стал причиной аварии, должны ли нести ответственность производитель, владелец или сама система искусственного интеллекта (Руденко, 2020)? На кого следует возлагать ответственность, когда что-то идет не так в других случаях использования инструментов искусственного интеллекта? Кто несет ответственность за несчастные случаи или ошибки, вызванные системой искусственного интеллекта: программист, владелец или сам искусственный интеллект? Это юридические вопросы уже сегодняшнего дня.

– Вопросы, касающиеся интеллектуальной собственности на продукты, созданные технологиями искусственного интеллекта (Lee et al., 2021). Например, кого считать изобретателем или художником, если система искусственного интеллекта создает произведение искусства или изобретает новую технологию?

– Принципиальным моментом становится использование инструментов искусственного интеллекта (таких, например, как ChatGPT) для юридической интерпретации документов и применения юридических норм, особенно в отношении сложного и нюансированного характера юридического обоснования того или иного решения. Есть все основания опасаться, что искусственный интеллект не сможет полностью охватить человеческие рассуждения и суждения, необходимые для эффективного принятия юридических решений (Цветков, 2021).

– Отсутствие общения и реальных человеческих контактов. Это очень существенный недостаток для юридических практик, которые по определению могут затрагивать экзистенциальные проблемы жизни и смерти, ограничение свободы человека. Сами судьи отмечают: правосудие невозможно без целостного видения ситуации, включающего моральные и эмоциональные аспекты, – без того, что недоступно искусственному интеллекту (Быков, Нарская, 2022).

Следует отметить, что сам вопрос, следует ли регулировать технологии искусственного интеллекта и как именно это делать, является предметом обсуждения (Etzioni & Etzioni, 2017). Принципы здесь только начинают вырабатываться, причем «пионерами» часто выступают законодатели Европейского союза (Hickman & Petrin, 2021; Fink & Finck, 2022; Ulnicane, 2022). В России также ведется работа по юридическому регулированию технологий искусственного интеллекта: в 2019 г. была принята Национальная стратегия развития искусственного интеллекта на период до 2030 г.³, в которой прописаны базовые определения и общие принципы использования технологий искусственного интеллекта.

Таким образом, развитие искусственной социальности представляет и практическую, и концептуальную проблему для юриспруденции. Практическую – в отношении того, как технологии искусственного интеллекта будут менять юридические практики, концептуальную – в отношении их правового регулирования.

Нам представляется, что набор проблем, которые уже возникли и обязательно будут возникать в самом ближайшем будущем в юридической практике, можно решать более продуктивно с помощью подхода, который получил название «человеко-ориентированный искусственный интеллект» (Human-Centered AI).

³ О развитии искусственного интеллекта в Российской Федерации: Указ Президента Российской Федерации. <http://static.kremlin.ru/media/events/files/ru/AH4x6HgKWANwVtMO6PDhcbRpvdt1HCCsv.pdf>

3. Человеко-ориентированный искусственный интеллект

Подход, который получил в научной литературе название Human-Centered AI (Ford et al., 2015; Shneiderman, 2021)⁴, предполагает в первую очередь понимание очень простого факта: люди и машины – не одно и то же⁵. Нет необходимости ориентироваться на то, чтобы сделать абсолютно подобный человеку инструмент искусственного интеллекта. Напротив, скорее всего, успех будет в противоположном направлении, когда человек останется человеком со своим интеллектом, сознанием, подсознанием и эмоционально-духовным миром, а машины и алгоритмы будут разрабатываться человеком и в «самообучении» следовать своей, отличной от человека логике развития.

К сожалению, данным обстоятельством, как, впрочем, и в целом человеко-ориентированным подходом, пренебрегают. Большинство технологических лидеров в США и других странах продолжают тратить значительные средства на разработку программного обеспечения, которое может делать то же самое, с чем прекрасно справляются люди. Разработчики отчетливо осознают, что могут легко заработать, продавая свои продукты корпорациям, у которых нет других ориентаций в своем развитии, кроме тех, что заданы логикой рынка и получения прибыли (Зубофф, 2022). Все сосредоточено на использовании искусственного интеллекта для сокращения затрат на рабочую силу, мало заботясь о сути социального прогресса и развитии нравственного человека и справедливого общества.

Human-Centered AI требует уделять первоочередное внимание сбору и анализу высококачественных данных для организации разработок инструментов искусственного интеллекта. Алгоритмы искусственного интеллекта эффективны ровно настолько, насколько эффективны данные, на которых они обучаются, а предвзятые или неполные данные могут привести не только к несправедливым/ложным, но и к прямо противоположным первоначальной цели результатам. Сбор информации для самообучающихся моделей должен быть не только разнообразным, но и репрезентативным, данные должны отражать реальный мир, в котором мы живем, и людей, с которыми мы работаем, вне зависимости от их социальных и классовых различий.

Мы уже обращали внимание на тот факт, что при современном этапе развития капитализма, в ситуации предельной ориентации на финансовые показатели, прибыль и функциональную эффективность, практически невозможно обеспечить решение этих проблем (Rezaev, 2021)⁶. Тем не менее вообще их не рассматривать, не пытаться предлагать варианты их решения – неправильная тактика развития социальных наук.

⁴ Следует отметить, что в 2019 г. в Стэнфордском университете США был организован Институт человеко-ориентированного искусственного интеллекта (Stanford University Human-Centered AI) – крупнейший исследовательский центр по данной проблематике.

⁵ В философии и социальных науках это суждение высказывалось неоднократно. См. (Дрейфус, 1978; Wolfe, 1993; Esposito, 2017).

⁶ Так, Илон Маск (на его деньги была создана компания OpenAI, которая разработала ChatGPT) с очевидным сожалением сказал, что он давал деньги (1 млрд долларов) на то, чтобы была создана открытая платформа, ориентированная на бесплатный и открытый доступ, а сейчас ChatGPT представляет собой прямо противоположную модель – закрытую и полностью ориентированную на прибыль. Причем в настоящий момент Илон Маск уже не имеет никакого контроля ни над OpenAI, ни над ChatGPT. *Elon Musk at the 2023 World Government Summit in Dubai*. https://www.youtube.com/watch?v=jmNrINgXx_U&ab_channel=ElonAlerts

Рынок никогда не был и не может быть (даже в условиях искусственной социальности) высшим мерилom красоты, добра и истины. Стратегически социальное знание обосновало невозможность гармоничного, нравственного и справедливого мира, где нет эксплуатации человека человеком и отсутствует социальное и культурное неравенство, в системе координат капиталистического хозяйствования⁷. Но вот какие существуют проблемы для общества и каковы варианты траекторий развития общества в условиях пока еще бесконтрольного распространения инструментов искусственного интеллекта – темы, которые явно с запозданием только начинают рассматриваться.

Характеризуя особенности развития искусственного интеллекта, следует иметь в виду, что данные технологии не являются нейтральными по своей сути. Они созданы людьми и отражают ценности, предубеждения, предрассудки их создателей. Отсюда при создании инструментов искусственного интеллекта следует с необходимостью придерживаться этического и ориентированного на человека подхода к разработке и развертыванию систем искусственного интеллекта. В том числе это означает учет различных точек зрения и мнений в процессе проектирования, обеспечение прозрачности и подотчетности, а также уделение приоритетного внимания человеческой личности, благополучию общества в целом, а не отдельных субъектов или технологических систем.

Принципиальным моментом остается понимание того, что инструменты искусственного интеллекта действительно уже сейчас являются мощным средством для решения некоторых из самых насущных проблем, стоящих перед обществом, но это не панацея. Определяя и формулируя направления общественного развития, нельзя полагаться исключительно на искусственный интеллект при решении социальных и экономических, культурных и политических проблем. Даже в условиях «искусственной социальности» люди должны оставаться в реальности человеческого опыта и признавать, что социальный прогресс требует большего, чем простые технологические решения.

⁷ К примеру, «Богатство народов» Адама Смита. Хотя Смита часто считают первым теоретиком политической экономии и основным защитником капитализма, его работы критически относятся ко многим аспектам капиталистического хозяйствования. Так, он утверждает, что погоня за прибылью может привести к отсутствию заботы о благополучии рабочих и общества в целом и что для обеспечения справедливого и равноправного общества необходима определенная форма государственного вмешательства. Карл Полаanyi в работе «Великая трансформация» утверждает, что капитализм – исторически недавно возникшее явление, которое породило абсолютно негативные последствия для развития общества, включая превращение труда в товар, разрушение традиционного образа жизни и подъем националистических и фашистских движений. Торстейн Веблен в «Теории праздного класса» показывает, что капитализм создает культуру демонстративного потребления и расточительства, при которой люди ценятся за их способность потреблять и демонстрировать богатство, а не за их вклад в общество. Современная исследовательница из Канады Наоми Кляйн утверждает, что капитализм часто навязывается обществу посредством насилия и принуждения и что он часто используется богатой элитой как средство сохранения политической и экономической власти (Klein, 2007). Более подробно см. (Harvey, 2014).

Выводы

Настоящее рассуждение начиналось цитатой Норберта Винера – отца-основателя кибернетики. Завершим его суждениями одного из основателей области исследований искусственного интеллекта Джозефа Вейценбаума. Вейценбаум утверждал, что использование вычислительных машин следовало бы запретить или, по меньшей мере, ограничить в двух случаях (Вейценбаум, 1982). Первый случай связан с попытками заменить человека машиной в сферах, связанных с межличностными отношениями, любовью, пониманием. Второй случай – использование компьютеров в ситуации, когда оно может привести к необратимым последствиям. Вейценбаум, как нам представляется, удачно формулирует базовые положения человеко-ориентированного искусственного интеллекта, которые касаются распространения технологий искусственного интеллекта в целом и использования их в теории и практике юриспруденции в частности.

В заключение сформулируем три принципа включения искусственного интеллекта в юридическую теорию и практику с точки зрения человеко-ориентированного подхода к искусственному интеллекту.

Первое. Человек всегда должен оставаться в цепочке принятия/исполнения юридических решений. Данный тезис неоднократно формулировался и отечественными (Степанов и др., 2021), и зарубежными (Лессиг, 2019) правоведами. Технологии искусственного интеллекта могут брать на себя многие задачи в юридической практике, но контролировать, проверять, осмысливать, взвешивать действия и решения искусственного интеллекта должен человек.

Второе. Сегодня мы стоим перед необходимостью разработки законов, определяющих рациональный и понятный *modus Vivendi* для деятельности искусственного интеллекта в социальных системах, ориентированных на человека, а не на прибыль и рынок. Это почти невозможно сделать в рамках одного государства, особенно если это капиталистическое государство. Вот почему мир стоит перед необходимостью создания международного права в области искусственного интеллекта. Закон, как и любое правило, может быть нарушен – по недоразумению или по злему умыслу. Однако нарушение закона не отменяет самого закона, нарушение просто выявляет злонамеренных лиц, извративших закон.

Третье. Развитие искусственного интеллекта ставит вопрос о необходимости запретов, в том числе юридических, своего рода табу на использование искусственного интеллекта в некоторых сферах жизни человека (Rezaev, 2021). Прежде всего, это сферы, связанные с экзистенциальными вопросами. Например, большой вопрос – следует ли использовать искусственный интеллект, чтобы установить, лжет ли человек (Oravec, 2022), или может ли искусственный интеллект выступать автономным оружием (International Committee, 2020). Определение таких сфер на международном, национальном и локальном уровнях, формулировка правовых запретов и создание механизмов правоприменения – одна из важнейших задач для человеко-ориентированного искусственного интеллекта.

Список литературы

Андреев, В. К., Лаптев, В. А., Чуча, С. Ю. (2020). Искусственный интеллект в системе электронного правосудия при рассмотрении корпоративных споров. *Вестник Санкт-Петербургского университета. Право*, 1, 19–34. <https://doi.org/10.21638/spbu14.2020.102>

- Батурин, Ю. М., Полубинская, С. В. (2022). Искусственный интеллект: правовой статус или правовой режим? *Государство и право*, 10, 141–154. <https://doi.org/10.31857/s102694520022606-7>
- Быков, А. В., Нарская, А. И. (2022). Закон, мораль и машинное обучение: взгляд судей на сущность и перспективы роботизации правосудия. *Мониторинг общественного мнения: экономические и социальные перемены*, 5, 278–298. <https://doi.org/10.14515/monitoring.2022.5.2137>
- Вавилин, Е. В. (2021). Искусственный интеллект как участник гражданских отношений: трансформация права. *Вестник Томского государственного университета. Право*, 42, 135–146. <https://doi.org/10.17223/22253513/42/11>
- Вейценбаум, Дж. (1982). *Возможности вычислительных машин и человеческий разум. От суждений к вычислениям*. Москва: Радио и связь.
- Винер, Н. (1983). *Кибернетика, или управление и связь в животном и машине* (2-е изд.). Москва: Наука; Главная редакция изданий для зарубежных стран.
- Войниканис, Е. А. (2020). Регулирование больших данных и право интеллектуальной собственности: общие подходы, проблемы и перспективы развития. *Закон*, 7, 135–156.
- Горохова, С. С. (2021). Искусственный интеллект: инструмент обеспечения кибербезопасности финансовой сферы или киберугроза для банков. *Банковское право*, 1, 35–46. <https://doi.org/10.18572/1812-3945-2021-1-35-46>
- Греггер, Р. (2017). Судья как интернет-серфер. Выяснение обстоятельств дела в Интернете. *Вестник гражданского процесса*, 7(4), 161–173. <https://doi.org/10.24031/2226-0781-2017-7-4-161-173>
- Дрейфус, Х. (1978). *Чего не могут вычислительные машины: Критика искусственного разума*. Москва: Прогресс.
- Ефимова, Л. Г., Михеева, И. В., Чуб, Д. В. (2020). Сравнительный анализ доктринальных концепций правового регулирования смарт-контрактов в России и в зарубежных странах. *Право. Журнал Высшей школы экономики*, 4, 78–105. <https://doi.org/10.17323/2072-8166.2020.4.78.105>
- Зубофф, Ш. (2022). *Эпоха надзорного капитализма. Битва за человеческое будущее на новых рубежах власти*. Москва: Изд-во Института Гайдара.
- Лессиг, Л. (2019). Искусственный интеллект вытеснит широкий пласт юристов. *Закон*, 5, 8–30.
- Ли, К.-Ф. (2019). *Сверхдержавы искусственного интеллекта. Китай, Кремниевая долина и новый мировой порядок*. Москва: Манн, Иванов и Фербер.
- Мажорина, М. В. (2020). Киберпространство и методология международного частного права. *Право. Журнал Высшей школы экономики*, 2, 230–253. <https://doi.org/10.17323/2072-8166.2020.2.230.253>
- О'Нил, К. (2018). *Убийственные большие данные. Как математика превратилась в оружие массового поражения*. Москва: АСТ.
- Панченко, В. Ю. (2012). Информационная доступность юридической помощи: идеальная модель и реальное состояние. *Аграрное и земельное право*, 11(95), 95–102.
- Рассел, С., Норвиг, П. (2007). *Искусственный интеллект: современный подход* (2-е изд.). Москва: Вильямс.
- Резаев, А. В., Трегубова, Н. Д. (2019). «Искусственный интеллект», «онлайн-культура», «искусственная социальность»: определение понятий. *Мониторинг общественного мнения: Экономические и социальные перемены*, 6, 35–47. <https://doi.org/10.14515/monitoring.2019.6.03>
- Руденко, Н. И. (2020). Социотехнические барьеры разработки беспилотных автомобилей в России. В кн. Л. В. Земнухова и др. *Приключения технологий: барьеры цифровизации в России* (с. 17–70). Москва – Санкт-Петербург: ФНИСЦ РАН. <https://doi.org/10.31119/978-5-89697-339-3>
- Русакова, Е. П. (2020). Интегрирование современных цифровых технологий в судопроизводство Китайской Народной Республики и Сингапура. *Государство и право*, 9, 102–109. <https://doi.org/10.31857/s102694520011323-6>
- Степанов, О. А., Печегин, Д. А., Дьяконова, М. О. (2021). К вопросу о цифровизации судебной деятельности. *Право. Журнал Высшей школы экономики*, 5, 4–23. <https://doi.org/10.17323/2072-8166.2021.5.4.23>
- Талапина, Э. В. (2021). Искусственный интеллект и правовые экспертизы в государственном управлении. *Вестник Санкт-Петербургского университета. Право*, 4, 865–881. <https://doi.org/10.21638/spbu14.2021.404>
- Талапина, Э. В. (2022). Право на информационное самоопределение: на грани публичного и частного. *Право. Журнал Высшей школы экономики*, 15(5), 24–43.
- Утехин, И. (2019). Поиск и его интерфейсы. *Laboratorium: журнал социальных исследований*, 11(1), 152–165. <https://doi.org/10.25285/2078-1938-2019-11-1-152-165>
- Цветков, Ю. А. (2021). Искусственный интеллект в правосудии. *Закон*, 4, 91–107.

- Esposito, E. (2017). Artificial Communication? The Production of Contingency by Algorithms. *Zeitschrift für Soziologie*, 46(4), 249–265. <https://doi.org/10.1515/zfsoz-2017-1014>
- Etzioni, A., & Etzioni, O. (2017). Should Artificial Intelligence Be Regulated? *Issues in Science and Technology*, 33(4), 32–36.
- Fink, M., & Finck, M. (2022). Reasoned A(I)administration: explanation requirements in EU law and the automation of public administration. *European Law Review*, 47(3), 376–392.
- Ford, K. M., Hayes, P. J., Glymour, C., & Allen, J. (2015). Cognitive Orthoses: Toward Human-Centered AI. *AI Magazine*, 36(4), 5–8. <https://doi.org/10.1609/aimag.v36i4.2629>
- Harvey, D. (2014). *Seventeen Contradictions and the End of Capitalism*. Oxford: Oxford University Press.
- Hickman, E., & Petrin, M. (2021). Trustworthy AI and Corporate Governance: The EU's Ethics Guidelines for Trustworthy Artificial Intelligence from a Company Law Perspective. *European Business Organization Law Review*, 22, 593–625. <https://doi.org/10.1007/s40804-021-00224-0>
- International Committee of the Red Cross (2020). Artificial intelligence and machine learning in armed conflict: A human-centred approach. *International Review of the Red Cross*, 102(913), 463–479. <https://doi.org/10.1007/s40804-021-00224-0>
- Klein, N. (2007). *The Shock Doctrine: The Rise of Disaster Capitalism*. New York: Henry Holt.
- Lee, J.-A., Hilty, R. M., & Liu, K.-C. (Eds.). (2021). *Artificial Intelligence and Intellectual Property*. Oxford: Oxford University Press.
- Oravec, J. A. (2022). The emergence of “truth machines”? Artificial intelligence approaches to lie detection. *Ethics and Information Technology*, 24, 6. <https://doi.org/10.1007/s10676-022-09621-6>
- Rezaev, A. V. (2021). Twelve Theses on Artificial Intelligence and Artificial Sociality. *Monitoring of Public Opinion: Economic and Social Changes*, 1, 20–30. <https://doi.org/10.14515/monitoring.2021.1.1894>
- Shneiderman, B. (2021). Human-centered AI. *Issues in Science and Technology*, 37(2), 56–61.
- Ulnicane, I. (2022). Artificial Intelligence in the European Union: policy, ethics and regulation. In T. Hoerber, I. Cabras, G. Weber (Eds.). *Routledge Handbook of European Integrations* (pp. 254–269). London: Routledge. <https://doi.org/10.4324/9780429262081-19>
- Wolfe, A. (1993). *The Human Difference: Animals, Computers, and the Necessity of Social Science*. Berkeley: University of California Press.

Сведения об авторах



Резаев Андрей Владимирович – доктор философских наук, профессор, руководитель Международной исследовательской лаборатории ТАНДЕМ, Санкт-Петербургский государственный университет

Адрес: 191124, Российская Федерация, г. Санкт-Петербург, ул. Смольного, 1/3

E-mail: rezaev@hotmail.com

ORCID ID: <https://orcid.org/0000-0002-3918-835X>

Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=13004674100>

Web of Science Researcher ID:

<https://www.webofscience.com/wos/author/record/K-3472-2013>

Google Scholar ID: <https://scholar.google.ru/citations?user=Uzv39ccAAAAJ>

РИНЦ Author ID: https://elibrary.ru/author_items.asp?authorid=648768



Трегубова Наталья Дамировна – кандидат социологических наук, доцент кафедры сравнительной социологии, Санкт-Петербургский государственный университет

Адрес: 191124, Российская Федерация, г. Санкт-Петербург, ул. Смольного, 1/3

E-mail: n.tregubova@spbu.ru

ORCID ID: <https://orcid.org/0000-0003-3259-5566>

Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=56645016900>

Web of Science Researcher ID:

<https://www.webofscience.com/wos/author/record/K-3487-2013>

Google Scholar ID: <https://scholar.google.com/citations?user=8dhGr3gAAAAJ&hl>

РИНЦ Author ID: https://elibrary.ru/author_items.asp?authorid=832705

Вклад авторов

А. В. Резаев и Н. Д. Трегубова внесли равный вклад как в формулировку ключевых положений статьи, так и в подготовку рукописи к публикации.

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

Финансирование

Исследование выполнено при финансовой поддержке РФФИ и Министерства по науке и технологиям Тайваня в рамках научного проекта № 21-511-52002.

Тематические рубрики

Рубрика OECD: 5.05 / Law

Рубрика ASJC: 3308 / Law

Рубрика WoS: OM / Law

Рубрика ГРНТИ: 10.07.49 / Планирование и прогнозирование в праве

Специальность ВАК: 5.1.1 / Теоретико-исторические правовые науки

История статьи

Дата поступления – 3 марта 2023 г.

Дата одобрения после рецензирования – 4 мая 2023 г.

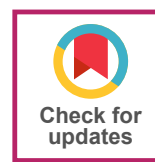
Дата принятия к опубликованию – 16 июня 2023 г.

Дата онлайн-размещения – 20 июня 2023 г.



Research article

DOI: <https://doi.org/10.21202/jdtl.2023.24>



Possibility and Necessity of the Human-Centered Artificial Intelligence in Legal Theory and Practice

Andrey V. Rezaev ✉

Saint Petersburg State University
Saint Petersburg, Russian Federation

Natalia D. Tregubova

Saint Petersburg State University
Saint Petersburg, Russian Federation

Keywords

Algorithm,
artificial intelligence,
artificial sociality,
digital economy,
digital technologies,
human,
human-centered artificial
intelligence,
law,
regulation,
sociology

Abstract

Objective: the paper aims to define the problems juridical theory and practice face with the progress of AI technologies in everyday life and correlate these problems with the human-centered approach to exploring artificial intelligence (Human-Centered AI).

Methods: the research critically analyzes the relevant literature from various disciplines: jurisprudence, sociology, philosophy, and computer sciences.

Results: the article articulates the prospects and problems the legal system confronts with the advancement of digital technologies in general and the tools of AI specifically. The identified problems are correlated with the provisions of the human-centered approach to AI. The authors acknowledge the necessity for AI inventors, as well as the owners of companies participating in the race to develop artificial intelligence technologies, to place humans, not machines, into the focus of attention as a primary value. In particular, special effort should be directed towards collecting and analyzing high-quality data for the organization of artificial intelligence tools development, taking into account that nowadays, the tools of AI are as practical as the data on which they are trained are effective.

✉ Corresponding author

© Rezaev A. V., Tregubova N. D., 2023

The English translation of the original text has been provided by the Editorial of the Journal of Digital Technologies and Law.



This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

The authors formulate three principles of human-centered AI for the legal sphere: 1) a human as a necessary link in the chain of making and executing legal decisions; 2) the need to regulate artificial intelligence at the international law level; 3) formulating “a taboo” for introducing the artificial intelligence technologies.

Scientific novelty: the article manifests one of the first attempts in the Russian-language scientific literature to outline the prospects of developing human-centered AI methodology in jurisprudence. Based on an analysis of special literature, the authors formulate three principles of including artificial intelligence into juridical theory and practice according to the assumptions of a human-centered approach to AI.

Practical significance: the principles and arguments the article advances can be helpful in the legal regulation of artificial intelligence technologies and their harmonious inclusion into legal practices.

For citation

Rezaev, A. V., Tregubova, N. D. (2023). The Possibility and Necessity of the Human-Centered AI in Legal Theory and Practice. *Journal of Digital Technologies and Law*, 1(2), 564–580. <https://doi.org/10.21202/jdtl.2023.24>

References

- Andreev, V. K., Laptev, V. A., & Chucha S. Yu. (2020). Artificial intelligence in the system of electronic justice by consideration of corporate disputes. *Vestnik of Saint Petersburg University. Law*, 11(1), 19–34. (In Russ.). <https://doi.org/10.21638/spbu14.2020.102>
- Baturin, Yu. M., & Polubinskaya, S. V. (2022). Artificial intelligence: legal status or legal regime? *Gosudarstvoiparvo*, 10, 141. (In Russ.). <https://doi.org/10.31857/s102694520022606-7>
- Bykov, A. V., & Narskaya, A. I. (2022). Law, Morality, and Machine Learning: Judges' Perspective on the Essence of Justice and the Prospects of Its Robotization. *Monitoring of Public Opinion: Economic and Social Changes Journal (Public Opinion Monitoring)*, 5, 278–298. (In Russ.). <https://doi.org/10.14515/monitoring.2022.5.2137>
- Dreyfus, H. (1978). *What computers can't do: A critique of artificial reason*. Moscow: Progress. (In Russ.).
- Efimova, L., Mikheeva, I., & Chub, D. (2020). Comparative Analysis of Doctrinal Concepts of Legal Regulating Smart Contracts in Russia and Foreign States. *Journal of the Higher School of Economics*, 4, 78–105. (In Russ.).
- Esposito, E. (2017). Artificial Communication? The Production of Contingency by Algorithms. *Zeitschrift für Soziologie*, 46(4), 249–265. <https://doi.org/10.1515/zfsoz-2017-1014>
- Etzioni, A., & Etzioni, O. (2017). Should Artificial Intelligence Be Regulated? *Issues in Science and Technology*, 33(4), 32–36.
- Fink, M., & Finck, M. (2022). Reasoned A(I) administration: explanation requirements in EU law and the automation of public administration. *European Law Review*, 47(3), 376–392.
- Ford, K. M., Hayes, P. J., Glymour, C., & Allen, J. (2015). Cognitive Orthoses: Toward Human-Centered AI. *AI Magazine*, 36(4), 5–8. <https://doi.org/10.1609/aimag.v36i4.2629>
- Gorokhova, S. S. (2021). Artificial intelligence: an instrument ensuring cybersecurity of the financial sphere or a cyber threat to banks? *Banking Law*, 1, 35–46. (In Russ.). <https://doi.org/10.18572/1812-3945-2021-1-35-46>
- Greger, R. (2017). Judge as an Internet Surfer. Identification of the Circumstances of the Case on the Internet. *Herald of Civil Procedure*, 7(4), 161–173. (In Russ.). <https://doi.org/10.24031/2226-0781-2017-7-4-161-173>
- Harvey, D. (2014). *Seventeen Contradictions and the End of Capitalism*. Oxford: Oxford University Press.
- Hickman, E., & Petrin, M. (2021). Trustworthy AI and Corporate Governance: The EU's Ethics Guidelines for Trustworthy Artificial Intelligence from a Company Law Perspective. *European Business Organization*

- Law Review, 22, 593–625. <https://doi.org/10.1007/s40804-021-00224-0>
- International Committee of the Red Cross (2020). Artificial intelligence and machine learning in armed conflict: A human-centred approach. *International Review of the Red Cross*, 102(913), 463–479. <https://doi.org/10.1007/s40804-021-00224-0>
- Klein, N. (2007). *The Shock Doctrine: The Rise of Disaster Capitalism*. New York: Henry Holt.
- Lee, J.-A., Hilty, R. M. & Liu, K.-C. (Eds.). (2021). *Artificial Intelligence and Intellectual Property*. Oxford: Oxford University Press.
- Lee, K.-F. (2019). *AI Superpowers: China, Silicon Valley and the new world order*. Moscow: Mann, Ivanov i Ferber. (In Russ.).
- Lessig, L. (2019). Artificial intelligence is going to oust a wide circle of lawyers. *Zakon*, 5, 8–30. (In Russ.).
- Mazhorina, M. (2020). Cyberplace and Methodology of International Private Law. *Journal of the Higher School of Economics*, 2, 230–253. (In Russ.).
- Oravec, J. A. (2022). The emergence of “truth machines”? Artificial intelligence approaches to lie detection. *Ethics and Information Technology*, 24, 6. <https://doi.org/10.1007/s10676-022-09621-6>
- O’Neil, C. (2018). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Moscow: AST. (In Russ.).
- Panchenko, V. Yu. (2012). Information availability of legal assistance: ideal and real state. *Agrarnoye i zemelnoye parvo*, 11(95), 95–102. (In Russ.).
- Rezaev, A. V. (2021). Twelve Theses on Artificial Intelligence and Artificial Sociality. *Monitoring of Public Opinion: Economic and Social Changes*, 1, 20–30. <https://doi.org/10.14515/monitoring.2021.1.1894>
- Rezaev, A. V. & Tregubova, N. D. (2019). Artificial intelligence, On-line Culture, Artificial Sociality: Definition of the Terms. *Monitoring of Public Opinion: Economic and Social Changes*, 6, 35–47. <https://doi.org/10.14515/monitoring.2019.6.03>
- Rudenko, N. (2020). Sociotechnical barriers to developing autonomous vehicles in Russia. In L. Zemnukhova, K. Glazkov, O. Logunova, A. Maksimova, D. Sivkov, & N. Rudenko, *The Adventures of Technologies: Digitalization Barriers in Russia* (17–70). Moscow – Saint Petersburg: FNISTS RAN. (In Russ.). <https://doi.org/10.31119/978-5-89697-339-3>
- Rusakova, E. (2020). The integration of modern digital technologies to the legal proceedings of People’s Republic of China and Singapore, *Gosudarstvo i parvo*, 9, 102. (In Russ.). <https://doi.org/10.31857/s102694520011323-6>
- Russell, S., & Norvig, P. (2007). *Artificial Intelligence: A Modern Approach* (2d ed.). Moscow: Vilyams. (In Russ.).
- Shneiderman, B. (2021). Human-centered AI. *Issues in Science and Technology*, 37(2), 56–61.
- Stepanov, O., Pechegin, D., & Diakonova, M. (2021). Towards the Issue of Digitalization of Judicial Activities. *Journal of the Higher School of Economics*, 5, 4–23. (In Russ.). <https://doi.org/10.17323/2072-8166.2021.5.4.23>
- Talapina, E. V. (2021). Artificial intelligence and legal expertise in public administration. *Vestnik of Saint Petersburg University. Law*, 12(4), 865–881. (In Russ.). <https://doi.org/10.21638/spbu14.2021.404>
- Talapina, E. V. (2022). The right to informational self-determination: on the edge of public and private. *Law. Journal of the Higher School of Economics*, 15(5), 24–43. (In Russ.).
- Tsvetkov, Yu. A. (2021). Artificial Intelligence in Justice. *Zakon*, 4, 91–107. (In Russ.).
- Ulnicane, I. (2022). Artificial Intelligence in the European Union: policy, ethics and regulation. In T. Hoerber, I. Cabras, G. Weber (Eds.). *Routledge Handbook of European Integrations* (pp. 254–269). London: Routledge. <https://doi.org/10.4324/9780429262081-19>
- Utekhin, I. (2019). Search and Interfaces for Search. *Laboratorium: Russian Review of Social Research*, 11(1), 152–165. (In Russ.). <https://doi.org/10.25285/2078-1938-2019-11-1-152-165>
- Vavilin, E. V. (2021). Artificial intelligence as a participant in civil relations: the transformation of law. *Vestnik Tomskogo gosudarstvennogo universiteta. Pravo*, 42, 135–146. (In Russ.). <https://doi.org/10.17223/22253513/42/11>
- Voinikanis, E. A. (2020). Regulation of big data and intellectual property right: common approaches, problems and prospects of development. *Zakon*, 7, 135–156. (In Russ.).
- Weizenbaum, J. (1982). *Computer Power and Human Reason: From Judgment to Calculation*. Moscow: Radio i svyaz. (In Russ.).
- Wiener, N. (1983). *Cybernetics: Or Control and Communication in the Animal and the Machine* (2d ed.). Moscow: Nauka; Glavnaya redaktsiya izdaniy dlya zarubezhnykh stran. (In Russ.).
- Wolfe, A. (1993). *The Human Difference: Animals, Computers, and the Necessity of Social Science*. Berkeley: University of California Press.
- Zuboff, Sh. (2022). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Moscow: Izd-vo Instituta Gaidara. (In Russ.).

Authors information



Andrey V. Rezaev – Doctor of Philosophical Sciences, Professor, Head of the International Research Laboratory TANDEM at the Faculty of Sociology, St Petersburg State University

Address: 1/3 Smolnogo Str., 191124 Saint Petersburg, Russian Federation

E-mail: rezaev@hotmail.com

ORCID ID: <https://orcid.org/0000-0002-3918-835X>

Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=13004674100>

Web of Science Researcher ID:

<https://www.webofscience.com/wos/author/record/K-3472-2013>

Google Scholar ID: <https://scholar.google.ru/citations?user=Uzv39ccAAAAJ>

РИНЦ Author ID: https://elibrary.ru/author_items.asp?authorid=648768



Natalia D. Tregubova – PhD (Sociology), Associate Professor of the Department of Comparative Sociology, St Petersburg State University

Address: 1/3 Smolnogo Str., 191124 Saint Petersburg, Russian Federation

E-mail: n.tregubova@spbu.ru

ORCID ID: <https://orcid.org/0000-0003-3259-5566>

Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=56645016900>

Web of Science Researcher ID:

<https://www.webofscience.com/wos/author/record/K-3487-2013>

Google Scholar ID: <https://scholar.google.com/citations?user=8dhGr3gAAAAJ&hl>

РИНЦ Author ID: https://elibrary.ru/author_items.asp?authorid=832705

Authors' contributions

A. V. Rezaev and N. D. Tregubova contributed equally to the formulation of the article's key provisions and preparation of the manuscript for publication.

Conflict of interests

The authors declare no conflict of interests.

Financial disclosure

The study was supported by RFBR and MOST, the research project No. 21-511-52002.

Thematic rubrics:

OECD: 5.05 / Law

PASJC: 3308 / Law

WoS: OM / Law

Article history

Date of receipt – March 3, 2023

Date of approval – May 4, 2023

Date of acceptance – June 16, 2023

Date of online placement – June 20, 2023

