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.

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

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, укоренные 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; Claeys, 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; Claeys, 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 (Schiavone, 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 (Schiavone, 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» (Schiavone, 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 оправданна 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; самим 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 интрузивных 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 взором 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 соотнести 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 несуществующим[1], т. е. ничтожным и не имеющим юридической силы[2], 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 непременным 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 внутрипроцессуальная 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 конституциональные 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 сувереном: 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 окончательным[4].

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 неэксклюзивной 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 найме, 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)(a) Общего регламента по защите данных), кроме случав, 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 ЮНЕСКО[8]. 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 черт[10].

---

9 Там же.

10 Bocc, 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 укоренённых 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 уравновешивания 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

Mannozzi, 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

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.

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 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, 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

Mannozzi, 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

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