**Научная статья**
**УДК 34:004:17:004.8:342.7**
**EDN:** https://elibrary.ru/egkppn
**DOI:** https://doi.org/10.21202/jdtl.2025.7

# Искусственный интеллект в здравоохранении: баланс инноваций, этики и защиты прав человека

**Педро Мигель Алвес Рибейро Коррейя** ✉
Коимбрский университет, Коимбра, Португалия

**Рикардо Лопес Динис Педро**
Лиссабонский университет, Лиссабон, Португалия

**Сусана Видейра**
Лиссабонский университет, Лиссабон, Португалия

## Ключевые слова

защита данных,
здравоохранение,
искусственный интеллект,
права человека,
право,
правовое регулирование,
предиктивная аналитика,
фундаментальные права,
этика,
этическое регулирование

## Аннотация

**Цель**: определить ключевые этические, правовые и социальные вызовы, связанные с использованием искусственного интеллекта в здравоохранении, а также разработать рекомендации для создания адаптивных правовых механизмов, способных обеспечить баланс между инновациями, этическим регулированием и защитой фундаментальных прав человека.

**Методы**: в ходе исследования был реализован многоаспектный методологический подход, интегрирующий классические правовые методы анализа с современными инструментами сравнительного правоведения. Данное исследование охватывает как фундаментальные основы правового регулирования цифровых технологий в медицинской сфере, так и глубокий анализ этических, правовых и социальных импликаций внедрения искусственного интеллекта в систему здравоохранения. Такой комплексный подход позволил обеспечить всестороннее понимание проблематики и сформировать обоснованные выводы относительно перспектив развития данной области.

**Результаты**: выявлен ряд серьезных проблем, связанных с использованием искусственного интеллекта в здравоохранении. К ним относятся необъективность данных, непрозрачность сложных алгоритмов и риски нарушения неприкосновенности частной жизни. Эти проблемы могут подорвать доверие общества к технологиям искусственного интеллекта и усугубить неравенство в доступе к медицинским услугам. Авторы приходят к выводу, что интеграция искусственного интеллекта в систему здравоохранения должна осуществляться с учетом фундаментальных прав, таких как защита данных и запрет дискриминации, а также соответствовать этическим нормам.

✉ Контактное лицо

**Научная новизна**: состоит в предложении эффективных механизмов управления для снижения рисков и максимизации потенциала искусственного интеллекта в кризисных ситуациях. Особое внимание уделяется регулятивным мерам, таким как оценка влияния, предусмотренная Законом об искусственном интеллекте. Эти меры играют ключевую роль в выявлении и минимизации рисков, связанных с высокорисковыми системами искусственного интеллекта, обеспечивая соблюдение этических норм и защиту основных прав.

**Практическая значимость**: заключается в разработке адаптивных правовых механизмов, которые поддерживают демократические нормы и оперативно реагируют на возникающие вызовы в области общественного здравоохранения. Предложенные механизмы позволяют достичь баланса между использованием искусственного интеллекта для управления кризисными ситуациями и сохранением прав человека. Это способствует укреплению доверия к системам искусственного интеллекта и их устойчивому положительному влиянию на общественное здравоохранение.

## Для цитирования

Коррейя, П. М. А. Р., Педро, Р. Л. Д., & Видейра, С. (2025). Искусственный интеллект в здравоохранении: баланс инноваций, этики и защиты прав человека. *Journal of Digital Technologies and Law*, *3*(1), 143–180. https://doi.org/10.21202/jdtl.2025.7

## Содержание

## Введение

В настоящее время борьба с болезнями и пандемиями ведется на уровне врачей-практиков, учреждений здравоохранения и целых государств. Может ли искусственный интеллект (далее – ИИ) помочь нам в этой сфере? Какие ограничения накладывает при этом необходимость соблюдения фундаментальных прав?

На протяжении всей истории человечество сталкивалось с разрушительными последствиями эпидемий и пандемий, от бубонной чумы до испанского гриппа. Вспышки болезней изменяли целые сообщества и даже ставили под сомнение само существование человека. Сегодня, когда мы живем во все более взаимосвязанном мире, все более реальной становится угроза новых и быстро распространяющихся болезней. Глобализация все чаще выступает в роли обоюдоострого меча, способствуя как сотрудничеству, так и быстрой передаче патогенов через границы государств (Jones et al., 2008; Morse et al., 2012).

Искусственный интеллект сегодня рассматривается как новое оружие в этой извечной борьбе. Эта мощная революционная технология (или совокупность технологий), некогда относившаяся к области научной фантастики, обладает огромным потенциалом для кардинального изменения методов борьбы с эпидемиями и пандемиями. Она может стать мощным оружием в нашем арсенале для борьбы с болезнями. С помощью методов искусственного интеллекта можно будет анализировать огромные массивы данных, предсказывать вспышки и исход болезней, ускорять поиск лекарств и даже персонализировать стратегии лечения (Syrowatka et al., 2021). По крайней мере, так принято считать.

К примеру, искусственный интеллект можно использовать для прогнозирования чрезвычайных ситуаций, поскольку он способен просеивать огромный объем информации из социальных сетей, новостных сообщений и даже спутниковых снимков, выявляя ранние признаки вспышки заболевания до того, как она перерастет в полномасштабную пандемию. Нетрудно представить, как система искусственного интеллекта обнаруживает резкий рост числа запросов о симптомах, похожих на грипп, в определенном регионе и немедленно начинает исследовать это явление, что потенциально может в зародыше пресечь вспышку заболевания (Wong et al., 2023).

Другой пример – использование искусственного интеллекта для ускорения поиска лекарств (если фармацевтические компании вообще собираются делать это, а не только облегчать хронические заболевания, что приносит им постоянный доход). Открытие химических соединений (в том числе вакцин) для медицинских целей всегда было медленным и сложным процессом; часто проходят годы, пока не будет получен результат, а иногда он и вовсе не достигается. Искусственный интеллект может революционизировать этот процесс путем анализа молекулярных структур веществ для выявления потенциальных кандидатов в лекарственные препараты или для перепрофилирования существующих лекарств в новые. Это значительно сократит время, необходимое для того, чтобы жизненно важные лекарства попали к пациентам (Matsuzaka & Yashiro, 2022).

Еще одна возможность состоит в применении индивидуального подхода к лечению. Искусственный интеллект может анализировать индивидуальные генетические особенности и историю болезни пациента, предсказать его реакцию на различные варианты лечения. Это откроет пути к персонализированной медицине и позволит врачам адаптировать планы лечения для достижения максимальной эффективности (Topol, 2019).

Также искусственный интеллект может использоваться в данном контексте для предсказания траектории развития вспышки заболевания или пандемии. Модели ИИ способны анализировать данные о развитии и характеристиках заболевания, что позволяет чиновникам системы здравоохранения стратегически распределять ресурсы и осуществлять целенаправленные меры для сдерживания распространения болезни (Ferguson et al., 2006).

Последний пример (из множества других, не перечисленных здесь) – использование искусственного интеллекта для улучшения отслеживания контактов, поскольку ИИ может анализировать данные о контактах и поездках, точно определяя лиц с высоким риском заражения. Это поможет медицинским работникам определить приоритетность тестирования и карантинных мер, потенциально сдерживая распространение патогена (Fetzer & Graeber, 2021).

Однако этот путь не лишен преград. Предвзятость данных, «черного ящика», или «серого ящика», сложные алгоритмы, а также постоянно присутствующий риск чрезмерной зависимости от искусственного интеллекта – все это создает потенциальные и значительные трудности (DeCamp & Tilburt, 2019).

Хотя мы перечислили ряд потенциальных возможностей искусственного интеллекта, данная работа посвящена не столько его преимуществам в борьбе с болезнями, эпидемиями или пандемиями, сколько критическим соображениям и предостережениям, которые следует учитывать пользователям по мере того, как использование этих подходов все чаще становится предметом размышлений. Мы рассмотрим различные проблемы, с которыми столкнемся, прежде чем искусственный интеллект сможет стать маяком надежды в мире, над которым постоянно нависает угроза широкомасштабных, всеобщих и острых вспышек заболеваний.

Кроме того, мы проанализируем влияние ИИ на фундаментальные права, а также проблему использования искусственного интеллекта для борьбы с пандемией и ее связь с соблюдением фундаментальных прав.

## 1. Использование искусственного интеллекта в борьбе с пандемиями и причины неудач в этой области

Потенциальные возможности ИИ неразрывно связаны с проблемами, которые необходимо решать. Широко известно высказывание Джорджа Бокса о том, что «все модели ошибочны, но некоторые из них полезны» (Box, 1979). Неудивительно, что, несмотря на сильные стороны моделей искусственного интеллекта, их эффективность сдерживается рядом ограничений, которые могут превратить эти решения в обоюдоострый меч.

Один из главных недостатков моделей искусственного интеллекта заключается в том, что эти модели хороши лишь настолько, насколько хороши данные, на которых они обучаются. Неточные, неполные или предвзятые данные могут привести к ненадежным и потенциально вредным результатам (Gianfrancesco et al., 2018). Ограниченный доступ к медицинским данным в режиме реального времени в некоторых регионах или проблемы с конфиденциальностью еще больше снижают возможности искусственного интеллекта.

Еще одна проблема широко известна как свойство «черного ящика» или, в более мягкой формулировке, свойство «серого ящика». Внутренняя работа сложных алгоритмов чрезвычайно сложна, и люди не всегда способны понять процессы принятия

решений искусственным интеллектом. Эта непрозрачность, в свою очередь, подрывает доверие и усложняет задачу обнаружения и устранения скрытых предубеждений и различных других проблем (Mittelstadt et al., 2016).

Ошибкой было бы также чрезмерно полагаться на значительные возможности моделей искусственного интеллекта, не признавая их фундаментальных ограничений. Другими словами, необходимо уделять должное внимание основополагающим стратегиям общественного здравоохранения, таким как отслеживание контактов, вакцинация и кампании по повышению осведомленности населения. Все эти подходы проверены временем и должны продолжать играть жизненно важную роль в эффективном управлении вспышками заболеваний (Silva et al., 2022).

## 2. Примеры неудач искусственного интеллекта: извлеченные уроки и дальнейшие действия

Пандемия COVID-19 стала полигоном для испытания искусственного интеллекта в борьбе с болезнями, но результаты оказались неоднозначными. Рассмотрим ряд примеров.

В самом начале пандемии некоторые модели искусственного интеллекта сильно переоценили распространение вируса из-за ограниченности исходных данных и быстро меняющейся ситуации. Это привело к панике и выделению ненужных ресурсов. В других случаях чат-боты на базе искусственного интеллекта, созданные для ответов на вопросы в области здравоохранения, были перегружены и иногда предоставляли неверную информацию. Это подчеркивает необходимость наличия надежных обучающих данных и четких ограничений для приложений с искусственным интеллектом (Bajwa et al., 2021; Gürsoy & Kaya, 2023).

Таким образом, можно утверждать, что, только признав ограничения искусственного интеллекта и сосредоточившись на принципах ответственного развития, человечество сможет использовать его возможности для более здорового будущего. Для создания надежных моделей искусственного интеллекта крайне важны приоритетность качества данных и ответственная практика их сбора, а также устранение предвзятости данных и обеспечение их конфиденциальности. Разработка надежных методов снижения предвзятости алгоритмов искусственного интеллекта, основанных на проверке справедливости и расширении спектра данных, также должна помочь в выявлении и устранении потенциальной предвзятости на самых начальных этапах. Необходимо поддерживать исследования в области объяснимого искусственного интеллекта (explainable artificial intelligence, XAI). Это поможет заинтересованным сторонам понять, как модели искусственного интеллекта приходят к своим выводам, укрепит доверие и позволит выявлять потенциальные проблемы на ранней стадии (Jobin et al., 2019).

Важно придерживаться сбалансированных подходов, когда искусственный интеллект используется наряду с традиционными мерами в области общественного здравоохранения и дополняет, а не заменяет их. Решая эти проблемы и поощряя ответственное развитие искусственного интеллекта, заинтересованные стороны смогут пользоваться всеми его возможностями и стать лучше подготовленными к будущим пандемиям (Benke & Benke, 2018). Искусственный интеллект может стать мощным оружием в арсенале человечества, но только при условии его разумного использования, как показано ниже.

## 3. Проблема «мусор на входе – мусор на выходе»

Фундаментальный принцип искусственного интеллекта, в частности машинного обучения, можно сформулировать так: «мусор на входе – мусор на выходе» (Breiman, 2001).

Рассмотрим основные типы «мусорных» данных. Во-первых, данные могут быть неточными. К ним относятся орфографические ошибки, опечатки, фактические ошибки, устаревшая информация (Halevy et al., 2009). Представим себе искусственный интеллект, обученный на новостных статьях с большим количеством опечаток; ему будет трудно понимать язык. Во-вторых, данные могут быть неполными. Отсутствующие значения или отдельные данные будут искажать понимание модели (Little & Rubin, 2019). Например, модель для прогнозирования оттока клиентов (причин, почему пациенты решают не возвращаться в данную больницу) пропустит важные данные, если не будет учитывать отзывы клиентов. В-третьих, данные могут быть необъективными. Данные, необъективно представляющие определенную группу, приведут к дискриминационным результатам (Berk, 1983). Так, если искусственный интеллект, используемый для принятия решений о приеме на работу, обучали в основном на резюме мужчин, то он будет отдавать предпочтение кандидатам-мужчинам. И в-четвертых, данные могут быть нерелевантными. Информация, не имеющая отношения к решаемой задаче, исказит работу модели (Greiner et al., 1997). Так, искусственный интеллект для анализа настроения (понимания эмоций в тексте) в психиатрической клинике может быть перегружен нерелевантными эмодзи в представленном наборе данных.

Это также помогает понять, каковы могут быть последствия «мусора на входе». Во-первых, поддерживается предвзятость: искусственный интеллект может усиливать существующие в обществе предубеждения, если таковые содержатся в обучающих данных (Bazarkina & Pashentsev, 2020). Это может негативно отразиться на результатах деятельности ИИ в таких областях, как одобрение кредитов, распознавание лиц и прогнозирование в сфере уголовного правосудия. Далее, снижается точность и надежность, поскольку модели, обученные на неточных данных, будут выдавать ненадежные результаты (Shin & Park, 2019). Представьте себе искусственный интеллект для прогнозирования патологий, обученный на неверных показаниях температуры у пациентов; его диагноз будет неточным. Кроме того, при обучении моделей на неверных данных значительные ресурсы будут потрачены впустую (Hulten, 2018).

Необходимо знать основные методы борьбы с «мусором на входе». С одной стороны, нужно вкладывать средства в очистку и обработку данных. Для обеспечения качества данных используются такие методы, как проверка данных, исправление ошибок и фильтрация. Это трудоемкий процесс, но крайне важный для получения надежного искусственного интеллекта (Wang & Shi, 2011). С другой стороны, для решения такой проблемы, как неполнота данных, необходимо использовать синтетические данные в дополнение к существующим базам данных (Mumuni & Mumuni, 2022). Например, создание реалистичных изображений различных лиц помогает уменьшить предвзятость при распознавании лиц. Другой пример – использование алгоритмических методов обнаружения предвзятости для выявления и снижения предвзятости в самих алгоритмах искусственного интеллекта. Сюда входит анализ процесса принятия решений моделью с целью выявления скрытых предубеждений

(Kordzadeh & Ghasemaghaei, 2022). Еще одна техника, направленная на повышение прозрачности моделей искусственного интеллекта, – это объясняемый искусственный интеллект. Она позволяет понять, как модель приходит к своим выводам, и выявить потенциальные предубеждения или ошибки (Arrieta et al., 2020).

Решение проблемы «мусор на входе – мусор на выходе» критически важно для создания надежного и этичного искусственного интеллекта в будущем. По мере того как искусственный интеллект все больше интегрируется в жизнь каждого человека, обеспечение качества данных и уменьшение их предвзятости приобретает огромное значение (Jobin et al., 2019). В этом направлении уже прилагаются определенные усилия. Стандартизация и регулирование не решат проблему полностью, но могут помочь в ее решении. Разработка руководящих принципов и правил ответственной разработки и внедрения искусственного интеллекта повышает качество и надежность данных[1]. Также ведется просветительская и информационная работа с населением. Повышение осведомленности о потенциальных недостатках искусственного интеллекта и важности ответственного подхода к разработке способствует укреплению общественного доверия (Kandlhofer et al., 2023). Кроме того, междисциплинарное сотрудничество между разработчиками искусственного интеллекта и экспертами, включая специалистов в области этики, науки о данных, государственных деятелей, имеет решающее значение для создания надежных и ответственных систем искусственного интеллекта (Bisconti et al., 2023). Это еще одна тенденция, которая может предотвратить или замедлить попадание в ловушку «мусор на входе – мусор на выходе».

## 4. Проблема устойчивости, о которой почти не говорят

Ключевым фактором должна стать устойчивость решений в области искусственного интеллекта.

Первостепенное значение имеет энергопотребление. Обучение большой языковой модели, такой как GPT-3, требует столько же энергии, сколько несколько автомобилей за весь срок службы. По оценкам некоторых исследований, энергопотребление при обучении одной большой языковой модели составляет около 1,5 МВт·ч[2]. Центры обработки данных, в которых размещаются системы искусственного интеллекта, по самым скромным оценкам, потребляют от 1 до 3 % мирового объема электроэнергии[3].

Кроме того, все большую озабоченность вызывает потребление воды. Центры обработки данных в значительной степени зависят от воды для охлаждения: по оценкам, только в Соединенных Штатах Америки они потребляют до 1,7 млрд галлонов воды в год[4]. Расход воды при работе искусственного интеллекта может быть значительным даже для отдельных пользователей. Один запрос к большой языковой модели расходует небольшую бутылку воды[5].

---

1 European Commission. (2021). Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Legislative Instruments (COM(2021) 206 final). https://clck.ru/3DzaGK

2 Luccioni, S. (2023, April 12). The mounting human and environmental costs of generative AI. Ars Technica. https://clck.ru/3DzaKM

3 AI Now Institute. (2023). Algorithmic Accountability: Moving Beyond Audits. https://clck.ru/3DzaLM

4 Meredith, S. (2023, December 6). A 'thirsty' generative AI boom poses a growing problem for Big Tech. CNBC. https://clck.ru/3DzaLy ; Microsoft (2022). 2022 Environmental Sustainability Report. https://clck.ru/3DzaLy

5 Там же.

Однако главным ограничивающим фактором станут, вероятно, требования к хранению данных (Susskind, 2020). Объем данных, генерируемых во всем мире, растет экспоненциально, удваиваясь примерно каждые два года. Нынешние технологии хранения данных, такие как жесткие диски, приближаются к своим физических пределам с точки зрения миниатюризации и емкости[6].

Важно отметить, что эти характеристики постоянно меняются по мере развития технологий. Исследователи активно разрабатывают более энергоэффективные модели искусственного интеллекта, водосберегающие системы охлаждения для центров обработки данных и новые способы хранения данных с большей емкостью (Chen, 2016).

Предсказать, когда именно возможности для хранения данных будут исчерпаны, довольно сложно. Рост объема данных происходит по экспоненте, но и технологии хранения постоянно развиваются. Однако можно с уверенностью сказать, что в ближайшем будущем физическое пространство всей планеты для хранения данных не понадобится. Мы становимся свидетелями гонки между экспоненциальным ростом и законом Мура (Theis & Wong, 2017). Объем создаваемых данных действительно растет экспоненциально, удваиваясь примерно каждые два года. Даже если рост замедлится до 20 % в год (сейчас он составляет около 70 %), то в долгосрочной перспективе темп роста все равно будет неустойчивым. Существующая инфраструктура и энергетические ограничения будут осложнять поддержание таких темпов[7]. Однако емкость систем хранения данных также быстро растет, следуя тенденции, аналогичной закону Мура (удвоение плотности транзисторов в интегральных схемах примерно каждые два года). Методы сжатия позволяют значительно сократить физическое пространство, необходимое для хранения информации. Хотя в какой-то момент рост объема данных может превысить емкость хранилища, развитие технологий хранения данных, таких как твердотельные накопители, и совершенствование методов сжатия данных могут решить эту проблему (Chen, 2016). Также приводятся аргументы, что не все данные нужно вечно хранить в этих системах. Значительная часть данных не требует постоянного хранения. Через некоторое время можно удалить блоги, временные файлы и часть развлекательного контента. Эффективное управление данными и определение приоритетных категорий данных могут значительно сократить потребности в системах хранения (Arass & Souissi, 2018). Однако в этом случае предполагаемый сверхинтеллект может опуститься до человеческого уровня. Менее совершенная память влечет за собой несовершенные решения, большее количество ошибок и меньшую надежность: иными словами, человеческую производительность. Совершенно новые, пока еще неизвестные технологии хранения данных могут отсрочить эту неизбежность; исследователи изучают альтернативные решения с гораздо большей емкостью, чем традиционные жесткие диски. К ним относятся такие технологии, как ДНК-хранилища, позволяющие хранить огромные объемы данных на очень компактном пространстве (как это делают все живые организмы в своем геноме). Пока эти технологии находятся на ранних стадиях разработки, но обладают огромным потенциалом для долгосрочного архивирования данных (Goldman et al., 2013). К ним относятся и гологра-

6 Rydning, D., Reinsel, J., & Gantz, J. (2018). The digitization of the world from edge to core. International Data Corporation. https://clck.ru/3DzaNN

7 Там же.

фические хранилища, использующие лазерную технологию для хранения данных в трех измерениях, обеспечивая гораздо более высокую плотность, чем традиционные методы (Lin et al., 2020).

Несмотря на все надежды, исследователь Vopson (2020) убедительно подсчитал, сколько лет потребуется для того, чтобы вся масса Земли была отдана под хранение данных при нескольких сценариях ежегодного роста объемов информации. Его подсчет основан на количестве доступных атомов и не зависит от гипотетических новых технологий и методов управления эффективностью хранения данных. Автор приводит следующие значения: 4500 лет при росте объемов информации на 1 % в год, 918 лет при росте на 5 %, 246 лет при росте на 20 % и около 110 лет при росте на 50 % в год. Земная кора составляет всего 0,7 % от общего объема планеты, и это тот объем, который человечество может, хоть и гипотетически, использовать для хранения информации. Таким образом, у нас осталось менее столетия (скорее всего, 30–50 лет или даже меньше) до информационной катастрофы. В соответствующих исследованиях Института AI Now[8] и Стэндфордского института искусственного интеллекта, ориентированного на человека[9], приводятся другие цифры и используются иные подходы, но говорится о тех же тенденциях.

## 5. Управление как ключ к решению проблем: государственные показатели и доступность данных укрепляют организационные ценности и способствуют устойчивости национальной системы здравоохранения

Мы утверждаем, что недостающее звено, которое позволит адекватно преодолеть разрыв между традиционной практикой здравоохранения и подходом к реагированию на пандемии с использованием искусственного интеллекта, – это эффективное управление. Только государственные меры, соответствующие принципам надлежащего управления, могут стать основой для интеграции искусственного интеллекта с более традиционными методами.

В работе Correia с соавторами (2020a) были рассмотрены традиционные меры, лежащие в основе надежной национальной системы здравоохранения во время пандемий. В частности, авторы рассматривают локдауны, отслеживание контактов и кампании по вакцинации. Эти меры могут быть реализованы для замедления распространения вирусов, защиты уязвимых групп населения и достижения коллективного иммунитета. Они воплощают в себе приоритетность общественного здоровья и демонстрируют, что правительство несет ответственность перед своими гражданами. Второй важный момент, который подчеркивают авторы, заключается в том, что данные играют решающую роль в мониторинге уровня заражения, отслеживании распределения ресурсов и понимании результатов лечения пациентов. Именно это позволяет принимать решения, что, в свою очередь, укрепляет ценность доказательной практики и в конечном итоге способствует эффективному использованию ресурсов в системе здравоохранения.

---

8 AI Now Institute. (2023). Algorithmic Accountability: Moving Beyond Audits. https://clck.ru/3DzaQC

9 Stanford Institute for Human-Centered Artificial Intelligence. (2023). Sustainability and AI. https://clck.ru/3DzaRd

Таким образом, появляется возможность достичь устойчивости через эффективность, при этом искусственный интеллект может усилить действие традиционных мер. Хотя традиционные меры по-прежнему важны и, вероятно, всегда будут важны, искусственный интеллект дает возможность значительного повышения их эффективности и усиления устойчивости национальных систем здравоохранения. Одним из непосредственных воплощений этой идеи может стать использование моделей искусственного интеллекта для анализа исторических данных, выявления закономерностей и прогнозирования возникновения или распространения вспышек заболеваний, эпидемий и пандемий. Это позволит принимать упреждающие меры в системе здравоохранения, включая раннее оповещение, создание запасов жизненно важных материалов и стратегическое развертывание ресурсов. В частности, оптимизация распределения ресурсов с помощью алгоритмов искусственного интеллекта может быть легко использована в реальном времени при анализе данных об уровне инфекций, пропускной способности больниц и наличии материальных ресурсов. Это даст возможность динамически распределять медицинский персонал, оборудование и критически важные материалы в районах, испытывающих наибольшую нагрузку (Correia et al., 2021, 2022), и, следовательно, обеспечивать эффективное управление ресурсами. Более продвинутые и, соответственно, сложные варианты применения предусматривают разработку персонализированных планов лечения. Это потребует анализа индивидуальных особенностей пациента, таких как история болезни и генетические данные. Искусственный интеллект потенциально может помочь медицинским работникам в разработке планов лечения для достижения максимальной эффективности, способствуя ускорению сроков выздоровления, улучшению результатов лечения пациентов и снижению нагрузки на систему здравоохранения (Jiang et al., 2017).

Таким образом, становится очевидным, что эффективность управления в борьбе с пандемией (независимо от того, используется искусственный интеллект или нет) зависит от наличия высококачественных, всеобъемлющих данных, собранных с помощью традиционных мер и методов, таких как отслеживание контактов и истории болезней пациентов (Wu et al., 2022). Однако при правильной организации управления должна учитываться также проблема конфиденциальности данных. Это касается сбора и использования данных о пациентах, в том числе для обучения искусственного интеллекта. Необходимо также обеспечить анонимность данных и надежность протоколов информационной безопасности для поддержания доверия со стороны общества (Smidt & Jokonya, 2021).

Эффективное использование искусственного интеллекта в здравоохранении требует беспрепятственного обмена данными между различными медицинскими учреждениями и совместимости между операционными системами (O'Reilly-Shah et al., 2020). Первостепенное значение имеет наличие стандартизированных форматов данных и безопасных каналов связи для обмена данными (Sass et al., 2020).

В заключение следует отметить, что для создания устойчивых систем здравоохранения можно реализовать симбиотические отношения. Традиционные меры общественного здравоохранения, доступность данных и искусственный интеллект не являются отдельными сущностями, но могут стать взаимосвязанными элементами в борьбе с пандемиями. Существующая инфраструктура данных и опыт применения традиционных мер создают благоприятную почву для интеграции (Baclic et al., 2020). Используя возможности искусственного интеллекта в сочетании с устоявшейся практикой, национальные системы здравоохранения могут добиться

большей эффективности, персонализировать подходы к лечению и в конечном итоге обеспечить свою долгосрочную устойчивость перед лицом будущих событий (Gunasekeran et al., 2021). Другими словами, надежные методы управления и твердые организационные принципы должны стать залогом для будущей интеграции искусственного интеллекта в эту важнейшую область.

## 6. Управление как ключ к решению проблем: управление укрепляет организационные ценности и способствует эффективности антикризисных мер

Приведенное выше утверждение имеет огромное значение в контексте применения искусственного интеллекта в борьбе с пандемиями. Это связано с тем, что эффективные методы управления обеспечивают основу и руководящие принципы для ответственного и этичного использования искусственного интеллекта в управлении кризисами, ярким примером которых являются пандемии.

Управление предполагает открытую коммуникацию и ответственность лиц, принимающих решения, за свои действия. Это крайне важно для укрепления доверия общественности к решениям на базе искусственного интеллекта, используемым во время пандемий, например, приложениям для отслеживания контактов. Чтобы избежать опасений общественности, необходимо четко объяснять, как используется искусственный интеллект и как обеспечивается конфиденциальность данных (Galetsi et al., 2022). Эффективное управление также способствует сотрудничеству, включая обмен данными, и координации действий различных заинтересованных сторон, в том числе помогает при распределении ресурсов, разработке вакцин и стратегий коммуникации. Такое сотрудничество и взаимодействие охватывают государственные органы, медицинские учреждения, исследовательские организации и частный сектор (Bulled, 2023).

Эффективное управление воплощается в конкретных организационных принципах, которые должны определять развитие искусственного интеллекта и его использование в условиях пандемии. При правильном применении оно способствует равенству и справедливости, адекватному распределению ресурсов и преодолению цифрового разрыва. Это, в свою очередь, может гарантировать, что инструменты искусственного интеллекта не будут усугублять существующее социальное неравенство (Margetts, 2022). Например, приложения для отслеживания контактов с помощью искусственного интеллекта должны быть доступны для всех групп населения и не должны несправедливо воздействовать на определенные группы. Управление также может стать определяющим фактором в создании надежных протоколов конфиденциальности и безопасности данных. Это способствует защите информации граждан, обеспечивая при этом ответственный сбор данных и их использование для разработки искусственного интеллекта при ликвидации последствий пандемии. Во время пандемии крайне важно найти баланс между инновационными решениями и безопасностью данных (Zhang et al., 2022). Кроме того, надлежащее управление необходимо при принятии решений на основе фактических данных, так как оно создает традицию опоры на научные данные и доказательства для обоснования решений. Это прекрасно согласуется с основным принципом искусственного интеллекта, который использует анализ данных для выработки выводов и рекомендаций, в том числе для руководителей системы здравоохранения (Rubin et al., 2021).

Добавим, что эффективное и устойчивое реагирование в условиях пандемии требует перспективного подхода и долгосрочного планирования. Эффективные методы управления способствуют устойчивости антикризисного управления в нескольких направлениях. Задача руководства – обеспечить долгосрочные инвестиции в инфраструктуру и поддержание аппаратного и программного обеспечения, а также экспертных знаний, необходимых для разработки и внедрения искусственного интеллекта в здравоохранении. Это включает в себя инвестиции в научные исследования и разработки, программы подготовки специалистов по искусственному интеллекту в медицинских учреждениях и создание надежных систем управления данными (Balog-Way & McComas, 2022). Необходимо также разрабатывать перспективные стратегии, создавать гибкие механизмы, способные адаптироваться к меняющимся угрозам и пандемиям с новыми характеристиками. Это гарантирует, что искусственный интеллект останется актуальным и полезным для решения будущих проблем здравоохранения. Например, алгоритмы искусственного интеллекта для прогнозирования пандемий должны быть адаптируемыми для работы с новыми штаммами и вариациями вирусов. Эффективное управление также способно формировать и укреплять доверие общества к государственным учреждениям и использованию ими искусственного интеллекта во время пандемии (Romano et al., 2021). Такое доверие способствует сотрудничеству в области инициатив искусственного интеллекта, таких как приложения для отслеживания контактов и симптомов.

В работе Correia с соавторами (2020b) исследовались принципы, которые создают прочную основу для решений в области борьбы с пандемиями. Способствуя сотрудничеству, ставя во главу угла этические ценности и обеспечивая долгосрочную устойчивость, практика управления открывает путь к тому, чтобы стать мощным оружием в арсенале борьбы с пандемиями и построить более устойчивое будущее для общественного здравоохранения. Авторы предлагают модель, включающую шесть измерений и восемь гипотез, которая уже прошла проверку в конкретных обстоятельствах. Модель представлена на рис. 1.

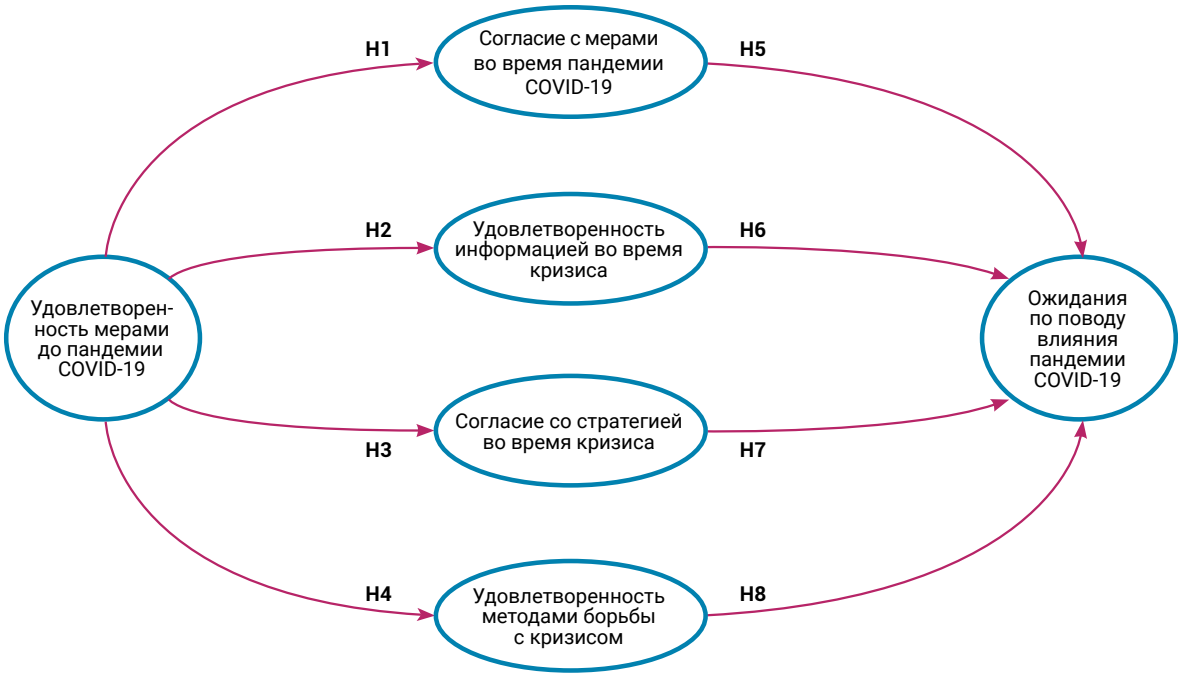



**Рис. 1. Кризисное управление: структурная модель в условиях пандемии COVID-19**
(Correia et al., 2020b)

Можно утверждать, что включение этих исходных данных (моделей, показателей и связей между показателями), человеческих факторов является тем звеном, которое необходимо для создания благоприятной среды с помощью эффективного управления, чтобы искусственный интеллект использовался ответственно и этично, одновременно повышая его эффективность в таких сферах, как прогнозирование, подготовка и предупреждение вспышек заболеваний, прогнозирование и подготовка реакции людей на меры общественного здравоохранения, оптимизация распределения стратегических ресурсов в режиме реального времени на основе показателей инфицирования, разработка целевых, персонализированных мер для лиц с высоким риском, а также отслеживание, локализация, изоляция и ограничение распространения патогенов.

Взаимосвязь между эффективным управлением, моделями кризисного управления и решениями искусственного интеллекта, а также синергетический эффект от их применения обладают огромным потенциалом для повышения готовности к будущим пандемиям и реагирования на них, что в конечном итоге позволит спасти жизни людей и обеспечить более устойчивое будущее для здравоохранения во всем мире.

## 7. Критические положения

Искусственный интеллект представляет собой одну из величайших технологических инноваций современной эпохи, способную кардинально изменить общество во многих аспектах (Bostrom, 2014). Однако эта трансформация также несет с собой серьезные вызовы и опасения по поводу хрупкости современных человеческих обществ.

Цифровое неравенство представляет собой насущную проблему, поскольку интеграция искусственного интеллекта в различных областях может увеличить социально-экономические разрывы, усиливая неравенство между людьми, имеющими доступ к технологическим достижениям и умеющими их использовать, и теми, кто не имеет таких ресурсов (Eubanks, 2018). Это явление способно усилить уже существующее неравенство и одновременно породить новые проявления цифрового отчуждения.

Появление автоматизации, основанной на искусственном интеллекте, также создает значительный риск вытеснения многих традиционных профессий, особенно тех, которые характеризуются рутинными и предсказуемыми действиями. Такой переход может привести к массовой безработице и вызвать глубокий экзистенциальный кризис, основанный на вытеснении человека технологическими инновациями. Сектор здравоохранения, видимо, не является исключением в отношении этой опасности (Hazarika, 2020).

Более того, алгоритмы искусственного интеллекта, особенно в таких важных сферах, как правосудие, здравоохранение и финансы, подвержены манипуляциям и предвзятости, что создает риск несправедливых и вредных последствий, особенно для маргинализированных и уязвимых сообществ (Obermeyer, 2019).

Опасения по поводу конфиденциальности и безопасности данных усиливаются в связи с обширной практикой сбора и анализа данных, на основе которых работают алгоритмы искусственного интеллекта. Отсутствие прозрачности и контроля использования данных представляет собой предсказуемую и значительную угрозу общественному доверию к технологиям, государственной политике и государственным институтам (Larsson & Heintz, 2020).

Кроме того, во всем мире люди все чаще сталкиваются с поляризацией и дезинформацией. Это проявляется в злоупотреблении цифровыми платформами, управляемыми искусственным интеллектом, и может подорвать социальную сплоченность и доверие к демократическим институтам, приводя к все более фрагментированному и расколотому обществу (Kavanagh & Rich, 2018).

Нельзя игнорировать и потенциальные опасности, связанные с чрезмерной зависимостью от технологий. Чем больше человечество зависит от искусственного интеллекта в принятии решений и выполнении задач, тем слабее становится способность человека функционировать самостоятельно (Bostrom, 2014). Такая уязвимость к сбоям и системным отказам в работе искусственного интеллекта может иметь разрушительные последствия.

Кроме того, искусственный интеллект все больше вторгается в нашу жизнь, или, выражаясь более техническим языком, все заметнее становится снижение автономии и самостоятельности человека (Ettlinger, 2022). Растущая интеграция искусственного интеллекта в жизнь людей может привести к разрушению человеческой автономии и самостоятельности, поскольку мы все чаще оставляем принятие важных решений на усмотрение автоматизированных систем. Это поднимает вопросы о том, кто контролирует технологии и кому можно доверить принятие решений, влияющих на нашу жизнь.

Наконец, этот первоначальный, поверхностный анализ проблем должен охватывать не только сиюминутные опасения, но и экзистенциальные риски, о которых часто говорят. Они представляют собой долгосрочные опасения по поводу развития искусственного интеллекта, включая распространенные сценарии искусственных сверхразумов, не контролируемых человеком и угрожающих выживанию человечества (Bostrom, 2014).

По сути, слияние искусственного интеллекта с хрупкой природой современных человеческих обществ заставляет глубоко задуматься над вопросами этики, управления, справедливости и человеческих принципов. Для решения этих вопросов жизненно важен комплексный подход, чтобы при разработке и внедрении искусственного интеллекта во главу угла ставились благосостояние людей и устойчивое развитие (Jobin et al., 2019).

Попытаемся, однако, чуть глубже взглянуть на осознанное стремление человечества использовать искусственный интеллект ответственным, продуктивным и безопасным образом.

При правильной нагрузке системы искусственного интеллекта оказываются весьма хрупкими и уязвимыми. Это может проявляться по-разному, в зависимости от контекста и характера систем искусственного интеллекта. Например, возможны враждебные атаки, связанные с намеренным манипулированием входными данными этих систем таким образом, чтобы заставить их совершать ошибки или выдавать неверные результаты. Такие атаки могут использовать уязвимости алгоритмов искусственного интеллекта, например, нейронных сетей глубокого обучения, приводя к неожиданным и потенциально опасным действиям (Ruan et al., 2021). Еще один подобный пример, рассматриваемый с другой точки зрения, – это использование нерепрезентативных данных для обучения моделей искусственного интеллекта, что также приводит к неверным решениям или прогнозам. При определенных условиях может усиливаться предвзятость, например, когда система искусственного интеллекта сталкивается с новыми данными, которые значительно отличаются от тех,

на которых ее обучали. В результате ИИ может оказаться неспособным к эффективному обобщению, что приведет к нестабильности его работы (Navigli et al., 2023). Другой пример – так называемое катастрофическое забывание: некоторые системы искусственного интеллекта, в частности основанные на искусственных нейронных сетях, могут демонстрировать катастрофическое забывание при получении новых данных. Это происходит, когда система искусственного интеллекта забывает ранее изученную информацию по мере поступления новой, что приводит к снижению производительности или точности с течением времени. Такая уязвимость ограничивает способность системы адаптироваться к изменяющимся условиям или задачам (Kirkpatrick et al., 2017). Еще один пример – хрупкость модели, т. е. неустойчивость системы искусственного интеллекта к небольшим изменениям входных данных или параметров. Например, небольшие возмущения входных изображений не позволяют системе распознавания образов правильно классифицировать объекты. Это может привести к опасным последствиям в таких областях, как автономные транспортные средства или медицинская диагностика (Chen et al., 2020). Последний пример уязвимости искусственного интеллекта – это присущее ему свойство сложности системы. По мере того как системы искусственного интеллекта становятся более сложными и взаимосвязанными, они все более подвержены сбоям или отказам отдельных компонентов. Сбой в одной части системы каскадно отражается на других частях, что приводит к общесистемным отказам или поломкам (Chen et al., 2020). Сложность – это обоюдоострый меч, поскольку мощь многих систем обусловлена их сложностью и способностью обрабатывать огромные объемы данных. Однако эта сложность является и их слабостью, поскольку она увеличивает возможности для потенциальных атак и делает систему более недоступной для понимания и защиты. Эта уязвимость подчеркивает первостепенное значение надежности и устойчивости при разработке и развертывании систем искусственного интеллекта. Решение проблемы уязвимости систем искусственного интеллекта требует пристального внимания к процессам разработки, тестирования и валидации, а также постоянного мониторинга и обслуживания таких систем. Это также подчеркивает необходимость прозрачности, подотчетности и соблюдения этических принципов при разработке и внедрении технологий искусственного интеллекта. Решив эти проблемы, отдельные личности, организации и сообщества смогут работать над созданием систем искусственного интеллекта, которые будут более устойчивыми, надежными и заслуживающими доверия в широком спектре приложений. Если же подобные проблемы, возникающие по отдельности или совокупно, не будут решены, это приведет к катастрофическим последствиям в системе здравоохранения в случае эпидемического или пандемического кризиса.

Еще одно противоречие состоит в том, что развитие искусственного интеллекта, даже если оно кажется несомненным, часто нарушается наличием «узких мест» – критических точек, где сбой или нарушение работы может иметь каскадные последствия для всей системы. Эта двойственность, когда сосуществуют прочность и хрупкость, присуща многим сложным системам искусственного интеллекта (Zhou et al., 2024). Например, как мы уже видели, модели искусственного интеллекта часто включают в себя сложные комплексы взаимосвязанных компонентов, таких как слои в глубоких нейронных сетях или узлы в моделях на основе графов. Такая взаимосвязанность повышает надежность системы, обеспечивая избыточность и отказоустойчивость, но она также создает «узкие места», когда отказ или нарушение работы критического

компонента может распространиться по всей сети (Villegas-Ch et al., 2024). Другой пример – возникновение критических зависимостей. В таком качестве могут выступать определенные компоненты приложений искусственного интеллекта, от которых зависит функциональность всей системы. К таким «критическим точкам» относятся определенные слои или узлы нейронных сетей, играющие ключевую роль в обработке информации или принятии решений. Если эти компоненты выйдут из строя или дадут сбой, будет нарушена работа всей системы (Macrae, 2022). Еще одна тенденция – чувствительность систем искусственного интеллекта к входным данным, особенно в таких областях, как распознавание изображений или обработка естественного языка. Небольшие возмущения или ошибочные входные данные в «критической точке» могут привести к значительным изменениям в результатах работы системы. Такая чувствительность подчеркивает уязвимость в отношении определенных типов действий с входными данными (Dhingra & Gupta, 2017). Последний пример, иллюстрирующий опасность «критических точек», – это особенности дизайна, построенного на компромиссе между надежностью и эффективностью. Стратегии, направленные на повышение надежности, такие как добавление избыточности или механизмов исправления ошибок, создают дополнительные «критические точки» или увеличивают стоимость вычислений. И наоборот, оптимизация для повышения эффективности может непреднамеренно увеличить уязвимость системы за счет уменьшения избыточности или отказоустойчивости. Решение этих специфических проблем требует многогранного подхода, включая выявление и смягчение проблемных моментов и повышение устойчивости за счет избыточности и разнообразия (Goodfellow et al., 2016). Понимая этот тонкий баланс, ученые, инженеры и практики могут работать над созданием более устойчивых и надежных технологий искусственного интеллекта.

Сложность систем искусственного интеллекта еще более повышается, когда мы расширяем и детализируем понятие «потеря контроля» – иными словами, когда это понятие действительно начинает отражать все сложности и проблемы, связанные с разработкой и внедрением технологий искусственного интеллекта. Одно из таких проявлений связано с автономностью и принятием решений (Wallach & Allen, 2008). По мере того как системы искусственного интеллекта становятся все более автономными и способными принимать решения без непосредственного вмешательства человека, возникает опасность утраты контроля над результатами этих решений. Это особенно актуально для высокорисковых областей, таких как автономные транспортные средства или медицинские приборы, где действия систем искусственного интеллекта могут иметь угрожающие жизни последствия в реальном мире. Еще одно проявление указанного свойства – это непрозрачность, присущая системам искусственного интеллекта, особенно основанных на глубоком обучении и нейронных сетях, что затрудняет понимание и интерпретацию их внутренней работы человеком. Отсутствие прозрачности приводит к потере контроля над тем, как системы ИИ принимают свои решения, что вызывает опасения по поводу подотчетности и доверия (Chiao, 2019). Еще одно проявление заключается в том, что системы искусственного интеллекта могут демонстрировать эмерджентное поведение, когда сложные модели или модели поведения возникают в результате взаимодействия простых компонентов. Такое эмерджентное поведение трудно предсказать или контролировать, что приводит к неопределенности в отношении поведения систем в новых или непредвиденных ситуациях. Это само по себе может привести к непредсказуемым

последствиям, поскольку действия или решения функций искусственного интеллекта приводят к результатам, которые не были предусмотрены или запланированы их создателями. Причина может лежать в неожиданных взаимодействиях с окружающей средой или других факторах (Bostrom, 2014). И последнее, но не менее важное: необходимо учитывать этические и социальные последствия. Потеря контроля над технологиями искусственного интеллекта может иметь и более широкие этические и общественные последствия, такие как влияние искусственного интеллекта на занятость, частную жизнь, безопасность и неравенство, о чем мы будем говорить далее (Thomsen, 2019). Эти проблемы подчеркивают необходимость ответственного развития искусственного интеллекта и управления им. Важно обеспечить внедрение технологий таким образом, чтобы они приносили пользу всему обществу, а не только олигархии, крупным технологическим компаниям и правящей элите. Решение этих проблем требует целостного подхода, включающего технические, этические и нормативные аспекты. Сюда относятся и обеспечение прозрачности и подотчетности систем искусственного интеллекта, а также постоянный диалог и сотрудничество между заинтересованными сторонами для снижения рисков и максимизации выгоды.

Еще одна сложность возникает в аспекте предпосылки надежности, особенно в периоды напряженности или неопределенности, когда она требует срочной переоценки. Одним из примеров неверной предпосылки является взаимосвязанность. Системы искусственного интеллекта представляют собой сложный набор взаимосвязанных компонентов, каждый из которых вносит свой вклад в общую функциональность. При возникновении напряженности или неожиданных условий, таких как атаки противника, аномалии данных или изменения окружающей среды, сложность этих систем увеличивает вероятность сбоя. Отсюда необходимость более тонкого понимания надежности, выходящей за рамки традиционных показателей (Macrae, 2022). Еще один пример – непредсказуемость. Эмерджентное поведение систем ИИ приводит к непредсказуемости их реакции на стрессовые факторы. Даже незначительные возмущения или вариации входных данных могут привести к неожиданным результатам, что подчеркивает сложность обеспечения надежности в различных условиях. Это говорит о важности тестирования на устойчивость и планирования различных сценариев для выявления потенциальных точек отказа и смягчения возможных последствий (Bostrom, 2014). Именно эти опасения были показаны выше. Еще один пример этого класса явлений связан с адаптивными и развивающимися средами, поскольку системы искусственного интеллекта работают в динамичной и постоянно меняющейся обстановке, где условия могут меняться быстро и непредсказуемо. В такой ситуации понятие надежности как статичного свойства становится неадекватным. Вместо этого надежность следует рассматривать как динамическое свойство, которое адаптируется к изменяющимся обстоятельствам, требуя постоянного мониторинга, адаптации и механизмов обратной связи (Sundar, 2020). Последний пример посвящен тесной взаимосвязи между этими типами систем и человеко-машинным взаимодействием. Люди-операторы играют важнейшую роль в мониторинге производительности системы, интерпретации результатов и вмешательстве в случае необходимости. Однако в условиях стресса или высокого напряжения операторы-люди могут быть склонны к ошибкам или когнитивным предубеждениям, что влияет на надежность систем (Hoff & Bashir, 2015). В свете этих проблем переосмысление предпосылок надежности в системах ИИ требует более адаптивных,

устойчивых и учитывающих контекст подходов. Среди них – соблюдение принципов количественной оценки неопределенности, надежности решений и человекоориентированного проектирования при разработке и внедрении моделей и приложений искусственного интеллекта. Приняв на вооружение более широкое понимание надежности и заблаговременно устранив факторы, способствующие сбоям, человечество сможет создать более надежные и безотказные технологии искусственного интеллекта. Первостепенной задачей должна стать подготовка плана действий в непредвиденных обстоятельствах без участия искусственного интеллекта, чтобы общество могло продолжать функционировать в случае технологического кризиса или коллапса. Если правосудие «дематериализуется», сохранится ли система правосудия в большинстве развитых стран, если не во всех, если завтра Интернет выйдет из строя? Хотим ли мы идти на риск, который может привести к остановке или полному краху общественных отношений?

Еще более глубокая проблема, еще одно препятствие, которое необходимо преодолеть, – это переход от предпосылки надежности к предпосылке риска в отношении искусственного интеллекта. Это равносильно переходу от неподвижности к быстрому реагированию на меняющиеся обстоятельства. Один из примеров – понимание риска. Надежность часто ассоциируется с понятием детерминированности результатов и предсказуемости поведения. Однако в сложных и динамичных условиях, с которыми сталкиваются системы искусственного интеллекта, полная надежность может оказаться недостижимой. Если признать это, то акцент смещается на понимание и управление рисками; отсюда аспекты вероятности и влияния неблагоприятных событий или неопределенностей (Bigham et al., 2019). Принятие риска подразумевает признание неопределенности, присущей системам искусственного интеллекта, и использование адаптивных стратегий для ее преодоления. Вместо того чтобы стремиться к абсолютной надежности, системы искусственного интеллекта должны быть спроектированы таким образом, чтобы быть устойчивыми и быстро реагировать на изменяющиеся обстоятельства. Это подразумевает механизмы мониторинга в реальном времени, динамической корректировки и обучения на опыте (Syed et al., 2023). Такой способ реагирования на изменяющиеся обстоятельства требует от этих приложений быстроты и гибкости, включая способность быстро оценивать риски, выявлять возможности и соответствующим образом адаптировать поведение или стратегии принятия решений. Гибкие системы искусственного интеллекта будут обладать способностью все более динамично распределять ресурсы, расставлять приоритеты и адаптироваться к новой информации или целям по мере их появления. Переход от установки на надежность к установке на риск требует разработки надежных принципов управления рисками. Эти принципы должны обеспечивать систематический подход к выявлению, оценке, снижению и мониторингу рисков на протяжении всего жизненного цикла использования искусственного интеллекта. Проактивно управляя рисками, организации могут повысить устойчивость и снизить вероятность неблагоприятных исходов (Jobin et al., 2019). Несомненно, что использование искусственного интеллекта с учетом рисков должно ориентироваться на непрерывное обучение и совершенствование, используя петли обратной связи, эксперименты и данные для итеративного повышения эффективности и адаптации к меняющимся проблемам. Такой итеративный подход позволит системам искусственного интеллекта со временем совершенствовать свои стратегии и становиться более эффективными в управлении рисками. Последний фрагмент

этой головоломки появится в результате признания ограниченности возможностей систем ИИ в сложных и неопределенных ситуациях и, как следствие, усиления акцента на подходе «человек в контуре» (Russell & Norvig, 2021). Благодаря интеграции человеческих суждений, опыта и надзора системы искусственного интеллекта смогут более адекватно дополнять процесс принятия решений человеком, снижать риски и повышать общую эффективность системы (Parasuraman & Riley, 1997). В целом переход от парадигмы надежности к парадигме риска отражает более широкое признание неопределенности и сложности, присущих реальным приложениям. Принимая риски и развивая гибкость в реагировании на изменяющиеся обстоятельства, модели искусственного интеллекта могут лучше ориентироваться в ситуации неопределенности, адаптироваться к изменяющимся задачам и в конечном итоге демонстрировать большую ценность и влияние в различных областях, к которым не в последнюю очередь относится здравоохранение.

Следующий уровень проблемы состоит в том, что искусственный интеллект может маскировать свои слабые стороны под сильные, особенно когда речь идет об определенных типах моделей или алгоритмов машинного обучения. Самым вопиющим примером является дисбаланс между обобщением и чрезмерной подгонкой, когда модель учится хорошо работать на обучающих данных, но не может обобщить их на новые, неизвестные данные (Igual & Seguí, 2024). Это создает иллюзию надежности, поскольку модель кажется исключительно хорошо работающей на данных, на которых она была обучена. Однако при изучении новых данных слабые стороны модели становятся очевидными, поскольку она не может делать точные предсказания. Так, система визуального распознавания легко учится определять галстуки на фотографиях и ассоциировать их с мужчинами, если ее обучить на наборе данных о высокопоставленных лицах с Уолл-стрит. Это происходит потому, что модели искусственного интеллекта нацелены обобщать шаблоны из обучающих данных и делать предсказания по неизвестным данным. Способность к обобщению необходима, однако чрезмерная зависимость от конкретных паттернов в обучающих данных может привести к подгонке, когда модель не может эффективно обобщать. Понимание баланса между обобщением и чрезмерной подгонкой имеет решающее значение для обеспечения надежности. Тщательное тестирование систем искусственного интеллекта на различных наборах данных, внимательный анализ процессов принятия решений и устранение предвзятости и уязвимостей позволяют пользователям выявлять и устранять слабые стороны, замаскированные под сильные, что приводит к созданию более надежных и заслуживающих доверия систем.

По мере углубления анализа приходит понимание того, что искусственный интеллект культивирует нестабильность. Коллектив, привыкший к тому, что все работает и будет работать без перебоев, а блокировки либо случайны, либо незначительны по своему воздействию, несомненно, гораздо менее подготовлен к тому, что эти принципы окажутся под угрозой. Зависимость от искусственного интеллекта все больше внедряется в различные аспекты жизни общества, и растет зависимость от его функциональности. Отдельные лица, организации, правительства и наднациональные институты (такие как Организация Объединенных Наций и Всемирная организация здравоохранения) все больше полагаются на искусственный интеллект в принятии решений, автоматизации и оптимизации процессов (Bostrom, 2014). Однако такая зависимость может привести к нестабильности, если эти функции будут давать сбои или нарушаться. Повсеместное внедрение решений на основе

искусственного интеллекта формирует коллективные ожидания непрерывности и бесперебойности работы. Когда эти решения функционируют так, как ожидается, они укрепляют представление о том, что риски минимальны. Однако это может привести к самоуспокоенности и уязвимости, если системы столкнутся с неожиданными проблемами или сбоями (Parasuraman et al., 2000). Из этого следует, что, когда приложения искусственного интеллекта выходят из строя или блокируются, последствия будут значительными, особенно если на них полагаются при выполнении критически важных задач или услуг. Перебои в технологических процессах могут разрушить цепочки поставок, финансовые рынки, коммуникационные сети и другие важнейшие функции и системы (например, атомные электростанции), что приведет к экономическим потерям, социальным беспорядкам и даже угрозе безопасности. Для повышения устойчивости и адаптивности перед лицом этой потенциальной нестабильности необходимы проактивные меры по прогнозированию и снижению рисков, о чем уже говорилось выше. Это может включать диверсификацию технологических зависимостей, создание избыточности в критически важных системах и разработку человекоориентированных подходов к принятию решений и решению проблем (Bigham et al., 2019; Jobin et al., 2019). Таким образом, технологии искусственного интеллекта также создают проблемы, связанные со стабильностью и устойчивостью. Признавая потенциальную нестабильность, присущую этим системам, и принимая упреждающие меры по устранению рисков, заинтересованные стороны могут лучше ориентироваться в сложностях мира, управляемого искусственным интеллектом, и создавать более надежные и устойчивые системы.

Наконец, добравшись до самой глубины – до центра Вселенной и ее самого холодного места, по Аристотелю, или центра Земли и Ада, по Данте, – мы должны рассмотреть искусственный интеллект в свете того, что можно назвать люциферианской семиотикой, путешествием в символические или метафорические последствия искусственного интеллекта. В различных мифологиях и системах верований Люцифер (люциферианская символика) часто ассоциируется с темами бунтарства, просвещения и стремления к знаниям. Люцифер часто изображается как носитель света (Hanegraaff, 2013). Таким образом, термин «люциферианец» может означать стремление к знаниям или силе, которое бросает вызов устоявшимся нормам или структурам власти. Семиотика относится к изучению знаков и символов и их интерпретации. В контексте искусственного интеллекта семиотика охватывает символические значения, связанные с искусственным интеллектом, включая понятия разума, автономии и контроля (Binder, 2024). Искусственный интеллект часто воспринимается как символ силы и возможностей, учитывая его способность обрабатывать огромные объемы данных, принимать сложные решения и эффективно автоматизировать задачи. Такое представление о могуществе подкрепляется впечатляющими деяниями ИИ в различных областях. Однако, как было показано выше, объекты или сущности, которые кажутся сильными, на самом деле могут обладать уязвимостями или слабостями, которые не сразу бросаются в глаза. Такое изменение ожиданий на противоположное можно рассматривать как проявление люциферианской символики, когда стремление к знаниям или власти приводит к переоценке устоявшихся истин или предположений. Модели и системы искусственного интеллекта, несмотря на их кажущуюся мощь, в определенных контекстах демонстрируют уязвимость или ограниченность. Эти слабости могут стать более заметными с течением времени, когда технологии искусственного интеллекта будут

подвергаться тщательному изучению, экспериментам и внедрению в реальный мир. Исследование люциферианской семиотики поднимает более широкие этические и философские вопросы (Bostrom, 2014) о природе власти, знания и контроля в эпоху искусственного интеллекта. Это побуждает задуматься о непредвиденных последствиях технологического прогресса и необходимости ответственного отношения к любым технологиям. Таким образом, люциферианская семиотика, примененная к искусственному интеллекту, предлагает нам рассмотреть символические значения и последствия искусственного интеллекта, включая то, как восприятие силы и власти может быть подорвано или поставлено под сомнение более глубоким изучением и пониманием. Это подчеркивает важность критического исследования и этических размышлений при изучении сложностей искусственного интеллекта и его влияния на общество, государственную политику и политические системы.

Каковы же перспективы использования искусственного интеллекта в борьбе с пандемиями с учетом вышесказанного? Возможно, лучший вариант действий для человечества – продолжать полагаться на человеческий фактор. Люди совершают больше ошибок, в этом нет сомнений. Но большинство этих ошибок мелкие и незначительные. Они могут привести к отдельным трагедиям, но не к глобальным. Модели искусственного интеллекта для борьбы со вспышками заболеваний, эпидемиями и пандемиями, а также другие медицинские приложения, основанные на искусственном интеллекте, могут быть практически безошибочными. Но одна ошибка может погубить всех.

## 8. Искусственный интеллект и фундаментальные права

Влияние ИИ на основные права настолько актуально, что не осталось незамеченным законодателем. Статья 27 Закона об ИИ требует, чтобы системы с высоким уровнем риска проходили оценку в отношении их влияния на основные права. Основная цель такой оценки – выявить и смягчить потенциальные угрозы, которые эти системы могут представлять для основных прав человека. Это особенно важно, когда речь идет о системах ИИ с высоким уровнем риска, которые способны существенно повлиять на жизнь и благосостояние людей. В целом можно сказать, что, осознавая негативные последствия ИИ, человечество решило не отказываться от него, а, напротив, воспользоваться его положительными свойствами, классифицировав системы ИИ по степени риска и осуществляя их предварительный, сопутствующий и последующий контроль.

Существует несколько направлений взаимосвязей между ИИ и основными правами. В частности, это влияние, которое ИИ может оказать на осуществление и, напротив, на нарушение основных прав. В любом случае есть основания полагать, что, как отмечает Агентство Европейского союза по основным правам[10], даже в ограниченном контексте отсутствие большого объема эмпирических данных по широкому спектру прав, связанных с ИИ, затрудняет задачу обеспечения необходимых гарантий для того, чтобы использование ИИ эффективно соответствовало основным правам.

[10] FRA – European Union Agency for Fundamental Rights, Getting the future right – Artificial intelligence and fundamental rights – Report, Publications Office of the European Union, 2020.

Главный аргумент в пользу использования ИИ – его эффективность (Pedro, 2023). Что касается проблем, то основное беспокойство вызывает нарушение фундаментальных прав. Так, среди основных прав, которым потенциально может навредить ИИ, выделяют право на защиту персональных данных (Gómez Abeja, 2022) и право на недискриминацию (Gómez Abeja, 2022), право на эффективную судебную защиту (Shaelou & Razmetaeva, 2023), право на свободу информации, избирательное право и право на доступ к публичной информации (Gómez Abeja, 2022).

Возвращаясь к работе Агентства Европейского союза по основным правам[11], следует отметить, что использование ИИ может оказывать влияние на основные права, вызывая необходимость гарантировать недискриминационное использование ИИ (право на недискриминацию); требование законной обработки данных (право на защиту персональных данных); возможность подачи жалоб на решения, основанные на ИИ, и подачи апелляций (право на эффективные средства правовой защиты и беспристрастный суд).

Наконец, следует также подчеркнуть, что связь между ИИ и основными правами может быть более тесной, по крайней мере, в следующих аспектах: нарушение неявных основных прав (Gómez Colomer, 2023), таких как принцип верховенства закона и право на рассмотрение дела судьей-человеком, а также появление «новых» или «обновленных» основных прав (Shaelou & Razmetaeva, 2023), таких как право на забвение (Gómez Abeja, 2022), «право не быть объектом автоматических решений и автоматических действий» в широком смысле (Shaelou & Razmetaeva, 2023); «право влиять на свой цифровой след» (Shaelou & Razmetaeva, 2023), а также новые права, такие как «право не быть объектом манипуляции», «право на нейтральное информирование в Сети» и «право на значимый человеческий контакт», «право не быть объектом измерения, анализа или обучения» (Shaelou & Razmetaeva, 2023).

## 9. Пандемии и фундаментальные права

Ситуация пандемии, как в случае с COVID-19, потребовала введения публично-правовых режимов исключительности (наряду с режимами нормальности). Это не является чем-то новым (Gomes & Pedro, 2020) – вспомним латинское высказывание «У необходимости нет закона, но она сама устанавливает его для себя». Именно так обосновывались чрезвычайные полномочия в римском праве, которые могли быть использованы в случаях, когда необходимо было справиться с непредвиденной ситуацией, требующей немедленного решения, без возможности отсрочки.

Потребность в правовом режиме исключительности и, соответственно, его мобилизация стали более очевидными в последнее время. Этому в значительной степени способствуют высокорисковая конфигурация современного общества (Beck, 1986) и тот факт, что мы живем в экономически и социально глобализированном мире. Несмотря на физические расстояния, здесь все оказывается близким – так, продолжающийся кризис здравоохранения был вызван вспышкой COVID-19, которая за несколько месяцев распространилась из своего источника (китайский город Ухань) на весь мир (Pedro, 2022).

[11] Там же.

Перед лицом катастроф такого рода публичное право не могло и не может оставаться в стороне. Иными словами, учитывая пагубные последствия, которые общественные бедствия оказывают на salus populi – здоровье народа, становится очевидно, что государство должно использовать все имеющиеся в его распоряжении средства для восстановления нормальной жизни (Alvarez Garcia, 1996). Поэтому, чтобы гарантировать верховенство закона, необходимо предусмотреть режимы, достаточно гибкие для соответствия публичным интересам, находящимся под угрозой, режимы, позволяющие реагировать на общественную необходимость, или, другими словами, публично-правовые режимы исключительности.

В рамках реальной нормативности публичное право руководствуется принципом законности публичных действий, что соответствует условиям правовой нормативности. Проблема возникает, когда реальность временно меняется радикальным образом, создавая ситуации неминуемой или реальной опасности для общества. На такие ситуации нормативное публичное право не может дать адекватного ответа, и идея поддержания демократического верховенства права навязывает необходимость введения в действие исключительных правовых режимов – jus extremae necessitatis, чтобы в кратчайшие сроки восстановить нормальность и вернуть в действие нормативные правовые режимы. Речь идет об альтернативной законности, исключительной законности для исключительной ситуации (Correia, 1987) – о замещающей и временной законности.

Таким образом, как правило, в исключительных ситуациях, в условиях чрезвычайного положения, при соблюдении принципа пропорциональности, действие некоторых основных прав может быть приостановлено. Несмотря на это, следует отметить, что действие не всех основных прав может быть приостановлено, например, право на жизнь, личную неприкосновенность, личную идентичность, гражданскую правоспособность и гражданство, отсутствие обратной силы уголовного закона, право на защиту обвиняемых, свобода совести и религии.

## 10. Возможные взаимосвязи между искусственным интеллектом и фундаментальными правами в борьбе с пандемиями

В демократических правовых государствах рассмотрение вопроса об использовании ИИ для борьбы с пандемиями обычно связано с соблюдением основных прав. Это требует, с одной стороны, рассмотрения воздействия использования ИИ на определенные основные права с учетом рисков, которые несет в себе каждая конкретная система ИИ, а с другой – того, что контекст пандемии, как это произошло с COVID-19, требует признать законность исключений, когда определенные основные права должны быть ограничены с целью защиты таких ценностей, как общественное здоровье, до восстановления нормальной ситуации.

## Заключение

В современном мире борьба с болезнями и пандемиями требует комплексного подхода, объединяющего усилия врачей, медицинских учреждений и различных государств. ИИ представляет собой перспективный инструмент, способный кардинально изменить методы противодействия эпидемиям и пандемиям. Его потенциал заключается в анализе больших данных, прогнозировании вспышек заболеваний,

ускорении разработки лекарств, персонализации лечения и оптимизации распределения ресурсов. Примеры использования ИИ, такие как раннее выявление вспышек через анализ данных из социальных сетей, ускорение поиска лекарственных препаратов и улучшение отслеживания контактов, демонстрируют его значимость в борьбе с глобальными угрозами здоровью.

Однако внедрение ИИ в сферу здравоохранения сопряжено с рядом вызовов. К ним относятся проблемы предвзятости данных, сложность алгоритмов, риски чрезмерной зависимости от технологий и этические дилеммы, связанные с соблюдением фундаментальных прав. Использование ИИ для борьбы с пандемиями требует тщательного баланса между инновациями, этикой и защитой прав человека, включая право на приватность, свободу и равный доступ к медицинской помощи.

Таким образом, ИИ, несмотря на свои революционные возможности, не является панацеей. Его применение должно сопровождаться критическим анализом потенциальных рисков и разработкой правовых и этических механизмов, которые обеспечат безопасное и справедливое использование технологий. Только при условии учета этих аспектов ИИ сможет стать эффективным инструментом в борьбе с болезнями, не ставя под угрозу фундаментальные права и свободы человека.

## Список литературы


Arass, M., & Souissi, N. (2018). Data lifecycle: from big data to SmartData. In *2018 IEEE 5th International Congress on Information Science and Technology* (pp. 80–87). IEEE. https://doi.org/10.1109/CIST.2018.8596547

Alvarez Garcia, V. (1996). *El concepto de necesidad en derecho público* (1st ed.). Madrid: Civitas. (In Spanish).

Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Baclic, O., Tunis, M., Young, K., Doan, C., Swerdfeger, H., & Schonfeld, J. (2020). Challenges and opportunities for public health made possible by advances in natural language processing. *Canada Communicable Disease Report*, *46*(6), 161–168. https://doi.org/10.14745/ccdr.v46i06a02

Bajwa, J., Munir, U., Nori, A., & Williams, B. (2021). Artificial intelligence in healthcare: transforming the practice of medicine. *Future Healthcare Journal*, *8*(2), e188-e194. https://doi.org/10.7861/fhj.2021-0095

Beck, U. (1986). *Risikogesellschaft: Auf dem Weg in eine andere Moderne*. Frankfurt am Main: Suhrkamp Verlag.

Balog-Way, D., & McComas, K. (2022). COVID-19: Reflections on trust, tradeoffs, and preparedness. In *COVID-19* (pp. 6–16). Routledge.

Bazarkina, D. Y., & Pashentsev, E. N. (2020). Malicious use of artificial intelligence. *Russia in Global Affairs*, *18*(4), 154–177. https://doi.org/10.31278/1810-6374-2020-18-4-154-177

Benke, K., & Benke, G. (2018). Artificial Intelligence and Big Data in Public Health. *International Journal of Environmental Research and Public Health*, *15*(12), 2796. https://doi.org/10.3390/ijerph15122796

Berk, R. A. (1983). An introduction to sample selection bias in sociological data. *American Sociological Review*, *48*(3), 386–398. https://doi.org/10.2307/2095230

Bigham, G., Adamtey, S., Onsarigo, L., & Jha, N. (2019). Artificial Intelligence for Construction Safety: Mitigation of the Risk of Fall. In K. Arai, S. Kapoor, R. Bhatia (Eds.). *Intelligent Systems and Applications*. Springer. https://doi.org/10.1007/978-3-030-01057-7_76

Binder, W. (2024). Technology as (dis-)enchantment. AlphaGo and the meaning-making of artificial intelligence. *Cultural Sociology*, *18*(1), 24–47. https://doi.org/10.1177/17499755221138720

Bisconti, P., Orsitto, D., Fedorczyk, F., Brau, F., Capasso, M., De Marinis, L., ... & Schettini, C. (2023). Maximizing team synergy in AI-related interdisciplinary groups: an interdisciplinary-by-design iterative methodology. *AI & Society*, *38*(4), 1443–1452. https://doi.org/10.1007/s00146-022-01518-8

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Box, G. (1979). Robustness in the strategy of scientific model building. In R. Launer & G. Wilkinson (Eds.), *Robustness in Statistics* (pp. 201–236). Academic Press. https://doi.org/10.1016/B978-0-12-438150-6.50018-2

Breiman, L. (2001). Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3), 199–231. https://doi.org/10.1214/ss/1009213726

Bulled, N. (2023). "Solidarity:" A failed call to action during the COVID-19 pandemic. *Public Health in Practice*, *5*, 100379. https://doi.org/10.1016/j.puhip.2023.100379

Chen, A. (2016). A review of emerging non-volatile memory (NVM) technologies and applications. *Solid-State Electronics*, *125*, 25–38. https://doi.org/10.1016/j.sse.2016.07.006

Chen, J., Zhang, R., Han, W., Jiang, W., Hu, J., Lu, X., Liu, X., & Zhao, P. (2020). Path Planning for Autonomous Vehicle Based on a Two-Layered Planning Model in Complex Environment. *Journal of Advanced Transportation*, 2020, 6649867. https://doi.org/10.1155/2020/6649867

Chiao, V. (2019). Fairness, accountability and transparency: notes on algorithmic decision-making in criminal justice. *International Journal of Law in Context*, *15*(2), 126–139. https://doi.org/10.1017/S1744552319000077

Correia, P., Mendes, I., Pereira, S., & Subtil, I. (2020a). The combat against COVID-19 in Portugal: How state measures and data availability reinforce some organizational values and contribute to the sustainability of the National Health System. *Sustainability*, *12*(18), 7513. https://doi.org/10.3390/su12187513

Correia, P., Mendes, I., Pereira, S., & Subtil, I. (2020b). The combat against COVID-19 in Portugal, Part II: how governance reinforces some organizational values and contributes to the sustainability of crisis management. *Sustainability*, *12*(20), 8715. https://doi.org/10.3390/su12208715

Correia, P., Pereira, S., Mendes, I., & Subtil, I. (2022). COVID-19 Crisis management and the Portuguese regional governance: Citizens perceptions as evidence. *European Journal of Applied Business Management*, *8*(1), 1–12.

Correia, P., Pereira, S., Mendes, I., & Subtil, I. (2021). COVID-19 Crisis management and the Portuguese regional governance: Citizens perceptions as evidence. In *European Consortium for Political Research General Conference* (pp. 1–18). United Kingdom.

Correia, J. M. C. (1987). *Legalidade e autonomia contratual nos contratos administrativos* (pp. 283, 768). Lisboa: Almedina.

DeCamp, M., & Tilburt, J. (2019). Why we cannot trust artificial intelligence in medicine. *The Lancet Digital health*, *1*(8), e390. https://doi.org/10.1016/S2589-7500(19)30197-9

Dhingra, M., & Gupta, N. (2017). Comparative analysis of fault tolerance models and their challenges in cloud computing. *International Journal of Engineering & Technology*, *6*(2), 36–40. https://doi.org/10.14419/ijet.v6i2.7565

Ettlinger, N. (2022). *Algorithms and the Assault on Critical Thought: Digitalized Dilemmas of Automated Governance and Communitarian Practice* (1st ed.). Routledge. https://doi.org/10.4324/9781003109792

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. New York: Picador, St Martin's Press.

Ferguson, N., Cummings, D., Fraser, C., Cajka, J., Cooley, P., & Burke, D. (2006). Strategies for mitigating an influenza pandemic. *Nature*, *442*(7101), 448–452. https://doi.org/10.1038/nature04795

Fetzer, T., & Graeber, T. (2021). Measuring the scientific effectiveness of contact tracing: Evidence from a natural experiment. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(33), e2100814118. https://doi.org/10.1073/pnas.2100814118

Galetsi, P., Katsaliaki, K., & Kumar, S. (2022). The medical and societal impact of big data analytics and artificial intelligence applications in combating pandemics: A review focused on Covid-19. *Social Science & Medicine*, *301*, 114973. https://doi.org/10.1016/j.socscimed.2022.114973

Gianfrancesco, M., Tamang, S., Yazdany, J., & Schmajuk, G. (2018). Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data. *JAMA Internal Medicine*, *178*(11), 1544–1547. https://doi.org/10.1001/jamainternmed.2018.3763

Goldman, N., Bertone, P., Chen, S., Dessimoz, C., LeProust, E. M., Sipos, B., & Birney, E. (2013). Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature*, *494*(7435), 77–80. https://doi.org/10.1038/nature11875

Gomes, C. A., & Pedro, R. (Coords.). (2020). *Direito administrativo de necessidade e de excepção*. Lisboa: AAFDL.

Gómez Abeja, L. (2022). Inteligencia artificial y derechos fundamentales. In F. H. Llano Alonso (Dir.), J. Garrido Martín & R. Valdivia Jiménez (Coords.), *Inteligencia artificial y filosofía del derecho* (1.ª ed., pp. 91–114, 93). Murcia: Ediciones Laborum. (In Spanish).

Gómez Colomer, J.-L. (2023). *El juez-robot: La independencia judicial en peligro*. Valencia: Tirant lo Blanch. (In Spanish).

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Greiner, R., Grove, A., & Kogan, A. (1997). Knowing what doesn't matter: exploiting the omission of irrelevant data. *Artificial Intelligence*, *97*(1–2), 345–380. https://doi.org/10.1016/S0004-3702(97)00048-9

Gunasekeran, D., Tseng, R., Tham, Y., & Wong, T. (2021). Applications of digital health for public health responses to COVID-19: a systematic scoping review of artificial intelligence, telehealth and related technologies. *NPJ Digital Medicine*, *4*(1), 40. https://doi.org/10.1038/s41746-021-00412-9

Gürsoy, E., & Kaya, Y. (2023). An overview of deep learning techniques for COVID-19 detection: methods, challenges, and future works. *Multimedia Systems*, *29*(3), 1603–1627. https://doi.org/10.1007/s00530-023-01083-0

Hanegraaff, W. (2013). *Western Esotericism: A Guide for the Perplexed*. Bloomsbury Publishing.

Halevy, A., Norvig, P., & Pereira, F. (2009). The unreasonable effectiveness of data. *IEEE Intelligent Systems*, 24(2), 8–12. https://doi.org/10.1109/MIS.2009.36

Hazarika, I. (2020). Artificial intelligence: opportunities and implications for the health workforce. *International Health*, *12*(4), 241–245. https://doi.org/10.1093/inthealth/ihaa007

Hoff, K., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Hulten, G. (2018). Building Intelligent Systems: A Guide to Machine Learning Engineering. Apress.

Igual, L., & Seguí, S. (2024). *Supervised learning. In Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications* (pp. 67–97). Springer International Publishing.

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular Neurology*, *2*(4), 230–243. https://doi.org/10.1136/svn-2017-000101

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*, 389–399. https://doi.org/10.1038/s42256-019-0088-2

Jones, K., Patel, N., Levy, M., Storeygard, A., Balk, D., Gittleman, J., & Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature*, *451*(7181), 990–993. https://doi.org/10.1038/nature06536

Kandlhofer, M., Weixelbraun, P., Menzinger, M., Steinbauer-Wagner, G., & Kemenesi, Á. (2023). Education and Awareness for Artificial Intelligence. In *International Conference on Informatics in Schools: Situation, Evolution, and Perspectives* (pp. 3–12). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-44900-0_1

Kavanagh, J., & Rich, M. (2018). *Truth Decay: An Initial Exploration of the Diminishing Role of Facts and Analysis in American Public Life*. RAND Corporation. https://doi.org/10.7249/RR2314

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., ... & Hadsell, R. (2017). Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, *114*(13), 3521–3526. https://doi.org/10.1073/pnas.1611835114

Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: review, synthesis, and future research directions. E*uropean Journal of Information Systems*, *31*(3), 388–409. https://doi.org/10.1080/0960085X.2021.1927212

Larsson, S., & Heintz, F. (2020). Transparency in artificial intelligence. *Internet Policy Review*, *9*(2). https://doi.org/10.14763/2020.2.1469

Lin, X., Liu, J., Hao, J., Wang, K., Zhang, Y., Li, H., ... & Tan, X. (2020). Collinear holographic data storage technologies. *Opto-Electronic Advances*, *3*(3), 190004. https://doi.org/10.29026/oea.2020.190004

Little, R. J., & Rubin, D. B. (2019). Statistical analysis with missing data. John Wiley & Sons.

Macrae, C. (2022). Learning from the failure of autonomous and intelligent systems: Accidents, safety, and sociotechnical sources of risk. *Risk Analysis*, *42*(9), 1999–2025. https://doi.org/10.1111/risa.13850

Margetts, H. (2022). Rethinking AI for good governance. *Daedalus*, *151*(2), 360–371. https://doi.org/10.1162/daed_a_01922

Matsuzaka, Y., & Yashiro, R. (2022). Applications of Deep Learning for Drug Discovery Systems with BigData. *BioMedInformatics*, *2*(4), 603–624. https://doi.org/10.3390/biomedinformatics2040039

Mittelstadt, B., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2). https://doi.org/10.1177/2053951716679679

Morse, S., Mazet, J., Woolhouse, M., Parrish, C., Carroll, D., Karesh, W., Zambrana-Torrelio, C., Lipkin, W., & Daszak, P. (2012). Prediction and prevention of the next pandemic zoonosis. *Lancet*, *380*(9857), 1956–1965. https://doi.org/10.1016/S0140-6736(12)61684-5

Mumuni, A., & Mumuni, F. (2022). Data augmentation: A comprehensive survey of modern approaches. *Array*, *16*, 100258. https://doi.org/10.1016/j.array.2022.10025

Navigli, R., Conia, S., & Ross, B. (2023). Biases in Large Language Models: Origins, Inventory, and Discussion. *Journal of Data and Information Quality*, *15*(2), 10. https://doi.org/10.1145/3597307

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

O'Reilly-Shah, V., Gentry, K., van Cleve, W., Kendale, S., Jabaley, C., & Long, D. (2020). The COVID-19 pandemic highlights shortcomings in US health care informatics infrastructure: a call to action. *Anesthesia & Analgesia*, *131*(2), 340–344. https://doi.org/10.1213/ANE.0000000000004945

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors*, *39*(2), 230–253. https://doi.org/10.1518/001872097778543886

Parasuraman, R., Sheridan, T., & Wickens, C. (2000). A model for types and levels of human interaction with automation. *Systems and Humans*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

Pedro, R. (2022). Traços gerais da indemnização civil extracontratual pública em contextos de excecionalidade. In *Impactos da pandemia da Covid-19 nas estruturas do direito público* (pp. 379–413). Coimbra: Almedina. (In Portuguese).

Pedro, R. (2023). Inteligência artificial e arbitragem de direito público: Primeiras reflexões. In R. Pedro, & P. Caliendo (Coords.), *Inteligência artificial no contexto do direito público: Portugal e Brasil* (1.ª ed., pp. 105–127). Coimbra: Almedina. (In Portuguese).

Romano, A., Spadaro, G., Balliet, D., Joireman, J., van Lissa, C., Jin, S., ... & Leander, N. P. (2021). Cooperation and trust across societies during the COVID-19 pandemic. *Journal of Cross-Cultural Psychology*, *52*(7), 622–642. https://doi.org/10.1177/00220221209889

Ruan, W., Yi, X., & Huang, X. (2021). Adversarial robustness of deep learning: Theory, algorithms, and applications. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management* (pp. 4866–4869). https://doi.org/10.48550/arXiv.2108.10451

Rubin, O., Errett, N., Upshur, R., & Baekkeskov, E. (2021). The challenges facing evidence-based decision making in the initial response to COVID-19. *Scandinavian Journal of Public Health*, *49*(7), 790–796. https://doi.org/10.1177/140349482199722

Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.

Sass, J., Bartschke, A., Lehne, M., Essenwanger, A., Rinaldi, E., Rudolph, S., ... & Thun, S. (2020). The German Corona Consensus Dataset (GECCO): a standardized dataset for COVID-19 research in university medicine and beyond. *BMC Medical Informatics and Decision Making*, *20*, 341. https://doi.org/10.1186/s12911-020-01374-w

Shin, D., & Park, Y. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior*, *98*, 277–284. https://doi.org/10.1016/j.chb.2019.04.019

Shaelou, S. L., & Razmetaeva, Y. (2023). Challenges to fundamental human rights in the age of artificial intelligence systems: Shaping the digital legal order while upholding rule of law principles and European values. *ERA Forum*, *24*(3), 567–587. https://doi.org/10.1007/s12027-023-00777-2

Silva, M., Flood, C., Goldenberg, A., & Singh, D. (2022). Regulating the Safety of Health-Related Artificial Intelligence. *Healthcare Policy*, *17*(4), 63–77. https://doi.org/10.12927/hcpol.2022.26824

Smidt, H., & Jokonya, O. (2021). The challenge of privacy and security when using technology to track people in times of COVID-19 pandemic. *Procedia Computer Science*, *181*, 1018–1026. https://doi.org/10.1016/j.procs.2021.01.281

Sundar, S. (2020). Rise of machine agency: A framework for studying the psychology of human – AI interaction (HAII). *Journal of Computer-Mediated Communication*, *25*(1), 74–88. https://doi.org/10.1093/jcmc/zmz026

Susskind, D. (2021). A world without work: Technology, automation and how we should respond. New Technology, *Work and Employment*, *36*(1), 114-117. https://doi.org/10.1111/ntwe.12186

Syed, R., Ulbricht, M., Piotrowski, K., & Krstic, M. (2023). A Survey on Fault-Tolerant Methodologies for Deep Neural Networks. *Pomiary Automatyka Robotyka*, *27*(2), 89–98. https://doi.org/10.14313/PAR_248/89

Syrowatka, A., Kuznetsova, M., Alsubai, A., Beckman, A., Bain, P., Craig, K., ... & Bates, D. (2021). Leveraging artificial intelligence for pandemic preparedness and response: a scoping review to identify key use cases. *npj Digital Medicine*, *4*(1), 96. https://doi.org/10.1038/s41746-021-00459-8

Theis, T., & Wong, H. (2017). The end of Moore's law: A new beginning for information technology. *Computing in Science & Engineering*, *19*(2), 41–50. https://doi.org/10.1109/MCSE.2017.29

Thomsen, K. (2019). Ethics for artificial intelligence, ethics for all. *Paladyn, Journal of Behavioral Robotics*, *10*(1), 359–363. https://doi.org/10.1515/pjbr-2019-0029

Topol, E. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, *25*(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

Villegas-Ch, W., Jaramillo-Alcázar, A., & Luján-Mora, S. (2024). Evaluating the Robustness of Deep Learning Models against Adversarial Attacks: An Analysis with FGSM, PGD and CW. *Big Data and Cognitive Computing*, *8*(1), 8. https://doi.org/10.3390/bdcc8010008

Vopson, M. (2020). The information catastrophe. *AIP Advances*, *10*(8), 085014. https://doi.org/10.1063/5.0019941

Wallach W., & Allen, C. (2008). *Moral Machines: Teaching Robots Right from Wrong*. Oxford University Press.

Wang, S., & Shi, W. (2011). Data Mining and Knowledge Discovery. In W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*. Springer Handbooks. https://doi.org/10.1007/978-3-540-72680-7_5

Wong, F., de la Fuente-Nunez, C., & Collins, J. (2023). Leveraging artificial intelligence in the fight against infectious diseases. *Science*, *381*(6654), 164–170. https://doi.org/10.1126/science.adh1114

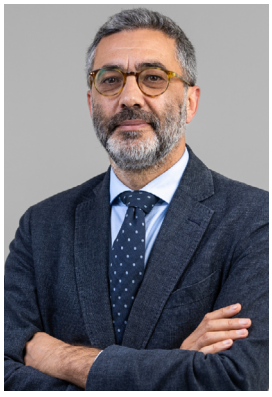
Wu, D., Xu, H., Yongyi, W., & Zhu, H. (2022). Quality of government health data in COVID-19: definition and testing of an open government health data quality evaluation framework. *Library Hi Tech*, *40*(2), 516–534. https://doi.org/10.1108/LHT-04-2021-0126

Zhang, Q., Gao, J., Wu, J., Cao, Z., & Dajun, D. (2022). Data science approaches to confronting the COVID-19 pandemic: a narrative review. *Philosophical Transactions of the Royal Society A*, *380*(2214), 20210127. https://doi.org/10.1098/rsta.2021.0127

Zhou, J., Zheng, W., Wang, D., & Coit, D. W. (2024). A resilient network recovery framework against cascading failures with deep graph learning. *Journal of Risk and Reliability*, *238*(1), 193–203. https://doi.org/10.1177/1748006X22112886

## Сведения об авторах

**Коррейя Педро Мигель Алвес Рибейро** – PhD в области общественных наук (государственное управление), приглашенный доцент, юридический факультет, Коимбрский университет; приглашенный профессор, ICET/CUA/UFMT, Барра до Гарсас
**Адрес**: Португалия, 3004-528, г. Коимбра, Патио да Универсидаде; Бразилия, 78605-091, Авенида Валдон Варжан, 6390, Барра до Гарсас – MT, CEP
**E-mail**: pcorreia@fd.uc.pt
**ORCID ID**: https://orcid.org/0000-0002-3111-9843
**Scopus Author ID**: https://www.scopus.com/authid/detail.uri?authorId=58223408400
**WoS Researcher ID**: https://www.webofscience.com/wos/author/record/B-2753-2015
**Google Scholar ID:** https://scholar.google.pt/citations?user=KABKPuUAAAAJ



**Рикардо Лопес Динис Педро** – PhD в области права, научный сотрудник, Лиссабонский исследовательский центр в области публичного права, юридический факультет, Лиссабонский университет
**Адрес:** Португалия, 1649-014, г. Лиссабон, Аламеда де Универсидаде
**E-mail**: ricardopedro@fd.ulisboa.pt
**ORCID ID**: https://orcid.org/0000-0001-6339-5140
**Scopus Author ID**: https://www.scopus.com/authid/detail.uri?authorId=57879177700
**WoS Researcher ID**: https://www.webofscience.com/wos/author/record/AEN-4511-2022
**Google Scholar ID:** https://scholar.google.com/citations?hl=en&user=oJ1ImgUAAAAJ

**Сусана Видейра** – PhD в области права, доцент, юридический факультет, Лиссабонский университет; координатор по науке и образованию, Европейский университет в Лиссабоне
**Адрес**: Португалия, 1649-014, г. Лиссабон, Аламеда де Универсидаде; Португалия, 1500-210, г. Лиссабон, Эстрада да Коррейя, 53
**E-mail**: susanavideira@fd.ulisboa.pt
**ORCID ID**: https://orcid.org/0000-0002-9246-2557

## Вклад авторов

Авторы внесли равный вклад в разработку концепции, методологии, валидацию, формальный анализ, проведение исследования, подбор источников, написание и редактирование текста, руководство и управление проектом.

## Конфликт интересов

Авторы сообщают об отсутствии конфликта интересов.

## Финансирование

Участие автора Рикардо Лопес Динис Педро частично финансировалось Фондом науки и технологий Португалии (Foundation for Science and Technology, FCT) в рамках проекта UIDP/04310/2020. Исследование также было поддержано тем же Фондом в рамках проекта UIDB/04643/2020.

## Тематические рубрики

**Рубрика OECD**: 5.05 / Law
**Рубрика ASJC**: 3308 / Law
**Рубрика WoS**: OM / Law
**Рубрика ГРНТИ**: 10.15.59 / Права и свободы человека и гражданина
**Специальность ВАК**: 5.1.2 / Публично-правовые (государственно-правовые) науки

## История статьи

**Дата поступления** – 15 июня 2024 г.
**Дата одобрения после рецензирования** – 27 июня 2024 г.
**Дата принятия к опубликованию** – 25 марта 2025 г.
**Дата онлайн-размещения** – 30 марта 2025 г.

Research article
UDC 34:004:17:004.8:342.7
EDN: https://elibrary.ru/egkppn
DOI: https://doi.org/10.21202/jdtl.2025.7

# Artificial Intelligence in Healthcare: Balancing Innovation, Ethics, and Human Rights Protection

**Pedro Miguel Alves Ribeiro Correia** ✉
University of Coimbra, Coimbra, Portugal

**Ricardo Lopes Dinis Pedro**
Lisbon Public Law Research Centre University of Lisbon

**Susana Videira**
University of Lisbon, Lisboa, Portugal

## Keywords

artificial intelligence,
data protection,
ethical regulation,
ethics,
fundamental rights,
healthcare,
human rights,
law,
legal regulation,
predictive analytics

## Abstract

**Objective**: to identify key ethical, legal and social challenges related to the use of artificial intelligence in healthcare; to develop recommendations for creating adaptive legal mechanisms that can ensure a balance between innovation, ethical regulation and the protection of fundamental human rights.

**Methods**: a multidimensional methodological approach was implemented, integrating classical legal analysis methods with modern tools of comparative jurisprudence. The study covers both the fundamental legal regulation of digital technologies in the medical field and the in-depth analysis of the ethical, legal and social implications of using artificial intelligence in healthcare. Such an integrated approach provides a comprehensive understanding of the issues and well-grounded conclusions about the development prospects in this area.

**Results**: has revealed a number of serious problems related to the use of artificial intelligence in healthcare. These include data bias, non-transparent complex algorithms, and privacy violation risks. These problems can undermine public confidence in artificial intelligence technologies and exacerbate inequalities in access to health services. The authors conclude that the integration of artificial intelligence into healthcare should take into account fundamental rights, such as data protection and non-discrimination, and comply with ethical standards.

✉ Corresponding author

**Scientific novelty**: the work proposes effective mechanisms to reduce risks and maximize the potential of artificial intelligence under crises. Special attention is paid to regulatory measures, such as the impact assessment provided for by the Artificial Intelligence Act. These measures play a key role in identifying and minimizing the risks associated with high-risk artificial intelligence systems, ensuring compliance with ethical standards and protection of fundamental rights.

**Practical significance**: adaptive legal mechanisms were developed, that support democratic norms and respond promptly to emerging challenges in public healthcare. The proposed mechanisms allow achieving a balance between using artificial intelligence for crisis management and human rights. This helps to build confidence in artificial intelligence systems and their sustained positive impact on public healthcare.

## For citation

Correia, P. M. A. R., Pedro, R. L. D., & Videira, S. (2025). Artificial Intelligence in Healthcare: Balancing Innovation, Ethics, and Human Rights Protection. *Journal of Digital Technologies and Law, 3*(1), 143–180. https://doi.org/10.21202/jdtl.2025.7

## References

Arass, M., & Souissi, N. (2018). Data lifecycle: from big data to SmartData. In *2018 IEEE 5th International Congress on Information Science and Technology* (pp. 80–87). IEEE. https://doi.org/10.1109/CIST.2018.8596547

Alvarez Garcia, V. (1996). *El concepto de necesidad en derecho público* (1st ed.). Madrid: Civitas. (In Spanish).

Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Baclic, O., Tunis, M., Young, K., Doan, C., Swerdfeger, H., & Schonfeld, J. (2020). Challenges and opportunities for public health made possible by advances in natural language processing. *Canada Communicable Disease Report*, *46*(6), 161–168. https://doi.org/10.14745/ccdr.v46i06a02

Bajwa, J., Munir, U., Nori, A., & Williams, B. (2021). Artificial intelligence in healthcare: transforming the practice of medicine. *Future Healthcare Journal*, *8*(2), e188-e194. https://doi.org/10.7861/fhj.2021-0095

Beck, U. (1986). *Risikogesellschaft: Auf dem Weg in eine andere Moderne*. Frankfurt am Main: Suhrkamp Verlag.

Balog-Way, D., & McComas, K. (2022). COVID-19: Reflections on trust, tradeoffs, and preparedness. In *COVID-19* (pp. 6–16). Routledge.

Bazarkina, D. Y., & Pashentsev, E. N. (2020). Malicious use of artificial intelligence. *Russia in Global Affairs*, *18*(4), 154–177. https://doi.org/10.31278/1810-6374-2020-18-4-154-177

Benke, K., & Benke, G. (2018). Artificial Intelligence and Big Data in Public Health. *International Journal of Environmental Research and Public Health*, *15*(12), 2796. https://doi.org/10.3390/ijerph15122796

Berk, R. A. (1983). An introduction to sample selection bias in sociological data. *American Sociological Review*, *48*(3), 386–398. https://doi.org/10.2307/2095230

Bigham, G., Adamtey, S., Onsarigo, L., & Jha, N. (2019). Artificial Intelligence for Construction Safety: Mitigation of the Risk of Fall. In K. Arai, S. Kapoor, R. Bhatia (Eds.). *Intelligent Systems and Applications*. Springer. https://doi.org/10.1007/978-3-030-01057-7_76

Binder, W. (2024). Technology as (dis-)enchantment. AlphaGo and the meaning-making of artificial intelligence. *Cultural Sociology*, *18*(1), 24–47. https://doi.org/10.1177/17499755221138720

Bisconti, P., Orsitto, D., Fedorczyk, F., Brau, F., Capasso, M., De Marinis, L., ... & Schettini, C. (2023). Maximizing team synergy in AI-related interdisciplinary groups: an interdisciplinary-by-design iterative methodology. *AI & Society*, *38*(4), 1443–1452. https://doi.org/10.1007/s00146-022-01518-8

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Box, G. (1979). Robustness in the strategy of scientific model building. In R. Launer & G. Wilkinson (Eds.), *Robustness in Statistics* (pp. 201–236). Academic Press. https://doi.org/10.1016/B978-0-12-438150-6.50018-2

Breiman, L. (2001). Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3), 199–231. https://doi.org/10.1214/ss/1009213726

Bulled, N. (2023). "Solidarity:" A failed call to action during the COVID-19 pandemic. *Public Health in Practice*, *5*, 100379. https://doi.org/10.1016/j.puhip.2023.100379

Chen, A. (2016). A review of emerging non-volatile memory (NVM) technologies and applications. *Solid-State Electronics*, *125*, 25–38. https://doi.org/10.1016/j.sse.2016.07.006

Chen, J., Zhang, R., Han, W., Jiang, W., Hu, J., Lu, X., Liu, X., & Zhao, P. (2020). Path Planning for Autonomous Vehicle Based on a Two-Layered Planning Model in Complex Environment. *Journal of Advanced Transportation*, 2020, 6649867. https://doi.org/10.1155/2020/6649867

Chiao, V. (2019). Fairness, accountability and transparency: notes on algorithmic decision-making in criminal justice. *International Journal of Law in Context*, *15*(2), 126–139. https://doi.org/10.1017/S1744552319000077

Correia, P., Mendes, I., Pereira, S., & Subtil, I. (2020a). The combat against COVID-19 in Portugal: How state measures and data availability reinforce some organizational values and contribute to the sustainability of the National Health System. *Sustainability*, *12*(18), 7513. https://doi.org/10.3390/su12187513

Correia, P., Mendes, I., Pereira, S., & Subtil, I. (2020b). The combat against COVID-19 in Portugal, Part II: how governance reinforces some organizational values and contributes to the sustainability of crisis management. *Sustainability*, *12*(20), 8715. https://doi.org/10.3390/su12208715

Correia, P., Pereira, S., Mendes, I., & Subtil, I. (2022). COVID-19 Crisis management and the Portuguese regional governance: Citizens perceptions as evidence. *European Journal of Applied Business Management*, *8*(1), 1–12.

Correia, P., Pereira, S., Mendes, I., & Subtil, I. (2021). COVID-19 Crisis management and the Portuguese regional governance: Citizens perceptions as evidence. In *European Consortium for Political Research General Conference* (pp. 1–18). United Kingdom.

Correia, J. M. C. (1987). *Legalidade e autonomia contratual nos contratos administrativos* (pp. 283, 768). Lisboa: Almedina.

DeCamp, M., & Tilburt, J. (2019). Why we cannot trust artificial intelligence in medicine. *The Lancet Digital health*, *1*(8), e390. https://doi.org/10.1016/S2589-7500(19)30197-9

Dhingra, M., & Gupta, N. (2017). Comparative analysis of fault tolerance models and their challenges in cloud computing. *International Journal of Engineering & Technology*, *6*(2), 36–40. https://doi.org/10.14419/ijet.v6i2.7565

Ettlinger, N. (2022). *Algorithms and the Assault on Critical Thought: Digitalized Dilemmas of Automated Governance and Communitarian Practice* (1st ed.). Routledge. https://doi.org/10.4324/9781003109792

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. New York: Picador, St Martin's Press.

Ferguson, N., Cummings, D., Fraser, C., Cajka, J., Cooley, P., & Burke, D. (2006). Strategies for mitigating an influenza pandemic. *Nature*, *442*(7101), 448–452. https://doi.org/10.1038/nature04795

Fetzer, T., & Graeber, T. (2021). Measuring the scientific effectiveness of contact tracing: Evidence from a natural experiment. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(33), e2100814118. https://doi.org/10.1073/pnas.2100814118

Galetsi, P., Katsaliaki, K., & Kumar, S. (2022). The medical and societal impact of big data analytics and artificial intelligence applications in combating pandemics: A review focused on Covid-19. *Social Science & Medicine*, *301*, 114973. https://doi.org/10.1016/j.socscimed.2022.114973

Gianfrancesco, M., Tamang, S., Yazdany, J., & Schmajuk, G. (2018). Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data. *JAMA Internal Medicine*, *178*(11), 1544–1547. https://doi.org/10.1001/jamainternmed.2018.3763

Goldman, N., Bertone, P., Chen, S., Dessimoz, C., LeProust, E. M., Sipos, B., & Birney, E. (2013). Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature*, *494*(7435), 77–80. https://doi.org/10.1038/nature11875

Gomes, C. A., & Pedro, R. (Coords.). (2020). *Direito administrativo de necessidade e de excepção*. Lisboa: AAFDL.

Gómez Abeja, L. (2022). Inteligencia artificial y derechos fundamentales. In F. H. Llano Alonso (Dir.), J. Garrido Martín & R. Valdivia Jiménez (Coords.), *Inteligencia artificial y filosofía del derecho* (1.ª ed., pp. 91–114, 93). Murcia: Ediciones Laborum. (In Spanish).

Gómez Colomer, J.-L. (2023). *El juez-robot: La independencia judicial en peligro*. Valencia: Tirant lo Blanch. (In Spanish).

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.
Greiner, R., Grove, A., & Kogan, A. (1997). Knowing what doesn't matter: exploiting the omission of irrelevant data. *Artificial Intelligence*, *97*(1–2), 345–380. https://doi.org/10.1016/S0004-3702(97)00048-9
Gunasekeran, D., Tseng, R., Tham, Y., & Wong, T. (2021). Applications of digital health for public health responses to COVID-19: a systematic scoping review of artificial intelligence, telehealth and related technologies. *NPJ Digital Medicine*, *4*(1), 40. https://doi.org/10.1038/s41746-021-00412-9
Gürsoy, E., & Kaya, Y. (2023). An overview of deep learning techniques for COVID-19 detection: methods, challenges, and future works. *Multimedia Systems*, *29*(3), 1603–1627. https://doi.org/10.1007/s00530-023-01083-0
Hanegraaff, W. (2013). *Western Esotericism: A Guide for the Perplexed*. Bloomsbury Publishing.
Halevy, A., Norvig, P., & Pereira, F. (2009). The unreasonable effectiveness of data. *IEEE Intelligent Systems*, 24(2), 8–12. https://doi.org/10.1109/MIS.2009.36
Hazarika, I. (2020). Artificial intelligence: opportunities and implications for the health workforce. *International Health*, *12*(4), 241–245. https://doi.org/10.1093/inthealth/ihaa007
Hoff, K., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407–434. https://doi.org/10.1177/0018720814547570
Hulten, G. (2018). Building Intelligent Systems: A Guide to Machine Learning Engineering. Apress.
Igual, L., & Seguí, S. (2024). *Supervised learning. In Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications* (pp. 67–97). Springer International Publishing.
Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular Neurology*, *2*(4), 230–243. https://doi.org/10.1136/svn-2017-000101
Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*, 389–399. https://doi.org/10.1038/s42256-019-0088-2
Jones, K., Patel, N., Levy, M., Storeygard, A., Balk, D., Gittleman, J., & Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature*, *451*(7181), 990–993. https://doi.org/10.1038/nature06536
Kandlhofer, M., Weixelbraun, P., Menzinger, M., Steinbauer-Wagner, G., & Kemenesi, Á. (2023). Education and Awareness for Artificial Intelligence. In *International Conference on Informatics in Schools: Situation, Evolution, and Perspectives* (pp. 3–12). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-44900-0_1
Kavanagh, J., & Rich, M. (2018). *Truth Decay: An Initial Exploration of the Diminishing Role of Facts and Analysis in American Public Life*. RAND Corporation. https://doi.org/10.7249/RR2314
Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., ... & Hadsell, R. (2017). Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, *114*(13), 3521–3526. https://doi.org/10.1073/pnas.1611835114
Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: review, synthesis, and future research directions. E*uropean Journal of Information Systems*, *31*(3), 388–409. https://doi.org/10.1080/0960085X.2021.1927212
Larsson, S., & Heintz, F. (2020). Transparency in artificial intelligence. *Internet Policy Review*, *9*(2). https://doi.org/10.14763/2020.2.1469
Lin, X., Liu, J., Hao, J., Wang, K., Zhang, Y., Li, H., ... & Tan, X. (2020). Collinear holographic data storage technologies. *Opto-Electronic Advances*, *3*(3), 190004. https://doi.org/10.29026/oea.2020.190004
Little, R. J., & Rubin, D. B. (2019). Statistical analysis with missing data. John Wiley & Sons.
Macrae, C. (2022). Learning from the failure of autonomous and intelligent systems: Accidents, safety, and sociotechnical sources of risk. *Risk Analysis*, *42*(9), 1999–2025. https://doi.org/10.1111/risa.13850
Margetts, H. (2022). Rethinking AI for good governance. *Daedalus*, *151*(2), 360–371. https://doi.org/10.1162/daed_a_01922
Matsuzaka, Y., & Yashiro, R. (2022). Applications of Deep Learning for Drug Discovery Systems with BigData. *BioMedInformatics*, *2*(4), 603–624. https://doi.org/10.3390/biomedinformatics2040039
Mittelstadt, B., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2). https://doi.org/10.1177/2053951716679679
Morse, S., Mazet, J., Woolhouse, M., Parrish, C., Carroll, D., Karesh, W., Zambrana-Torrelio, C., Lipkin, W., & Daszak, P. (2012). Prediction and prevention of the next pandemic zoonosis. *Lancet*, *380*(9857), 1956–1965. https://doi.org/10.1016/S0140-6736(12)61684-5
Mumuni, A., & Mumuni, F. (2022). Data augmentation: A comprehensive survey of modern approaches. *Array*, *16*, 100258. https://doi.org/10.1016/j.array.2022.10025
Navigli, R., Conia, S., & Ross, B. (2023). Biases in Large Language Models: Origins, Inventory, and Discussion. *Journal of Data and Information Quality*, *15*(2), 10. https://doi.org/10.1145/3597307

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

O'Reilly-Shah, V., Gentry, K., van Cleve, W., Kendale, S., Jabaley, C., & Long, D. (2020). The COVID-19 pandemic highlights shortcomings in US health care informatics infrastructure: a call to action. *Anesthesia & Analgesia*, *131*(2), 340–344. https://doi.org/10.1213/ANE.0000000000004945

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors*, *39*(2), 230–253. https://doi.org/10.1518/001872097778543886

Parasuraman, R., Sheridan, T., & Wickens, C. (2000). A model for types and levels of human interaction with automation. *Systems and Humans*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

Pedro, R. (2022). Traços gerais da indemnização civil extracontratual pública em contextos de excecionalidade. In *Impactos da pandemia da Covid-19 nas estruturas do direito público* (pp. 379–413). Coimbra: Almedina. (In Portuguese).

Pedro, R. (2023). Inteligência artificial e arbitragem de direito público: Primeiras reflexões. In R. Pedro, & P. Caliendo (Coords.), *Inteligência artificial no contexto do direito público: Portugal e Brasil* (1.ª ed., pp. 105–127). Coimbra: Almedina. (In Portuguese).

Romano, A., Spadaro, G., Balliet, D., Joireman, J., van Lissa, C., Jin, S., ... & Leander, N. P. (2021). Cooperation and trust across societies during the COVID-19 pandemic. *Journal of Cross-Cultural Psychology*, *52*(7), 622–642. https://doi.org/10.1177/00220221209889

Ruan, W., Yi, X., & Huang, X. (2021). Adversarial robustness of deep learning: Theory, algorithms, and applications. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management* (pp. 4866–4869). https://doi.org/10.48550/arXiv.2108.10451

Rubin, O., Errett, N., Upshur, R., & Baekkeskov, E. (2021). The challenges facing evidence-based decision making in the initial response to COVID-19. *Scandinavian Journal of Public Health*, *49*(7), 790–796. https://doi.org/10.1177/140349482199722

Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.

Sass, J., Bartschke, A., Lehne, M., Essenwanger, A., Rinaldi, E., Rudolph, S., ... & Thun, S. (2020). The German Corona Consensus Dataset (GECCO): a standardized dataset for COVID-19 research in university medicine and beyond. *BMC Medical Informatics and Decision Making*, *20*, 341. https://doi.org/10.1186/s12911-020-01374-w

Shin, D., & Park, Y. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior*, *98*, 277–284. https://doi.org/10.1016/j.chb.2019.04.019

Shaelou, S. L., & Razmetaeva, Y. (2023). Challenges to fundamental human rights in the age of artificial intelligence systems: Shaping the digital legal order while upholding rule of law principles and European values. *ERA Forum*, *24*(3), 567–587. https://doi.org/10.1007/s12027-023-00777-2

Silva, M., Flood, C., Goldenberg, A., & Singh, D. (2022). Regulating the Safety of Health-Related Artificial Intelligence. *Healthcare Policy*, *17*(4), 63–77. https://doi.org/10.12927/hcpol.2022.26824

Smidt, H., & Jokonya, O. (2021). The challenge of privacy and security when using technology to track people in times of COVID-19 pandemic. *Procedia Computer Science*, *181*, 1018–1026. https://doi.org/10.1016/j.procs.2021.01.281

Sundar, S. (2020). Rise of machine agency: A framework for studying the psychology of human – AI interaction (HAII). *Journal of Computer-Mediated Communication*, *25*(1), 74–88. https://doi.org/10.1093/jcmc/zmz026

Susskind, D. (2021). A world without work: Technology, automation and how we should respond. New Technology, *Work and Employment*, *36*(1), 114-117. https://doi.org/10.1111/ntwe.12186

Syed, R., Ulbricht, M., Piotrowski, K., & Krstic, M. (2023). A Survey on Fault-Tolerant Methodologies for Deep Neural Networks. *Pomiary Automatyka Robotyka*, *27*(2), 89–98. https://doi.org/10.14313/PAR_248/89

Syrowatka, A., Kuznetsova, M., Alsubai, A., Beckman, A., Bain, P., Craig, K., ... & Bates, D. (2021). Leveraging artificial intelligence for pandemic preparedness and response: a scoping review to identify key use cases. *npj Digital Medicine*, *4*(1), 96. https://doi.org/10.1038/s41746-021-00459-8

Theis, T., & Wong, H. (2017). The end of Moore's law: A new beginning for information technology. *Computing in Science & Engineering*, *19*(2), 41–50. https://doi.org/10.1109/MCSE.2017.29

Thomsen, K. (2019). Ethics for artificial intelligence, ethics for all. *Paladyn, Journal of Behavioral Robotics*, *10*(1), 359–363. https://doi.org/10.1515/pjbr-2019-0029

Topol, E. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, *25*(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

Villegas-Ch, W., Jaramillo-Alcázar, A., & Luján-Mora, S. (2024). Evaluating the Robustness of Deep Learning Models against Adversarial Attacks: An Analysis with FGSM, PGD and CW. *Big Data and Cognitive Computing*, *8*(1), 8. https://doi.org/10.3390/bdcc8010008

Vopson, M. (2020). The information catastrophe. *AIP Advances*, *10*(8), 085014. https://doi.org/10.1063/5.0019941

Wallach W., & Allen, C. (2008). *Moral Machines: Teaching Robots Right from Wrong*. Oxford University Press.

Wang, S., & Shi, W. (2011). Data Mining and Knowledge Discovery. In W. Kresse, D. Danko (Eds.), *Springer Handbook of Geographic Information*. Springer Handbooks. https://doi.org/10.1007/978-3-540-72680-7_5

Wong, F., de la Fuente-Nunez, C., & Collins, J. (2023). Leveraging artificial intelligence in the fight against infectious diseases. *Science*, *381*(6654), 164–170. https://doi.org/10.1126/science.adh1114

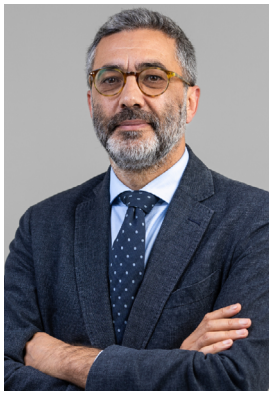
Wu, D., Xu, H., Yongyi, W., & Zhu, H. (2022). Quality of government health data in COVID-19: definition and testing of an open government health data quality evaluation framework. *Library Hi Tech*, *40*(2), 516–534. https://doi.org/10.1108/LHT-04-2021-0126

Zhang, Q., Gao, J., Wu, J., Cao, Z., & Dajun, D. (2022). Data science approaches to confronting the COVID-19 pandemic: a narrative review. *Philosophical Transactions of the Royal Society A*, *380*(2214), 20210127. https://doi.org/10.1098/rsta.2021.0127

Zhou, J., Zheng, W., Wang, D., & Coit, D. W. (2024). A resilient network recovery framework against cascading failures with deep graph learning. *Journal of Risk and Reliability*, *238*(1), 193–203. https://doi.org/10.1177/1748006X22112886

## Authors information

**Pedro Miguel Alves Ribeiro Correia** – PhD in Social Sciences (Specialty in Public Administration), Invited Associate Professor, Faculty of Law, University of Coimbra; Visiting Full Professor, ICET/CUA/UFMT, Barra do Garças
**Address**: Pátio da Universidade, 3004-528 Coimbra, Portugal;
Avenida ValdonVarjão, n. 6390, Barra do Garças – MT, CEP: 78605-091, Brazil
**E-mail**: pcorreia@fd.uc.pt
**ORCID ID**: https://orcid.org/0000-0002-3111-9843
**Scopus Author ID**: https://www.scopus.com/authid/detail.uri?authorId=58223408400
**WoS Researcher ID**: https://www.webofscience.com/wos/author/record/B-2753-2015
**Google Scholar ID:** https://scholar.google.pt/citations?user=KABKPuUAAAAJ



**Ricardo Lopes Dinis Pedro** – PhD (Law), Researcher, Lisbon Public Law Research Centre, Faculty of Law, University of Lisbon
**Address**: Alameda da Universidade, 1649-014 Lisbon, Portugal
**E-mail**: ricardopedro@fd.ulisboa.pt
**ORCID ID**: https://orcid.org/0000-0001-6339-5140
**Scopus Author ID**: https://www.scopus.com/authid/detail.uri?authorId=57879177700
**WoS Researcher ID**: https://www.webofscience.com/wos/author/record/AEN-4511-2022
**Google Scholar ID:** https://scholar.google.com/citations?hl=en&user=oJ1ImgUAAAAJ

**Susana Videira** – PhD (Law), Associate Professor, Faculty of Law, University of Lisbon; Scientific and Pedagogical Coordinator, Law Degree and the Master's Degree in Judicial Law, European University
**Address**: Faculdade de Direito da Universidade de Lisboa, Alameda da Universidade, 1649-014 Lisbon, Portugal; Universidade Europeia, Estrada da Correia, n.º 53, 1500-210, Lisbon, Portugal
**E-mail**: susanavideira@fd.ulisboa.pt
**ORCID ID**: https://orcid.org/0000-0002-9246-2557

## Authors' contributions

The authors have contributed equally into the concept and methodology elaboration, validation, formal analysis, research, selection of sources, text writing and editing, project guidance and management.

## Conflict of interests

The authors declare no conflict of interests.

## Financial disclosure

Regarding the participation of the Author Ricardo Pedro, it should be noted that, to the exact extent of his participation, the work is financed (or partially financed) by national funds through FCT–Foundation for Science and Technology, I.P., under the project UIDP/04310/2020. This work was also supported by Portuguese national funds through FCT–Foundation for Science and Technology, I.P., under project UIDB/04643/2020.

## Thematic rubrics

**OECD**: 5.05 / Law
**PASJC**: 3308 / Law
**WoS**: OM / Law

## Article history

**Date of receipt** – June 15, 2025
**Date of approval** – June 27, 2025
**Date of acceptance** – March 25, 2025
**Date of online placement** – March 30, 2025