Право. Журнал Высшей школы экономики. 2025. Том 18. № 3.
Law. Journal of the Higher School of Economics. 2025. Vol. 18, no 3.

Правовая мысль: история и современность

Научная статья
УДК: 349
JEL: K3
DOI:10.17323/2072-8166.2025.3.4.27

Прозрачность алгоритмов искусственного интеллекта

Эльвира Владимировна Талапина
Институт государства и права РАН, Россия, Москва 119019, Знаменка, 10,
talapina@mail.ru, https://orcid.org/0000-0003-3395-3126

Аннотация

В эпоху активного освоения искусственного интеллекта (ИИ) перед юристами встает вопрос, как решить проблему «черного ящика», непонятности и непредсказуемости решений, принимаемых искусственным интеллектом. Разработка правил, обеспечивающих прозрачность и понятность алгоритмов ИИ, позволяет вписать его в классические правоотношения, устраняя угрозу институту юридической ответственности. В частном праве защита потребителей перед лицом крупных онлайн-платформ выдвигает прозрачность алгоритмов на первый план, меняя само обязательство по предоставлению информации потребителю, которое теперь описывается формулой «знать + понимать». Аналогично и в праве публичном — государства не в состоянии должным образом защитить граждан от вреда, причиняемого зависимостью от алгоритмов при оказании государственных услуг. Противопоставить этому можно только знание и понимание функционирования алгоритмов. Требуется принципиально новое регулирование, позволяющее ввести использование искусственного интеллекта в легальные рамки, в которых следует формулировать требования к прозрачности алгоритмов. Эксперты активно обсуждают создание нормативной базы для формирования системы наблюдения за внедрением и использованием технологий ИИ. Разрабатываются меры «политики алгоритмической подотчетности», фреймворк «Прозрачность через проектирование» с акцентом на постоянном взаимодействии с заинтересованными сторонами и организационной открытости, обосновывается внедрение объяснимых систем ИИ. В целом предлагаемые подходы к регулированию ИИ и обеспечению прозрачности схожи, равно как и прогнозы о смягчающей роли прозрачности алгоритмов ИИ в вопросах доверия к ИИ. Интересна концепция «алгоритмического суверенитета», относящаяся к способности государства управлять разработкой,

развертыванием и воздействием систем ИИ в соответствии с собственными правовыми, культурными и этическими нормами. Обеспечение прозрачности алгоритмов ИИ — направление общей политики управления в области ИИ, важнейшей частью которой является этика ИИ. Несмотря на кажущуюся универсальность, этика ИИ не всегда учитывает все разнообразие этических конструкций в разных уголках мира, что демонстрирует африканский пример и опасения алгоритмической колонизации.

Ключевые слова

алгоритм; искусственный интеллект; этика искусственного интеллекта; прозрачность; доверие; подотчетность; ответственность.

Для цитирования: Талапина Э.В. Прозрачность алгоритмов искусственного интеллекта // Право. Журнал Высшей школы экономики. 2025. Том 18. № 3. С. 4–27. DOI:10.17323/2072-8166.2025.3.4.27

Legal Thought: History and Modernity

Research article

Transparency of Artificial Intelligence Algorithms

Elvira V. Talapina

Institute of State and Law, 10 Znamenka St., Moscow 119019, Russian Federation,

talapina@mail.ru

https://orcid.org/0000-0003-3395-3126

Abstract

In the modern era of active development of artificial intelligence (AI), lawyers are faced with the question: how to solve the “black box” matter, the incomprehensibility and unpredictability of decisions made by artificial intelligence. The development of rules ensuring transparency and explainability of AI algorithms allows artificial intelligence to be integrated into classical legal relations, eliminating the threat to the institution of legal liability. In Private Law consumer protection in front of large online platforms brings the algorithms transparency to the forefront, changing the very obligation to provide information to the consumer, which is now described by the formula: know + understand. Similarly, in Public Law, states are unable to properly protect citizens from harm caused by dependence on algorithmic applications in the provision of public services. It can only be countered by knowledge and understanding of the functioning of algorithms. A fundamentally new regulation is required to introduce the artificial intelligence use into a legal framework in which requirements for the transparency of algorithms should be formulated. Researchers are actively discussing creation of a regulatory framework for the formation of a system of observation, monitoring and preliminary permission for the AI technologies use. The paper analyzes “algorithmic accountability policies” and a “Transparency by Design” framework (problem solving throughout the entire AI

development process) and the implementation of explainable AI systems. Overall, the proposed approaches to AI regulation and transparency are quite similar, as are the predictions about the mitigating role of AI algorithm transparency in matters of trust in AI. The concept of “algorithmic sovereignty” which refers to the ability of a democratic State to govern the development, deployment, and impact of AI systems in accordance with its own legal, cultural, and ethical norms, is also analyzed. This model is designed for the harmonious coexistence of different states, leading to an equally harmonious coexistence between humanity and AI. At the same time, ensuring the AI algorithms transparency is a direction of the general AI governance policy, the most important part of which is AI ethics. Despite its apparent universality, artificial intelligence ethics does not always take into account the diversity of ethical constructs in different parts of the world, as the African example demonstrates as well as fears of algorithmic colonization.

Keywords

algorithm; artificial intelligence; AI ethics; transparency; trust; accountability; liability.

For citation: Talapina E.V. (2025) Transparency of Artificial Intelligence Algorithms. *Law. Journal of the Higher School of Economics,* vol.18, no. 3, pp. 4–27. DOI:10.17323/2072-8166.2025.3.4.27

Введение

Современная стадия технологического развития общества уникальна — мы уже сосуществуем в повседневной жизни с искусственным интеллектом (далее — ИИ), наблюдаем непосредственно его преимущества и минусы, пытаемся сформировать свою точку зрения и отношение на собственном индивидуальном уровне. В то же время на государственном уровне идет конкуренция за «обуздание» ИИ, предполагающее одновременно его развитие и использование по правилам. Правилам, которых пока нет. Освоение ИИ до сих пор остается «серой зоной», в которой действие традиционного права ограничено (специально или фактически); формирование нормативного контура происходит медленно и осторожно, оно требует тесного взаимодействия с техническими специалистами. Одна из нерешенных пока проблем — проблемы «черного ящика», непонятности и непредсказуемости решений, принимаемых ИИ. Право движется по пути установления правил разработки и применения алгоритмов ИИ, обеспечивающих прозрачность и понятность, поскольку это позволяет обнаружить ошибки и отклонения, обжаловать и исправить их, привлечь к ответственности виновных, и тем самым вписать ИИ в классические правоотношения. Полезно рассмотреть, какие пути предлагает современная наука в деле обеспечения прозрачности алгоритмов ИИ.

1. В поисках баланса публичного и частного права

В вопросах прозрачности и открытости информации публичное и частное право давно конкурируют. С одной стороны, частное право заинтересовано в сохранении закрытости, установлении режима тайн, конфиденциальной информации и пр., с другой — защита прав потребителей диктует необходимость раскрывать определенные виды информации. Публичное право в свою очередь тоже поощряло склонность государства хранить информацию «при себе», в глубинах бюрократического аппарата. Однако административные реформы побудили бюрократию к переменам, внедряя политику открытости в деятельность государственных органов. Технологические изменения выступили катализатором всех этих процессов, в результате которых изменились и площадка, и форма сообщения информации, а также быстрота ее обращения, расширились возможности верификации. На данной стадии технологического развития, характеризуемой активным использованием алгоритмов ИИ, вопросы правомерности и этичности его применения выдвигаются на первый план.

Первоначально прозрачность алгоритмов заботила частное право, имея в виду защиту потребителей перед лицом крупных онлайн-платформ. Изначально, до эпохи алгоритмов ИИ, в основе отношений потребителей и бизнеса лежала классическая экономическая теория, рассматривающая потребителя как рационального экономического субъекта, способного совершить верный выбор. Для этого профессиональный продавец должен сообщить потребителю всю необходимую информацию, что долгое время и делалось в рамках информационной прозрачности в отношениях типа B2C [Sposini L., 2024: 2]. В нормативном плане это нашло отражение в законодательстве о защите прав потребителей, содержащем целый диапазон правовых решений потенциальных и реальных потребительских споров, центральным элементом которых выступает информация.

Однако вскоре было подмечено, что человек не настолько рационален, как это может показаться, и на его выбор влияют социальный контекст, окружающая среда, предубеждения и др., что в результате может сделать выбор неудачным. В 1978 году Герберт Саймон разработал теорию ограниченной рациональности, в которой подчеркивается, что эмоциональная составляющая при принятии решения действует раньше рациональной. В наши дни другой нобелевский лауреат по экономике — Ричард Талер получил премию за теорию новой поведенческой экономики, противопоставленную теории эффективного рынка, доказав иррациональность человеческого

поведения в казалось бы очевидных экономических ситуациях. Поведенческая экономика стала мейнстримом, хотя, по утверждению самого Талера, «процесс развития обогащенной версии экономической науки, в центре внимания которой просто Люди, еще далек от завершения» [Талер Р., 2018: 355]. Для юристов такой экономический тренд означает отказ от доминантной призмы рациональной эффективности при разработке и анализе правовых норм. Если добавить в эту смесь технологический компонент, получим еще одно подтверждение нового тренда в виде человекоцентричности использования технологий, повсеместно провозглашаемой в числе принципов применения ИИ. Таким образом, человек как существо иррациональное возвращается в центр внимания разных наук.

В плане отношений с потребителями указанные тенденции побудили к разработке алгоритмов ИИ, способных предсказывать и предугадывать черты характера пользователей, чтобы затем использовать их для персонализации предложений с высокой степенью внушаемости. Если смотреть на это шире, то речь идет о глобальной тенденции к персонализации отношений, в которых используется ИИ, в том числе в сфере публичного права. О персонализации коммерческих услуг известно немало, но не следует оставлять без внимания и персонализацию услуг государственных.

Индивидуальное предложение продукта или услуги — многоступенчатый процесс. Сначала алгоритмы собирают данные, которые пользователь поставляет (при согласии с условиями договора или при регистрации) в качестве «встречного обязательства» за использование услуг провайдера. На деле собираются не только данные, поставленные с согласия, поскольку алгоритмы также изучают следы, неосознанно оставленные пользователями во время интернет-серфинга (например, время, на которое курсор компьютерной мыши застыл на продукте прежде совершения покупки). Это выходит далеко за рамки классической массовой коммуникации, направленной на анонимную массу получателей, поскольку компании адаптируют продукты и услуги к потребностям и желаниям каждого потребителя.

Алгоритмы обрабатывают собранные персональные данные для анализа поведения и предпочтений потребителей и таким образом создают их цифровые профили. Этот путь не лишен отклонений, поскольку помимо угадывания поведения алгоритмы способны целенаправленно использовать уязвимости пользователей и манипулировать ими (как это произошло в известном кейсе с манипуляцией новостной лентой миллионов пользователей посредством

изменения эмоционального содержания постов[1]). И хотя упомянутому кейсу уже десять лет, государства как регуляторы до сих пор не нашли действенных механизмов для предотвращения подобных злоупотреблений.

Очевидно, в частном праве обязательство сообщения информации потребителю меняется — речь должна идти не только о том, чтобы пользователь знал все элементы информации, необходимые для выбора, но и о действительном их понимании (формула «знать + + понимать»). Можно формально обеспечить прозрачность, забросав потребителя как можно бо́льшими объемами информации, однако цель состоит в качестве этой информации. Уже на данной стадии исследования можно обозначить элементы, составляющие контур алгоритмической прозрачности — достоверная информация, изложенная понятно и своевременно. Начинается прозрачность с того, что человек всегда знает о применении в отношении него алгоритмов.

Проблема, обнаруженная в практике частного права, находит все больше подтверждений и в практике публичного права. Государства часто не в состоянии должным образом защитить граждан от вреда, причиняемого зависимостью от алгоритмических приложений при оказании государственных услуг. Можно привести множество примеров неудачных технологических новаций в сфере государственных услуг. Так, в Нидерландах получатели пособий по уходу за детьми были ложно обвинены в мошенничестве на основе модели определения рисков [Rachovitsa A., Johann N., 2022: 2–3]. Аналогичным образом, программа Robodebt в Австралии ложно приписывала гражданам долги на основе автоматически рассчитанных выплат[2]. Ложные обвинения и последующие суровые наказания могли привести и в некоторых случаях привели к серьезному ущербу невиновных граждан, вплоть до разрушения их жизни, одновременно являясь серьезным нарушением прав человека.

Если государственные органы используют алгоритмы ИИ для повышения эффективности работы или решения политических задач, их институциональные и административные практики также подвергаются изменениям. Например, практики изменяются за счет непрозрачности или изменения динамики власти. Разрабатываемые в сфере оказания государственных услуг алгоритмы ИИ могут затруднять индивидуумам построение собственных прогнозов и

[1] Available at: URL: https://www.forbes.ru/tekhnologii/internet-i-svyaz/261401-eksperiment-tsukerberga-kak-facebook-protestirovala-emotsii-700 (дата обращения: 10.08.2025)

[2] Royal Commission into the Robodebt Scheme. Report (2023). Available at: URL: https://robodebt.royalcommission.gov.au/publications/report (дата обращения: 10.08.2025)

понимание, например, путем автоматизации определения размера социальных выплат. Серьезную угрозу может нести смещение дискреционных полномочий от государственного служащего к разработчику алгоритмов ИИ. В целом обозначается опасность устоявшемуся балансу власти и отношениям между гражданами и ее органами, поскольку при автоматизированном принятии решений государственные органы могут перекладывать ответственность за подтверждение правомочности и получение льгот на самих граждан.

Традиционное право не позволяет решить указанные проблемы — ни частное, ни публичное. Очевидно, требуется принципиально новое регулирование, позволяющее ввести использование ИИ в легальные границы. Именно в рамках такого регулирования логично формулировать требования к прозрачности алгоритмов.

2. Риски использования искусственного интеллекта и пути его регулирования

Итак, развитие технологий, способных распознавать человеческие эмоции и использовать их для управления поведением потребителей, требует соответствующего законодательства для обеспечения надежной системы, уважающей устоявшиеся фундаментальные ценности.

Нужно сказать, что постепенно в мире выстраиваются подходы к регулированию ИИ, пока на национальном или региональном уровне. Одни исследователи выделяют следующие модели: регулирование, основанное на правах человека в Европейском союзе, централизованный контроль в Китае и невмешательство и минимальные ограничения в США. Каждый подход предлагает разные приоритеты: прозрачность, безопасность или инновации, но в совокупности наблюдается отсутствие единой глобальной системы управления в области ИИ [Badawy W., 2025: 4403].

Другие исследователи добавляют, что в Азиатско-Тихоокеанском регионе наблюдается разнообразие: если Китай придерживается государственного подхода, используя ИИ для социального управления и государственной безопасности, то такие страны, как Сингапур и Япония, сосредоточены на ответственной разработке ИИ, балансируя между этикой и технологическим прогрессом. В сингапурской модели ИИ-управления повышенное внимание уделяется подотчетности и прозрачности, в то время как в Японии принципы ИИ, ориентированные на человека, делают акцент на инклюзивности и этичном использовании. Австралия аналогично придерживается

проактивного подхода в своей системе этики ИИ, уделяя большое внимание прозрачности, справедливости и предотвращению дискриминации, соответствию этическим принципам и содействию инновациям. Эти региональные различия отражают глубинные геополитические предубеждения: Европа отдает приоритет этическим стандартам, США — инновациям и рыночной свободе, а страны Азиатско-Тихоокеанского региона — балансу между государственным контролем и экономическим развитием. Австралийская система, в которой особое внимание уделяется как этическим стандартам, так и инновациям, предлагает золотую середину [Batool A. et al., 2025: 3271].

Проводя анализ национальных стратегий, утвержденных США, Китаем, Индией, Великобританией, Германией и Канадой, а также нормативных актов и кодексов поведения, в доктрине говорят о трех основных подходах в отношении ИИ: это «мягкое право», экспериментальные правовые режимы (ЭПР) и техническое регулирование. Делается вывод, что подавляющее большинство стран, включая Россию, выбрали «мягкое право» (кодексы поведения, декларации), которое обеспечивает гибкое регулирование без чрезмерных административных барьеров. При этом отмечается, что экспериментальные правовые режимы имеют решающее значение для валидации приложений ИИ, позволяя тестировать технологии в контролируемой среде [Buriaga V.O., Djuzhoma V.V., Artemenko E.A., 2025: 50–68] .

По какому бы пути ни развивалось регулирование ИИ, его неизбежность очевидна. Важно понять, как в это регулирование встроить требования к прозрачности алгоритмов. В качестве потенциального средства защиты потребителей–пользователей эксперты активно обсуждают создание нормативной базы для формирования системы наблюдения, мониторинга и предварительного разрешения на использование технологий ИИ. Для принятия обоснованных решений о безопасности и надежности продукции эти механизмы требуют анализа конфиденциальной информации поставщиков, например, программного обеспечения и наборов данных для обучения технологии ИИ.

По этой причине неизбежно обозначается противоречие между коммерческими интересами поставщиков в сохранении тайны их технологий, с одной стороны, и общественными интересами в справедливой, прозрачной и подотчетной нормативной среде, с другой. Очевидно, что демократический контроль над государственными решениями может осуществляться только через доступ общественности к документации, хранящейся в государственных органах [Spina Alì G., Yu R., 2021: 5–6]. В то же время в частноправовых отно-

шениях существует больше возможностей договорного обеспечения конфиденциальности информации, не препятствующего общему режиму прозрачности.

Другая проблема — контроль над содержанием алгоритмов ИИ, чтобы предотвратить риски и ошибки, которым подвержено его использование. Примеров ошибок, совершаемых ИИ в разных сферах, накопилось уже достаточно — доказано, что такие системы втрое чаще ошибаются при отказе в государственных пособиях гражданам, имеющим на это право, или при принудительном взыскании незаконных долгов с налогоплательщиков. Причина — плохое проектирование алгоритмов. ИИ не только далеко не безошибочен, он способен обучаться и изменяться без ведома программистов (т.е. отличается непредсказуемостью). Причем ошибки ИИ гораздо опаснее, чем человеческие, — это массовость, дискриминация, самовоспроизводство предубеждений и стереотипов, наконец, в большей степени неконтролируемость и невозможность обжалования решений ИИ.

Многочисленные доказательства того, что инструменты на основе ИИ далеки от совершенства в вопросах прозрачности, подотчетности, манипулирования, публичности и справедливости алгоритмических решений [Han S.J., 2025: 2], побуждают некоторых специалистов к описанию негативных последствий деятельности ИИ как «оружия математического разрушения» [O'Neil C., 2016]. Автор этого термина О'Нил описывает ИИ как абстрактные компьютерные модели, которые используются для вынесения суждений на основе несовершенных статистических моделей, часто с катастрофическими последствиями для тех, на кого эти суждения влияют. Логика, лежащая в основе оружия математического разрушения, непрозрачна для всех, кроме программиста, пишущего алгоритм, и специалиста по изучению данных, создающего модель. Это оружие — самоисполняющиеся пророчества, лишенные значимого способа учиться на ошибках, которые они допускают в моделировании мира. Вместо этого они создают петли обратной связи, которые парадоксальным образом делают их неверные прогнозы неизбежными. Алгоритмы используются: для отказа в кредите кредитоспособным лицам; в таргетировании коммерческими колледжами рекламы для охвата групп риска; они создают статистическую путаницу корреляции с причинно-следственными связями в алгоритмах, применяемых страховой индустрией. Кроме того, вызывает беспокойство роль алгоритмов в конструировании нашей гражданской жизни с помощью социальных сетей. В книге О'Нила показано, как алгоритмы наказывают, в частности, низкодоходных граждан, а также пагубно влияют на общество в целом.

Приведенное описание — крайность, во всяком случае, хочется так думать. Человечество надеется «обуздать» алгоритмы именно путем регулирования их разработки и использования. Обычно сбои зависят либо от недостатков в разработке алгоритмов, либо от наборов обучающих данных. В отличие от традиционных вычислений алгоритмы ИИ обладают способностью к самомодификации на основе опыта, подобно биологическому мозгу, совершенствуясь со временем. Это происходит с помощью вычислительных техник, таких как обратное распространение, которое позволяет алгоритму вернуться от нежелательного результата к истокам ошибки и улучшить процесс с этого момента.

Кроме того, в процессе использования ИИ необходимо тщательно фиксировать и анализировать любые отклонения и ошибки. В частности, уже создаются инструменты для оценки рисков применения систем ИИ. Например, Лаборатория компьютерных наук и искусственного интеллекта Массачусетского технологического института опубликовала репозиторий рисков ИИ[3], который состоит из трех частей: база данных, охватывающая более 700 выявленных рисков; аналитический инструмент, объясняющий, как, когда и почему возникают эти риски; подсистему, которая классифицирует эти риски. В данном репозитории выделено семь категорий рисков: 1) дискриминация и токсичность; 2) защита персональных данных и безопасность; 3) дезинформация; 4) злоупотребление технологией; 5) взаимодействие человека и компьютера; 6) социально-экономический и экологический ущерб; 7) безопасность системы ИИ, сбои и ограничения.

Схема регулирования использования ИИ довольно стандартна. Как любое регулирование, оно будет состоять из двух этапов. Первый — установление новых правовых стандартов, в которые закладывается соблюдение прав человека, требования к обучающим данным и др. Второй этап связан с соблюдением стандартов, что включает механизмы прозрачности и доступности для обеспечения соответствия поставщиков ИИ. Речь идет об ответственности автоматизированных процессов принятия решений, возможности предотвратить потенциально вредные действия и исправить любой источник неравноправного, незаконного или нежелательного поведения.

Уместно привести в качестве иллюстрации одну из попыток глобального исследования, посвященного анализу первой волны алго-

[3] AI Risk Repository. Available at: URL: https://airisk.mit.edu/ (дата обращения: 10.08.2025)

ритмической подотчетности в государственном секторе. Его провели совместно Институт Ады Лавлейс (Ada Lovelace Institute), Институт искусственного интеллекта (AI Now Institut) и партнерство Открытое правительство (Open Government Partnership)[4]. Исследователи структурируют данный отчет через шесть ключевых уроков:

Четкие институциональные стимулы и обязательная правовая база могут поддержать последовательное и полезное внедрение механизмов подотчетности, подкрепленное репутационным давлением, благодаря освещению в СМИ и активности гражданского общества.

Алгоритмические политики подотчетности должны четко определять объекты управления, а также установить общую терминологию для всех государственных ведомств.

Установление соответствующей сферы применения политики способствует ее принятию. Подходы к определению сферы применения, распределение по уровням риска, должны быть усовершенствованы, чтобы предотвратить недостаточность охвата или же его чрезмерность.

Механизмы политики прозрачности должны быть подробными и соответствовать аудитории, чтобы обеспечить подотчетность.

Участие общественности поддерживает политику, отвечающую потребностям затронутых сообществ. В политике участие общественности должно быть приоритетной целью, подкрепленной соответствующими ресурсами и официальными стратегиями.

Политика выигрывает от институциональной координации между секторами и уровнями управления для последовательности в применении и использовании разнообразного опыта.

В цитируемом исследовании термин «политика алгоритмической подотчетности» используется для обозначения совокупности мер, направленных на то, чтобы лица, создающие, покупающие и использующие алгоритмы, в конечном счете несли ответственность за их воздействие. Создатели типологии предложили восемь различных политических механизмов, используемых для алгоритмической подотчетности, которые различаются в соответствии с их целями, требованиями и предположениями: принципы и руководства; запреты и моратории; общественная прозрачность; оценка воздействия; аудиты и инспекции регулирующих органов; внешние/независимые надзорные органы; право на слушания и обжалование; условия закупок.

[4] Ada Lovelace Institute, AI Now Institute and Open Government Partnership Algorithmic Accountability for the Public Sector. Available at: URL: https://www.opengovpartnership.org/documents/algorithmic-accountability-public-sector/(дата обращения: 10.08.2025)

Приведенная стратегия рассчитана на универсальное применение. Однако у ее истоков — решение частноправовых проблем. Так, обеспечение прозрачности и подотчетности имеет решающее значение в сфере генеративного ИИ в финтехе для укрепления доверия, снижения рисков и поддержания моральных стандартов. Имеет смысл кратко привести стратегию повышения прозрачности в этой сфере. Заинтересованным сторонам в финтех-сфере, основанной на ИИ, для повышения прозрачности необходимо предпринимать проактивные шаги, поощряющие ответственность, открытость и доверие. Это включает:

а) объяснимость: создание понятных и интерпретируемых алгоритмов ИИ, чтобы заинтересованные стороны могли понимать обоснование решений, принятых с использованием ИИ;

б) раскрытие информации: включение подробностей об источниках данных, входных данных модели и критериях принятия решений в четкое и полное раскрытие информации об использовании ИИ в финтех-товарах и услугах;

в) аудит и валидация: регулярное проведение независимых оценок и аудитов третьими сторонами для оценки результативности, беспристрастности и надежности алгоритмов ИИ;

г) взаимодействие с заинтересованными сторонами: получение отзывов, решение проблем и укрепление доверия к финансовым приложениям на базе ИИ со стороны клиентов, регулирующих органов и других заинтересованных сторон [Saleem M. et al., 2025: 67].

Как можно видеть, предлагаемые подходы к регулированию ИИ и обеспечению прозрачности довольно схожи в разных источниках.

3. Прозрачность как защита от предвзятости и ошибок

Прозрачность имеет решающее значение для защиты от системных предвзятостей, особенно в сферах, где справедливое распределение ресурсов играет первостепенную роль. По сути такая гипотеза автоматически повышает степень прозрачности в государственном управлении, а прозрачность принятия решений ИИ позволяет заинтересованным сторонам выявлять и устранять потенциальные предвзятости.

Можно выделить три основных аспекта, обеспечивающие эффективную прозрачность ИИ. Во-первых, пользователи должны понимать, как он принимает решения, включая уровни уверенности и ограничения. ИИ должен усиливать, а не заменять человеческое суждение. Во-вторых, хотя системы ИИ часто функционируют как

«черные ящики», применяющие его организации должны находить баланс между прозрачностью и безопасностью данных. Объяснимый ИИ (XAI) предлагает решения посредством понятных человеку объяснений. В-третьих, для внедрения требуются общеотраслевые руководящие принципы прозрачности, человеческий контроль, сотрудничество между разработчиками ИИ, специалистами по этике и отраслевыми экспертами, а также информирование о возможностях и ограничениях системы. Зависимость здесь простая — прозрачные системы ИИ укрепляют доверие пользователей.

Часто подчеркиваются два аспекта принципа прозрачности, а именно — объяснимость и интерпретируемость. Термин «объяснимость» относится к способности давать при условии технической осуществимости и в рамках общепризнанного уровня развития понятные объяснения того, почему система ИИ поставляет информацию, производит прогнозы, контент, рекомендации или решения. Это особенно важно в таких чувствительных сферах, как здравоохранение, финансы, иммиграция, пограничные службы и уголовное правосудие (здесь понимание обоснованности решений, принятых с помощью системы ИИ, имеет жизненное значение). В таких случаях прозрачность может, например, существовать в форме списка факторов, которые система ИИ принимает во внимание при информировании или принятии решения.

Другим важным аспектом прозрачности является интерпретируемость, которая относится к способности понимать, как система ИИ создает прогнозы или решения. Другими словами, результаты работы систем ИИ могут быть сделаны доступными и понятными как экспертам, так и неспециалистам. Это включает понятность и доступность внутренней работы, логики и процессов принятия решений систем ИИ для пользователей, включая разработчиков, заинтересованных лиц и конечных пользователей. При этом важно, что прозрачность в контексте систем ИИ имеет технологические ограничения — путь к результату системы ИИ не всегда доступен даже тем, кто ее проектирует или внедряет.

Для обеспечения прозрачности исследователями предлагается фреймворк «Прозрачность через проектирование» (решение проблем на протяжении всей разработки ИИ) с упором на постоянное взаимодействие с заинтересованными сторонами и организационную открытость [Visave J., 2025: 3967–3980]. На этой базе сформулированы ключевые принципы прозрачной разработки ИИ, фокусирующиеся на контекстуальной релевантности и интересах сторон на протяжении всего процесса разработки и эксплуатации системы ИИ.

Обеспечивать прозрачность предлагается с нуля, начиная с этапа разработки алгоритма. Прозрачность должна быть основным принципом проектирования, а не модификацией. На таком начальном этапе приоритет отдается проактивному внедрению мер прозрачности, демонстрируя, как каждый компонент системы влияет на принятие решений. Далее, на этапе обработки и анализа данных нужно «освещать» операции ИИ. На данном этапе основное внимание уделяется формированию у заинтересованных сторон ясного понимания операций ИИ, руководствуясь четырьмя принципами:

подотчетность (четко определить и объяснить процессы принятия решений всем заинтересованным сторонам);

адаптированная коммуникация (адаптировать объяснения как для технической, так и для нетехнической аудитории);

стандарты принятия решений (разъяснить критерии принятия решений и их обоснование);

управление рисками (открыто раскрывать потенциальные системные риски, предубеждения и стратегии смягчения).

Затем на этапе организационного управления предстоит поддерживать прозрачность и взаимодействие. Здесь повышенное внимание уделяется постоянной прозрачности и взаимодействию с заинтересованными сторонами посредством надежных возможностей системного аудита и инспекций, своевременных ответов на запросы заинтересованных сторон, регулярной отчетности об эксплуатационных показателях и воздействии [Visave J., 2025: 3970].

Таким образом, для прочного доверия и подотчетности организации должны использовать комплексную систему прозрачности, выходящую за рамки простого раскрытия данных. Эта система требует четких алгоритмических объяснений, осмысленного человеческого контроля и протоколов систематической оценки. Основные элементы включают подробные методы документирования, внедрение объяснимых систем ИИ (XAI), надежные механизмы контроля и структурированные каналы коммуникации. Подобный механизм применим на микроуровне (в организации) при проектировании алгоритмов определенной отрасли и под определенные задачи. Что же касается нормативных решений более общего порядка, то это задача государства.

В целом специалисты прогнозируют смягчающую роль прозрачности алгоритмов ИИ [Park K., Yoon Ho Y., 2025: 1160] в вопросах доверия к нему. На уровне бизнеса прозрачность часто ассоциируется с раскрытием организационной информации общественности. Прозрачность в системах ИИ действует на двух отдельных, но вза-

имосвязанных уровнях: это техническая прозрачность, касающаяся алгоритмического функционирования, и организационная прозрачность, касающаяся корпоративной подотчетности. Эта двухуровневая структура разъясняет, как различные типы прозрачности влияют на восприятие пользователей через отдельные, но взаимодополняющие пути.

Основанная на информатике прозрачность алгоритмов ИИ — это техническая прозрачность, относящаяся к ясности и понятности механизмов и процессов, управляющих работой чат-бота и его взаимодействием с пользователями. Техническая прозрачность подчеркивает объяснимость механизмов работы систем ИИ, включающую: раскрытие источников данных для обучения и методологий, объяснение процессов принятия решений и разъяснение ограничений системы. Такая техническая прозрачность непосредственно решает проблему «черного ящика», позволяя пользователям развивать доверие к ИИ посредством понимания работы системы. Когда пользователи понимают, как алгоритмы обрабатывают входные данные и генерируют выходные данные, они могут формировать более точные ментальные модели возможностей системы, уменьшая неприятие, вызванное неопределенностью.

Организационная прозрачность — уровень, отражающий корпоративную практику раскрытия информации о разработке и внедрении ИИ. Сюда входят: этические принципы, регулирующие использование ИИ, меры ответственности за системные ошибки и реагирование на опасения заинтересованных сторон. В отличие от прямого влияния технической прозрачности на процесс, организационная прозрачность действует главным образом через механизм призмы, формируя восприятие корпоративной репутации, которое косвенно влияет на доверие к продукту. Например, отчеты Google о прозрачности демонстрируют организационную подотчетность, не требуя от пользователей понимания технических деталей.

Взаимодействие между этими измерениями прозрачности имеет существенное значение при изучении формирования доверия. Техническая прозрачность обеспечивает центральный маршрут, поставляя оценочную информацию о системных операциях, в то время как организационная прозрачность облегчает периферийный маршрут через сигналы корпоративной репутации. Это объясняет, почему для всестороннего построения доверия требуется и то, и другое: одна лишь техническая прозрачность не может компенсировать слабой корпоративной подотчетности (что видно, когда технически сложные системы сталкиваются с общественным недовольством из-за этиче-

ских проблем), а организационная прозрачность не может преодолеть фундаментального непонимания возможностей системы.

Разнообразие подходов и решений проблемы прозрачности алгоритмов свидетельствует как о многомерности данной категории, так и о возможном сочетании разных механизмов прозрачности в зависимости от условий [Kabytov P.P., Nazarov N.A., 2025: 171, 180]. Например, можно различать прозрачность в зависимости от цели, группируя элементы, подпадающие под требование прозрачности и объяснимости, по двум основным аспектам: прозрачность и объяснимость процесса принятия решений (алгоритма) подразумевает раскрытие информации о самой системе, ее архитектуре, логике функционирования и используемых данных (например, какие факторы и как система учитывает при принятии решений); прозрачность и объяснимость результата (решения) фокусируется на сообщении информации, обосновывающей решение, принятое системой в отношении субъекта или ситуации (например, почему в данном случае было принято данное решение и какие данные о субъекте на него повлияли).

Кроме того, можно сконцентрироваться на времени объяснения, различая механизмы прозрачности в зависимости от этапа жизненного цикла систем автоматизированного принятия решений и ИИ, на котором они внедряются. В этом смысле предлагается различать механизмы *ex ante* (реализуются до принятия автоматизированных решений и независимо от конкретного решения; их цель — предотвратить риски, обеспечить предсказуемость работы системы и информировать общественность и заинтересованных лиц о принципах ее работы и возможных последствиях) и механизмы *ex post* (применяются после принятия автоматизированного решения, особенно если оно затрагивает права и законные интересы субъектов; их цель — обеспечить подотчетность, возможность обжалования, исправление ошибок и анализ эффективности системы для дальнейшего совершенствования) [Kabytov P.P., Nazarov N.A., 2025: 171–174].

Прозрачности алгоритмов придается весомое значение с самых разных сторон — технической, организационной, правовой. Полноценное регулирование предполагает их обобщение на нормативном уровне.

4. К понятию алгоритмического суверенитета?

Масштабные и разноцелевые использования ИИ на государственном уровне требуют формирования ясной политики в этой сфере. Хотя страны по-разному подходят к регулированию ИИ, сохраняются общие проблемы: дезинформация, пробелы в подотчетности

и неконтролируемая централизация полномочий в сфере ИИ. Приведем данные исследования, которое выявило противоречия между прозрачностью, контролем, инновациями и доверием в управлении генеративным ИИ [Badawy W., 2025: 4405-4407]. На основе объединения анализа политических документов, интервью с экспертами и сравнительных глобальных исследований был выявлен ряд пробелов в управлении, в частности, связанных с подрывом доверия, сложностью аудита и геополитической фрагментацией. Был сделан вывод: хотя генеративный ИИ, возможно, и не является первопричиной таких проблем, как дезинформация или институциональное недоверие, он, безусловно, усиливает и ускоряет их. Проведенный опрос показал, что общественное доверие к демократическим институтам неуклонно снижалось за последние шесть лет, что частично совпало с ростом дезинформации, генерируемой ИИ, и непрозрачных практик модерации контента. Опрошенные связывали это снижение со следующими факторами: распространение фейкового контента во время выборов; целенаправленная дезинформация с использованием рекомендательных систем, оптимизированных для ИИ; непонимание в обществе, какой контент создан человеком, а какой — машиной.

Хотя ИИ не является единственной причиной констатируемого снижения доверия, он действует как усилитель в уже существующей хрупкой медиасреде. В качестве решения автором исследования предложена многоуровневая структура управления для достижения алгоритмического суверенитета. Новое понятие «алгоритмический суверенитет» относится по сути к способности демократического государства управлять разработкой, развертыванием и воздействием систем ИИ в соответствии с собственными правовыми, культурными и этическими нормами.

Понятие алгоритмического суверенитета определяется не только как технологическая самодостаточность, но и как способность демократического общества формировать траекторию развития и использования ИИ этичными, ответственными и прозрачными способами. Это выходит за рамки традиционного цифрового суверенитета, который фокусируется главным образом на локализации данных или технологической независимости. Вместо этого он требует следующих компонентов: права на аудит сложных моделей ИИ; права вмешиваться в автоматизированные решения; институциональной компетенции регулировать, не подрывая инновации.

Важно, что алгоритмический суверенитет заключается не в контроле кода как такового, а в формировании того, как алгоритмические системы работают в общественной жизни, включая то, кого

они наделяют полномочиями, а кого исключают, и насколько прозрачной остается их логика. В совокупности результаты цитируемого исследования показывают, что достижение алгоритмического суверенитета требует согласования трех противоречий: инновации против регулирования, непрозрачность против подотчетности, государственный контроль против гражданской автономии.

Многоуровневая структура управления, на основе которой должна разрабатываться надежная стратегия управления ИИ, включает три уровня. Первый уровень — нормативная инфраструктура. К числу основных задач, решаемых на первом уровне, относятся: необходимость обязательных классификации рисков и требований к аудиту (в стиле Евросоюза); определить области применения ИИ, требующие участия человека (например, здравоохранение, правоохранительные органы); создать независимые органы надзора за ИИ с полномочиями по обеспечению соблюдения.

Второй уровень управления создает этические и гражданские стандарты. К числу решаемых задач относятся: содействие повышению уровня грамотности в области ИИ в школах, журналистике и государственных структурах; введение обязательной прозрачной маркировки контента, создаваемого ИИ; стимулирование разработки инклюзивных моделей, чтобы избежать углубления социального неравенства.

Третий уровень управления связан с выстраиванием международного сотрудничества. На этом уровне ставятся такие трудные задачи, как заключение и выполнение многосторонних соглашений об управлении ИИ, ориентированных на прозрачность и подотчетность; гармонизация стандартов ИИ между странами, чтобы избежать регуляторного арбитража; финансирование некоммерческих исследований в области ИИ с открытым исходным кодом во всем мире, чтобы сбалансировать корпоративную власть.

Описанная многоуровневая структура управления раскрывает, как нормативная инфраструктура, этические нормы и международное сотрудничество могут работать в тандеме для достижения алгоритмического суверенитета. Это видение требует большего, чем законодательства — оно требует участия в контроле, цифровой грамотности и переосмысления глобального управления в эпоху ИИ. В конечном счете алгоритмический суверенитет — это защита демократических перспектив в мире, который все больше формируется машинами. Необходимо, чтобы невидимые архитектуры цифровой власти оставались видимыми, оспариваемыми и подотчетными лицам, на которых они влияют [Badawy W., 2025: 4406]. В этом смысле многоуровневая модель не заменяет национальной автономии,

а укрепляет ее, встраивая гибкость, демократическое участие и трансграничное согласование в управление ИИ. В таком понимании алгоритмический суверенитет — не возврат к контролю сверху вниз, а демократическое обновление — способ гарантировать, что системы ИИ поддерживают ценности, права и потребности обществ, в которых они функционируют.

Несмотря на очевидные плюсы, рассмотренная модель, похоже, оптимистически рассчитана на гармоничное сосуществование государств, приводящее к не менее гармоничному сосуществованию человечества и ИИ. Однако мир более разнообразен и несговорчив, доказательством чему является отсутствие значимых международных договоренностей в области регулирования и использования ИИ. Кроме того, значение фактора глобализации все ощутимее уменьшается, а диапазон универсальных регуляторов сужается.

5. Универсальность этики искусственного интеллекта

Практически все проблемы и их исследования, рассмотренные выше, базировались на мысли, что ИИ — технически универсальное средство, требующее одинаковых мер регулирования. В этом смысле универсальность технологий, казалось бы, обусловливает универсальность регулирования. Но это не совсем так, поскольку в мире уже оформились разные подходы к регулированию ИИ — от либерального до жесткого. Обеспечение прозрачности алгоритмов ИИ — это направление общей политики управления в области ИИ, важнейшей частью которой является этика ИИ. Можно ли утверждать, что этика ИИ универсальна?

В этой области тоже намечаются трудности. Обратимся к любопытному африканскому примеру. В литературе, исследующей применение ИИ на Африканском континенте, активно обосновываются утверждения о надвигающейся «цифровой» или «алгоритмической» колонизации Африки из-за отсутствия африканских ценностей в этике ИИ [Adams R., 2021: 190]; [Birhane A., 2023: 251]. В качестве одного из способов урегулирования проблемы все больше исследователей подчеркивает необходимость утверждения «африканских ценностей» в этике ИИ для устранения «эпистемической несправедливости» [Mohamed S., 2020: 672].

В качестве жизнеспособной дополнительной африканской этической основы этики ИИ часто предлагается концепция «Убунту» (слово «убунту» толкуется как человечность по отношению к другим, а также вера во вселенские узы общности, связывающие все

человечество). Концепция является общеафриканской этической системой, и ее нормативные структуры ясны, чтобы послужить основой этики ИИ и, следовательно, решать проблему «эпистемической несправедливости» в этике ИИ.

Реляционная природа «Убунту» выступает в двух формах. Одна из них в том, что, согласно «Убунту», чтобы человек полностью стал личностью, его позитивные отношения с другими должны стать основополагающими. В основе концепции лежит распространенная поговорка «человек является человеком через других», а не индивидуально как автономное существо. Личность постигается через межличностные и общинные отношения, а не через индивидуалистические, рациональные и атомистические усилия. «Убунту» уделяет значительное внимание роли общества в формировании индивидуальной идентичности и определении личности. Это делает концепцию общинной этикой. Такое понимание контрастирует с западной философией, где индивидуальная автономия, рациональность и благоразумие считаются решающими для хорошей жизни или личности.

Кроме того, реляционность «Убунту» заключается в понятии о том, что человек должен активно стремиться к гармонии с сообществом, чтобы считаться личностью. Как форма относительной автономии, индивидуум в концепции рассматривается как социально обусловленная и встроенная в социальную среду единица. Западная же философия подчеркивает центральное значение взаимного уважения прав среди других добродетелей для мирного сосуществования членов сообщества. В отличие от «Убунту», это пассивный подход, при котором личность человека зависит не от активного поиска социальной гармонии, а от способности руководить своими действиями или жизнью в соответствии с принятыми стандартами. В этом смысле западные этические системы основаны на правах, тогда как «Убунту» больше фокусируется на обязанностях.

В современном мире технологии ИИ разрабатываются и управляются на основе ценностей, приоритетных в западных обществах — таких, как индивидуальная автономия. В контексте ИИ игнорирование незападных систем ценностей и знаний при разработке систем ИИ, а также структур нормативного управления можно рассматривать как форму эпистемической несправедливости [Nihei M., 2022: 42]. Исследователи все чаще предупреждают, что эпистемическая несправедливость может привести к «цифровой колонизации» континента. Такая форма колонизации, также называемая «этическим колониализмом», следует исключительной ориентации на евро-атлантическую моральную систему, которая ставит в центр рациональ-

ных личностей. Конечный результат такой формы эпистемической несправедливости в том, что она приведет к состоянию, когда технологии ИИ развиваются таким образом, что увековечивают маргинализацию пользователей из стран глобального Юга.

Однако, несмотря на популярность утверждений, что понятия о современной этике ИИ не учитывают этических принципов, присущих незападным культурам, включая Африку, остается неясным, как общая концепция «Убунту» будет трансформирована в руководящий принцип этики ИИ. Например, неясно, как разработчик ИИ должен перевести концепцию в дизайн, чтобы сделать технологию более социально конституированной (что, впрочем, не исключает полностью перспектив внедрения ценностей, заимствованных из Африки, включая «Убунту», в технологический дизайн) [Yilma K., 2025: 3, 5]. Кроме того, смысл «Убунту» различался в африканском пространстве, и в разных обществах юга Африки ему придавались неодинаковые значения. Это свидетельствует об изменчивой природе концепции. Такие проблемы ограничивают ее роль в предложении африканской этической перспективы ИИ.

В последние годы в Африке появляется множество инициатив в области этики ИИ, в основном в форме национальных и континентальных стратегий ИИ, в которые в той или иной степени встраиваются «африканские ценности». Например, в обзоре Африканской группы высокого уровня новых технологий (2023) подчеркнута необходимость базировать этику ИИ на «африканских ценностях». Стратегия развития ИИ в Бенине также упоминает об «Убунту». Это пока не создает последовательной и практичной основы «африканской этики ИИ» [Yilma K., 2025; 11], но само направление демонстрирует, насколько глубоки назревающие противоречия даже в вопросах этики ИИ, имеющей, казалось бы, универсальную основу.

Заключение

Использование ИИ в современном обществе — источник проблем, размышлений, прогнозов, дискуссий. Классическое право стремится рассматривать ИИ в качестве объекта регулирования, предъявляя к нему ряд требований, исходящих из определенной «подчиненности» человеку. В действительности неизученных рисков и проблем еще больше, нежели обнародованных. Одним из всерьез анализируемых в науке прогнозов является возможность войны между человечеством и ИИ. Автор прогноза приходит к выводу о вероятности информационных сбоев и проблем с обязательствами

в конфликте между ИИ и человеком. Информационные сбои будут обусловлены затрудненностью измерения возможностей ИИ, неинтерпретируемостью систем ИИ и различиями в том, как ИИ и человек анализируют информацию. Проблемы с обязательствами затруднят для ИИ и людей заключение надежных сделок. Вероятность войны предлагается уменьшить, улучшив измерение возможностей ИИ, ограничив развитие его возможностей, разработав системы ИИ, аналогичные человеческим[5].

Понимание функционирования алгоритмов ИИ выглядит неплохим способом для думающего и ответственного человека принять решение об использовании ИИ. Главное, чтобы у человека сохранялось право выбора — использовать ИИ или нет. Поэтому обеспечение прозрачности алгоритмов ИИ, являющееся частью общей политики управления в области ИИ, является важной задачей ближайшего периода.

Список источников

1. Талер Р. Новая поведенческая экономика. Почему люди нарушают правила традиционной экономики и как на этом заработать. М.: Эксмо, 2018. 367 с.

2. Adams R. Can artificial intelligence be decolonized? Interdisciplinary Science Reviews, 2021, no 46, pp. 176–197.

3. Badawy W. Algorithmic sovereignty and democratic resilience: rethinking AI governance in the age of generative AI. AI and Ethics. Springer. 2025, vol. 5, pp. 4402–4410.

4. Batool A. et al. AI governance: a systematic literature review. AI and Ethics, 2025, vol. 5, pp. 3265–3279.

5. Birhane A. Algorithmic colonization of Africa. In: S. Cave S. & K. Dihal (eds.). Imagining AI: How the world sees intelligent machines. Oxford: University Press, 2023, pp. 247–260.

6. Buriaga V.O., Djuzhoma V.V., Artemenko E.A. Shaping an Artificial Intelligence Regulatory Model: International and Domestic Experience. Legal Issues in the Digital Age, 2025, vol. 6, no. 2, pp. 50–68.

7. Goldstein S. Will AI and humanity go to war? AI & SOCIETY. 2025. Available at: URL: https://link.springer.com/article/10.1007/s00146-025-02460-1 (дата обращения: 10.08.2025)

8. Han S.J. The Question of AI and Democracy: Four Categories of AI Governance. Philosophy & Technology, 2025, no. 38, pp. 1–26.

9. Kabytov P.P., Nazarov N.A. Transparency in Public Administration in the Digital Age: Legal, Institutional and Mechanisms. Legal Issues in the Digital Age, 2025, vol. 6, no. 2, pp. 161–182.

[5] AI & SOCIETY. Available at: URL: https://link.springer.com/article/10.1007/s00146-025-02460-1 (дата обращения: 10.08.2025)

10. Mohamed S., Png M.T., Isaac W. Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. Philosophy & Technology, 2020, no. 33, pp. 659–684.

11. Nihei M. Epistemic injustice as a philosophical conception for considering fairness and diversity in human-centered AI principles. Interdisciplinary Information Sciences, 2022, no 28, pp. 25–43.

12. O'Neil C. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. New York: Crown Publishers, 2016, 272 p.

13. Park K., Yoon Ho Y. AI algorithm transparency, pipelines for trust not prisms: mitigating general negative attitudes and enhancing trust toward AI. Humanities and Social Sciences Communications, 2025, no. 12. Article number 1160.

14. Rachovitsa A., Johann N. The Human Rights Implications of the Use of AI in the Digital Welfare State: Lessons Learned from the Dutch SyRI Case. Human Rights Law Review, 2022, no. 22, pp. 1–15.

15. Saleem M. et al. Responsible AI in Fintech: Addressing Challenges and Strategic Solutions. In: Dutta S. et al. (eds.) Generative AI in FinTech: Revolutionizing Finance Through Intelligent Algorithms. Cham: Springer, 2025, pp. 61–72.

16. Spina Alì G., Yu R. Artificial Intelligence between Transparency and Secrecy: From the EC Whitepaper to the AIA and Beyond. European Journal of Law and Technology, 2021, vol. 12, no. 3, pp. 1–25.

17. Sposini L. The governance of algorithms: profiling and personalization of online content in the context of European Consumer Law. Nordic journal of European Law, 2024, vol. 7, no 1, pp. 1–22.

18. Visave J. Transparency in AI for emergency management: building trust and accountability. AI and Ethics, 2025, no 5, pp. 3967–3980.

19. Yilma K. Ethics of AI in Africa: Interrogating the role of Ubuntu and AI governance initiatives. Ethics and Information Technology, 2025, vol. 27, article no 24. P. 1–14.

References

1. Adams R. (2021) Can artificial intelligence be decolonized? *Interdisciplinary Science Reviews*, vol. 46, pp. 176–197.

2. Badawy W. (2025) Algorithmic sovereignty and democratic resilience: rethinking AI governance in the age of generative AI. *AI and Ethics*, vol. 5, pp. 4402-4410.

3. Batool A. et al. (2025) AI governance: a systematic literature review. *AI and Ethics*, vol. 5, pp. 3265–3279.

4. Birhane A. (2023) Algorithmic colonization of Africa. In: S. Cave S., K. Dihal (eds.). Imagining AI: How the world sees intelligent machines. Oxford: University Press, pp. 247–260.

5. Buriaga V.O., Djuzhoma V.V., Artemenko E.A. (2025) Shaping an artificial intelligence regulatory model: international and domestic Experience. *Legal Issues in the Digital Age*, vol. 6, no. 2, pp. 50–68

6. Goldstein S. (2025) Will AI and humanity go to war? AI & Society. Accepted: 22 June. Available at: https://doi.org/10.1007/s00146-025-02460-1

7. Han S. (2025) The question of AI and democracy: four categories of AI governance. *Philosophy & Technology*, vol. 38, pp. 1–26.

8. Kabytov P.P., Nazarov N.A. (2025) Transparency in public administration in the digital age: legal, institutional and mechanisms. *Legal Issues in the Digital Age,* vol. 6, no. 2, pp. 161–182

9. Mohamed S. et al. (2020) Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology*, vol. 33, pp. 659–684.

10. Nihei M. (2022) Epistemic injustice as a philosophical conception for considering fairness and diversity in human-centered AI principles. *Interdisciplinary Information Sciences*, vol. 28, pp. 25–43.

11. O'Neil C. (2016) *Weapons of math destruction: how big data increases inequality and threatens democracy*. New York: Crown Publishers, 272 p.

12. Park K., Yoon Ho Y. (2025) AI algorithm transparency, pipelines for trust not prisms: mitigating general negative attitudes and enhancing trust toward AI. *Humanities and Social Sciences Communities*, no. 12.

13. Rachovitsa A., Johann N. (2022) The human rights implications of the use of AI in the digital welfare state: lessons learned from the dutch SyRI case. *Human Rights Law Review*, no. 22, pp. 1–15.

14. Saleem M. et al. (2025) Responsible AI in fintech: addressing challenges and strategic solutions. In: S. Dutta et al. Generative AI in FinTech: Revolutionizing Finance through Intelligent Algorithms. Cham: Springer, pp. 61–72.

15. Spina Alì G., Yu R. (2021) Artificial Intelligence between transparency and secrecy: from the EC Whitepaper to the AIA and beyond. *European Journal of Law and Technology*, vol. 12, no. 3, pp. 1–25.

16. Sposini L. (2024) The governance of algorithms: profiling and personalization of online content in the context of European Consumer Law. *Nordic Journal of European Law,* vol. 7, no. 1, pp. 1–22.

17. Thaler R. (2018) *New behavioral economics. Why people break the rules of traditional economics and how to make money on it.* Moscow: Eksmo, 367 p. (in Russ.)

18. Visave J. (2025) Transparency in AI for emergency management: building trust and accountability. *AI and Ethics*, no. 5, pp. 3967–3980.

19. Yilma K. (2025) Ethics of AI in Africa: interrogating the role of Ubuntu and AI governance initiatives. *Ethics and Information Technology*, vol. 27, pp. 1–14.

Информация об авторе:

Э.В. Талапина– доктор юридических наук, главный научный сотрудник.

Information about the author:

E.V. Talapina — Doctor of Sciences (Law), Chief Researcher.

Статья поступила в редакцию 31.07.2025; одобрена после рецензирования 11.08.2025; принята к публикации 18.08.2025.

The article was submitted to editorial office 31.07.2025; approved after reviewing 11.08.2025; accepted for publication 18.08.2025.