



**ТЕМА АНАЛИТИЧЕСКОГО МАТЕРИАЛА:
ПРОБЛЕМА "ЧЕРНОГО ЯЩИКА" В ИИ И ПРАВО НА ОБЪЯСНЕНИЕ РЕШЕНИЙ**

Эркинбаева Жасмина Фарход кизи

Выполнила: студентка 2-го курса, Факультета МПСР По

модулю: ИИ, робототехника и право

АННОТАЦИЯ

В данной статье исследуется одна из наиболее актуальных проблем современного права и информационных технологий — феномен «чёрного ящика» в системах искусственного интеллекта и связанное с ним право граждан на получение объяснений автоматизированных решений. Методологической основой работы выступает сравнительно-правовой анализ международного опыта регулирования прозрачности алгоритмических систем, включая Общий регламент защиты данных Европейского союза (GDPR), Акт ЕС об ИИ 2024 года, а также национальное законодательство Республики Узбекистан, в частности Указ Президента № УП-189 от 22 октября 2025 года. Результаты исследования свидетельствуют о необходимости комплексного подхода, сочетающего правовые, технические и этические инструменты обеспечения алгоритмической прозрачности.

Ключевые слова: искусственный интеллект, чёрный ящик, право на объяснение, GDPR, Акт ЕС об ИИ, объяснимый ИИ, XAI, LIME, SHAP, алгоритмическая прозрачность, автоматизированное принятие решений, этика ИИ, алгоритмическая дискриминация

**TITLE OF THE ANALYTICAL PAPER:
THE “BLACK BOX” PROBLEM IN ARTIFICIAL INTELLIGENCE AND THE RIGHT
TO EXPLANATION OF DECISIONS**

Erkinbaeva Jasmina Farhod qizi

2nd-year student, Faculty of MPSP

Module: Artificial Intelligence, Robotics, and Law

ABSTRACT

This article examines one of the most pressing issues in modern law and information technology—the “black box” phenomenon in artificial intelligence systems and the associated right of individuals to obtain explanations of automated decisions.

The methodological basis of the study is a comparative legal analysis of international approaches to regulating the transparency of algorithmic systems, including the General Data Protection Regulation (GDPR) of the European Union, the EU Artificial Intelligence Act of 2024, as well as the national legislation of the Republic of Uzbekistan, in particular Presidential Decree No. UP-189 dated October 22, 2025.

The results of the study indicate the need for a comprehensive approach that combines legal, technical, and ethical tools to ensure algorithmic transparency.

Keywords: artificial intelligence, black box, right to explanation, GDPR, EU Artificial Intelligence Act, explainable AI, XAI, LIME, SHAP, algorithmic transparency, automated decision-making, AI ethics, algorithmic discrimination

ВВЕДЕНИЕ

Стремительное распространение технологий искусственного интеллекта во всех сферах общественной жизни — от медицинской диагностики и банковского кредитования до уголовного правосудия и государственного управления — ставит перед юридической наукой принципиально новые вопросы. Одним из центральных среди них является проблема так называемого «чёрного ящика»: ситуации, при которой алгоритм машинного обучения



принимает решение, затрагивающее права и законные интересы человека, однако логика этого решения остаётся непрозрачной как для самого субъекта, так и для оператора системы.

Проблема непрозрачности алгоритмов приобретает особую остроту в контексте глубокого обучения (deep learning), где нейронные сети с миллионами параметров обрабатывают данные способами, которые не поддаются интуитивному человеческому пониманию. Когда такая система отказывает гражданину в кредите, определяет риск рецидива преступления или распределяет медицинские ресурсы, отсутствие возможности получить содержательное объяснение подрывает фундаментальные принципы правового государства — надлежащую правовую процедуру, равенство перед законом и право на эффективное средство правовой защиты.

Мировое сообщество активно ищет правовые ответы на эти вызовы. Европейский союз закрепил элементы права на объяснение в GDPR ещё в 2016 году, а в 2024 году принял первый в мире комплексный Акт об искусственном интеллекте (AI Act), устанавливающий требования прозрачности для систем ИИ высокого риска. Республика Узбекистан также предпринимает значительные шаги в данном направлении: Указ Президента № УП-189 от 22 октября 2025 года определяет стратегический курс на развитие технологий ИИ при одновременном обеспечении их безопасности и подотчётности.

Цель настоящего исследования представляет собой: комплексный анализ проблемы «чёрного ящика» в правовом измерении, изучение международного опыта правового регулирования алгоритмической прозрачности и выработка рекомендаций по совершенствованию нормативной базы Узбекистана в данной области.

МЕТОДОЛОГИЯ

Методологическую основу исследования составляет комплекс общенаучных и специально-юридических методов. Сравнительно-правовой метод применяется для сопоставления подходов к правовому регулированию алгоритмической прозрачности в Европейском союзе, Соединённых Штатах Америки, Великобритании и Республике Узбекистан. Формально-юридический метод используется при анализе нормативных текстов — статей 13–15 и 22 GDPR, статей 13, 50 и 86 Акта ЕС об ИИ, а также положений Указа УП-189 и принятых на его основе подзаконных актов.

Системный подход позволяет рассмотреть проблему прозрачности ИИ в единстве её правового, технического и этического измерений. Междисциплинарный характер исследования обусловлен необходимостью привлечения данных информатики — в частности, описания технических методов объяснимого ИИ (Explainable AI, XAI), таких как LIME (Local Interpretable Model-Agnostic Explanations) и SHAP (SHapley Additive exPlanations). Эмпирическую базу исследования формируют: нормативные правовые акты ЕС, Великобритании и Узбекистана; практика Суда Европейского союза, в том числе решение по делу C-203/22 (Dun & Bradstreet Austria); доклады Рабочей группы по защите данных (Article 29 Working Party); научные публикации в области права, компьютерных наук и этики ИИ; а также практические случаи алгоритмической дискриминации (система COMPAS, рекрутинговый алгоритм Amazon, алгоритм распределения медицинских ресурсов Optum).

РЕЗУЛЬТАТЫ

Техническая природа проблемы «чёрного ящика». Анализ показывает, что непрозрачность алгоритмов обусловлена тремя основными барьерами. Во-первых, намеренное сокрытие информации разработчиком в целях защиты коммерческой тайны и интеллектуальной собственности. Во-вторых, недостаток технической грамотности у конечных пользователей и даже у юристов, призванных оценивать правомерность



алгоритмических решений. В-третьих, внутренняя непрозрачность (intrinsic opacity), вытекающая из самой природы методов глубокого обучения, где итоговое решение формируется взаимодействием миллионов числовых весов в скрытых слоях нейронной сети.

Современная наука предлагает два основных направления преодоления этой проблемы. Первое направление — разработка изначально интерпретируемых моделей (inherently interpretable models), таких как деревья решений или линейные регрессионные модели. Второе — создание пост-хок методов объяснения, применяемых к уже обученным непрозрачным моделям. К наиболее распространённым инструментам второго типа относятся метод LIME, который аппроксимирует поведение сложной модели локально понятной линейной моделью в окрестности конкретного предсказания, и метод SHAP, основанный на теории кооперативных игр Шепли и позволяющий оценить вклад каждого признака в итоговое решение как на локальном, так и на глобальном уровне. Вместе с тем исследования показывают, что оба метода имеют существенные ограничения. LIME обеспечивает лишь локальные объяснения и не учитывает нелинейные зависимости между признаками. SHAP, несмотря на более широкие возможности, чувствителен к мультиколлинеарности признаков и может давать различные результаты в зависимости от выбранной модели машинного обучения. Это означает, что полностью «открыть» чёрный ящик на сегодняшний день технически невозможно, и любое правовое регулирование должно учитывать эти объективные ограничения.

1. Международный опыт

Право на объяснение в GDPR. Общий регламент защиты данных ЕС содержит ряд положений, направленных на обеспечение прозрачности автоматизированного принятия решений. Статья 22 устанавливает право субъекта данных не подвергаться решению, основанному исключительно на автоматизированной обработке, если такое решение имеет для него правовые последствия или существенно затрагивает его иным образом. Статьи 13(2)(f), 14(2)(g) и 15(1)(h) обязывают контролёра предоставлять «значимую информацию о логике обработки» (meaningful information about the logic involved).

Долгое время в доктрине велась дискуссия о том, закрепляет ли GDPR подлинное «право на объяснение» или лишь право на информацию общего характера. Решение Суда ЕС по делу C-203/22 (Dun & Bradstreet Austria) внесло определённую ясность: суд подтвердил существование права на объяснение конкретных автоматизированных решений, при этом указал, что раскрытие самого алгоритма не является ни необходимым, ни достаточным средством обеспечения прозрачности. В качестве адекватного инструмента суд признал контрфактические объяснения — указание на то, какие изменения во входных данных привели бы к иному результату.

Акт ЕС об искусственном интеллекте принятый в июне 2024 года Акт об ИИ (Регламент 2024/1689) представляет собой первый в мире комплексный законодательный акт в данной области. Он вводит риск-ориентированный подход, устанавливая наиболее строгие требования для систем ИИ высокого риска. Статья 13 Акта обязывает разработчиков проектировать такие системы с достаточной прозрачностью, чтобы операторы могли надлежащим образом интерпретировать их результаты. Статья 86 закрепляет право лиц, затронутых решениями систем ИИ высокого риска, на получение объяснений. Важно отметить, что статья 50 Акта вводит дополнительные обязательства по прозрачности для определённых категорий систем ИИ: провайдеры обязаны информировать пользователей о взаимодействии с ИИ, а системы, генерирующие синтетический контент, должны маркировать



его соответствующим образом. За нарушение требований Акта предусмотрены штрафы до 35 миллионов евро или 7 процентов глобального годового оборота компании.

2. Случай алгоритмической дискриминации

Практика применения алгоритмических систем уже продемонстрировала серьёзные риски, связанные с их непрозрачностью. Система COMPAS, используемая судами США для оценки риска рецидива, была уличена в системном завышении рисков для афроамериканских подсудимых, что объясняется воспроизведением исторических предрассудков, заложенных в обучающих данных. Рекрутинговый алгоритм компании Amazon систематически дискриминировал женщин-кандидатов, поскольку обучался преимущественно на резюме мужчин.

В сфере здравоохранения исследование 2019 года выявило, что алгоритм компании Optum, применявшийся для распределения медицинских ресурсов, использовал стоимость лечения в качестве косвенного показателя тяжести заболевания. Поскольку афроамериканские пациенты исторически тратили на медицину меньше при сопоставимой тяжести состояния, алгоритм систематически занижал их потребность в медицинской помощи. Все эти случаи объединяет одно обстоятельство: непрозрачность алгоритма препятствовала своевременному обнаружению и устранению дискриминации.

3. Опыт Великобритании

Заслуживает внимания расхождение подходов Великобритании и ЕС после Brexit. Поправки к британскому GDPR, внесённые Актом о данных (Data Use and Access Act 2025), существенно ослабили защиту граждан в сфере автоматизированного принятия решений, введя размытый стандарт «отсутствия значимого человеческого участия» и ограничив ключевые гарантии случаями обработки особых категорий данных. Ряд исследователей квалифицирует эти изменения как создание серьёзных правовых лакун, позволяющих компаниям скрывать непрозрачные модели от контроля.

4. Опыт Республики Узбекистан.

Узбекистан демонстрирует последовательный курс на развитие технологий ИИ в сочетании с формированием соответствующей нормативной базы. Указ Президента № УП-189 от 22 октября 2025 года «О дополнительных мерах по дальнейшему развитию технологий искусственного интеллекта» является ключевым программным документом, определяющим стратегию страны в данной области. Указ предусматривает создание национальной инфраструктуры ИИ, в том числе единой цифровой платформы, объединяющей данные из различных ведомств, формирование Центра развития искусственного интеллекта и цифровой экономики, привлечение высококвалифицированных зарубежных специалистов, а также подготовку методических рекомендаций по безопасному применению ИИ в деятельности государственных органов. Особенно значимым представляется положение Указа УП-189 о присоединении к Международной программе сотрудничества по ИИ Организации экономического сотрудничества и развития (ОЭСР), что свидетельствует о стремлении Узбекистана интегрировать свою нормативную базу в международные стандарты алгоритмической этики и прозрачности. В развитие этой инициативы в марте 2026 года Министерство юстиции зарегистрировало Этические правила создания, внедрения и использования решений на основе ИИ, которые закрепляют принципы законности, защиты прав человека, ответственности и надёжности технологий. Принципиальным нововведением стал запрет на полностью автономное принятие решений системами ИИ в критически значимых сферах, затрагивающих права и здоровье граждан.



Кроме того, в январе 2026 года были приняты поправки в законодательство, устанавливающие административную ответственность за незаконную обработку персональных данных с использованием ИИ в виде штрафов от 50 до 100 базовых расчётных величин. Также Указом Президента № УП-140 от 21 августа 2025 года предусмотрено внедрение технологий ИИ в деятельность судов с одновременным повышением уровня доступа к правосудию, что ставит вопрос о прозрачности алгоритмических решений в судебной системе особенно остро.

ОБСУЖДЕНИЕ

Проведённый сравнительный анализ позволяет выделить несколько ключевых тенденций в правовом регулировании алгоритмической прозрачности. Первая и наиболее очевидная тенденция — переход от мягкого этического регулирования к юридически обязательным нормам. Если Рекомендации ОЭСР по ИИ 2019 года и Этические руководства Экспертной группы высокого уровня ЕС носили рекомендательный характер, то GDPR и особенно Акт ЕС об ИИ подкрепляют требования прозрачности значительными санкциями. Вторая тенденция — осознание невозможности обеспечить абсолютную прозрачность сложных алгоритмических систем. Суд ЕС по делу *Dun & Bradstreet* фактически признал, что раскрытие алгоритма не равнозначно объяснению решения, и предложил контрфактический подход как прагматичную альтернативу. Это решение имеет огромное практическое значение, поскольку снимает нереалистичное ожидание полного раскрытия внутренней архитектуры моделей и направляет правоприменительную практику в сторону функциональных объяснений, понятных обычному гражданину. Третья тенденция связана с необходимостью балансировки конкурирующих интересов. Право на объяснение неизбежно вступает в противоречие с правом на защиту коммерческой тайны разработчика. Рецитал 63 GDPR прямо указывает, что право на доступ к информации не должно нарушать права интеллектуальной собственности. Суд ЕС предлагает решать этот конфликт путём ограниченного раскрытия с применением защитных механизмов — например, процедуры камерального рассмотрения, при которой алгоритм изучается судом без передачи его стороне-заявителю.

Применительно к Узбекистану анализ показывает, что страна находится на этапе формирования базовой нормативной инфраструктуры регулирования ИИ. Указ УП-189 создаёт институциональные предпосылки для внедрения ИИ в государственное управление и экономику, однако вопросы алгоритмической прозрачности и права на объяснение пока не получили детальной проработки в нормативных актах. Этические правила 2026 года делают важный шаг, запрещая полностью автономное принятие решений в критических сферах, но не раскрывают конкретные механизмы реализации права граждан на получение объяснений. Между тем планируемое масштабное внедрение ИИ в судебную систему (не менее 70 процентов решений государственных исполнителей должны приниматься автоматически с 2027 года) делает разработку таких механизмов неотложной задачей. Опыт применения системы COMPAS в США наглядно демонстрирует, что без надлежащих гарантий прозрачности автоматизация правосудия может привести к системной дискриминации уязвимых групп населения. Важно также учитывать, что технические методы объяснимого ИИ, при всей их полезности, не являются панацеей.

Как показывают исследования, LIME и SHAP зависят от выбранной модели и чувствительны к корреляции признаков, что ставит под сомнение их надёжность в качестве единственного инструмента обеспечения прозрачности. Поэтому правовое регулирование должно сочетать требования технической объяснимости с организационными мерами и обязательным человеческим надзором, аудитом алгоритмических систем, оценкой



воздействия на фундаментальные права и процессуальными гарантиями: правом на обжалование автоматизированных решений и доступом к независимой экспертизе.

ЗАКЛЮЧЕНИЕ

Проведённое исследование позволяет сформулировать следующие основные выводы. Проблема «чёрного ящика» в системах искусственного интеллекта представляет собой не только техническое, но и фундаментальное правовое явление, затрагивающее базовые принципы правового государства. Невозможность получить содержательное объяснение автоматизированного решения подрывает право на справедливое разбирательство, создаёт условия для системной дискриминации и ограничивает возможности судебного контроля.

Международный опыт свидетельствует о формировании многоуровневого подхода к решению этой проблемы: На законодательном уровне закрепляются обязательные требования прозрачности с дифференциацией по степени риска. На техническом уровне развиваются методы объяснимого ИИ, обеспечивающие различные формы интерпретируемости — от локальных объяснений отдельных решений до глобального анализа поведения модели. На судебном уровне формируется практика применения контрфактических объяснений как прагматичного стандарта раскрытия. Республика Узбекистан, реализуя амбициозную стратегию цифровизации в рамках Указа УП-189 и Стратегии развития ИИ до 2030 года, нуждается в разработке специального нормативного механизма обеспечения алгоритмической прозрачности.

Исходя из выше перечисленных будет целесообразным рекомендовать: во-первых, законодательное закрепление права граждан на получение содержательного объяснения автоматизированных решений, затрагивающих их права и законные интересы; во-вторых, введение обязательной оценки воздействия систем ИИ высокого риска на фундаментальные права по аналогии с моделью Акта ЕС об ИИ; в-третьих, создание независимого органа по надзору за алгоритмическими системами при Министерстве цифровых технологий. Дальнейшие исследования целесообразно направить на разработку конкретных стандартов объяснимости для различных сфер применения ИИ в Узбекистане, изучение возможностей адаптации европейских регуляторных песочниц в области ИИ к условиям центральноазиатского региона, а также на анализ соотношения алгоритмической прозрачности и защиты персональных данных в контексте национального законодательства.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) Official Journal of the European Union. — L series. — 2024.
2. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data (GDPR) Official Journal of the European Union. — L 119. — 2016.
3. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689#>
4. Указ Президента Республики Узбекистан от 22.10.2025 г. № УП-189 «О дополнительных мерах по дальнейшему развитию технологий искусственного интеллекта» // Национальная база данных законодательства. — 27.10.2025. — № 06/25/189/0978.
5. Постановление Президента Республики Узбекистан от 14.10.2024 г. № ПП-358 «Об утверждении Стратегии развития технологий искусственного интеллекта до 2030 года».



6. Этические правила создания, внедрения и использования решений на основе искусственного интеллекта (утв. приказом Министра цифровых технологий, рег. № 3787 от 14.03.2026 г.).
7. Указ Президента Республики Узбекистан от 21.08.2025 г. № УП-140 «О дополнительных мерах по повышению уровня доступа к правосудию путём внедрения технологий ИИ в деятельность судов».
8. Goodman B., Flaxman S. European Union Regulations on Algorithmic Decision-Making and a «Right to Explanation» AI Magazine. — 2017. — Vol. 38. — No. 3.
9. Wachter S., Mittelstadt B., Floridi L. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation International Data Privacy Law. — 2017. — Vol. 7. — No. 2.
10. Wachter S., Mittelstadt B., Russell C. Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR. Harvard Journal of Law & Technology. — 2018. — Vol. 31. — No. 2.
11. Edwards L., Veale M. Slave to the Algorithm? Why a «Right to an Explanation» Is Probably Not the Remedy You Are Looking For Duke Law & Technology Review. — 2017. — Vol. 16.
12. Kaminski M.E. The Right to Explanation, Explained Berkeley Technology Law Journal. — 2019. — Vol. 34. — No. 1.
13. Brkan M. Do Algorithms Rule the World? Algorithmic Decision-Making in the Framework of the GDPR and Beyond. — 2017.
14. Ribeiro M.T., Singh S., Guestrin C. «Why Should I Trust You?»: Explaining the Predictions of Any Classifier Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. — 2016.
15. Lundberg S., Lee S.-I. A Unified Approach to Interpreting Model Predictions Advances in Neural Information Processing Systems. — 2017.
16. Salih A. et al. A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME Advanced Intelligent Systems. — 2025. — Vol. 7. — 2400304.
17. Panigutti C. et al. The Role of Explainable AI in the Context of the AI Act Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency. — 2023.
18. Papadimitriou E. The Right to Explanation in the Processing of Personal Data with the Use of AI Systems International Journal of Law in Changing World. — 2023.
19. Buruiană (Rusu) A. Black Box AI and the Sovereignty of Personal Data: Between GDPR and Digital Ethics European Journal of Law and Public Administration. — 2025.
20. Vorras A., Mitrou L. Unboxing the Black Box of Artificial Intelligence: Algorithmic Transparency and/or a Right to Functional Explainability EU Internet Law in the Digital Single Market. — Cham: Springer, 2021.
21. Angwin J. et al. Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks ProPublica. — 2016. — 23 May.
22. Dastin J. Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women Reuters. — 2018. — 10 October.
23. Obermeyer Z. et al. Dissecting racial bias in an algorithm used to manage the health of populations Science. — 2019. — Vol. 366. — No. 6464.
24. OECD Recommendation on Artificial Intelligence. OECD/LEGAL/0449. — Paris: OECD Publishing, 2019.
25. High-Level Expert Group on AI. Ethics Guidelines for Trustworthy AI. — Brussels: European Commission, 2019.



26. Molnar C. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. — 2nd ed. — Lulu.com, 2022.
27. Gunning D. et al. XAI — Explainable Artificial Intelligence Science Robotics. — 2019. — Vol. 4. — eaay7120.
28. Dwivedi R. et al. Explainable AI (XAI): Core Ideas, Techniques, and Solutions ACM Computing Surveys.2023.
29. Linardatos P., Papastefanopoulos V., Kotsiantis S. Explainable AI: A Review of Machine Learning Interpretability Methods Entropy. 2020.
30. Chaudhary G. Unveiling the Black Box: Bringing Algorithmic Transparency to AI Masaryk University Journal of Law and Technology. 2024.