Научная статья
УДК 34:004:340.1:004.8:004.051
EDN: https://elibrary.ru/uiujcv
DOI: https://doi.org/10.21202/jdtl.2025.26

Объяснимый искусственный интеллект и правовые традиции: разработка универсальных ключевых показателей эффективности для стран «Большой двадцатки»

Нилкант Бхатт ✉

Государственный инженерный колледж, Раджкот, Индия

Джайкишен Наталал Бхатт

Государственная корпорация страхования работников, Ахмедабад, Индия

Ключевые слова

искусственный интеллект,
общественные интересы,
объяснимый искусственный интеллект,
право,
прозрачность алгоритмов,
уголовное правосудие,
цифровые технологии,
экологическая устойчивость,
экономическое развитие,
этика

Аннотация

Цель: изучить концепцию «право на объяснение» в контексте доктрины PEEC (общественные интересы, экологическая устойчивость, экономическое развитие, уголовное правосудие) для разработки ключевых показателей эффективности, отражающих социокультурные особенности различных стран и обеспечивающих адаптивность, прозрачность и культурную релевантность в регулировании объяснимого искусственного интеллекта.

Методы: в исследовании применяется уникальный методологический подход, сочетающий итеративные процессы методологии мягких систем с теоретической базой, основанной на принципах PEEC. Подобная интеграция позволяет комплексно рассмотреть социальные, экономические, политические и правовые режимы крупнейших стран «Большой двадцатки»: Соединенных Штатов Америки, Федеративной Республики Германия, Японии, Республики Индия, Федеративной Республики Бразилия и Российской Федерации – при конструировании ключевых показателей эффективности. Предложенные ключевые показатели эффективности применимы для оценки прозрачности и подотчетности систем искусственного интеллекта, упрощая сбор данных и практическую имплементацию в различных культурных контекстах. Разработанная модель соответствует реальным общественным потребностям в принятии решений с использованием технологий искусственного интеллекта.

Результаты: в исследовании предлагается новая правовая модель регулирования объяснимого искусственного интеллекта, основанная на системе ключевых показателей эффективности. Помимо устранения

✉ Корреспондирующий автор

проблем регулирования объяснимого искусственного интеллекта в различных культурных, этических и правовых областях, данная модель гарантирует, что система регулирования объяснимого искусственного интеллекта должным образом учитывает антропоцентрические аспекты, поскольку ориентирована на раскрытие истинного потенциала искусственного интеллекта. Предложенный подход способствует максимально эффективному использованию технологий искусственного интеллекта на благо общества в перспективе устойчивого развития.

Научная новизна: в работе применен уникальный научный подход, учитывающий культурные, этические, социально-экономические и правовые различия при разработке правовой базы для регулирования объяснимого искусственного интеллекта, что позволяет адаптировать ее к различным национальным условиям, одновременно способствуя ответственному управлению искусственным интеллектом с системой сдержек и противовесов.

Практическая значимость: полученные результаты позволяют использовать предложенную правовую модель в практической деятельности государственных органов и разработчиков систем искусственного интеллекта для обеспечения прозрачности и объяснимости технологий. Эффективная корректировка предлагаемых ключевых показателей эффективности с учетом специфики конкретных государств позволит оптимизировать их для универсального применения. Хотя все пять ключевых показателей эффективности актуальны для крупнейших стран «Большой двадцатки», их относительная значимость зависит от социокультурных и правовых условий конкретного государства. Дальнейшие исследования должны охватывать более широкий спектр вопросов, включая другие развитые и развивающиеся страны, для адаптации регулирования объяснимого искусственного интеллекта к различным национальным и глобальным требованиям.

Для цитирования

Бхатт, Н., Бхатт, Дж. Н. (2025). Объяснимый искусственный интеллект и правовые традиции: разработка универсальных ключевых показателей эффективности для стран «Большой двадцатки». *Journal of Digital Technologies and Law*, *3*(4), 660–676. https://doi.org/10.21202/jdtl.2025.26

Содержание

Введение

Использование систем искусственного интеллекта (далее – ИИ) расширяется, и из-за их сложности и автономности эволюционируют такие аспекты, как подотчетность, прозрачность и юридическая ответственность этих систем. В случае сбоев в работе ИИ распределение обязанностей и понимание процесса принятия решений этими системами выдвигает на первый план вопрос об «объяснимости» ИИ (Gilpin et al., 2018; Hacker et al., 2020). Чтобы решить эту проблему, Общий регламент ЕС по защите данных (GDPR) предусматривает возможность частным лицам получать информацию о решениях, принимаемых системами искусственного интеллекта (Gilpin et al., 2018). Однако для такой страны, как Индия, с ее сложными и разнородными культурными и социальными условиями, применение этого права для регулирования систем искусственного интеллекта создает значительные проблемы.

Во всем мире, несмотря на успехи в исследованиях, направленных на повышение объяснимости систем искусственного интеллекта, предлагаемые до сих пор структуры по-прежнему не учитывают должным образом разнообразие культурных, социальных и этнических групп заинтересованных сторон. Большинство исследований показывают, что западные модели применимы повсеместно; при этом часто не учитываются незападные, коллективистские общества (Peters & Carman, 2024). Существующие структуры также предполагают, что прозрачность в отношении объяснимости систем искусственного интеллекта будет обеспечиваться действием политических и экономических идеологий. В результате такие системы являются культурно предвзятыми, а их глобальное использование может привести к противоречиям (Prabhakaran et al., 2022). Глобально применимые и согласованные нормы в области искусственного интеллекта должны отражать основные принципы прозрачности, подотчетности, безопасности и динамичной социальной адаптации (Bhatt, 2025).

Системы ИИ, адаптирующиеся к культурным особенностям и учитывающие интересы заинтересованных сторон, являются насущной необходимостью. Регулирование ИИ должно учитывать культурные, социально-политические, этические и правовые особенности различных регионов. Чтобы обеспечить равноправие и целесообразность в сфере «объяснимого ИИ» (Explainable Artificial Intelligence, XAI; далее – ОИИ), необходимо направить внимание на разработку моделей ОИИ, адаптированных к культурным особенностям и учитывающих мнение заинтересованных сторон. Одной из многообещающих идей является «Доктрина PEEC», предложенная Bhatt & Bhatt в 2023 г. Она объединяет такие известные положения, как теории общественных интересов, экологической устойчивости, экономического развития и уголовного права (Public Interest, Environmental Sustainability, Economic Development, and Criminal Law, PEEC), фокусируясь на создании реалистичного подхода к разработке ОИИ. Доктрина также ставит своей целью достижение прозрачности, учет более широких социальных, экономических, политических и правовых последствий результатов работы ИИ. Эта концепция обладает потенциалом для решения задач, связанных с объяснимым ИИ, за счет надлежащего учета элементов PEEC, способствующих устойчивому и простому доступу к параметру объяснимости, обеспечивая при этом достаточную подотчетность

с многомерным подходом к системам ОИИ для удовлетворения реальных потребностей общества.

Системы ИИ/ОИИ представляют собой сложные алгоритмы, включающие в себя социальные, этические и общечеловеческие ценности. Человеческое восприятие, ценности и интерпретации имеют решающее значение для успешного функционирования этих систем. Однако противоречия в целях и задачах, динамичная и непредсказуемая окружающая среда, ценностные проблемы, а также сложное взаимодействие между человеческими ценностями, технологиями и общественными нормами требуют структурированного и надежного методологического подхода, а не чисто технологической или юридической стратегии для решения данной проблемы.

Таким образом, целью нашего исследования является изучение «права на объяснение» и доктрины PEEC путем надлежащего учета различных культур и ценностей крупнейших стран «Большой двадцатки» (США, Германии, Японии, Индии, Бразилии и России) с использованием методологии мягких систем (Soft System Methodology, SSM; далее – ММС) и ключевых показателей эффективности. Это должно помочь в разработке адаптируемых, прозрачных и учитывающих культурные особенности норм ОИИ и повысить уровень доверия и эффективность систем искусственного интеллекта во всем мире.

1. Влияние этических и культурных аспектов на развитие объяснимого искусственного интеллекта в крупнейших странах «Большой двадцатки»

Значительные различия в культурных и этических ценностях стран мира отмечены во многих кросс-культурных исследованиях последних лет (Triandis, 2018; Jan et al., 2024). Это особенно заметно в таких аспектах, как индивидуализм (личная автономия и самоопределение), коллективизм (приоритетность групповой солидарности и общественного благополучия), доверие к технологиям (уверенность в цифровых инновациях и автоматизированных системах) и уважение к власти (приверженность институциональным иерархиям и структурам управления). К примеру, Соединенные Штаты и Германия считаются индивидуалистическими обществами, реализующими идеи личной автономии и уверенности в своих силах (Triandis, 2018). И наоборот, такие культуры, как Япония и Индия, также известные как коллективистские культуры, подчеркивают идеалы группового благополучия и социальной гармонии (Eckhardt, 2002). Данные исследований говорят о том, что распространение технологий в развивающихся странах редко достигает уровня внедрения технологий в развитых странах, в основном из-за социально-экономических проблем и ограничений в области цифровой грамотности (Comin & Hobijn, 2011). Культурные особенности влияют на политические решения, поведение в обществе и международные отношения.

Глубоко укоренившиеся общественные нормы, исторический контекст и национальные (региональные) идеологии и политика формируют основу для различий в культурных и этических ценностях в разных странах. В табл. 1 показаны различия ключевых культурных аспектов разных стран.

Таблица 1. Влияние культурных ценностей на регулирование ОИИ в крупнейших странах «Большой двадцатки»

Крупнейшие страны «Большой двадцатки»	Влияние этических и культурных ценностей на регулирование ОИИ					
	Индивидуализм (личная свобода и независимость)	Коллективизм (направленность на интересы сообщества)	Акцент на общественных выгодах (направленность на всеобщее процветание)	Доверие к технологиям (принятие ИИ, автоматизации и цифровых систем)	Требование прозрачности (подотчетность и открытость в управлении и принятии решений)	Уважение к власти (уважение к руководству и порядку)
США	Решающее	Минимальное	Минимальное	Значительное	Решающее	Низкое
Германия	Значительное	Ограниченное	Ограниченное	Значительное	Умеренное	Значительное
Япония	Умеренное	Решающее	Решающее	Умеренное	Ограниченное	Решающее
Индия	Ограниченное	Значительное	Значительное	Ограниченное	Значительное	Умеренное
Бразилия	Ограниченное	Значительное	Значительное	Ограниченное	Значительное	Ограниченное
Россия	Минимальное	Значительное	Значительное	Ограниченное	Минимальное	Решающее

Эти элементы в значительной степени влияют на формирование политики регулирования ОИИ, способствующей межкультурному сотрудничеству и созданию универсальных моделей регулирования ОИИ. Они также могут обеспечить строгое соответствие указанной политики ожиданиям и ценностям общества.

2. Действующие положения «права на объяснение» в крупнейших странах «Большой двадцатки»

«Право на объяснение» стало предметом пристального внимания в связи с расширением правовых и этических дискуссий о системах искусственного интеллекта. Многие исследователи придерживаются мнения, что право на объяснение не всегда отвечает принципам практичности и достаточности (Edwards & Veale, 2018; Taylor, 2023; Doshi-Velez et al., 2017). Этот механизм был создан в развитых странах, а именно в ЕС, для устранения сложностей в процессе принятия решений с использованием искусственного интеллекта. При этом развивающиеся страны сталкиваются с множеством препятствий, включая юридические и технические, при внедрении права на объяснение в существующее законодательство. Практическая реализация этого права остается сложной задачей для развивающихся стран. Хотя предусмотренное законом «право на объяснение» является мощным механизмом, позволяющим человеку понимать и оспаривать решения автоматизированных систем, его эффективность зависит от создания дополнительных механизмов, таких как оценка воздействия и судебный надзор. Чтобы избежать возможной необъективности и дискриминации при автоматизированном принятии решений, некоторые государства – члены ЕС включили в свое национальное законодательство обязательную оценку влияния таких решений (Malgieri, 2019). Судебный надзор обеспечивает дополнительный уровень контроля и подотчетности и соблюдение справедливости при автоматизированном принятии решений (Gacutan & Selvadurai, 2020; Malgieri, 2019).

Сложности машинного языка ограничивают возможности разработчиков и операторов ИИ по предоставлению содержательных и понятных объяснений для непрофессионалов. Это требует сбалансированного подхода, при котором ни чрезмерный контроль, ни невмешательство не будут препятствовать разработке систем искусственного интеллекта. Использование ИИ в таких секторах, как государственное управление и здравоохранение, должно соответствовать стандартам безопасности, прозрачности и подотчетности в различных социально-технических и правовых контекстах.

В табл. 2 обобщена информация о положениях права на объяснение в законодательстве крупнейших стран «Большой двадцатки» и об отраслях, в которых в настоящее время осуществляется или планируется работа ОИИ.

Таблица 2. Действующие положения «права на объяснение» в отношении систем ИИ в крупнейших странах «Большой двадцатки»

Страна	Право на объяснение	Действующие правовые нормы	Примеры ОИИ по отраслям
США	«Право на объяснение» не закреплено в явном виде, но подразумевается в существующих законах, таких как Закон об алгоритмической подотчетности (2022)	– Система управления рисками ИИ (AI RMF) принята Национальным институтом стандартов и технологий (NIST) в 2023 г.[1] – Федеральная торговая комиссия приняла пять типов правоохранительных мер (2024) против операций, применяющих рекламу искусственного интеллекта или продающих технологии искусственного интеллекта, которые могут быть использованы обманными и недобросовестными способами[2]	Финансовый сектор: Комиссия по ценным бумагам и биржам США (SEC) постановила (в 2023 г.), что финансовые учреждения должны внедрить надежные системы управления ИИ, основанные на принципах прозрачности, управления рисками и этичного принятия решений[3]
Германия	В Общем регламенте по защите данных (GDPR) от 2018 г. прямо указано, что пользователи имеют право на содержательные объяснения при автоматизированном принятии решений	Статьи 22, 71 и 13, 14 и 15 Общего регламента по защите данных предусматривают возможность отдельным лицам понимать и оспаривать решения ИИ[4]	Сектор здравоохранения: согласно GDPR, больницы обязаны разъяснять пациентам автоматизированные решения, касающиеся планов и логики рекомендуемого лечения
Япония	«Право на объяснение» не закреплено в явном виде, но прозрачность и подотчетность при использовании персональных данных обеспечивает Закон о защите личной информации (APPI)	Существует целый комплекс нормативных актов и руководящих принципов[5]. Такие акты, как «Социальные принципы ИИ, ориентированного на человека» (2019), «Руководящие принципы ИИ для бизнеса» (2024) и «Руководящие принципы Японского общества искусственного интеллекта» (JSAI) (2024), направлены на обеспечение соответствия развития ИИ общественным и этическим ценностям	Транспортный сектор: автономные транспортные средства, управляемые искусственным интеллектом, регулируются строгими требованиями безопасности и объяснимости (Irwan & Mursyid, 2025), но это не обеспечивает должного соблюдения прав потребителей
Индия	«Право на объяснение» не закреплено в явном виде, но Закон о защите персональных данных (2023) предусматривает нормы прозрачности ИИ	Закон о защите персональных данных (2023)[6] и «Политика в области ИИ NITI Aayog» (2023)[7] призваны обеспечить понятность, прозрачность, подотчетность и надежность систем искусственного интеллекта	Банковский сектор: резервный банк Индии разрабатывает «Основы ответственного и этичного внедрения искусственного интеллекта (FREE-AI)» в финансовом секторе[8]

1 National Institute of Standards and Technology (NIST). AI Risk Management Framework. https://clck.ru/3QmQ64

2 Federal Trade Commission. (2024). FTC announces crackdown on deceptive AI claims and schemes. Federal Trade Commission. https://clck.ru/3QmQ9Z

3 Essert. AI Governance Frameworks for Financial Institutions. https://clck.ru/3QmQAn

4 General Data Protection Regulation (GDPR). https://clck.ru/3QmQCt

5 Habuka, H. (2023). Japan's approach to AI Regulation and its impact on the 2023 G7 Presidency. Center for Strategic & International Studies. https://clck.ru/3QmQYX

6 Ministry of Law and Justice. (2023). Digital Personal Data Protection Act, 2023. The Gazette of India, CG-DL-E-12082023-248045. https://goo.su/m3v3Zp

7 NITI Aayog. (2023). National Strategy for Artificial Intelligence. NITI Aayog. https://goo.su/nfPaH

8 Reserve Bank of India. (2023). RBI mandates explainability in AI-driven loan approvals. Reserve Bank of India. https://goo.su/SWi8E

Страна	Право на объяснение	Действующие правовые нормы	Примеры ОИИ по отраслям
Бразилия	Законопроект о «праве на объяснение» 2383/2023[9]	Законопроект, одобренный Сенатом, гарантирует, что отдельные лица или группы, затрагиваемые ИИ высокого риска, будут иметь право на своевременное и понятное объяснение решений, рекомендаций и/или прогнозов, принятых с использованием систем ИИ. Законопроект[10] устанавливает национальную нормативную базу, регулирующую использование и разработку систем ИИ в Бразилии	Сектор охраны правопорядка: системы искусственного интеллекта используются для прогнозирования и предотвращения преступлений в крупных городах Бразилии (Ribeiro et al.,2024)
Россия	«Право на объяснение» не закреплено в явном виде, но принципы, изложенные в «Стратегии в области искусственного интеллекта», направлены на создание систем ИИ, обеспечивающих прозрачность при защите прав отдельных лиц	Национальная стратегия развития ИИ в России направлена на создание российских продуктов и услуг в области ИИ. Основной упор делается на «сильный ИИ» для военных операций и национальных разработок[11]	Военный сектор: использование искусственного интеллекта для анализа данных, позволяющего лучше и быстрее принимать решения в бою[12]

3. Оценка применимости доктрины PEEC к объяснимому искусственному интеллекту в крупнейших странах «Большой двадцатки»

Существуют различные подходы к регулированию искусственного интеллекта и конкретных мер в этой области. Эти подходы отражают уникальный социально-культурный и политический ландшафт каждой страны. Обосновать их помогает качественная оценка доктрины PEEC, предложенной в работе Bhatt & Bhatt (2023). Меры регулирования ИИ в разных странах определяются их соответствующими социокультурными приоритетами и идеологиями управления. Соответствующим образом должны оцениваться и элементы концепции PEEC, а именно: общественные интересы, экологическая устойчивость, экономическое развитие и уголовное законодательство.

Общественные интересы в разных странах также различаются. Соединенные Штаты уделяют первостепенное внимание защите прав потребителей, Германия – конфиденциальности данных, в то время как Индия и Япония больше нацелены на достижение социальной гармонии и равноправного доступа. Бразилия и Россия стремятся бороться с недостатками в управлении и обеспечивать государственную безопасность. Что касается экологической устойчивости, такие страны, как Соединенные Штаты и Германия, стараются использовать искусственный интеллект для инноваций в частном секторе и повышения производительности и эффективности промышленности. Япония и Индия более склонны к достижению долгосрочных целей в области интеллектуального планирования городов и управления водными

[9] Data Privacy Brazil Research Association. (2024). The artificial intelligence legislation in Brazil: Technical analysis of the text to be voted on in the Federal Senate plenary. https://clck.ru/3QmR23

[10] The Mattos Filho News Portal. (2024). Framework for artificial intelligence in the Senate. https://goo.su/lbFTrr

[11] CNA. (2020). Artificial intelligence in Russia: Issue 11. https://clck.ru/3QmRSw

[12] Boulanin, V., & Zerbo, L. (2023, July 20). Roles and implications of AI in the Russian-Ukrainian conflict. Russia Matters. https://clck.ru/3QmRTy

ресурсами. И Бразилия, и Россия понимают, что системы искусственного интеллекта могут помочь достичь экологической устойчивости. Бразилия сосредоточена на борьбе с климатическими проблемами, в то время как российский подход больше ориентирован на энергетический сектор.

В экономическом плане Россия и Бразилия преуспевают благодаря государственным инновациям и технологической модернизации, в то время как Соединенные Штаты и Германия поддерживают инновации и уделяют особое внимание охране труда. Япония и Индия отдают приоритет развитию робототехники и финансовых технологий. Использование искусственного интеллекта в уголовном законодательстве также значительно различается. В то время как Соединенные Штаты балансируют между безопасностью и личной свободой, Германия делает упор на надзор. Япония использует искусственный интеллект с системой сдержек и противовесов. Индийский подход заключается в разработке гарантий. Бразилия уделяет приоритетное внимание решению существующих проблем, а Россия – безопасности через системы наблюдения. Это разнообразие подходов подчеркивает сложные взаимосвязи, которые следует учитывать при разработке ОИИ во всем мире.

Целесообразно оценить, насколько принципы PEEC, предложенные Bhatt & Bhatt в 2023 г., способны обеспечить такое положение, чтобы регулирование ОИИ оставалось эффективным, учитывало контекст и соответствовало ожиданиям общества во всем мире. В табл. 3 показано соответствие принципов PEEC требованиям разработки ОИИ в ключевых экономиках «Большой двадцатки».

Таблица 3. «Тепловая карта» соответствия принципов PEEC в разных странах требованиям разработки ОИИ

№	Принцип PEEC	США	Германия	Япония	Индия	Бразилия	Россия
1	Прозрачность и подотчетность	Да	Да	Нет	Да	Да	Нет
2	Безопасность и конфиденциальность данных	Да	Да	Нет	Частично	Частично	Нет
3	Этические аспекты	Да	Да	Частично	Да	Да	Частично
4	Оценка влияния на экологию	Частично	Да	Да	Да	Да	Нет
5	Экономические стимулы и инновации	Да	Да	Да	Да	Да	Да
6	Управление рисками и ответственность	Частично	Да	Нет	Частично	Частично	Да
7	Участие общественности и консультации	Да	Да	Нет	Да	Да	Нет
8	Правоприменение и уголовное регулирование ИИ	Частично	Да	Нет	Частично	Частично	Частично
9	Междисциплинарное сотрудничество	Да	Да	Частично	Да	Да	Нет

4. Развитие интегрированных ключевых показателей эффективности в сфере объяснимого искусственного интеллекта

Чисто технический или количественный подход не может полностью охватить всю сложность, субъективность и этические аспекты системы регулирования ОИИ. Для создания действительно всеобъемлющей и надежной системы необходимо активно изучать и учитывать различные точки зрения всех заинтересованных сторон. Одним из интересных подходов к решению проблемных ситуаций различного характера, особенно связанных с человеческими системами, является методология мягких систем (ММС) (Checkland & Poulter, 2020).

Методология мягких систем позволяет пользователям организовать свое взаимодействие со сложными техническими, политическими и социокультурными проблемами и искать их комплексные решения. Интеграция принципов PEEC и MMC может стать мощным инструментом для разработки интегрированных ключевых показателей эффективности для практического регулирования политики в отношении ОИИ.

Чтобы по-настоящему разобраться в тонкостях объяснимости систем искусственного интеллекта, необходимо выйти за рамки чисто технических аспектов. Авторы предлагают новый подход к разработке комплексных ключевых показателей эффективности (KPI), которые учитывали бы общественные интересы, экологическую устойчивость, экономическое развитие и уголовное правосудие посредством структурированной процедуры MMC для целостной оценки влияния ИИ на социальные, экономические, политические и правовые режимы. На рисунке показана концептуальная модель, предлагаемая для достижения указанной цели.

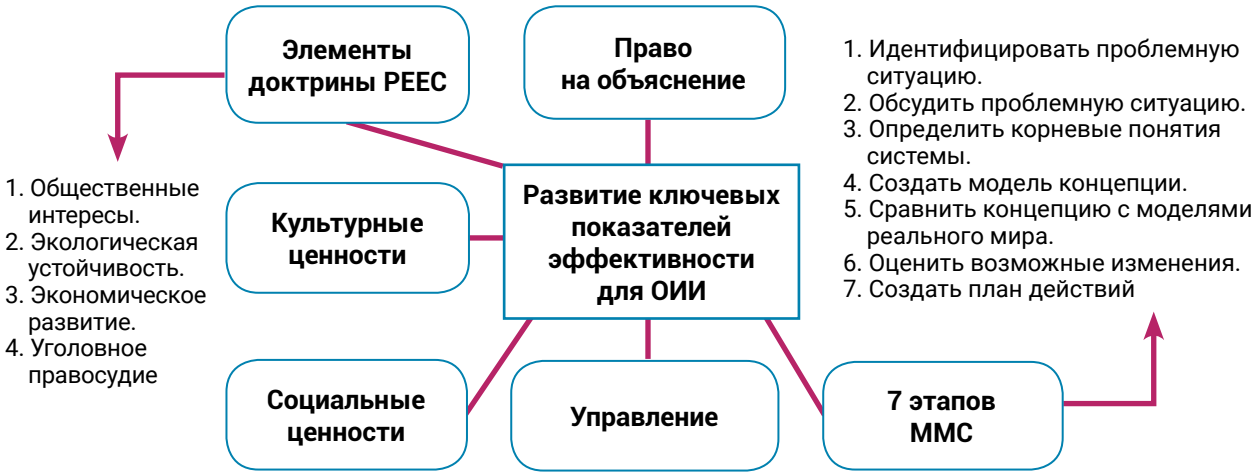

Рис. Концептуальная модель для разработки KPI

Для разработки надежных и универсально приемлемых ключевых показателей эффективности для ОИИ мы использовали подход MMC, в котором объединены ключевые аспекты общественных интересов, экологической устойчивости, экономического развития, правовых вопросов и управления. Структурированные и устойчивые процессы, характерные для MMC, гарантируют, что разработанные ключевые показатели эффективности будут соответствовать различным культурным и правовым контекстам. На первом этапе была проведена тщательная идентификация проблем прозрачности ИИ в различных условиях. Во-вторых, был проведен всесторонний обзор существующей политики, научных исследований и новостных статей по теме ИИ; этап критического всестороннего анализа позволил наглядно отобразить ожидания ключевых заинтересованных сторон в области ИИ, а именно: официальных лиц, общественности, отраслевых экспертов и юристов. В-третьих, для уточнения «Направлений и областей воздействия PEEC» мы использовали определение корневых понятий и моделирование концепции. На этом этапе было обеспечено соответствие данной концепции различным аспектам социально-экономической, правовой, этической и экологической устойчивости. В-четвертых, мы провели сравнительный анализ для подтверждения применимости регулирования ИИ в реальных условиях. Наконец, итеративный уточняющий цикл помог обеспечить, чтобы разработанные

ключевые показатели эффективности не только соответствовали потребностям, но и были практически реализуемы, что упрощает сбор данных и позволяет использовать измеримые критерии. В табл. 4 представлена предлагаемая комплексная структура ключевых показателей эффективности для ОИИ.

Для решений с высокой степенью риска, когда отсутствие объяснимого решения чревато серьезными последствиями, мы предлагаем установить «Индекс прозрачности и доверия» (Clarity and Trust Index, CTI; далее – ИПД) на уровне от 90 до 100 % в зависимости от предъявляемых требований. Для рутинных автоматизированных решений значение ИПД может составлять от 50 до 80 %, а для стратегических решений – от 70 до 90 %.

Таблица 4. Предлагаемая комплексная структура KPI для ОИИ

Направления и области воздействия PEEC	Предлагаемые KPI	Определение	Метод расчета
Общественные интересы: установление контроля над социальными и правовыми проблемами	Индекс прозрачности и доверия (Clarity and Trust Index, CTI)	Процент решений ИИ, содержащих четкие и понятные объяснения пользователю	CTI = (E ÷ T) × 100, где E = число объяснений, T = число решений
Общественные интересы: установление контроля над социальным и экономическим неравенством	Индекс снижения предвзятости (Bias Reduction Index, BRI)	Снижение предвзятости при принятии решений с помощью ИИ в разрезе демографических факторов	BRI = 1 – (BB ÷ MB), где BB = базовый уровень предвзятости = наблюдаемый уровень предвзятости решений ИИ (например, неравномерность выбора между группами), MB = максимальный уровень предвзятости = худший сценарий предвзятости (например, 100 % у одной группы и 0 % у другой). При BRI = 0 уровень предвзятости максимальный; при BRI = 100 предвзятость отсутствует
Экологическая устойчивость: обеспечение соблюдения требований экологии и зеленой экономики	Индекс углеродного следа ИИ (AI Carbon Footprint Index, AICFI)	Мера воздействия систем ИИ на окружающую среду с точки зрения их энергопотребления и выбросов парниковых газов	AICFI = ACF × TD, где ACF = энергия, потребляемая системой ИИ (кВт на одно решение) × фактор выбросов углерода (кг CO_2 на кВт), который зависит от источника энергии TD = число решений
Экономическое развитие: обеспечение положительного влияния искусственного интеллекта на региональную культуру и экономику	Соотношение социально-экономических выгод и затрат в области ИИ (AI Socio-economic Benefit-Cost Ratio, ASEBC)	Число созданных рабочих мест, экономических выгод и связанных с ними затрат на внедрение систем ИИ для анализа влияния этой технологии на экономику и культуру	ASEBC = EB ÷ CD, где EB = экономическая выгода от внедрения ИИ, CD = затраты на внедрение ИИ
Право и управление: отслеживание эффективности систем искусственного интеллекта в различных культурах и правовых системах	Оценка культурной и правовой подотчетности (Cultural & Legal Accountability Score, CLAS)	Число споров, жалоб и соответствующих решений, касающихся использования ИИ в контексте различных культур при наличии собственного правового механизма регулирования ИИ	CLAS = RD ÷ TG, где RD = общее число урегулированных споров и жалоб в области ИИ, TG = общее число споров и жалоб в области ИИ

В идеале предлагаемый «Индекс снижения предвзятости» (BRI) должен составлять 100 %, хотя в большинстве случаев его значение выше 90 % будет приемлемым. Теоретически «Индекс углеродного следа ИИ» (AICFI) должен быть как можно

ниже. Однако он также может меняться в соответствии с Целями устойчивого развития ООН. Большинство стран стремились бы, чтобы показатель «Соотношение социально-экономических выгод и затрат в области ИИ» (ASEBC) был выше 1,0, однако следует направить усилия на максимизацию реальных социально-экономических выгод в чистом выражении. Показатель «Оценка культурной и правовой подотчетности» (CLAS) в идеале должен составлять 1,0, хотя в большинстве случаев ожиданиям общественности будет соответствовать значение, превышающее 0,9. Осуществимость и желательность предлагаемых значений показателей KPI можно будет определить после тщательного анализа ситуации в каждой конкретной стране, при условии вовлеченности заинтересованных сторон и необходимого понимания факторов культурного контекста. Ответственность за разработку точных диапазонов, отражающих национальные условия и при этом остающихся приемлемыми в глобальном масштабе, лежит на директивных органах. Такой подход будет способствовать реализации истинного потенциала систем ОИИ на благо человечества.

Эффективная корректировка предлагаемых ключевых показателей с учетом специфики конкретной страны позволит оптимизировать их для универсального использования. Хотя все пять ключевых показателей эффективности актуальны для крупнейших стран «Большой двадцатки», их относительная важность зависит от социально-культурного и правового контекста конкретной страны. Давно идущие споры о регулировании и корпоративных инициативах в США требуют более высокого CTI. В то же время в Германии положения GDPR предполагают необходимость более высоких значений для показателя BRI и более низких – для AICFI. Политика Бразилии и Индии в большей степени ориентирована на более высокий уровень CLAS и ASEBC. Российская политика направлена на использование искусственного интеллекта в управлении и, таким образом, на более высокий уровень ASEBC, а также четкую стратегическую направленность на поддержание национального суверенитета и целостности.

Заключение

Различия в культурных, этических, социально-экономических и правовых подходах, которые выбирает общество, ставят перед директивными органами сложную задачу по разработке регулирующей системы ОИИ, соответствующей международным требованиям. Эти факторы, по сути, определяют успех или неудачу норм ОИИ. Поиск путей прогресса требует внимания не только к технологическим, но и в первую очередь к человеческим аспектам. Такие аспекты, как право на объяснение, общественные интересы, экологическая устойчивость, экономическое развитие и уголовное правосудие (PEEC), являются основополагающими и всегда будут оставаться центральными для регулирования технологий искусственного интеллекта.

В данном исследовании предлагается новая модель регулирования ОИИ, основанная на ключевых показателях эффективности и принципах PEEC и использующая структурированный подход методологии мягких систем. Чтобы по-настоящему использовать потенциал предлагаемой модели ключевых показателей эффективности, страны должны установить соответствующие национальным стандартам диапазоны показателей, которые имели бы международное значение. Помимо устранения проблем регулирования ОИИ в контексте различных культурных, этических и правовых норм, предлагаемая модель гарантирует, что система регулирования

ОИИ должным образом учитывает человеческие аспекты, поскольку она стремится использовать истинный потенциал искусственного интеллекта. Этот подход поможет максимально эффективно использовать ИИ на благо общества в будущем.

В работе рассматриваются только культурные различия крупнейших стран «Большой двадцатки». Дальнейшие исследования должны охватывать более широкий спектр вопросов, включая другие развитые и развивающиеся страны, чтобы сделать регулирующую систему ОИИ адаптируемой к различным национальным и глобальным требованиям.

Список литературы

Bhatt, N. (2025). Crimes in the Age of Artificial Intelligence: a Hybrid Approach to Liability and Security in the Digital Era. *Journal of Digital Technologies and Law*, *3*(1), 65–88. https://doi.org/10.21202/jdtl.2025.3

Bhatt, N., & Bhatt, J. (2023). Towards a novel eclectic framework for administering artificial intelligence technologies: A proposed 'PEEC' doctrine. *EPRA International Journal of Research and Development (IJRD)*, *8*(9), 27–36. https://doi.org/10.13140/RG.2.2.11434.18888

Checkland, P., & Poulter, J. (2020). Soft Systems Methodology. In M. Reynolds, S. Holwell (Retired) (Eds), *Systems Approaches to Making Change: A Practical Guide* (pp. 201–253). Springer, London. https://doi.org/10.1007/978-1-4471-7472-1_5

Comin, D., & Hobijn, B. (2011). An exploration of technology diffusion. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1116606

Doshi-Velez, F., Kortz, M., Budish, R., Bavitz, C., Gershman, S., O'Brien, D., Schieber, S., Waldo, J., Weinberger, D., Weller, A., & Wood, A. (2017). Accountability of AI Under the Law: The Role of Explanation. *ArXiv*, abs/1711.01134. https://doi.org/10.2139/SSRN.3064761

Eckhardt, G. (2002). Culture's Consequences: Comparing Values, Behaviors, Institutions and Organisations Across Nations. *Australian Journal of Management*, *27*(1), 89–94. https://doi.org/10.1177/031289620202700105

Edwards, L., & Veale, M. (2018). Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"? *IEEE Security & Privacy*, *16*, 46–54. https://doi.org/10.1109/MSP.2018.2701152b

Gacutan, J., & Selvadurai, N. (2020). A statutory right to explanation for decisions generated using artificial intelligence. *International Journal of Law and Information Technology*, *28*(3), 193–216. https://doi.org/10.1093/ijlit/eaaa016

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining Explanations: An Overview of Interpretability of Machine Learning. In *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, Turin, Italy, 2018 (pp. 80–89). https://doi.org/10.1109/DSAA.2018.00018

Hacker, P., Krestel, R., Grundmann, S., & Naumann, F. (2020). Explainable AI under contract and tort law: Legal incentives and technical challenges. *Artificial Intelligence and Law*, *28*(4), 415–439. https://doi.org/10.1007/s10506-020-09260-6

Irwan, M., & Mursyid, M. (2025). AI-Driven Traffic Accidents: A Comparative Legal Study. *Artes Libres Law and Social Journal*, *1*(1), 1–20. https://doi.org/10.12345/jxt3j717

Jan, J., Alshare, K. A., & Lane, P. L. (2024). Hofstede's cultural dimensions in technology acceptance models: a meta-analysis. *Universal Access in the Information Society*, *23*(2), 717–741. https://doi.org/10.1007/s10209-022-00930-7

Malgieri, G. (2019). Automated decision-making in the EU Member States: The right to explanation and other «suitable safeguards» in the national legislations. *Computer Law & Security Review*, *35*(5), 105327. https://doi.org/10.1016/J.CLSR.2019.05.002

Peters, U., & Carman, M. (2024). Cultural bias in explainable AI research: A systematic analysis. *Journal of Artificial Intelligence Research*, *79*, 971–1000. https://doi.org/10.1613/jair.1.14888

Prabhakaran, V., Qadri, R., & Hutchinson, B. (2022). Cultural incongruencies in artificial intelligence. *arXiv preprint arXiv:2211.13069*. https://doi.org/10.48550/arXiv.2211.13069

Ribeiro, L. H. da C., Silva, C. M. da, & Viana, P. W. P. (2024). Artificial intelligence as a tool for predicting crime in large Brazilian cities. *Revista FT*, *28*. https://doi.org/10.5281/zenodo.11100354

Taylor, E. (2023). Explanation and the Right to Explanation. *Journal of the American Philosophical Association*, *10*(3), 467–482. https://doi.org/10.1017/apa.2023.7

Triandis, H. C. (2018). *Individualism and collectivism*. Routledge. https://doi.org/10.4324/9780429499845

Сведения об авторах

Бхатт Нилкант – PhD, доцент, заведующий кафедрой гражданского строительства, Государственный инженерный колледж
Адрес: 360 005, Индия, штат Гуджарат, г. Раджкот, ул. Мавди-Канкот
E-mail: neelkanth78bhatt@gmail.com
ORCID ID: https://orcid.org/0000-0003-0315-2985
Scopus Author ID: https://www.scopus.com/authid/detail.uri?authorId=58919442100
WoS Researcher ID: https://www.webofscience.com/wos/author/record/KRO-8652-2024
Google Scholar ID: https://scholar.google.com/citations?user=L7K-e3IAAAAJ

Бхатт Джайкишен Наталал – бакалавр коммерции, сотрудник службы социального обеспечения в отставке, Государственная корпорация страхования работников
Адрес: 380 009, Индия, штат Гуджарат, г. Ахмедабад, ул. Ашрам, Панчдип Бхаван
E-mail: neelkanth.bhatt@gujgov.edu.in
ORCID ID: https://orcid.org/0009-0000-3645-829X

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

Исследование не имело спонсорской поддержки.

Тематические рубрики

Рубрика OECD: 5.05 / Law
Рубрика ASJC: 3308 / Law
Рубрика WoS: OM / Law
Рубрика ГРНТИ: 10.07.45 / Право и научно-технический прогресс
Специальность ВАК: 5.1.1 / Теоретико-исторические правовые науки

История статьи

Дата поступления – 18 июля 2025 г.
Дата одобрения после рецензирования – 2 августа 2025 г.
Дата принятия к опубликованию – 20 декабря 2025 г.
Дата онлайн-размещения – 25 декабря 2025 г.

Research article
UDC 34:004:340.1:004.8:004.051
EDN: https://elibrary.ru/uiujcv
DOI: https://doi.org/10.21202/jdtl.2025.26

Explainable Artificial Intelligence and Legal Ethos: Developing Key Performance Indicators for 'G20 Giants'

Neelkanth Bhatt ✉
Government Engineering College, Rajkot, India

Jaikishen Nathalal Bhatt
Employees' State Insurance Corporation, Ahmedabad, India

Keywords

algorithmic transparency,
artificial intelligence,
criminal justice,
digital technologies,
economic development,
environmental sustainability,
ethics,
explainable artificial intelligence,
law,
public interest

Abstract

Objective: to study the "right to explanation" in the context of the PEEC doctrine (public interest, environmental sustainability, economic development, criminal justice) in order to develop key performance indicators reflecting the socio-cultural characteristics of different countries and ensuring adaptability, transparency and cultural relevance in the regulation of explainable artificial intelligence.

Methods: the research uses a unique methodological approach that combines the iterative processes of soft systems methodology with a theoretical framework based on the PEEC principles. Such integration makes it possible to comprehensively study the social, economic, political and legal regimes of the 'G20 Giants' – the United States of America, the Federal Republic of Germany, Japan, the Republic of India, the Federal Republic of Brazil and the Russian Federation – when designing key performance indicators. The proposed key performance indicators are applicable to assess the transparency and accountability of artificial intelligence systems, simplifying data collection and practical implementation in various cultural contexts. The developed model corresponds to the actual social needs in decision-making using artificial intelligence technologies.

✉ Corresponding author

Results: the study proposes a new legal model for regulating explainable artificial intelligence based on a system of key performance indicators. In addition to eliminating the problems of regulating explainable artificial intelligence in various cultural, ethical and legal fields, this model ensures that the system of regulating explainable artificial intelligence properly takes into account anthropocentric aspects, since it is focused on unlocking the true potential of artificial intelligence. The proposed approach promotes the most effective use of artificial intelligence technologies for the benefit of society in the perspective of sustainable development.

Scientific novelty: the work applies a unique scientific approach that takes into account cultural, ethical, socio-economic and legal differences when developing a legal framework for regulating explainable artificial intelligence. This allows adapting the legal framework to various national conditions, while contributing to responsible management of artificial intelligence with a check-and-balance system.

Practical significance: the results obtained make it possible to use the proposed legal model in the practical activities of government agencies and developers of artificial intelligence systems to ensure transparency and explainability of technologies. Effective adjustment of the proposed key performance indicators, taking into account the specifics of states, will optimize them for universal use. Although all five key performance indicators are relevant for the 'G20 Giants', their relative significance depends on the socio-cultural and legal conditions of a particular state. Further research should cover a wider range of issues, including other developed and developing countries, in order to adapt the regulation of explainable artificial intelligence to various national and global requirements.

For citation

Bhatt, N., & Bhatt, J. N. (2025). Explainable Artificial Intelligence and Legal Ethos: Developing Key Performance Indicators for 'G20 Giants'. *Journal of Digital Technologies and Law, 3*(4), 660–676. https://doi.org/10.21202/jdtl.2025.26

References

Bhatt, N. (2025). Crimes in the Age of Artificial Intelligence: a Hybrid Approach to Liability and Security in the Digital Era. *Journal of Digital Technologies and Law*, *3*(1), 65–88. https://doi.org/10.21202/jdtl.2025.3

Bhatt, N., & Bhatt, J. (2023). Towards a novel eclectic framework for administering artificial intelligence technologies: A proposed 'PEEC' doctrine. *EPRA International Journal of Research and Development (IJRD)*, *8*(9), 27–36. https://doi.org/10.13140/RG.2.2.11434.18888

Checkland, P., & Poulter, J. (2020). Soft Systems Methodology. In M. Reynolds, S. Holwell (Retired) (Eds), *Systems Approaches to Making Change: A Practical Guide* (pp. 201–253). Springer, London. https://doi.org/10.1007/978-1-4471-7472-1_5

Comin, D., & Hobijn, B. (2011). An exploration of technology diffusion. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1116606

Doshi-Velez, F., Kortz, M., Budish, R., Bavitz, C., Gershman, S., O'Brien, D., Schieber, S., Waldo, J., Weinberger, D., Weller, A., & Wood, A. (2017). Accountability of AI Under the Law: The Role of Explanation. *ArXiv*, abs/1711.01134. https://doi.org/10.2139/SSRN.3064761

Eckhardt, G. (2002). Culture's Consequences: Comparing Values, Behaviors, Institutions and Organisations Across Nations. *Australian Journal of Management*, *27*(1), 89–94. https://doi.org/10.1177/031289620202700105

Edwards, L., & Veale, M. (2018). Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"? *IEEE Security & Privacy*, *16*, 46–54. https://doi.org/10.1109/MSP.2018.2701152b

Gacutan, J., & Selvadurai, N. (2020). A statutory right to explanation for decisions generated using artificial intelligence. *International Journal of Law and Information Technology*, *28*(3), 193–216. https://doi.org/10.1093/ijlit/eaaa016

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining Explanations: An Overview of Interpretability of Machine Learning. In *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, Turin, Italy, 2018 (pp. 80–89). https://doi.org/10.1109/DSAA.2018.00018

Hacker, P., Krestel, R., Grundmann, S., & Naumann, F. (2020). Explainable AI under contract and tort law: Legal incentives and technical challenges. *Artificial Intelligence and Law*, *28*(4), 415–439. https://doi.org/10.1007/s10506-020-09260-6

Irwan, M., & Mursyid, M. (2025). AI-Driven Traffic Accidents: A Comparative Legal Study. *Artes Libres Law and Social Journal*, *1*(1), 1–20. https://doi.org/10.12345/jxt3j717

Jan, J., Alshare, K. A., & Lane, P. L. (2024). Hofstede's cultural dimensions in technology acceptance models: a meta-analysis. *Universal Access in the Information Society*, *23*(2), 717–741. https://doi.org/10.1007/s10209-022-00930-7

Malgieri, G. (2019). Automated decision-making in the EU Member States: The right to explanation and other "suitable safeguards" in the national legislations. *Computer Law & Security Review*, *35*(5), 105327. https://doi.org/10.1016/J.CLSR.2019.05.002

Peters, U., & Carman, M. (2024). Cultural bias in explainable AI research: A systematic analysis. *Journal of Artificial Intelligence Research*, *79*, 971–1000. https://doi.org/10.1613/jair.1.14888

Prabhakaran, V., Qadri, R., & Hutchinson, B. (2022). Cultural incongruencies in artificial intelligence. *arXiv preprint arXiv:2211.13069*. https://doi.org/10.48550/arXiv.2211.13069

Ribeiro, L. H. da C., Silva, C. M. da, & Viana, P. W. P. (2024). Artificial intelligence as a tool for predicting crime in large Brazilian cities. *Revista FT*, 28. https://doi.org/10.5281/zenodo.11100354

Taylor, E. (2023). Explanation and the Right to Explanation. *Journal of the American Philosophical Association*, *10*(3), 467–482. https://doi.org/10.1017/apa.2023.7

Triandis, H. C. (2018). *Individualism and collectivism*. Routledge. https://doi.org/10.4324/9780429499845

Authors information

Neelkanth Bhatt – PhD, Head of the Department & Associate Professor, Department of Civil Engineering, Government Engineering College
Address: Near Mavdi-Kankot Road, Rajkot, Pin Code 360 005, Gujarat, India
E-mail: neelkanth78bhatt@gmail.com
ORCID ID: https://orcid.org/0000-0003-0315-2985
Scopus Author ID: https://www.scopus.com/authid/detail.uri?authorId=58919442100
WoS Researcher ID: https://www.webofscience.com/wos/author/record/KRO-8652-2024
Google Scholar ID: https://scholar.google.com/citations?user=L7K-e3IAAAAJ

Jaikishen Nathalal Bhatt – Bachelor of Commerce, Retired Social Security Officer, Employees' State Insurance Corporation
Address: Panchdeep Bhavan, Ashram Road, Ahmedabad, Pin 380 009, Gujarat, India
E-mail: neelkanth.bhatt@gujgov.edu.in
ORCID ID: https://orcid.org/0009-0000-3645-829X

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 interest

The authors declares no conflict of interest.

Financial disclosure

The research had no sponsorship.

Thematic rubrics

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

Article history

Date of receipt – July 18, 2025
Date of approval – August 2, 2025
Date of acceptance – December 20, 2025
Date of online placement – December 25, 2025