Допустимость использования объектов интеллектуальной собственности при обучении нейросетей

Тамаев Арсений Витальевич
аспирант, кафедра информационного и интеллектуального права, цифровых технологий и инноватики, Российский государственный университет правосудия имени В.М. Лебедева, ai-ip-law@mail.ru

Статья посвящена правовым аспектам использования объектов интеллектуальной собственности при обучении генеративных моделей искусственного интеллекта. В условиях быстрых технологических изменений, связанных с внедрением генеративного ИИ, возникает необходимость балансировать между инновациями и защитой авторских прав. Основной проблемой является использование больших объемов данных, многие из которых защищены авторским правом, без ведома и согласия правообладателей. Генеративные модели, обучающиеся на таких данных, могут случайно «запоминать» защищенный контент — воспроизводить его частично или полностью в выдаваемых результатах, что приводит к правовым спорам и судебным искам. В статье рассматриваются различные подходы к прозрачности данных для обучения ИИ, а также правовые нормы, которые регулируют использование таких данных в России. Особое внимание уделяется отсутствию четких норм в российском законодательстве, что создает правовую неопределенность. Работа анализирует возможные риски для разработчиков ИИ и пути их минимизации через совершенствование законодательства в области авторских прав. В заключении акцентируется внимание на необходимости формирования новых правовых норм, обеспечивающих баланс между инновациями и защитой интеллектуальной собственности.

Ключевые слова: искусственный интеллект, обучение нейросетей, объекты интеллектуальной собственности, авторское право, исключительное право.

С момента дебюта ChatGPT в конце 2022 года внимание политиков по всему миру привлек генеративный искусственный интеллект (далее — ИИ) [1]. Он широко рассматривается как технология с преобразующим потенциалом, однако ставит перед политиками серьезные вопросы: как реагировать на потенциал технологии воспроизводить или усиливать существующие предубеждения и предрассудки, распространять дезинформацию и угрожать средствам к существованию людей (особенно в творческих индустриях) [2]. Перед законом стоит сложный вопрос о том, как можно смягчить вред генеративного ИИ, не подавляя инновации.

Генеративный ИИ поднимает глубокие вопросы относительно основополагающих положений, лежащих в основе авторского права. Один из наиболее острых из них касается данных, используемых для обучения моделей генеративного ИИ [3]. Эти модели требуют огромных объемов данных: миллионов текстовых документов, изображений, аудиофрагментов и других форм контента. Большая часть этих материалов защищена авторским правом, но разработчики ИИ зачастую практически не прилагают усилий для получения разрешения правообладателей на использование своих работ.

Правообладатели возбудили десятки судебных исков против разработчиков ИИ в различных юрисдикциях на том основании, что несанкционированное использование их работ является нарушением авторских прав [4, 5]. Разработчики ИИ возражают, что такое использование подпадает под действие различных исключений из исключительного права правообладателя разрешать воспроизведение своих работ [6]. Если правообладатели добьются успеха хотя бы в некоторых из этих случаев, то понесенный ущерб может оказаться достаточным, чтобы обанкротить даже крупнейших разработчиков ИИ [7]. Таким образом, хотя генеративный ИИ служит «стресс-тестом» для законодательства об авторском праве, законодательство об авторском праве, в свою очередь, представляет собой потенциальную «экзистенциальную угрозу» для генеративного ИИ. Это делает актуальным исследование допустимости использования объектов интеллектуальной собственности при обучении генеративного ИИ.

Целью работы является изучение допустимости использования объектов интеллектуальной собственности при обучении нейросетей. Для ее достижения были использованы структурно-функциональные и теоретические методы исследования: анализ, обобщение и синтез нормативно-правовых документов и литературы в области использования генеративного ИИ.

Для понимания значения споров вокруг авторских прав и данных для обучения ИИ необходимо рассмотреть, как эти данные используются. Современная волна популярных генеративных моделей ИИ, таких как серия GPT от OpenAI или генераторы изображений Stability AI, основана на подтипе машинного обучения, известном как «глубокое обучение» [8]. Как и другие формы машинного обучения, глубокое обучение использует «нейронные сети», то есть связанные блоки узлов, вдохновленные структурой человеческого мозга. Глубокое обучение отличается использует несколько слоев узлов, называемых «глубокими» слоями. Информация, проходя через них, обрабатывается на разных уровнях сложности: первые слои обычно идентифицируют простые закономерности, а последующие, опираясь на эту основу, распознают закономерности все большей сложности. Это позволяет моделям ИИ «обучаться» на больших объемах данных.

В генеративной модели ИИ закономерности и правила, выявленные в процессе обучения, могут быть затем использованы для создания нового контента [9]. Степень, в которой окончательная генеративная модель ИИ сохраняет данные, на которых она была обучена, не совсем ясна. Хотя общепризнанно, что модели ИИ кодируют закономерности, полученные из данных в процессе глубокого обучения, как числовые параметры, а не сохраняют весь набор данных для обучения, в некоторых случаях генеративные модели ИИ могут воссоздавать идентичные или почти идентичные копии материала, найденного в их тренировочных данных — явление, известное как «запоминание» [10].

Современные генеративные модели ИИ требуют огромных объемов обучающих данных. К примеру, GPT-3 был обучен на наборах данных Common Crawl и WebText2, двух наборах данных книг и англоязычной версии Википедии [11]. В дополнение к этим общим обучающим данным, модель ИИ может затем дополнительно обучаться на меньшем, курируемом наборе данных для тонкой настройки. В результате генеративная модель ИИ может быть обучена на миллионах отдельных произведений.

Разработчики моделей ИИ широко призывают к прозрачности содержания наборов обучающих данных. Хотя во многих случаях эти требования обусловлены опасениями по поводу несанкционированного использования контента, защищенного авторским правом, стоит отметить, что существуют и аргументы в пользу прозрачности обучающих данных, не связанные с авторским правом. Например, призывы к прозрачности обучающих данных также мотивированы опасениями, что их содержание может привести к предвзятым или иным несправедливым результатам [12]. Хотя все разработчики ИИ заявляют о проверке своих моделей на наличие такой предвзятости перед выпуском, прозрачность обучающих данных позволяет внешним сторонам изучать обучающие данные способами, выходящими за рамки компетенции любой отдельной команды. Прозрачность может способствовать укреплению общественного доверия к моделям ИИ, уменьшая информационную асимметрию между поставщиками и потребителями моделей.

Однако, несмотря на эти аргументы в пользу прозрачности, разработчики ИИ в последние годы стали заметно более скрытными в отношении своих тренировочных данных. Многие крупные разработчики ИИ перешли от подробных объяснений тренировочных данных, используемых для обучения конкретной модели, к описаниям отдельными предложениями. Например, OpenAI раскрыла основные источники данных для GPT-3, но в документе, представляющем GPT-4, было указано только, что данные, на которых обучалась модель, представляли собой смесь «общедоступных данных (таких как интернет-данные) и данных, лицензированных у сторонних поставщиков» [13]. OpenAI обосновала свое решение не публиковать более подробную информацию о GPT-4 опасениями относительно «конкурентной среды и последствий для безопасности крупномасштабных моделей».

Хотя эти аргументы имеют определенные основания, существуют явные возражения против них с точки зрения государственной политики. Предотвращение копирования конкурентами инновационных технологий без инвестирования сопоставимых ресурсов является общей целью для многих компаний, но лишь изредка отвечает общественным интересам; более того, отсутствие требований к прозрачности обучающих данных само по себе может способствовать антиконкурентной практике, например, позволяя крупнейшим разработчикам ИИ заключать преференциальные лицензионные соглашения с организациями, имеющими доступ к большим пулам обучающих данных. Неясно, почему ограничение доступа к обучающим данным будет особенно эффективным способом предотвращения попадания опасных инструментов ИИ в руки злоумышленников, особенно по сравнению с ограничением доступа к более конфиденциальной информации, такой как весовые коэффициенты конкретной модели ИИ [14]. Кроме того, нераскрытие информации об обучающих данных ИИ может причинить вред, затрудняя регулирующим органам и третьим лицам выявление потенциально вредоносного или дискриминационного поведения, возникающего на основе этих данных.

Таким образом, официальные аргументы против прозрачности в значительной степени неубедительны; широко распространено мнение, что основной причиной растущей непрозрачности данных для обучения является желание разработчиков ИИ избежать или минимизировать ответственность за нарушение авторских прав, содержащихся в данных для обучения. Расследование, проведенное газетой Washington Post в апреле 2024 года, показало, что многие компании, занимающиеся разработкой ИИ, даже не ведут внутренние записи своих данных для обучения из-за опасений, что это может быть использовано в качестве доказательства нарушения авторских прав или закона о защите данных [15].

Существует несколько подходов, которые предоставляют разный уровень информации о содержании рассматриваемых данных [16]. Подход, обеспечивающий максимальную степень прозрачности, заключается в том, что разработчики ИИ делают наборы обучающих данных полностью общедоступными. В рамках такого подхода с «полным доступом» третьи стороны (включая правообладателей) могут самостоятельно просматривать данные для обучения и проверять использованный контент. Некоторые разработчики уже используют этот подход, публикуя полные копии своих данных для обучения в интернете [17].

Однако предоставление полного доступа к обучающим данным вряд ли осуществимо для большинства моделей ИИ. Во-первых, существуют логистические проблемы, связанные с размещением доступного репозитория обучающего набора данных, содержащего сотни тысяч или миллионы отдельных работ. Во-вторых, такой подход, как ни парадоксально, может противоречить закону об авторском праве; даже если правообладатель согласился на включение своей работы в набор данных посредством лицензионного соглашения, он вряд ли будет рад, если его работы фактически станут свободно доступны другим разработчикам через полностью доступный набор данных. Аналогичные проблемы возникают и в отношении любых персональных данных, которые могут содержаться в наборе данных.

Один из способов избежать этих проблем — разрешить пользователям запрашивать доступ к ограниченным частям наборов данных, который затем может быть одобрен или отклонен владельцем набора данных после предоставления и проверки учетных данных. Это называется «закрытый доступ». Однако создание и поддержка инфраструктуры, необходимой для управления закрытым доступом к набору данных, усугубляет и без того значительные логистические проблемы, связанные с полным доступом.

Большинство дискуссий о повышении прозрачности обучающих данных сосредоточены на предоставлении некоего резюме, содержащего ключевую информацию о наборе данных. Это гораздо менее обременительно, чем предоставление прямого доступа к содержимому набора данных, но его полезность во многом зависит от информации, содержащейся в резюме [18]. Теоретически такое резюме может содержать метаданные по каждому элементу набора данных, например, название, URL (если применимо), автора, дату публикации и т. д., что позволяет идентифицировать отдельные работы. Однако такие данные часто неточны или отсутствуют, особенно в случае данных, полученных непосредственно из интернета. Предоставление или обеспечение точности даже базовой информации, такой как название или автор для отдельных элементов, потребует значительных ресурсов и вытеснит мелких разработчиков, увеличив концентрацию рынка. В сообществе ИИ уже разработан ряд фреймворков для предоставления подробных резюме обучающих данных без перечисления отдельных элементов, например, The Dataset Nutrition Label и Datasheets for Datasets.

Подавляющее большинство контента, используемого для обучения большинства моделей генеративного ИИ, будет защищено законом об авторском праве. Это неизбежное следствие того, что авторское право автоматически возникает для любой работы, отвечающей минимальному набору требований, и что срок этой защиты длителен — в России, согласно ст. 1281 ГК РФ, в течение жизни автора плюс 70 лет [19]. Хотя существуют альтернативы [20] использованию работ, поднимающих вопросы авторского права, они не являются жизнеспособной заменой: материалы, находящиеся в общественном достоянии по истечению срока действия авторских прав, датируются периодом до 1950-х годов — следовательно, любой ИИ, обученный исключительно на таких работах, будет безнадежно устаревшим.

Некоторые материалы, защищенные авторским правом, такие как Википедия, лицензированы по лицензии Creative Commons (CC) или другим лицензиям «копилефт», которые разрешают использование этого контента для обучения моделей генеративного ИИ. Однако, поскольку лишь небольшая часть контента, защищенного авторским правом, доступна по таким лицензиям, маловероятно, что передовые генеративные модели ИИ могут обучаться исключительно на нем. Кроме того, выходные данные моделей ИИ, обученных с использованием такого контента, могут быть ограничены обязательствами «Share Alike».

Поскольку генеративные модели ИИ, основанные на глубоком обучении, не хранят копии данных для обучения, многие ученые-комментаторы пришли к выводу, что в большинстве случаев генеративные модели ИИ не могут считаться нарушающими авторские права на произведения, на которых они были обучены [21]. Однако эта точка зрения осложняется явлением запоминания, при котором генеративные модели ИИ иногда могут воспроизводить дословно

или почти дословно части своих данных для обучения. Бюро по авторским правам США в своем отчете указало, что в случаях, когда результаты, генерируемые ИИ, существенно схожи с входными данными для обучения, существует «веский аргумент» в пользу того, что сами веса моделей нарушают права на воспроизведение и создание производных работ, основанных на оригинальных работах [22].

Несанкционированное воспроизведение работ, защищенных авторским правом, составляет основу обвинений против разработчиков ИИ в большинстве текущих судебных разбирательств, связанных с данными для обучения [23]. Защита разработчиков ИИ от обвинений в массовом нарушении прав заключается в исключениях и ограничениях из защиты авторских прав. Однако сфера применения этих исключений и ограничений существенно различается в разных юрисдикциях, что влияет на потенциальное воздействие введения требований к прозрачности данных для обучения [24].

В действующем российском законодательстве отсутствуют нормы, которые прямо разрешали бы использовать чужие произведения для обучения систем ИИ. Ст. 1274 ГК РФ не предусматривает подобных исключений, а перечень способов использования, закрепленный в ст. 1270 ГК РФ, не охватывает сам процесс обучения моделей. Иными словами, любое действие, которое может быть квалифицировано как использование произведения по смыслу гражданского законодательства, потенциально влечет нарушение исключительных прав [25].

К таким действиям относится не только открытое воспроизведение чужого произведения, но и его фиксация в цифровой форме: запись на носитель, сохранение в памяти ЭВМ, размещение в репозитории владельца ИИ-сервиса. Уже на этом этапе возникает риск признания факта незаконного использования объекта авторского права. Еще более очевидное нарушение возникает в момент, когда сервис начинает выдавать результаты, воспроизводящие чужие произведения полностью или фрагментарно, либо создает их переработанные версии [26].

Тем не менее, на практике вероятность привлечения к ответственности до того момента, как ИИ-сервис начнет предоставлять пользователям конкретные результаты генерации, остается крайне низкой. С одной стороны, обучение как процесс не имеет четкой правовой квалификации в рамках ст. 1270 ГК РФ, а с другой — доказать сам факт использования произведений без согласия правообладателя на этапе внутреннего обучения почти невозможно.

Таким образом, если ИИ воспроизводит индивидуализирующие и узнаваемые элементы исходного произведения, то возникает правовое основание для квалификации такого результата как нарушения исключительных авторских прав (ст. 1255 ГК РФ). Узнаваемость элементов рассматривается в правовой доктрине и судебной практике как ключевой критерий при отграничении самостоятельного произведения от производного (ст. 1260 ГК РФ).

Дальнейшее распространение результата, содержащего указанные элементы, в цифровой среде (например, посредством размещения на интернет-платформах) подпадает под понятие доведения произведения до всеобщего сведения (ст. 1270 ГК РФ). В отсутствие согласия правообладателя такие действия образуют нарушение его исключительных прав и влекут предусмотренные законом последствия — как в виде гражданско-правовой ответственности, так и, в отдельных случаях, административной либо уголовной (ст. 1301, 1252 ГК РФ, ст. 146 УК РФ).

Литература

1. Zewe A. *How do powerful generative AI systems like ChatGPT work, and what makes them different from other types of artificial intelligence?* (2023). MIT Schwarzman College of Computing Massachusetts Institute of Technology. Available at: https://computing.mit.edu/news/explained-generative-ai/ (accessed 30.08.2025).

2. Турп Э. Потенциал и опасность искусственного интеллекта [Электронный ресурс] // Международный валютный фонд. – 2023. – URL: https://www.imf.org/ru/Publications/fandd/issues/2023/12/B2B-Artificial-Intelligence-promise-peril-Tourpe (дата обращения: 30.08.2025).

3. Овчинников И.В. Генеративный искусственный интеллект и авторское право // Журнал Суда по интеллектуальным правам. – 2025. – № 1 (47). – С. 74-82. – DOI: 10.58741/23134852_2025_1_1

4. Авторы роликов на YouTube подали в суд на OpenAI за обучение на них ИИ [Электронный ресурс] // РИА Новости. – 2024. – URL: https://ria.ru/20240806/openai-1964404112.html (дата обращения: 30.08.2025).

5. Thomson Reuters Enterprise Centre GMBH v. Ross Intelligence Inc. No. 1:20-cv-613-SB. Memorandum Opinion February 11, 2025. Available at: https://www.ded.uscourts.gov/sites/ded/files/opinions/20-613_5.pdf (accessed 30.08.2025).

6. Response to USCO Inquiry on Artificial Intelligence and Copyright. Docket No. 2023-6. Available at: file:///C:/Users/user/Downloads/COLC-2023-0006-9044_attachment_1.pdf (accessed 30.08.2025).

7. Lopatto E. *OpenAI Searches for an Answer to its Copyright Problems* (2024). The Verge. Available at: https://www.theverge.com/2024/8/30/24230975/openai-publisher-deals-web-search (accessed 30.08.2025).

8. Жиленков А.А. Сравнительный анализ систем глубокого обучения с подкреплением и систем обучения с учителем / А.А. Жиленков, А.А. Силкин, М.Ю. Серебряков, С.В. Колесова // Известия Тульского государственного университета. Технические науки. – 2022. – № 10. – С. 109-112. – DOI: 10.24412/2071-6168-2022-10-109-112

9. Стори В., Юэ В.Т., Чжао Д.Л., Лукьяненко Р. Генеративный искусственный интеллект: развивающиеся технологии, растущее влияние на общество и возможности для исследований / пер. В. Лукьяненко // Границы информационных систем. – 2025.

10. Emanuilov I., Margoni T. Forget me not: memorisation in generative sequence models trained on open source licensed code. *Zenodo*, 2024.

11. Brown Tom B. Language Models are Few-Shot Learners. *arXiv*, 2020, v4. DOI: 10.48550/arXiv.2005.14165

12. Бакшеев В.В. Принцип прозрачности искусственного интеллекта // Аграрное и земельное право. – 2024. – № 12 (240). – С. 154-156. – DOI: 10.47643/1815-1329_2024_12_154

13. OpenAI. GPT-4 Technical Report. *arXiv*, 2024, v6. DOI: 10.48550/arXiv.2303.08774

14. Смирнова М. Модели с открытым исходным кодом и модели с открытым весом [Электронный ресурс] // ФГУП «ГРЧЦ». – 2025. – URL: https://rdc.grfc.ru/2025/04/open_source_vs_open_weights/ (дата обращения: 30.08.2025).

15. Schaul K., Chen S.Y., Tiku N. *Inside the secret list of websites that make AI like ChatGPT sound smart* (2025). The Washington Post. Available at: https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/ (accessed 30.08.2025).

16. Андрада Г. Разновидности прозрачности: изучение свободы действий в системах ИИ / Г. Андрада, Р.В. Клоуз, П.Р. Смарт // KANT: Social Sciences & Humanities. – 2022. – № 1 (9). – С. 35-51. – DOI: 10.24923/2305-8757.2022-9.4

17. Liu Z. LLM360: Towards Fully Transparent Open-Source LLMs. *arXiv*, 2023. DOI: 10.48550/arXiv.2312.06550

18. Харитонова Ю.С. Правовые средства обеспечения принципа прозрачности искусственного интеллекта // Journal of Digital Technologies and Law. – 2023. – Т. 1, № 2. – С. 337-358. – DOI: 10.21202/jdtl.2023.14

19. Гражданский кодекс Российской Федерации (часть четвертая) от 18.12.2006 N 230-ФЗ (ред. от 23.07.2025) [Электронный ресурс] // СПС КонсультантПлюс. – URL: https://www.consultant.ru/document/cons_doc_LAW_64629/ (дата обращения: 30.08.2025).

20. Васильева А.С. Использование объектов авторских прав для обучения искусственного интеллекта // Журнал Суда по интеллектуальным правам. – 2024. – № 4 (46). – С. 80-84. – DOI: 10.58741/23134852_2024_4_8

21. Mallesons K., Mallesons W. *Navigating Copyright Challenges in AI Model Training: A Cross-Border Perspective*. Law Business Research. Available at: https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/ (accessed 30.08.2025).

22. Levi S.D., Ghaemmaghami M. *Copyright Office Weighs In on AI Training and Fair Use* (2025). Skadden. Available at: https://www.skadden.com/insights/publications/2025/05/copyright-office-report (accessed 30.08.2025).

23. Самарцева М.В. Использование охраняемых произведений при машинном обучении генеративного искусственного интеллекта // Труды по интеллектуальной собственности (Works on Intellectual

Property). – 2025. – Т. 53, № 2. – С. 127-138. – DOI: 10.17323/tis.2025.27374

24. Медведев А.И. Принцип прозрачности искусственного интеллекта в системе принципов российского права // Вестник Удмуртского университета. Серия Экономика и право. – 2023. – Т. 33, № 3. – С. 516-526. – DOI: 10.35634/2412-9593-2023-33-3-516-526

25. Ходаков И. Три вопроса к легальному использованию нейросетей/ИИ: обучение, применение и охрана // ЭЖ-Юрист. – 2024. – № 21 (1322).

26. Свиридова Е.А. О правомерности использования произведений при применении технологий искусственного интеллекта // Проблемы экономики и юридической практики. – 2024. – Т. 20, № 1. – С. 70-75.

Acceptability of using intellectual property objects in training neural networks

Tamaev A.V.
Lebedev Russian State University of Justice

The article is devoted to the legal aspects of using intellectual property objects in training generative models of artificial intelligence. In the context of rapid technological changes associated with the introduction of generative AI, there is a need to balance between innovation and copyright protection. The main problem is the use of large amounts of data, many of which are protected by copyright, without the knowledge and consent of the copyright holders. Generative models trained on such data can accidentally "remember" protected content - reproduce it partially or completely in the output results, which leads to legal disputes and lawsuits. The article considers various approaches to data transparency for AI training, as well as legal norms that regulate the use of such data in Russia. Particular attention is paid to the lack of clear norms in Russian legislation, which creates legal uncertainty. The paper analyzes possible risks for AI developers and ways to minimize them by improving copyright legislation. The conclusion focuses on the need to form new legal norms that ensure a balance between innovation and intellectual property protection.

Keywords: artificial intelligence, neural network training, intellectual property, copyright, exclusive right.

References

1. Zewe A. How do powerful generative AI systems like ChatGPT work, and what makes them different from other types of artificial intelligence? (2023). MIT Schwarzman College of Computing Massachusetts Institute of Technology. Available at: https://computing.mit.edu/news/explained-generative-ai/ (accessed 30.08.2025).
2. Turp E. The potential and danger of artificial intelligence [Electronic resource] // International Monetary Fund. – 2023. – URL: https://www.imf.org/ru/Publications/fandd/issues/2023/12/B2B-Artificial-Intelligence-promise-peril-Tourpe (date of access: 08/30/2025).
3. Ovchinnikov I.V. Generative artificial intelligence and copyright // Journal of the Intellectual Property Rights Court. – 2025. – № 1 (47). – Pp. 74-82. – DOI: 10.58741/23134852_2025_1_1
4. The authors of YouTube videos sued OpenAI for training AI on them [Electronic resource] // RIA Novosti. - 2024. – URL: https://ria.ru/20240806/openai-1964404112.html (accessed: 08/30/2025).
5. Thomson Reuters Enterprise Centre GMBH v. Ross Intelligence Inc. No. 1:20-cv-613-SB. Memorandum Opinion February 11, 2025. Available at: https://www.ded.uscourts.gov/sites/ded/files/opinions/20-613_5.pdf (accessed 08/30/2025).
6. Response to USCO Inquiry on Artificial Intelligence and Copyright. Docket No. 2023-6. Available at: file:///C:/Users/user/Downloads/COLC-2023-0006-9044_attachment_1.pdf (accessed 30.08.2025).
7. Lopatto E. OpenAI Searches for an Answer to its Copyright Problems (2024). The Verge. Available at: https://www.theverge.com/2024/8/30/24230975/openai-publisher-deals-web-search (accessed 30.08.2025).
8. Zhilenkov A.A., Silkin A.A., Serebryakov M.Yu., Kolesova S.V. Comparative analysis of deep learning systems with reinforcement and learning systems with a teacher // Proceedings of Tula State University. Technical sciences. – 2022. – No. 10. – pp. 109-112. – DOI: 10.24412/2071-6168-2022-10-109-112
9. Story V., Yue V.T., Zhao D.L., Lukyanenko R. Generative artificial intelligence: developing technologies, growing impact on society and research opportunities / trans. V. Lukyanenko // Frontiers of information systems. – 2025.
10. Emanuilov I., Margoni T. Forget me not: memorisation in generative sequence models trained on open source licensed code. Zenodo, 2024.
11. Brown Tom B. Language Models are Few-Shot Learners. arXiv, 2020, v4. DOI:10.48550/arXiv.2005.14165
12. Baksheev V.V. Principle of transparency of artificial intelligence // Agrarian and Land Law. – 2024. – № 12 (240). – С. 154-156. – DOI: 10.47643/1815-1329_2024_12_154
13. OpenAI. GPT-4 Technical Report. arXiv, 2024, v6. DOI: 10.48550/arXiv.2303.08774
14. Smirnova M. Open source models and open weight models [Electronic resource] // FSUE "GRCHTS". – 2025. – URL: https://rdc.grfc.ru/2025/04/open_source_vs_open_weights / (accessed: 08/30/2025).
15. Schaul K., Chen S.Y., Tiku N. Inside the secret list of websites that make AI like ChatGPT sound smart (2025). The Washington Post. Available at: https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/ (accessed 30.08.2025).
16. Andrada G. Varieties of transparency: the study of freedom of action in AI systems / G. Andrada, R.V. Close, P.R. Smart // KANT: Social Sciences & Humanities. – 2022. – № 1 (9). – Pp. 35-51. – DOI: 10.24923/2305-8757.2022-9.4
17. Liu Z. LLM360: Towards Fully Transparent Open-Source LLMs. arXiv, 2023. DOI:10.48550/arXiv.2312.06550
18. Kharitonova Yu.S. Legal means of ensuring the principle of transparency of artificial intelligence // Journal of Digital Technologies and Law. – 2023. – Vol. 1, No. 2. – pp. 337-358. – DOI: 10.21202/jdtl.2023.14
19. The Civil Code of the Russian Federation (part four) of 12/18/2006 N 230-FZ (as amended on 07/23/2025) [Electronic resource] // SPS ConsultantPlus. – URL: https://www.consultant.ru/document/cons_doc_LAW_64629 / (date of access: 08/30/2025).
20. Vasilyeva A.S. The use of copyright objects for artificial intelligence training // Journal of the Intellectual Property Rights Court. – 2024. – № 4 (46). – Pp. 80-84. – DOI: 10.58741/23134852_2024_4_8
21. Mallesons K., Mallesons W. Navigating Copyright Challenges in AI Model Training: A Cross-Border Perspective. Law Business Research. Available at: https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/ (accessed 30.08.2025).
22. Levi S.D., Ghaemmaghami M. Copyright Office Weighs In on AI Training and Fair Use (2025). Skadden. Available at: https://www.skadden.com/insights/publications/2025/05/copyright-office-report (accessed 30.08.2025).
23. Samartseva M.V. The use of protected works in machine learning of generative artificial intelligence // Works on Intellectual Property. – 2025. – Vol. 53, No. 2. – pp. 127-138. – DOI: 10.17323/tis.2025.27374
24. Medvedev A.I. The principle of transparency of artificial intelligence in the system of principles of Russian law // Bulletin of the Udmurt University. Economics and Law series. – 2023. – Vol. 33, No. 3. – pp. 516-526. – DOI: 10.35634/2412-9593-2023-33-3-516-526
25. Khodakov I. Three questions for the legal use of neural networks/AI: training, application and protection // EJ-Lawyer. – 2024. – № 21 (1322).
26. Sviridova E.A. On the legality of the use of works in the application of artificial intelligence technologies // Problems of economics and legal practice. - 2024. – Vol. 20, No. 1. – pp. 70-75.