

СТАТЬИ

КОПИРОВАТЬ НЕЛЬЗЯ ОБУЧАТЬ: ПРОБЛЕМА ОБУЧЕНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА С ПОЗИЦИЙ АВТОРСКОГО ПРАВА

А. А. Никифоров^{1,2}

¹Московская высшая школа социальных и экономических наук
(МВШСЭН)

125009, Россия, Москва, Газетный пер., 3-5/1

²Яндекс

119021, Россия, Москва, ул. Льва Толстого, 16

Аннотация

Что можно считать использованием в авторском праве? Охватывает ли исключительное право правообладателя любое взаимодействие с объектом? Настоящая статья посвящена исследованию правовых аспектов обучения искусственного интеллекта (ИИ) на объектах авторских и смежных прав. Целью исследования является комплексный правовой анализ обучения ИИ на охраняемых объектах авторского и смежного права с позиций российского и зарубежного законодательства, в первую очередь через толкование понятий «использование», «воспроизведение» и юридическую квалификацию таких действий, как интеллектуальный анализ данных (*text and data mining*). В статье также анализируются исключения и ограничения, предусмотренные правом ЕС, США и Японии и демонстрирующие различные модели правового баланса между интересами разработчиков ИИ и правообладателей. Методологически работа основывается на междисциплинарном подходе, сочетающем техническое описание алгоритмов обучения нейросетей с догматическим и сравнительно-правовым исследованием подходов к обучению ИИ и интеллектуальному анализу данных в разных странах. При этом при подготовке статьи для редактуры, корректуры и повышения ясности текста был использован *ChatGPT*¹, однако все мысли, идеи, примеры и выводы являются чистым результатом работы автора и не сгенерированы. Также рассматриваются политико-правовые аргументы за и против свободного обучения ИИ на объектах авторского права. В результате проведенного исследования автор приходит к выводу, что процесс обучения ИИ сам по себе не является использованием произведения в смысле ст. 1270 ГК РФ, поскольку не связан с воспроизведением охраняемой формы произведения и не приводит к непосредственному восприятию произведения

¹ В соответствии с Политикой публикации Американской психологической ассоциации необходимо раскрывать использование нейросетей даже в случае, если они были применены чисто для редакционных целей, без заимствования результатов исследовательских генераций. См.: American Psychological Association. (n.d.). *Publishing policies*. <https://www.apa.org/pubs/journals/resources/publishing-policies>. На первом этапе был подготовлен расширенный черновик, объем которого превышал итоговую версию в полтора-два раза. Далее отдельные фрагменты текста передавались ИИ для улучшения логики изложения, повышения читаемости и упрощения синтаксиса. Особое внимание уделялось преобразованию чрезмерно длинных предложений: в первоначальной версии некоторые из них занимали почти целую страницу и напоминали упражнения по синтаксису из учебника русского языка.

человеком или его функциональной эксплуатации («впечатляющее использование»). При этом в право-порядке целесообразно предусмотреть исключения, позволяющие создавать временные копии произве-дений без согласия правообладателя, если это необходимо для интеллектуального анализа данных. Также важно установить механизмы, которые сделали бы возможным использование закрытых для обучения данных без необходимости вести переговоры с каждым правообладателем. Исключением будет ситуация обучения на базах данных, которые специально были собраны, подготовлены и обработаны правообла-дателями для целей обучения ИИ.

Ключевые слова

искусственный интеллект, авторское право, смежные права, обучение ИИ, интеллектуальный анализ данных

Конфликт интересов

Автор сообщает об отсутствии конфликта интересов.

Финансирование

Исследование не имеет спонсорской поддержки.

Для цитирования

Никифоров, А. А. (2025). Копировать нельзя обучать: проблема обучения искусственного интеллекта с позиций авторского права. *Цифровое право*, 6(1), 74–128. <https://doi.org/10.38044/2686-9136-2025-6-6>

Поступила: 25.01.2025, принята в печать: 24.02.2025, опубликована: 31.03.2025

ARTICLES

REPRODUCING OR DATA MINING: THE COPYRIGHT LAW DILEMMA OF AI TRAINING

Artem A. Nikiforov^{1,2}

¹Moscow School of Social and Economic Sciences (MSSES)
3-5/1, Gazetny Lane, Moscow, Russia, 125009

²Yandex
16, Lev Tolstoy St., Moscow, Russia, 119021

Abstract

What constitutes “use” under Copyright Law? Does the exclusive right of the copyright holder encompass any interaction with a protected work? This article explores the legal dimensions of training artificial intelligence (AI) based on works protected by copyright and related rights. The aim of this study is to conduct a comprehensive legal analysis of AI training based on protected subject matter, focusing on the interpretation of key terms such as “use”, “reproduction”, and the legal qualification of activities such as text and data mining, within both Russian and foreign legal systems. The article examines the relevant statutory exceptions and limitations provided under EU, U.S., and Japanese law, illustrating divergent models of legal balance between the interests of AI developers and copyright holders. Methodologically, the research adopts an interdisciplinary approach, combining a technical description of neural network training algorithms with doctrinal and comparative legal

analysis of regulatory approaches to AI training and text and data mining across jurisdictions. During the editing and proofreading stages, ChatGPT was used to improve clarity and coherence. However, all ideas, reasoning, examples, and conclusions are entirely the author's own and were not generated by AI. The article further engages with normative and policy-based arguments for and against permitting AI systems to train freely based on copyrighted content. As a result of the analysis, the author concludes that the act of training an AI model, in itself, does not constitute “use” of a work within the meaning of Article 1270 of the Russian Civil Code. This is because such training does not involve reproduction of the protected expression of the work, nor does it entail perceptible access by a human or functional exploitation of the work (i.e., expressive use). Nevertheless, it is advisable for the legal system to establish exceptions which allow the creation of temporary copies of works without the right holder's consent, when such copying is necessary for legitimate text and data mining purposes. Additionally, the law should provide mechanisms which enable the use of data that is otherwise restricted for training, without requiring individual negotiations with every rights holder. An exception to this rule should apply to databases which have been specifically curated, structured, and prepared by rights holders for the purpose of AI training.

Keywords

artificial intelligence, copyright, related rights, AI training, text and data mining

Conflict of interest The author declares no conflict of interest.

Financial disclosure The study has no sponsorship.

For citation Nikiforov, A. A. (2025). Reproducing or data mining: The copyright law dilemma of AI training. *Digital Law Journal*, 6(1), 74–128. <https://doi.org/10.38044/2686-9136-2025-6-6>

Submitted: 25 Jan. 2025, accepted: 24 Feb. 2025, published: 31 Mar. 2025

Введение

Искусственный интеллект — это «новый черный» в юриспруденции, особенно в сфере интеллектуальной собственности. По влиянию на право его появление сравнимо с такими вехами, как изобретение Гутенбергом книгопечатного станка или создание первых компьютерных программ. Эти технологические достижения не только изменили правовую систему, но и способствовали развитию новых доктрин.

Поэтому неудивительно, что сегодня множество статей, конференций и круглых столов посвящены вопросам права искусственного интеллекта. Основное внимание уделяется тому, как существующее законодательство регулирует отношения, связанные с искусственным интеллектом (далее также — ИИ), и требуется ли его адаптация.

Большинство российских исследований сосредоточено на анализе выходных данных ИИ (*output*): охраняются ли результаты его работы авторским правом и кто несет ответственность за них². Однако вопросы обучения ИИ (*input*) и связанные с этим юридические аспекты в доктринальных обсуждениях пока практически не затрагиваются, а если и попадают в фокус,

² См. широкий круг источников: Abramova & Khamidullina, 2024; Anikin, 2022; Borisova, 2024; Entin, 2017; Gurko, 2017, 2024; Kalyatin, 2024; Kartskhiya, 2024; Kharitonova, 2018; Kirsanova, 2023; Kozlova, 2024; Morkhat, 2018; Orlova, 2022; Rakhmatulina, 2025; Rolinson et al., 2018; Roznina, 2021; Safin et al., 2018; Shpakovskaya, 2023; Vitko, 2019a, 2019b; Vorozhevich, 2025.

то с применением большого количества антропоморфных и иных риторических метафор, вместо точного технического и юридического анализа (Murray, 2025). Настоящая статья стремится восполнить этот пробел, хотя бы для отечественной доктрины, ввести ключевые понятия и обозначить правовые проблемы обучения ИИ. Для этого она будет разделена на следующие части.

1. Технические особенности работы и обучения генеративной нейросети.

С учетом того что любой существующий на данный момент искусственный интеллект — это по сути программа для ЭВМ, пусть и более сложная по сравнению с другими, технологические особенности диктуют правовое осмысление. Невозможно понять, когда происходит нарушение прав на программу, не разобравшись в категории исходного кода, и невозможно корректно применить положения открытых лицензий (*open source*) без понимания, что такое статическая и динамическая компоновка такого кода.

2. Формально-юридический анализ обучения ИИ с точки зрения гражданского права.

Несмотря на отсутствие прямого и однозначного ответа на вопрос о правомерности обучения искусственного интеллекта на объектах авторского права, действующее законодательство и правоприменительная практика создают предпосылки для формирования обоснованной позиции посредством системного толкования. Анализируя нормы авторского права в совокупности с исключениями, ограничениями и целями правовой охраны, можно сформулировать правовую оценку допустимости использования охраняемых произведений в процессе машинного обучения, даже при отсутствии специальных положений, прямо регулирующих данную сферу.

3. Анализ мировых подходов к обучению ИИ.

Третья часть будет посвящена зарубежной практике, в частности примеру Евросоюза, США и Японии, где уже неоднократно поднималась проблема влияния новых технологий на интересы авторов и общества.

4. Политико-правовые аргументы.

Четвертая часть статьи включает анализ ключевых политико-правовых аргументов, выдвигаемых как в поддержку, так и против идеи свободного обучения искусственного интеллекта на объектах авторского права и смежных прав. В центре внимания окажутся вопросы баланса между стимулированием инноваций и сохранением справедливой правовой охраны творческого труда, а также соотношение публичных и частных интересов в цифровую эпоху.

1. Технические особенности работы и обучения генеративной нейросети

Для начала нужно договориться о терминах. Так, в дальнейшем под искусственным интеллектом (ИИ), нейросетями будет иметься в виду генеративный искусственный интеллект вроде *ChatGPT*, *Midgorney* или *GitHub Copilot*³.

³ Генеративный ИИ — это общее название для массивной экосистемы слабо связанных технологий. В их число входят разговорные текстовые чат-боты, такие как *ChatGPT*, генераторы изображений, в частности *Midjourney* и *DALL-E*, помощники по кодированию, например *GitHub Copilot*, и системы, которые сочиняют музыку, создают идеи и предлагают молекулы для новых медицинских препаратов. Модели генеративного ИИ имеют различные технические архитектуры и обучаются на разных видах и источниках данных с помощью различных алгоритмов. Обучение одних занимает месяцы и стоит миллионы долларов, другие можно создать за выходные. Эти модели также приобретаются пользователями разными способами. Одни предлагаются через платные онлайн-сервисы, другие распространяются с открытым исходным кодом, чтобы любой мог скачать и изменить их (Lee et al., 2024).

Разберемся, что именно подразумевается под процессом генерации ответа и обучения⁴.

1.1. Описание процесса генерации ответа

Как нейросеть создает ответ? Лучше всего это можно понять из сравнения ИИ с интернет-поисковиком. В отличие от обычного поисковика, который просто ищет информацию по ключевым словам и может лишь «угадывать» запрос на основе истории поиска и популярных запросов, генеративный ИИ в виде чат-бота, как следует из названия, может вести диалог и генерировать идеи и тексты по запросу. Его революционность заключается в способности не только составлять формально «осмысленные» предложения, но и учитывать предыдущие диалоги. После получения запроса нейросеть не извлекает похожие фрагменты текста из базы данных, а каждый раз генерирует новый ответ, исходя из контекста. ИИ анализирует вес каждого слова в предложении и их порядок, что помогает ему «понимать» смысл (Garbacea & Mei, 2020).

Процесс генерации ответа можно описать следующим образом.

1. Токенизация текста: исходный запрос разбивается на токены (минимальные смысловые единицы), например слова в единственном числе и именительном падеже (Gastaldi et al., 2024).

Нейросеть разбивает текст на отдельные слова и превращает их в токены, минимальные смысловые единицы. Например, запрос «Напиши мне список самых известных женщин в истории науки» разбивается на токены вроде «на», «пиши», «список», «женщин», «история» и т. д. Затем ИИ удаляет лишние части, чтобы уменьшить величину запроса, а также определяет части речи. В итоге остается что-то вроде «список (существительное), самый (прилагательное), известный (прилагательное), женщина (существительное), наука (существительное)». Такой подход помогает ускорить процесс генерации ответа.

2. Преобразование токенов в векторы: каждому токеноу соответствует числовое представление, позволяющее анализировать связи между словами⁵.

Токены преобразуются в векторы — числовые наборы, которые нейросеть заранее назначает каждому слову в процессе обучения. По сути, нейросеть создает связи между словами на основе того, насколько осмысленно они могут быть использованы вместе. Например, токены «мама», «родить» и «дочь» будут находиться близко друг к другу, а «мама» и «люминесценция» — далеко.

3. Алгоритмическая обработка: векторы проходят через множество алгоритмов, которые оценивают их значение в контексте запроса.

Далее эти векторы проходят через внутренние алгоритмы ИИ, которые строят черновые версии будущего ответа. Это можно представить как процесс, в котором слова пропускаются через множество фильтров, чтобы в итоге создать цельную фразу. На этом этапе ИИ взвешивает значение каждого слова и сопоставляет их между собой, чтобы лучше понять, как они связаны и какой контекст создается.

4. Генерация ответа: ИИ поочередно формирует слова ответа, проверяя их соответствие запросу и контексту (Khoshnoodi et al., 2024).

В конце нейросеть начинает последовательно генерировать каждое слово, учитывая все предыдущие этапы: введенный запрос, векторные представления слов, знания из базы данных

⁴ Для более глубокого погружения в технические вопросы см.: Bloch (2019), Drexler et al. (2019); Sharifani & Amini (2023), Zhao et al. (2023); Hadi et al. (2025).

⁵ См. подробнее: NickyP. (2023). LLM Basics: Embedding Spaces — Transformer Token Vectors Are Not Points in Space. <https://lesswrong.com/posts/pHPmMGEMYefk9JLeh/llm-basics-embedding-spaces-transformer-token-vectors-are>

и другие факторы. Важно, что после добавления каждого слова ИИ проверяет, насколько оно соответствует исходному запросу, предыдущим словам и возможным следующим словам. Этот процесс продолжается до тех пор, пока нейросеть не решит, что текст достаточно подходит для ответа на вопрос.

Таким образом, генерация ответов нейросетью чем-то напоминает процесс формирования человеческой речи: шаг за шагом, слово за словом, пусть и зависит от заранее определенных алгоритмов. Благодаря усилиям разработчиков и людей, участвовавших в обучении ИИ, современный ИИ научился распознавать не только грамматику и пунктуацию, но даже форматирование, фонетические особенности и использование слов в различных контекстах.

1.2. Описание технического процесса обучения ИИ

Как видно из описания процесса генерации ответов ИИ, ему не требуется доступ к оригинальным произведениям для работы. Нужна не сама охраняемая форма, а скорее то, что находится за этой формой произведения. ИИ необходимы метаданные, такие как статистические закономерности, паттерны, принципы и методы, на основе которых создаются произведения. Иными словами, обучаясь, генеративный искусственный интеллект не загружает себе в память целую базу копий произведений, используемых для обучения, а раскладывает эти произведения на составляющие для формирования абстрактных данных, из которых в конце концов и получаются веса ИИ, т. е. абстрактные закономерности, необходимые для генерации ответов. Это можно сравнить с обучением человека: мы изучаем методы, идеи и принципы, не запоминая дословно каждую страницу книги. Достаточно понять идеи, технические приемы и методы, которые использовались для их создания⁶.

Нейросети, как и люди, могут обучаться двумя способами: с учителем (*supervised learning*) и без него (*unsupervised learning*).

Обучение с учителем — это тип машинного обучения, при котором нейросеть обучается на размеченных данных, т. е. на данных, где входные данные и правильные ответы заранее известны (Zhang et al., 2023; Gunel et al., 2020).

Предположим, у нас есть набор данных с текстами, и каждый текст уже имеет метку, показывающую его категорию, например предложения:

Входные данные (текст): «Юрист пишет».

Правильный ответ (метка): «Исковое заявление».

Нейросеть будет обучаться на таких примерах, чтобы правильно предсказывать ответ для нового текста, основываясь на уже известных примерах.

Обучение без учителя — это тип машинного обучения, при котором нейросеть работает с неразмеченными данными, т. е. не содержащими меток или классов (Ge et al., 2023)⁷. В отличие от обучения с учителем, где данные имеют известные метки (например, изображение юриста, пишущего исковое заявление, и человека, пишущего картину), в этом случае нейросеть должна самостоятельно находить структуру и закономерности в данных.

⁶ При этом, конечно, сами процессы не идентичны: обучение человека — это более сложный и неисследованный процесс, чем обучение ИИ.

⁷ См. также: Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners [Technical report]. OpenAI. https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

На протяжении обучения параметры нейросети (веса или векторы) обновляются так, чтобы лучше отражать структуру данных, но не сами данные⁸.

Именно потому, что ИИ не зависит от прямого воспроизведения исходных данных, он может имитировать диалог с человеком по любому вопросу. После того как модель усваивает внутреннюю структуру данных, она может использовать эту структуру для создания ответов на другие вопросы и в различных областях. Однако это является как преимуществом, так и ограничением любого ИИ. Будучи ограниченным структурой данных, на которой он обучался, ИИ может применять эту структуру и к другим ситуациям, что приводит к ошибкам и различным смешным генерациям. Несмотря на то что современные ИИ могут имитировать понимание контекста, они на самом деле не могут истинно понимать его или придавать ему значимость, как это способен делать человек⁹.

Следует понимать, что процесс обучения искусственного интеллекта практически никогда не завершается на этапе общего (*pretraining*) обучения. После него нередко следует этап

⁸ Это можно проиллюстрировать следующим образом: у нас есть коллекция текстов или документов, но каждый из них не имеет ярлыка или категории (например, не указано, о каком типе текста идет речь). Нейросеть будет анализировать эти тексты, чтобы выявить скрытые закономерности, такие как схожесть между словами или фразами. Во время обучения нейросеть обновляет свои параметры (веса или векторы), чтобы лучше отразить структуру данных. Однако нейросеть учится не на самих данных или метках данных, а на структуре и взаимосвязях между объектами. В процессе обучения она пытается определить, какие элементы данных могут быть сгруппированы, какие являются аномальными или какие признаки наиболее важны для указанных данных.

⁹ В этом контексте всё, что генерирует нейросеть является «бессмысленным». Несмотря на то, что тот же ChatGPT выглядит, будто он может вести осмысленный диалог или понимает реальный смысл того, что пишет в ответах, на самом деле нет. И дело не в том, что «это всего лишь машина, неспособная написать симфонию», и не в навязном биологическом шовинизме — вере в то, что только биологические существа (или даже только люди) могут обладать сознанием и его субститутами вроде интеллекта или возможности работы с абстрактными смыслами. Вероятно, когда-нибудь появится настоящий искусственный интеллект, обладающий самосознанием, имеющий свободу воли и понимающий происходящее в той же степени, в которой сейчас эти элементы имеют средние разумные люди. Но ChatGPT пока не такой. Он не просто не понимает, что говорит, но и является неиссякаемым генератором откровенной чуши. В ответах нейросети не просто нет смысла: он там отсутствует даже на уровне концепции, подменяясь внешними признаками осмысленности. Российский философ Антон Беседин приводит идеальный пример для понимания: чат не умеет играть в игру «Покажи и назови». «Представьте, что вы столкнулись с человеком из другого времени, культуры, с другой планеты. Вы начинаете беседу на абстрактную тему — например, высказываете свои соображения об инфляции — и вдруг обнаруживаете, что ваш визави совершенно не понимает, о чем речь. И вам приходится с уровня экономической теории спускаться на уровень тыканья пальцем: “Это деньги, это товар. Я тебе — ты мне”. ChatGPT похож на такого инопланетянина: не будучи в состоянии показать, где деньги, а где товар, он научился ловко имитировать разговоры на сложные темы» (Беседин, А. (2023). О чем говорит ChatGPT. *Эксперт*, (8), 13). ChatGPT не работает с реальным миром. Он имеет дело с текстами, на которых обучался. У него на данный момент нет даже теоретической возможности соотносить свои тексты с реально существующими вещами. И из-за этого, кроме понятий «смысл» и «значение», для чат-бота отсутствуют такие важные категории для осмысленного и серьезного разговора, как «правда» или «ложь». Их просто не существует внутри алгоритма, да и непонятно, как их туда заложить. Поэтому он может сгенерировать в ответе несуществующие книги и даже законы либо давать откровенно вредные советы. Не потому, что он «злой», а просто потому, что алгоритм решил, что цепочка выстроенных слов лучше всего подойдет для тех слов, которые ввели в качестве запроса. И если человека можно припереть к стенке и спросить: «Это действительно так?», после чего ответ — как да, так и нет — будет не просто звуками, но социальным действием по взятию на себя ответственности за сказанное, то с чат-ботом такое не пройдет: он, как деревенский дурачок, уличенный во лжи, лишь улыбнется и сделает вид, что ничего не было. Разве только в письменном виде и более вежливо.

специализированного дообучения (*fine-tuning*) — адаптации модели к конкретной задаче, стилю или предметной области. При этом модель сохраняет общие языковые и логические паттерны, приобретенные ранее, но дополнительно уточняет свои алгоритмы на основе паттернов целевого корпуса данных. Как правило, *fine-tuning* реализуется в форме обучения с учителем (*supervised learning*), где модели предъявляются пары «ввод — желаемый ответ», отдельные методы которого могут снижать способность модели к обобщению, усиливая вероятность воспроизведения фрагментов обучающей выборки или прививая предвзятость в отношении конкретных формулировок и ответов (см., например: Li et al., 2025; Chen et al., 2025). Однако в случае узкоспециализированных моделей, разработанных для решения конкретных задач, подобные техники целевого обучения могут существенно влиять на качество, этичность и правовую нейтральность результатов. Следует сразу уточнить, что то, что называется «запоминанием» модели, даже в случае большинства вариантов дообучения фактически является необычно высоковероятностями последовательностями (при этом уже вероятность в 0,5% или даже ниже может считаться необычно высокой) (Cooper, Gokaslan, et al., 2025, pp. 22–21). Поэтому даже тогда нельзя говорить о фактическом воспроизведении работ в памяти моделей на этапах до генерации. Исключением здесь является совмещение генеративной модели ИИ с *Retrieval-Augmented Generation (RAG)* — архитектурой для генерации текста, в которой языковая модель дополняется модулем поиска. Вместо того чтобы полагаться только на свою «память» (внутренние параметры), модель во время генерации динамически извлекает информацию из внешней базы знаний и использует ее для формирования более точных, обоснованных и актуальных ответов (Lewis et al., 2020; Izacard & Grave, 2021). В этом случае дообучение модели может осуществляться для того, чтобы модель лучше классифицировала и находила информацию во внешней базе данных и выводила ее пользователю по запросу.

Для улучшения работы нейросетей и их выходных данных применяются различные алгоритмы. Приведу несколько из них.

Градиентный спуск (*Gradient Descent*) — это основной способ уменьшения ошибок при генерации ответа (Zhang, 2024). Под градиентом понимается математическая величина, отражающая то, насколько предсказания нейросети отличаются от реальных фактов. В процессе обучения нейросеть строит предсказания, и последние сравниваются с реальными значениями. Разница между ними — это ошибка. Например, если нейросеть пытается предсказать цену товара, ошибка — это разница между предсказанной и реальной ценами. Градиентный спуск указывает направление изменения весов нейросети, при котором ошибка уменьшается или увеличивается¹⁰.

Обратное распространение (*Backpropagation*) — метод, который помогает нейросети учиться на своих ошибках (Li, 2024). Когда сеть делает ошибку, она «отматывает» ее назад через свои уровни и корректирует настройки (веса), чтобы в следующий раз справиться лучше¹¹. Обычно используется вместе с градиентным спуском.

¹⁰ Чтобы лучше понять, как это работает, давайте представим следующее: вы стоите на вершине горы, и ваша цель — спуститься в долину. Градиент — это то, что показывает, в каком направлении и с какой силой нужно двигаться, чтобы быстрее всего спуститься. Градиентный спуск работает также: он направляет модель к минимальной ошибке.

¹¹ Продолжаем визуализировать разные ситуации, чтобы лучше понимать, как работают различные техники. Представьте, что нейросеть — это ребенок, который учится писать цифры. Сначала он ошибается, рисуя цифры не совсем правильно. Когда видит свою ошибку (например, «3» вместо «8»), ребенок «отматывает» события и думает, что именно в процессе написания было неправильным. Он корректирует свои действия, чтобы в следующий раз изобразить цифру правильно.

Адаптивные методы оптимизации (*Adaptive Optimization Methods*). Так называются более умные способы обучения. Алгоритмы вроде *Adam*¹² или *RMSprop*¹³ могут менять скорость обучения¹⁴ в зависимости от ситуации, помогая сети быстрее находить решения¹⁵.

Стохастический градиентный спуск с импульсом (*SGD with Momentum*) представляет собой улучшение стандартного метода. В классическом градиентном спуске веса модели обновляются, двигаясь в направлении минимизации ошибки. Однако на практике в больших наборах данных используется стохастический градиентный спуск (*SGD*), который обновляет веса не на основе всего набора данных, а на основе случайно выбранных подмножеств данных. Это ускоряет обучение, но оно может стать нестабильным, так как каждый шаг может быть немного «шумным» и не всегда направлен к оптимальному решению.

Метод стохастического градиентного спуска с импульсом решает проблему нестабильности, добавляя «память» о предыдущих обновлениях. Вместо того чтобы делать шаги в случайном направлении (как при обычном *SGD*), он учитывает предыдущие шаги, добавляя некоторую инерцию или импульс, чтобы сделать обучение более стабильным и быстрым¹⁶.

Таким образом, нейросеть представляет собой сложную систему взаимосвязей между минимальными единицами данных (например, словами в тексте или пикселями в изображении), но не хранит оригинальные произведения в структурированном виде. Как видно из алгоритмов улучшения работы нейросетей, после загрузки исходных закономерностей эти данные не могут быть использованы в «голом виде». ИИ необходимо проделать еще множество этапов работы, которые зависят не только от исходных данных, но и от конкретной нейросети, ее внутренних алгоритмов, а также от количества ошибок, выявленных в процессе генерации.

¹² *Adam* — это один из самых популярных методов, который комбинирует два других подхода: *Momentum* (инерция) и *RMSprop* (адаптивная коррекция скорости обучения). В *Adam* вычисляются два момента: первый (среднее значение градиента) помогает сглаживать движения в процессе обучения. Второй момент (средний квадрат градиентов), который позволяет алгоритму изменять скорость обучения в зависимости от величины градиента. Это делает *Adam* быстрым и стабильным методом, особенно в сложных и многослойных нейросетях, где стандартный градиентный спуск может быть слишком медленным или нестабильным (Zhang, 2019).

¹³ Этот метод также изменяет скорость обучения, но делает это на основе среднего квадрата градиентов. Он помогает нейросети избегать проблем с большими колебаниями градиентов, которые могут возникать при обучении сложных моделей. *RMSprop* хорош для задач, где нужно стабилизировать обучение в условиях изменяющихся или нестабильных градиентов. Он уменьшает проблемы с большими шагами обновления весов и помогает быстрее сходиться старым и новым данным (Zhang, 2019).

¹⁴ Обычный градиентный спуск использует одну фиксированную скорость обучения для всех параметров. Это может быть проблемой, потому что слишком высокая скорость может привести к нестабильности, а чересчур низкая — к медленному обучению. Адаптивные методы динамически подстраивают скорость обучения, увеличивая эффективность этого процесса. Каждый из этих методов имеет свои особенности в том, как именно он изменяет эту скорость, что помогает в решении разных типов задач.

¹⁵ Представьте, что Вы учитесь кататься на велосипеде. Сначала Вы можете крутить педали быстро, чтобы научиться управлять. Но чем лучше Вы едете, тем медленнее нужно вращать педали, чтобы сохранить баланс и не упасть. *Adam* или *RMSprop* делают то же самое для нейросети: сначала предпринимают шаги для быстрого улучшения, а потом замедляют процесс, чтобы добиться точности.

¹⁶ А теперь слезьте с велосипеда, так как наступила зима. Вы катаетесь на санках по холмистому снегу. Без импульса (или инерции) нужно каждый раз, когда Вы меняете направление, начинать с нуля. Это очень замедляет процесс, так как приходится постоянно бороться с сопротивлением. Теперь представьте, что есть немного инерции: Вы катитесь по склону и получаете небольшое ускорение в направлении, которое уже выбрали. Это помогает двигаться быстрее, так как Вы не тратите силы на каждое изменение направления.

Поэтому утверждать, что обученная нейросеть содержит произведение, — это всё равно что сказать, будто в каждом букваре содержится «Война и мир» только потому, что эта книга состоит из тех же букв, что и букварь.

Это подтверждается результатами недавнего исследования (Cooper, Gokaslan, et al., 2025), в котором была протестирована способность нескольких генеративных моделей (включая *LLaMA*, *DeepSeek*, *Gemini* и др.) воспроизводить фрагменты различных произведений, как находящихся в общественном достоянии, так и охраняемых авторским правом на момент проведения эксперимента. В большинстве случаев книги, использованные в тестировании, демонстрировали минимальное присутствие в «памяти»¹⁷ моделей¹⁸.

1.3. Этапы обучения ИИ

Соответственно, процесс обучения ИИ нужно разделить на несколько этапов.

Первый этап — это сбор данных, которые содержат полезную информацию, необходимую для дальнейшего обучения модели.

¹⁷ Напомню, что под «запоминанием» тут опять же имеются ввиду высоковероятностные последовательности. «Паттерн [который модель усвоила из обучающих данных. — А.Н.] и есть запомненные обучающие данные» (Cooper & Grimmelmann, 2025, p. 42).

¹⁸ Наиболее часто воспроизводились типовые и тривиальные фрагменты, такие как уведомления об авторском праве, адреса издательств, оглавления и краткие биографии авторов. Подобные блоки текста широко распространены в открытых источниках, многократно дублируются в книгах и на веб-страницах, и с высокой вероятностью повторяются в составе обучающих датасетов. Поэтому если модель генерирует, например, уведомление об авторском праве, это вовсе не означает, что оно было заимствовано из конкретного произведения; скорее оно всё еще представляет собой обобщенный результат статистического усреднения большого числа сходных источников.

Реже, но всё же воспроизводились фрагменты из всемирно известных и часто цитируемых источников, таких как Библия, философские тексты Джона Стюарта Милля, «Божественная комедия» Данте Алигьери, а также официальные документы Правительства США.

Некоторое повышение вероятности генерации наблюдалось и в отношении особо популярных художественных произведений, таких как «Гарри Поттер» или «Властелин колец». Однако даже в этих случаях модели лишь эпизодически генерировали краткие, хотя и дословные, фрагменты, и то лишь при большом количестве попыток. В отдельных экспериментах такие фрагменты были единственными, которые удавалось воспроизвести с ненулевой вероятностью. Причем даже эти воспроизведения не возникали как результат одного запроса, а представляли собой накопленные статистически возможные совпадения, для получения которых требовались сотни или тысячи генераций. Для извлечения одного фрагмента длиной, например, в 50 токенов потребовались тысячи целенаправленных итераций запросов. Но даже в случаях, когда такое минимальное воспроизведение происходит, его степень может существенно различаться как между авторами, так и между произведениями одного и того же автора.

По моему мнению, наиболее вероятным объяснением этого феномена также является многократное дублирование таких текстов в Интернете — как в виде прямых цитат, так и в переработанном виде, — что статистически увеличивает вероятность генерации определенных фрагментов при правильной формулировке запроса и большом количестве попыток.

С одной стороны, это показывает, что воссоздание части содержания некоторых книг путем «взлома» самой модели технически возможно. Но это, по сути, тривиальное следствие самой природы генеративных моделей: инструмент, предназначенный для создания нового текста, неизбежно имеет потенциал для генерации элементов охраняемых форм.

Следовательно, сам факт генерации таких фрагментов не может служить достаточным доказательством нарушения авторских прав или устойчивого воспроизведения конкретной книги в модели. Более того, требуемое для этого количество запросов и невозможность последовательного воссоздания значительных фрагментов текста подрывают аргумент о том, что генерация может заменить чтение произведения как пользовательский опыт.

Второй этап — интеллектуальный анализ данных (*text and data mining*)¹⁹, процесс автоматизированного анализа больших объемов данных с целью извлечения полезной информации, такой как закономерности, тенденции, связи и корреляции. Может включать обработку как структурированных, так и неструктурированных данных (например, текстов, изображений, видеозаписей и других типов данных). В результате этого анализа создается база данных, которая содержит не сами произведения, а лишь извлеченную информацию — метаданные.

Третий этап — это обучение модели. На основе полученных метаданных создается система взаимосвязей, позволяющая модели делать предсказания и генерировать результаты.

Первый и второй этапы взаимосвязаны между собой и могут быть реализованы разными способами. Например, можно собрать «сырые данные», включая охраняемые произведения, на сервере, а затем провести их анализ. В этом случае, очевидно, на сервере будут создаваться копии этих данных, что может повлечь за собой нарушение права на воспроизведение объектов авторского права.

С другой стороны, если объект доступен для «технического» свободного посещения, т. е. он был доведен до всеобщего сведения, можно предоставить доступ к такому объекту программе для анализа. В этом случае программе не нужно создавать копию произведения в своей базе данных. Ей достаточно просто просканировать объект и сохранить только метаданные, которые затем могут быть использованы на третьем этапе в процессе обучения ИИ.

Следовательно, неверно утверждать, что ИИ «содержит произведения, на которых он был обучен». На самом деле ИИ хранит только информацию, полученную в результате анализа этих произведений (Cooper & Grimmelmann, 2024, p. 14–44).

Тем не менее тот факт, что модель ИИ не сохраняет сами произведения, на которых она обучалась, не означает, что она не может воспроизвести их при составлении ответа. Здесь снова уместно сравнение с человеком. Любое лицо, обладающее определенными навыками, может воспроизвести картину или текст, посмотрев на них и разобравшись в технике их выполнения. Однако интенция воспроизведения и само воспроизведение — это разные ситуации.

2. Формально-юридический анализ обучения ИИ с точки зрения гражданского права

2.1. Использование объекта авторских прав для обучения ИИ

Для определения возможного нарушения авторских прав при обучении ИИ необходимо понять, что охраняется российским законодательством в сфере авторских прав и какие действия могут считаться их нарушением. Так, согласно п. 1 ст. 1270 Гражданского кодекса РФ²⁰ (ГК РФ) автору произведения или иному правообладателю принадлежит исключительное право

¹⁹ «Текстовый анализ данных — это общее наименование для вычислительных процессов, направленных на структурирование неструктурированных электронных текстов и применение статистических методов для выявления новой информации и закономерностей в обработанных данных. Иными словами, текстовый анализ данных включает любые процессы, задействующие компьютеры, которые генерируют метаданные из информации, изначально не предназначенной для использования в качестве данных. Процесс текстового анализа данных может быть применен для получения статистики и фактов о произведениях, защищенных авторским правом. Однако он также может преобразовывать защищенные авторским правом тексты, звуки и изображения в абстракции, которые не являются теми же самыми произведениями и даже непохожи на оригинал. Вместо этого данные абстракции могут быть полезны для получения инсайтов и новых идей, касающихся оригинального выражения» (Sag, 2019, pp. 294–295).

²⁰ Собрание законодательства РФ, 05.12.1994, № 32, ст. 3301.

использовать произведение в любой форме и любым не противоречащим закону способом. Здесь можно выделить два критерия: что использоваться должно произведение и что произведение должно использоваться.

Общеизвестный постулат российского авторского права: «авторское право защищает форму произведения, но не его содержание»²¹. Данная позиция исходит из п. 5 ст. 1259 ГК РФ, не распространяющего авторско-правовую охрану на идеи, концепции, принципы, процессы, системы, способы, решения технических, организационных или иных задач и пр.²² К этой норме также примыкают и общие положения о том, что произведение подлежит авторско-правовой охране, если оно получено в результате творческого труда (ст. 1257 ГК РФ) и *выражено в объективной форме* (п. 3 ст. 1259 ГК РФ), при этом часть произведения будет охраняться только в том случае, если она может быть признана самостоятельным результатом творческого труда автора (п. 7 ст. 1259 ГК РФ).

Эти положения российского законодательства отражают результаты давней дискуссии о том, что и как охраняется авторским правом. Начиная с работ И. Фихте, произведение стало рассматриваться как абстрактный (идеальный) объект, не привязанный к какой-либо материальной форме (Peukert, 2021 p. 258). После создания оно существует вне времени и пространства, а каждая его манифестация в объективном мире (будь то телесная, цифровая или пространственная форма) всего лишь копия этого идеального объекта. В данной парадигме исключительное право на такое произведение — это прежде всего право контролировать его коммерческое использование (Dozortsev, 2003, p. 112). Однако, поскольку этот объект абстрактен и не является уникальным, как, например, движимые вещи, где вопросы незаконного заимствования одного объекта в другом не могут возникнуть, распространение абсолютных прав на произведение в целом — на каждую его часть — привело бы либо к полной невозможности охраны, либо к полной невозможности дальнейшего творчества. Как шутят филологи, каждое европейское произведение включает в себя отсылки или повторения сюжетов христианской Библии. Конечно, это шутка, но в каждой шутке есть доля утки (правды). Действительно, любое произведение не создается в полном вакууме, а является, как выразился Р. Барт, «текстом в тексте»²³, т. е. оно остается внутренне связанным с прошлыми произведениями.

Поэтому произошло разделение на охраняемую форму и неохраняемое содержание. Любое произведение, как правило, состоит из совокупности как охраняемых, так и неохраняемых элементов. Исходя из этого, использование произведения и, соответственно, нарушение прав на произведение могут иметь место только в тех случаях, когда задействованы именно

²¹ Надо отметить, что в литературе указанный подход критикуется в основном потому, что невозможно точно определить, где проходит линия демаркации между неохраняемым содержанием и охраняемой формой, см.: Рожкова, М. А. (2021). «Авторское право охраняет форму, патентное — содержание» — правильно ли это утверждение и при чем здесь «оригинальность произведения» и «плагиат»? *Закон.ру*. https://zakon.ru/blog/2021/2/10/avtorskoe_pravo_ohranyaet_formu_patentnoe_sode; см. также Kashanin (2010b). В этом смысле указанная доктрина должна выступать скорее грубым ориентиром, чем способом точного расчета. И мы в целом согласны с указанной критикой, однако в данном случае достаточно даже этого грубого критерия, чтобы увидеть, что извлечение и перемножение миллионов корреляций между словами из миллионов книг не могут считаться копированием творческих элементов.

²² И в целом то, что идеи не подлежат защите авторским правом, — это общая парадигма для всех стран, имеющих хоть какую-то доктрину авторского права, неважно, строится ли она на англо-американском понимании, делающем упор на экономическую часть отношений, или на континентальном подходе к авторскому праву с личными неимущественными правами (Ducato & Strowel, 2021, p. 24).

²³ См. подробнее: Barthes (1989, pp. 413–423).

охраняемые элементы. Неохраняемые элементы, в свою очередь, не являются частью произведения в юридическом смысле этого слова.

Как можно было заметить, слово «произведение» использовалось в двух разных значениях. Во-первых, как результат творческого труда в обыденном языке и, во-вторых, как объект гражданских прав. Причем два соответствующих понятия не совпадают по объему. Точно так же стоит рассматривать и слово «использование». Несмотря на то, что согласно ст. 1270 ГК РФ правообладатель имеет исключительное право контролировать использование произведения любым способом и для любой цели, как в доктрине, так и в судебной практике значение термина «использование» оказывается более узким, чем в повседневной речи²⁴. Так, в соответствии с п. 3 ст. 1270 ГК РФ «практическое применение положений, составляющих содержание произведения, в том числе положений, представляющих собой техническое, экономическое, организационное или иное решение, не является использованием произведения...». Эта норма подчеркивает, что простое использование идей или решений, заложенных в произведении, не является нарушением авторских прав, если эти идеи или решения не выражены в конкретной форме произведения.

Кроме того, существует различие между использованием и потреблением произведений (Pavlova, 2018, p. 238). Под потреблением понимается сам процесс восприятия произведения, например чтение книги, просмотр произведения искусства или аудиовизуального произведения. Этот процесс не воспринимается доктриной как использование, несмотря на отсутствие явного положения об этом в позитивном праве.

Анализируя пункт 2 ст. 1270 ГК РФ, в котором приводится открытый перечень способов использования произведений, можно свести все эти способы либо к фактическому созданию копии, либо к доведению копий до потенциальных потребителей. Все виды использования, указанные в статье, за исключением, возможно, сдачи в прокат и импорта²⁵, могут быть рассмотрены как различные формы воспроизведения произведения.

Это могут быть долговечные копии произведения или его частей (например, воспроизведение по смыслу подп. 1 и 2 п. 2 ст. 1270 ГК РФ, распространение, доведение до всеобщего сведения, такое как загрузка произведения в Интернет, переработка, практическая реализация архитектурного проекта и т. д.). Также это могут быть кратковременные копии, например публичный показ, сообщение в эфир, сообщение по кабелю, ретрансляция и другие формы временного использования.

²⁴ Этот подход можно найти и в судебной практике других стран с континентальной системой права. Например, в решении Федерального суда Германии (Bundesgerichtshof [BGH], Oct. 4, 1990, I ZR 139/89, 453) по делу *G. Radio-Werke GmbH, F. i. Bay., v. GEMA* суд постановил следующее: «Объектом охраны авторского права является нематериальное благо, которое по своему прямому назначению, как правило, служит прежде всего интеллектуальному или эстетическому наслаждению личности, которое по своей природе имеет место в сугубо частной сфере в случае многих интеллектуальных произведений, а если действие не предшествует или не обеспечивает это интеллектуальное удовлетворение, оно не имеет значения для авторского права» (Kollár, 2021, p. 3).

Во Франции точно так же статья L.122-3 Code de la propriété intellectuelle [C. Pro. Int.] [Кодекс интеллектуальной собственности] определяет право на воспроизведение как «материальную фиксацию произведения всеми средствами, которые позволяют довести его до сведения общественности косвенным образом». «Здесь мы видим, что французское определение права на воспроизведение предполагает доведение до сведения общественности, и формулировка статьи, по-видимому, отражает чувство наслаждения или принуждение другого человека к наслаждению как предварительное условие нарушения авторских прав» (Dusollier, 2018, p. 166).

²⁵ Но сдачу в прокат и импорт можно рассмотреть как частный случай распространения (как это было в Законе РФ «Об авторском праве и смежных правах» от 09.07.1993. № 5351-1. Российская газета, 03.08.1993), а само распространение как ввод экземпляров в оборот неразрывно связано с воспроизведением (Pavlova, 2018, p. 238–239).

Тем не менее в любом случае при использовании произведения появляется новая копия в материальном мире либо эта копия представляется на показ так, чтобы множество людей имело доступ ко всему произведению (т. е. могли потребить его, не создавая собственную копию, что, как мы отметили ранее, не является использованием в юридическом смысле).

Соответственно, применение произведения в этих двух случаях (создание копии или представление копии на показ) можно считать использованием в юридическом смысле, и нарушение авторских прав имеет место, если такое использование происходит без согласия правообладателя.

Это также косвенно подтверждается тем, что в ст. 1273–1280.1 ГК РФ устанавливаются исключения для создания копий или предоставления доступа к кратковременным копиям (например, публичное исполнение). Это логично, поскольку исключения для свободного использования произведений вводятся только в тех случаях, когда по общему правилу было бы нарушение авторских прав. Если же в юридическом смысле какое-то действие не подпадает под определение использования, то и ограничений на него не требуется²⁶.

Из этого можно сделать вывод, что нарушением имущественного права на произведение будет такое использование охраняемых элементов произведения, при котором создается новая копия этого элемента либо элемент становится доступным для публичного потребления, если это произошло без согласия автора.

Ввиду того что процесс обучения ИИ можно разделить на три этапа, нарушение прав правообладателей может произойти только в случае, если к произведению был незаконно получен доступ или оно было воспроизведено без согласия правообладателя. Метаданные, полученные на втором этапе, и сам процесс обучения ИИ на этих метаданных не могут нарушать исключительное право, так как метаданные и сам процесс обучения не являются частью юридического объекта, на который распространяется исключительное право. Таким образом, использование метаданных для обучения ИИ не нарушает авторские права, поскольку метаданные не являются охраняемым произведением, а процесс обучения — не использование произведения в юридическом смысле.

2.2. Использование базы данных как объекта смежных прав для обучения ИИ

Впрочем, ситуация меняется, если второй этап обучения нейросети осуществляется на базе данных, которая является объектом смежного права. В отличие от авторского права смежные права представляют собой объект охраны *sui generis* (самостоятельной, особенной охраны), поскольку они защищают не творческий труд и связь автора с произведением, а коммерческий интерес, который оформлен в виде исключительного права.

Пункт 1 ст. 1334 ГК РФ устанавливает особое содержание исключительного права на базу данных. Теперь это не использование любым способом и для любой цели, как в случае с авторским правом, а только существенное извлечение материалов из базы данных или полное копирование базы данных. При этом в п. 1 и 3 ст. 1335.1 ГК РФ специально оговорено, что извлечение незначительной части базы данных не считается нарушением исключительного права

²⁶ Это необязательно означает отсутствие такого ограничения в законе, поскольку в ГК РФ нередко встречаются так называемые декларативные нормы, которые не изменяют общего регулирования института, но подчеркивают возможность или невозможность определенных действий. Как мы увидим далее, в других юрисдикциях, когда речь идет об интеллектуальном анализе данных, в законодательстве часто устанавливаются ограничения авторских прав, касающиеся создания временных копий для целей анализа, но эти ограничения не затрагивают сам процесс анализа, так как его нельзя назвать использованием.

изготовителя базы данных, если такие действия не противоречат нормальному использованию базы данных и не ущемляют необоснованно законные интересы изготовителя базы данных.

Для того чтобы понять, является ли извлечение метаданных нарушением прав правообладателя базы данных, нужно ответить на следующие вопросы:

- являются ли метаданные, используемые для обучения ИИ, частью материалов базы данных;
- если да, то является ли извлечение этих метаданных существенным;
- если нет, то ущемляют ли такие действия необоснованно законные интересы правообладателя базы данных?

Само определение базы данных подразумевает, что это структурированная информация, организованная по определенному принципу. Для целей охраны базы данных эта информация может быть разнообразной, начиная от сметных цен на ресурсы и заканчивая охраняемыми объектами авторских прав, независимо от того, защищены ли они другими правовыми режимами (Kalyatin, 2023, pp. 20–21).

Метаданные также могут быть собраны и структурированы таким образом, чтобы они могли получить охрану в качестве базы данных. Однако для этого необходимо, чтобы база данных изначально создавалась с намерением структурировать информацию, содержащую метаданные, которые будут использованы для обучения ИИ. Это подтверждается пунктом 1 ст. 1334 ГК РФ, который распространяет исключительное право только на те базы данных, создание которых «требует существенных финансовых, материальных, организационных или иных затрат». Логично, что исключительное право на базу данных принадлежит ее создателю только в том случае, если формирование этой базы данных потребовало значительных затрат. Исключительное право не возникает на данные, которые потенциально могут быть включены в базу. Если база данных содержит информацию, которая была собрана не с целью обучения ИИ, а для других нужд, то для появления необходимых метаданных нужно провести ее анализ. Метаданные — это всегда результат анализа, осуществляемого внешним наблюдателем (в данном случае специальной программой), а не просто данные, хранящиеся в базе.

Из этого следует и ответ на вопрос, является ли интеллектуальный анализ данных существенным извлечением материалов из базы данных: нет, за исключением случаев, когда сама база состоит из таких метаданных. Ведь, как уже говорилось, правовой режим охраны базы данных защищает коммерческие усилия правообладателя, связанные с ее сбором. Если метаданные не собирались в базу изначально, это означает, что составитель базы данных не вкладывал достаточных коммерческих усилий, которые подлежат защите в рамках смежного права²⁷.

Частично это касается и ответа на вопрос, ущемляют ли такие действия законные интересы правообладателя базы данных. С одной стороны, ущемления не будет, если предположить, что владелец обычной базы данных не вкладывал коммерческих усилий в создание метаданных, необходимых для обучения ИИ. В таком случае действия, направленные на использование этих метаданных, не могут считаться нарушением интересов правообладателя.

С другой стороны, можно представить случаи, когда сам процесс анализа (шаг 2) может нарушить законный интерес правообладателя. Этот интерес заключается в обеспечении нормального использования базы данных. Если интеллектуальный анализ, необходимый для извлечения метаданных, создает чрезмерную нагрузку на базу и это приводит к значительным трудностям

²⁷ Хотя здесь логика должна быть такой же, что и с воспроизведением объекта авторских прав. Если для анализа требуется сначала извлечь материал, а только потом проанализировать, то де-юре это уже будет нарушением прав на базу данных.

в ее использовании (например, она начинает слишком долго загружаться, перезагружается или замедляется), то такой интеллектуальный анализ можно рассматривать как нарушение законных интересов правообладателя.

Резюмируем: для обучения модели ИИ нужны не сами материалы базы данных, а только абстрактные метаданные, полученные в процессе интеллектуального анализа. Извлечение таких метаданных не является извлечением существенной части базы данных, поскольку таковым считается извлечение материалов, составляющих саму базу данных. Следовательно, извлечение метаданных, как правило, не нарушает законных прав и интересов изготовителя базы данных, за исключением нескольких специфических случаев:

- если база данных была специально собрана для обучения ИИ и выдача лицензий на такое использование является основным имущественным интересом правообладателя;
- если анализ создает слишком большую нагрузку на базу данных, затрудняя её использование другими пользователями (например, замедляя ее работу);
- если доступ к базе данных был получен неправомерным путем, например с обходом технических средств защиты.

В остальных случаях такое использование базы данных не должно считаться нарушением прав.

3. Анализ мировых подходов к обучению ИИ

3.1. Европейский подход

В Европе была принята Директива об авторском праве на Едином цифровом рынке²⁸, которая определяет интеллектуальный анализ данных (*text and data mining*) как «любой автоматизированный аналитический метод, направленный на анализ текста и данных в цифровой форме с целью получения информации, включая закономерности, тенденции и корреляции, но не ограничиваясь ими» (п. 2 ст. 2). Это можно рассматривать как юридическое описание части процесса обучения искусственного интеллекта (Rosati, 2019, pp. 200–201). Как уже упоминалось, нейросеть при обучении не воспроизводит объекты анализа напрямую, а извлекает из них различные корреляции и паттерны, чтобы затем использовать эти метаданные для предсказания и генерации новых работ.

Пункт 1 ст. 4 Директивы об авторском праве в целом разрешает временное воспроизведение и извлечение произведений²⁹ «для целей интеллектуального анализа текста и данных» при условии, что

- совершается анализ произведения, доступ к которому был получен правомерно;
- создаваемые копии используются только в целях анализа;

²⁸ Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC. 2019 O.J. (L 130), 92–125. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019L0790> (далее — Директива об авторском праве).

²⁹ Тот факт, что исключение сформулировано в части «временного воспроизведения и извлечения», также подтверждает, что сам по себе анализ и полученные данные в случае, если они не воспроизводят охраняемую часть, не являются нарушением авторских прав. Однако есть аргументы, почему разумно предусмотреть такое же исключение для постоянных копий: как минимум для того, чтобы понимать, на чем именно обучался ИИ и как он пришел к тому или иному решению, а также в целом повысить прозрачность работы таких систем (Margoni & Kretschmer, 2022, pp. 693–694).

- правообладатель не предусмотрел явного и ясного отказа от такого интеллектуального анализа в машиночитаемом виде³⁰.

Таким образом, данное положение Директивы не ограничивает интеллектуальный анализ какой-либо конкретной целью³¹. Анализ может проводиться для любых целей, будь то коммерческие или некоммерческие.

³⁰ Согласно п. 18 преамбулы Директивы об авторском праве «целесообразно сохранять эти права только за счет использования машиночитаемых средств, включая метаданные и условия использования веб-сайта или сервиса». Однако следует помнить, что преамбула не создает юридически обязательных норм, но является одним из инструментов толкования текста обозначенной Директивы.

³¹ В литературе можно встретить иное мнение, что положения Директивы об авторском праве не охватывают интеллектуальный анализ данных во время обучения искусственного интеллекта, поскольку обучение генеративного ИИ выходит за рамки традиционного интеллектуального анализа данных: оно не просто извлекает информацию, но и воспроизводит, интернализирует и синтезирует выразительные элементы произведений: структуру, стиль, композицию. По мнению авторов, юридически и технически обучение ИИ является актом воспроизведения, а не аналитического использования и потому выходит за пределы целей исключений (Lucchi, 2025).

Их аргументация о неприменимости исключения интеллектуального анализа данных к обучению генеративного ИИ опирается на методологическую путаницу между двумя последовательными этапами одной цепочки работы с данными. Фаза интеллектуального анализа данных ограничивается автоматизированным извлечением статистических сигналов (повторяемых элементов, корреляций, эмбедингов) из законно доступных произведений; обучение модели начинается после этой аналитики и использует уже обезличенные числовые представления, а не сами тексты или изображения (хотя, конечно, в реальности эти процессы могут происходить почти одновременно). Подменяя разные этапы (процесс получения аналитических данных (TDM) и использование этих данных программой (генеративным ИИ)), критикуемые авторы необоснованно объявляют сам факт обучения воспроизведением (а значит, нарушением Директивы, ст. 2 Директивы *Infosoc* (Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society. 2001 O.J. (L 167) 44, 10. <https://eur-lex.europa.eu/eli/dir/2001/29/oj/eng>) и решения суда ЕС по делу *Pelham* (Case C-476/17, *Pelham GmbH v Hütter*, EU:C:2019:624, 29.07.2019), хотя в модели остается лишь результат линейных и нелинейных преобразований — матрицы весов, которые не позволяют реконструировать исходное произведение в читаемом виде. Формально «копий» произведений в смысле решения *Infopaq* (Case C-5/08, *Infopaq Int'l A/S v Danske Dagblades Forening*, EU:C:2009:465, E.C.R. I-6569, 19.07.2009) здесь не существует, поскольку в *Pelham* (на которое авторы также ссылаются) Суд ЕС указал, что воспроизведение имеет место лишь тогда, когда потенциальный потребитель может в обыденном языке узнать и распознать первоначальное произведение. Разумеется, увидев одни векторы, он этого сделать не сможет. Юридической защитой охватывается только фактическое воспроизведение конкретной формы выражения, а не сама по себе способность системы или человека — при удачном запросе — теоретически сгенерировать схожий фрагмент. Такая возможность сопоставима с гипотетическим демоном Лапласа в искусстве: всеведущее существо, зная точное положение каждой молекулы кисти и холста, могло бы восстановить любое живописное полотно прошлого, но правонарушением это станет лишь в тот момент, когда демон действительно нанесет копию на холст. Точно так же художник, разобравший чужие произведения на приемы и решения, отвечает только за конкретное воспроизведение, а не за наличие в памяти абстрактных паттернов, позволяющих при желании приблизиться к чьему-то произведению.

Наконец, указанный авторы фактически пытаются осуществить телеологическое толкование нормы Директивы. Если, согласно Директиве, процесс интеллектуального анализа данных является аналитическим, то и использование результатов должно быть исключительно аналитическим. Однако ни статья 3, ни в статье 4 Директивы не увязывают правомерность копирования для интеллектуального анализа данных с последующим использованием результатов такого анализа. Исключение для интеллектуального анализа данных — это, по сути, эксцепция для работы с большими данными, в рамках которой в момент анализа есть воспроизведение, но в результате самого анализа его уже нет. Иное прочтение сделало бы норму, позволяющую анализ в коммерческих целях, фактически «мертворожденной», ведь аргумент «Ваше использование не является аналитическим» оставляет возможность только для научного и образовательного использования.

Другое исключение содержится в ст. 3 Директивы, которая устанавливает возможность свободного интеллектуального анализа правомерно доступных произведений³² в научных и исследовательских целях, если он совершается организациями культурного наследия или исследовательскими организациями³³. В отличие от исключения ст. 4 правообладатели не могут ограничить возможность использования произведений в научных целях, что обусловлено конкретной целью этого исключения.

Соответственно, можно выделить следующие различия между ст. 3 и 4 Директивы об авторском праве:

- **Цель использования.** Статья 3 ограничена только научными и исследовательскими целями, в то время как статья 4 охватывает все цели, включая коммерческие.
- **Ограничения для правообладателей.** В ст. 3 правообладатели не могут ограничить использование произведений, тогда как в ст. 4 они могут установить ограничения с помощью машиночитаемых средств или условий использования.
- **Тип использования.** Статья 3 ориентирована на некоммерческую деятельность³⁴, а статья 4 допускает коммерческое использование.

Европейский законодатель ясно выразил намерение распространить ограничения, содержащиеся в Директиве, на создание баз данных для обучения искусственных нейронных сетей в Регламенте об искусственном интеллекте³⁵. Согласно подп. (с) п. 1 ст. 53 этого Регламента

³² Т. Маргони проводит обстоятельный анализ понятия «законный доступ» (*lawful access*) в контексте ст. 3 и 4 Директивы об авторском праве (Magroni, 2025). Автор подчеркивает, что, несмотря на центральную роль этого понятия в допустимости использования охраняемых произведений в рамках анализа данных, ни в Директиве, ни в корпусе авторского права ЕС оно не получает четкого нормативного определения. В правоприменительной и научной практике часто происходит смешение категорий «законный доступ» (*lawful access*) и «законный источник» (*lawful source*). Маргони предлагает четко развести их и интерпретировать требования законного доступа как характеристику поведения пользователя, получающего доступ к информации (например, исследователя), а не как характеристику правового статуса источника, к которому осуществляется доступ (например, сайта с нелегальным контентом). Такая интерпретация опирается на буквальное и телеологическое прочтение ст. 3 Директивы и преамбулы к ней: согласно п. 14 преамбулы доступ может быть признан законным, если осуществляется на основе политики открытого доступа, по подписке, иным правомерным способом либо в отношении контента, свободно доступного онлайн. Это позволяет утверждать, что законодатель намеренно вывел законный доступ за пределы связки с законным источником. Автор подчеркивает, что в противном случае реализация ст. 3 станет невозможной: исследовательские учреждения не способны технически и финансово проверять законность каждого отдельного произведения, тем более что после обучения модели на огромном массиве данных невозможно удалить отдельный след или пересобрать модель без огромных затрат (Cooper, et al, 2024). Такая буквальная и телеологическая трактовка позволяет сохранить эффективность ст. 3 Директивы, особенно с учетом приоритетов научной свободы, закрепленной в праве ЕС. Маргони утверждает, что, в то время как для статьи 4, ориентированной на коммерческое использование, уместна более строгая связка законности доступа и законности источника, для ст. 3 такая интерпретация недопустима, поскольку она разрушает баланс между публичными целями научного исследования и охраной авторского права.

³³ По поводу определения исследовательской организации см.: Quintais (2025).

³⁴ «Некоммерческая» не трактуется как «бесплатная». Если некоммерческая организация взимает плату за доступ к результатам, но при этом обеспечивает режим равного доступа для всех желающих, это не считается нарушением Директивы. См. судебное решение по спору *Robert Kneschke v. LAION e.V.* (LG Hamburg [Региональный суд Гамбурга], 27.09.2024, 310 O 227/23, § 1–56), анализ этого решения представлен ниже.

³⁵ Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No. 300/2008, (EU) No. 167/2013, (EU) No. 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (text with EEA relevance), O.J. L, 2024/1689, 12.7.2024, <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>

«поставщики моделей искусственного интеллекта общего пользования обязаны разработать стратегию, в частности, для выявления и соблюдения ограничений, предусмотренных пунктом 3 ст. 4 Директивы об авторском праве на Едином цифровом рынке». Было бы странным утверждать, что эти ограничения не применяются, при этом обязывая всех разработчиков следовать стратегии по их соблюдению.

Таким образом, в Директиве уже предусмотрен баланс интересов, выбранный Европейским Союзом. Извлечение метаданных с помощью технических средств разрешено, и их дальнейшее использование также допускается, за исключением случаев, когда правообладатель явно запретил такой анализ. Либо разрешение правообладателя не имеет значения, но в этом случае лицо, которое производит интеллектуальный анализ, должно делать это в некоммерческих научных целях.

3.1.1. Дело «Кнешке против LAION»

Данный подход уже был частично апробирован в немецкой судебной практике³⁶.

Немецкий фотограф Роберт Кнешке обнаружил, что несколько его фотографий были включены в набор данных, который использовала исследовательская организация LAION для обучения искусственного интеллекта. Р. Кнешке утверждал, что процесс создания набора данных включал временное копирование изображений для извлечения информации и что такие действия равносильны нарушению его авторских прав.

Позиция истца (Кнешке)

Создание временных копий его фотографий без согласия нарушает авторское право, так как эти действия подпадают под понятие «воспроизведение» в соответствии с Законом об авторском праве Германии³⁷ и не покрываются существующими ограничениями авторского права согласно § 44a³⁸, 44b³⁹, 60d *UrhG*⁴⁰.

Временное копирование в целях анализа данных не может считаться правомерным: оно, во-первых, не является достаточно кратковременным (§ 44a *UrhG*), а во-вторых, не соответствует определению интеллектуального анализа данных (§ 44b

³⁶ LG Hamburg [Региональный суд Гамбурга], 27.09.2024, 310 O 227/23.

³⁷ Urheberrechtsgesetz [UrhG] [Закон об авторском праве и смежных правах], 09.09.1965, Bundesgesetzblatt [BGBl] I, 1273 (далее — *UrhG*).

³⁸ § 44a *UrhG* «Временное копирование» разрешает временное копирование произведений, если оно:

- кратковременное или сопутствующее;
- является частью технического процесса (например, кэширование или буферизация);
- не имеет самостоятельного экономического значения.

³⁹ § 44b *UrhG* «Анализ текста и данных» позволяет интеллектуальный анализ произведений (*text and data mining*) для выявления закономерностей, трендов или корреляций. Исключение действует, только если:

- доступ к произведению был получен правомерно;
- результаты анализа не используются для копирования содержания;
- правообладатель не установил явного запрета на такие действия.

⁴⁰ § 60d *UrhG* разрешает использование произведений для целей научного исследования, включая интеллектуальный анализ данных. Исключение распространяется на:

- некоммерческие исследовательские организации;
- произведения, используемые исключительно для научных целей;
- деятельность, не нарушающую законный интерес правообладателей или нормальную коммерческую эксплуатацию произведений.

UrhG), поскольку предполагает использование самого содержания произведений, что прямо запрещено.

Деятельность ответчика не может рассматриваться как некоммерческая. Члены ассоциации, включая руководство, связаны с коммерческими структурами и работают в области технологий и искусственного интеллекта. Созданная база данных размещена в открытом доступе и используется коммерческими организациями. Более того, проект частично финансировался из коммерческих источников, включая выплату в размере 5000 долл. США за предоставление предварительной версии базы одному из членов ассоциации. Также отсутствуют признаки научной деятельности. База данных не содержит оригинального научного вклада и лишь агрегирует материалы из других источников. Ответчик не представил доказательств того, что база применялась при подготовке научных публикаций или других исследований с признанной научной ценностью. Истец утверждает, что работа по сбору и обработке данных имела исключительно технический, а не исследовательский, характер и не была направлена на создание нового знания, как научная деятельность.

Использование его фотографий для обучения нейросети противоречит действующим исключениям, в том числе по причине наличия на сайте явного запрета на подобное использование. Кроме того, массовое применение этих изображений в целях обучения ИИ подрывает нормальную коммерческую эксплуатацию произведений.

Позиция ответчика (LAION)

Использование изображений для анализа не затрагивало содержания произведений и было ограничено обработкой метаданных. Это соответствовало технической цели — сопоставлению визуальной информации с текстовыми описаниями — без вмешательства в охраняемую форму выражения.

Отдельно следует указать на отсутствие машиночитаемого запрета. Условия использования, на которые ссылается истец, не были сформулированы явно и не соответствовали требованиям § 44b(3) *UrhG*. Они были изложены на естественном языке, что делает их доступными только для восприятия человеком, но не машинной системой. Следовательно, отсутствует юридически значимый запрет в форме, поддающейся автоматическому распознаванию.

Деятельность ответчика можно обоснованно квалифицировать как научную в смысле § 60d *UrhG*. Создание базы данных было направлено на развитие технологий искусственного интеллекта и поддержку исследовательских инициатив. Сам факт, что база распространяется свободно и безвозмездно, свидетельствует о некоммерческом характере проекта. Более того, ее открытое предоставление способствует возможности других исследователей разрабатывать, тестировать и совершенствовать ИИ-модели, используя общедоступный научный ресурс.

База данных не только служит техническим инструментом, но и сама по себе представляет основу для научных исследований. Процесс ее создания и последующего использования соответствует общепринятым критериям научной деятельности. В 2022 г. была опубликована рецензируемая статья, описывающая методологию формирования базы, и она получила широкое признание в академическом сообществе: количество ее цитирований превысило 1400 (Schuhmann et al., 2022). Разработанные на базе этого ресурса инструменты применяются для анализа механизмов обучения ИИ и углубленного понимания его моделей.

Наконец, нельзя согласиться с утверждением о коммерческом характере деятельности ответчика. Связи отдельных участников проекта с коммерческими структурами не означают институциональную коммерческую заинтересованность: их работа в ассоциации носит добровольный,

неприбыльный характер. Финансовая поддержка от одной из компаний (в размере, например, 5000 долл. США) не предполагает управленческого контроля или влияния на принятие решений.

Публичный характер распространения базы данных также не свидетельствует о коммерческом подходе. Использование базы третьими лицами, включая коммерческие организации, происходит вне контроля и участия ответчика, без предварительных договоренностей и без получения прибыли. Ассоциация не занимается продажей базы и не использует ее в целях создания собственных коммерческих продуктов.

Решение суда

Суд первой инстанции Гамбурга постановил, что действия *LAION* не нарушают авторских прав истца, так как они полностью подпадают под исключение, предусмотренное § 60d *UrhG* на следующих основаниях.

- **Цель научного исследования** (§ 60d(1) *UrhG*). Суд установил, что создание базы данных и анализ данных проводились для целей научного исследования. Эти действия способствовали развитию технологий искусственного интеллекта и созданию открытого ресурса для исследователей. Суд отклонил довод о том, что сам сбор данных и извлечение метаданных нельзя признать научной деятельностью⁴¹.
- **Некоммерческий характер деятельности** (§ 60d(2) *UrhG*). Суд принял во внимание, что ответчик является некоммерческой ассоциацией, которая предоставляет данные бесплатно. Доказательств коммерческой направленности деятельности ответчика истцом представлено не было. Использование базы данных коммерческими структурами не определяет деятельность ответчика как коммерческую, так как данные предоставлялись на равных условиях всем пользователям⁴².
- **Отсутствие контролирующего влияния коммерческих организаций** (§ 60d(2)(3) *UrhG*). Суд отклонил доводы истца о том, что коммерческие компании имели определяющее влияние на деятельность ответчика. Наличие связей между отдельными членами ассоциации и коммерческими компаниями не свидетельствует о зависимости ассоциации или предоставлении преимуществ этим компаниям.

⁴¹ «Термин „научное исследование“ включает методическое и систематическое стремление к получению новых знаний. Поэтому его не следует трактовать слишком узко, ограничивая только этапами, непосредственно связанными с приобретением знаний. Скорее достаточно, чтобы рассматриваемый этап работы был направлен на последующее получение знаний, как это, например, происходит при сборе данных, которые необходимо сначала обработать, чтобы затем сделать эмпирические выводы. В частности, концепция научного исследования не требует обязательного достижения успеха. Таким образом, вопреки мнению истца, создание набора данных, аналогичного спорному, который может служить основой для обучения систем искусственного интеллекта, безусловно, может быть квалифицировано как научное исследование. Хотя само по себе создание набора данных еще не связано с непосредственным приобретением знаний, оно является фундаментальным этапом работы, направленным на дальнейшее использование такого набора с целью получения знаний. В настоящем деле достаточно того, что набор данных был бесспорно опубликован бесплатно и, таким образом, предоставлен исследователям, включая тех, кто работает в области искусственных нейронных сетей» (LG Hamburg [Региональный суд Гамбурга], 27.09.2024, 310 O 227/23).

⁴² «Тот факт, что оспариваемый набор данных также может использоваться коммерческими компаниями для обучения или развития их систем искусственного интеллекта, не влияет на квалификацию деятельности ответчика как научной. Более того, тот факт, что отдельные члены ассоциации, помимо работы в ее составе, также осуществляют оплачиваемую деятельность в коммерческих структурах, не является достаточным основанием для того, чтобы приписать деятельность этих компаний ассоциации» (LG Hamburg [Региональный суд Гамбурга], 27.09.2024, 310 O 227/23).

Фактически дело было разрешено без рассмотрения по существу других аргументов истца о незаконности интеллектуального анализа данных, т. е. применения § 44a *UrhG*. Однако в своем *obiter dictum* суд имплицитно обсудил возможность использования ст. 4 Директивы об авторском праве к данной ситуации.

Суд рассмотрел, как Германия имплементировала ст. 4 Директивы в национальное законодательство, включая § 44b *UrhG*. В решении подчеркивалось, что:

- § 44b *UrhG* основывается на ст. 4 Директивы, обеспечивая исключения для интеллектуального анализа данных, если произведения были законно доступны и отсутствуют ограничения со стороны правообладателя;
- национальные правила соответствуют духу статьи 4 Директивы, направленной на развитие технологий искусственного интеллекта и поддержку научных исследований.

Суд отметил, что ограничения, введенные правообладателем, должны соответствовать требованиям ст. 4 Директивы в части четкости и недвусмысленности, причем сделаны они должны быть на машиночитаемом языке. Однако опубликованные на сайте фотобанка ограничения соответствовали этим требованиям, даже несмотря на то, что они были изложены на естественном, а не машиночитаемом, языке⁴³.

Кроме того, суд подчеркнул необходимость разделения разных этапов процесса обучения ИИ. В настоящем деле суд указал следующие этапы:

- 1) создание набора данных (единственного предмета спора здесь), который также может быть использован для обучения ИИ;
- 2) последующее обучение искусственной нейронной сети в том числе на этом наборе данных;
- 3) последующее использование обученного ИИ с целью создания нового графического контента.

Каждый этап использования ИИ должен быть подвергнут различной правовой оценке. Вполне можно сказать, что каждый этап будет иметь свои критерии того, что законно и что нет.

Следует также подчеркнуть, что в германском законе анализ текстов и данных в научных целях (§ 60d *UrhG*) и интеллектуальный анализ данных (§ 44b *UrhG*) — это, вслед за Директивой, разные способы взаимодействия с произведениями. Так же как и Директива об авторском праве, § 60d *UrhG* не дает возможности правообладателям отказаться от анализа, но ограничивает цели анализа только научными, а § 44b устанавливает возможность создания временных копий для интеллектуального анализа, но при условии, что нет запрета со стороны правообладателя.

Кроме того, § 44b не говорит о том, что законным использованием является именно сам процесс интеллектуального анализа. Норма указывает именно на разрешение воспроизведения законно доступных произведений для целей такого анализа. Поэтому вполне можно сказать, что и германское право не рассматривает интеллектуальный анализ данных как процесс в качестве использования произведения. Использование появляется только в том случае, если необходимо создание копии произведения для такого анализа.

⁴³ «Суд склонен считать “машиночитаемым” также и чисто “естественный” язык... Однако вопрос, может ли и при каких конкретных условиях оговорка, заявленная на “естественном языке”, также считаться “машиночитаемой”, всегда должен решаться в зависимости от технических достижений, существовавших на момент использования произведения... Однако “современные технологии” также однозначно включают в себя приложения ИИ, которые способны понимать текст, написанный на естественном языке... В этом отношении всё говорит о том, что законодатель в Регламенте ИИ имел в виду именно такие приложения ИИ, когда говорил о “современных технологиях”» (LG Hamburg [Региональный суд Гамбурга], 27.09.2024, 310 O 227/23).

Очевидно, что если Директива об авторском праве или национальное законодательство не устанавливает специальных ограничений на сам интеллектуальный анализ, то эти ограничения не могут волшебным образом появиться на этапе использования полученных метаданных. Для этого сам интеллектуальный анализ должен быть ограничен определенными целями или способами на законодательном уровне.

3.1.2. Европейский подход — идеальный баланс интересов?

Может показаться, что европейский подход пытается учесть одновременно интересы как разработчиков, так и правообладателей. Действительно, с одной стороны, у разработчиков остается достаточно большое количество материалов, с другой — авторы и правообладатели, которые не хотят, чтобы на их материалах обучались (читай: хотят, чтобы им платили за то, что их материалы подвергнутся интеллектуальному анализу), могут зарезервировать свои права. Но это только на первый взгляд.

Самая большая проблема — это транзакционные издержки⁴⁴ (Senftleben, 2023). Отсутствие возможности обучения без переговоров с авторами означает следующее.

Во-первых, разработчики ИИ должны будут вступать в переговоры с каждым автором по отдельности, чтобы получить его данные для обучения.

Во-вторых, каждый автор постарается получить довольно высокую цену, желательно в виде отчислений, процента от дохода.

В-третьих, с учетом того, что нереально определить, насколько та или иная работа повлияла на конкретную генерацию (и повлияла ли вообще), запрашиваемые цены практически никогда не будут в действительности коррелировать с вкладом конкретного произведения в работу ИИ.

В-четвертых, это приведет к тому, что разработчики либо будут отказываться от использования произведений под запретом, либо пойдут по пути наименьшего сопротивления и будут заключать соглашения только с крупными игроками.

Это, в свою очередь, ставит в невыгодное положение владельцев нишевого и небольшого контента и фактически приносит прибыль только крупным игрокам (Senftleben et al., 2025, p. 15). Легко представить себе сценарий, в котором крупные издательства будут принуждать авторов включать в произведения запрет на обучение, но при этом не будут делиться с ними прибылью, которые они получили от лицензирования произведений.

А с учетом того, что это, в общем, не правовое решение, а политическое, т. е. апеллирует к балансу интересов, как правильно указал М. Сенфтлебен, «механизм резервирования прав п. 3 ст. 4 Директивы не запускает автоматически новый поток доходов для авторов и правообладателей. Это очень простой механизм: заявляя об отказе, владелец авторских прав нейтрализует разрешение на интеллектуальный анализ данных абзаца 1 ст. 4 и восстанавливает исключительное право запрещать использование человеческих произведений для обучения ИИ. С точки зрения двух центральных общественных целей (лучший ИИ и справедливое вознаграждение) это проигрышный сценарий: никаких учебных ресурсов для тренировки ИИ (владелец авторских прав сказал нет) и никакого

⁴⁴ Транзакционные издержки — это юридико-экономическая категория, обозначающая составные затраты, косвенно связанные с заключением, исполнением и обеспечением сделок (например, поиск контрагента, сбор информации, переговоры, мониторинг, разрешение споров). В экономике они особенно важны для обоснования распределения прав на имущество и объяснения, почему важны институциональные механизмы. См. подробнее: Coase (1960), Rindfleisch (2020).

дополнительного дохода для авторов и правообладателей (статья 4 не предусматривает выплату вознаграждения)»⁴⁵. В этом смысле даже если общество решает, что право на информацию и прогресс должны быть в какой-то степени ограничены для поддержки авторов (по крайней мере в той части, в которой они сами отказываются от участия их работ в обучении), то для таких «исключений из исключений» просто возможность отказа максимально неэффективна, особенно с учетом того, что отсутствует консенсус по ключевым аспектам отказа: кто и как должен выражать отказ, какие технологии признаны допустимыми, где и в какой момент отказ действует (Mezei, 2025).

В таких случаях система государственного/принудительного лицензирования (режим «можно, но платно»), которая принята в некоторых странах, например, для создания каверов⁴⁶, выглядит более эффективной, и к ее закреплению призывают многие авторы, исследующие эти ограничения (Geiger & Iaia, 2024; Riccio, 2024; Senftleben, 2023; Senftleben, 2025)⁴⁷. Однако с учетом чисто правовой стороны вопроса, а также того факта, что на самом деле от лицензирования опять же будут выигрывать скорее крупные правообладатели, чем собственно авторы, гораздо убедительнее те исследователи, которые говорят о необходимости введения «широкого, но при этом четко определенного» теста, при котором можно было бы определять, что то или иное использование является справедливым и добросовестным (Tuagi, 2024, p. 569)⁴⁸, или указывают на неохраноспособность того, что извлекается из произведений при интеллектуальном анализе (Dermawan, 2024; Sag et al., 2024).

Права авторов в данном случае, с учетом публичной природы их интереса, были бы гораздо эффективнее реализованы через публичные механизмы, такие как налогообложение или целевые сборы⁴⁹, взимаемые с разработчиков ИИ. Эти средства могли бы направляться специально на поддержку творчества и свободы самовыражения. Причем взимание сборов должно быть

⁴⁵ Senftleben, M. (2025, April 22). The TDM opt-out in the EU — Five problems, one solution. *Kluwer Copyright Blog*. <https://copyrightblog.kluweriplaw.com/2025/04/22/the-tdm-opt-out-in-the-eu-five-problems-one-solution/>

⁴⁶ Например, в США существует так называемая *Compulsory Mechanical License*, позволяющая использовать опубликованное музыкальное произведение без согласия правообладателя, при соблюдении некоторых условий и выплате фиксированного вознаграждения специализированной платформе или напрямую издателю (ст. 115 Закона об авторском праве 1976 г. (The Copyright Act of 1976, 17 U.S.C. §§ 101–810)). Аналогичный подход существует в Германии, где можно получить согласие на кавер не у автора, а у Общества по коллективному управлению авторскими правами (*Gesellschaft für musikalische Aufführungs- und mechanische Vervielfältigungsrechte*).

⁴⁷ Но существует и альтернативная позиция, что рыночное лицензирование всё равно предпочтительно, см.: Stratton (2025), Padilla & Prasad (2025).

⁴⁸ В данном случае такое использование рассматривается с точки зрения концепции прав пользователя. Как только пользователи определяются как законные, их пользование произведением определяется не как исключение из авторского права, а как реализация собственных правовых возможностей в хофельдовском смысле (Tuagi, 2024, p. 569).

⁴⁹ Отличие от принудительного лицензирования здесь в том, что при системе лицензирования мы признаем, что корреляции и метаданные подпадают под контроль правообладателя, что на самом деле открывает другую бездну ограничений прав на образование, информацию и исследования. Кроме того, лицензирование предполагает, что собранные средства справедливо распределяются между правообладателями в зависимости от того, насколько их конкретный вклад полезен для цели лицензирования. Здесь же вклад конкретной работы ничтожен. В этом смысле невозможно реально пропорциональное вознаграждение с точки зрения подхода частного права. Однако публичный элемент, т. е. целевые сборы, как раз может разрешить эти проблемы, так как, раз мы будем не распределять сборы между авторами, а сразу направлять на стимулирование творчества в целом, то и устанавливать размер сборов можно просто как часть получаемой прибыли (например, при условии уменьшения общего налога на прибыль в этой части), не задумываясь о конкретном вкладе.

ограничено использованием произведений, чьи авторы выразили отказ от использования их работ для обучения ИИ. Тем самым можно создать ситуацию, при которой отказ будет реально выгоден только тем авторам, которые действительно хотят достигнуть полезного социального эффекта, а не пытаются в очередной раз распространить авторское право на неохраняемые элементы.

Впрочем, даже при этом кажется, что сам по себе полезный эффект от развития ИИ и открытия возможностей для творчества большому количеству людей, у которых есть творческая искра, но отсутствуют необходимые навыки, гораздо больше, чем возможные потери доходов тех авторов, чьи произведения имеют настолько низкую оригинальность, что могут быть реально заменены ИИ, который в принципе не умеет создавать что-то творческое, а только будет делать самое банальное, если его не направит реальный творец.

3.2. Подход США

На июль 2025 г. счет 2:1 в пользу разработчиков ИИ. Несмотря на то что еще множество важных и больших дел⁵⁰ все еще находятся на рассмотрении американских судов, уже по трем из них первая инстанция вынесла свои решения. Два из них касаются генеративных ИИ общего назначения (*Bartz v. Anthropic PBC*⁵¹ и *Kadrey v. Meta Platforms, Inc.*⁵²), и одно — генеративного ИИ специального назначения, для работы с юридическими документами и подбором судебной практики (*Thomson Reuters Enterprise Centre GmbH et al. v. ROSS Intelligence Inc.*⁵³). Однако надо учитывать, что это лишь первая инстанция и сторонам предстоит еще долгий путь в апелляции. В связи с этим гораздо более продуктивным будет обсудить американский подход к авторскому праву и интеллектуальному анализу данных в целом, чем концентрироваться только на указанных делах.

В связи с этим в данной части статьи мы сосредоточимся на следующих вопросах:

- 1) конституционный анализ обучения ИИ с точки зрения оснований авторского права;
- 2) можно ли считать использованием включение в базу данных для обучения объектов авторского права в американском праве;
- 3) если признавать это использованием, то можно ли считать такое использование справедливым (*fair use*)?

3.2.1. Конституционные основы авторского права в США

Для того чтобы разобраться, как американское право смотрит на интеллектуальный анализ данных и обучение нейросетей, для начала нужно понять, что вообще фундаментально представляет собой американский подход к интеллектуальной собственности. В отличие от континентальных стран, в том числе России, где авторское право представляет собой одно из основных конституционных прав гражданина в силу того, что оно мыслится тем, что Кант называл «естественными правами», т. е. правами, которые неотъемлемо принадлежат любому лицу, исходя из его автономии воли и формального равенства, данная идея была рассмотрена и отвергнута отцами-основателями (Stallman, 2024, pp. 77–86). Вместо этого американский

⁵⁰ См. подробнее: *Master List of lawsuits v. AI, ChatGPT, OpenAI, Microsoft, Meta, Midjourney & other AI cos.* <https://chatgptiseatingtheworld.com/2024/08/27/master-list-of-lawsuits-v-ai-chatgpt-openai-microsoft-meta-midjourney-other-ai-cos/>

⁵¹ *Bartz v. Anthropic PBC*, No. 3:24-cv-05417 (N.D. Cal. June 23, 2025).

⁵² *Kadrey v. Meta Platforms, Inc.*, No. 3:23 cv 03417 (N.D. Cal. June 25, 2025).

⁵³ *Thomson Reuters Enterprise Centre GmbH et al. v. ROSS Intelligence Inc.*, No. 1:20-cv-00613-SB (D. Del. Feb. 11, 2025).

фундамент авторского права пронизан одновременно утилитарными⁵⁴ и локковскими идеями⁵⁵, в соответствии с которыми, с одной стороны, представляется справедливым, что каждый должен получать достойное вознаграждение за свой интеллектуальный труд, а с другой — форма защиты этого вознаграждения должна способствовать прогрессу наук и искусств. В связи с этим положение об интеллектуальной собственности в американской Конституции не носит характера неотъемлемого права, а сформулировано через полномочие Конгресса «содействовать прогрессу науки и полезных искусств, закрепляя на ограниченный срок за авторами и изобретателями исключительные права на их соответствующие произведения и открытия»⁵⁶. Казалось бы, данная разница должна существенно влиять на правовую форму достижения этой утилитарной конституционной цели, однако, к сожалению, она почти не получила должного конституционного осмысления⁵⁷. Вместо этого авторское право начало рассматриваться в качестве простого имущественного права (Carrier, 2004).

Несмотря на отсутствие полной последовательности в американском законодательстве по данному вопросу, можно говорить об имплицитном утилитарном подходе, оказывающем влияние на судебную практику: американские суды, как правило, проявляют крайнюю осторожность в вынесении решений, которые могли бы препятствовать развитию новых технологий или сдерживать научно-технический прогресс. Например, в чрезвычайно важном для развития доктрины добросовестного использования деле *Google LLC v. Oracle America, Inc.* (2021)⁵⁸ решался вопрос о возможности использования API-интерфейсов⁵⁹ во время разработки для достижения совместимости разных программ, написанных на Java. Судья Томас Брейер прямо указал на то, что если решение было бы принято в пользу компании Oracle, это «рисковало бы нанести вред общественности»: «[если бы] Oracle одна держала бы ключи, результат вполне мог бы оказаться прибыльным для Oracle (или других фирм, имеющих авторские права на программные интерфейсы) ... [но] блокировка помешала бы, а не способствовала, основным творческим целям авторского права»⁶⁰. Судья Брейер определил, что использование Google API соответствовало всем четырем факторам добросовестного использования и что Google использовала

⁵⁴ «Авторские права и патенты призваны мотивировать творческую деятельность авторов и изобретателей путем особого вознаграждения и предоставления общественности доступа к продуктам их гениальности после истечения ограниченного периода исключительного контроля» (*Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 US 417, 429 (1984); «Непосредственным результатом нашего закона об авторском праве является обеспечение справедливого вознаграждения за творческий труд „автора“. Но конечной целью выступает стимулирование художественно-го творчества посредством этого для всеобщего блага» (*Twentieth Century Music Corp. v. Aiken*, 422 US 151, 156 (1975).

⁵⁵ Mossoff, A. (2021, March 8). The constitutional protection of intellectual property (Legal Memorandum No. 282). *The Heritage Foundation*. <https://www.heritage.org/sites/default/files/2021-03/LM282.pdf> См. также: Solberg, Th. (Comp.). (1906). *Copyright enactments of the United States, 1783-1906*. U.S. Government Printing Office, 18-19. В деле *Ruckelshaus v. Monsanto*, 467 US 986 (1984) Верховный суд прямо цитировал Джона Локка.

⁵⁶ U.S. Const., art. I, § 8, cl. 8.

⁵⁷ Как пишет A. Mossoff, практически все авторитетные сборники по конституционному праву не обсуждают положение об авторском праве и патентах как в целом, так и в части конституционного измерения „прогресса“» (Mossoff, 2007, p. 955; см. также о проблематике: Mossoff, A. (2021, March 8). The constitutional protection of intellectual property (Legal Memorandum No. 282). *The Heritage Foundation*. <https://www.heritage.org/sites/default/files/2021-03/LM282.pdf>).

⁵⁸ *Google LLC v. Oracle America, Inc.*, 593 U.S. 1 (2021).

⁵⁹ *Application Programming Interface (API)* — это набор правил и протоколов, которые позволяют различным программам взаимодействовать друг с другом. Это своего рода «контракт», который описывает, как программы могут обмениваться данными и использовать функции друг друга.

⁶⁰ *Google LLC v. Oracle America, Inc.* (2021).

«только то, что было необходимо, чтобы позволить пользователям применить накопленные таланты для работы в новой и преобразованной программе».

3.2.2. Можно ли считать обучение использованием с точки зрения американского права?

Федеральное законодательство США не имеет такого широкого определения исключительного права, как континентальные страны. Согласно ст. 106 Закона об авторском праве США владелец авторских прав имеет исключительные права на выполнение и разрешение конкретных действий, установленных в Законе⁶¹.

Иными словами, говорить об использовании по смыслу Закона об авторском праве США можно в том случае, если использование подпадает под один из вышеперечисленных способов. Это позволяет сделать вывод о том, что потребление не будет использованием в том числе и в США⁶². Однако американское понятие воспроизведения включает не столько копирование защищаемых элементов произведения (хотя это тоже учитывается в анализе), сколько конкурентное сходство двух объектов и потенциальную возможность замещения одного объекта другим, если такое сходство было обнаружено. В этом проявляется различие между концепциями охраняемой формы и неохраняемого содержания, принятыми в континентальном праве, и неохраняемой идеи и охраняемого выражения в США. Мы пытаемся выделить конкретные элементы охраняемой формы творчества; для США важно, чтобы существенное сходство было обозначено в зависимости от ответов «обычного разумного человека»⁶³. Как указывал в одном

⁶¹ Статья 106 американского Закона об авторском праве устанавливает следующие способы использования:

- 1) воспроизведение защищенного авторским правом произведения в виде копий или фонограмм;
- 2) создание производных произведений на основе произведений, защищенных авторским правом;
- 3) распространение копий или фонограммы защищенного авторским правом произведения среди общественности путем продажи или иной передачи права собственности, а также путем сдачи в аренду, лизинга или предоставления займы;
- 4) в случае литературных, музыкальных, драматических и хореографических произведений, пантомим, кинофильмов и других аудиовизуальных произведений — публичное исполнение защищенного авторским правом произведения;
- 5) в случае литературных, музыкальных, драматических и хореографических произведений, пантомим, а также живописных, графических или скульптурных произведений, включая отдельные кадры кинофильма или другого аудиовизуального произведения, — публичная демонстрация произведения, защищенного авторским правом, и
- 6) в случае звукозаписей — публичное исполнение защищенного авторским правом произведения посредством цифровой аудиопередачи.

⁶² Однако есть и альтернативное мнение. Р. Матюлёните предлагает переосмыслить исключительное право на воспроизведение в контексте генеративного ИИ, утверждая, что процесс «попадания» произведений в параметры модели представляет собой функциональный эквивалент воспроизведения, пусть и не тождествен ему в техническом смысле (Matulionyte, 2025). Однако, как видно, это все-таки переосмысление подхода к авторскому праву, тем более с применением доктрины функциональной эквивалентности из других отраслей права. Это возможно в американском праве, но губительно для континентального подхода, так как в этом случае функциональную эквивалентность можно будет находить у произведений с принципиально разной формой, а значит, констатировать нарушение.

⁶³ См.: *Sid & Marty Krofft Television v. McDonald's Corp.*, 562 F. 2d 1157; *International Luggage Registry v. Avery Products Corp.*, 541 F.2d 831; *Harold Lloyd Corp. v. Witwer*, 65 F.2d 1, 18–19 (9 Cir. 1933); *Twentieth Century-Fox Film Corp. v. Stonesifer*, 140 F.2d 579, 582 (9 Cir. 1944). Однако в Россию тоже проник этот же взгляд на нарушение исключительных прав в вопросе нарушений прав на персонажа. Пункт 82 постановления Пленума Верховного Суда РФ «О применении части четвертой Гражданского кодекса Российской Федерации» от 23 апреля 2019 г. № 10, согласно которому нарушение можно установить исходя из того, «сохранил ли персонаж свою узнаваемость» или нет, конечно, отход от континентальной концепции. С другой стороны, в случае персонажей это более оправдано. См. более подробный анализ: Nikiforov (2020).

из прецедентов судья Хэнд, авторское право «не может быть ограничено буквально текстом, в противном случае плагиатор избежит наказания за счет несущественных изменений»⁶⁴. Это расширительное, «экономическое» толкование нарушения уравнивается более абстрактными принципами справедливого использования, а также другими ограничивающими доктринами вроде *de minimis*⁶⁵.

В судах практически не ставятся вопросы о том, что представляет собой юридическое использование, вместо этого суды концентрируются на проблеме отнесения того или иного фактического использования к понятию справедливого использования. Даже такой очевидный пример, как использование неохраняемых абстрактных идей, рассматривается в качестве случая справедливого использования⁶⁶.

Поэтому с точки зрения американского авторского права любое взаимодействие с объектом — неважно, какое — будет называться использованием. В связи с этим становится более важным различие между справедливым (законным) и несправедливым (незаконным) использованием.

3.2.3. Справедливое использование

В отличие от континентального подхода, где ограничения авторских прав должны быть четко сформулированными и конкретными согласно трехступенчатому тесту Бернской конвенции по охране литературных и художественных произведений 1886 г. (далее — Бернская конвенция)⁶⁷, американское справедливое использование представляет собой набор принципов и тестов, согласно которым определяется та или иная возможность. При определении добросовестного использования проверяются четыре фактора:

- 1) цель использования, в том числе коммерческий или некоммерческий характер;
- 2) характер произведения;
- 3) объем и существенность заимствования;
- 4) воздействие, которое такое использование оказывает на потенциальный рынок или ценность охраняемого произведения.

Кроме этого, можно выделить еще две дополнительные доктрины справедливого использования, которые частично вытекают из этих критериев, а частично расширяют их понимание: это использование, которое обосновывается другим неотъемлемым правом, в первую очередь

⁶⁴ *Nicholas v. Universal Pictures Corp.*, 45 F.2d 119, 121 (2 Cir. 1930).

⁶⁵ См. подробнее: Astrachan (2008).

⁶⁶ В том же деле *Google LLC v. Oracle America, Inc.*, несмотря на то что аргументация *Google* строилась в первую очередь на неохраняемой *API* в целом, Верховный суд уклонился от ответа на этот вопрос, вместо этого сосредоточившись на анализе справедливого использования. «В Соединенных Штатах необходимость сохранения различия между идеями и выражением играет важную роль в применении доктрины добросовестного использования» (*Sag et al.*, 2024, p. 10). В деле *Sony Computer Entertainment, Inc. v. Connectix Corp.*, 203 F.3d 596 (9th Cir. 2000), суд прямо признал, что «доктрина добросовестного использования сохраняет общественный доступ к идеям и функциональным элементам, встроенным в защищенные авторским правом программные продукты».

⁶⁷ Трехступенчатый тест Бернской конвенции — это международно-правовой критерий, ограничивающий возможность государств вводить исключения и ограничения из авторского права. Он был впервые закреплён в ст. 9(2) Бернской конвенции (Парижский акт 1971 г.), согласно которой национальные законодательства могут разрешать воспроизведение произведений лишь в особых случаях, если такое воспроизведение не наносит ущерба нормальной эксплуатации произведения и не ущемляет необоснованно законные интересы автора. Эти три критерия образуют так называемый трехступенчатый тест (*three-step test*), который должен соблюдаться кумулятивно, т. е. все три элемента должны быть выполнены одновременно.

правом на свободу слова (Первая поправка к Конституции США), и преобразующее использование. Последнее касается прежде всего производных произведений, и является их характеристикой, при которой такое произведение выходит за смысловые рамки изначального произведения или позволяет представить его в совершенно новом ключе, недоступном при потреблении изначального произведения (Carroll, 2019, pp. 916–922)⁶⁸. Соответственно, включение объектов авторских прав в процесс обучения ИИ будет рассматриваться с точки зрения этих критериев.

Цель использования, в том числе коммерческий или некоммерческий характер

В ст. 107 Закона об авторском праве США уже приводятся некоторые случаи, когда цель использования считается справедливой: это критика, комментарии, репортажи, обучение (включая множественные копии для использования в классе), научные исследования. При этом коммерческий или некоммерческий характер использования, хотя и учитывается при анализе, не является решающим фактором, т. е. само по себе коммерческое использование производного объекта не может считаться нарушением априори. В деле *Campbell v. Acuff-Rose Music, Inc.* было установлено, что коммерческое использование становится тем менее значимым, чем более преобразованным является новое произведение⁶⁹.

В этом смысле выделение метаданных для обучения, даже если и признавать его использованием, будет использованием в высшей степени преобразующим⁷⁰. То же самое происходит

⁶⁸ См. также: *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569 (1994); *Kelly v. Arriba Soft Corp.*, 336 F.3d 811 (9th Cir. 2003); *Perfect 10, Inc. v. Google, Inc.*, 487 F.3d 701 (9th Cir. 2007).

⁶⁹ Но см. также: *Andy Warhol Found. for the Visual Arts v. Goldsmith*, 143 S. Ct. 1258, 1276 (2023), где указывалось, что преобразование должно быть достаточным, а цель нового произведения должна сильно отличаться от оригинального.

⁷⁰ Действительно, сложно представить большее преобразование, чем превращение художественных текстов в одну единую базу данных, состоящую из математических матриц и цифр, выражающих корреляции между блоками или пикселями, при взгляде на которые даже нельзя сказать, какие произведения использовались для получения этих корреляций. См. другие примеры цифрового преобразования: «Суд считает, что цель и характер использования произведений истцов компанией *iParadigms* являются высоко преобразовательными. Истцы изначально создали и произвели свои работы с целью образования и творческого выражения. *iParadigms* через *Turnitin* использует эти работы для совершенно другой цели, а именно для предотвращения плагиата и защиты письменных работ студентов от плагиата. *iParadigms* достигает этого, архивируя работы студентов в виде цифрового кода и не использует никакого конкретного выразительного или творческого содержания работы, кроме ограниченного использования для сравнения с другими работами» (*A.V. v. iParadigms, Ltd. Liab. Co.*, 544 F. Supp. 2d 473, 482 (E.D. Va. 2008)); «Суд первой инстанции, на наш взгляд, правильно определил, что архивирование работ истцов было преобразовательным и способствовало выводу о «добросовестном использовании». Использование этих работ компанией *iParadigms* было совсем не связанным с выразительным содержанием и было направлено на выявление и предотвращение плагиата» (*A.V. ex rel. Vanderhye v. iParadigms, LLC*, 562 F. 3d 630, 640 (4th Cir, 2009)); «...мы приходим к выводу, что создание полнотекстовой базы данных с возможностью поиска является квинтэссенцией преобразующего использования» (*Authors Guild, Inc. v. HathiTrust*, 755 F. 3d 87, 97 (2nd Cir. 2014)); «У нас нет никаких сомнений в том, что создание *Google* цифровых копий книг истцов с целью упрощения поиска для идентификации книг, содержащих интересующий термин, подразумевает высоко преобразовательную цель в смысле, подразумеваемом Кэмпбеллом; «...с помощью инструмента *Ngrams Google* позволяет читателям узнать частоту использования выбранных слов в совокупном корпусе опубликованных книг за разные исторические периоды. Мы не сомневаемся, что цель этого копирования является тем самым преобразовательным назначением, описанным Кэмпбеллом, которое сильно способствует удовлетворению первого фактора» (*Authors Guild, Inc. v. Google, Inc.*, 804 F.3d 202, 216–217 (2d Cir. 2015)) Указанные выдержки цит. по: Sag et al. (2024, pp. 14–15). Впрочем, есть также позиция, что использование авторско-правовых материалов для обучения не является преобразующим, но при этом всё равно выдерживает данный критерий, так как цели ИИ и обычного потребителя кардинально разные (Lemley & Casey, 2021).

с целью. В случае «невпечатляющего использования» разработчик ИИ преследует иную цель, не схожую с целями правообладателя объекта авторских прав. Для первого важно именно извлечение неохраняемых метаданных, для второго — получение вознаграждения за передачу копии работы для восприятия ее потенциальным потребителем.

Суд в деле *Thomson Reuters Enter. Ctr. GmbH et al. v. ROSS Intelligence Inc.*⁷¹ пришел к выводу, что использование истцом аннотированного контента (*headnotes*) не является преобразующим в смысле первого критерия. Основным аргументом послужило то, что цели использования ROSS и Thomson Reuters совпадают: обе стороны разрабатывают конкурирующие юридические поисковые системы. Росс применял аннотации не для создания новой выразительной формы, а для извлечения статистических данных, используемых при обучении негенеративного ИИ. Несмотря на то что копирование происходило на промежуточной стадии и данные не попадали в финальный продукт, суд отметил, что прецеденты, допускавшие промежуточное копирование в сфере программного обеспечения⁷², неприменимы, поскольку они касались функционального кода и для них был необходим доступ к нефункциональным элементам системы. В отличие от этих случаев копирование текстовых аннотаций ROSS не было функционально необходимым и не способствовало достижению новой цели пользователя, как того требует стандарт, установленный в деле *Andy Warhol Foundation v. Goldsmith*⁷³. Таким образом, первый этап анализа был решен не в пользу ответчика. Однако суд специально указал, что сделанные им выводы не применяются к генеративному ИИ.

В деле *Bartz v. Anthropic PBC*⁷⁴ окружной суд признал, что использование авторских произведений для обучения больших языковых моделей является высокопреобразующим в рамках первого критерия. Суд фактически использовал ту же логику, которая была описана выше. Раз модели обрабатывали тексты не для воспроизведения или замещения оригиналов, а для построения статистических связей с целью генерации новых текстов, это аналогично процессу обучения человека письму через чтение литературы. Обучение на объектах авторского права не может контролировать правообладатель: единственное, что он может требовать, — это плата за доступ к самому тексту⁷⁵. Суд отклонил аргумент, что компьютерам нельзя разрешать то же, что разрешено людям, и подчеркнул, что раз в результатах генерации не было копирования охраняемого содержания, то фактически речь идет об использовании стилистических приемов, методов и принципов, традиционно не охраняемых авторским правом⁷⁶. Это отличает указанное дело от *Thomson Reuters v. ROSS Intelligence Inc.*, где модель была натренирована не на генерацию новых текстов, а на поиск судебных решений, т. е. ИИ создавался с тем же функционалом, что и у оригинального продукта.

Кроме того, суд рассмотрел преобразования законно приобретенных печатных книг в цифровую форму для создания из них базы данных для обучения ИИ. С учетом того что бумажные копии были уничтожены, цифровые копии не распространялись⁷⁷ и не создавались избыточные экземпляры, такое изменение формата не затрагивает исключительные права

⁷¹ *Thomson Reuters Enter. Ctr. GmbH et al. v. ROSS Intelligence Inc.*, No. 1:20-cv-00613-SB (2025).

⁷² См.: *Google LLC v. Oracle America, Inc.*, 2021; *Sony Computer Entertainment, Inc. v. Connectix Corp.*, 2000; *Sega Enterprises Ltd. v. Accolade, Inc.*, 1992.

⁷³ *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508 (2023).

⁷⁴ *Bartz v. Anthropic PBC*, 3:24-cv-05417, (N.D. Cal.) (2025).

⁷⁵ См.: *Bartz v. Anthropic PBC* (2025) at 12.

⁷⁶ *Ibid* at 12–13.

⁷⁷ *Ibid* at 17.

автора, так как смена формата не связана с творческой модификацией содержания⁷⁸. Несмотря на коммерческий характер деятельности *Anthropic*, такая конвертация не носит самостоятельного экономического характера и сравнима с существующими прецедентами (и даже превосходит их по уровню преобразований с учетом обстоятельств)⁷⁹. Тем не менее скачивание книг с пиратских (теневых) библиотек справедливым использованием суд не признал, поскольку нет аргументов, что такое скачивание было разумно необходимо в противовес легальному приобретению копий⁸⁰.

В деле *Kadrey et al. v. Meta Platforms, Inc.*, суд также дал развернутую оценку первому критерию. Основной вывод такой же: обучение ИИ — это высокопреобразующее использование⁸¹. Однако в отличие от дела *Anthropic* суд занял позицию, что само по себе скачивание материалов с теневых библиотек не может признаваться нарушением только в силу того, что копия была пиратской⁸², и чтобы ответить на этот вопрос, нужно проанализировать добросовестность такого использования⁸³ или вред рынку.

Характер произведения

Данный критерий концентрируется на творческой составляющей объекта авторского права. Как правильно отметила А. Ворожевич, «границы исключительного права расширяются или сужаются в зависимости от ценности объекта авторского права, творческого права автора»⁸⁴.

Второй фактор требует признания того, что одни произведения находятся ближе к «ядру» охраны, предусмотренной авторским правом, чем другие. Соответственно, доказать справедливое использование труднее, когда объектом копирования становятся именно такие произведения или их существенные части. Так, документальные произведения подлежат меньшей степени охраны по сравнению с художественными или фантастическими⁸⁵, однако меньшая защита не означает ее полного отсутствия: даже оригинальная подборка и компоновка фактов, самих по себе не охраняемых, могут соответствовать низкому порогу оригинальности, необходимому для признания произведения охраноспособным⁸⁶. При этом второй фактор, как правило, «играет незначительную роль в анализе справедливого использования»⁸⁷ и имеет меньшее значе-

⁷⁸ Ibid.

⁷⁹ Ibid at 15.

⁸⁰ Дополнительным аргументом для суда здесь стало то, что после обучения *Anthropic* не стала удалять копии пиратских библиотек, а продолжила их хранить. Данный факт суд признал самостоятельным нарушением, не связанным с собственно процессом обучения. См.: *Bartz v. Anthropic PBC*, 18–24.

⁸¹ См.: *Kadrey et al. v. Meta Platforms, Inc.* (2025), at 15–16.

⁸² Ibid at 19.

⁸³ Вопрос, насколько должна учитываться добросовестность при анализе справедливого использования, дискуссионный: в американском праве отсутствует доктрина «плодов отравленного дерева» (Sag et al., 2024, p. 25), и существуют аргументы в пользу отказа от проверки добросовестности (Frankel & Kellogg, 2012; Leval, 2015; Carroll, 2019). Но см. альтернативную позицию: (Mukherjee, 2025). См. также о проблематике применения этой доктрины к интеллектуальным правам: (Lemley, 2017). Тем не менее можно представить и другие доводы, которые сделают использование по данному пункту несправедливым, вроде обхода средств защиты, игнорирования исключений, *robots.txt* и подобных механизмов (Sag et al., 2024, p. 7).

⁸⁴ Ворожевич, А. (2022). *Fair use и доктрина «расширяющего полезность использования»*. Гардиум. <https://legal-support.ru/information/blog/ip-daidzhest/fair-use-i-doktrina-rasshiryaushhego-poleznost-ispolzovaniya/>

⁸⁵ *Harper & Row, Publishers, Inc. v. Nation Enterprises*, 471 U.S. 539, 539 (1985).

⁸⁶ *Authors Guild v. Google*, 804 F.3d 202 (2nd Cir. 2015)

⁸⁷ Ibid.

ние в случаях, когда произведение уже опубликовано и, следовательно, исключительное право автора на первое обнародование не нарушается⁸⁸.

Основное назначение второго фактора заключается в том, чтобы обеспечить контекст для оценки прочих критериев добросовестного использования. Он позволяет выявить различия между характером оригинального произведения и вторичного, а также установить взаимосвязь между объемом заимствования и его значимостью в контексте создаваемого зависимого материала⁸⁹.

В данном случае выводы судов были противоположны позициям в первом критерии. В деле *ROSS* суд признал: материалы ответчика, хоть и содержат творческий элемент, это элемент достаточно низкого уровня и не идет в сравнение с художественными произведениями, что ограничивает степень его защиты. В силу этого он признал второй критерий благоприятствующим разработчикам.

В двух других делах суды, наоборот, признали, что данный критерий благоприятствует истцу по той причине, что все использованные произведения, несомненно, являются высокотворческими и оригинальными⁹⁰. Однако в деле *Meta* суд пошел дальше и предложил дополнительную аргументацию, на которой сто́ит остановиться поподробнее. Суд указал, что использование *Meta* книг истцов все-таки зависит от их творческого выражения и не сводится только к статистическим неохранным данным. Логика суда была следующей: «раз генеративный ИИ обучается на сборе статистических данных о порядке слов, частотности [какие слова и как часто используются. — А.Н.], грамматике и синтаксисе», а при этом «выбор слов, грамматика и синтаксис — это и есть формы выражения мысли»⁹¹, то статистика об этих взаимосвязях тоже есть продукт творческого выражения. Но, как кажется, здесь происходит путаница между двумя разными уровнями: с одной стороны, порядком слов, грамматикой, синтаксисом как формальными средствами выражения авторской мысли; с другой — статистическими данными, извлеченными из анализа этих средств в миллионах текстов. Да, порядок слов и остальное — это охраняемая форма литературного произведения. Однако это не значит, что агрегированные статистические зависимости, выявленные на их основе, тоже входят в состав этой формы. Статистика — это результат повторения. Она возникает там, где множество текстов демонстрируют общие черты. Но можем ли мы утверждать, что такая статистика является продуктом творческого выражения, если она становится видимой только в результате обработки тысяч объектов? Сам по себе этот факт говорит о том, что мы имеем дело не с индивидуальным творчеством, а с его усредненным следом. Если утверждать, что статистические взаимосвязи содержатся в каждом произведении и потому охраняются наряду с формой, возникает правовая ловушка. Тогда любое произведение, в котором воспроизводятся те же статистические зависимости, потенциально нарушает права авторов более ранних работ. А поскольку статистика по определению выявляет то, что повторяется, получается, что тысячи произведений нарушают права авторов тысяч других просто потому, что все они существуют в одном и том же культурном и языковом пространстве. Именно здесь стирается граница между неохранными идеями и охраняемой формой. Если статистические зависимости — это продукт творческого выражения, то почему бы не признать таковыми и идеи, и темы, и нарративные структуры, и типовые интонации? Ведь они тоже в каком-то смысле

⁸⁸ *VHT v. Zillow*, 918 F.3d 723, 744 (9th Cir. 2019), 336 F.3d at 820.

⁸⁹ *Campbell v. Acuff-Rose Music, Inc.* (1994); *Kelly v. Arriba Soft Corp.* (9th Cir. 2003); *Authors Guild v. Google* (2015); *Authors Guild v. HathiTrust* (2014); *Bill Graham Archives v. Dorling Kindersley Ltd.*, 448 F.3d 605, 612–613 (2d Cir. 2006).

⁹⁰ См.: *Bartz v. Anthropic PBC* (2025) at 24; *Kadrey et al. v. Meta Platforms, Inc.* (2025) at 23–24.

⁹¹ *Kadrey et al. v. Meta Platforms, Inc.* (2025) at 22–23.

выражаются творчески. Этот аргумент был бы убедительнее, если бы можно было доказать прямую зависимость влияния конкретного произведения на результаты обучения модели, но тогда можно было бы просто взять «самый великий» роман и обучить модель только на нем, однако такое обучение не дает таких же результатов, как обучение на большом количестве текста. Конкретное произведение не оказывает существенного влияния на дальнейшую работу ИИ, так как для его обучения требуются именно большие данные.

Объем и существенность использования

Третий фактор заключается в том, был ли объем заимствования разумным по отношению к той цели, которая была определена в первом критерии в качестве добросовестного использования. Даже небольшая часть может оказаться чрезмерной, если она представляет собой «сердце» произведения. Когда продуктивное использование возможно только за счет заимствования из конкретного произведения, доктрина справедливого использования достигает апогея. Когда использование возможно и без такого заимствования, она, наоборот, сводится к минимуму и заимствование требует особенно убедительного обоснования⁹².

В большинстве случаев признается, что заимствование целого произведения сильно снижает вероятность признания использования добросовестным⁹³. Однако технические копирования, которые не доносят оригинальное выражение автора до новой аудитории⁹⁴, суды США признают подчас справедливым (Sag et al., 2024). Определяющее значение здесь имеет не столько объем использованного материала, сколько объем охраняемого авторским правом материала, представленного публике⁹⁵, который может служить конкурентной заменой оригинального произведения⁹⁶.

Примерами такого «невпечатляющего использования»⁹⁷ в практике судов США будут:

- обратное проектирование (*reverse engineering*) программного обеспечения⁹⁸, в котором суды признали справедливым использованием копирование для дальнейшего изучения и извлечения неохранных элементов программ;
- копирование студенческих работ для проверки на плагиат⁹⁹;
- копирование библиотечных книг для их доступности в поиске, а также индексации¹⁰⁰;
- создание цифровых версий печатных библиотечных книг в исследовательских целях, включая анализ текстовых данных и машинное обучение¹⁰¹.

⁹² *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 547 (2023).

⁹³ *Worldwide Church of God v. Philadelphia Church of God, Inc.*, 227 F.3d 1110, 1118 (9th Cir. 2000), цитируя *Hustler Mag. Inc. v. Moral Majority Inc.*, 796 F.2d 1148, 1155 (9th Cir. 1986)); см. также: *Campbell v. Acuff-Rose Music, Inc.*, 587 (1994).

⁹⁴ Такое использование было метко названо *Matthew Sag* «невпечатляющим» (*non-expressive use*) (Sag, 2009, p. 3).

⁹⁵ *Fox News Network, LLC v. TVEyes*, 883 F.3d 169, 179 (2d Cir. 2018).

⁹⁶ *Google*, 804 F.3d 222.

⁹⁷ Следует отметить, что слово «впечатляющий» в данном контексте использовано для перевода английского *expressive* не в значении «хороший» или «качественный», а как «вызывающий впечатление», независимо от эмоциональной окраски или отношения воспринимающего к объекту. Соответственно, «невпечатляющее» использование подразумевает любое «техническое» использование, которое не позволяет воспринимающему ощутить произведение и получить какие-либо впечатления, обычно возникающие при его непосредственном восприятии органами чувств.

⁹⁸ *Sega Enters. v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992); *Sony Computer Ent. v. Connectix Corp.*, 203 F.3d 596 (9th Cir. 2000). Суды охарактеризовали копирование для извлечения неохранных элементов как «законную, по сути неэксплуатирующую цель» (Sag et al., 2024, p. 11).

⁹⁹ См.: *A.V. ex rel. Vanderhye v. iParadigms, LLC*, 562 F.3d 630, 644–645 (4th Cir. 2009).

¹⁰⁰ См.: *Authors Guild, Inc. v. Google, Inc* (2015).

¹⁰¹ См.: *Authors Guild, Inc. v HathiTrust* (2014).

Особенно в этом смысле важен вопрос интеллектуального анализа данных в силу того, что как уже указывалось, интеллектуальный анализ данных — это процесс, неразрывно связанный с обучением ИИ. Однако, как указывают некоторые американские ученые, обучение ИИ — это даже не отдельное технологическое явление, а скорее «приложение» к интеллектуальному анализу данных (Sag et al., 2024, p. 7). Более того, отнесение интеллектуального анализа данных к возможному справедливому использованию в том числе было подтверждено Бюро авторского права США¹⁰². Такие действия признавались справедливым использованием даже в случае, если при «невпечатляющем использовании» создавалось большое количество «технических» копий¹⁰³.

В этом смысле третий пункт не просто благоприятствует возможным ответчикам по делам, связанным с обучением ИИ, он также устанавливает возможность создания дополнительных копий в целях дальнейшего интеллектуального анализа, без получения согласия на воспроизведение у правообладателя, конечно, при условии соблюдения остальных условий.

Здесь суды были единодушны: во всех трех делах было признано, что третий критерий благоприятствует разработчикам ИИ. Главным аргументом здесь было то, что при дальнейшем использовании ИИ не генерируются произведения, применяемые при обучении, но при этом пользователи получают мощную выгоду от обучения на хороших данных. Кроме того, для хорошего обучения требуются миллиарды входных данных, в силу чего аргументы истца о том, что разработчики могли бы выбрать другие произведения или обучать только на произведениях, находящихся в общественном достоянии были отклонены. Как указали суды, использование каждого отдельного произведения является столь же разумным, как и любое другое¹⁰⁴.

Воздействие использования на потенциальный рынок или ценность охраняемого произведения

Четвертый критерий справедливого использования единогласно признается судами самым важным¹⁰⁵. Как указал суд в деле *Meta*: «Если первый фактор однозначно благоприятен

¹⁰² См.: The Register of Copyrights of the U.S.A. (2021). Section 1201 Rulemaking: Eighth Triennial Proceeding, Recommendation of the Register of Copyrights (pp. 121–124). https://cdn.loc.gov/copyright/1201/2021/2021_Section_1201_Registers_Recommendation.pdf «В процессе оценки предлагаемого исключения по § 1201 Закона об авторском праве в цифровую эпоху (*Digital Millennium Copyright Act*) для обхода технологических мер защиты на DVD и электронных книгах с целью проведения анализа текстовых данных Бюро авторского права заявило: „Учитывая четыре фактора добросовестного использования и обсуждаемые ограничения, регистратор приходит к выводу, что предлагаемое использование, вероятно, является добросовестным использованием“» (Sag et al., 2024, p. 17).

¹⁰³ «Полное неизменное копирование неоднократно считалось обоснованным как добросовестное использование, когда оно было разумно уместным для достижения преобразующей цели копирующего и было выполнено таким образом, что не предлагало конкурентной замены оригиналу» (*Authors Guild, Inc. v. Google, Inc.* (2015)); «Как и в случае с *HathiTrust*, копирование целого оригинала не только разумно уместно для преобразующей цели *Google*, но и фактически необходимо для ее достижения. <...> Хотя *Google* создает несанкционированную цифровую копию всей книги, она не раскрывает эту цифровую копию публике. Копия создается для того, чтобы функции поиска могли раскрыть ограниченную, но важную информацию о книгах» (*Authors Guild, Inc. v. HathiTrust* (2014)).

¹⁰⁴ *Bartz v. Anthropic PBC* (2025) at 25–27; *Kadrey et al. v. Meta Platforms, Inc.* (2025) at 24–25. Следует отметить, что при анализе третьего критерия суд в деле *Anthropic* еще раз подчеркнул, что в отношении пиратских библиотечных копий *Anthropic* не имела правовых оснований на их хранение: несмотря на заявленную цель обучения больших языковых моделей, компания стремилась собрать «все книги мира» и сохраняла их даже после отказа от использования. Такое масштабное несанкционированное копирование, включающее полные тексты миллионов книг, включая произведения истцов, выходит за пределы допустимого.

¹⁰⁵ *Harper & Row v. Nation Enterprises* (1985) at 566.

ответчику, у истца остается шанс победить только при безусловном выигрыше по четвертому фактору»¹⁰⁶.

Данный критерий в первую очередь проверяет наличие двух возможных негативных эффектов от использования произведения:

- 1) происходит ли прямое замещение одного произведения другим¹⁰⁷;
- 2) снижает ли новая работа потенциальную возможность лицензирования первоначального произведения заинтересованным лицом (косвенное замещение или *market dilution*)¹⁰⁸.

При этом, если вторичное произведение уменьшает спрос на оригинал из-за критики или пародии, такой ущерб не учитывается¹⁰⁹. Также принимаются во внимание «общественные блага, которые может принести копирование»¹¹⁰.

Фактически это самый неюридический критерий из всех, так как он отсылает к эмпирическому, в первую очередь экономическим, данным, которые показывают влияние фактического использования на рынок¹¹¹.

И в случае «непечатляющего использования» не происходит ни первый, ни второй вариант, ведь рынок и стоимость, о которых идет речь в четвертом факторе добросовестного использования, «не являются просто любыми выгодами, которые правообладатель может выбрать для ссылки» (Sag et al., 2024, p. 23). Чуть подробнее мы рассмотрим этот момент дальше в части про политико-правовые аргументы, а пока будет достаточно того, что потери правообладателя здесь носят такой же косвенный характер, как возможные потери правообладателя от появления на рынке произведения — новой оригинальной формы выражения той же идеи. Американское авторское право защищает оригинальное выражение, но не правообладателя от угрозы конкуренции с произведениями, которые не являются существенно сходными в глазах потребителей¹¹².

¹⁰⁶ *Kadrey et al. v. Meta Platforms, Inc.* (2025) at 26.

¹⁰⁷ *American Geophysical Union v. Texaco Inc.*, 60 F.3d 913 (2d Cir. 1994) — в отношении воспроизведенных копий; *Dr. Seuss Enters., L.P. v. ComicMix LLC*, 983 F.3d 443, 461 (9th Cir. 2020) — в отношении производных копий.

¹⁰⁸ *Princeton University Press v. Michigan Document Services*, 99 F.3d 1381 (6th Cir. 1996).

¹⁰⁹ *Campbell v. Acuff-Rose Music, Inc.* (1994) at 590.

¹¹⁰ *Google LLC v. Oracle America, Inc.* (2021) at 35.

¹¹¹ Надо уточнить, что в некоторых случаях предоставление эмпирических данных необязательно, если речь идет о не-преобразующем использовании — простом дублировании в коммерческих целях. В деле *Hachette Book Group, Inc. v. Internet Archive*, No. 20-cv-4160 (2023), например, было достаточно одного вывода: читатель может предпочесть скачать книгу бесплатно, а не купить ее. Это короткая и очевидная логическая цепочка. Но в случае с ИИ такую простую логическую цепочку сделать нельзя. Здесь решение зависит от ряда гипотетических условий: что именно ИИ способен генерировать замещающие тексты, что он будет использоваться с этой целью, что потребители предпочтут такие тексты работам живых авторов, что это предпочтение затронет книги истцов и что качество генераций обусловлено обучением на охраняемых книгах (см.: *Kadrey et al. v. Meta Platforms, Inc.* (2025) at 35–36).

¹¹² Впрочем, суд в деле *Meta* акцентировал внимание на том, что даже косвенный рыночный ущерб (например, в форме размывания рынка произведений массового характера) подлежит оценке в рамках четвертого критерия добросовестного использования — с точки зрения того, приводит ли поведение ответчика к снижению мотивации к творчеству, которую система авторского права призвана стимулировать и охранять. При этом суд особо подчеркнул, что релевантной для анализа является не гипотетическая ситуация полного отсутствия ИИ, а сценарий, при котором те же языковые модели были бы обучены исключительно на произведениях, находящихся в общественном достоянии (см.: *Kadrey et al. v. Meta Platforms, Inc.* (2025) at 30).

Однако, как указал суд, истцы в своем иске не заявили аргумент о рыночном размывании как самостоятельную теорию ущерба. Они сослались лишь на два типа вреда: (1) возможность воспроизведения фрагментов их книг

Надо отметить, что в деле *Thomson Reuters Enter. Ctr. GMBH v. Ross Intel, Inc.* суд первой инстанции пришел к выводу, что обучение специализированного ИИ для решения юридических проблем и вопросов на авторско-правовых материалах — заметках правовой базы *Westlaw* и ее системе заголовков, аннотаций, ранжирования и поиска — будет нарушением, в первую очередь исходя из четвертого критерия¹¹³. Следует признать, что данное дело имеет некоторую специфику, которая, по нашему мнению, не дает возможности признать его прецедентом в сфере обучения ИИ¹¹⁴.

Во-первых, в пользовательском соглашении *Westlaw* был прямо описан запрет на использование для создания конкурирующих платформ (а не в целом для любого обучения).

Во-вторых, ответчик нанял третью сторону, чтобы она скопировала аннотации и систему ранжирования, т. е. речь идет в том числе о незаконном копировании, а не просто об обучении.

В-третьих, самым важным фактором, на мой взгляд, стало то, что инструмент ответчика фактически являлся прямым конкурентом и рыночным заменителем для базы истца. В силу этого здесь роль сыграла в первую очередь недобросовестность ведения конкуренции в целом, а также существенное копирование «сердца» базы данных — их системы ранжирования, которое, в свою очередь, сняло заявление о «справедливости».

Данный вывод подтверждается при анализе двух других дел: *Meta*¹¹⁵ и *Antropic*¹¹⁶, в которых суды также пришли к выводу, что четвертый критерий благоприятствует признанию использо-

в ходе генераций и (2) нарушение развивающегося рынка лицензирования произведений для целей обучения ИИ. В результате основное внимание суд уделил именно этим двум теориям.

При этом *Meta* представила доказательства того, что выпуск модели *Llama 3* не оказал значимого негативного влияния на продажи книг истцов, равно как и на аналогичные книги из обучающего корпуса. Истцы же не предоставили никаких эмпирических данных, опровергающих эти выводы, ни относительно уже наступившего ущерба, ни относительно вероятности его возникновения в будущем. Как отметил суд, в подобных ситуациях, когда ответчик демонстрирует отсутствие рыночного вреда, а истец не представляет доказательств обратного, четвертый критерий подлежит оценке в пользу ответчика, что в данном случае и повлияло на исход дела. По поводу оценки эмпирических доказательств в таких ситуациях см.: *Seltzer v. Green Day, Inc.*, 725 F.3d 1170, 1179 (9th Cir. 2013); *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1168; *Anheuser-Busch, Inc. v. Natural Beverage Distributors*, 69 F.3d 337, 345 (9th Cir. 1995).

Таким образом, анализ четвертого критерия добросовестного использования в части косвенного размывания рынка в рамках американского права не может считаться завершенным или исчерпанным. Остается открытым вопрос, достаточна ли общественная и технологическая польза, получаемая за счет обучения моделей на охраняемых авторским правом произведениях, чтобы компенсировать потенциальный ущерб, наносимый тем категориям авторов, чьи произведения характеризуются низким уровнем творческой оригинальности, например создателям стоковых фотографий, авторам массовой жанровой прозы (включая определенные сегменты *LitRPG* или романтической литературы), которые и до появления ИИ зачастую воспринимались как шаблонные, а также начинающим писателям, чье творчество еще не обрело читательской аудитории, но может быть вытеснено генеративными аналогами.

¹¹³ Суд признал, что первый и четвертый критерий добросовестного использования (цель и характер использования и влияние на потенциальный рынок) говорят в пользу правообладателя, а второй и третий (характер использования, объем и значимость) — в пользу ответчика. При этом он еще раз подчеркнул, что главный критерий — четвертый, про влияние на потенциальный рынок.

¹¹⁴ Также вызывает сомнение отнесение заголовков к объектам, которые могут быть защищены авторским правом. См.: Frye, B. L. (2025, April 21). *Thomson Reuters v. ROSS: Brief for Amici Curiae in Support of Appellant-Defendant's Petition for Certification under § 1292(c)*. <https://ssrn.com/abstract=5225271>. Однако в данном случае вопрос защиты решал не столько судья, сколько суд присяжных, которые ориентировались скорее на мораль, чем на право. Здесь стрелка морального компаса повернулась в сторону правообладателя. О дополнительных проблемах, следующих из этого решения см.: Schwartz-croft (2024).

¹¹⁵ Ibid, pp. 25–39.

¹¹⁶ Ibid, pp. 27–30.

вания справедливым¹¹⁷. Принципиально, что суды отклонили аргумент, прозвучавший в деле *Thomson Reuters*, согласно которому само по себе существование «потенциального рынка лицензирования для обучения ИИ достаточно для признания рыночного ущерба». Верховный суд США, а также федеральные апелляционные суды последовательно подчеркивают, что ущерб в рамках четвертого критерия не может быть признан только на основании утраты гипотетического дохода от лицензирования: в противном случае любое несанкционированное, но преобладающее использование автоматически считалось бы недобросовестным¹¹⁸.

Следовательно, за пределами конкретных дел и фактов, с точки зрения абстрактных норм, обучение ИИ должно признаваться добросовестным использованием, особенно исходя из практики американских судов, по аналогии с предшествовавшим этому явлению интеллектуальным анализом данных.

3.2. Японский «рай» искусственного интеллекта

1 января 2019 г. в Японии были приняты поправки¹¹⁹ в Закон об авторском праве¹²⁰ (далее — Японский закон об авторском праве). В результате было установлено, пожалуй, самое широкое исключение для интеллектуального анализа данных на произведениях авторского права и объектах смежных прав¹²¹ во всем мире¹²².

Японское исключение, связанное с анализом данных, считается самым широким в мире по следующим причинам:

- оно применимо как для коммерческих, так и для некоммерческих целей;
- вне зависимости от возражений правообладателей;
- позволяет использовать произведения любым необходимым способом, если это нужно для анализа;
- не требует законного доступа к произведению.

Хотя поправка включает только три исключения, при которых интеллектуальный анализ данных не считается нарушением авторского права, в ней также предусмотрена специальная оговорка. Эта оговорка распространяет действие исключений на любые другие аналогичные способы использования и применяется *mutatis mutandis* (с учетом обстоятельств) к смежным правам. По сути, японское правительство гарантировало, что авторское право не станет препятствием для развития искусственного интеллекта, тем самым укрепив позиционирование Японии как «рая для ИИ и машинного обучения» (Dermawan, 2024, p. 54).

Статья 30-4 Японского закона об авторском праве регулирует использование произведений, не направленное на получение удовольствия от мыслей или чувств, выраженных в этих

¹¹⁷ Однако суды опять разошлись в оценке этого критерия с точки зрения теневого библиотек.

¹¹⁸ См.: *Tresóna Multimedia, LLC v. Burbank High School Vocal Music Association*, 953 F.3d 638 (9th Cir. 2020); *Bill Graham Archives v. Dorling Kindersley Ltd.*, 448 F.3d 605, 612–613 (2d Cir. 2006); *Google LLC v. Oracle America* (2021).

¹¹⁹ См. подробнее: Ueno (2021a, p. 213), см. также Japanese Copyright Research and Information Center (CRIC). (б.д.). *The history of copyright systems in Japan*. <https://www.cric.or.jp/english/csj/csj2.html>

¹²⁰ Chosakukenhō [Copyright Act], Act No. 48 of May 6, 1970, английский перевод доступен в (Japanese Law Translation [JLT DS]), <https://www.japaneselawtranslation.go.jp/en/laws/view/4207>

¹²¹ Но см.: Japan Copyright Office. (2024). *General understanding on AI and copyright in Japan — Overview*. https://www.bunka.go.jp/english/policy/copyright/pdf/94055801_01.pdf. Офис по интеллектуальной собственности Японии опубликовал проект документа, который должен разъяснить и, вероятно, сузить применение данной оговорки к обучению ИИ, по крайней мере в части обучения на материалах, полученных незаконным путем.

¹²² CRIC. (б.д.). *History of copyright systems in Japan*. <https://www.cric.or.jp/english/csj/csj2.html>

произведениях¹²³. И, как правильно указывает в своем анализе А. Дермаван, ключевой момент в понимании японского подхода к использованию произведений для обучения — различие между «удовольствием» (享受) и «использованием не ради удовольствия» (不見転) (Dermawan, 2024, p. 54)¹²⁴. Японское ведомство по авторским правам определяет удовольствие как «принятие и оценку эмоционально значимого или физически интересного» (Ueno, 2021b, p. 150). В случае литературных и художественных произведений «удовольствие» означает «наслаждение произведением через его восприятие органами чувств», а для компьютерных программ — «наслаждение функциями программы при ее выполнении» (Ueno, 2021, p. 150). Таким образом, действие характеризуется «удовольствием от идей или эмоций, выраженных в произведении», если оно направлено на удовлетворение интеллектуальных или эмоциональных потребностей зрителя (Ueno, 2021, p. 147). Проще говоря, японцы считают, что цель авторского права — гарантировать, что автор может получать компенсацию от тех, кто хочет наслаждаться его произведением. Однако защита авторским правом не требуется, если использование произведения не направлено на получение удовольствия или на то, чтобы его получил другой человек (например, интеллектуальный анализ данных) (Ueno, 2021b, p. 150–153).

Различение на использование «для удовольствия» и «не для удовольствия» помогает ограничению авторского права в части интеллектуального анализа данных пройти трехступенчатый тест Бернской конвенции (Dermawan, 2024). Этот тест включает три кумулятивных условия, которые в том числе дублируются в ст. 13 Соглашения по торговым аспектам прав интеллектуальной собственности¹²⁵: государства-участники должны ограничивать исключения и ограничения (1) «некоторыми специальными (достаточно определенными) случаями» (*certain*¹²⁶

¹²³ Статья разрешает эксплуатацию произведений любым способом и в необходимом объеме при выполнении одного из следующих условий или в аналогичных случаях, когда цель пользователя не заключается в получении личного удовольствия или доставлении удовольствия другим от содержания произведения:

- если произведение используется для тестирования, разработки или внедрения технологий, связанных с записью звуков, визуальных эффектов, или иной формы использования произведения;
- если произведение применяется для анализа данных, включая извлечение, сравнение, классификацию или иной статистический анализ языковых, звуковых, визуальных или других элементарных данных, извлеченных из большого количества произведений или объема данных (аналогичные положения содержатся в ст. 47-5(1)(ii));
- если произведение используется в процессе компьютерной обработки данных или иным способом, который не подразумевает, что его содержание воспринимается человеческими чувствами.

Исключение не применяется, если такие действия необоснованно ущемляют интересы правообладателей с учетом характера, цели произведения или условий его использования.

¹²⁴ Здесь нужно сделать ремарку, что японский язык многогранен и глубок, из-за чего одни и те же иероглифы в зависимости от контекста могут выражать разный смысл. На английский язык данный термин обычно переводят как *enjoyment*, но это тоже не является точной передачей смысла. Здесь имеется в виду скорее «чувственное восприятие» само по себе, вне зависимости от того, нравится произведение воспринимающему или нет и получает ли он при восприятии удовольствие в исконном смысле этого слова. Поэтому для описания японского подхода мы продолжим использовать термин «удовольствие» с учетом вышесказанного, но лучше использовать в этом контексте слово «потребление».

¹²⁵ Agreement on Trade-Related Aspects of Intellectual Property Rights [TRIPS], April 15, 1994. Annex 1C of the Marrakesh Agreement Establishing the World Trade Organization, Apr. 15, 1994, 1867 U.N.T.S. 154 [Marrakesh Agreement].

¹²⁶ В докладе панели Всемирной торговой организации слово *certain* было интерпретировано как указание на то, что любое исключение или ограничение должно быть представлено в ограниченной и четко определенной форме (Panel Report, *United States — Section 110(5) of the US Copyright Act*, WTO Doc. WT/DS160/R (adopted June 15, 2000) at § 6.62). См.: Hugenholtz, P.B., & Okediji, R. (2008, March 6). *Conceiving an international instrument on limitations and exceptions to copyright* (Amsterdam Law School Research Paper No. 2012-43). <https://ssrn.com/abstract=2017629>

*special*¹²⁷ cases), (2) «которые не вступают в противоречие с обычным использованием произведения» и (3) «необоснованно не ущемляют законный интерес правообладателя».

Ограничение для интеллектуального анализа в этом смысле является и определенным, и особым, так как оно является количественным, и качественным, ведь использование «не ради удовольствия» имеет довольно узкий охват возможных действий¹²⁸. Кроме того, в соответствии с количественным критерием условие использования «не ради удовольствия» ориентировано

¹²⁷ «Термин *special* должен пониматься так, что исключение или ограничение должно быть узким „по своему охвату и значению“» (Panel Report, *United States — Section 110(5) of the US Copyright Act*, WTO Doc. WT/DS160/R (adopted June 15, 2000) at § 6.112).

¹²⁸ Существует и противоположная точка зрения: ряд исследователей (Lucchi, 2025; Dornis & Stober, 2024, 2025) утверждают, что применение исключения для интеллектуального анализа данных к обучению генеративных моделей нарушает трехступенчатый тест. Во-первых, обучение ИИ на массово собранных корпусах утрачивает признак «специального случая» — это уже не исключительная исследовательская практика, а широкомасштабная и систематическая деятельность, ставшая индустриальной нормой. Во-вторых, такие системы, способные воспроизводить стилиевые и структурные особенности оригинальных произведений, нарушают нормальное использование охраняемого контента: они подменяют спрос как на само произведение, так и на рынок произведений работ: иллюстраций, переводов, дубляжа, адаптаций, аудиокниг и т. п., тем самым разрушая экономическую модель лицензионной эксплуатации. В-третьих, само обучение осуществляется без ведома и согласия авторов, без какого-либо механизма компенсации, что означает неоправданное ущемление их законных интересов как в имущественном, так и в морально-правовом аспекте. Однако такая аргументация, несмотря на убедительность на уровне этико-экономических оценок, вызывает сомнения по следующим причинам. Во-первых, для того чтобы вообще начать применять трехступенчатый тест, нужно занять позицию, что обучение — использование в юридическом смысле, в пользу чего, на наш взгляд, пока нет правовых аргументов, и для такого признания нужно специальное указание в законодательстве по этому поводу. Во-вторых, если специальный случай должен быть нераспространенным, то тогда большинство исключений, даже такие фундаментальные, как образовательное или научное цитирование, пародия, также неизбежно нарушают тест Бернской конвенции, поскольку эти исключения широко применяются и представляют собой нормализованную практику. Тем не менее международное и национальные законодательства давно признают их допустимыми, не ставя под сомнение их соответствие трехступенчатому тесту. Следовательно, более обоснованно трактовать признак «специальность» не как требование ограниченного или редкого применения исключения, а как необходимость **четкого и определенного очерчивания его границ** — как по цели, так и по субъектам, способам и условиям использования. Такой подход позволяет отличить произвольное, общее исключение от строго структурированного, правомерного механизма, встроенного в систему авторского права для обеспечения баланса между защитой и свободой использования. Оставшиеся два аргумента критики применения исключений для интеллектуального анализа данных к обучению ИИ взаимосвязаны и фактически сводятся к предположению, что такое обучение неизбежно приведет к снижению спроса на произведения живых авторов. Технически это вполне достоверный экономический аргумент, и он требует отдельного рассмотрения в следующем разделе (политико-правовой части статьи). Однако на этапе анализа соответствия трехступенчатому тесту необходимо подчеркнуть, что этот тест оценивает воздействие исключения на рынок существующих произведений, а не на гипотетический рынок произведений, которые могут быть созданы авторами в будущем. Так, на данный момент нет оснований полагать, что пользователи массово откажутся от оригинальных литературных или художественных произведений в пользу синтетического контента, сгенерированного на основе статистических моделей. Вряд ли кто-то перестанет покупать, например, «Властелина колец» Толкина только потому, что *ChatGPT* способен создать текст о путешествии группы героев с магическим артефактом в стиле фэнтези. В сфере художественного творчества личность автора, его реальное существование и уровень оригинальности (или по крайней мере личный «голос» автора, т. е. передача личного опыта посредством произведения) остаются ключевыми факторами, определяющими воспринимаемую ценность произведения и покупательское поведение.

на цели общественных интересов, для достижения баланса между авторским правом и общественными интересами, включая образование, исследования ИИ, прозрачность информации и право на свободу выражения¹²⁹. Из этого следует, что использование «не для удовольствия» само по себе не противоречит второй части теста, так как оно и есть критерий отделения нормального использования от ненормального.

Как отмечалось экспертами панели Всемирной торговой организации¹³⁰ по этому вопросу: «Исключение или ограничение исключительного права <...> вступает в конфликт с нормальным использованием произведения, если действия, которые подпадают под это право, но исключены в рамках таких ограничений, начинают составлять экономическую конкуренцию способам, с помощью которых правообладатели обычно получают доход от использования этого права (например, авторского), и тем самым лишают их значительной или ощутимой коммерческой выгоды»¹³¹.

Выше была описана разница между потенциальной возможностью воспроизвести произведение и воспроизведением произведения. Выведенный из существующего законодательства запрет на возможность обучения на произведениях ударит не только по разработчикам ИИ, но и по обычным людям, изучающим литературное и художественное мастерство на примерах других работ. Экономическая возможность извлекать выгоду из произведения подразумевает воспроизведение его для потребления, т. е. для потенциального получения удовольствия, каким бы оно ни было, а не для его анализа.

Точно так же можно рассуждать и о возможном законном интересе правообладателя. Отдельный правообладатель не имеет законного интереса в использовании его произведения для обучения ИИ, поскольку в контексте обучения одно произведение оказывает минимальное влияние на дальнейшую работу модели. Большие данные получают свою ценность именно в совокупности, когда их много. Таким образом, если рассматривать использование отдельных произведений в контексте обучения ИИ, они не обладают самостоятельной экономической ценностью (Kollár, 2021 p. 28; Directorate General for the Internal Market and Services & De Wolf & Partners, 2014, p. 47) и их участие в обучении не может само по себе привести к необоснованной утрате дохода правообладателя от использования конкретного произведения.

Впрочем, в самой Японии продолжаются дискуссии о правильном толковании оговорки о «наслаждении» и ее соответствии третьему пункту трехступенчатого теста Бернской конвенции.

Проблема рыночного замещения, вызванная генеративным ИИ, проявляется прежде всего в специфических сферах, таких как фотостоки, автоматизированный копирайтинг и иные форматы с минимальным уровнем оригинального творчества. Но именно в силу невысокой творческой составляющей сложно утверждать, что интересы соответствующих авторов нарушаются в смысле, предусмотренном Бернской конвенцией: они теряют позиции скорее как поставщики рутинного или стандартизированного контента, нежели как носители уникального авторского выражения, достойного исключительной охраны. На июль 2025 г. только 4,6% страниц, находящихся в топе поисковой выдачи Google, полностью сгенерированы нейросетями. Более того, по данным исследования Ahrefs, страницы, занимающие первые позиции, как правило, содержат меньше ИИ-контента, чем страницы более низкого ранга (см.: Southern, M. G. (2025, July 8). Ahrefs study finds no evidence Google penalizes AI content. *Search Engine Journal*. <https://www.searchenginejournal.com/ahrefs-study-finds-no-evidence-google-penalizes-ai-content/550656/>).

¹²⁹ По проблематике соотношения частных и общих интересов в Бернской конвенции см.: Ricketson (1987, p. 535).

¹³⁰ Панели Всемирной торговой организации — это временные экспертные группы, создаваемые в рамках Всемирной торговой организации для разрешения споров между странами-членами. Эти панели играют ключевую роль в механизме урегулирования споров данной организации, помогая интерпретировать правила и соглашения и обеспечивать их соблюдение.

¹³¹ Panel Report, *United States — Section 110(5) of the US Copyright Act*, WTO Doc. WT/DS160/R. (adopted June 15, 2000) at § 6.183

Японское агентство по авторским правам (*Japan Copyright Office*) недавно, в 2024 г., опубликовало проект документа, в котором предложены уточнения в применении этой оговорки.

Агентство выделяет две специфические ситуации, которые могут рассматриваться как использование произведения для наслаждения, несмотря на то что между таким использованием и восприятием нет прямой связи:

- **Точная настройка модели** (*fine-tuning*) или использование данных для генерации на основе извлечения (RAG). Если ИИ специально настраивается для копирования конкретных произведений или воссоздания ключевых черт, защищаемых авторским правом.
- **Целенаправленное копирование стиля.** Когда ИИ интенсивно анализирует и воспроизводит уникальные элементы стиля конкретного автора, чтобы воссоздавать их в новых произведениях.

Таким образом, если модель специально адаптируют для создания произведений, которые полностью или частично имитируют оригинальные элементы автора, такие действия могут быть квалифицированы как использование произведения для наслаждения.

Фактически Агентство подчеркивает, что копирование самого произведения является нарушением, если было произведено без согласия, а появление охраняемых элементов произведения на этапе генерации нового произведения через ИИ тоже может привести к применению средств авторско-правовой защиты. При этом не зря делается акцент на том, что такая точная настройка или копирование стиля должны быть целенаправленными. Фактически данные действия можно рассматривать как разновидность шиканы, или недобросовестного поведения. Одно дело, если во время обычного обучения ИИ появляется потенциальная возможность воспроизведения защищаемых авторским правом произведений. При таком подходе разработчик освобождается от ответственности, а вопрос правомерности использования, например генерации текста или изображения, возлагается на пользователя, которому предстоит обосновать, соответствует ли это использованию критериям добросовестного или справедливого использования в зависимости от применимой юрисдикции. Другое дело, если разработчик специально настраивает свою модель таким образом, чтобы она целенаправленно нарушала права конкретного лица. Такое, конечно же, недопустимо. Тем не менее должна сохраняться прямая причинная связь между действием, вредом и обходом средств защиты: нарушение не должно быть только потенциальным. Следует еще раз подчеркнуть, что регулирование обучения и дальнейшего использования ИИ — это два разных вопроса.

С другой стороны, в документе указывается на необходимость вывести из-под ограничения, предусмотренного статьей 30-4 Японского закона об авторском праве, интеллектуальный анализ, который осуществляется на различных пиратских источниках, т. е. ввести требование законного доступа. В этом, конечно, есть своя логика. Если доступ к копии был получен незаконно, то здесь появляется прямой и непосредственный вред правообладателю, ведь он не получил компенсацию за нормальное использование (т. е. за возможность потребления). Самого по себе нарушения авторских прав со стороны пользователя пиратского сайта может и не быть, если он сам не создал себе новую копию на основе копии из нелегитимного источника, и вред остается чисто экономическим, но введение такого критерия позволяет избежать появления даже такого чисто экономического вреда¹³².

Впрочем, это скорее вопрос политики права, чем догматики. Догматически не может считаться использованием произведения копирование его неохраняемых элементов (т. е. абстрактных метаданных), даже если у лица, совершающего копирование, появляется возможность такого

¹³² Противоположное мнение о том, что такое исключение необходимо, см.: Frankel & Kellogg (2012).

использования неохранных элементов, которое приведет к нарушению уже охраняемых элементов¹³³.

Японский подход можно рассматривать как развитие концепции «невпечатляющего использования», известной в американской правовой доктрине, но закрепленной на законодательном уровне. При этом японское авторское право основывается преимущественно на континентальной правовой традиции, признавая личные неимущественные права авторов, защищая форму произведения и исключая охрану содержания. Поэтому и в японском праве выводы о допустимости использования произведений «вне удовольствия» могли быть сделаны и на основании общих положений. Однако специальные нормы об интеллектуальном анализе данных всё же полезны, так как позволяют уточнять рамки правоприменения, сужая или расширяя подход, например в вопросах обучения на неправомерно полученных данных или создания временных копий для целей интеллектуального анализа данных.

4. Политико-правовые аргументы

В дискуссии о возможности обучения нейросетей на объектах авторского права также приводят несколько довольно сильных политико-правовых аргументов в пользу запрета какого-либо обучения без прямо выраженного согласия.

Аргумент 1: «авторский апокалипсис»

Некоторые утверждают, что искусственный интеллект и искусственное творчество в конце концов полностью захватят мир, рынок и любую работу, оставив авторов без средств к существованию. В результате люди перестанут творить, так как нейросети будут делать всё за них.

Впрочем, на данный момент эта перспектива скорее выглядит как сценарий фильма ужасов, чем как реальная угроза, и вот почему. Современный искусственный интеллект — это далеко не тот интеллект, который мы видим в фантастических кинокартинах. Сегодня это просто новый инструмент, расширяющий возможности пользователей. Да, возможно, обидно учиться 5–10 лет рисовать цифровые картины, если теперь для того, чтобы сделать то же самое, нужно потратить гораздо меньше времени. Но это и есть суть прогресса.

Если подходить к этому аргументу серьезно, то получается, что современные авторы — это не просто люди, а некая особая каста, обладающая исключительным знанием и способностями. Имплицитно в этом аргументе содержится мысль о том, что те, кто использует нейросети, не являются творцами, в отличие от авторов, работающих без них. Но современные нейросети не способны генерировать что-то совершенно новое с точки зрения внутреннего смысла. Они могут лишь создать новую форму, которая качественно не отличается от того, что сделал бы достаточно опытный художник при условии, что ему дадут четкое техническое задание.

Наделение результатов творчества смыслом и создание уникальных выражений существующих идей всё еще остаются прерогативой человека. Если раньше для того, чтобы научиться писать и рисовать, нужно было потратить огромное количество времени и усилий, то с развитием нейросетей всё больше людей с уникальным видением, но без достаточных умений смогут реализовываться как творцы. Обидно, когда то, на что вы потратили десятую часть жизни, другой человек может освоить за пару лет обучения, но это не является достаточным аргументом для того, чтобы тормозить прогресс.

¹³³ Что не отменяет, конечно, того факта, что хранение копии, полученной незаконным путем, может расцениваться как самостоятельное нарушение.

Ведь в улучшении нейросети есть и большой социальный интерес. Нейросети открывают огромный простор для деятельности людей с ограниченными возможностями, позволяя им преодолевать физические и когнитивные барьеры. Эти технологии помогают улучшить качество жизни, обеспечивая большую автономию и участие в социальной жизни, что, в свою очередь, способствует интеграции в общество и личностной реализации. Нейросети могут помочь людям с ограниченными возможностями в творчестве, предоставляя им инструменты для выражения своих идей и чувств, независимо от физических ограничений, например людям с нарушениями зрения — программы для генерации аудиовизуальных произведений, а людям с нарушениями слуха — способ создавать музыку или использовать текстовые программы для работы с языком. Люди с моторными нарушениями посредством нейросети могут использовать жесты или движения глаз для создания цифровых картин, музыки или текста и тем самым участвовать в творчестве, несмотря на ограничения. Эти технологии открывают новые горизонты для самовыражения и инклюзивного участия в искусстве.

Аргумент 2: коммерческое использование произведений

Еще популярен довод о том, что коммерческое использование ИИ, обученного на авторских произведениях, влечет коммерческое использование самих произведений. Этот аргумент строится на идее, что если продукт, созданный с использованием обученной модели, приносит доход, авторы произведений, использованных для обучения, также должны получить компенсацию.

Другими словами, если ты обучаешься на чем-то, а потом используешь полученные знания в предпринимательской деятельности, то ты таким образом используешь в хозяйственной деятельности те объекты, на которых обучался. С одной стороны, это сильный аргумент, так как построен на базовом принципе свободного рынка: если у тебя есть ценность, то за ее использование другим людям нужно платить. С другой стороны — это то, что можно назвать «выстрелом себе в ногу» или «пилением сука, на котором сидишь».

Ведь процессы обучения нейросети и обычного автора не настолько сильно различаются. А если нельзя нейросети, то почему можно людям? Даже сейчас, если посмотреть на то, как учатся или пишут картины художники, они всегда используют, а точнее, вдохновляются работами других. Те, кто менее этичен, могут рисовать то же самое, но не точь-в-точь, чтобы не выглядело полным плагиатом. Высокоморальные повторяют только технические приемы, чтобы улучшить собственные навыки. Но в любом случае использовать картины других, чтобы отработать конкретный навык, — нормальный этап обучения любого художника в современном мире. Но даже потом, на этапе собственного творчества, художники всё равно обращаются к референсам, т. е. к картинам других авторов, в качестве дополнительного источника визуальной информации¹³⁴. Конечно, правильное применение референсов — это не копирование, и поэтому сейчас оно не приводит к нарушению авторских прав. Фактически то, как художники используют референсы, — это то же самое, что делает программа при обработке произведений для включения в базу данных, которая будет использоваться в обучении. Они точно так же снимают с референсов метаданные, которые потом используют для реализации уже собственных идей.

А современный ИИ на данный момент даже не может использовать метаданные для реализации собственных идей в силу отсутствия их в частности, да и сознания в целом. Всегда

¹³⁴ См. подробнее: Artland. (2023). *Референс в рисовании*. <https://izokurs.ru/blog/referens-v-risovanii/>; CG Speak. (2021). *Референсы в рисовании 2*. YouTube. <https://www.youtube.com/watch?v=4VUVAZ2JSYA>; SmirnovSchool. (2017). *Рисунок и референс: как избежать копирования. Ответы на вопросы*. YouTube. <https://www.youtube.com/watch?v=6AZNNpVctgg>

нужен другой человек, который уже будет применять вычислительные мощности машины в качестве инструментов. Поэтому этот аргумент может точно так же быть обращен против всей индустрии, если мы не будем придерживаться четкого юридического критерия, при котором нарушение происходит только при условии воспроизведения, а не потенциальной возможности воспроизведения.

Ведь легко помыслить налог на обучение, сходный с правом на вознаграждение за свободное воспроизведение фонограмм и аудиовизуальных произведений в личных целях, именуемым «налогом на болванки»¹³⁵. Вознаграждение за свободное воспроизведение фонограмм и аудиовизуальных произведений в личных целях никак не учитывает, что многие носители, за которые взимается данный сбор, не используются для воспроизведения, а могут быть применены для других целей. То есть сбором облагается не столько само воспроизведение, сколько потенциальная возможность¹³⁶. А значит, такой же подход можно перенести на обучение. Даже если конкретный художник не обучался на картинах других, многие всё равно это делали. Даже если конкретный художник не переносит части произведений, на которых он обучался, или использует свои произведения в виде референсов (некоторые могут это делать неосознанно, хоть это и несущественно).

Но главное — ИИ сам по себе не креативен: его «творчество» — это на самом деле всегда максимальная банальность. В некоторых случаях она и требуется, например при определении болезни или ситуации на дороге. Но в случае с художественными и научными произведениями, банальность не нужна никому. Человек создает произведение искусства, опираясь на личный жизненный опыт, эмоциональный мир и культурные смыслы. Нейросеть же функционирует как статистическая модель, выбирающая и комбинирующую фрагменты из существующих данных. Человек, наоборот, является носителем авторского замысла и самосознания. По данным экспериментов, творцы используют воображение и оригинальные художественные метафоры, тщательно продумывают сюжет и композицию, придавая работе уникальную цельность. AI-модель же не имеет собственных предпочтений и не стремится выразить личные смыслы: она просто реагирует на входной запрос (Komariah, 2024; Rebelo et al., 2022). Проще говоря, алгоритм не знает любви, не знает утраты и сострадания: слова и образы для него лишь статистические выборки. Поэтому, чтобы ИИ выдал что-то действительно творческое, нужно, чтобы его использовал как инструмент тот, кто возможностью творчества обладает, — человек. И даже в этом случае человеку придется приложить усилия; впрочем, по-иному в искусстве и не бывает.

Указанные особенности делают «искусственные» произведения скорее технической симуляцией творчества. Следовательно, ИИ-произведения не могут выполнять ту же функцию, что искренние работы человека: они не вызывают того же эмоционального отклика и редко имеют глубокий культурный смысл. Поэтому ИИ-произведения никогда не заменят для людей глубоких, эмоциональных и, главное, оригинальных произведений, создаваемых другими людьми. Опасность здесь появляется только для тех авторов, чье творчество изначально было достаточно банальным или неоригинальным. Лучший пример — фотографии с фотостоков или продающие тексты на маркетинговых сайтах. Действительно, те, кто работал в такой сфере, могут потерять часть заработка, если не смогут показать, в чем их преимущество перед машиной. Однако здесь скорее вопрос в том, почему произведения с очень низким или вообще

¹³⁵ См., например: «Налог на болванки» расширят на все ввозимые цифровые носители (2022). *Habr*. <https://habr.com/ru/news/676714/>

¹³⁶ Из-за этого очень большие вопросы возникают к справедливости такого сбора. См. подробнее: Gupta & Devaiah (2016); Poort & Quintais (2013).

нулевым уровнем творчества подпадают под защиту авторского права¹³⁷. Но даже если не вдаваться в такие экзистенциальные вопросы, это всё равно выглядит как требование создателей повозок и кучеров платить им роялти за каждый сбор и дальнейшее использование машин и самолетов.

Аргумент о коммерческом использовании был бы убедительнее, если бы использование авторских произведений было необходимым и существенным элементом работы ИИ и если бы эта работа составляла основную коммерческую ценность технологии. Однако проблема заключается в том, что влияние одного произведения на работу конкретного ИИ крайне мало.

Даже для того, чтобы научить ИИ имитировать стиль конкретного художника, нужно как минимум 200 его произведений (Verma et al., 2024). При этом чем новее модель и чем больше данных ей «скормлено» для обучения, тем выше этот порог (Verma et al., 2024). Это показывает, что даже большое количество произведений одного автора оказывает непропорционально малое влияние на работу модели в целом. Важнейшим фактором здесь являются большие данные, часть которых может вообще не иметь авторских прав, даже если они не используются для извлечения метаданных.

Если модель не была создана специально для имитации произведений конкретного автора и не становилась его «цифровой личностью», то справедливая компенсация за использование произведений будет исчисляться крошечными суммами. Установление фиксированного сбора за каждое произведение, с учетом эффекта масштаба и необходимости переработки огромного количества данных, приведет к нерентабельности разработки ИИ.

Аргумент 3: эффект масштаба

Здесь можно было бы возразить, что существует эффект масштаба, которого не достигает использование произведений для обучения обычными людьми. Эффект масштаба — это экономический эффект, при котором с ростом объема производства или использования технологии снижаются издержки на единицу результата, а влияние на рынок или среду усиливается непропорционально. В контексте ИИ это означает, что система, обученная на чужих произведениях, может генерировать массовые и мгновенные копии, что радикально отличается от ограниченных действий одного человека. Хотя человек может научиться копировать чужие работы или создавать похожие работы в определенном творческом стиле, это не приведет к такому же негативному эффекту, как использование нейросети. Человек ограничен только своими физическими возможностями и количеством копий, которые он может создать. В отличие от этого с помощью нейросети огромное количество людей может одновременно создавать копии чужих произведений, что значительно увеличивает масштабы возможных нарушений авторских прав, либо достаточно профессиональные картины, чтобы заменить авторов на существующем рынке.

Это ощущение можно считать эфемерным, поскольку с техническим прогрессом копирование чужих произведений и так стало значительно проще. Нейросеть в этом контексте не представляет собой нечто революционное по сравнению с обычными инструментами, такими как сканеры или программы вроде *Photoshop*. Этот аргумент смещает фокус с момента нарушения на потенциальную возможность нарушения, т. е. мы уже предполагаем, что определенная технология будет использоваться недобросовестно. Однако всё что угодно может быть

¹³⁷ О проблематике защиты произведений с низким уровнем творчества см. работы А. В. Кашанина (Kashanin, 2007b, 2007a, 2009, 2010b, 2010a, 2011, 2012), О том, что многие проблемы авторского права могут быть решены с введением требования «оригинальности», см. Н. В. Щербак (Shcherbak, 2023).

использовано недобросовестно, главное — является ли недобросовестное использование единственно возможным или хотя бы преимущественно возможным способом применения технологии. В данном случае это не так.

ИИ можно использовать множеством способов, не нарушая авторских прав других людей. Практически все современные нейросети оснащены специальными надстройками — этическими правилами, которые значительно затрудняют создание контента, способного нарушать чьи-то права. Конечно, эти правила могут быть обойдены, но для этого нужно иметь прямой умысел на совершение правонарушения. Как в случае с преступлением, когда наказывают не того, кто продал нож, а того, кто его использовал во вред, здесь нельзя запрещать технологию просто потому, что ее могут применить во вред.

С другой стороны, если обратить внимание на эмпирические данные, можно заметить, что для многих людей важен источник произведения. Многие предпочитают произведения, созданные человеком, и довольно негативно относятся к необоснованному использованию нейросетей¹³⁸. В отличие от ситуации с машинами и лошадьми, где машины практически полностью вытеснили лошадей как средство передвижения, цифровая живопись не вытеснила традиционное искусство. Оба направления по-прежнему существуют параллельно.

Что касается текстов, то пока нет прогнозов, что нейросеть сможет создать художественное произведение на уровне, позволяющем говорить о вкладе в профессиональную литературу или даже о создании бестселлера (Komariah, 2024; Rebelo et al., 2022). Оба этих достижения требуют работы со смыслами и генерации новых идей, а не простого переписывания одних и тех же текстов по несколько раз. Это остается фундаментальным ограничением современного ИИ до тех пор, пока не появится ИИ, обладающий чем-то похожим на человеческое сознание.

Как видно, политико-правовые аргументы в пользу взимания средств за использование произведений для обучения ИИ являются, по крайней мере, спорными. Эти аргументы явно недостаточны для того, чтобы игнорировать формально-юридический анализ и наделить правообладателей правом полного контроля за произведениями в процессе обучения ИИ. Однако они могут оказать влияние на возможные точечные ограничения, направленные на регулировку баланса интересов.

Заключение

Какие выводы можно сделать из анализа процесса обучения искусственного интеллекта?

1. Процесс обучения ИИ состоит из трех последовательных шагов: интеллектуальный анализ данных, сбор полученных данных в базу и обучение нейросети на этом датасете.
2. Утверждение о том, что ИИ обучается на объектах авторского права, не представляется справедливым в юридическом смысле. ИИ обучается на метаданных (паттернах, синтаксисе, методах и т. д.), полученных благодаря интеллектуальному анализу данных.
3. Следовательно, вопрос возможного нарушения прав правообладателей с точки зрения авторского права должен рассматриваться применительно к анализу данных, а не к использованию результатов для обучения ИИ, ведь только в момент анализа данных происходит взаимодействие с объектом авторских или смежных прав.
4. С учетом того что благодаря интеллектуальному анализу данных из произведений или баз данных выделяются метаданные, которые сами по себе не являются охраноспособными,

¹³⁸ См., например: *Wizards of the Coast* попала в скандал с AI-артом из новой книги по D&D (2023). *Мир фантастики и фэнтези*. <https://www.mirf.ru/news/wizard-of-the-coast-popala-v-skandal-s-ai-artom-iz-novoj-knigi-po-d-d>

такой анализ нельзя назвать использованием в правовом смысле, а значит, и дальнейшее применение таких метаданных для обучения ИИ не может считаться нарушением прав правообладателей.

5. Исключением здесь будет использование баз данных, специально созданных для обучения ИИ, так как это единственный случай, когда в обучении ИИ будут использоваться сами материалы базы данных, а не результат стороннего интеллектуального анализа базы данных.
6. При этом всё равно разумно включить интеллектуальный анализ данных в законодательство для снятия вопроса о возможности создания временных копий для такого анализа (и других случаев «невпечатляющего воспроизведения»).
7. Ради достижения баланса интересов государства могут вводиться дополнительные условия в отношении такого анализа данных и дальнейшего обучения.

Первое условие — это правомерный доступ к произведению. Разработчик ИИ должен сначала приобрести копию, тем самым реализовывая коммерческий интерес правообладателя перед обучением.

Второе условие — это запрет на целенаправленное обучение ИИ (обучение с учителем) в целях подражания конкретному автору, его стилю через генерацию охраняемых элементов его работ как главный способ использования ИИ.

Третье условие — это ограничение возможности интеллектуального анализа через резервирование прав, т. е. при помощи прямого отказа правообладателя от участия его произведений в интеллектуальном анализе данных, который в случае распространения в цифровой среде должен быть выполнен машиночитаемыми средствами. Однако это условие будет эффективным, только если будут предусмотрены внедоговорные возможности для разработчиков ИИ использовать такие произведения для обучения.

Кроме этого, можно выделить еще несколько результатов анализа, которые являются частью аргументации позиции о том, что обучение ИИ не является использованием с точки зрения авторских или смежных прав.

1. Необходимо различать «невпечатляющее» и «впечатляющее использование».
2. «Впечатляющее использование» — это такое использование, которое приводит к потреблению произведения людьми с помощью их органов чувств либо эксплуатации объекта по его функциональному назначению (программы для ЭВМ). «Невпечатляющее использование», соответственно, не приводит к такому восприятию или эксплуатации функциональных свойств объекта.
3. Только «впечатляющее использование» (воспроизведение, публичное исполнение, доведение до всеобщего сведения и т. д.) или шаги, которые очевидно в конце концов ведут к такому виду использования (распространение, экспорт, передача по кабелю и т. д.), можно считать использованием в правовом смысле, или, иными словами, только контроль за возможностью «впечатляющего использования» составляет исключительное право правообладателя на его произведение.

Послесловие

Как показывает анализ, вопрос возможности обучения ИИ на объектах авторских и смежных прав остается одним из самых острых и дискуссионных в современной правовой практике. Развитие ИИ действительно способно существенно изменить рынок литературных и художественных произведений: некоторые работы могут утратить востребованность. Однако,

как часто бывает, закрытие одной двери открывает другую. Искусственный интеллект становится новым инструментом, который дает возможность большему числу людей проявить себя в качестве творца. Особенно это важно для людей с ограниченными возможностями, которые не в состоянии использовать кисть или ручку так, как это делают другие. И если мы говорим, что прогресс и удобство — это хорошо, то мы не можем чинить препятствия к развитию генеративных нейросетей, только исходя из того, что это может как-то повлиять на чьи-то потенциальные доходы.

В этом как раз помогает доктринальный анализ, а также законодательные решения. Большая часть стран имеют различные ограничения авторского права, которые касаются интеллектуального анализа данных, пусть в некоторых случаях и недостаточно широкие, чтобы говорить о любом анализе (Fiil-Flynn, 2022). Даже Европейский союз, известный своим потребительским, прочеловеческим подходом, например в сфере защиты прав потребителей или в вопросах защиты персональных данных, в данном вопросе занял позицию скорее в пользу разработчика. Это неудивительно в силу того, что интеллектуальный анализ данных сейчас — это базовый инструмент в цифровую эпоху и исходит из фундаментальной свободы научной и исследовательской деятельности (Geiger, 2021). Европейский законодатель попытался найти золотую середину, позволив правообладателям отказываться от возможности обучения на данных, извлекаемых из их произведений, путем проставления необходимых запретов на машинночитаемом языке. Возможно, как и при введении любого условно-запретительного порядка, гораздо больше будет тех, кто не отказался, чем отказавшихся, т. е. откажутся только те, кто чувствует себя несправедливо обиженным и обделенным, так что эффект будет не таким плохим. Но авторское право и так дает достаточную защиту произведениям и не стоит ее расширять еще больше там, где это не следует из его сущности.

Исследования показывают, что просто «голый» режим *opt-out* неэффективен и не позволяет действительно ни уравновесить разнонаправленные интересы, ни способствовать прогрессу. Интеллектуальный анализ — это не то, что должно по умолчанию входить в сферу регулирования авторского права. Это не столько вопрос частной автономии и связи автора с его произведением, сколько политическое решение для баланса интересов между инновациями и статусом-кво. Нужна дополнительная политическая воля, чтобы запретить то, что сейчас прямо не запрещено, ведь это основной метод регулирования гражданско-правовых отношений. Осталось только добавить, что потребление — это не юридическое использование, но любое юридическое использование имеет конечной целью потребление, чтобы понять одну из границ, где заканчивается объект авторского права и начинаются его неохранные элементы.

С другой стороны, и интерес, и опасения правообладателей тоже понятны. Однако сама по себе потеря чьей-то прибыли не может быть достаточным аргументом для ограничения фундаментальных свобод и замедления прогресса. Распространить авторское право на метаданные, содержащиеся в произведении, — это всё равно, что подвести под авторское право стиль или идею. Кому-то это, конечно, окажется выгодным в короткой перспективе, но вот в отдаленной — хуже станет абсолютно всем.

Список литературы / References

1. Abramova, E. N., & Khamidullina, E. V. (2024). Prava na rezul'taty intellektual'noy deyatel'nosti, sozdannyye s ispol'zovaniyem iskusstvennogo intellekta [Rights to intellectual deliverables created with the use of artificial intelligence]. *Khozyaystvo i Pravo*, 10, 71–83. <https://doi.org/10.18572/0134-2398-2024-10-71-83>

2. Anikin, A. S. (2022). K voprosu ob okhranosposobnosti rezul'tatov deyatel'nosti iskusstvennogo intellekta kak ob'yekta intellektual'noy sobstvennosti [On the protectability of the results of artificial intelligence activity as an object of intellectual property]. *Tsivilist*, 2, 25–31.
3. Astrachan, J. B. (2008). *De minimus copyright infringement* (SSRN Scholarly Paper No. 1625037). Social Science Research Network. <https://doi.org/10.2139/ssrn.1625037>
4. Barthes, R. (1989). *Ot proizvedeniya k tekstu. Izbrannyye raboty: Semiotika i poetika* [From work to text. Selected works: Semiotics and poetics]. Progress.
5. Benhamou, E. (2022). *Machine learning fundamentals: Unsupervised learning part 2. Data & AI reskilling seminar slides*. Social Science Research Network. <https://doi.org/10.2139/ssrn.4234521>
6. Bloch, D. A. (2023). *Machine learning: Models and algorithms* (SSRN Scholarly Paper No. 4493977). Social Science Research Network. <https://doi.org/10.2139/ssrn.4493977>
7. Borisova, L. B. (2024). O ponyatii iskusstvennogo intellekta i pravovom rezhime proizvedeniy, sozdannykh im bez tvorcheskogo uchastiya cheloveka [On the concept of artificial intelligence and the legal regime of AI-generated results without the creative participation of an individual]. *Actual Problems of Russian Law*, 19(8), 100–113. <https://doi.org/10.17803/1994-1471.2024.165.8.100-113>
8. Carrier, M. (2004). Cabining intellectual property through a property paradigm. *Duke Law Journal*, 54(1), 1–145.
9. Carroll, M. W. (2019). Copyright and the progress of science: Why text and data mining is lawful. *UC Davis Law Review*, 53, 893–964.
10. Chander, A., & Sunder, M. (2004). The romance of the public domain. *California Law Review*, 92(5), 1331–1373. <https://doi.org/10.2307/3481419>
11. Chen, W., Zhang, L., Zhong, L., Peng, L., Wang, Z., & Shang, J. (2025). *Memorize or generalize? Evaluating LLM code generation with evolved questions* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2503.02296>
12. Coase, R. H. (1960). The problem of social cost. *The Journal of Law & Economics*, 3, 1–44.
13. Cooper, A. F., & Grimmelmann, J. (2024). *The files are in the computer: Copyright, memorization, and generative AI* (Version 6). arXiv. <https://doi.org/10.48550/ARXIV.2404.12590>
14. Cooper, A. F., Choquette-Choo, C. A., Bogen, M., Jagielski, M., Filippova, K., Liu, K. Z., Chouldechova, A., Hayes, J., Huang, Y., Miresghallah, N., Shumailov, I., Triantafillou, E., Kairouz, P., Mitchell, N., Liang, P., Ho, D. E., Choi, Y., Koyejo, S., Delgado, F., ... Lee, K. (2024). *Machine unlearning doesn't do what you think: Lessons for generative AI policy, research, and practice* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2412.06966>
15. Cooper, A. F., Gokaslan, A., Ahmed, A., Cyphert, A. B., De Sa, C., Lemley, M. A., Ho, D. E., & Liang, P. (2025). *Extracting memorized pieces of (copyrighted) books from open-weight language models* (Version 2). arXiv. <https://doi.org/10.48550/ARXIV.2505.12546>
16. Dermawan, A. (2024). Text and data mining exceptions in the development of generative AI models: What the EU member states could learn from the Japanese “nonenjoyment” purposes? *The Journal of World Intellectual Property*, 27(1), 44–68. <https://doi.org/10.1111/jwip.12285>
17. Dornis, T. W., & Stober, S. (2024). *Urheberrecht und Training generativer KI-Modelle: Technologische und juristische Grundlagen* [Copyright and training of generative AI models: technological and legal foundations]. Nomos. <https://doi.org/10.5771/9783748949558>
18. Dornis, T. W., & Stober, S. (2025). *Generative AI training and copyright law* (No. arXiv:2502.15858). arXiv. <https://doi.org/10.48550/arXiv.2502.15858>
19. Dozortsev, V. A. (2003). *Intellektual'nyye prava: Ponyatiye. Sistema. Zadachi kodifikatsii* [Intellectual rights: Concept. System. Codification challenges]. Statut.
20. Drexl, J., Hilty, R., Beneke, F., Desautettes-Barbero, L., Finck, M., Globocnik, J., Gonzalez Otero, B., Hoffmann, J., Hollander, L., Kim, D., Richter, H., Scheuerer, S., Slowinski, P. R., & Thonemann, J. (2019). *Technical aspects of*

- artificial intelligence: An understanding from an intellectual property law perspective* (SSRN Scholarly Paper No. 3465577). Social Science Research Network. <https://doi.org/10.2139/ssrn.3465577>
21. Ducato, R., & Strowel, A. (2021). Ensuring text and data mining: Remaining issues with the EU copyright exceptions and possible ways out. *European Intellectual Property Review*, 43(5), 322–337.
 22. Dusollier, S. (2018). Realigning economic rights with exploitation of works: The control of authors over the circulation of works in the public sphere. In B. Hugenholtz (Ed.), *Copyright reconstructed: Rethinking copyright's economic rights in a time of highly dynamic technological and economic change* (pp. 163–201). Wolters Kluwer Law International.
 23. Entin, V. L. (2017). *Avtorskoye pravo v virtual'noy real'nosti: Novyye vozmozhnosti i vyzovy tsifrovoy epokhi* [Copyright law in virtual reality: New opportunities and challenges of the digital age]. Statut.
 24. European Parliament. Directorate General for Citizens' Rights, Justice and Institutional Affairs. (2025). *Generative AI and copyright :training, creation, regulation*. Publications Office. <https://data.europa.eu/doi/10.2861/0365517>
 25. Directorate General for the Internal Market and Services & De Wolf & Partners. (2014). *Study on the legal framework of text and data mining (TDM)*. European Commission. Publications Office. <https://data.europa.eu/doi/10.2780/1475>
 26. Fil-Flynn, S. M., Butler, B., Carroll, M., Cohen-Sasson, O., Craig, C., Guibault, L., Jaszi, P., Jütte, B. J., Katz, A., Quintais, J. P., Margoni, T., De Souza, A. R., Sag, M., Samberg, R., Schirru, L., Senftleben, M., Tur-Sinai, O., & Contreras, J. L. (2022). Legal reform to enhance global text and data mining research. *Science*, 378(6623), 951–953. <https://doi.org/10.1126/science.add6124>
 27. Frankel, S., & Kellogg, M. (2012). Bad faith and fair use. *Journal of the Copyright Society of the USA*, 60, 1–36.
 28. Garbacea, C., & Mei, Q. (2020). *Neural language generation: Formulation, methods, and evaluation* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2007.15780>
 29. Gastaldi, J. L., Terilla, J., Malagutti, L., DuSell, B., Vieira, T., & Cotterell, R. (2024). *The foundations of tokenization: Statistical and computational concerns* (Version 4). arXiv. <https://doi.org/10.48550/ARXIV.2407.11606>
 30. Ge, J., Tang, S., Fan, J., & Jin, C. (2023). *On the provable advantage of unsupervised pretraining* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2303.01566>
 31. Geiger, C. (2021). The missing goal-scorers in the artificial intelligence team: Of big data, the fundamental right to research and the failed text and data mining limitations in the CSDM Directive. In M. Senftleben, J. Poort, M. M. van Eechoud, S. van Gompel, & N. Helberger (Eds.), *Intellectual property and sports: Essays in honour of P. Bernt Hugenholtz* (pp. 383–394). Wolters Kluwer Law International. <https://doi.org/10.2139/ssrn.3829768>
 32. Geiger, C., & Iaia, V. (2024). The forgotten creator: Towards a statutory remuneration right for machine learning of generative AI. *Computer Law & Security Review*, 52, 105925. <https://doi.org/10.1016/j.clsr.2023.105925>
 33. Gupta, I., & Devaiah, V. H. (2016). CJEU addresses private copying and fair compensation issues. *Journal of Intellectual Property Law & Practice*, 11(6), 403–406. <https://doi.org/10.1093/jiplp/jpw035>
 34. Gurko, A. V. (2017). *Iskusstvennyy intellekt i avtorskoye pravo: Vzgl'yad v budushcheye* [Artificial intelligence and copyright: A look into the future]. *Intellektual'naya Sobstvennost'. Avtorskoye Pravo i Smezhnyye Prava*, 12, 7–18.
 35. Gurko, A. V. (2024). O vozmozhnosti avtorsko-pravovoy okhrany ob'yektov, generiruyemykh sistemami iskusstvennogo intellekta [On a possibility of copyright protection of objects generated by artificial intelligence systems]. *Journal of Russian Law*, 28(7), 64–77. <https://doi.org/10.61205/S160565900027569-9>
 36. Hadi, M. U., Tashi, Q. A., Qureshi, R., Shah, A., Muneer, A., Irfan, M., Zafar, A., Shaikh, M. B., Akhtar, N., Hassan, S. Z., Shoman, M., Wu, J., Mirjalili, S., & Shah, M. (2025). *Large language models: A comprehensive survey of its applications, challenges, limitations, and future prospects*. TechRxiv. <https://doi.org/10.36227/techrxiv.23589741.v8>

37. Horwitz, M. (2004). Conceptualizing the right to access to technology. *Washington Law Review*, 79(1), 105–118.
38. Hugenholtz, P. B., & Okediji, R. (2012). *Conceiving an International instrument on limitations and exceptions to copyright* (SSRN Scholarly Paper No. 2017629). Social Science Research Network. <https://doi.org/10.2139/ssrn.2017629>
39. Izacard, G., & Grave, E. (2021). Leveraging passage retrieval with generative models for open domain question answering. *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 874–880. <https://doi.org/10.18653/v1/2021.eacl-main.74>
40. Kalyatin, V. O. (2023). *Pravo intellektual'noy sobstvennosti. Pravovoye regulirovaniye baz dannykh* [Intellectual property law. Legal regulation of databases]. Yurait.
41. Kalyatin, V. O. (2024). Gotovo li patentnoye zakonodatel'stvo k ispol'zovaniyu iskusstvennogo intellekta? [Is patent law ready for artificial intelligence?]. *Intellektual'naya sobstvennost'. Promyshlennaya sobstvennost'*, 3, 58–62.
42. Kartschiya, A. A. (2024). Pravovaya okhrana dostizheniy iskusstvennogo intellekta [Legal protection of artificial intelligence achievements]. *Intellektual'naya Sobstvennost'. Avtorskoye Pravo i Smezhnyye Prava*, 4, 4–16.
43. Kashanin, A. V. (2007a). Problema minimal'nykh standartov okhranosposobnosti proizvedeniy v avtorskom prave [The problem of minimum standards of copyrightability of works]. *Vestnik Grazhdanskogo Prava*, (4), 23–62.
44. Kashanin, A. V. (2007b). Tvorcheskoy kharakter kak usloviye okhranosposobnosti proizvedeniya v rossiyskom i inostrannom avtorskom prave [Creative character as a condition of protectability of a work in Russian and foreign copyright law]. *Vestnik Grazhdanskogo Prava*, (2), 75–119.
45. Kashanin, A. V. (2009). Minimal'nyy uroven' tvorcheskogo kharaktera proizvedeniy v avtorskom prave Germanii [Minimum level of creative character of works in German copyright law]. *Zakonodatel'stvo i Ekonomika*, (12). <https://publications.hse.ru/articles/publications.hse.ru/articles/74677719>
46. Kashanin, A. V. (2010a). Nepovtorimost' proizvedeniya kak kriteriy ego sposobnosti k avtorsko-pravovoy okhrane [The uniqueness of a work as a criterion for its suitability for copyright protection]. *Vestnik Arbitrazhnogo Suda Goroda Moskvy*, (6), 32–41.
47. Kashanin, A. V. (2010b). Razvitiye ucheniya o forme i soderzhanii proizvedeniya v doktrine avtorskogo prava. Problema okhranosposobnosti soderzhaniya nauchnykh proizvedeniy [Development of the doctrine of form and content of a work in copyright law: The problem of protectability of the content of scientific works]. *Vestnik Grazhdanskogo Prava*, (2), 68–138.
48. Kashanin, A. V. (2010c). Obespecheniye lichnogo neimushchestvennogo interesa avtora proizvedeniya v Velikobritanii i SSHA [Ensuring the personal non-property interest of the author of a work in the UK and the USA]. *Journal of Foreign Legislation and Comparative Law*, (2), 23–39.
49. Kashanin, A. V. (2011). Razvitiye mekhanizma demarkatsii v avtorskom prave kontinental'noy Evropy [Development of the demarcation mechanism in the copyright law of continental Europe]. *Vestnik Grazhdanskogo Prava*, (3), 61–101.
50. Kashanin, A. V. (2012). Uroven' trebovaniy k tvorcheskomy kharakteru proizvedeniya v rossiyskom yuridicheskom disкурse [The level of requirements for the creative nature of a work in Russian legal discourse]. *Zakony Rossii: Opyt, Analiz, Praktika*, (9), 92–102.
51. Kharitonova, Yu. S. (2018). Pravovoy rezhim rezul'tatov deyatel'nosti iskusstvennogo intellekta [Legal regime of the results of artificial intelligence activities]. In E. B. Lauts (Ed.), *Sovremennyye informatsionnyye tekhnologii i pravo* (pp. 68–83). Statut.
52. Khoshnoodi, M., Jain, V., Gao, M., Srikanth, M., & Chadha, A. (2024). *A comprehensive survey of accelerated generation techniques in large language models* (Version 2). arXiv. <https://doi.org/10.48550/ARXIV.2405.13019>
53. Kirsanova, E. E. (2023). Obzor osnovnykh teoriy opredeleniya pravovogo rezhima ob'yektov, sozdannykh iskusstvennym intellektom [Review of the main theories of determining the legal regime of objects created by artificial intelligence]. *Zakon*, 9, 36–46.

54. Kollár, P. (2021). *Mind if I mine? A study on the justification and sufficiency of text and data mining exceptions in the european union* [Master Dissertation, Maastricht University]. <https://doi.org/10.2139/ssrn.3960570>
55. Komariah, S. (2024). Assessment of AI's creativity in the literary text: In comparison with AI's creative compositions on visual art. In N. Haristiani, Y. Yulianeta, Y. Wirza, W. Gunawan, A. A. Danuwijaya, E. Kurniawan, S. Suharno, N. Nafisah, & E. D. A. Imperiani (Eds.), *Proceedings of the 7th International Conference on Language, Literature, Culture, and Education (ICOLLITE 2023)* (pp. 430–438). Atlantis Press International BV. https://doi.org/10.2991/978-94-6463-376-4_57
56. Kozlova, A. A. (2024). Prava na intellektual'nyuyu sobstvennost', sozdannuyu s primeneniym iskusstvennogo intellekta [Rights to Intellectual property created using artificial intelligence]. *Intellektual'naya Sobstvennost'. Avtorskoye Pravo i Smezhnyye Prava*, 2, 33–40.
57. Lee, K., Cooper, A. F., & Grimmelmann, J. (2024). *Talkin' 'bout AI generation: Copyright and the generative-AI supply chain* (No. arXiv:2309.08133). arXiv. <https://doi.org/10.48550/arXiv.2309.08133>
58. Lemley, M. A. (1997). Romantic authorship and the rhetoric of property. *Texas Law Review*, 75, 873–923. <https://doi.org/10.2139/ssrn.44418>
59. Lemley, M. A. (2017). The fruit of the poisonous tree in IP law. *Iowa Law Review*, 103, 245–269.
60. Lemley, M. A., & Casey, B. (2021). Fair learning. *Texas Law Review*, 99(4), 743–785.
61. Leval, P. N. (1990). Toward a fair use standard. *Harvard Law Review*, 103(5), 1105–1136. <https://doi.org/10.2307/1341457>
62. Leval, P. N. (2015). *Campbell as fair use blueprint?* *Washington Law Review*, 90(2), 597–614.
63. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). *Retrieval-augmented generation for knowledge-intensive NLP tasks* (Version 4). arXiv. <https://doi.org/10.48550/ARXIV.2005.11401>
64. Li, M. (2024). Comprehensive review of backpropagation neural networks. *Academic Journal of Science and Technology*, 9(1), 150–154. <https://doi.org/10.54097/51y16r47>
65. Li, Z., Chen, C., Xu, T., Qin, Z., Xiao, J., Luo, Z.-Q., & Sun, R. (2025). Preserving diversity in supervised fine-tuning of large language models [Conference poster]. In *Proceedings of the 13th International Conference on Learning Representations (ICLR 2025)*. <https://openreview.net/forum?id=NQEe7B7bSw>
66. Lucchi, N. (2025, July). *Generative AI and copyright: Training, creation, regulation* (Study No. PE 774.095). Directorate-General for Citizens' Rights, Justice and Institutional Affairs, European Parliament. <https://doi.org/10.2861/0365517>
67. Margoni, T. (2025). *TDM and generative AI: Lawful access and opt-outs*. Social Science Research Network. <https://doi.org/10.2139/ssrn.5036164>
68. Margoni, T., & Kretschmer, M. (2022). A deeper look into the EU text and data mining exceptions: Harmonisation, data ownership, and the future of technology. *GRUR International*, 71(8), 685–701. <https://doi.org/10.1093/grurint/ikac054>
69. Matulionyte, R. (2025). *Reconceptualising the reproduction right in the age of AI*. Social Science Research Network. <https://doi.org/10.2139/ssrn.5041741>
70. Mezei, P. (2025). *The multi-layered regulation of rights reservation (opt-out) under EU copyright law and the AI Act -for the benefit of whom?* (Version 3.0). Social Science Research Network. <https://doi.org/10.2139/ssrn.5064018>
71. Morkhat, P. M. (2018). Yunit iskusstvennogo intellekta v kontekste prava intellektual'noy sobstvennosti: Avtor, soavtor, nayomnyy rabotnik ili instrument [Artificial intelligence unit in the context of intellectual property law: Author, co-author, employee or tool]. *Intellektual'naya Sobstvennost'. Avtorskoye Pravo i Smezhnyye Prava*, 8, 35–42.
72. Mossoff, A. (2007). Who cares what Thomas Jefferson thought about patents? Reevaluating the patent privilege in historical context. *Cornell Law Review*, 92(5), 953–1012.

73. Mukherjee, A. (2025). *The AI ouroboros and copyright laundering: Why copyright needs a “fruit of the poisonous tree” doctrine for generative AI* (SSRN Scholarly Paper No. 5256625). Social Science Research Network. <https://doi.org/10.2139/ssrn.5256625>
74. Murray, M. D. (2025). *AI pirated my art and birthed infringing works, and other metaphors that confound copyright law*. Social Science Research Network. <https://doi.org/10.2139/ssrn.5116714>
75. Nikiforov, A. A. (2020). Personazh kak ob'yekt avtorskogo prava [Character as an object of copyright]. In *Intellektual'nyye prava: Sbornik rabot vypusnikov Rossiyskoy shkoly chastnogo prava, posvyashchennyy 90-letiyu so dnya rozhdeniya Viktora Abramovicha Dozortseva* (pp. 154–219). Statut.
76. Orlova, T. E. (2022). Avtorskoye pravo na rezul'taty deyatel'nosti iskusstvennogo intellekta [Copyright on the results of artificial intelligence activities]. *Intellektual'naya Sobstvennost'. Avtorskoye Pravo i Smezhnyye Prava*, (11), 62–71.
77. Pavlova, E. A. (Ed.). (2018). *Kommentariy k chasti chetvertoy Grazhdanskogo kodeksa Rossiyskoy Federatsii* [Commentary on Part Four of the Civil Code of the Russian Federation]. Issledovatel'skiy Tsentr Chastnogo Prava.
78. Peukert, A. (2021). *A critique of the ontology of intellectual property law* (G. Mertens, Trans.). Cambridge University Press. <https://doi.org/10.1017/9781108653329>
79. Poort, J., & Quintais, J. P. (2013). The levy runs dry: A legal and economic analysis of EU private copying levies. *Journal of Intellectual Property, Information Technology and E-Commerce Law*, 4(3), 205–224. <https://nbn-resolving.de/urn:nbn:de:0009-29-38466>
80. Prasad, K., & Padilla, J. (2025). *Generative AI models at the gate: Licensing frameworks for the effective and efficient protection of copyright protected content in an AI world* (SSRN Scholarly Paper No. 5263547). Social Science Research Network. <https://doi.org/10.2139/ssrn.5263547>
81. Quintais, J. P. (2025). *The concept of “research organisation” and its implications for text and data mining and AI research*. Social Science Research Network. <https://doi.org/10.2139/ssrn.5155685>
82. Rakhmatulina, R. S. (2025). Kontsept «ischezayushchego avtora» i kontseptsiya otsutstvuyushchego avtora rezul'tatov, sozdannykh tekhnologiyami iskusstvennogo intellekta [The concept of the “disappearing author” and the concept of the absent author of results generated by artificial intelligence technologies]. *Grazhdanskoye Pravo*, (1), 14–17. <https://doi.org/10.18572/2070-2140-2025-1-14-17>
83. Rebelo, A. D. P., Inês, G. D. O., & Damion, D. E. V. (2022). The impact of artificial intelligence on the creativity of videos. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 18(1), 1–27. <https://doi.org/10.1145/3462634>
84. Riccio, G. M. (2024). AI, data mining and copyright law: Remarks about lawfulness and efficient choices. 2024 47th MIPRO ICT and Electronics Convention (MIPRO), 1457–1463. <https://doi.org/10.1109/MIPRO60963.2024.10569189>
85. Ricketson, S. (1987). *The Berne convention for the protection of literary and artistic works: 1886-1986*. Centre for Commercial Law Studies, Queen Mary College.
86. Rindfleisch, A. (2020). Transaction cost theory: Past, present and future. *AMS Review*, 10(1–2), 85–97. <https://doi.org/10.1007/s13162-019-00151-x>
87. Rolinson, P., Ariyevich, E. A., & Ermolina, D. E. (2018). Ob'yekty intellektual'noy sobstvennosti, sozdavayemye s pomoshch'yu iskusstvennogo intellekta: Osobennosti pravovogo rezhima v Rossii i za rubezhom [Intellectual property objects created with the help of artificial intelligence: Features of the legal regime in Russia and abroad]. *Zakon*, 5, 63–71.
88. Rosati, E. (2019). Copyright as an obstacle or an enabler? A European perspective on text and data mining and its role in the development of AI creativity. *Asia Pacific Law Review*, 27(2), 198–217. <https://doi.org/10.1080/10192557.2019.1705525>
89. Roznina, A. M. (2021). Avtorskiye prava na ob'yekty, sozdannyye iskusstvennym intellektom [Copyrights for objects created by artificial intelligence]. In E. A. Morgunova & P. V. Savishchev (Eds.), *Sbornik luchshikh*

- dokladov po grazhdanskomu pravu i grazhdanskomu protsessu uchastnikov studencheskikh nauchnykh meropriyatiy i chlenov studencheskikh nauchnykh klubov MGYUA: Materialy v ramkakh I Mezhdunarodnogo kongressa po tsivilisticheskoy komparativistike (pp. 101–104). RG-Press.
90. Safin, R. R., Maskin, K. A., & Povarov, Yu. S. (2018). Pravovoye regulirovaniye ob'yektov avtorskogo prava, sozdannykh s ispol'zovaniyem "neyroseti" [Legal regulation of copyright objects created using a "neural network"]. In O. V. Sushkova (Ed.), *Pravovoye regulirovaniye intellektual'noy sobstvennosti i innovatsionnoy deyatel'nosti: Sbornik statey uchastnikov nauchno-metodologicheskogo seminara* (pp. 154–158). RG-Press.
91. Sag, M. (2009). Copyright and copy-reliant technology. *Northwestern University Law Review*, 103(4), 1607–1682.
92. Sag, M. (2019). The new legal landscape for text mining and machine learning. *Journal of the Copyright Society of the U.S.A.*, 66, 291–367.
93. Sag, M. (2023). Copyright safety for generative AI. *Houston Law Review*, 61(2), 295–347.
94. Sag, M., Samuelson, P., & Sprigman, C. J. (2024). *Comments in response to the Copyright Office's notice of inquiry on artificial intelligence and copyright*. Social Science Research Network. <https://doi.org/10.2139/ssrn.4976391>
95. Samuelson, P. (1989). Information as property: Do Ruckelshaus and Carpenter signal a changing direction in intellectual property law? *Catholic University Law Review*, 38(2), 365–400.
96. Schuhmann, C., Beaumont, R., Vencu, R., Gordon, C., Wightman, R., Cherti, M., Coombes, T., Katta, A., Mullis, C., Wortsman, M., Schramowski, P., Kundurthy, S., Crowson, K., Schmidt, L., Kaczmarczyk, R., & Jitsev, J. (2022). *LAION-5B: An open large-scale dataset for training next generation image-text models* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2210.08402>
97. Schwartz-croft, L. (2024, February 15). *Effects of ROSS Intelligence and NDAs, highlighting the need for AI regulation*. Social Science Research Network. <https://ssrn.com/abstract=4727662>
98. Senftleben, M. (2023). Generative AI and author remuneration. *IIC — International Review of Intellectual Property and Competition Law*, 54(10), 1535–1560. <https://doi.org/10.1007/s40319-023-01399-4>
99. Senftleben, M. (2025). Win-win: How to remove copyright obstacles to AI training while ensuring author remuneration (and why the AI act fails to do the magic). *Chicago-Kent Law Review*, 100(1), 7.
100. Senftleben, M., Poort, J., van Eechoud, M. M. M., van Gompel, S., & Helberger, N. (Eds.). (2021). *Intellectual property and sports: Essays in honour of P. Bernt Hugenholtz*. Wolters Kluwer Law International.
101. Senftleben, M., Szkalej, K., Sganga, C., & Margoni, T. (2025). Towards a European Research Freedom Act: A reform agenda for research exceptions in the EU copyright acquis. *IIC — International Review of Intellectual Property and Competition Law*. <https://doi.org/10.1007/s40319-025-01604-6>
102. Sharifani, K., & Amini, M. (2023). Machine Learning and Deep Learning: A Review of Methods and Applications. *World Information Technology and Engineering Journal*, 10(7), 3897–3904.
103. Shcherbak, N. V. (2023). Mnozhestvennost' doktrin okhranosposobnosti proizvedeniya kak variant resheniya problemy pravovoy okhrany proizvedeniya po rossiyskomu pravu [Doctrines of protectability as a solution to the problem of the protectability of a work under russian law]. *Vestnik Grazhdanskogo Prava*, (5), 79–101. <https://doi.org/10.24031/1992-2043-2023-23-5-79-101>
104. Shpakovskaya, V. V. (2023). Iskusstvennyy intellekt v grazhdanskom prave [Artificial intelligence in civil law]. *Intellektual'naya Sobstvennost'. Avtorskoye Pravo i Smezhnyye Prava*, 4, 54–60.
105. Stallman, R. (2002). Misinterpreting Copyright—A Series of Errors. In J. Gay (Ed.), *Free software, free society: Selected essays of Richard Stallman* (1st ed, pp. 77–86). Free Software Foundation. <https://www.gnu.org/philosophy/misinterpreting-copyright.html>
106. Stratton, M. (2025). Market-based licensing for publishers' works is feasible. big tech agrees. *The Columbia Journal of Law & The Arts*, 48(4), 434–449. <https://doi.org/10.52214/jla.v48i4.13925>

107. Tyagi, K. (2024). Copyright, text & data mining and the innovation dimension of generative AI. *Journal of Intellectual Property Law & Practice*, 19(7), 557–570. <https://doi.org/10.1093/jiplp/jpae028>
108. Ueno, T. (2021a). A general clause on copyright limitations in civil law countries: Recent discussion on japanese-style fair use clause. In H. Sun, N.-L. Wee Loon, & S. Balganes (Eds.), *The Cambridge Handbook of Copyright Limitations and Exceptions* (pp. 211–215). Cambridge University Press. <https://doi.org/10.1017/9781108671101.016>
109. Ueno, T. (2021b). The flexible copyright exception for ‘non-enjoyment’ purposes – recent amendment in japan and its implication. *GRUR International*, 70(2), 145–152. <https://doi.org/10.1093/grurint/ikaa184>
110. Verma, S., Rassin, R., Das, A., Bhatt, G., Seshadri, P., Shah, C., Bilmes, J., Hajishirzi, H., & Elazar, Y. (2024). *How many Van Goghs does it take to Van Gogh? Finding the imitation threshold* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2410.15002>
111. Vitko, V. S. (2019a). Analiz nauchnykh predstavleniy ob avtore i pravakh na rezul'taty deyatel'nosti iskusstvennogo intellekta (Chast' 1) [Analysis of scientific views of authorship and right for results of AI activity (Part 2)]. *Intellektual'naya Sobstvennost'. Avtorskoye Pravo i Smezhnyye Prava*, 2, 5–20.
112. Vitko, V. S. (2019b). Analiz nauchnykh predstavleniy ob avtore i pravakh na rezul'taty deyatel'nosti iskusstvennogo intellekta (Chast' 2) [Analysis of scientific views of authorship and right for results of AI activity (Part 2)]. *Intellektual'naya Sobstvennost'. Avtorskoye Pravo i Smezhnyye Prava*, 3, 5–22.
113. Vorozhevich, A. S. (2025). Kartiny, sgenerirovannyye neyroset'yu: Ob'yekty avtorskikh prav ili net? [Pictures generated by a neural network: Objects of copyright or not?]. *Intellektual'naya Sobstvennost'. Avtorskoye Pravo i Smezhnyye Prava*, 1, 24–33.
114. Zhang, J. (2019). *Gradient descent based optimization algorithms for deep learning models training* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.1903.03614>
115. Zhang, S. (2024). Gradient descent algorithm optimization and its application in linear regression model. *Academic Journal of Natural Science*, 1(1), 1–5. <https://doi.org/10.5281/ZENODO.13753916>
116. Zhang, S., Dong, L., Li, X., Zhang, S., Sun, X., Wang, S., Li, J., Hu, R., Zhang, T., Wu, F., & Wang, G. (2023). *Instruction tuning for large language models: A survey* (Version 9). arXiv. <https://doi.org/10.48550/ARXIV.2308.10792>
117. Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., ... Wen, J.-R. (2023). *A survey of large language models* (Version 16). arXiv. <https://doi.org/10.48550/ARXIV.2303.18223>

Сведения об авторе:

Никифоров А. А. — магистр права (Российская школа частного права), преподаватель юридического факультета, Московская высшая школа социальных и экономических наук, ведущий юрисконсульт группы правового сопровождения сделок с программным обеспечением, технологиями, брендом и данными, Яндекс, Москва, Россия.

Information about the author:

Artem A. Nikiforov — LL.M. (Russian School of Private Law), Lecturer, Moscow School of Social and Economic Sciences (MSSES), Senior Legal Counsel, Software, Technology, Brand, and Data Transactions Legal Support Group, Yandex, Moscow, Russia.