R

WIPO
ВСЕМИРНАЯ ОРГАНИЗАЦИЯ
ИНТЕЛЛЕКТУАЛЬНОЙ
СОБСТВЕННОСТИ

SCCR/49/6
ОРИГИНАЛ: АНГЛИЙСКИЙ
ДАТА: 23 ОКТЯБРЯ 2026 ГОДА

Постоянный комитет по авторскому праву и смежным правам

Сорок девятая сессия

Женева, 30 ноября — 4 декабря 2026 года

КОМПЛЕКСНОЕ ИССЛЕДОВАНИЕ ВОПРОСОВ ОБУЧЕНИЯ МОДЕЛЕЙ ИИ И АВТОРСКОГО ПРАВА

подготовлено Даниэлем Дж. Жерве, Чой Сун Джаем, Мелиссой Омино, Родриго Копке Салинасом, Цянь Ваном

Комплексное исследование вопросов обучения моделей ИИ и авторского права

ОГОВОРКА ОБ ОГРАНИЧЕНИИ ОТВЕТСТВЕННОСТИ:

Настоящее исследование подготовлено по заказу Секретариата Всемирной организации интеллектуальной собственности (ВОИС). Мнения, выраженные в настоящем документе, принадлежат авторам и необязательно отражают позицию государств-членов или Секретариата.

БЛАГОДАРНОСТИ

ДАНИЭЛЬ ДЖ. ЖЕРВЕ — ведущий соавтор настоящего исследования. Почетный профессор права имени Милтона Р. Андервуда в Vanderbilt Law School (Школа права Вандербильта), Даниэль Дж. Жерве занимает должность ведущего доцента в Свободном университете Амстердама и входит в состав Американского института права в качестве избранного члена. С его исследованиями можно ознакомиться на веб-сайтах www.collectivemanagement.net, www.tripsagreement.com и www.thehumancause,.org.

ЧОЙ СУН ДЖАЙ — профессор права в Sejong University (Университет Седжон). Получил степень магистра права (LL.M.) в Колумбийской школе права. В сферу его научных интересов входят вопросы права интеллектуальной собственности и конкурентного права. Г-н Чой Сун Джай — соавтор книги «Авторское право» (Copyright Law, 2024 г.), лауреат Корейской арбитражной премии 2021 года в номинации «Лидер нового поколения».

МЕЛИССА ОМИНО возглавляет Центр права интеллектуальной собственности и информационных технологий (CIPIT) при Университете Стратмора в Найроби. Имеет докторскую степень по праву интеллектуальной собственности и торговому праву, специализируется на вопросах интеллектуальной собственности, искусственного интеллекта, защиты данных и цифровых прав в странах Глобального Юга.

РОДРИГУ КОПКЕ САЛИНАС — профессор в Школе права фонда Жетулиу Варгаса в Сан-Паулу (GVlaw), получил степень магистра права (LL.M.), в Колумбийском университете. Входит в состав Специального совета по авторскому праву Коллегии адвокатов Бразилии и специализируется на праве средств массовой информации и индустрии развлечений.

ЦЯНЬ ВАН— профессор права, заместитель председателя Академического комитета в Восточно-Китайском университете политологии и права. Он также занимает пост вице-председателя Китайской ассоциации авторского права и входил в состав экспертной комиссии по подготовке третьей редакции закона «Об авторском праве» КНР.

ПРИМЕЧАНИЕ: Языковые модели применялись для проверки синтаксиса, грамматики и содержания исследования, а также для внесения редакционных предложений с целью гармонизации настоящего текста, подготовленного в соавторстве.

СОДЕРЖАНИЕ

Введение 5
Часть I. Использование охраняемого контента при разработке моделей генеративного ИИ 7
Раздел A. Получение данных 7
1. Модели использования данных в зависимости от типа системы 7
2. Автоматизированный сбор данных с веб-страниц 8
3. Генерация с расширенным поиском 10
Раздел B. Использование данных при обучении моделей 11
1. Обучение модели 11
2. Оптимизация и обучение с подкреплением 13
Раздел C. По завершении обучения 14
1. Последующее использование обучающего корпуса 14
2. Сохранение копий в модели 14
3. Соотношение входных данных и выходного результата 18
Часть II. Разрешение, лицензирование и компенсация 20
Раздел A. Существующие варианты получения разработчиками охраняемых авторским правом объектов для обучения моделей 20
1. Пять каналов получения данных 20
2. Формы сотрудничества 23
Раздел B. Основные элементы государственной политики 23
1. Права, затрагиваемые обучением моделей генеративного ИИ 23
(a) Право на воспроизведение 23
(b) Право на переработку и производные произведения 24
(c) Сообщение для всеобщего сведения и доведение до всеобщего сведения 25
(d) Личные неимущественные права 25
(e) Информация об управлении правами и меры против обхода технических средств 26
2. Добровольное (коллективное и индивидуальное) лицензирование 27
(a) Индивидуальное и коллективное лицензирование использования произведений в системах ИИ 27
(b) Основные модели коллективного лицензирования (добровольное, расширенное и принудительное) 32
3. Лицензирование первичных языковых данных на основе интересов сообщества и ответственного управления данными 37
4. Предусмотренные законом исключения в отношении интеллектуального анализа текста и данных и общие нормы 39
5. Системы вознаграждения 44
(a) Разновидности систем вознаграждения 44

. (b) Формы и основа расчета вознаграждения 45
Часть III. Обеспечение соблюдения прав 47
Раздел A. Выводы из текущей судебной практики 48
1. Вопрос о признании модели копией 48
2. Исключение в отношении интеллектуального анализа текста и данных 51
3. Юрисдикция в отношении обучения моделей за пределами соответствующей территории 52
4. Вопрос о замещении произведений результатами работы модели 55
Раздел B. Обеспечение соблюдения прав в отношении разработчиков ИИ 59
1. Структурные особенности системы обеспечения соблюдения прав 59
2. Неравенство сторон 60
Раздел C. Проблемы обеспечения соблюдения прав 61
1. Транспарентность 61
2. Атаки на данные, фильтрация и распространение стандартов отказа от использования 63
3. Судебная практика в целом 67
Заключение 68

ВВЕДЕНИЕ

Настоящее исследование подготовлено по запросу Постоянного комитета по авторскому праву и смежным правам (ПКАП). На своей сорок седьмой сессии (Женева, 1–5 декабря 2025 г.) Постоянный комитет попросил Секретариат подготовить исследование о стратегических и регуляторных подходах к соотношению между обучением моделей ИИ и авторским правом, а также смежными нормами, и о применимой практике получения разрешения, обеспечения соблюдения прав и выплаты компенсации за использование. Этот запрос зафиксирован в отчете Председателя по итогам указанной сессии (документ SCCR/47/SUMMARY[1] от 5 декабря 2025 г., пункт 21).

Исследование носит описательный и сравнительный характер. В нем рассматриваются методы, принятые в ряде юрисдикций, и формирующаяся практика, а также анализируются предложенные варианты без формулирования конкретных рекомендаций для государств-членов и без выражения какой-либо позиции по спорным вопросам, которые им еще предстоит решить.

Исследование посвящено обучению моделей генеративного ИИ, то есть использованию контента для создания и совершенствования таких моделей, а также вопросам получения разрешения, обеспечения соблюдения прав и выплаты компенсации, возникающим в связи с таким использованием, если контент охраняется авторским правом или смежными правами. В более широком плане взаимодействие искусственного интеллекта и авторского права выходит далеко за рамки обучения моделей. В частности, оно охватывает правовой статус результатов работы развернутой модели как с точки зрения того, нарушает ли конкретный результат права на произведения, на которых обучалась модель, так и с точки зрения возможности и условий признания самого результата объектом, способным охраняться авторским правом. Вопросы, связанные с результатами работы моделей, затрагивают другие аспекты, рассмотрение которых Постоянный комитет перед данным исследованием не ставил. Они упоминаются лишь постольку, поскольку это требуется для формирования целостного представления о вопросе обучения моделей или для создания необходимого контекста.

Исследование состоит из трех частей. В части I на техническом и описательном уровне представлены операции, посредством которых охраняемый контент используется при разработке моделей генеративного ИИ, а также определены этапы, на которых такие операции способны затрагивать исключительные права, признаваемые национальным законодательством об авторском праве и международными договорами в области авторского права. В части II рассматриваются объем и пределы затрагиваемых прав, а также механизмы их реализации: добровольное лицензирование, индивидуальное или коллективное, и применимые исключения и ограничения. Часть III посвящена обеспечению соблюдения прав и анализу выводов, которые можно сделать на основе совокупности судебной практики, начинающей складываться в ряде юрисдикций. В ней также анализируются трансграничный характер обучения моделей и связанные с ним вопросы обеспечения соблюдения прав, включая прозрачность и стандарты выражения отказа от использования там, где это применимо. Приведенный в исследовании перечень юрисдикций служит иллюстративным целям и не может считаться исчерпывающим. В их число входят Бразилия, Канада, Китай, Европейский союз и многие его

[1] См. https://www.wipo.int/meetings/ru/doc_details.jsp?doc_id=653196

государства-члены, Кения, Республика Корея, Соединенное Королевство и Соединенные Штаты Америки.

ЧАСТЬ I. ИСПОЛЬЗОВАНИЕ ОХРАНЯЕМОГО КОНТЕНТА ПРИ РАЗРАБОТКЕ МОДЕЛЕЙ ГЕНЕРАТИВНОГО ИИ

Разработка модели генеративного искусственного интеллекта (ИИ) осуществляется в несколько последовательных этапов, и на каждом из них обычно используется сторонний контент. Часть контента может охраняться авторским правом и смежными правами; часть может находиться в общественном достоянии либо вследствие истечения срока охраны, либо в силу того, что правовая охрана ему изначально не предоставлялась. Технические операции, производимые с таким контентом, обычно одинаковы вне зависимости от его правового статуса. Одни и те же процессы, включая веб-краулинг, копирование, токенизацию и статистический анализ, описанные ниже, могут в равной мере применяться как к недавно вышедшему роману, так и к тексту XVIII века, перешедшему в общественное достояние. В настоящей части описываются эти операции и на описательном уровне показаны моменты, в которых они способны затрагивать права, признаваемые национальным законодательством об авторском праве и международными договорами по авторскому праву и смежным правам.

Последовательно представлены три фазы процесса обучения. В разделе А рассматривается получение данных, то есть процесс формирования разработчиком модели ИИ корпуса (зачастую в значительной мере состоящего из охраняемых объектов), на котором предстоит обучать модель. В разделе В освещается непосредственное использование этих данных в ходе обучения, включая основной цикл обучения и последующие разновидности тонкой настройки. В разделе С рассматривается ситуация, складывающаяся по завершении обучения, включая дальнейшие действия с обучающим корпусом и вопрос о том, содержит ли сама обученная модель копии произведений[2], на которых она обучалась.

РАЗДЕЛ А. ПОЛУЧЕНИЕ ДАННЫХ

Разработка большой языковой модели начинается с формирования корпуса, как правило, имеющего весьма значительный объем и разнородное происхождение. Разработчики получают данные несколькими способами. Некоторые данные могут лицензироваться, и в таком случае контент может предоставляться вместе с лицензией на его использование. При таком сценарии разработчик вправе приобрести доступ к курируемому набору данных, сформированному третьей стороной, либо заключить прямое лицензионное соглашение с издателем, новостной организацией или организацией коллективного управления (ОКУ). Если лицензия на использование охраняемых произведений не получена, использование может опираться на предположение о применимости исключения или ограничения в отношении соответствующего исключительного права. С точки зрения объема и порождаемых авторско-правовых проблем основным каналом сбора обучающих материалов служит автоматизированное извлечение материалов из открытой части Интернета. Для обучения могут также применяться синтетические данные, то есть данные, созданные существующей моделью ИИ, а не полученные от авторов-людей, особенно на более поздних стадиях разработки.

1. Модели использования данных в зависимости от типа системы

Потребность системы в данных обычно не ограничивается исходным циклом обучения: данные используются для переобучения и тонкой настройки, для интеллектуального анализа текста и данных на базе постоянно обновляемой

[2] Для упрощения изложения термин «произведения» следует понимать как включающий в себя объекты смежных прав там, где это уместно.

поисковой библиотеки, для генерации с расширенным поиском (RAG)[3], а также для создания превью и фрагментов описания (сниппетов). Эти способы использования различаются по требуемому объему материалов и по степени сложности отслеживания их источников. Тип системы обычно определяет характер потребления данных: системы генерации медиаконтента и чат-боты на основе больших языковых моделей потребляют данные преимущественно на этапе загрузки данных для обучения; специализированные корпоративные системы, надстраиваемые над языковой (называемой также базовой) моделью, дополняются предметно-ориентированным корпусом и зачастую поисковой библиотекой, к которой обращаются непосредственно в момент использования; а диалоговые поисковые системы требуют доступа к интернет-источникам в момент генерации. Положение системы в этом спектре влияет на набор исключительных прав, затрагиваемых использованием данных, а также на структуру любой соответствующей лицензии. Подробнее этот вопрос раскрывается в части II, раздел B.2.

2. Автоматизированный сбор данных с веб-страниц

Автоматизированный сбор данных с веб-страниц представляет собой распространенный подход, к которому разработчики базовых моделей прибегают при формировании своего обучающего корпуса. Впрочем, единой общепринятой терминологии в этой области не сложилось. В ряде источников разграничиваются «автоматизированный обход веб-сайтов» (веб-краулинг (*англ*. web crawling) — автоматический переход программы по сайтам по цепочкам ссылок со страницы на страницу) и «автоматизированный сбор данных с веб-страниц» (веб-скрейпинг (*англ*. web scraping) — выборочное извлечение определенных материалов со страницы, которая уже загружена). На практике эти процессы могут совмещаться: разработчик запускает программу-краулер, переходящую по ссылкам в сети, и на каждой веб-странице идентифицирует и извлекает текстовый, визуальный или иной контент, подлежащий сохранению. Результатом зачастую выступает копия такого контента, сохраняемая, как правило, в структурированном виде на собственной инфраструктуре разработчика.

В контексте настоящего анализа значение имеют четыре особенности. Во-первых, этот метод предполагает создание копий. Копии могут сохраняться на сервере или в кэше либо обрабатываться в оперативной памяти перед последующим преобразованием. Указанные действия способны затрагивать право на воспроизведение, закрепленное в статье 9(1) Бернской конвенции, в статье 1(4) Договора ВОИС по авторскому праву (ДАП) и в корреспондирующих положениях национального законодательства. В Согласованном заявлении в отношении статьи 1(4) ДАП подтверждено, что хранение охраняемого произведения в цифровой форме представляет собой воспроизведение и что право на воспроизведение и допускаемые исключения из него в полной мере применяются в цифровой среде[4]. Вопрос о том, меняют ли правовую оценку временный характер определенных копий или действие такого исключения, как исключение в отношении случайного и кратковременного

[3] Генерация с расширенным поиском (retrieval-augmented generation, обычно сокращенно RAG) более подробно рассматривается в разделе 3 настоящей части.

[4] Договор ВОИС по авторскому праву (20 декабря 1996 г.), Согласованное заявление в отношении статьи 1(4): «Право на воспроизведение, как оно определено в Статье 9 Бернской конвенции, и допускаемые этой статьей исключения полностью применяются в цифровой среде и, в частности, в отношении использования произведений в цифровой форме. Понимается, что хранение охраняемого произведения в цифровой форме в электронном средстве является воспроизведением в смысле Статьи 9 Бернской конвенции».

воспроизведения, рассматривается далее в настоящем исследовании. Статья 7 Договора ВОИС по исполнениям и фонограммам (ДИФ) и аналогичная статья Пекинского договора по аудиовизуальным исполнениям также содержат норму о праве разрешать прямое или косвенное воспроизведение «любым образом и в любой форме».

Во-вторых, извлечение материалов может носить невыборочный характер. Веб-краулер не всегда различает контент, перешедший в общественное достояние, материалы, распространяемые на условиях открытой лицензии, контент, размещенный вовсе без каких-либо лицензионных указаний, и охраняемые материалы, доступные на условиях пользовательских соглашений или защищенные техническими средствами защиты. Так, лишь некоторые краулеры соблюдают инструкции файла robots.txt, тогда как другие их игнорируют[5]. Подобные различия порой учитываются уже после совершения соответствующего действия, путем фильтрации или дедупликации корпуса, однако первичный акт копирования далеко не всегда носит избирательный характер.

В-третьих, правообладатели располагают ограниченными возможностями обнаружения подобных действий. Правообладатель может не знать, подвергалось ли его произведение автоматизированному сбору, каким именно краулером, в какое время и для каких последующих целей. Обязательства по раскрытию информации, начинающие появляться в законодательстве некоторых стран и, в частности, в Европейском союзе в рамках Регламента ЕС об искусственном интеллекте 2024 года, требуют от поставщиков моделей ИИ общего назначения публиковать «достаточно подробное резюме» о контенте, использованном для обучения[6]. Оценивать практические последствия этих обязательств пока преждевременно. Сам вопрос о том, как совместить раскрытие информации с соблюдением законных коммерческих интересов разработчиков, служит предметом продолжающихся дискуссий[7].

В-четвертых, создание и использование копий охраняемых произведений для обучения моделей может повлечь за собой удаление или изменение информации об управлении правами и обход эффективных технических средств, применяемых правообладателями в связи с осуществлением их прав по ДАП, ДИФ и Пекинскому договору.

---

[5] Протокол исключения поисковых роботов (Robots Exclusion Protocol), стандартизированный как IETF RFC 9309 (2022 г.), представляет собой договоренность, в соответствии с которой администратор сайта размещает в его корневом каталоге текстовый файл (robots.txt), указывающий, к каким разделам сайта разрешен доступ тем или иным автоматизированным краулерам. Соблюдение этого протокола носит добровольный характер; механизм принудительного обеспечения его выполнения отсутствует.

[6] Регламент (ЕС) 2024/1689 Европейского парламента и Совета от 13 июня 2024 г., устанавливающий гармонизированные правила в отношении искусственного интеллекта (Регламент ЕС об искусственном интеллекте), статья 53(1)(d) (обязанность поставщиков моделей ИИ общего назначения «составлять и делать общедоступным достаточно подробное резюме о контенте, использованном для обучения модели ИИ общего назначения, в соответствии с типовой формой, предоставляемой EU AI Office (Бюро ЕС по искусственному интеллекту)»).

[7] Тем не менее иногда происхождение собранных материалов проявляется в поведении итоговой модели: ранние версии известной модели генерации изображений воспроизводили узнаваемый водяной знак фотобанка, что свидетельствовало о несанкционированной загрузке значительного массива снимков этого агентства. Этот факт получил широкую огласку и фигурировал в судебном разбирательстве, рассмотренном в части III, раздел A.1 (дело Getty Images (US) Inc. v. Stability AI Ltd.). Это наглядно подтверждает, что хотя сбор конкретного произведения правообладателю обычно не виден, сама обученная модель способна выдаваемыми результатами обнаружить факт использования охраняемых материалов при обучении.

3. Генерация с расширенным поиском

Самостоятельным, но смежным техническим приемом выступает генерация с расширенным поиском (RAG)[8]. Этот метод состоит в том, что система ИИ формирует ответ на запрос путем предварительного обращения к отдельному, заранее сформированному массиву документов, не полагаясь исключительно на массив данных, на котором она обучалась. В качестве наглядной аналогии можно привести различие между человеком, отвечающим на вопрос по памяти, и тем же человеком, делающим паузу, чтобы заглянуть в справочник на полке около рабочего стола. При поступлении запроса система осуществляет поиск по этой коллекции, извлекает фрагменты, наиболее релевантные вопросу, сопоставляет их с запросом и использует их для формирования ответа. Коллекция создается заранее посредством создания и сохранения копий документов, а соответствующие фрагменты нередко повторно копируются в систему в момент генерации ответа; эти обстоятельства имеют существенное значение для авторско-правового анализа, изложенного ниже.

В RAG-системе поддерживается внешний корпус, который обычно индексируется в базе данных после преобразования каждого документа или его фрагмента в так называемое векторное представление[9]. При получении пользовательского запроса система формирует его векторное представление и на его основе извлекает наиболее релевантные фрагменты из корпуса. Затем система формирует запрос (промпт) для передачи в языковую модель, объединяя исходный запрос пользователя с извлеченными фрагментами, на основе которых и данных собственного обучения модель формирует итоговый ответ.

В рамках рассматриваемых целей иногда отмечают, что RAG меняет характер авторско-правовых вопросов, возникающих при обучении языковых моделей. Обосновывается это тем, что сама модель на таком корпусе не обучается, поскольку он хранится отдельно и опрашивается только на этапе логического вывода. Вместе с тем построение индексированного корпуса требует предварительной обработки и, при стандартной архитектуре RAG, хранения исходных документов либо их фрагментов[10]. Преобразование каждого документа в векторное представление порождает производное числовое представление документа, хотя, в отличие от сохраненной копии источника, восстановить исходный текст из векторного представления в общем

---

[8] См. сноску 2 выше.

[9] Векторные представления слов представляют каждый токен в виде плотного многомерного вектора вещественных чисел. Векторные представления именуются «контекстуальными», когда представление, присвоенное данному токену, зависит от окружающих его токенов. См. Heart of the Matter, выше, сноска 10, часть II.C.

[10] С более строгой технической точки зрения следует разграничивать векторные представления фрагментов текста, служащие для векторного поиска, и исходный текст, на основе которого они созданы. Векторное представление представляет собой плотное числовое представление, созданное кодировщиком; точной копией текста оно не выступает и обычным декодированием вектора, как правило, не восстанавливается. Тем не менее специальные методы инверсии векторных представлений способны восстанавливать исходные короткие фрагменты текста по их векторам частично или, в специфических экспериментальных условиях, полностью. В классической системе RAG исходные документы или их фрагменты сохраняются также и в текстовом виде — либо непосредственно в векторном хранилище, либо в отдельном хранилище документов, связанном с векторами. Векторные представления служат для поиска релевантных фрагментов, после чего именно связанный с ними текст, а не вектор представления, извлекается и передается языковой модели в составе входного контекста. См. Моррис Дж. Икс. и др. Векторные представления текста раскрывают (почти) столько же, сколько сам текст (Morris John X. et al. Text Embeddings Reveal (Almost) As Much As Text), Proc. 2023 Conf. on Empirical Methods in Natural Language Processing (EMNLP) (2023 г.). Отраслевое описание конвейера поиска приведено в публикации: https://infohub.delltechnologies.com/en-us/p/demystifying-on-device-intelligent-search-using-rag-architecture/

случае невозможно. В процессе работы (на этапе логического вывода) релевантные фрагменты копируются в контекстное окно модели и могут воспроизводиться в результатах ее работы полностью или в существенной части. Эти действия затрагивают право на воспроизведение, а также могут затрагивать право на переработку (статья 12 Бернской конвенции) в зависимости от применимого права и характеристик представления. Возможное применение указанных прав рассматривается ниже.

РАЗДЕЛ B. ИСПОЛЬЗОВАНИЕ ДАННЫХ ПРИ ОБУЧЕНИИ МОДЕЛЕЙ

Обучение модели генеративного ИИ охватывает последовательность действий, которую принято делить на четыре этапа: предварительное обучение, тонкая настройка с учителем, моделирование вознаграждения и обучение с подкреплением[11]. Первый из них обычно выступает наиболее ресурсоемким как по вычислительным затратам, так и по объему используемых данных; три последующих этапа в совокупности нередко называют «тонкой настройкой» или «дообучением после основного этапа обучения». Самостоятельное, но взаимосвязанное направление деятельности — оценка и оптимизация модели — сопровождает процесс обучения и продолжается на этапе развертывания модели.

С точки зрения авторского права ключевой вывод заключается в том, что каждый из этих этапов обычно связан с хранением и обработкой всего исходного корпуса или его части. Процесс обучения устроен иначе, чем краткое обращение к данным с их последующим удалением: данные сохраняются, загружаются, преобразуются и используются для обновления модели. Если такие данные содержат охраняемые произведения, указанные операции сопряжены с созданием копий или преобразованиями, имеющими значение для анализа воспроизведения, изложенного в части II.

1. Обучение модели

Основной цикл обучения, обычно именуемый предварительным обучением, представляет собой этап, на котором модель приобретает основной массив сведений о языке (в случае языковой модели) либо об изображениях, аудио- и видеоматериалах (в случае мультимодальной модели). Предварительное обучение включает несколько технических шагов, каждый из которых выполняется на сохраненной копии корпуса. Текстовый контент сначала подвергается токенизации, то есть преобразованию слов и иных последовательностей в поток числовых токенов, выбираемых из фиксированного словаря[12]. Современные токенизаторы обычно работают на уровне субсловных

---

[11] Использованное здесь четырехэтапное описание заимствовано из технической литературы и подробно рассматривается в статье «The Heart of the Matter: Copyright, AI Training, and LLMs» 71 J Cop Soc'y of the USA 482 (2024 г.) (далее — Heart of the Matter), размещенной по адресу https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4963711. Аналогичный четырехэтапный процесс (сбор данных, предварительная обработка данных, обучение модели, а также оценка и оптимизация модели) представлен в документе Ministry of Culture, Sports and Tourism (Министерство культуры, спорта и туризма Республики Корея) и Korea Copyright Commission (Корейская комиссия по авторскому праву) «Руководство по применимости доктрины добросовестного использования к обучению моделей генеративного ИИ» (Guide on Applicability of the Fair Use Doctrine to Training of Generative AI Models, февраль 2026 г.) (далее — Руководство Кореи по добросовестному использованию), с. 10.

[12] По вопросу токенизации см. Heart of the Matter, выше, сноска 10, часть II.B. В Руководстве Кореи по добросовестному использованию токенизация отнесена к техническим процедурам предварительной обработки данных наряду с очисткой и нормализацией. См. Руководство Кореи по добросовестному использованию, выше, сноска 10, с. 10.

единиц, а не целых слов[13]. Этот шаг имеет существенное значение для целей настоящего анализа, поскольку в случае байтовых токенизаторов, применяемых ведущими современными моделями, он протекает без потерь и обратим: текст, закодированный таким токенизатором, при последующем декодировании дает исходный текст, поэтому токенизированный корпус представляет собой не краткое изложение или абстракцию произведений, а их перекодирование в иную форму.

Затем токены представляются в виде векторных представлений[14]. Подход, занимающий доминирующее положение в этой сфере с 2017 года — архитектура трансформера (буква «T» в аббревиатуре GPT, означающей Generative Pre-trained Transformer), — обрабатывает такие представления с помощью механизма, известного как механизм внимания[15]. Квалификация векторных представлений и самой обученной модели, предполагающих более глубокую степень преобразования, рассматривается ниже в разделе C.2.

После инициализации модель обучается путем многократной подачи на нее последовательностей токенов, извлеченных из корпуса. Модель последовательно предсказывает каждый токен на основе предшествующих ему токенов и корректируется путем множества малых приращений, чтобы приблизить ее прогнозы к исходному тексту; таким образом, корпус опрашивается не единожды, а используется токен за токеном на протяжении всего цикла обучения. После обработки колоссального объема токенов, взятых из корпуса, модель усваивает масштабное распределение вероятностей по всему пространству возможных последовательностей токенов. Это распределение, закодированное в параметрах модели, и обеспечивает способность модели генерировать связный текст в ответ на поступивший запрос.

Соотношение между обучающими данными и обученной моделью служит предметом оживленных дискуссий. Согласно одной из точек зрения, поскольку обученная модель содержит лишь параметры, а не исходные обучающие данные, по завершении обучения никаких копий охраняемых произведений не остается[16]. Параметры описывают распределение вероятностей в чрезвычайно обширном векторном пространстве. Хотя такое распределение позволяет модели обобщать информацию и генерировать совершенно новые последовательности, оно способно приписывать непропорционально высокие, концентрированные вероятности конкретным точным последовательностям, неоднократно встречавшимся в процессе обучения. В силу того, что распределение имеет значительный перекос в сторону этих конечных обучающих цепочек, модель во многих случаях можно побудить к их дословному или почти дословному воспроизведению. Данное явление подробно описано в технической литературе и дополнительно рассматривается ниже в разделе C. Как разъясняется

---

[13] Словарь формируется начиная с отдельных символов путем последовательного объединения наиболее часто встречающихся соседних пар в более крупные единицы, вследствие чего распространенные слова начинают обозначаться отдельными токенами, а более редкие делятся на несколько. Байтовое парное кодирование (Byte-Pair Encoding) применяется компанией OpenAI для моделей Generative Pretrained Transformer (GPT), а также рядом других ведущих разработчиков. См. Heart of the Matter, выше, сноска 10, часть II.B.

[14] Точнее, они отображаются в многомерное векторное пространство, в котором каждый из них выражается в виде векторного представления.

[15] Механизм внимания позволяет модели при вычислении представления заданного токена оценивать значимость каждого другого токена в последовательности, благодаря чему она способна улавливать взаимосвязи на протяжении длинных фрагментов текста. Васвани А. и др. Внимание — это все, что вам нужно (Vaswani Ashish et al. Attention Is All You Need), 31 Conf. Neural Inf. Processing Syst. (NeurIPS) (2017 г.). Архитектура трансформера лежит в основе большинства современных больших языковых моделей.

[16] См. Heart of the Matter, выше, сноска 10, часть II.C.

ниже, между этим явлением (нередко обозначаемым как «запоминание обучающих данных моделью») и вопросом о наличии внутри модели «воспроизведения» всех использованных при обучении произведений существует принципиальное различие.

Среди моделей, генерирующих изображения, аудио или видео, наибольшее распространение получили так называемые «диффузионные модели». Они обучаются путем постепенного добавления шума к обучающим примерам с последующим освоением обратного процесса для синтеза новых объектов из шума[17]. Этот механизм отличается от принципа функционирования языковых (текстовых) моделей, однако в обоих случаях процесс обучения формирует сложное представление входных данных.

2. Оптимизация и обучение с подкреплением

Этапы, следующие за предварительным обучением, предназначены для адаптации предварительно обученной модели к конкретным задачам, конкретным сценариям использования и конкретным стандартам поведения. Терминология у разных разработчиков различается, однако эта последовательность охватывает тонкую настройку с учителем, моделирование вознаграждения и обучение с подкреплением[18].

Тонкая настройка с учителем предусматривает дальнейшее обучение на меньшем по объему, курируемом наборе данных, состоящем из пар «входные данные — результат» того типа, который ожидается получить от модели. Такой набор данных существенно уступает по масштабу корпусу предварительного обучения, однако отбирается по критерию качества и зачастую составляется или редактируется людьми. Каждый обучающий пример этапа тонкой настройки, подобно примерам предварительного обучения, загружается и используется для обновления параметров модели.

Моделирование вознаграждения предполагает создание отдельной модели вознаграждения, оценивающей варианты результатов работы основной модели по степени их предпочтительности. Сама модель вознаграждения обучается на сопоставлениях вариантов выдачи, составленных людьми.

Обучение с подкреплением на основе обратной связи от человека (RLHF), представляющее собой широко распространенный метод согласования моделей генеративного ИИ с предпочтениями человека, использует модель вознаграждения для направления последующих обновлений основной модели.

---

[17] Об истоках диффузионных моделей см. Сол-Дикштейн Дж. и др. Глубокое обучение без учителя с использованием неравновесной термодинамики (Sohl-Dickstein Jascha and others. Deep Unsupervised Learning using Nonequilibrium Thermodynamics), Proc. 32nd Int'l Conf. on Machine Learning (ICML) (2015 г.); Хо Дж., Джайн А., Аббил П. Шумоподавляющие диффузионные вероятностные модели (Ho Jonathan, Jain Ajay and Abbeel Pieter. Denoising Diffusion Probabilistic Models), 34 Conf. Neural Inf. Processing Syst. (NeurIPS) (2020 г.); а также, касательно архитектуры латентных диффузионных моделей, используемой в Stable Diffusion и фигурирующей в судебном разбирательстве по иску Getty, рассмотренном в части III, Ромбах Р. и др. Синтез изображений высокого разрешения с использованием латентных диффузионных моделей (Rombach Robin and others. High-Resolution Image Synthesis with Latent Diffusion Models), IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022 г.). В качестве доступного вводного материала см. также Сун Я., Эрмон С. Генеративное моделирование путем оценки градиентов распределения данных (Song Yang and Ermon Stefano. Generative Modeling by Estimating Gradients of the Data Distribution), 33 Conf. Neural Inf. Processing Syst. (NeurIPS) (2019 г.), доступно по адресу http://papers.neurips.cc/paper/9361-generative-modeling-by-estimating-gradients-of-the-data-distribution.pdf. Общая идея настоящего обсуждения не зависит от математических особенностей какого-либо конкретного класса моделей.

[18] См. Heart of the Matter, выше, сноска 10, часть II.D. Термины «предварительное обучение», «тонкая настройка с учителем», «моделирование вознаграждения» и «обучение с подкреплением» широко применяются разработчиками ИИ.

Каждый из этих этапов требует хранения и обработки обучающих материалов. Хотя их объемы скромнее требуемых для предварительного обучения, сам обучающий материал может включать охраняемые произведения, в особенности когда сопоставления, подготовленные разметчиками-людьми, опираются на такие произведения или содержат их. Акцент на качестве материалов, а не на их количестве не меняет правовой оценки с позиций авторского права.

Оценка и оптимизация модели, сопровождающие этапы обучения и продолжающиеся после развертывания, предполагают использование оценочных наборов данных для проверки работоспособности модели и выявления аспектов, требующих дальнейшего обучения. Оценочные наборы данных могут и сами содержать охраняемые произведения, в частности когда тестовые наборы создаются на основе общедоступных тестовых корпусов.

РАЗДЕЛ С. ПО ЗАВЕРШЕНИИ ОБУЧЕНИЯ

После завершения обучения возникают три самостоятельных вопроса. Первый касается того, сохраняется ли корпус, кем именно и для каких последующих целей. Второй состоит в том, содержит ли модель копии произведений, на которых она обучалась, и в каком именно смысле. Третий — воспроизводят ли результаты работы модели произведения, использованные при обучении, перерабатывают ли их или иным образом затрагивают права на них.

1. Последующее использование обучающего корпуса

Обучающий корпус сохраняет свое значение и по завершении цикла обучения. Разработчик может сохранять его для различных целей: для оценки будущих версий модели, обеспечения дальнейшей тонкой настройки на отдельных подмножествах, проведения текущих исследований или предоставления доступа третьим лицам по лицензии либо на иных основаниях. Он может объединяться со вновь собранными материалами для формирования расширенного корпуса моделей следующего поколения. Его можно продавать или передавать другим лицам (например, аффилированным организациям). В ряде случаев производные материалы на основе корпуса, включая отфильтрованные подмножества или подмножества после дедупликации, размещаются в открытом доступе на условиях разрешительных лицензий.

Каждый из этих способов последующего использования может повлечь за собой акты воспроизведения, распространения и доведения до всеобщего сведения по смыслу применимых национальных и международных норм. Иными словами, корпус, выполнивший задачу обучения, может продолжать использоваться на протяжении многих лет, и правила, регулирующие такое последующее использование, согласуются с нормами, определявшими правомерность его первоначального формирования.

2. Сохранение копий в модели

Вопрос о том, содержит ли сама обученная модель «копии» произведений, на которых она обучалась, относится к числу наиболее сложных с технической точки зрения аспектов, рассмотренных в настоящей части, и отличает генеративный ИИ от предшествующих технологий. Позиции, занимаемые разработчиками ИИ, с одной стороны, и правообладателями, с другой стороны, до недавнего времени носили диаметрально противоположный характер[19]. Разработчики ИИ обычно настаивали на

[19] Разногласия выходят за рамки процессуальных позиций в судебных разбирательствах и затрагивают приоритеты проектирования систем. Многоэтапное исследование методом Дельфи с участием

том, что обученная модель содержит лишь параметры, а не сами произведения, и что для целей авторского права никаких копий в ней не сохраняется. Правообладатели утверждали, что модель фактически представляет собой копию всего корпуса. Эмпирические данные, накопленные за последние два года, свидетельствуют о том, что в обеих позициях взаимосвязь между запоминанием данных и моделью представлена в чрезмерно упрощенном виде. Кроме того, каким бы ни оказался исход этого спора, он не затрагивает предшествующие действия: формирование корпуса и его использование при обучении сопряжены с актами воспроизведения, правовая квалификация которых не зависит от того, содержит ли итоговая модель копии.

В ходе анализа можно опереться на две группы технических доказательств. Запоминание обучающих данных моделью представляет собой свойство обученной модели, позволяющее восстанавливать конкретные обучающие последовательности по параметрам модели с достаточной степенью точности. Поскольку непосредственное изучение внутренней структуры модели сопряжено со сложностями, запоминание исследуется путем извлечения запомненных данных, то есть подтвержденной способности модели при формировании подходящих запросов воспроизводить последовательности из обучающих данных дословно или почти дословно. Извлечение запомненных данных служит в таком случае доказательством факта запоминания. Безрезультатность конкретной процедуры извлечения еще не доказывает, что произведение не было запомнено моделью. В рамках комплексного исследования Купер и соавторы применили метод вероятностного извлечения к 200 книгам в четырнадцати языковых моделях с открытыми весами и установили, что степень запоминания существенно различается в зависимости как от модели, так и от произведения[20]. Большинство наиболее крупных моделей не запомнили большинство книг ни полностью, ни частично. Однако там, где запоминание имело место, его масштабы могли оказаться значительными: в исследовании отмечено, что одна из широко распространенных моделей запомнила первый роман о Гарри Поттере и произведение Джорджа Оруэлла «1984» практически целиком; в отношении первого

специалистов по этике ИИ, экспертов в области права, разработчиков музыкального ИИ и деятелей искусства, проведенное для проверки концепции ответственного проектирования систем генеративной музыки, выявило согласованность защитных («протекционистских») позиций авторов и юристов в вопросах суверенитета данных, контрастирующую с ориентацией разработчиков-практиков на техническую осуществимость; в исследовании сделан вывод, что подобные концептуальные структуры наиболее эффективно служат связующим звеном между расходящимися приоритетами заинтересованных сторон (включая поиск компромисса между прозрачностью модели и интересами интеллектуальной собственности), а не инструментами принудительного консенсуса. Де Берардинис Дж. и др. RAIM: Operationalising Trustworthy AI in Music Generation (de Berardinis J. and others. RAIM: Operationalising Trustworthy AI in Music Generation), IEEE Transactions on Technology and Society, Special Issue on Ethical Innovation with/in Music Technology (готовится к публикации, 2026 г.), сайт проекта: https://casbahlab.github.io/raim-framework/, дата обращения: 25 августа 2026 г.

[20] Федер Купер А., Гокаслан А., Ахмед А., Сайферт Э. Б., Де Са К., Лемли М. А., Хо Д. Э., Лян П. Извлечение запомненных фрагментов (охраняемых авторским правом) книг из языковых моделей с открытыми весами (Feder Cooper A., Gokaslan Aaron, Ahmed Ahmed, Cyphert Amy B., De Sa Christopher, Lemley Mark A., Ho Daniel E. and Liang Percy. Extracting Memorized Pieces of (Copyrighted) Books from Open-Weight Language Models), arXiv:2505.12546 (2025 г.), версия от 30 ноября 2025 г. (опубликовано также как Stanford Public Law Working Paper / WVU College of Law Research Paper No. 2025-005). Формулировку данного вопроса авторами (на с. 1) целесообразно привести полностью: «Истцы и ответчики по судебным спорам об авторском праве в сфере генеративного ИИ нередко выдвигают масштабные взаимоисключающие утверждения относительно того, в какой мере большие языковые модели (LLM) запомнили охраняемую форму выражения истцов. Опираясь как на теорию машинного обучения, так и на нормы авторского права, мы показываем, что столь полярные позиции крайне упрощают связь между запоминанием данных и авторским правом». По тезису о том, что запомненный контент неизбежно закодирован в параметрах модели, поскольку «ему больше просто негде находиться», см. также Федер Купер А. и Гриммельманн Дж. (Cooper A. Feder and Grimmelmann James, цитируется там же на с. 25).

романа авторы доказали возможность почти дословной генерации практически всей книги.

Впрочем, эти эмпирические подтверждения запоминания вызывают возражения. Разработчики ИИ возражают, указывая, что методы составления запросов, необходимые для вызова дословного воспроизведения, представляют собой состязательное формирование запросов, не отражают реальную практику применения систем и могут образовывать манипуляцию или неправомерное использование продукта, выходящее за рамки обычного использования (а в отдельных случаях — за рамки условий обслуживания). В рамках судебного разбирательства по иску The New York Times против OpenAI ответчик подал ходатайство об отклонении иска в части требований о прямом нарушении авторского права на том основании, что истец применил усложненные манипулятивные запросы для искусственного провоцирования дословного воспроизведения обучающих данных, ставшего предметом жалобы[21]. В апреле 2025 года Окружной суд Южного округа Нью-Йорка отказался отклонить требования по этому основанию, и в ходе стадии раскрытия доказательств предмет спора сместился с вопроса о возможности искусственной фабрикации дословного воспроизведения на вопрос о частоте его естественного возникновения на практике, что завершилось постановлением о раскрытии доказательств в виде репрезентативной выборки из 20 млн журналов реального взаимодействия пользователей с системой[22]. Данное судебное разбирательство и значение вынесенного судебного постановления подробнее рассматриваются в части III, раздел A.4, настоящего исследования.

Таким образом, эмпирические результаты ставят под сомнение обе полярные позиции: как правило, модель не представляет собой «копию» своего обучающего корпуса, но и полностью «свободной» от него ее назвать нельзя. Точный объем таких данных различается в зависимости от модели.

Вторую группу доказательств образует обширная литература, посвященная атакам с целью извлечения обучающих данных, в которой зафиксированы условия, позволяющие восстанавливать отдельные обучающие примеры из обученной модели путем целенаправленного формирования запросов[23]. Среди извлеченных примеров встречаются фрагменты текста значительного объема, узнаваемые изображения и персональные данные. Эти исследования приводятся в качестве подтверждения того,

---

[21] Дело The New York Times Company v. Microsoft Corp., № 1:23-cv-11195, Memorandum in Support of Defendants' Motion to Dismiss (S.D.N.Y., 26 февраля 2024 г.); Opinion and Order (S.D.N.Y., 4 апреля 2025 г.) (судья Стайн) (отклонение в соответствующей части ходатайств об отклонении требований по основаниям прямого нарушения авторского права и содействия нарушению авторского права). Официальная публичная позиция OpenAI по данному вопросу изложена в материале: OpenAI, «Reporting the Facts About the New York Times' Lawsuit», размещенном по адресу openai.com/new-york-times.

[22] Дело In re OpenAI, Inc. Copyright Infringement Litigation, № 1:25-md-03143-SHS-OTW (S.D.N.Y.), Opinion and Order (судья Стайн, 5 января 2026 г.), подтверждающее Opinion and Order (мировой судья Ван, 2 декабря 2025 г.) (об отказе в удовлетворении ходатайства OpenAI о пересмотре решения) и предшествующее постановление о раскрытии доказательств от ноября 2025 г. Суд предписал предоставить рандомизированную обезличенную выборку из 20 млн журналов взаимодействия пользователей ChatGPT, сформированную из сохраненного массива данных, исходя из того, что общие данные о результатах работы модели, предоставляемых потребителям, имеют значение для анализа добросовестного использования и, в особенности, для ответа на вопрос о том, выступают ли результаты работы модели рыночным замещением произведений истцов.

[23] К числу более ранних работ в этом направлении относится публикация: Карамолегку А. и др. Нарушения авторских прав и большие языковые модели (Karamolegkou Antonia et al. Copyright Violations and Large Language Models), arXiv:2310.13771 (2023 г.) (демонстрирующая на примере ряда языковых моделей, что отдельные примеры обучающих данных можно восстановить посредством атак с целью извлечения обучающих данных).

что возможность извлечения может выступать структурным свойством крупных генеративных моделей, а не исключительно следствием манипуляций пользователя. Ряд компаний-разработчиков ИИ возражают, что подобные запросы образуют недобросовестную манипуляцию или нарушение условий обслуживания[24]. Впрочем, за последний год подобные оценки существенно усложнились: в развернутых системах между моделью и пользователем все чаще устанавливаются фильтры и иные защитные ограничения, снижающие возможность извлечения запомненного контента; в последних исследованиях состязательные методы формирования запросов требовались лишь для некоторых систем, тогда как другие выдавали охраняемый контент и без их применения[25].

Для лучшего понимания ситуации можно привести аналогию. Речь идет об аналогии с компьютерными программами[26]. Когда исходный код программы (нередко написанный на языке, напоминающем естественный) компилируется в исполняемую форму, полученный результат отличается от исходного кода и не поддается (легкому) прочтению человеком. Вместе с тем результат компиляции не свободен от оригинала: исполняемый файл сохраняет информацию исходной программы, и путем обратного проектирования оригинал может быть восстановлен в существенной степени. В этом отношении исполняемый код сохраняет сходство с исходной программой. Действия, позволяющие получить обученную модель из обучающего корпуса, если следовать этой аналогии, можно признать сопоставимыми. В обоих случаях понятный для восприятия источник (программа в виде исходного кода; корпус произведений) преобразуется в функциональный числовой объект, который внешне уже не похож на оригинал, но продолжает нести производную от него информацию. Тем не менее у этой аналогии есть пределы: компиляция представляет собой детерминированный процесс, полностью переносящий исходную программу в исполняемый файл, тогда как кодирование обучающих данных в модели носит избирательный и отчасти вероятностный характер, вследствие чего аналогия применима максимум к тому подмножеству произведений, запоминание которых доказано; в недавних научных публикациях утверждается, что именно вокруг этой разницы между детерминированной и чисто вероятностной возможностью восстановления, скорее всего, и будет строиться правовая квалификация модели[27].

---

[24] См. Карлини Н. и др. Извлечение обучающих данных из больших языковых моделей (Carlini Nicholas et al. Extracting Training Data from Large Language Models), 30th USENIX Security Symposium (2021 г.) 2633; Наср М. и др. Масштабируемое извлечение обучающих данных из промышленных языковых моделей, прошедших процедуру согласования (Nasr Milad et al. Scalable Extraction of Training Data from Aligned, Production Language Models), ICLR 2025 (расширенная версия публикации: Nasr et al., arXiv:2311.17035 (2023 г.)).

[25] Ахмед А. и др. Извлечение книг из промышленных языковых моделей (Ahmed Ahmed and others. Extracting Books from Production Language Models), arXiv:2601.02671 (2026 г.) (где установлено, что для извлечения запомненного текста книг из одних развернутых систем требовалось состязательное формирование запросов, тогда как для других этого не требовалось).

[26] См. Heart of the Matter, выше, сноска 10, часть II.C (где проводится аналогия между обучением языковой модели и компиляцией компьютерной программы в исполняемую форму и отмечается, что в обоих случаях «оригинальное произведение, охраняемое авторским правом» «переводится в иную форму, напрямую не понятную человеку», но при этом содержание оригинала «[не] отсутствует в этой понятной для машины версии»).

[27] Лемли М. А., Федер Купер А. Вероятностные «копии» в моделях генеративного ИИ (Cooper A.F. Probabilistic ‘Copies’ in Generative AI Models), Berkeley Technology Law Journal (2026 г.), доступно по адресу https://ssrn.com/abstract=7067878 (где на основе эмпирических исследований по извлечению данных, рассмотренных в настоящей части, обосновывается, что модель следует считать содержащей «копию» конкретного произведения лишь в том случае, когда произведение удается извлечь прямым и, в наиболее очевидных ситуациях, детерминированным способом; что в случаях, когда извлечение запомненных данных происходит лишь с определенной долей вероятности после множества попыток, сохраняется не

Рассмотренные технические доказательства свидетельствуют о том, что обученная модель сохраняет информацию, полученную из произведений в составе ее обучающих данных, пусть даже глубина преобразования существенна, а доступ к любому конкретному произведению может быть частичным и вероятностным. То, как это техническое наблюдение переносится в плоскость правовой квалификации модели и соотносится с вопросом о том, представляет ли собой сама модель копию либо переработку / производное произведение от произведений, на которых она обучалась с точки зрения авторского права, служит предметом продолжающегося анализа в судебной практике и научной литературе, как поясняется далее в настоящем исследовании.

На данном этапе можно сформулировать два замечания. Во-первых, довод о запоминании обучающих данных моделью используется как на этапе обучения, где он служит доказательством того, что обучение модели было сопряжено с кодированием копий, так и на этапе получения результатов, выступая подтверждением способности развернутой модели генерировать охраняемые произведения в форме, замещающей оригиналы. Во-вторых, наглядные примеры, на которые ссылаются истцы, касаются конкретных произведений в конкретных моделях. Они скорее подтверждают, что в случае конкретных, отдельных и часто встречающихся произведений запоминание может достигать значительных объемов, однако не доказывают, что оно выступает неизменным свойством всех моделей или что параметры модели в целом сводятся к сохраненным копиям. Правовая квалификация этих вопросов рассматривается в части III[28].

Наконец, стоит отметить, что после завершения основного обучения модель может меняться в результате тонкой настройки, обучения для согласования поведения модели или попыток «машинного отучения» в отношении конкретного контента. Если после такого вмешательства модель перестает воспроизводить определенное произведение, это само по себе не доказывает, что оно изначально в ней не кодировалось. В технической литературе отмечается, что надежное удаление конкретных произведений из обученной модели по-прежнему сопряжено с трудностями, а подобные вмешательства способны непредвиденным образом ухудшать качество работы модели[29].

3. Соотношение входных данных и выходного результата

После развертывания модель функционирует, принимая входные данные от пользователя (запрос) и формируя выходной результат. Результатом ее работы могут служить текст, программный код, изображение, аудио-, видеозапись или иная форма

копия произведения, а набор «ингредиентов»; и что данный вопрос, соответственно, относится к вопросам доказывания и требует ответа в отношении каждого отдельного произведения и каждой отдельной модели).

[28] По вопросу объединенного производства см. дело In re OpenAI, Inc. Copyright Infringement Litigation, № 1:25-md-03143-SHS-OTW (S.D.N.Y.), объединяющее иски, поданные The New York Times и группами авторов против OpenAI и Microsoft. Квалификация обученной модели в качестве механизма хранения и извлечения, а параметров — в качестве сжатой формы загруженных произведений, отражает позицию истцов в поданных исках; противоположная квалификация параметров как статистических обобщений, кодирующих языковые закономерности, выражает позицию ответчиков. Об эмпирических публикациях, предостерегающих от обеих крайних квалификаций, см. исследование Купера и соавторов, цитируемое выше в данном разделе.

[29] Федер Купер А. и др. Машинное отучение работает не так, как вы думаете: уроки для политики, исследований и практики в сфере генеративного ИИ (Feder Cooper A. and others. Machine Unlearning Doesn’t Do What You Think: Lessons for Generative AI Policy, Research, and Practice), 39 Conf. Neural Inf. Processing Syst. (NeurIPS) (2025 г.).

контента. Соотношение между входным запросом и выходным результатом в определенных аспектах аналогично связи поискового запроса и выдачи в поисковой системе, однако эта аналогия носит лишь частичный характер. Модель не обращается к индексированному корпусу и не извлекает оттуда фрагменты; она генерирует новый контент путем выборки из распределения вероятностей, усвоенного в ходе обучения. В ряде случаев генерируемый ею результат тесно связан с конкретными произведениями из обучающих данных; в других случаях результат представляет собой оригинальное произведение, не обнаруживающее явного сходства ни с одним отдельным обучающим примером.

Особое значение имеют три замечания. Во-первых, результат работы модели может представлять собой дословное или почти дословное воспроизведение произведения из обучающих данных; в таком случае выходной результат образует воспроизведение оригинального произведения[30]. Это было продемонстрировано в исследованиях языковых моделей, воспроизводивших фрагменты из охраняемых авторским правом книг, а также моделей генерации изображений, воспроизводивших охраняемые фотографии и произведения изобразительного искусства[31]. Будет ли такое воспроизведение признано нарушением, зависит от ряда дополнительных вопросов, включая применимость исключений и ограничений, роль установленных разработчиком фильтров выходных данных и распределение ответственности между разработчиком модели и пользователем, сформировавшим запрос. Эти вопросы кратко упоминаются далее в настоящем исследовании[32].

Во-вторых, результат работы модели может иметь существенное сходство с произведением из обучающих данных, не выступая его дословным воспроизведением. Это способно образовывать нарушение, например, в силу доктрины существенного сходства — известного стандарта в законодательстве об авторском праве ряда государств.

В-третьих, результат работы модели может оказаться полностью новым в том смысле, что ни одно конкретное произведение из обучающих данных нельзя идентифицировать в качестве его источника. Генеративная модель способна создавать контент, не совпадающий ни с одним отдельным обучающим примером. Более того, в большинстве случаев именно так она и работает. В подобных ситуациях связь между выходным результатом и конкретным охраняемым произведением ослабевает. Остается ли такой результат «основанным на» обучающих данных в каком-либо юридически значимом для авторского права смысле — этот вопрос ставился перед судами нескольких юрисдикций, однако консенсус по нему до сих пор не выработан.

Таким образом, в части I выявлены факты использования охраняемых произведений, а также вопросы их сохранения или воспроизведения на нескольких этапах разработки

---

[30] Если результат работы LLM схож с ранее существовавшим охраняемым произведением, которое модель не использовала при обучении и к которому не имела доступа иным образом, факт создания «копии» представляется гораздо менее очевидным.

[31] Дословное воспроизведение языковыми моделями зафиксировано в научных исследованиях и материалах судебных дел; воспроизведение узнаваемых фотографий и произведений изобразительного искусства моделями генерации изображений подтверждено на примере диффузионных моделей. Количество задокументированных примеров напрямую не предопределяет правовую квалификацию какого-либо конкретного результата, которая зависит от применимого национального законодательства и фактических обстоятельств дела.

[32] О взаимосвязи между исходным запросом, моделью и выходным результатом, а также об остающемся открытым вопросе распределения ответственности между разработчиком и пользователем см. часть III ниже.

и функционирования модели генеративного ИИ: в корпусе, формируемом на начальном этапе, в операциях по хранению и обработке корпуса в процессе обучения, в самой обученной модели и в результатах ее работы после развертывания. На каждом этапе могут затрагиваться основные исключительные права, признаваемые национальным законодательством об авторском праве и международными договорами. Вопрос о том, может ли каждый акт воспроизведения образовывать нарушение prima facie (при условии неприменимости исключения или ограничения), а также вопрос о том, какие механизмы лицензирования соответствующих видов использования существуют или могут быть разработаны, рассматриваются в последующих частях.

ЧАСТЬ II. РАЗРЕШЕНИЕ, ЛИЦЕНЗИРОВАНИЕ И КОМПЕНСАЦИЯ

В части I настоящего исследования на техническом уровне были определены моменты, в которых разработка и функционирование модели генеративного ИИ способны затрагивать основные исключительные права, признаваемые национальным законодательством об авторском праве и международными договорами в области авторского права. В настоящей части фокус внимания смещается с самих прав на механизмы, с помощью которых они могут осуществляться, прекращаться в силу отказа от них или вытесняться в контексте ИИ. Вопрос о правомерности соответствующих способов использования при отсутствии разрешения и, в случае их неправомерности, вопрос о том, как разработчики могут перевести свою деятельность в правовое поле и как обеспечить выплату компенсации правообладателям в ходе этого процесса, приводят к единому доктринальному блоку: институту лицензирования в авторском праве (индивидуального и коллективного, добровольного и установленного законом), во взаимосвязи с исключениями и ограничениями, обусловливающими саму необходимость получения лицензии.

Настоящая часть построена следующим образом. В разделе А рассматриваются основные каналы, по которым разработчики ИИ в настоящее время получают объекты, способные охраняться авторским правом, для обучения своих моделей, а также правовой статус каждого из таких каналов. В разделе В анализируются ключевые стратегические вопросы, определяющие текущую практику лицензирования: права, затрагиваемые процессом обучения моделей, добровольное лицензирование, коллективное лицензирование в трех его основных формах, статус предусмотренного в праве ЕС исключения в отношении интеллектуального анализа текста и данных, научные дискуссии о его применимости к обучению моделей ИИ, а также системы вознаграждения. Отдельные государства приводятся в качестве иллюстрации подходов, принятых в различных регионах мира; они не служат самостоятельным предметом исследования, и данный обзор исчерпывающим не выступает.

РАЗДЕЛ А. СУЩЕСТВУЮЩИЕ ВАРИАНТЫ ПОЛУЧЕНИЯ РАЗРАБОТЧИКАМИ ОХРАНЯЕМЫХ АВТОРСКИМ ПРАВОМ ОБЪЕКТОВ ДЛЯ ОБУЧЕНИЯ МОДЕЛЕЙ

1. Пять каналов получения данных

На практике разработчики ИИ получают данные для обучения одним из пяти способов либо путем их комбинирования. Первый заключается в использовании открытых или курируемых наборов данных, созданных третьими лицами и предоставляемых как бесплатно, так и на возмездной основе. Ярким примером выступает Common Crawl — постоянно обновляемый репозиторий веб-страниц, собранных и размещенных некоммерческим фондом, который прямо или косвенно лежит в основе текстовых обучающих корпусов большинства ведущих языковых моделей. Аналогом для

моделей генерации изображений служат набор LAION-5B и предшествовавшие ему версии[33]. Такие наборы данных сами выступают результатом масштабных операций по автоматизированному извлечению материалов и могут воспроизводить авторско-правовые проблемы, связанные с исходным сбором данных. Их правовой статус будет определяться основаниями получения исходного материала и правом, применимым к последующему использованию набора данных. В государственном секторе начинает формироваться альтернатива подобным курируемым наборам данных: в Республике Корея государственная система лицензирования открытых данных (Gonggong Nuri) была дополнена категорией лицензий, в рамках которой наборы государственных данных содержат прямое указание на то, что их использование для обучения моделей ИИ разрешено[34].

Другим источником служат так называемые «теневые библиотеки» — масштабные несанкционированные хранилища книг, статей, музыки и других произведений, к которым, согласно имеющимся сведениям или утверждениям сторон, прямо либо через включающие их наборы данных обращался целый ряд широко применяемых моделей. Судебные выводы о фактах пиратства в США, рассмотренные ниже в разделе A части III, наглядно иллюстрируют правовые риски, возникающие при опоре на подобные репозитории.

В качестве сравнительного примера стоит отметить, что некоторые суды проводят разграничение между неправомерным получением произведений и самостоятельным вопросом о том, подпадает ли использование этих произведений при обучении под исключение или ограничение. В одном из рассмотренных ниже судебных решений судов США скачивание и сохранение пиратских копий не было признано добросовестным использованием, тогда как обучение модели на правомерно приобретенных копиях тех же произведений суд счел трансформирующим и добросовестным. В этом отношении законность получения данных и законность обучения с аналитической точки зрения не зависят друг от друга. Это различие существенно, поскольку в рамках одной и той же модели обучающие данные могут поступать как из законных, так и из незаконных источников, порождая сложности при лицензировании и обеспечении соблюдения прав. Следовательно, с учетом применимого национального права разработчику могут требоваться как правомерное приобретение корпуса, так и правовое основание для обучения модели как такового.

Третий подход охватывает индивидуальное и коллективное лицензирование. Разработчик может приобрести доступ к курируемому корпусу либо договориться о его предоставлении с издателем, новостной организацией, вещательной организацией, фотобанком, оператором научной базы данных или иными институциональными правообладателями. За последние три года лицензирование стало привычной практикой: ведущие разработчики ИИ заключают сделки с крупнейшими новостными

---

[33] Информация о Common Crawl приведена по адресу https://commoncrawl.org. Набор данных LAION-5B (использованный при создании Stable Diffusion и других моделей генерации изображений) и предшествовавший ему LAION-400M представляют собой общедоступные наборы пар «изображение — текст», сформированные организацией Large-scale Artificial Intelligence Open Network. Сам набор данных содержит не исходные изображения, а пары из URL-адресов и текстовых описаний; изображения загружаются из мест их первоначального размещения непосредственно во время обучения. О правовой квалификации действий по формированию наборов данных в качестве интеллектуального анализа текста и данных см. решение по делу Kneschke v. LAION e.V., LG Hamburg, 27 сентября 2024 г. (310 O 227/23), а также постановление суда апелляционной инстанции: Kneschke v. LAION e.V., OLG Hamburg, 10 декабря 2025 г. (5 U 104/24).

[34] Имеется в виду система Gonggong Nuri (Korea Open Government License), в рамках которой государственные учреждения предоставляют открытые данные для повторного использования по стандартизированным категориям лицензий; обновленная система дополнена категорией, прямо разрешающей использование данных для обучения моделей ИИ.

агентствами и фотобанками Европы, Америки и Азии[35]. Конструктивная особенность лицензирования состоит в том, что разработчик получает юридически выверенную цепочку перехода прав на лицензируемый корпус, как минимум в объеме собственных прав лицензиара. Ограничения такой цепочки рассматриваются ниже в разделе B.2. В ряде случаев договор предусматривает непосредственную передачу контента.

Четвертым, и ведущим по объему каналом выступает прямой сбор материалов из открытой части Интернета путем автоматизированного обхода веб-сайтов и сбора данных с веб-страниц, выполняемого самим разработчиком, а не третьей стороной. Технические процедуры описаны в разделе A.2 части I настоящего исследования. Масштабы такого автоматизированного извлечения материалов весьма внушительны: по некоторым оценкам, на долю автоматизированных агентов в настоящее время приходится большая часть мирового интернет-трафика[36]. Автоматизированный обход зачастую осуществляется без предварительного согласия правообладателей извлекаемых материалов на основании одной из трех правовых концепций: что материал находится в общественном достоянии либо надлежащим образом лицензирован неопределенному кругу лиц на разрешительных условиях; что операция подпадает под действие исключения или ограничения в применимом авторском праве (в Европейском союзе — исключение в отношении интеллектуального анализа текста и данных по статье 4 Директивы об авторском праве на едином цифровом рынке (CDSM); в США и, в силу статьи 35-5 закона «Об авторском праве», в Республике Корея — общая доктрина добросовестного использования; в Сингапуре и Японии — специальные законодательные нормы); либо что операция вовсе не затрагивает право на воспроизведение, поскольку создаваемые копии носят временный или кратковременный характер, либо что в ряде юрисдикций такие копии подпадают под прямо предусмотренное исключение в отношении временных актов воспроизведения (как, например, в соответствии со статьей 5(1) Директивы об авторском праве в информационном обществе (InfoSoc Directive) в Европейском союзе). Эти концепции не исключают друг друга, и разработчик может опираться на различные теории применительно к разным частям корпуса.

Пятый канал — синтетические данные: данные, сгенерированные существующей моделью ИИ и используемые для обучения или тонкой настройки другой модели. Синтетические данные приобретают все большее значение в качестве инструмента, чаще всего применяемого на более поздних этапах разработки модели, в частности при тонкой настройке с учителем и моделировании вознаграждения[37]. С позиций авторского права синтетические данные ставят принципиально иной вопрос. Если модель, сгенерировавшая такие данные, обучалась на охраняемых произведениях,

---

[35] О двусторонних соглашениях о лицензировании между разработчиками ИИ и издателями см. (среди многих других) договоры, заключенные OpenAI с Associated Press, News Corp, Le Monde, Prisa Media, Axel Springer, Vox Media, The Atlantic и Hearst; компанией Google — с Reddit и Associated Press; а также компанией Anthropic — с рядом издательств. Сведения о сделках раскрываются с различной степенью детализации, а их финансовые условия составляют коммерческую тайну.

[36] См. Imperva, Отчет о вредоносных ботах за 2025 год (Bad Bot Report, 12-е издание, 2025 г.), где сообщается, что в 2024 году автоматизированный трафик составил 51 процент от всего интернет-трафика, впервые за десятилетие превысив трафик, генерируемый людьми, при этом на долю так называемых «вредоносных ботов» пришлось 37%.

[37] Синтетические данные представляют собой данные, созданные существующей моделью ИИ и используемые при обучении или тонкой настройке другой модели. Они порождают особые вопросы в области авторского права в тех случаях, когда создавшая их модель сама обучалась на охраняемых произведениях. О расширении их использования на более поздних этапах разработки моделей см. Лю Р. и др. Передовой опыт и извлеченные уроки в отношении синтетических данных (Liu R. and others. Best Practices and Lessons Learned on Synthetic Data), arXiv:2404.07503 (2024 г.), доклад представлен на Первой конференции по языковому моделированию (First Conference on Language Modeling).

воспроизводят ли синтетические результаты охраняемую форму выражения либо представляют ли собой ее переработку? Этот вопрос еще не передавался на рассмотрение суда таким образом, чтобы получить обязательное для применения судебное решение.

2. Формы сотрудничества

Наконец, помимо этих пяти основных путей, в ряде юрисдикций наблюдаются формы организованного сотрудничества между разработчиками ИИ и институциональными или государственными правообладателями, имеющими наибольшее значение для конкретных отраслевых сфер применения. Одним из таких примеров выступает Китай. При разработке больших моделей для специализированных предметных областей наблюдается сотрудничество между разработчиком моделей ИИ общего назначения и телерадиокомпаниями Гонконга (для получения голосовых данных на кантонском диалекте), между разработчиками и онлайн-издательствами, обладающими обширным репертуаром авторов по контрактам (для сбора массивов цифровых романов), между разработчиками и крупными фотобанками с миллиардами изображений, полученных по договорам об отчуждении прав или исключительным лицензиям (для получения визуальных обучающих данных), а также между разработчиками и операторами академических баз данных при совместном создании специализированных моделей. По имеющимся данным, эта модель носит распространенный характер, хотя сами соглашения зачастую не озаглавлены как лицензии на использование объектов авторского права, а их точный объем раскрывается далеко не всегда. Как показано ниже в разделе B.2, структурное ограничение подобного подхода заключается в том, что издатель, вещательная организация или хранилище обычно обладают лишь определенными правами, предоставленными первоначальным автором, при этом остается неясным, охватывают ли имеющиеся права обучение моделей ИИ.

РАЗДЕЛ B. ОСНОВНЫЕ ЭЛЕМЕНТЫ ГОСУДАРСТВЕННОЙ ПОЛИТИКИ

1. Права, затрагиваемые обучением моделей генеративного ИИ

(a) Право на воспроизведение

Технические операции, описанные в части I, prima facie затрагивают ряд исключительных прав, признаваемых национальным законодательством об авторском праве и международными договорами по авторскому праву. Центральное место в анализе занимает право на воспроизведение[38]. Формирование обучающего корпуса предполагает создание копий; хранение и обработка этого корпуса на различных этапах обучения связаны с созданием дополнительных копий — как долговременных, так и кратковременных; отдельные аспекты самой обученной модели, согласно техническому анализу, приведенному в разделе C.2 части I, могут рассматриваться в качестве копий произведений, на которых модель обучалась. Каждая из этих операций prima facie подпадает под сферу действия права на воспроизведение, закрепленного в статье 9(1) Бернской конвенции, в статье 1(4) Договора ВОИС по авторскому праву, статье 7 ДИФ и Пекинского договора, а также в соответствующих положениях национального законодательства.

Режим временных и кратковременных копий требует отдельного рассмотрения, поскольку два различных правовых приема приводят к схожим практическим результатам: в одних юрисдикциях такие копии не признаются подпадающими под

[38] См. часть I, разделы A.1 и A.2.

сферу действия права на воспроизведение, тогда как в других они подпадают под действие прямо предусмотренного исключения. Основным примером применения первого приема служит Китай, где временное воспроизведение последовательно рассматривается как выходящее за рамки действия этого права. В китайской судебной практике исключаемые таким образом временные воспроизведения представляют собой кратковременные копии, создаваемые в оперативной памяти или в кэше и строго необходимые для процесса обработки данных; при этом устоявшаяся позиция относительно того, как долго копия может существовать, чтобы по-прежнему считаться временной, отсутствует. В статье 10(1)(5) закона «Об авторском праве» воспроизведение определяется как право создавать один или более экземпляров произведения такими способами, как полиграфическое воспроизведение, фотокопирование, гравирование, звукозапись, видеозапись, дублирование, фотографирование или оцифровка[39], и китайская делегация, в числе прочих, проголосовала против согласованного заявления на Дипломатической конференции, которое однозначно распространило бы сферу действия права на временное хранение в цифровой форме[40]. Положение «Об охране права на сообщение для всеобщего сведения по информационным сетям» (Regulation on the Protection of the Right of Communication to the Public on Information Networks, 2006 г.) не содержало нормы, признающей временное воспроизведение актом воспроизведения, и китайские суды применяют такой подход в делах, связанных с транскодированием и аналогичными кратковременными цифровыми операциями. С этой точки зрения сам процесс обучения модели в той мере, в какой он предполагает лишь кратковременный доступ к произведениям в оперативной памяти или кэше, исчезающий по завершении цикла обучения, может не затрагивать право на воспроизведение по праву Китая, хотя формирование базовой базы данных произведений, безусловно, способно затрагивать такое право. В других юрисдикциях тот же практический результат достигается посредством прямо предусмотренного исключения: в Европейском союзе статья 5(1) Директивы об авторском праве в информационном обществе (InfoSoc Directive) распространяется на кратковременные и случайные копии через узко сформулированное исключение, а не путем их исключения из сферы действия права ab initio. По существу рассматриваемые копии в обеих системах одинаковы; различие состоит лишь в применяемом правовом приеме[41]. В большинстве других рассмотренных юрисдикций процесс обучения, как правило, признается затрагивающим право на воспроизведение, и аналитический вопрос сводится к наличию исключения или ограничения.

(b) Право на переработку и производные произведения

Помимо права на воспроизведение, затрагиваются или могут затрагиваться еще три права. Право на переработку, закрепленное в статье 12 Бернской конвенции и в корреспондирующих нормах национального права, приводится на этапе получения результатов, поскольку генерация выдачи рассматривается как приводящая к

---

[39] Закон КНР «Об авторском праве», статья 10(1)(5).

[40] См. документ ВОИС CRNR/DC/101, Summary Minutes, Plenary, paras 497-502. Тем не менее Согласованное заявление в отношении статьи 1(4) ДАП было принято Дипломатической конференцией; Китай проголосовал против него.

[41] Статья 5(1) Директивы 2001/29/ЕС Европейского парламента и Совета от 22 мая 2001 г. о гармонизации определенных аспектов авторского права и смежных прав в информационном обществе, OJ L 167/10 [часто именуемой «Директивой InfoSoc»]. Исключение охватывает временные акты воспроизведения, которые носят кратковременный или случайный характер, составляют неотъемлемую и существенную часть технологического процесса, преследуют единственную цель сделать возможной передачу в сети между третьими лицами через посредника либо правомерное использование, и не имеют самостоятельного экономического значения.

трансформации или изменению первоначальных произведений. Например, бразильское законодательство прямо относит право на переработку к формам использования, требующим предварительного и прямого согласия[42], а Секретариат по авторскому праву и интеллектуальной собственности Ministry of Culture (Министерство культуры Бразилии) занял позицию, согласно которой генерация результатов работы модели затрагивает право на переработку наряду с правом на воспроизведение. Позиция в канадском, французском и немецком праве в целом с этим согласуется.

(c) Сообщение для всеобщего сведения и доведение до всеобщего сведения

Право на доведение до всеобщего сведения, предусмотренное статьей 8 Договора ВОИС по авторскому праву и статьями 10 и 14 Договора ВОИС по исполнениям и фонограммам, задействуется прежде всего на этапе выдачи результатов, когда развернутая модель выдает контент пользователю, если такой контент содержит копию охраняемой формы выражения. В Бразилии, которая не участвует в ДАП или ДИФ, *Superior Tribunal de Justiça* (Высший суд правосудия) в 2017 году по делу REsp 1.559.264-RJ постановил, что потоковая передача музыки через Интернет затрагивает право на сообщение для всеобщего сведения, охватывающее то, что в других юрисдикциях может квалифицироваться как право на доведение до всеобщего сведения[43]. Это соответствует «зонтичному решению», принятому на Дипломатической конференции 1996 года, согласно которому право на доведение до всеобщего сведения может быть имплементировано через право на сообщение для всеобщего сведения, право на распространение либо их сочетание. Применение аналогичной аргументации к результатам работы развернутой модели ИИ в бразильских судах пока не проверялось[44]. Исполнители и производители фонограмм обладают параллельными правами на доведение до всеобщего сведения согласно статьям 10 и 14 ДИФ, упомянутым выше, благодаря чему анализ выходит за пределы исключительно авторских прав.

(d) Личные неимущественные права

Личные неимущественные права также могут затрагиваться. Чаще всего заявляется о праве на неприкосновенность произведения. В Бразилии фрагментация произведений в ходе процесса обучения рассматривается как потенциальное посягательство на право на неприкосновенность произведения, а отсутствие указания авторов использованных материалов — как потенциальное нарушение права авторства[45].

---

[42] Закон Бразилии «Об авторском праве» (Закон № 9.610/1998), статья 29, пункты III и IV (переработка, музыкальная аранжировка, перевод и иные преобразования). Позиция Ministry of Culture (Министерство культуры Бразилии) была сформулирована, в частности, в технической записке, направленной в Autoridade Nacional de Proteção de Dados (Национальный орган по защите данных) в октябре 2024 г. в отношении обучения моделей одного из ведущих разработчиков, со ссылкой на статью 29(III) (преобразование для создания производных произведений); см. публикацию: https://nucleo.jor.br/curtas/2024/10/10/ministerio-cultura-prejuizos-direitos-ia-meta/, дата обращения: 7 июля 2026 г.

[43] Superior Tribunal de Justiça, REsp 1.559.264-RJ (2017 г.) (где установлено, что потоковая передача музыки через Интернет представляет собой форму публичного исполнения и затрагивает право на сообщение для всеобщего сведения, и что ECAD имеет право взимать вознаграждение на этом основании).

[44] Закон Бразилии «Об авторском праве» (Закон № 9.610/1998), статья 5(V) (определение «сообщения для всеобщего сведения») и статья 29(VII) (распространение посредством любой системы, позволяющей пользователю выбирать произведение).

[45] Закон Бразилии «Об авторском праве» (Закон № 9.610/1998), статья 24, пункты I–IV. Исполнители также обладают личными неимущественными правами на неприкосновенность и авторство в отношении своих исполнений в силу статьи 92.

Закон Канады «Об авторском праве» признает и право на неприкосновенность произведения, и право авторства в качестве личных неимущественных прав автора[46]. В ряде юрисдикций традиций *droit d'auteur* вопрос о том, образует ли использование произведения в обучающем корпусе без указания авторства и способом, способным генерировать результаты, воспринимаемые как искаженные или умаляющие честь и достоинство версии оригинала, нарушение личных неимущественных прав, поднимался в научных комментариях, но пока не получил авторитетного окончательного решения.

Наконец, в Европейском союзе и в некоторых юрисдикциях, установивших аналогичную охрану, особое право sui generis на базу данных, охраняемую Директивой 96/9/ЕС, может затрагиваться в тех случаях, когда обучающий корпус формируется путем извлечения и повторного использования существенных частей охраняемой базы данных[47]. Это право не зависит от авторского права и опирается на собственные критерии существенных инвестиций и «существенной части» (для «права на извлечение»). Пределы его применения к формированию обучающих корпусов ИИ остаются неопределенными, в том числе применительно к базам данных правовой информации, агрегаторам научных журналов и крупным фотобанкам.

Следовательно, права, потенциально затрагиваемые обучением модели генеративного ИИ, могут выходить за рамки права на воспроизведение. Очищает ли двусторонняя или коллективная лицензия соответствующие виды использования, зависит от круга охватываемых ею прав; лицензия, ограниченная воспроизведением, может не защищать от претензий, основанных на переработке, доведении до всеобщего сведения, личных неимущественных правах или, где применимо, особом праве sui generis на базу данных. Следующий аспект, рассматриваемый в части III в контексте трансграничного обеспечения соблюдения прав, заключается в том, что совокупность правомочий и точный объем каждого права различаются от юрисдикции к юрисдикции, в связи с чем лицензионная договоренность, призванная очистить права в одной юрисдикции, может оставить разработчика уязвимым в другой[48].

(e) Информация об управлении правами и меры против обхода технических средств

В соответствии со статьями 11 и 12 ДАП, статьями 18 и 19 ДИФ и статьями 15 и 16 Пекинского договора правообладатели имеют право на надлежащую правовую защиту и эффективные средства правовой защиты от обхода действующих технических средств, используемых ими в связи с осуществлением своих прав по указанным договорам и Бернской конвенции, а также на надлежащие и эффективные средства правовой защиты в отношении любого лица, умышленно совершающего следующие действия, зная или, применительно к гражданско-правовым санкциям, имея разумные основания знать, что такие действия повлекут, позволят, облегчат или скроют нарушение любого права, охватываемого этими договорами или Бернской конвенцией:

(i) устранение или изменение любой электронной информации об управлении правами без разрешения;

---

[46] Copyright Act, R.S.C. 1985, c. C-42, пункты 14.1 и 28.1.

[47] Директива 96/9/ЕС Европейского парламента и Совета от 11 марта 1996 г. о правовой защите баз данных, OJ L 77/20, статья 7 (особое право sui generis). См. также решение Суда Европейского союза (CJEU), C-203/02, *British Horseracing Board v. William Hill*, ECLI:EU:C:2004:695, по вопросу условий применения критерия существенных инвестиций.

[48] См. *Toronto Star Newspapers Ltd v. OpenAI Inc.*, 2025 ONSC 6217 (судья Киммел, 7 ноября 2025 г.) (где отклонено ходатайство OpenAI по вопросу юрисдикции на основании наличия реальной и существенной связи с Онтарио).

(ii) распространение, ввоз для распространения, передачу в эфир или сообщение для всеобщего сведения без разрешения произведений или экземпляров произведений, фонограмм или исполнений (в соответствии с положениями соответствующих договоров) при осведомленности о том, что электронная информация об управлении правами была устранена или изменена без разрешения.

Насколько нам известно, на момент подготовки настоящего документа успешных исков против компаний в связи с обучением моделей ИИ по данным основаниям подано не было.

2. Добровольное (коллективное и индивидуальное) лицензирование

(a) Индивидуальное и коллективное лицензирование использования произведений в системах ИИ

Добровольное лицензирование предполагает согласование разрешения между разработчиком ИИ и правообладателем либо представителем множества правообладателей, включая ОКУ или иного посредника, при отсутствии каких-либо установленных законом полномочий. Во всех исследованных юрисдикциях такой порядок служит правилом по умолчанию.

Практическую форму добровольного лицензирования определяют две структурные особенности. Первая представляет собой асимметрию между небольшим числом крупных институциональных правообладателей, способных вести переговоры с разработчиками на коммерчески значимых условиях, и весьма значительным числом индивидуальных авторов и мелких правообладателей, для которых подобные переговоры зачастую практически неосуществимы вследствие трансакционных издержек, нехватки специальных знаний или иных факторов. Крупные двусторонние соглашения, заключенные за последние три года, подписывались между разработчиками ИИ и крупными издательствами, новостными организациями, вещательными компаниями и фотобанками; сопоставимая инфраструктура для отдельных авторов, фотографов, иллюстраторов или музыкантов отсутствует, в связи с чем получение ими какой-либо доли создаваемой стоимости нередко зависит от сделок, заключаемых их издателями и другими лицензиатами.

Масштабы некоторых из этих договоренностей в настоящее время весьма существенны. Сумма одного пятилетнего соглашения между крупным новостным издателем и разработчиком ИИ, по сообщениям, превысила 250 млн долл. США[49], а лицензирование контента для компьютерного анализа стало существенной статьей доходов для крупных фотобанков: В годовом отчете Shutterstock, Inc. за 2025 год доход по направлению «Данные, дистрибуция и услуги» составил 203,3 млн долл. США, что соответствует примерно 20,5% от совокупной выручки в размере 989,9 млн долл. США; компания Getty Images отдельно сообщила о годовой выручке за 2025 год в размере 981,3 млн долл. США. Компании объявили о слиянии на равных 7 января 2025 г.; 30 июня 2026 г. Getty Images объявила о расторжении соглашения о слиянии

[49] О пятилетнем соглашении между News Corp и OpenAI издание The Wall Street Journal сообщало как о соглашении на сумму свыше 250 млн долл. США. Финансовые условия таких соглашений обычно конфиденциальны, и указанная цифра приведена в прессе, а не подтверждена сторонами публично. См. «OpenAI, WSJ Owner News Corp Strike Content Deal Valued at Over $250 Million» (The Wall Street Journal, 22 мая 2024 г.) <https://www.wsj.com/business/media/openai-news-corp-strike-deal-23f186ba>, дата обращения: 3 июля 2026 г.; «Sam Altman's OpenAI signs content agreement with News Corp» (Reuters, 22 мая 2024 г.) <https://finance.yahoo.com/news/sam-altmans-openai-signs-content-201808782.html>, дата обращения: 3 июля 2026 г.

после того, как Competition and Markets Authority (CMA) (Управление по конкуренции и рынкам Великобритании) обусловило одобрение сделки выделением и продажей редакционного бизнеса Shutterstock. По имеющимся сведениям, соглашение было расторгнуто 7 июля 2026 г.[50] Избирательный характер этого рынка иллюстрирует пример французской прессы: в марте 2024 года газета Le Monde заключила соглашение с OpenAI, охарактеризованное ею как первое соглашение между французским СМИ и крупным разработчиком генеративного ИИ[51]. Позже в том же году ассоциация APIG попыталась инициировать коллективные переговоры с OpenAI, однако опубликованные сообщения указывают на то, что к более широкой лицензионной договоренности в тот период эти переговоры не привели[52]. Ответом на эту асимметрию стало появление посредников, агрегирующих права более мелких издателей: один поставщик деловой информации сообщает, что несколько тысяч издателей из его партнерской сети (отобранных из гораздо более широкого пула) дали согласие на лицензирование своего контента для использования в системах ИИ[53], а разработчик лицензированной вопросно-ответной поисковой системы заявляет о сопоставимом объединении участвующих изданий[54].

Второй особенностью выступает вопрос о цепочке перехода прав. Во многих юрисдикциях, когда издателю, вещательной организации или хранилищу были предоставлены определенные права первоначальным автором на основании лицензионного договора или договора об отчуждении права, заключенного до появления генеративного ИИ, вопрос о распространении этих прав на использование произведения при обучении модели ИИ характеризуется правовой неопределенностью и условиями договора не урегулирован. Подобная неопределенность присуща не только сфере ИИ: аналогичные вопросы возникают по

---

[50] Shutterstock, «Shutterstock Reports Full Year 2025 and Fourth Quarter Financial Results» (17 февраля 2026 г.) <https://investor.shutterstock.com/node/14551/pdf>, дата обращения: 3 июля 2026 г.; Getty Images Holdings, Inc, «Getty Images Reports Fourth Quarter and Full Year 2025 Results» (16 марта 2026 г.) <https://investors.gettyimages.com/node/15726/pdf>, дата обращения: 3 июля 2026 г.; Getty Images Holdings, Inc and Shutterstock, Inc, «Getty Images and Shutterstock to Merge, Creating a Premier Visual Content Company» (7 января 2025 г.) <https://www.sec.gov/Archives/edgar/data/1898496/000110465925001732/tm252186d1_ex99-1.htm>, дата обращения: 3 июля 2026 г. О прекращении сделки по слиянию см. Competition and Markets Authority (CMA) (Управление по конкуренции и рынкам Великобритании), «Getty abandons Shutterstock merger plans» (июль 2026 г.) <https://www.gov.uk/government/news/getty-abandons-shutterstock-merger-plans>, дата обращения: 31 июля 2026 г.

[51] Соглашение, заключенное в марте 2024 г., охарактеризовано Le Monde как первое соглашение между французским медиаизданием и крупным разработчиком моделей генеративного ИИ. См. «Le Monde signs artificial intelligence partnership agreement with OpenAI» (Le Monde, 13 марта 2024 г.) <https://www.lemonde.fr/en/about-us/article/2024/03/13/le-monde-signs-artificial-intelligence-partnership-agreement-with-open-ai_6615418_115.html>, дата обращения: 3 июля 2026 г.

[52] Речь идет об Alliance de la presse d'information générale (APIG) (Альянс прессы общей информационной направленности), которая совместно с Syndicat des éditeurs de la presse magazine (SEPM) (Профсоюз издателей журнальной прессы) в июне 2024 г. официально запросила начало переговоров с OpenAI; по имеющимся данным, разработчик ответил, что на данном этапе не видит необходимости заключать новые лицензионные соглашения с французскими издателями. См. «Intelligence artificielle : un accord de partenariat entre Le Monde et OpenAI» (Le Monde, 13 марта 2024 г.) <https://www.lemonde.fr/le-monde-et-vous/article/2024/03/13/intelligence-artificielle-un-accord-de-partenariat-entre-le-monde-et-openai_6221836_6065879.html>, дата обращения: 3 июля 2026 г.

[53] Служба Factiva компании Dow Jones сообщила, что к концу 2025 года около 5 000 издателей из сети, насчитывающей примерно 30 000 участников, приняли решение лицензировать свой контент для использования в системах генеративного ИИ. См. Dow Jones, «Factiva Sees Growing Publisher Participation in AI Licensing» (Dow Jones, 2025 г.) <https://www.dowjones.com>, дата обращения: 3 июля 2026 г.

[54] Компания ProRata.ai, по ее собственным утверждениям, объединяет в своей сети «более 1 000 изданий и создателей контента».

договорам между авторами и музыкальными издательствами, между исполнителями и производителями фонограмм, а также между аудиовизуальными продюсерами и творческим составом проекта во всех случаях, когда конкретный способ использования не охватывался намерениями сторон при заключении договора; специфику контекста ИИ составляют масштабы возникающей проблемы и практические трудности пересмотра договорных условий. В Китае, например, большинство авторов передают своим издательствам исключительное право на издание произведений в печатной и, все чаще, в цифровой форме. Сюда относятся права воспроизводить произведение в бумажной форме, распространять бумажные экземпляры, а также воспроизводить и распространять цифровые копии. Ни один из этих способов использования нельзя с очевидностью толковать как охватывающий использование произведения при обучении модели ИИ. Поэтому в положении поставщиков корпусов и разработчиков ИИ, получающих авторско-правовые лицензии от издателей, законность лицензируемых действий по отношению к первоначальным авторам по-прежнему вызывает сомнения. Даже когда разработчик получил лицензию от институционального правообладателя, сохраняется остаточный риск предъявления требований со стороны первоначального автора. В зависимости от юрисдикции такие риски не ограничиваются лишь лицензиаром: лицензиар может нести ответственность за нарушение договора или закона, а разработчик как сторона, в конечном счете извлекающая выгоду из использования, может сам оставаться уязвимым перед требованиями в связи с нарушением права со стороны первоначальных авторов, несмотря на наличие лицензии. В аудиовизуальном секторе, по имеющимся сведениям, типовые лицензионные договоры решают эту проблему договорным путем: лицензиар заверяет и гарантирует наличие у него всех прав, необходимых для использования материалов при обучении моделей. В отношении газетных статей, написанных журналистами и редакторами до вступления в силу соответствующих поправок к закону «Об авторском праве» КНР 1 июня 2021 г., авторские права на такие статьи, как правило, принадлежали журналисту или редактору, а не новостному агентству. Это означало, что даже лицензия от новостного агентства не очищала использование статьи для обучения моделей ИИ. Подобные ограничения могут сами по себе препятствовать формированию рынка добровольного лицензирования, поскольку получение лицензии от посредника с неопределенным объемом собственных прав создает дополнительный уровень потенциальных нарушений в цепочке перехода прав. Они имеют также распределительное измерение: когда лицензирование осуществляется при посредничестве издательств, вещательных организаций и хранилищ контента, неравенство переговорных позиций между отдельными авторами или исполнителями и этими институциональными субъектами приводит к тому, что выгоды от лицензирования могут не доходить до конкретных создателей, чьи произведения используются, а это соображение существенно для выбора между моделями лицензирования, рассматриваемыми в разделе B.

Вместе с тем китайская практика показывает, что даже в рамках таких структурных ограничений масштабное добровольное лицензирование имеет место в отраслевом контексте и в специализированных предметных областях. Гонконгские радио- и телевещательные станции сотрудничают с разработчиками ИИ для предоставления высококачественных голосовых данных на кантонском диалекте. Цифровые романы предоставляются в рамках сотрудничества разработчиков с онлайн-издательствами, привлекающими большое число писателей на контрактной основе. Изображения предоставляются в рамках сотрудничества между разработчиками и крупными фотобанками, владеющими ими на основании договоров об отчуждении прав или исключительных лицензий от фотографов. Научные статьи предоставляются в ходе взаимодействия разработчиков с операторами академических баз данных, которые в ряде случаев выступают соразработчиками создаваемой предметно-

специализированной модели. Некоторые поставщики корпусов также создают собственные обучающие материалы, заключая договоры с фотографами, моделями и школьными учителями об отчуждении прав или предоставлении лицензий на созданные произведения для использования в обучающих наборах данных.

В других юрисдикциях наблюдаются схожие тенденции, хотя институциональные механизмы там отличаются. В Бразилии правовой режим добровольного лицензирования подчиняется общим нормам закона «Об авторском праве». Необходимость получения предварительного и прямого согласия в соответствии со статьей 29, а также ограничительное толкование сделок с авторскими правами согласно статьям 4, 31 и 49 означают, что даже при полном отчуждении имущественных прав до появления генеративного ИИ использование произведения для обучения моделей ИИ останется спорным, поскольку договор не мог прямо предусматривать технологию, которой на тот момент еще не существовало. Аналогичные доктрины буквального толкования встречаются во французской, итальянской и других юрисдикциях традиции *droit d'auteur*.[55]

Если закон прямо закрепляет авторское право на постановки, исполнения или иные определенные объекты, ситуация с добровольным лицензированием может оказаться более простой, поскольку исходные правомочия определены более четко. Более серьезные сложности могут возникать в отношении категорий произведений, создаваемых множеством участников, таких как аудиовизуальные произведения и фонограммы, когда неопределенность в праве или в договорах между изготовителями и отдельными авторами способна либо затруднять добровольное лицензирование, либо лишать этих участников причитающихся выгод. Примером служит законодательство Сенегала, прямо признающее постановки охраняемыми произведениями[56], хотя практическое применение этой нормы к обучению моделей ИИ пока не проверялось. В Кении закон «Об авторском праве» (Copyright Act, Cap 130) закрепляет стандартную совокупность правомочий исключительных прав и предусматривает коллективное управление авторскими правами через ОКУ, однако институциональная инфраструктура не рассчитывалась на компьютерный анализ данных, и добровольное лицензирование для обучения моделей ИИ в качестве зафиксированной практики пока отсутствует. В Канаде добровольное лицензирование служит основным механизмом, а согласование условий договоров выступает центральным звеном этой деятельности. Основное текущее судебное разбирательство, дело Toronto Star Newspapers Ltd v. OpenAI Inc., возникло в связи со спорным несанкционированным использованием, а не с вопросами лицензирования как такового, и рассматривается в части III настоящего исследования.

Характерной особенностью добровольного лицензирования, будь то индивидуального или коллективного, выступает адаптация условий лицензии под профиль использования данных той системы, для которой эти данные запрашиваются, причем этот профиль существенно варьируется в зависимости от категории применяемых сегодня систем генеративного ИИ. Когда данные используются однократно — для первоначального обучения или тонкой настройки модели, веса которой затем фиксируются, — типичной структурой выступает единовременный платеж либо платеж в виде доли дохода за определенный корпус. В ситуациях, когда системе требуется непрерывный доступ к лицензированному материалу при генерации результатов

---

[55] О действовавшем до 2021 года правиле принадлежности прав в авторском праве Китая в отношении газетных статей, созданных журналистами и редакторами, см. закон КНР «Об авторском праве» (в редакции 2020 г., действующей с 1 июня 2021 г.), статья 18 (служебные произведения).

[56] Закон Сенегала «Об авторском праве» (Loi n° 2008-09 du 25 janvier 2008), статья 2. Аналогичные положения содержатся в Бангийском соглашении OAPI (пересмотренный Акт 2015 г.), приложение VII.

работы (как в случае с RAG и системами, осуществляющими постоянный интеллектуальный анализ текста и данных на базе библиотеки), доступ представляет собой не разовое действие, а длящиеся отношения. Сложившиеся на практике модели компенсации отражают эту специфику: от периодических платежей до платы за каждое сканирование или платы за каждый запрос на этапе логического вывода. Помимо цены, ключевыми вопросами, требующими урегулирования сторонами, выступают разрешенный объем использования, срок действия лицензии, ее территориальная сфера действия, последствия ее прекращения и возможность отчуждения модели третьим лицам. В контексте ИИ возникает вопрос о том, означает ли прекращение действия лицензии обязанность разработчика прекратить использование модели, уже обученной на лицензированном материале, переобучить ее или уничтожить сохраненные копии; при этом ни один из указанных вариантов не вписывается четко в средства правовой защиты, выработанные для традиционных лицензионных отношений. Прекращение лицензии наглядно иллюстрирует эту сложность: при традиционном лицензировании расторжение договора восстанавливает контроль правообладателя над будущими способами использования, однако правообладатель, прекращающий действие лицензии на обучение моделей ИИ, с практической точки зрения не может принудить разработчика прекратить использование того, что обучение уже зафиксировало в модели, особенно когда соответствующий каталог мал по сравнению со всем корпусом; средством защиты в таком случае — если оно вообще применимо — должны были бы стать переобучение либо попытки машинного отучения, ограничения которых отмечены в части I. Даже если такие меры окажутся осуществимыми, они влекут за собой эксплуатационные и вычислительные затраты, распределение которых сторонам может потребоваться закрепить непосредственно в договоре, что создает дополнительную сложность при согласовании подобных лицензий. Когда веса такой модели публикуются или продаются последующим пользователям, возникает вопрос о возможности передачи прав: Следует ли договорное обязательство, связанное с лицензией, за весами модели? Смежным, но самостоятельным выступает вопрос о том, может ли вознаграждение на последующих этапах рассчитываться на основе самой модели, например с учетом вклада лицензированного материала в ее веса. Эти вопросы пока не получили разрешения в ставших достоянием гласности договорах и имеют значение для возможного выбора между моделями вознаграждения, рассматриваемыми ниже в разделе B.6.

В высокодоходном сегменте добровольное лицензирование в музыкальной и аудиовизуальной сферах преимущественно принимает форму индивидуализированных партнерств, объединяющих выдачу лицензии с коммерческими условиями, выходящими далеко за рамки простого использования охраняемых произведений. На протяжении 2025 года ведущие звукозаписывающие компании перешли от судебных разбирательств к лицензированию с основными разработчиками музыкальных инструментов ИИ: два таких разработчика урегулировали предъявленные к ним требования в связи с нарушением права и согласились создать новые лицензированные платформы (причем одно урегулирование сопровождалось приобретением разработчиком сервиса поиска концертов у звукозаписывающей компании), а стриминговый сервис отдельно договорился с тремя крупнейшими звукозаписывающими компаниями и еще двумя правообладателями о разработке музыкальных продуктов ИИ на основе предварительного согласия[57]. Этим договоренностям об урегулировании

[57] Universal Music Group и Warner Music Group урегулировали свои судебные споры против Udio в конце 2025 г. и заключили лицензионные соглашения в отношении новых платформ; Warner Music Group дополнительно урегулировала спор с Suno в рамках соглашения, предусматривавшего приобретение

предшествовали совместные заявления о принципах в данной отрасли: коалиции отраслевых структур, организаций авторов и технологических компаний сформулировали принципы, обусловливающие использование охраняемого репертуара при разработке систем ИИ получением разрешения и лицензированием, а также призывающие к обеспечению прозрачности в практике обучения моделей[58]. В аудиовизуальном секторе объявленная в декабре 2025 года лицензионная договоренность, по условиям которой киностудия должна была предоставить более двухсот своих персонажей для платформы генерации видео и приобрести долю в капитале разработчика на сумму 1 млрд долл. США, по имеющимся сведениям, была аннулирована в марте 2026 года в связи с отзывом платформы разработчиком[59]. Эти договоренности отчасти затрагивают статус результатов работы моделей, выходящий за рамки настоящего исследования; здесь они приводятся лишь в качестве иллюстрации формирующихся индивидуализированных дорогостоящих структур лицензирования и присущей им нестабильности.

(b) Основные модели коллективного лицензирования (добровольное, расширенное и принудительное)

Наиболее распространенной формой коллективного лицензирования служит добровольное коллективное управление. ОКУ получает по договору полномочия по управлению определенными правами от своих членов; она проводит переговоры и заключает лицензии в рамках своих полномочий; полученные доходы распределяются между ее членами. В большинстве юрисдикций с развитой системой авторского права функционируют добровольные ОКУ, однако управляемые ими права и категории произведений существенно различаются. Наиболее универсальным примером выступают общества по управлению правами на публичное исполнение музыки. В Китае действуют либо находятся в процессе выхода на полноформатное функционирование пять ОКУ: Music Copyright Society of China (Китайское общество музыкальных авторских прав), China Audio-Video Copyright Association (Китайская ассоциация авторских прав на аудиовизуальные произведения), China Written Works Copyright Society (Китайское общество авторских прав на письменные произведения), Images Copyright Society of China (Китайское общество авторских прав на изображения) и China Film Copyright Association (Китайская ассоциация кинематографических авторских прав)[60]. ОКУ публично заявили о готовности выдавать

компанией Suno сервиса поиска концертов Songkick. Еще один разработчик, KLAY Vision, заключил соглашения с крупнейшими звукозаписывающими компаниями и их издательскими подразделениями. О сотрудничестве в сфере стриминга компания Spotify совместно с Sony Music Group, Universal Music Group, Warner Music Group, Merlin и Believe объявила в октябре 2025 г.

[58] Кампания Human Artistry Campaign, созданная в 2023 г. коалицией из более чем 150 организаций, включая ведущие объединения музыкальной индустрии и группы представителей авторов, провозглашает семь основных принципов <https://www.humanartistrycampaign.com>, дата обращения: 25 августа 2026 г.; Принципы создания музыки с использованием ИИ (Principles for Music Creation with AI), представленные Roland Corporation и Universal Music Group и поддержанные более чем пятьюдесятью организациями <https://aiformusic.info>, дата обращения: 25 августа 2026 г. Обе инициативы рассматриваются и анализируются в контексте более широкой концепции ответственного проектирования в публикации Де Берардиниса и др. «RAIM: Operationalising Trustworthy AI in Music Generation», цитируемой в части I, сноска 17, настоящего исследования.

[59] Договоренность между The Walt Disney Company и OpenAI, о которой было объявлено 11 декабря 2025 г., касалась более чем 200 персонажей и планируемых инвестиций в акционерный капитал в размере 1 млрд долл. США; как сообщалось, от нее отказались после отзыва компанией OpenAI своего продукта для генерации видео Sora 24 марта 2026 г., при этом инвестиции осуществлены не были.

[60] Первые три обычно сокращенно именуются MCSC, AVCA и CWWCS; Images Copyright Society of China и China Film Copyright Association — две организации, пока не действующие в полном масштабе.

лицензии для обучения моделей ИИ, однако на дату подготовки настоящего исследования ни одного крупного соглашения о коллективном лицензировании для обучения моделей ИИ в Китае заключено не было[61]. Структурная причина этого в контексте Китая состоит в том, что полномочия ОКУ ограничены правами ее членов, и разработчик продолжает нести остаточный риск в отношении произведений, первоначальные авторы которых членами ОКУ не состоят.

Схожие особенности можно наблюдать в Республике Корея на примере Korea Music Copyright Association (Корейская ассоциация авторских прав на музыку) и аналогичных организаций. Ограниченный характер репертуара, охватываемого коллективным управлением, представляет собой, по-видимому, лишь часть объяснения причин незаключения лицензий. В условиях действующих регуляторных режимов у разработчиков может быть мало стимулов стремиться к добровольному получению лицензий. Два других фактора, судя по всему, объясняют, почему сделки тем не менее заключаются в Соединенных Штатах. Во-первых, структура рынка позволяет небольшому числу крупных правообладателей и поставщиков контента вести прямые переговоры с разработчиками. Во-вторых, риск судебного разбирательства, включая возможность взыскания предусмотренной законом компенсации, создает стимул для достижения соглашения. Лицензионные соглашения, заключенные в рамках урегулирования споров и описанные в разделе B.2, служат тому наглядными примерами. Поэтому при объяснении причин, по которым лицензирование еще не набрало обороты в Китае и Республике Корея, необходимо сопоставлять эти рыночные и правовые факторы с ограничениями полномочий по коллективному управлению. На дату подготовки настоящего исследования Ministry of Culture, Sports and Tourism (Министерство культуры, спорта и туризма Республики Корея) и Korea Copyright Commission (Корейская комиссия по авторскому праву) в Республике Корея вели совместную работу с целью совершенствования регуляторной базы.

С институциональной точки зрения Соединенные Штаты располагают разветвленной сетью отраслевых организаций коллективного управления. Организации по защите прав исполнителей (ASCAP, BMI, SESAC, Global Music Rights и AllTrack) управляют правами на публичное исполнение музыкальных произведений; Mechanical Licensing Collective (Организация по механическому лицензированию), учрежденная в соответствии с законом «О модернизации музыки» 2018 года (Music Modernization Act), осуществляет выдачу комплексных лицензий на механическое воспроизведение для цифрового использования музыкальных произведений;[62] SoundExchange управляет предусмотренными законом лицензионными отчислениями за цифровое исполнение звукозаписей по статье 114 закона «Об авторском праве»; Copyright Clearance Center, Inc. (CCC) (Центр по очистке авторских прав) управляет правами на воспроизведение опубликованных текстовых материалов; a Artists Rights Society (ARS) (Общество прав художников) управляет определенными правами на воспроизведение произведений изобразительного искусства. Каждая организация действует в рамках своего сектора и на условиях, согласованных с представляемыми ею правообладателями. В ответ на появление практики обучения моделей ИИ как формы использования охраняемых произведений CCC расширил действие своей ежегодной корпоративной лицензии на

[61] Ведущие китайские ОКУ публично выразили готовность лицензировать использование произведений для обучения моделей ИИ, однако к настоящему времени ни одного значимого лицензионного соглашения заключено не было.

[62] Orrin G. Hatch–Bob Goodlatte Music Modernization Act, Pub. L. No. 115-264, 132 Stat. 3676 (2018 г.), кодифицирован в 17 U.S.C. §§ 115 и сл. Mechanical Licensing Collective назначена в соответствии с 17 U.S.C. § 115(d)(3).

определенные виды использования в системах ИИ и объявил о планах введения нескольких дополнительных специальных лицензий для использования в ИИ[63].

За пределами Китая и Соединенных Штатов за последние два года ОКУ и другие посредники предприняли ряд практических шагов по включению обучения моделей ИИ в сферу своих полномочий по лицензированию, причем о некоторых из них было объявлено лишь в конце 2025 года. В Швеции организация STIM в сентябре 2025 года запустила механизм коллективного лицензирования с участием музыкальной ИИ-компании Songfox и поставщика решений по атрибуции Sureel, технология которого позволяет отслеживать результаты работы ИИ вплоть до оказавших на них влияние произведений[64]. В Норвегии организация Kopinor в декабре 2025 года заключила соглашение, по которому Национальная библиотека Норвегии при финансовой поддержке норвежского государства получила право использовать газетный контент для обучения моделей норвежского языка[65]. В Дании организация Danish Press Publications' Collective Management Organization (Организация коллективного управления издателей датской прессы) также в декабре 2025 года заключила партнерское соглашение о лицензировании на основе предварительного согласия с ProRata AI в целях использования контента ее членов в вопросно-ответной поисковой системе на базе извлечения данных, в то время как еще одна датская организация, Tekst&Node, разрабатывает модель лицензирования в рамках датской системы расширенного коллективного лицензирования[66]. В Нидерландах ассоциация частных новостных издателей совместно с новостным агентством ANP с июля 2025 года сотрудничает с государственными организациями в целях разработки национальной языковой модели GPT-NL[67]. В Испании организация CEDRO присоединилась к

---

[63] См. www.copyright.com/ai.

[64] STIM, «STIM launches world's first collective AI music licence» (9 сентября 2025 г.), сообщение CISAC, https://www.cisac.org/Newsroom/society-news/stim-launches-worlds-first-collective-ai-music-licence. Лицензия на основе предварительного согласия охватывает как обучение моделей, так и последующее использование результатов работы ИИ; компенсация сочетает авансовый платеж с долей от дохода разработчика, а поставщик решений по атрибуции Sureel отслеживает результаты работы вплоть до оказавших на них влияние произведений.

[65] «Kopinor signs historic agreement on newspaper content for AI in Norway» (19 декабря 2025 г.), сообщение IFRRO, https://ifrro.org/page/article-detail/ifrro-member-kopinor-signs-historic-agreement-on-newspaper-content-for-ai-in-norway/. Заключенное Kopinor от имени Norwegian Media Businesses' Association (Норвежская ассоциация медиапредприятий) соглашение предоставляет Национальной библиотеке Норвегии доступ к газетным материалам старше одного года для обучения моделей норвежского языка и финансируется норвежским государством в объеме 45 млн норвежских крон в год. Оно отличается от более раннего соглашения «Книжная полка» (Bookshelf) о цифровом распространении книг.

[66] ProRata AI и Danish Press Publications' Collective Management Organisation (DPCMO), «ProRata Partners with Danish Publishers Group DPCMO» (10 декабря 2025 г.), https://www.prnewswire.com/news-releases/prorata-partners-with-danish-publishers-group-dpcmo-to-launch-the-first-decentralized-sovereign-ai-answer-engine-302637182.html. DPCMO, заявляющая о представительстве 99% датской новостной индустрии, заключила основанное на предварительном согласии партнерское соглашение о коллективном лицензировании, позволяющее использовать контент ее участников в поисково-ответной системе ProRata на условиях распределения доходов. Еще одна датская организация (Tekst&Node) разрабатывает модель лицензирования в области ИИ в рамках датской системы расширенного коллективного лицензирования, позволяющей правообладателям лицензировать отдельные виды интеллектуального анализа текста и данных и использования литературных произведений системами ИИ при условии возможности отказа от использования и одобрения Министерством культуры. См. Фогед Т. Первое одобрение расширенной коллективной лицензии для интеллектуального анализа текста и данных и использования произведений системами ИИ (Foged Terese. [DK] First extended collective licence approval for text and data mining and the use of works by AI), IRIS Legal Observations of the European Audiovisual Observatory, 2026 г., <https://merlin.obs.coe.int/article/10505>, дата обращения: 3 июля 2026 г.

[67] NDP Nieuwsmedia and TNO, «Large dataset news organizations for Dutch AI language model GPT-NL» (17 июля 2025 г.), https://www.ndpnieuwsmedia.nl/2025/07/17/large-dataset-news-organizations-for-dutch-ai-language-model-gpt-nl/. Члены ассоциации голландских частных новостных издателей NDP Nieuwsmedia

инициативе Really Simple Licensing с целью разработки модели лицензирования для систем ИИ на испанском рынке[68]. В Соединенном Королевстве Copyright Licensing Agency совместно с Publishers' Licensing Services и Authors' Licensing and Collecting Society разрабатывает специальную комплексную лицензию для генеративного ИИ, охватывающую обучение, тонкую настройку и RAG[69]. В других странах организации по защите прав на воспроизведение включили права на повторное использование произведений в системах ИИ в свои действующие лицензии на цифровое использование объектов авторского права либо разрабатывают сопоставимые решения: в Японии JAC добавила такие права в свою лицензию на цифровое использование авторских прав в апреле 2025 г.[70]; Copyright Agency в Австралии разрабатывает аналогичные решения[71]; а в Германии VG Wort распространила действие своей цифровой лицензии на повторное использование в системах ИИ, в то время как общество по сбору авторских отчислений GEMA отдельно предложило специализированную модель, которая на дату подготовки настоящего исследования еще не привела к заключению соглашения[72]. Эти процессы развиваются неравномерно и в ряде случаев носят экспериментальный характер, однако они

совместно с новостным агентством ANP предоставляют части своих архивов для обучения некоммерческой модели нидерландского языка GPT-NL, разрабатываемой TNO, NFI и SURF; предоставившие материалы издатели должны получить вознаграждение после выпуска модели.

[68] CEDRO, пресс-релиз о соглашении о сотрудничестве с RSL Collective (январь 2026 г.), https://www.cedro.org/sala-de-prensa/noticias. CEDRO, испанская организация по правам на репрографическое воспроизведение, присоединилась к RSL Collective в качестве ее партнера в Испании и участвует в работе технического комитета RSL, адаптируя машиночитаемый стандарт лицензирования к испанскому рынку, чтобы ее члены могли определять условия использования своих произведений при обучении моделей ИИ. 16 декабря 2025 г. CEDRO отдельно опубликовала рекомендации по передаче прав для обучения моделей ИИ.

[69] Copyright Licensing Agency, «CLA announces development of Generative AI Training Licence» (2025 г.), https://cla.co.uk/development-of-cla-generative-ai-licence/; Publishers' Licensing Services, «PLS and ALCS agree to development of pioneering CLA generative AI licence» (апрель 2025 г.), https://www.pls.org.uk/news-events-policy/news/pls-and-alcs-agree-to-development-of-pioneering-cla-generative-ai-licence/. CLA совместно с PLS и ALCS разрабатывает коллективную лицензию на основе предварительного согласия на использование текстов в генеративном ИИ, охватывающую обучение, тонкую настройку и генерацию с расширенным поиском; правообладатели участвуют в ней путем регистрации своих произведений.

[70] Japan Academic Association for Copyright Clearance (JAC) (Японская академическая ассоциация по очистке авторских прав) и RightsDirect Japan, «JAC and RightsDirect Japan Announce the Availability of AI Re-Use Rights for Digital Copyright License» (9 апреля 2025 г.), https://www.businesswire.com/news/home/20250409032666/en/. JAC добавила права на повторное использование в системах ИИ в свою лицензию Digital Copyright License — общекорпоративную лицензию для внутреннего использования компаниями со штаб-квартирой в Японии, опирающуюся на репертуар произведений издательств и научных обществ со всего мира.

[71] Copyright Agency, «AI and Copyright in Australia» и сопутствующие вопросы и ответы, https://www.copyright.com.au/membership/ai-and-copyright-in-australia/. Copyright Agency, входящее в состав Справочной группы правительства Австралии по вопросам авторского права и ИИ, заявляет о разработке лицензионных решений для использования охраняемых материалов в связи с генеративным ИИ в Австралии.

[72] Относительно VG Wort: действие лицензии Digital Copyright License, администрируемой через RightsDirect (дочернюю компанию CCC), было распространено на внутреннее повторное использование в системах ИИ; см. https://www.rightsdirect.com/license/vg-wort-digital-license-2/. Общество по сбору авторских отчислений GEMA предложило модель лицензирования для генеративного ИИ в 2024 году; на дату подготовки настоящего исследования соглашений заключено не было, и GEMA параллельно инициировала судебные разбирательства по факту нарушения прав против разработчиков ИИ (см. дело *GEMA v. OpenAI*, *выше*). См. GEMA, «AI and music: GEMA calls for music creators to share in AI providers' revenues and presents first licensing model» (26 сентября 2024 г.) <https://www.cisac.org/Newsroom/society-news/ai-and-music-gema-calls-music-creators-share-ai-providers-revenues-and>, дата обращения: 3 июля 2026 г.; GEMA, «Suno AI and Open AI: GEMA sues for fair compensation» (20 января 2025 г.) <https://www.gema.de/en/news/ai-and-music/ai-lawsuit>, дата обращения: 3 июля 2026 г.

свидетельствуют о том, что механизм коллективного лицензирования активно проверяется на практике, а не только служит предметом теоретических дискуссий.

Расширенное коллективное лицензирование (ECL) представляет собой модель лицензирования, возникшую в странах Северной Европы в 1960-х годах и воспринятую рядом других государств. Оно решает проблему остаточного риска за счет распространения действия лицензии ОКУ на правообладателей, не являющихся ее членами, в пределах определенного репертуара, и предусматривает право на отказ от использования. В европейских стратегических дискуссиях по вопросам обучения моделей ИИ эта модель упоминалась как способ очистки прав на крупные текстовые и визуальные корпуса, необходимые разработчику, индивидуальное лицензирование которых практически неосуществимо. В исследовании Европейского парламента 2025 года, посвященном генеративному ИИ и авторскому праву, оно упоминалось прямо[73]. В Соединенных Штатах US Copyright Office (Бюро авторского права США) в третьей части своего отчета об авторском праве и искусственном интеллекте определило расширенное коллективное лицензирование в качестве возможного целевого промежуточного варианта для обучения моделей ИИ, отличного от принудительного лицензирования[74]. То же предложение с отдельными вариациями находит отражение в научной литературе, посвященной вознаграждению: В исследовании профессора Зенфтлебена о генеративном ИИ и вознаграждении авторов предлагается система, в рамках которой организации коллективного управления собирали бы вознаграждение за использование сервисов генеративного ИИ, чьи результаты работы замещают произведения конкретных авторов или конкурируют с ними, и распределяли бы полученные средства среди своих членов[75].

В дополнение к уже приведенным примерам эту возможность иллюстрируют еще два примера того, как существующее добровольное коллективное управление можно адаптировать для обучения моделей ИИ в рамках действующих институциональных структур, не дожидаясь принятия нового законодательства. В Бразилии режим публичного исполнения музыки централизован в рамках единой структуры — ECAD, опирающейся на законодательную и судебную базу для объединения функций по сбору и распределению отчислений за публичное исполнение музыкальных произведений и фонограмм как в отношении авторских, так и смежных прав[76]. В рамках действующего закона «Об авторском праве» и при отсутствии законодательных изменений существовала бы юридическая возможность осуществлять специальное право в отношении использования произведений и фонограмм для обучения моделей ИИ на коллективной основе по модели, аналогичной действующей сегодня для права на публичное исполнение музыки. Для этого потребовалось бы вступление правообладателей в ОКУ и закрепление в ее уставе положений о представлении интересов правообладателя в отношении иных

---

[73] Европейский парламент, Генеральный директорат по внутренней политике, исследование «Генеративный ИИ и авторское право: обучение, создание, регулирование» (Generative AI and Copyright: Training, Creation, Regulation), IUST_STU(2025)774095_EN (2025 г.), где установленное законом лицензирование рекомендовано в качестве оптимального механизма для обучения моделей ИИ в ЕС.

[74] US Copyright Office (Бюро авторского права США), *Copyright and Artificial Intelligence, Part 3: Generative AI Training* (2025 г.).

[75] Зенфтлебен М. Р. Ф. Генеративный ИИ и вознаграждение авторов (Senftleben M.R.F. Generative AI and Author Remuneration), 54 IIC 1535 (2023 г.).

[76] О структуре бразильской системы коллективного управления и централизации сбора отчислений за публичное исполнение в ECAD см. закон Бразилии «Об авторском праве» (Закон № 9.610/1998), статьи 97–100, и Декрет № 9.574/2018. Конституционность этой модели подтверждена Supremo Tribunal Federal (Федеральный верховный суд) по делу ADI 2.054 (2003 г.) и в совместном решении по делам ADI 5062 и 5065 (2016 г.).

авторских прав, а не только прав на публичное исполнение музыки; кроме того, потребовалось бы предварительное разрешение от Ministry of Culture (Министерство культуры Бразилии) — федерального органа исполнительной власти, осуществляющего надзор за деятельностью ОКУ. Статья 97 закона Бразилии «Об авторском праве» позволяет ОКУ при условии получения необходимых разрешений расширять сферу своей деятельности на лицензирование прав, затрагиваемых обучением и внедрением модели (права на воспроизведение, сообщения для всеобщего сведения или доведения до всеобщего сведения, а согласно позиции Ministry of Culture (Министерство культуры Бразилии) — и права на переработку на этапе формирования выдачи). Требование о едином сборе, однако, распространяется только на права на публичное исполнение музыки, в связи с чем даже в случае централизации в ECAD или ее аффилированных лицах полномочий по представлению права на обучение моделей ИИ это не исключало бы создание других коллективных организаций с аналогичной целью.

В Кении регистрацию, функционирование и надзор за деятельностью ОКУ регулирует Положение «Об авторском праве (коллективное управление)» 2020 года (Copyright (Collective Management) Regulations 2020).[77] Эта нормативная база носит общий характер и специально под компьютерный анализ данных не адаптировалась, однако она формирует институциональную основу, на которой может выстраиваться добровольное коллективное лицензирование для обучения моделей ИИ. Проект закона «Об авторском праве и смежных правах» 2026 года (Draft Copyright and Related Rights Bill 2026), опубликованный для общественного обсуждения Kenya Copyright Board (Совет по авторскому праву Кении) в марте 2026 года со сроком представления замечаний до конца того же месяца, содержит статью (статья 34), разрешающую анализ текста и данных для некоммерческих научных исследований; вместе с тем законопроект не дает четкого законодательного определения понятия «некоммерческий», не вводит механизм отказа от использования для правообладателей и не содержит специальных норм ни об использовании охраняемых авторским правом произведений при обучении моделей ИИ, ни о способности произведений, созданных ИИ, охраняться авторским правом. Ситуация в Кении показательна для более широкой тенденции в юрисдикциях, где адаптация системы авторского права осуществляется поэтапно: существующая инфраструктура ОКУ выступает возможным инструментом коллективного лицензирования для обучения моделей ИИ, однако такая адаптация требует конкретных законодательных или договорных мер, которые пока не предприняты.

3. Лицензирование первичных языковых данных на основе интересов сообщества и ответственного управления данными

Описанные выше механизмы лицензирования применяются к обучающим данным, которые к моменту обращения к ним разработчика уже зафиксированы и опубликованы. Особая категория обучающих данных формируется принципиально иным образом. Модели распознавания речи и языковые модели для большинства африканских языков невозможно построить на базе текстовых корпусов масштаба Интернета в силу отсутствия достаточного объема таких текстов. Вместо этого данные собираются как первичный материал: аудиозаписи и транскрипции, получаемые от носителей языка в ходе полевых исследований, коллективных сессий звукозаписи и в рамках партнерств с вещательными организациями. Авторское право возникает в

[77] Положение «Об авторском праве (коллективное управление)» 2020 г. (The Copyright (Collective Management) Regulations 2020, Кения). Инфраструктура лицензирования закреплена в статьях 33 и 34 и части VII закона «Об авторском праве» (Cap 130).

момент фиксации, а не в момент сбора данных краулером. Запись охраняется в качестве звукозаписи (фонограммы), транскрипция — в качестве литературного произведения, а говорящий может обладать правами исполнителей[78]. Поэтому вопросы получения разрешения и компенсации возникают в момент сбора, задолго до того, как может встать вопрос об интеллектуальном анализе текста и данных.

Эта категория данных затрагивает также дополнительный пласт интеллектуальной собственности. Записи устного народного творчества, сказок и пословиц одновременно представляют собой произведения, охраняемые авторским правом и принадлежащие зафиксировавшему их лицу, и выражения культуры, коллективными правами на которые обладают сообщества в соответствии с законодательством sui generis. В Кении закон «Об охране традиционных знаний и выражений культуры» № 33 от 2016 года (Protection of Traditional Knowledge and Cultural Expressions Act), принятый во исполнение статей 11, 40 и 69(1)(c) Конституции Кении 2010 года, признает сообщества правообладателями в отношении их выражений культуры, требует получения предварительного осознанного согласия и заключения соглашений с уполномоченными пользователями для их использования, а также предусматривает выплату компенсации сообществам; статья 11(3)(a) Конституции обязывает Парламент обеспечить получение сообществами лицензионных отчислений за использование их культуры и культурного наследия[79]. Когда языковые данные такого рода попадают в обучающий корпус, цепочка перехода авторских прав и права сообщества действуют параллельно, и лицензия от лица, зафиксировавшего запись, сама по себе не решает вопрос о закрепленном в законе статусе сообщества.

В ответ на это сложилась особая практика лицензирования. Стандартные открытые лицензии, и в особенности отказ от прав в пользу общественного достояния, включают подобные данные в процессы обучения, не предоставляя предоставившему их сообществу полномочий по контролю или экономической отдачи. Лицензии, ориентированные на интересы сообщества, меняют такое положение дел, сохраняя при этом доступность данных. Лицензия на открытые данные Nwulite Obodo (Nwulite Obodo Open Data Licence) устанавливает дифференцированные условия использования в зависимости от географического контекста пользователя и уровня развития его региона, а также включает обязательства по совместному использованию выгод; она применялась к наборам речевых данных африканских языков, включая набор речевых данных DhoNam Dholuo, разработанный совместно с Университетом Масено[80]. Лицензия Esethu (Esethu License), разработанная для набора речевых данных Vuk'uzenzele isiXhosa, разрешает бесплатное использование в научных целях и коммерческое использование африканскими организациями, предусматривая взимание лицензионных платежей с иных коммерческих пользователей, при этом полученные доходы реинвестируются в создание новых

---

[78] Copyright Act (Cap. 130), Laws of Kenya, статья 2 (содержащая определения, в числе прочего, понятий «sound recording» и «literary work») и нормы о правах исполнителей. О правах исполнителей на международном уровне см. статьи 10 и 14 ДИФ, рассмотренные выше в разделе B.1.

[79] Закон «Об охране традиционных знаний и выражений культуры» № 33 от 2016 г. (Protection of Traditional Knowledge and Cultural Expressions Act, Кения); Конституция Кении 2010 г., статьи 11, 40(5) и 69(1)(c).

[80] Nwulite Obodo Open Data Licence, Version 1.0 (Data Science Law Lab, University of Pretoria, 2025 г.) <https://datasciencelawlab.africa/nwulite-obodo-open-data-license/>, дата обращения: 31 июля 2026 г.; Окори С. И., Омино М. Устранение несправедливой открытости в лицензиях на совместное использование африканских данных и наборов данных посредством открытой лицензии Nwulite Obodo (Okorie C.I. and Omino M. Addressing Inequitable Openness in Licences for Sharing African Data and Datasets Through the Nwulite Obodo Open Data Licence) (2025 г.) 7(3) Law, Technology and Humans 94. Набор речевых данных DhoNam Dholuo (Maseno University Centre for African AI) опубликован под этой лицензией: <https://mozilladatacollective.com/datasets/cmjepxo6t08nmmk07iauvua6v>, дата обращения: 31 июля 2026 г.

языковых данных[81]. Лицензия Te Hiku Media Kaitiakitanga (Kaitiakitanga License), регулирующая данные языка маори, отражает тот же подход в другом регионе[82]. Эти инструменты образуют третий механизм — наряду с индивидуальным и коллективным лицензированием, — посредством которого организуются выдача разрешений и выплата компенсаций за использование первичных языковых данных при обучении моделей ИИ, и иллюстрируют, как национальная правовая база в отношении коллективных культурных ценностей может служить основой для договорной практики. В Кении этот уровень правового регулирования будет также взаимодействовать с исключением для анализа текста и данных, предложенным в проекте закона «Об авторском праве и смежных правах» 2026 года, рассмотренном в предыдущем разделе.

4. Предусмотренные законом исключения в отношении интеллектуального анализа текста и данных и общие нормы

В законодательстве Европейского союза об авторском праве содержатся прямо сформулированные исключения в отношении интеллектуального анализа текста и данных, применимые к операциям, используемым при обучении моделей ИИ, хотя их применение и достаточность в этом контексте вызывают споры. Статьи 3 и 4 Директивы (ЕС) 2019/790 об авторском праве и смежных правах на едином цифровом рынке (Директивы CDSM) устанавливают два обязательных исключения в отношении интеллектуального анализа текста и данных. Статья 3 закрепляет исключение для исследовательских целей в пользу исследовательских организаций и учреждений культурного наследия без права отказа от использования. Статья 4 устанавливает общее исключение, применимое к любым пользователям, при условии реализации правообладателем механизма отказа от использования «надлежащим образом, в частности с помощью машиночитаемых средств в случае общедоступного контента, размещенного в сети Интернет»[83]. Регламент ЕС об искусственном интеллекте отсылает к этой системе норм, обязывая поставщиков моделей ИИ общего назначения внедрять политику соблюдения законодательства Союза об авторском праве, выявлять и соблюдать заявления о резервировании прав, сделанные в соответствии со статьей 4(3) Директивы CDSM, а также публиковать достаточно подробное резюме о контенте, использованном для обучения моделей[84].

Два судебных решения в Германии коснулись применения этих исключений к обучению моделей ИИ. В сентябре 2024 года по делу Kneschke v. LAION e.V. Земельный суд Гамбурга (Landgericht Hamburg) постановил, что использование

---

[81] Lelapa AI, Way With Words и Data Science for Social Impact (University of Pretoria), «Концепция Esethu: новое видение устойчивого управления и кураторства наборов данных для языков с ограниченными ресурсами» (The Esethu Framework: Reimagining Sustainable Dataset Governance and Curation for Low-Resource Languages), Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (2025 г.) <https://aclanthology.org/2025.acl-long.1487/>, дата обращения: 31 июля 2026 г.

[82] Te Hiku Media, Kaitiakitanga License <https://tehiku.nz/te-hiku-tech/te-hiku-dev-korero/25141/data-sovereignty-and-the-kaitiakitanga-license>, дата обращения: 31 июля 2026 г.

[83] Директива (ЕС) 2019/790 Европейского парламента и Совета от 17 апреля 2019 г. об авторском праве и смежных правах на едином цифровом рынке, OJ L 130/92 («Директива CDSM»), статья 3 (обязательное исключение в отношении интеллектуального анализа текста и данных для целей научных исследований исследовательскими организациями и учреждениями культурного наследия) и статья 4 (общее исключение с возможностью отказа от использования в отношении интеллектуального анализа текста и данных из правомерно доступных произведений).

[84] Регламент (ЕС) 2024/1689 Европейского парламента и Совета от 13 июня 2024 г., устанавливающий гармонизированные правила в отношении искусственного интеллекта (Регламент ЕС об искусственном интеллекте), OJ L, 2024/1689, 12.7.2024, статья 53(1)(c) и (d), а также пункты 105 и 106 преамбулы.

фотографии при формировании набора данных LAION-5B, применяемого для обучения моделей генерации изображений, подпадало под имплементированное в германское право исключение статьи 3 CDSM в отношении интеллектуального анализа текста и данных для научных исследований. Суд признал LAION научно-исследовательской организацией по смыслу указанной нормы и счел, что спорный компьютерный анализ представлял собой интеллектуальный анализ текста и данных в научных целях. Суд также попутно отметил (obiter dictum), что общее исключение статьи 4 в отношении интеллектуального анализа текста и данных в любом случае не подлежало применению, поскольку заявление о резервировании прав в условиях использования исходной платформы, по мнению суда, являлось машиночитаемым с учетом возможностей современных больших языковых моделей[85]. В суде апелляционной инстанции по делу Kneschke v. LAION e.V. Высший земельный суд Гамбурга (Oberlandesgericht Hamburg) подтвердил данный результат, но представил более подробное толкование требования о машиночитаемости по смыслу статьи 4(3)[86]. Суд апелляционной инстанции постановил, что машиночитаемость требует не только возможности машинного считывания заявления о резервировании, но и способности автоматизированного процесса интерпретировать его таким образом, чтобы контент, охватываемый резервированием, не подвергался обработке. Другими словами, заявление о резервировании прав должно быть одновременно машиночитаемым и пригодным для автоматизированного исполнения. Суд далее указал, что машиночитаемость не представляет собой статичное свойство, а подлежит оценке применительно к технологическим возможностям, существовавшим на момент спорного использования; применительно к фактическим обстоятельствам дела это означало уровень развития техники во второй половине 2021 года. Логика суда апелляционной инстанции оставляет открытым вопрос о том, что именно образует машиночитаемость в 2025 году или позднее, когда модели, способные извлекать смысл из неструктурированного текста на естественном языке, получили широкое распространение[87].

Примерно за месяц до этого по делу GEMA v. OpenAI Земельный суд Мюнхена I (Landgericht München I) пришел к иному выводу по смежному, но самостоятельному вопросу[88]. Германское общество по сбору авторских отчислений GEMA предъявило иск к OpenAI, утверждая, что ChatGPT запомнил тексты девяти известных немецких

---

[85] *Kneschke v. LAION e.V.*, LG Hamburg, 27 сентября 2024 г. (310 O 227/23) (где установлено, что создание набора данных LAION-5B для использования при обучении моделей генерации изображений охватывалось имплементированным в законодательство Германии исключением статьи 3 CDSM для интеллектуального анализа текста и данных в научных целях (§ 60d UrhG); также попутно отмечено, что общее исключение статьи 4 для интеллектуального анализа текста и данных не подлежало бы применению на том основании, что заявление о резервировании прав в условиях исходной платформы являлось машиночитаемым в соответствующем смысле). О рассмотрении в апелляционной инстанции см. дело *Kneschke v. LAION e.V.*, OLG Hamburg, 10 декабря 2025 г. (5 U 104/24), рассматриваемое ниже.

[86] *Kneschke v. LAION e.V.*, OLG Hamburg, 10 декабря 2025 г. (5 U 104/24).

[87] Высший земельный суд Гамбурга указал, что машиночитаемость требует, чтобы заявление о резервировании «могло быть интерпретировано машиной таким образом... чтобы в автоматизированном процессе контент, охватываемый резервированием [прав], не подвергался обработке» (перевод авторов). См. также Келлер П. Второй раунд LAION: машиночитаемый, но все еще непригодный для исполнения — отсутствие прогресса по отказам от использования при анализе текста и данных (Keller P. LAION Round 2: Machine-Readable but Still Not Actionable - The Lack of Progress on TDM Opt-Outs), Kluwer Copyright Blog (17 и 18 декабря 2025 г.), часть 1.

[88] GEMA v. OpenAI, LG München I, 11 ноября 2025 г. (42 O 14139/24), Endurteil. Истцом выступает германское общество по сбору авторских отчислений GEMA; ответчиками — компании OpenAI OpCo LLC и OpenAI Ireland Limited. Суд вынес судебный запрет, обязал предоставить информацию и присудил возмещение убытков.

песен и может быть побужден к их воспроизведению. Мюнхенский суд квалифицировал как загрузку текстов песен в обучающий корпус, так и их последующее запоминание в модели в качестве актов воспроизведения по смыслу имплементированной в право Германии Директивы InfoSoc; он отклонил применение общего исключения в отношении интеллектуального анализа текста и данных в § 44b закона Германии об авторском праве, основываясь на том, что генеративные модели запоминают тексты песен на постоянной основе, а это выходит за пределы анализа текстовой информации и структурных закономерностей, разрешать который и было призвано данное исключение; а дословно воспроизведенные моделью результаты суд расценил как доказательство того, что тексты зафиксированы внутри модели способом, сопоставимым с цифровой библиотекой. Суд вынес судебный запрет, обязал предоставить информацию и присудил возмещение убытков[89]. Решение обжалуется в суде апелляционной инстанции. В определенных кругах это постановление восприняли как первое крупное европейское судебное решение по вопросам обучения моделей ИИ в пользу правообладателей; другие подвергли его критике за сведение к единому акту воспроизведения того, что с технологической точки зрения представляет собой последовательность самостоятельных операций, а также за неверное применение исключения для интеллектуального анализа текста и данных из-за смешения доказательств запоминания на стадии выдачи с техническими реалиями процесса обучения[90].

Помимо судебной практики, в научных кругах ведется спор о том, применимо ли исключение статьи 4 Директивы CDSM к обучению моделей ИИ вообще. В юридической литературе широко распространена точка зрения, согласно которой обучение моделей ИИ представляет собой форму анализа данных, охватываемую статьями 3 и 4 Директивы CDSM. Сторонники этой позиции нередко ссылаются на то, что к указанным статьям отсылает Регламент ЕС об искусственном интеллекте. Противоположную позицию убедительно обосновал ряд европейских ученых, включая профессора Тима Дорниса, который в статье, опубликованной в журнале European Intellectual Property Review, и в более подробном юридическом заключении, подготовленном в соавторстве со специалистом в области информатики Штобером, доказывает, что детальный анализ технологических основ обучения моделей генеративного ИИ свидетельствует: выполняемые операции выходят за рамки того, что входит в понятие «исключение в отношении интеллектуального анализа текста и данных». Логика этой аргументации в общих чертах сводится к следующему: исключение для интеллектуального анализа текста и данных разрешает извлечение из произведений смысловой, фактической и статистической информации; обучение же модели генеративного ИИ, напротив, извлекает и кодирует также синтаксические и выразительные особенности таким образом, что они затем используются для создания результатов, замещающих оригинальные произведения; при этом

---

[89] Там же. Суд квалифицировал как загрузку текстов песен в обучающий корпус, так и их последующее запоминание в модели в качестве актов воспроизведения по смыслу § 16 UrhG, имплементирующего статью 2 Директивы 2001/29/ЕС.

[90] Критический комментарий по данному вопросу см. в публикации: Фрозио Дж. Авторское право в формальдегиде: как дело GEMA против OpenAI замораживает доктрину и сдерживает развитие ИИ (Frosio G. How GEMA v OpenAI Freezes Doctrine and Chills AI), Kluwer Copyright Blog (10 и 11 декабря 2025 г.), части 1 и 2, где утверждается, что мюнхенский суд смешивает доказательства запоминания на стадии выдачи с техническими реалиями обучения, и что более обоснованно рассматривать правомерные промежуточные технические копии как подпадающие под статью 5(1) Директивы InfoSoc либо под исключения для интеллектуального анализа текста и данных.

возникающее ущемление законных интересов правообладателей выходит за пределы дозволенного данным исключением[91].

Профессор Элеонора Розати выдвинула схожий тезис, но исходя из иных предпосылок. Она отстаивает позицию, согласно которой интеллектуальный анализ текста и данных может входить в процессы обучения моделей ИИ, но синонимом обучения моделей ИИ не выступает и далеко не исчерпывает всего того, что такое обучение за собой влечет; что ограничение европейской дискуссии статьей 4 Директивы CDSM чрезмерно сужает проблему с учетом множества охватываемых авторским правом действий, которые способно затрагивать обучение моделей ИИ; и что в любом случае трехступенчатый тест, закрепленный в статье 7(2) Директивы CDSM и статье 5(5) Директивы InfoSoc, может служить основанием для выведения обучения моделей ИИ из-под действия исключения, поскольку такое использование вступает в противоречие с обычным использованием произведений и необоснованно ущемляет законные интересы правообладателей. Если эта позиция верна, лицензирование остается обязательным элементом анализа[92]. Профессор Маттиас Ляйстнер отдельно указал, что механизм отказа от использования, предусмотренный статьей 4(3), скорее всего, станет ключевой проблемой авторского права ЕС и в настоящее время на практике неосуществим; эта точка зрения также вошла в европейские политические дискуссии[93].

Профессор Зенфтлебен утверждает, напротив, что статья 4 не только применима в контексте обучения моделей ИИ, но и может иметь непосредственное значение для определения того, что образует обычное использование произведения в контексте ИИ, что существенно при применении трехступенчатого теста[94].

Европейская комиссия в своем толковании Регламента ЕС об искусственном интеллекте исходит из применимости статьи 4 Директивы CDSM к обучению моделей ИИ в том смысле, что ее подход сосредоточен на внедрении механизма отказа от использования, а не на исключении обучения моделей ИИ из сферы действия данного исключения. Инфраструктура отказа от использования рассматривается в части III

---

[91] Дорнис Т. В. Обучение генеративного ИИ не представляет собой интеллектуальный анализ текста и данных (Dornis T.W. The Training of Generative AI Is Not Text and Data Mining), 47 European Intellectual Property Review 65 (2025 г.). См. также: Дорнис Т. В., Штобер С. Авторское право и обучение моделей генеративного ИИ — технологические и правовые основы (Dornis T.W. and Stober S. *Urheberrecht und Training generativer KI-Modelle - technologische und juristische Grundlagen*), Initiative Urheberrecht (Nomos, сентябрь 2024 г.).

[92] Розати Э. Выступает ли интеллектуальный анализ текста и данных синонимом обучения ИИ? (Rosati E. Is text and data mining synonymous with AI training?), 19 Journal of Intellectual Property Law and Practice 851 (2024 г.); Розати Э. Решение по делу LAION в Германии: проблематичное понимание сферы действия авторско-правовых исключений для интеллектуального анализа текста и данных и переход от TDM к обучению ИИ (Rosati E. The German LAION decision: A problematic understanding of the scope of the TDM copyright exceptions and the transition from TDM to AI training), IPKat (7 октября 2024 г.); Розати Э. Исключения в авторском праве и возражения о добросовестном использовании при обучении ИИ для целей «исследований» и «обучения», или неизбежный горизонт лицензирования (Rosati E. Copyright Exceptions and Fair Use Defences for AI Training Done for ‘Research’ and ‘Learning’, or the Inescapable Licensing Horizon), SSRN Working Paper No. 5331405 (2025 г.).

[93] Ляйстнер М. Интеллектуальный анализ текста и данных и обучение ИИ в Европейском союзе (Leistner M. TDM und KI-Training in der Europäischen Union), GRUR 2024, 1665 (где утверждается, что механизм отказа от использования, по всей видимости, станет центральной проблемой права авторского права ЕС и в настоящее время на практике неосуществим).

[94] Зенфтлебен М. Интеллектуальный анализ текста и данных, генеративный ИИ и трехступенчатый тест в авторском праве (Senftleben, M. Text and Data Mining, Generative AI, and the Copyright Three-Step Test). IIC 57, 67, at 79 (2026 г.). https://doi.org/10.1007/s40319-026-01680-2

настоящего исследования в контексте прозрачности и обеспечения соблюдения прав[95].

Законодательный орган Республики Корея рассматривал, но так и не принял исключение в отношении интеллектуального анализа текста и данных; соответствующие законопроекты находятся на рассмотрении Национального собрания с 2021 года[96]. При отсутствии специальной законодательной нормы вопрос обучения моделей подлежит оценке на основании общей нормы о добросовестном использовании, закрепленной в статье 35-5 закона «Об авторском праве», которая была сформулирована по образцу статьи 107 закона США об авторском праве и введена по итогам заключения соглашения о свободной торговле между Республикой Корея и Соединенными Штатами. Внутригосударственные дискуссии во многом повторяют соображения, рассмотренные в настоящем разделе. Сторонники законодательного закрепления исключения указывают на правовую определенность, которую обеспечили бы прямые формулировки закона; оппоненты возражают, что законодательный текст порождает собственные трудности толкования, как показывает рассмотренный выше судебный процесс в Германии по вопросу машиночитаемости, что сформулированные Верховным судом Республики Корея в 2024 году критерии применения статьи 35-5 наполняют общую норму практически применимым содержанием, а широкое исключение способно принести непропорциональную выгоду зарубежным разработчикам при одновременном перераспределении создаваемой стоимости не в пользу правообладателей контента на корейском языке. [97]

Применительно к обучению моделей общая норма получила практическое наполнение благодаря руководящим административным разъяснениям. В феврале 2026 года Ministry of Culture, Sports and Tourism (Министерство культуры, спорта и туризма Республики Корея) и Korea Copyright Commission (Корейская комиссия по авторскому праву) совместно опубликовали Руководство по применимости доктрины добросовестного использования к обучению моделей генеративного ИИ (упоминавшееся в части I настоящего исследования), в котором четыре фактора статьи 35-5 применяются к этапам процесса обучения, описанным в части I настоящего исследования. В Руководстве разъясняется, что способ получения обучающих материалов имеет значение: так, несанкционированный сбор данных с веб-страниц, обход технических средств защиты, преодоление платного доступа (пейволлов) и игнорирование заявлений о резервировании прав, выраженных через протокол исключения поисковых роботов, свидетельствуют против признания использования добросовестным; при этом приводятся иллюстративные примеры по обе стороны границы допустимого[98].

---

[95] Регламент ЕС об искусственном интеллекте, статья 53(1)(c) (обязанность внедрить политику соблюдения законодательства Союза об авторском праве, а также выявлять и соблюдать, в том числе с применением передовых технологий, заявления о резервировании прав, сделанные в соответствии со статьей 4(3) Директивы CDSM).

[96] Начиная с комплексного законопроекта о внесении изменений в закон «Об авторском праве», внесенного в Национальное собрание 15 января 2021 г., в котором содержалось исключение в отношении интеллектуального анализа текста и данных; срок рассмотрения законопроекта истек с окончанием полномочий Национального собрания двадцать первого созыва, а пришедшие ему на смену инициативы остаются на рассмотрении.

[97] Supreme Court of Korea (Верховный суд Республики Корея), решение 2021Da272001 (2024 г.), содержащее основополагающее толкование критериев применения статьи 35-5 закона «Об авторском праве».

[98] Руководство Кореи по добросовестному использованию, выше, сноска 10.

5. Системы вознаграждения

(a) Разновидности систем вознаграждения

В случаях, когда использование охраняемых произведений при обучении моделей ИИ разрешено на основании исключения, лицензии либо того и другого вместе, может вставать вопрос о вознаграждении. В настоящее время рассматриваются несколько моделей, опирающихся на более широкий опыт адаптации авторского права к новым формам коммерческого использования.

Первая модель представляет собой систему лицензионных платежей, согласовываемых индивидуально либо коллективно и уплачиваемых разработчиком в момент формирования корпуса или обучения модели. Преимущество этой модели заключается в том, что правообладатель получает оплату за использование произведения при обучении без каких-либо дополнительных условий; слабая же сторона при индивидуальном лицензировании состоит в том, что за рамками модели остается многочисленная группа мелких правообладателей, для которых прямые переговоры неосуществимы.

Вторая модель — система сборов, в рамках которой сбор взимается с разработчиков ИИ или с определенных категорий пользователей сервисов генеративного ИИ, а собранные средства распределяются между правообладателями через ОКУ или аналогичную структуру. В исследовании Европейского парламента, упомянутом выше в разделе B.3, для обучения моделей ИИ в Европейском союзе рекомендуется определенная форма установленного законом лицензирования, сочетающая в себе признаки системы сборов с обязательным коллективным управлением. Аналогичные предложения выдвигались и для юрисдикций за пределами Европейского союза, зачастую по аналогии со сборами за воспроизведение в личных целях, существующими во многих странах. Возникают вопросы относительно совместимости подобных схем с трехступенчатым тестом по статье 9(2) Бернской конвенции, статье 10 ДАП и статье 13 Соглашения ТРИПС, а также, в той мере, в какой затрагиваются права помимо воспроизведения, со специальными условиями статей 11*bis*(2) и 13(1) Бернской конвенции.

Третью, наиболее новую и концептуально самостоятельную модель представляет собой вознаграждение на уровне результатов работы модели. При таком подходе объектом взимания вознаграждения выступает не само использование произведений при обучении, а создание развернутой моделью результатов, которые замещают произведения конкретных авторов или конкурируют с ними. Данная модель напрямую увязывает вознаграждение с возможным замещением произведения правообладателя на соответствующем рынке. Ее практическая реализация, по-видимому, требует технических механизмов для выявления замещающих результатов и установления их связи с конкретными произведениями из обучающих данных.

Возникают еще два вопроса. В тех случаях, когда число выявленных произведений чрезвычайно велико, размер вознаграждения в расчете на одно произведение может оказаться настолько малым, что станет чисто символическим. Кроме того, уточнения требует основа расчета вознаграждения: в большинстве предложений вознаграждение опирается не на доказанность факта воспроизведения или переработки конкретного произведения в определенном результате работы модели, а на отчетность разработчика о способах использования произведений в системе, что предполагает наличие механизмов прозрачности, рассмотренных в разделе C

настоящей части[99]. Более того, авторы могут страдать от замещения независимо от того, использовалось ли их произведение при обучении модели. Как отмечалось выше, при современных технологиях фильтрации результаты работы модели на практике вряд ли будут нарушать права на какое-либо конкретное охраняемое авторским правом произведение, в связи с чем рассматриваемый ущерб носит скорее системный характер, нежели форму типичного нарушения авторского права. Воспроизведение, совершенное на этапе обучения или сохраняющееся в модели, еще не означает нарушения прав самими результатами работы: анализ нарушения в отношении любого конкретного результата строится на сходстве этого результата с охраняемым произведением.

Масштабы проблемы замещения все чаще находят документальное подтверждение в музыкальном секторе: в апреле 2026 года Deezer сообщила, что полностью созданные ИИ треки составили около 44 процентов от ежедневного объема загрузок (примерно 75 000 треков в день), тогда как Spotify сообщила об удалении более 75 млн спам-треков за предшествующие 12 месяцев[100]; при этом, согласно отраслевому исследованию, к 2028 году контент, созданный ИИ, способен вытеснить значительную долю доходов авторов музыки[101]. В ответ на это профессор Зенфтлебен указывает, что критерий замещения мог бы служить концептуальным инструментом разграничения правомерного использования ИИ в поддержку человеческого творчества, не требующего уплаты вознаграждения, и использования ИИ в целях замещения человеческого творчества, за которое вознаграждение должно выплачиваться. Аналогичный подход лежит, по-видимому, и в основе редакции бразильского законопроекта 2338/2023, одобренной Сенатом, которая сочетает механизм отказа от использования с правом на вознаграждение, осуществляемым правообладателем индивидуально или коллективно[102].

. (b) Формы и основа расчета вознаграждения

В Бразилии эффект замещения произведений человека синтетическими материалами наглядно прослеживается в конкретных отраслях, включая звуковое оформление пространств, где пользователи стремятся проигрывать музыку, полностью сгенерированную системами ИИ, с целью уклонения от уплаты сборов за публичное исполнение. Рассматриваемые в настоящее время судебные дела против ECAD по

---

[99] В музыкальной индустрии целостность метаданных на уровне произведений и на уровне правообладателей, привязанных к обучающим материалам, признана исходным техническим условием как для установления вклада произведений, так и для любого распределения доходов от сгенерированных результатов: Де Берардинис и др. «RAIM: Operationalising Trustworthy AI in Music Generation», цитируемой в части I, сноска 17, настоящего исследования.

[100] Deezer, «AI-generated tracks now represent 44% of all new uploaded music» (19 апреля 2026 г.) <https://newsroom-deezer.com/2026/04/ai-generated-tracks-represent-44-of-new-uploaded-music/>, дата обращения: 3 июля 2026 г.; Spotify, «Spotify Strengthens AI Protections for Artists, Songwriters, and Producers» (25 сентября 2025 г.) <https://newsroom.spotify.com/2025-09-25/spotify-strengthens-ai-protections/>, дата обращения: 3 июля 2026 г.

[101] Международная конфедерация обществ авторов и композиторов (CISAC), исследование PMP Strategy (декабрь 2024 г.), где прогнозируется, что к 2028 году под угрозой из-за генеративного ИИ может оказаться около 24 процентов доходов создателей музыки, при этом для цифровых доходов этот показатель прогнозируется еще более высоким.

[102] Законопроект Сената № 2338/2023 в редакции, одобренной Федеральным сенатом Бразилии 10 декабря 2024 г. и направленной в Палату депутатов. На дату подготовки настоящего исследования законопроект находится на рассмотрении Специального комитета Палаты депутатов; редакция, которая в конечном итоге поступит на подпись Президенту, окончательно еще не сформирована.

этому вопросу в свое время сформируют судебную практику относительно того, влечет ли публичное исполнение музыки, созданной ИИ, в коммерческих помещениях какое-либо обязательство по выплате вознаграждения. Этот пример показывает взаимосвязь между вознаграждением за обучение моделей ИИ и вознаграждением, связанным с результатами работы ИИ.

В ходе продолжающихся дискуссий выделились два вопроса касательно возможной системы вознаграждения. Первый касается того, к чему именно должно быть привязано вознаграждение: к обучению, к эффекту замещения (то есть к результатам) либо к тому и другому вместе. В исследовании Европейского парламента и ряде научных публикаций предпринимаются попытки объединить оба элемента; в то же время бразильский законопроект 2338/2023 в его нынешней редакции точнее толковать как привязывающий компенсацию к загрузке материалов для обучения: статья 65 предусматривает компенсацию за использование охраняемого контента при анализе данных, а также при обучении и разработке систем ИИ, вследствие чего законопроект закрепляет механизм компенсации за акт обучения, не распространяющийся на использование результатов работы модели[103]. . Второй вопрос заключается в том, предполагает ли система вознаграждения наличие признанного исключительного права, в отношении которого предоставлена лицензия либо заявлен отказ от права, или же ее следует рассматривать как требование sui generis о справедливом вознаграждении, возможно, выстроенное по модели вознаграждения за воспроизведение в личных целях, существующего во многих юрисдикциях[104]. Как отмечалось выше, действующее бразильское законодательство способно охватывать этап выдачи результатов через право на воспроизведение, право на переработку или право на сообщение для всеобщего сведения даже в случае отклонения законопроекта 2338/2023, и на этом основании потребуется предварительное и прямое согласие правообладателей[105]. Воспроизведение в личных целях, напротив, представляет собой форму компенсации, а не лицензию, поскольку действия по копированию в личных целях признаются принципиально не подлежащими лицензированию по целому ряду причин, включая защиту неприкосновенности частной жизни пользователей.

В Бразилии еще одно наблюдение касается построения системы вознаграждения: эффект замещения произведений человека синтетическими материалами прослеживается в конкретных секторах, включая звуковое оформление пространств, где пользователи пытаются проигрывать музыку, полностью созданную системами ИИ, стремясь избежать уплаты сборов за публичное исполнение. Находящиеся на рассмотрении судебные дела против ECAD по данному вопросу со временем приведут к формированию судебных разъяснений о том, влечет ли публичное исполнение музыки, сгенерированной ИИ, в коммерческих помещениях обязанность выплачивать вознаграждение членам ECAD. Этот пример подтверждает взаимосвязь между

---

[103] Именно такое толкование статьи 65 законопроекта 2338/2023 в одобренной Федеральным сенатом редакции принято в настоящем исследовании. Законопроект толкуется также как сочетающий основу загрузки данных с учетом замещения произведений результатами работы модели, в частности потому, что критерии расчета компенсации отсылают к рыночным последствиям. При толковании, принятом в основном тексте, выплата компенсации привязана к анализу данных, обучению и разработке.

[104] Всестороннее рассмотрение возможных режимов sui generis и круга охватываемых ими видов использования выходит за рамки настоящего исследования.

[105] Именно такое толкование статьи 65 законопроекта 2338/2023 в одобренной Федеральным сенатом редакции принято в настоящем исследовании. Законопроект толкуется также как сочетающий основу загрузки данных с учетом замещения произведений результатами работы модели, в частности потому, что критерии расчета компенсации отсылают к рыночным последствиям. При толковании, принятом в основном тексте, выплата компенсации привязана к анализу данных, обучению и разработке.

вознаграждением за обучение моделей ИИ и вознаграждением, связанным с результатами работы ИИ: решения в одной области могут влиять на построение системы в другой[106].

Наконец, в юрисдикциях, где нормативно-правовая база еще только формируется, вопрос о выплате вознаграждения в настоящее время остается открытым. Руководство по применимости доктрины добросовестного использования к обучению моделей генеративного ИИ 2026 года, совместно изданное Ministry of Culture, Sports and Tourism (Министерство культуры, спорта и туризма Республики Корея) и Korea Copyright Commission (Корейская комиссия по авторскому праву) и упоминавшееся в части I, посвящено главным образом вопросу о том, какие операции по обучению могут осуществляться без лицензии; сопутствующий же вопрос о вознаграждении оставлен для разрешения в рамках более широкого процесса реформ, разворачивающегося в настоящее время[107]. Вместе с тем значительная часть внутренних дискуссий исходит из посыла, что коммерческое обучение моделей ИИ должно сопровождаться выплатой компенсации создателям, поэтому открытым остается лишь вопрос о механизме выплаты вознаграждения. Ситуация в Кении, Бразилии и Китае при всех различиях сопоставима: в каждой из этих стран идет законодательный процесс, стратегические варианты остаются открытыми, а итоговая модель пока не определена. В части III настоящего исследования, посвященной обеспечению соблюдения прав, анализируются практические последствия такого состояния правовой неопределенности для правообладателей, стремящихся добиться защиты своих прав.

ЧАСТЬ III. ОБЕСПЕЧЕНИЕ СОБЛЮДЕНИЯ ПРАВ

Технические операции, описанные в части I, нередко затрагивают исключительные права, признаваемые национальным законодательством об авторском праве и международными договорами в области авторского права; рассмотренные же в части II лицензионные соглашения предлагают механизмы очистки этих прав в тех случаях, когда никакие исключения или ограничения не действуют. В части III рассматривается ситуация, когда не имел места ни один из этих вариантов: когда правообладатель полагает, что его произведение использовалось при обучении модели генеративного ИИ без разрешения и вне рамок применимого исключения, и стремится добиваться защиты своих прав в судебном либо административном порядке.

До недавнего времени судебная практика по этим вопросам практически отсутствовала. В настоящее время положение дел изменилось. По состоянию на 22 августа 2026 г. в Соединенных Штатах против разработчиков ИИ было подано 137 исков, связанных с использованием охраняемых произведений системами генеративного ИИ, и еще около 30 исков зафиксировано в других юрисдикциях[108].

---

[106] Законопроект № 2338/2023, статьи 62 и сл. (в редакции, направленной в Палату депутатов). Исключение в целях охраны коммерческой тайны ограничивает обязательство по раскрытию информации и служит предметом острых дискуссий в ходе законодательного процесса.

[107] Ministry of Culture, Sports and Tourism (Министерство культуры, спорта и туризма Республики Корея) и Korea Copyright Commission (Корейская комиссия по авторскому праву), *Руководство по применимости доктрины добросовестного использования к обучению моделей генеративного ИИ* (февраль 2026 г.); см. часть I, сноска 10, настоящего исследования.

[108] Согласно трекеру судебных разбирательств, который ведет профессор Эдвард Ли на сайте chatgptiseatingtheworld.com, на карте Соединенных Штатов по состоянию на 22 августа 2026 г. насчитывалось 156 исков, и еще около 30 исков было зафиксировано в других юрисдикциях. Этот

Суды Бразилии, Германии, Канады, Китая, Соединенного Королевства, Соединенных Штатов и других стран вынесли решения, затрагивающие самые разные аспекты анализа: признается ли обученная модель копией либо содержит ли она копии произведений, на которых обучалась; распространяются ли исключения европейского права на обучение моделей ИИ; обладают ли национальные суды юрисдикцией в отношении обучения моделей, проведенного за рубежом; затрагивают ли результаты работы, замещающие произведения из обучающего корпуса, принадлежащие правообладателю право на воспроизведение или право на доведение до всеобщего сведения; и способны ли сами результаты работы охраняться авторским правом (вопрос, выходящий за рамки настоящего исследования и упомянутый здесь лишь в качестве контекста).

Настоящая часть структурирована вокруг трех основных вопросов. В разделе А обобщаются выводы из совокупности судебной практики в ее текущем состоянии с распределением материала по рассматриваемым проблемам, а не по юрисдикциям. В разделе В рассматривается структурный вопрос обеспечения соблюдения прав в отношении разработчиков ИИ, включая трансграничные особенности, усложняющие эту процедуру. В разделе С определены основные практические трудности, с которыми сталкиваются правообладатели, стремящиеся защитить свои права, включая проблемы прозрачности, распространение множества технических стандартов выражения отказа от использования и процедурные сложности сбора доказательств.

РАЗДЕЛ А. ВЫВОДЫ ИЗ ТЕКУЩЕЙ СУДЕБНОЙ ПРАКТИКИ

Судебная практика, накопившаяся за последние полтора года, фрагментарна, неоднородна по юрисдикциям и пока не получила окончательного доктринального закрепления ни в одной правовой системе. Она ответила еще далеко не на все вопросы. Тем не менее она проливает свет как минимум на четыре самостоятельных вопроса, по каждому из которых подходы судов в разных юрисдикциях начали сближаться или, в ряде случаев, расходиться. Эти вопросы последовательно рассмотрены ниже.

1. Вопрос о признании модели копией

Первый вопрос состоит в том, содержит ли сама обученная модель ИИ с точки зрения авторского права копии произведений, на которых она обучалась. Он отличается от предшествующего вопроса о том, создаются ли воспроизведения в ходе формирования корпуса и обучения модели. Техническая сущность данного вопроса о запоминании обучающих данных моделью рассматривалась в части I настоящего исследования. Три судебных постановления, вынесенных за последние полтора года, дают существенно различающиеся ответы.

В решении по делу GEMA v. OpenAI, рассмотренном выше в разделе B.5 части II, Земельный суд Мюнхена I (Landgericht München I) в ноябре 2025 г. постановил, что обучение языковой модели на текстах песен и последующее запоминание этих текстов внутри модели представляют собой акты воспроизведения по смыслу имплементированной в право Германии Директивы InfoSoc, и отклонил применение исключения по статье 4 Директивы CDSM[109].

показатель постоянно меняется и приведен по состоянию на дату подготовки настоящего исследования. См. также карту судебных разбирательств, которую ведет альянс Copyright Alliance.

[109] *GEMA v. OpenAI*, LG München I, 11 ноября 2025 г. (42 O 14139/24), Endurteil. Это решение также рассматривается в части II, раздел B.5.

По смежному вопросу о признании модели копией, касавшемуся иной технологии, Высокий суд Англии и Уэльса (High Court of Justice of England and Wales) всего несколькими днями ранее пришел к иному выводу. По делу Getty Images (US) Inc. v. Stability AI Ltd. судья дама Джоанна Смит 4 ноября 2025 г. постановила, что веса модели Stable Diffusion — латентной диффузионной модели, применяемой для генерации изображений, — не выступают контрафактными копиями фотографий, на которых обучалась данная модель[110]. Истец отказался от своих требований в связи с прямым нарушением авторского права и нарушением прав на базы данных в завершение судебного процесса, признав, что доказательства не подтверждают факт проведения обучения на территории Соединенного Королевства; поэтому спор сосредоточился вокруг вторичного нарушения авторского права, потребовав от суда установить, является ли сама модель материальным объектом, представляющим собой контрафактную копию. Суд пришел к выводу, что представленные в процессе технические доказательства подтверждают отсутствие сохранения обучающих данных в весах диффузионной модели, и что при отсутствии доказательств запоминания данных в юридически значимом смысле модель не представляет собой контрафактную копию[111]. Суд попутно добавил (obiter dictum), что в случае противоположного вывода компания Stability AI несла бы ответственность по нормам о вторичном нарушении авторского права, поскольку разработчик знал или имел разумные основания полагать, что обучающий корпус содержал произведения, собранные из Интернета без согласия правообладателей[112].

Третье решение — постановление Ханчжоуского интернет-суда по так называемому делу об Ультрамене — касается смежной, но иной фактической ситуации. Ответчиком по этому делу выступала платформа, размещавшая модели ИИ с открытым исходным кодом, к которым пользователи имели доступ; некоторые пользователи обучили эти модели на охраняемой авторским правом японской манге об Ультрамене и загрузили обученные модели для общего доступа других пользователей. Обладатель авторских прав подал иск к платформе, и суд привлек платформу к ответственности за содействие нарушению авторского права. В попутных замечаниях суд выразил мнение о том, что сам акт обучения может квалифицироваться как добросовестное использование в соответствии с законодательством Китая, исходя из того, что цель использования охраняемых авторским правом произведений при обучении модели генеративного ИИ состоит в изучении и анализе языковых закономерностей, стилистических характеристик и других элементов, а не в воспроизведении оригинальной творческой формы выражения. Необходимость выносить решение по данному вопросу у суда отсутствовала, и эти рассуждения не влияли на итоговый результат спора, однако высказанные им замечания рассматривали обучение как

---

[110] *Getty Images (US) Inc. v. Stability AI Ltd.*, [2025] EWHC 2863 (Ch) (судья Смит, 4 ноября 2025 г.). Решение затрагивает требования, основанные как на авторском праве, так и на праве товарных знаков; настоящее рассмотрение касается исключительно анализа авторских прав. Разрешение на обжалование вывода о вторичном нарушении авторского права было предоставлено в декабре 2025 г. ([2025] EWHC 3343 (Ch)); на дату подготовки настоящего исследования апелляция находилась на рассмотрении.

[111] Об архитектуре диффузионных моделей см. часть I, раздел B.1, и приведенные там источники.

[112] Комментарий см. в публикации: Деркле Э. Getty Images против Stability AI: знаковое решение, подтверждающее необходимость изменения правительством Великобритании законодательства об авторском праве (Derclaye E. Getty Images v Stability AI: A landmark judgment reinforcing the need for the UK government to amend its copyright laws), Kluwer Copyright Blog (20 ноября 2025 г.), доступно онлайн: https://legalblogs.wolterskluwer.com/copyright-blog/getty-images-v-stability-ai-a-landmark-judgment-reinforcing-the-need-for-the-uk-government-to-amend-its-copyright-laws/.

потенциально способное подпадать под добросовестное использование по законодательству Китая[113].

Эти три решения не вполне согласуются между собой по сугубо техническим причинам, поскольку они касаются различных типов моделей (трансформерная языковая модель в Мюнхене, латентная диффузионная модель в Лондоне и платформа, размещающая дообученные пользователями модели генерации изображений, в Ханчжоу), а также разных процессуальных ситуаций. И мюнхенский, и лондонский суды детально анализировали литературу по информатике, посвященную запоминанию обучающих данных в крупных генеративных моделях. Сюда относится и цикл научных публикаций профессоров Купер, Гриммельманна, Лемли и других ученых по вопросам извлечения запомненного контента из моделей с открытыми весами[114]. При этом один и тот же цикл работ мюнхенский суд истолковал как доказывающий, что обучение представляет собой масштабное воспроизведение, тогда как лондонский суд счел его подтверждением того, что веса диффузионных моделей, как правило, контрафактными копиями не выступают. Такое сопоставление показывает, что имеющиеся эмпирические данные о запоминании данных сами по себе еще не определяют целиком исход правового анализа. Следует ли рассматривать подтвержденную способность модели воспроизводить последовательности из ее обучающих данных как доказательство того, что модель представляет собой копию обучающего корпуса, либо как поведенческий дефект, подлежащий устранению с помощью фильтров выходных данных и методов машинного отучения, — этот вопрос относится к сфере правовой квалификации, по которой различные суды пришли к противоположным выводам на основе доказательственных баз, опиравшихся на сходную научную литературу, пусть и различавшихся по формату и конкретному составу представленных каждому суду доказательств[115]. Как отмечалось в части I, на вопрос о том, содержит ли модель копию охраняемых произведений или объектов смежных прав, следует отвечать применительно к каждой отдельной модели.

В Соединенных Штатах вопрос о признании модели копией был поставлен в иной и процессуально специфической форме: некоторые истцы утверждали не просто то, что процесс обучения привел к созданию копий, но что сама обученная модель представляет собой контрафактное производное произведение. По делу Kadrey v. Meta Platforms, Inc. авторы утверждали, что модели LLaMA не могли бы функционировать без выразительной информации, извлеченной из их книг, что эта информация сохраняется в весах модели, и что модели, соответственно, выступают несанкционированными производными произведениями; подобная квалификация возлагала бы ответственность за саму модель без необходимости доказывать соответствие какого-либо конкретного выходного результата произведениям, на

---

[113] Ханчжоуский интернет-суд, (2024) Zhe 0192 Min Chu № 1587 (первая инстанция); Ханчжоуский народный суд средней ступени провинции Чжэцзян, (2024) Zhe 01 Min Zhong № 10332 (вторая инстанция).

[114] Федер Купер А., Гриммельманн Дж. Файлы находятся в компьютере: авторское право, запоминание данных и генеративный ИИ (Cooper A. Feder and Grimmelmann J. The Files are in the Computer: Copyright, Memorization, and Generative AI), SSRN Working Paper No. 4803118 (2024 г.). Об эмпирических исследованиях запоминания данных в языковых моделях см. также исследование Купер и соавторов, цитируемое в части I, сноска 18, настоящего исследования. См. также: Лемли М. А., Федер Купер А. Вероятностные «копии» в моделях генеративного ИИ (Lemley M.A. and Cooper A.F. Probabilistic ‘Copies’ in Generative AI Models), цитируется в части I, сноска 25, настоящего исследования.

[115] Критический комментарий к аргументации суда по делу GEMA в этой части см. в публикации: Фрозио Дж. Авторское право в формальдегиде: как дело GEMA против OpenAI замораживает доктрину и сдерживает развитие ИИ (Frosio G. How GEMA v OpenAI Freezes Doctrine and Chills AI), Kluwer Copyright Blog (10 и 11 декабря 2025 г.), части 1 и 2.

которых обучалась модель. На стадии рассмотрения заявленных требований суд отказался принять эту теорию. Ответчик в конечном счете выиграл дело на основании доктрины добросовестного использования (рассматриваемой ниже в разделе 4), доказав, что веса представляют собой производные от обучающих данных статистические параметры, а не сохраненную форму выражения[116].

Сопоставимая теория выдвигалась в отношении латентной диффузионной модели по делу Andersen v. Stability AI Ltd., где истцы охарактеризовали модель как альтернативный способ хранения копий обучающих изображений; эта квалификация столкнулась с доказательственными трудностями из-за отсутствия результата выдачи, имеющего близкое сходство с охраняемым произведением, что перекликается с выводом английского суда по делу Getty о том, что веса диффузионной модели не сохраняют в себе обучающие данные. Опыт Соединенных Штатов показывает, что доказательство загрузки охраняемых произведений для обучения само по себе, без дополнительных данных, не устанавливает, что обученная модель служит производным произведением от этих работ[117]. На момент подготовки настоящего документа решений судов апелляционной инстанции ни по одному из указанных выше вопросов вынесено не было.

2. Исключение в отношении интеллектуального анализа текста и данных

Второй вопрос, возникающий в Европейском союзе и юрисдикциях, на авторское право которых повлияла европейская система норм (например, в странах ЕЭЗ), состоит в том, распространяются ли исключения в отношении интеллектуального анализа текста и данных, закрепленные в статьях 3 и 4 Директивы CDSM, на обучение моделей генеративного ИИ. Существо этого доктринального спора рассматривается в части II, раздел B.5, настоящего исследования. Судебная практика Германии по данному вопросу кратко обобщается здесь для полноты изложения с упором на выводы, касающиеся защиты требований правообладателей.

Два судебных решения германских судов по делу Kneschke v. LAION e.V. о применении исключения в отношении интеллектуального анализа текста и данных к обучению моделей ИИ (в суде первой инстанции: Landgericht Hamburg, 27 сентября 2024 г., и в суде апелляционной инстанции: Oberlandesgericht Hamburg, 10 декабря 2025 г.) рассмотрены выше в разделе B.5 части II.

Из них вытекают два практических вывода для обеспечения соблюдения прав. Первый состоит в том, что в юрисдикциях, где исключение в отношении интеллектуального анализа текста и данных применяется к обучению моделей ИИ, позиция правообладателя при обеспечении соблюдения прав обусловлена действием механизма отказа от использования. Если действительный отказ от использования выражен в форме, которую автоматизированный процесс способен обнаружить и

---

[116] Изложение позиции разработчиков вне рамок судебных споров см. в материале: Google, «A Policy Agenda for Responsible Progress in Artificial Intelligence» (Белая книга по регулированию ИИ), размещенном по адресу: https://blog.google/company-news/outreach-and-initiatives/public-policy/white-paper-ai-regulation/, дата обращения: 7 июля 2026 г.

[117] О концепции признания модели производным произведением в Соединенных Штатах см. дело Kadrey v. Meta Platforms, Inc., № 3:23-cv-03417-VC (N.D. Cal., судья Чабрия, 20 ноября 2023 г.) (отклонение на стадии рассмотрения заявленных требований утверждения о том, что модели LLaMA сами представляют собой контрафактные производные произведения, при сохранении возможности рассмотрения требования о прямом нарушении авторского права путем копирования для обучения); а также Andersen v. Stability AI Ltd., № 3:23-cv-00201-WHO (N.D. Cal., судья Оррик), определения от 30 октября 2023 г. и 12 августа 2024 г. (где рассматривалось, среди прочего, утверждение о том, что модель сохраняет сжатые копии обучающих изображений). Решение по вопросу добросовестного использования по делу Kadrey рассматривается ниже в разделе A.4.

обработать, правообладатель вправе защищать свои права в рамках обычных процедур по факту нарушения. Если же такой отказ не был выражен либо заявление о резервировании прав не удовлетворяет применимым требованиям, исключение для интеллектуального анализа текста и данных может подлежать применению (при соблюдении иных условий), ограничивая требование в связи с нарушением права, основанное на праве на воспроизведение. В этом плане способность правообладателя зарезервировать права отчасти зависит от технической формы сигнала об отказе от использования.

Второй вывод заключается в том, что стандарт машиночитаемости не выступает статичным. Высший земельный суд Гамбурга прямо постановил, что машиночитаемость подлежит оценке применительно к технологическим возможностям, существовавшим на момент спорного использования. Применительно к фактическим обстоятельствам дела это означало уровень развития техники во второй половине 2021 года, то есть до широкого распространения больших языковых моделей, способных извлекать смысл из неструктурированного текста на естественном языке. Данное судебное решение показывает, что отказ от использования, не являвшийся машиночитаемым в 2021 году, может быть машиночитаемым сейчас, а отказ, не признаваемый машиночитаемым в 2025 году, способен стать таковым в будущем. Практическое последствие для обеспечения соблюдения прав состоит в том, что правообладателям может требоваться пересматривать и во многих случаях заново формулировать свои отказы от использования по мере развития технологий; практическое же следствие для разработчиков ИИ заключается в том, что вывод о правомерности операции на момент ее совершения не опровергается последующим изменением соответствующих технологических возможностей[118].

3. Юрисдикция в отношении обучения моделей за пределами соответствующей территории

Третий вопрос, приобретающий все большее практическое значение, касается юрисдикции национальных судов в отношении операций по обучению, выполненных на инфраструктуре, расположенной в другой стране. Ведущее канадское судебное решение по данному вопросу по делу Toronto Star Newspapers Ltd. v. OpenAI Inc. было вынесено 7 ноября 2025 г. судьей Киммел из Высшего суда справедливости Онтарио (Ontario Superior Court of Justice)[119]. Коалиция ведущих канадских медиаорганизаций подала иск к OpenAI, утверждая, что ответчик использовал их контент для обучения модели ChatGPT в нарушение канадского авторского права, технических средств защиты, охранявших этот контент, и условий использования соответствующих веб-сайтов. Компания OpenAI подала ходатайство об отклонении иска за неподсудностью, утверждая, что акты обучения проводились на серверах, расположенных в Соединенных Штатах, и поэтому выходили за пределы территориальной сферы действия канадского закона «Об авторском праве». Суд отклонил этот довод,

---

[118] Келлер П. Второй раунд LAION: машиночитаемый, но все еще непригодный для исполнения — отсутствие прогресса по отказам от использования при анализе текста и данных (Keller P. LAION Round 2: Machine-Readable but Still Not Actionable - The Lack of Progress on TDM Opt-Outs), Kluwer Copyright Blog (17 и 18 декабря 2025 г.), части 1 и 2.

[119] *Toronto Star Newspapers Ltd. v. OpenAI Inc.*, 2025 ONSC 6217 (судья Киммел, 7 ноября 2025 г.). Истцами выступают издания Toronto Star, The Globe and Mail, медиахолдинг Postmedia и вещательная корпорация Canadian Broadcasting Corporation. Ходатайство об отклонении иска за неподсудностью строилось на утверждении, что вменяемые действия по обучению модели совершались на серверах, находящихся в Соединенных Штатах, и потому выходили за пределы территориальной сферы действия канадского закона «Об авторском праве».

постановив, что между разбирательством и Онтарио существует реальная и существенная связь, а вопрос о том, затрагивали ли вменяемые действия авторское право Канады, относится к существу спора, а не к оценке юрисдикции. В начале 2026 года дело перешло на стадию раскрытия доказательств; ответчики обжаловали решение по вопросу юрисдикции, и на дату подготовки настоящего исследования апелляция находилась на рассмотрении.

Пределы применения критерия реальной и существенной связи применительно к обучению моделей ИИ пока окончательно не определены. В научных комментариях к решению по делу Toronto Star отмечалось, что данный критерий вырабатывался в канадском праве в контексте права на сообщение для всеобщего сведения в рамках таких дел, как SOCAN v. CAIP и SOCAN v. ESA, и его перенос на право на воспроизведение далеко не очевиден[120]. Для целей установления юрисдикции право на воспроизведение традиционно привязывается к месту создания копии или осуществления фиксации. Соотношение между территориальной сферой действия права на воспроизведение и принципом национального режима по статье 5 Бернской конвенции рассматривается ниже в разделе B.2.

Другие примеры из канадской практики иллюстрируют разнообразие доступных процессуальных мер. По делу CanLII v. Caseway AI организация CanLII в 2024 году подала иск против ванкуверского стартапа Caseway AI в связи с массовым скачиванием ее базы данных судебных решений с целью обучения ИИ-ассистента в области правовой информации. Организация CanLII утверждала, что хотя сами судебные решения и находятся в общественном достоянии, структурированная организация, метаданные и отбор материалов ее базы данных образуют сборник, охраняемый авторским правом. В марте 2026 года стороны заключили мировое соглашение на конфиденциальных условиях; в силу заключения сторонами мирового соглашения судебного решения по существу данных вопросов по делу вынесено не было. В производстве находятся и коллективные иски: так, в Британской Колумбии писатель Дж. Б. Маккиннон подал коллективные иски против Meta, Nvidia, Anthropic и других компаний, утверждая, что они использовали незаконные копии произведений (извлеченные из набора данных Books3) для обучения своих моделей; в Квебеке Анн Робийяр обратилась за разрешением на подачу коллективных исков против Meta и OpenAI, требуя предусмотренной законом компенсации в размере до 20 000 канадских долларов за каждое произведение, что составляет установленный законом максимальный размер за нарушение в коммерческих целях. То же структурное ограничение оказалось решающим в английском судебном разбирательстве по делу Getty, рассмотренном выше в разделе A.1, где от требований в связи с прямым нарушением авторского права отказались после того, как доказательства не подтвердили факт проведения обучения на территории Соединенного Королевства[121].

---

[120] Society of Composers, Authors and Music Publishers of Canada v. Canadian Association of Internet Providers, 2004 SCC 45 («SOCAN v. CAIP»); см. также Society of Composers, Authors and Music Publishers of Canada v. Entertainment Software Association, 2022 SCC 30. Доктринальное замечание о том, что данный критерий формировался применительно к праву на сообщение для всеобщего сведения и может быть трудно применим к праву на воспроизведение, высказывалось в академических комментариях к решению по делу Toronto Star.

[121] Дело *CanLII v. Caseway AI* завершилось конфиденциальным урегулированием спора в марте 2026 г. Коллективные иски, поданные Дж. Б. Маккинноном в Британской Колумбии против Meta, Nvidia, Anthropic и Databricks, а также Анн Робийяр в Квебеке против Meta и OpenAI, на дату подготовки настоящего исследования находятся на стадии сертификации группы истцов. О максимальном размере компенсации за нарушение в коммерческих целях, предусмотренном законом, см. Copyright Act, RSC 1985, c C-42, статья 38.1(1)(a).

В Индии Высокий суд Дели (High Court of Delhi) подтвердил свою юрисдикцию в отношении обучения, проводившегося на серверах в Соединенных Штатах, по делу ANI Media Pvt. Ltd. v. OpenAI OPCO LLC, постановив на основе оценки prima facie, что сохранение произведений за рубежом выступало лишь «заключительным этапом в цепочке событий», берущей начало с доступа из Индии, и что возражение разработчика «несостоятельно, поскольку оно привело бы к уклонению от применения индийского авторского права»[122]; существо данного решения рассматривается ниже в разделе A.4. В Бразилии вопрос об обучении модели за пределами соответствующей юрисдикции был поставлен в иске, поданном газетой Folha de São Paulo против OpenAI в августе 2025 года, в котором издание утверждает, что ответчик без разрешения использовал его журналистский контент при обучении ChatGPT и предоставлении услуг бразильским пользователям[123]. Разбирательство проходило в режиме судебной тайны; стороны урегулировали спор в 2026 году, и решения по существу дела вынесено не было. В Аргентине крупное новостное издательство Editorial Perfil в августе 2026 года обратилось с иском против OpenAI и Microsoft в Федеральные суды по гражданским и коммерческим делам, заявляя о несанкционированном использовании своего журналистского контента при обучении и эксплуатации сервисов генеративного ИИ[124]. В более широком плане по аналогии с судебной практикой Supremo Tribunal Federal (Федеральный верховный суд) в отношении зарубежных цифровых платформ, действующих в Бразилии, был выдвинут еще один довод, пока не проверявшийся бразильскими судами в контексте ИИ[125]. Согласно этому аргументу, предоставление услуг ИИ в Бразилии подчиняет лежащее в их основе обучение бразильскому авторскому праву, поскольку последующая генерация результатов бразильскими пользователями выступает продолжением акта обучения; с этой точки зрения, если обучение происходило неправомерно, предложение зависящего от него сервиса затрагивает законодательство Бразилии об авторском праве через принцип национального режима в соответствии со статьей 5 Бернской конвенции[126]. По существу данный аргумент представляет собой вариант принципа, согласно которому правонарушение, начавшееся за рубежом, но повлекшее

---

[122] ANI Media Pvt. Ltd. v. OpenAI OPCO LLC, High Court of Delhi, CS(COMM) 1028/2024 (судья Бансал, судебное решение от 24 июля 2026 г.), 2026 SCC OnLine Del 5291, пункт 51 для приведенных в тексте цитат. Постановление было вынесено по заявлению истца о предварительном судебном запрете; производство по иску продолжается, и на дату подготовки настоящего исследования апелляция находилась на рассмотрении коллегии из двух судей (Division Bench) Высокого суда.

[123] Folha de São Paulo v. OpenAI, дело № 1107237-96.2025.8.26.0100 (TJSP). Иск был подан в августе 2025 г. и рассматривается в закрытом режиме; публичного решения на дату подготовки настоящего исследования вынесено не было.

[124] Editorial Perfil S.A. v OpenAI, L.L.C., Microsoft Corporation and Microsoft de Argentina S.R.L. (Федеральные суды по гражданским и коммерческим делам, Аргентина, иск подан 24 августа 2026 г.). На момент подготовки настоящего документа в открытом доступе было мало информации об иске, однако см. Editorial Perfil, «Perfil demanda a OpenAI y Microsoft por el uso de contenidos periodísticos sin autorización» (24 августа 2026 г.) <https://www.perfil.com/noticias/judiciales/perfil-demanda-a-openai-y-microsoft-por-el-uso-de-contenidos-periodisticos-sin-autorizacion.phtml>, дата обращения: 29 сентября 2026 г. См. также EFE, «El diario argentino Perfil demanda a OpenAI y Microsoft por uso indebido de sus contenidos» (25 августа 2026 г.) <https://www.swissinfo.ch/spa/el-diario-argentino-perfil-demanda-a-openai-y-microsoft-por-uso-indebido-de-sus-contenidos/91953813>, дата обращения: 29 сентября 2026 г.

[125] STF, RE 1037396 и RE 1057258 (объединенные дела). Supremo Tribunal Federal (Федеральный верховный суд) постановил, что платформы могут нести ответственность за контент, размещенный пользователями, если они не принимают мер после получения уведомления, что платформы, предоставляющие услуги в Бразилии, обязаны соблюдать бразильское законодательство, и что такие компании должны иметь законного представителя в стране.

[126] Бернская конвенция об охране литературных и художественных произведений, статья 5(1) (национальный режим); см. также статью 5(2) (независимость охраны).

последствия в юрисдикции суда, может преследоваться в этой юрисдикции, применительно к технологическому контексту, в котором соответствующая фиксация происходит до наступления эффекта, воспринимаемого пользователем.

4. Вопрос о замещении произведений результатами работы модели

Четвертый вопрос касается правовой квалификации результатов работы модели, которые замещают охраняемый материал, на котором обучалась модель, или конкурируют с ним. Этот вопрос имеет существенное значение, поскольку определяет, может ли заявляемый правообладателями экономический ущерб быть устранен в рамках существующих категорий авторского права, привязанных к актам воспроизведения и сообщения для всеобщего сведения, либо требует применения отдельного аппарата (рассмотренного в части II) — вознаграждения за эффекты замещения, в отношении которых ни один отдельный результат не дает оснований для иска. Технические основы изложены в части I, раздел C.3.

В Соединенных Штатах ряд судов уже обращался к этой теме. В рамках дела The New York Times Company v. Microsoft Corp., ныне объединенного в производство *In re* OpenAI, Inc. Copyright Infringement Litigation, ответчик на ранней стадии разбирательства заявил ходатайство об отклонении иска в части требований о прямом нарушении авторского права на том основании, что истец применил усложненные манипулятивные запросы для искусственного провоцирования дословного воспроизведения обучающих данных, послужившего поводом для претензий. Окружной суд Южного округа Нью-Йорка отказался отклонить эти требования по указанному основанию в апреле 2025 года, и дело перешло на стадию раскрытия доказательств. Центральное место на стадии раскрытия доказательств занял спор о предоставлении журналов взаимодействия с конечными пользователями. В серии постановлений, завершившихся подтверждением судьей Стайном 5 января 2026 г. более раннего постановления мирового судьи Ван, суд предписал предоставить рандомизированную обезличенную выборку из 20 млн журналов взаимодействия пользователей с ChatGPT[127]. Это постановление смещает доказательственный фокус с вопроса о том, можно ли вызвать дословное воспроизведение обучающих данных через злонамеренно сформированные запросы, на то, как часто оно возникает при обычном потребительском использовании. Суд расценил широкие доказательства результатов выдачи пользователям в обычном режиме как имеющие значение для анализа добросовестного использования, в частности для ответа на вопрос о том, выступают ли результаты работы модели рыночным замещением произведений истцов.

Затем в двух решениях в порядке упрощенного судопроизводства, вынесенных Окружным судом Северного округа Калифорнии в июне 2025 года с интервалом в два дня по делам Bartz v. Anthropic PBC и Kadrey v. Meta Platforms, Inc., два федеральных окружных суда пришли к одинаковому результату по главному вопросу, однако на основе существенно различающейся аргументации. По делу Bartz судья Олсап постановил, что обучение большой языковой модели на правомерно приобретенных книгах носило чрезвычайно трансформирующий характер и образовывало добросовестное использование в соответствии со статьей 107 закона «Об авторском праве»; одновременно он указал, что скачивание и сохранение пиратских копий этих

[127] *In re OpenAI, Inc. Copyright Infringement Litigation*, № 1:25-md-03143-SHS-OTW (S.D.N.Y.). Данное объединенное производство по делам из нескольких судебных округов консолидирует ряд исков, поданных против OpenAI новостными организациями, авторами и другими истцами. О постановлении о раскрытии доказательств в виде 20 млн журналов ChatGPT см. источники, приведенные в части I, сноска 20, настоящего исследования.

книг из LibGen и PiLiMi добросовестным использованием не являлось, и 17 июля 2025 г. группа истцов была сертифицирована исключительно в отношении требований, связанных с пиратством. Впоследствии стороны достигли соглашения об урегулировании спора на сумму 1,5 млрд долл. США, получившего окончательное утверждение в июле 2026 г.128 По делу Kadrey судья Чабрия также вынес решение в порядке упрощенного судопроизводства в пользу ответчика по вопросу о добросовестном использовании[129], однако в обосновании сделал больший акцент на четвертом факторе добросовестного использования (ущерб рынку) и меньший — на первом (трансформирующее использование), повторяя методологию Верховного суда по делу Andy Warhol Foundation v. Goldsmith. В последующих комментариях дело трактуется как разрешившееся главным образом из-за неспособности истца представить доказательства замещения на рынке, из чего следует, что иная доказательственная база могла бы привести к другому результату[130]. Эти два решения не устанавливают общего правила о том, признается ли обучение моделей ИИ добросовестным использованием или нет.

По делу Thomson Reuters Enterprise Centre GmbH v. *Ross Intelligence Inc.* Апелляционный суд третьего округа США, оставив в силе частичное решение в порядке упрощенного судопроизводства в рамках промежуточной апелляции, вынес первое в практике США решение суда апелляционной инстанции по вопросу добросовестного использования и обучения ИИ. Ответчик использовал аннотации судебных актов (headnotes) истца — редакционные резюме правовых позиций, подготовленные для исследовательской платформы Westlaw, — для обучения негенеративной системы правового поиска на базе ИИ. Применительно к первому фактору суд счел использование в лучшем случае минимально трансформирующим, поскольку оно преследовало по существу ту же цель правового анализа, что и использование аннотаций истцом; этот вывод в сочетании с коммерческим характером использования привел к тому, что первый фактор свидетельствует против добросовестного использования. Суд провел различие с делами о промежуточном копировании, указав, что в них копирование было необходимо для доступа к неохраняемым функциональным элементам программного обеспечения, тогда как никакой сравнимой необходимости, оправдывающей копирование аннотаций, не существовало. Второй фактор складывался в пользу ответчика, но, как отметил суд, редко имеет решающее значение; третий фактор свидетельствовал против добросовестного использования, поскольку аннотации заимствовались целиком без достаточного трансформирующего обоснования. По четвертому фактору суд установил причинение вреда в трех аспектах: ценности аннотаций как фактора привлечения подписчиков на платформу истца; исходному рынку платформ правового анализа, на котором ответчик конкурировал напрямую; и, в завершение анализа

---

[128] *Bartz v. Anthropic PBC*, № 3:24-cv-05417 (N.D. Cal., судья Олсап, 23 июня 2025 г.) (частичное вынесение решения в порядке упрощенного судопроизводства в пользу ответчика по основаниям добросовестного использования в части обучения на правомерно приобретенных произведениях и отказ в вынесении решения в пользу ответчика в отношении скачивания и сохранения пиратских копий произведений из LibGen и PiLiMi). Группа истцов была сертифицирована исключительно в отношении требований, касающихся пиратства. Мировое соглашение на сумму 1,5 млрд долл. США получило предварительное одобрение 25 сентября 2025 г.; окончательное слушание по вопросу справедливости условий урегулирования состоялось 14 мая 2026 г. перед судьей Мартинес-Ольгин, оставившей вопрос на рассмотрение суда. На дату подготовки настоящего исследования ожидалось окончательное утверждение. Окружной суд окончательно утвердил соглашение об урегулировании в июле 2026 года, и производство по делу было прекращено.

[129] *Kadrey v. Meta Platforms, Inc.*, № 3:23-cv-03417-VC (N.D. Cal., судья Чабрия, 25 июня 2025 г.).

[130] О различиях в аргументации двух судей см. (среди многих комментариев) анализ в публикации Wiggin and Dana от 18 сентября 2025 г., доступной по адресу wiggin.com.

добросовестного использования, потенциальному производному рынку лицензирования аннотаций в качестве обучающих данных для ИИ, который суд охарактеризовал как стремительно развивающийся, отметив, что истец использовал свои аннотации для обучения собственных продуктов ИИ, отказ лицензировать их третьим лицам не опровергает существование такого рынка, а несанкционированное копирование лишило истца возможности выйти на него. Суд отграничил генеративные модели, фигурировавшие в делах Bartz и In re OpenAI, подчеркнув, что оценка трансформирующего характера в делах о генеративных системах может быть иной. Тем не менее данное решение подтверждает на уровне суда апелляционной инстанции, что формирующийся рынок лицензирования контента, который авторы обычно создают сами или разрешают создавать другим, представляет собой юридически значимое обстоятельство в рамках четвертого фактора[131]. Волна урегулирований споров в музыкальной сфере не положила конец судебным разбирательствам в этой отрасли. В январе 2026 года группа музыкальных издательств подала еще один иск против разработчика, участвующего в процессе по делу Bartz, заявив об использовании более двадцати тысяч музыкальных произведений и требуя предусмотренной законом компенсации на сумму свыше 3 млрд долл. США, а общества по сбору авторских отчислений в Германии и Дании продолжили преследование другого разработчика систем генерации музыки[132].

За пределами Соединенных Штатов суды также обращались к данной теме. Этот вопрос фигурировал в упомянутом выше деле Folha de São Paulo v. OpenAI до тех пор, пока стороны не урегулировали спор. Примером дела, остающегося на рассмотрении на момент подготовки настоящего документа, служит процесс в Пекинском интернет-суде, в рамках которого четыре иллюстратора подали иск против Xiaohongshu, утверждая, что платформа использовала их оригинальные произведения для обучения своей модели генерации изображений Trik без разрешения, и что модель

---

[131] *Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc.*, № 1:20-cv-613 (D. Del., судья Бибас, 11 февраля 2025 г.) (частичное удовлетворение ходатайства истца о вынесении решения в порядке упрощенного судопроизводства по прямому нарушению авторского права и отклонение довода о добросовестном использовании). Решение окружного суда было оставлено в силе в порядке промежуточной апелляции: № 25-2153 (3d Cir., слушания состоялись 11 июня 2026 г., решение вынесено 29 сентября 2026 г.) (мнение судьи Монтгомери-Ривз). В сноске апелляционный суд отметил заявление об интересе (Statement of Interest), поданное Министерством юстиции США по делу In re OpenAI, Inc. Copyright Infringement Litigation, № 1:25-md-3143 (S.D.N.Y., 1 сентября 2026 г.), где утверждалось, что обучение большой языковой модели генеративного типа представляет собой трансформирующее использование, и отграничил генеративный контекст от рассматриваемого им дела.

[132] Компании Universal Music Publishing Group, Concord и ABKCO подали второй иск против Anthropic в Окружной суд США Северного округа Калифорнии 28 января 2026 г., заявляя о нарушении прав на более чем 20 000 произведений. Этот иск подан отдельно от первоначального иска издателей к Anthropic 2023 года по поводу текстов песен, который был направлен в Окружной суд Среднего округа Теннесси, а затем передан в Окружной суд Северного округа Калифорнии. См. Music Business Worldwide, «UMG, Concord and ABKCO sue Anthropic for $3bn in what could be single-largest non-class action copyright case» (28 января 2026 г.) <https://www.musicbusinessworldwide.com/umg-concord-and-abkco-sue-anthropic-for-3bn-in-what-could-be-single-largest-non-class-act/>, дата обращения: 3 июля 2026 г.; о деле 2023 года: Rolling Stone, «Music Publishers Sue Amazon-backed AI Platform for Copyright Infringement» (18 октября 2023 г.) <https://www.rollingstone.com/music/music-news/umg-concord-abkco-sue-anthropic-ai-company-copyright-infringement-1234856761/>, дата обращения: 3 июля 2026 г.; Music Business Worldwide, «Universal, Concord and ABKCO ask court for injunction to stop Anthropic AI from using their lyrics» (16 ноября 2023 г.) <https://www.musicbusinessworldwide.com/universal-concord-and-abkco-ask-court-for-injunction-to-stop-anthropic-ai-from-using-their-lyrics/>, дата обращения: 3 июля 2026 г. Организации GEMA (Германия) и Koda (Дания), в числе прочих, продолжили судебные разбирательства против Suno, невзирая на урегулирования споров, достигнутые звукозаписывающими компаниями с этим разработчиком и с Udio в конце 2025 года.

способна генерировать изображения, имеющие существенное сходство с работами истцов[133].

В январе 2025 года в Республике Корея три ведущие эфирные вещательные компании — KBS, MBC и SBS — подали иск против Naver в Центральный районный суд Сеула, утверждая, что их новостной контент использовался без разрешения для обучения линейки языковых моделей HyperCLOVA от Naver. Это первое корейское судебное разбирательство, в котором вопрос обучения моделей поставлен напрямую, и в силу общего происхождения статьи 35-5 корейского закона «Об авторском праве» и статьи 107 закона США об авторском праве приведенная выше судебная практика США по вопросам добросовестного использования входит в состав аргументов, заявляемых перед корейским судом[134]. В Бразилии актер дубляжа подал в суд на торговый центр, заявляя о несанкционированном использовании своего голоса в рекламных роликах; торговый центр возражал, что спорный голос был сгенерирован ИИ, и этот довод был поддержан судом первой инстанции. Однако в 2024 году Апелляционный суд Сан-Паулу отменил это решение, указав, что «использование искусственного интеллекта само по себе не устраняет (а напротив, усугубляет) риск неправомерного использования прав третьих лиц», поскольку такие инструменты обучаются на основе контента, полученного в результате автоматизированного сбора данных с веб-страниц без разрешения правообладателей. Суд также признал необходимым сбор дополнительных доказательств и направил дело на новое рассмотрение в суд первой инстанции; на дату подготовки настоящего исследования окончательного решения по существу вынесено не было[135].

В Индии, напротив, вопрос обучения моделей получил к настоящему моменту наиболее развернутую судебную оценку за пределами Соединенных Штатов и Германии. По делу ANI Media Pvt. Ltd. v. OpenAI OPCO LLC новостное агентство заявило, что его произведения использовались без разрешения как для обучения языковых моделей ответчика, так и при формировании выходных результатов. Вынося решение по ходатайству агентства о предварительном судебном запрете и подтвердив свою территориальную юрисдикцию в отношении обучения, проводившегося на серверах в Соединенных Штатах (см. раздел A.3 выше), Высокий суд Дели постановил, что обучение затрагивает право на воспроизведение, поскольку статья 14(a)(i) закона «Об авторском праве» 1957 года распространяет это право на хранение произведения «на любом носителе электронными средствами», а удаление копий после завершения обучения не аннулирует факт хранения. Суд тем не менее счел на основе оценки prima facie, что такое хранение подпадало под норму о добросовестном использовании (fair dealing) статьи 52(1)(a), охватывающую использование в целях «частного или личного использования, включая научные исследования». Понятие «научные исследования» получило расширительное и

---

[133] Пекинский интернет-суд (иск подан в ноябре 2023 г.; находится на рассмотрении на дату подготовки настоящего исследования); четверо истцов-иллюстраторов заявляют, что Xiaohongshu использовала их оригинальные произведения для обучения модели генерации изображений Trik без получения разрешения.

[134] KBS, MBC and SBS v. Naver Corp., Центральный районный суд Сеула (иск подан 13 января 2025 г.). Требования заявлены на основании закона «Об авторском праве» и закона «О предотвращении недобросовестной конкуренции», при этом истцы требуют возмещения убытков и судебного запрета на дальнейшее использование. Те же вещатели подали аналогичный иск против OpenAI в тот же суд в феврале 2026 года.

[135] Tribunal de Justiça de São Paulo, Civil Appeal № 1119021-41.2023.8.26.0100 (6-я палата частного права, судья-докладчик Коста Нетто, судебное решение от 31 октября 2024 г.), представлено на Форуме судей ВОИС по интеллектуальной собственности 2025 г.; см. <https://www.wipo.int/wipolex/en/judgments/details/3217>, дата обращения: 19 сентября 2026 г.

осовремененное толкование: «процесс обучения LLM, лежащих в основе ChatGPT, можно рассматривать как форму исследований для получения новых знаний», а «акты дальнейших исследований не могут сводиться исключительно к действиям человека», поскольку «в конечном счете исследования проводятся по воле человека и на благо человека». Коммерческий характер деятельности ответчика также не вывел использование за пределы указанной нормы: суд охарактеризовал использование как «сугубо частное», поскольку обучающие копии оставались в закрытой среде, доступной моделям, а не широкой публике, при этом статья 52(1)(a) не содержит прямого запрета на коммерческое использование.

В рамках оценки критерия добросовестности суд установил, что использование произведений ограничивалось исключительно обучением моделей, что доказательственная база не содержит подтверждений замещения на рынке или упущенной выгоды, помимо голословных утверждений, и что общественная польза технологии склоняет чашу весов в ее пользу; кроме того, суд отметил, что истец мог заблокировать веб-краулеры ответчика, однако не сделал этого. В предварительном судебном запрете было отказано также на том основании, что любые убытки, если они будут доказаны в ходе судебного разбирательства, могут быть возмещены в виде компенсации, при этом суд указал: требование получать лицензии от каждого сопоставимого источника новостей сделало бы разработку больших языковых моделей экономически нецелесообразной. Данное судебное решение носит промежуточный характер: разбирательство по существу продолжается, и на дату подготовки настоящего исследования апелляция находилась на рассмотрении коллегии из двух судей (Division Bench). Тем не менее его сравнительно-правовая значимость весьма высока. В то время как германские суды задавались вопросом о том, распространяется ли специализированное исключение в отношении интеллектуального анализа текста и данных на обучение моделей ИИ, а суды США — о том, образует ли обучение добросовестное использование, Высокий суд Дели подошел к тому же вопросу через норму о добросовестном использовании (fair dealing), перечисляющую «частное или личное использование, включая научные исследования», и ответил на него на данном этапе в пользу разработчика, опираясь на аргументацию (отсутствие доказанного ущерба рынку, наличие неиспользованного механизма отказа от использования, общественная польза технологии), которая перекликается с факторами, оцененными в рассмотренных выше судебных решениях судов США.

РАЗДЕЛ B. ОБЕСПЕЧЕНИЕ СОБЛЮДЕНИЯ ПРАВ В ОТНОШЕНИИ РАЗРАБОТЧИКОВ ИИ

Судебная практика, рассмотренная в разделе A, иллюстрирует материально-правовые тенденции последних полутора лет. В разделе B внимание переносится на структурный вопрос о том, как правообладатели могут добиваться защиты своих прав против разработчиков ИИ с учетом трансграничных особенностей данной технологии и неравенства сторон, типично возникающего в подобных спорах.

1. Структурные особенности системы обеспечения соблюдения прав

Постоянно проявляются три структурные особенности. Первая заключается в постоянстве и остроте трансграничной проблемы, обозначенной выше в разделе A.3. Обучение модели генеративного ИИ, как правило, осуществляется на инфраструктуре, расположенной в юрисдикции, отличной от места нахождения правообладателя; затем модель развертывается и используется во многих юрисдикциях одновременно; результаты же формируются в юрисдикции каждого отдельного пользователя. Каждый из этих этапов способен затрагивать право другой страны. Сложившаяся к настоящему времени судебная практика пока не выработала единого ответа на вопрос

о взаимодействии этих различных национальных правовых режимов, и разработчики могут сталкиваться с противоречащими друг другу обязательствами в юрисдикциях, где предлагаются их услуги.

Вторая особенность — роль принципа национального режима в соответствии со статьей 5(1) Бернской конвенции. Согласно этому принципу, авторы пользуются в странах Союза, кроме страны происхождения произведения, правами, которые предоставляются законами этих стран их собственным гражданам, а также правами, особо предоставляемыми Конвенцией. Данный принцип представляет собой требование о недискриминации и сам по себе не определяет территориальную сферу действия национального права.

Третья состоит в использовании процессуальных механизмов объединения исковых требований, главным примером которых служит объединенное производство по делам из нескольких судебных округов в Соединенных Штатах. Консолидация исков, поданных против OpenAI новостными организациями, авторами и другими истцами в Южном округе Нью-Йорка, в единое объединенное производство по делам из нескольких судебных округов позволила централизовать разбирательства, которые иначе оказались бы рассредоточенными по отдельным делам. Аналогичным образом коллективный иск по делу Bartz объединил требования в отношении пиратских копий примерно 482 000 книг и привел к заключению мирового соглашения внушительного масштаба, как отмечалось выше. В юрисдикциях, где механизмы коллективных исков (или «массовых требований») либо объединенного производства по делам из нескольких округов отсутствуют или развиты слабее, издержки на индивидуальную защиту прав могут оказаться непомерными, особенно для отдельных авторов и мелких правообладателей. Коллективные иски в Квебеке, описанные в разделе A.3, представляют собой попытку задействовать этот же механизм в иных процессуальных условиях; сопоставимую картину демонстрируют и коллективные иски в Британской Колумбии. В юрисдикциях, лишенных подобных механизмов, реальные возможности отдельных правообладателей добиваться защиты своих прав против обладающих крупными ресурсами разработчиков существенно ограничены.

2. Неравенство сторон

Помимо этих характеристик, судебная практика выявляет ряд аспектов дисбаланса, практически осложняющих обеспечение соблюдения прав. Первый — информационная асимметрия: разработчики владеют технической документацией (манифестами наборов данных, журналами работы веб-краулеров, конфигурациями обучения), по которой в принципе можно установить состав обучающего корпуса, даже если эти записи неполны; правообладатели сопоставимым доступом не обладают, кроме ситуаций, когда косвенные доказательства удается получить путем извлечения запомненных данных или иных эмпирических демонстраций. Этому посвящен анализ вопросов прозрачности в разделе C.1 ниже. Второй — неравенство ресурсов сторон: разработчики, особенно создатели крупных передовых моделей, распоряжаются ресурсами, превосходящими возможности большинства правообладателей. Затраты на ведение длительного судебного процесса по защите прав, сопряженного с проведением технических экспертиз, ставших стандартом в подобных делах, весьма значительны. Третий — доктринальная неопределенность: в условиях неустоявшегося законодательства формирующаяся судебная практика может испытывать на себе эффект отбора, поскольку часть споров стороны урегулируют миром, а по другим судятся до конца, тогда как возможности правообладателей вести показательные прецедентные процессы ограничены судебными издержками. Поэтому складывающуюся судебную практику следует оценивать с обязательным учетом такого эффекта отбора.

РАЗДЕЛ С. ПРОБЛЕМЫ ОБЕСПЕЧЕНИЯ СОБЛЮДЕНИЯ ПРАВ

При современном состоянии права и технологий правообладатели, стремящиеся защитить свои права против разработчиков ИИ, сталкиваются с тремя ключевыми проблемами. Они последовательно рассмотрены ниже.

1. Транспарентность

Первая проблема состоит в ограниченной прозрачности процесса обучения. За редчайшими исключениями, разработчики ИИ не раскрывают состав своих обучающих корпусов; если же краткие сведения и публикуются, они носят настолько обобщенный характер, что не позволяют правообладателю определить, использовалось ли его конкретное произведение. Эту закрытость усиливают ссылки на коммерческую тайну: например, в начале 2026 года разработчик системы генерации музыки ходатайствовал об оставлении в закрытом доступе даже сведений об объеме своего обучающего набора данных в споре, инициированном крупными музыкальными компаниями, сославшись на угрозу причинения вреда конкурентоспособности.[136] Одна из причин кроется в том, что состав обучающего корпуса может служить коммерческим активом, и исчерпывающее раскрытие его содержания способно повлиять на конкурентное положение разработчика. В итоге возникает информационная асимметрия, которая может сдерживать практическую способность правообладателей выявлять потенциальные нарушения и предъявлять требования о защите права.

Европейский союз подошел к решению этого вопроса через прямое законодательное требование о раскрытии информации. Статья 53(1)(d) Регламента ЕС об искусственном интеллекте обязывает поставщиков моделей ИИ общего назначения составлять и делать общедоступным «достаточно подробное резюме» о контенте, использованном для обучения модели, в соответствии с типовой формой, предоставляемой EU AI Office (Бюро ЕС по искусственному интеллекту).[137] Практическое исполнение нормы зависит, в частности, от того, какая степень детализации признается достаточной для целей этого положения. Кроме того, данная норма ограничена прямым признанием необходимости соблюдения коммерческой тайны и охраны конфиденциальной деловой информации, а реализация этого требования выступает предметом текущей работы в сфере формирования политики и стандартизации.[138] Противоречие между обязательством по раскрытию информации и исключением в целях охраны коммерческой тайны служит поводом для острых дискуссий. Итоговый практический охват этого обязательства, скорее всего, будет определяться как применением типовой формы, опубликованной EU AI Office (Бюро ЕС по искусственному интеллекту) в июле 2025 года, так и тем толкованием, которое национальные судебные и иные органы дадут указанному исключению.

Другие юрисдикции подошли к обеспечению прозрачности с помощью иных правовых инструментов. В Китае в отношении созданного ИИ контента действует разветвленная

---

[136] Music Business Worldwide, «Suno moves to keep size of its AI training data sealed in UMG and Sony's copyright case, citing competitive harm» (2026 г.) <https://www.musicbusinessworldwide.com/suno-moves-to-keep-size-of-its-ai-training-data-sealed-in-umg-and-sonys-copyright-case-citing-competitive-harm/>, дата обращения: 7 июля 2026 г.

[137] Регламент (ЕС) 2024/1689, статья 53(1)(d), упоминаемая в части II, сноска 82 настоящего исследования.

[138] Регламент ЕС об искусственном интеллекте, пункт 107 преамбулы, где четко указано, что обязанность публикации резюме должна быть сформулирована с соблюдением коммерческой тайны и конфиденциальной деловой информации и не должна требовать их раскрытия.

система обязательств по маркировке[139] Китайский режим маркировки направлен главным образом не на защиту авторских прав, а на предотвращение мошенничества, обмана и введения общественности в заблуждение; тем не менее на практике для правообладателей он открывает возможность обратного установления источника на основании результата, поскольку маркировка сгенерированного ИИ контента позволяет идентифицировать конкретные результаты выдачи как созданные ИИ и проверять по ним вероятность использования охраняемых произведений в процессе обучения.

В Республике Корея Рамочный закон «О развитии искусственного интеллекта и формировании основы для доверия», действующий с 22 января 2026 г., предписывает предварительно уведомлять пользователей сервисов генеративного ИИ, а также требует обязательной маркировки контента, созданного ИИ[140]. Обязательства по маркировке, подобно китайским, ориентированы не столько на охрану авторского права; однако они способны облегчить аналогичное обратное установление источника на основании результата. Авторско-правовые вопросы регулируются законом «Об авторском праве», а не рамочным законодательством; такое разграничение между правовыми актами воспроизводит соотношение Регламента ЕС об искусственном интеллекте и правовой базы авторского права (acquis) в Европейском союзе.

В Бразилии законодательный ответ содержится в законопроекте № 2338/2023, который в текущей редакции обязывает разработчиков ИИ раскрывать факт использования охраняемых материалов при создании систем путем публикации резюме; это обязательство сопровождается оговоркой о необходимости соблюдения коммерческой и промышленной тайны141. Практическое действие обязательства по раскрытию информации будет зависеть от того, насколько широко суды истолкуют исключение в целях охраны коммерческой тайны. То же замечание mutatis mutandis применимо и к европейской системе норм.

К проблеме прозрачности в контексте защиты прав относятся два замечания. Первое заключается в том, что прозрачность может облегчать обеспечение соблюдения прав, но сама по себе достаточной не выступает. Публикация резюме, каким бы подробным оно ни было, сама по себе не порождает основания иска; правообладатель по-прежнему должен доказать, что его произведение использовалось, что такое использование затронуло его исключительные права и что никакие исключения не действуют. Обязательство по раскрытию информации сокращает информационную асимметрию, но не устраняет ее полностью. Второе замечание связано с тем, что структура обязательства по раскрытию информации взаимодействует с конструкцией механизма отказа от использования. Правообладатель, заявивший отказ от использования, чье произведение указано в отчете о прозрачности как включенное в обучающий корпус, может получить более четкие основания для предъявления требования о защите права с учетом применимого права и других условий

---

[139] Положение «Об администрировании сервисов алгоритмических рекомендаций в информационных службах Интернета» (КНР), действует с 1 марта 2022 г.; Положение «Об управлении технологиями глубокого синтеза в информационных службах Интернета» (КНР), действует с 10 января 2023 г.; Временные меры «По управлению сервисами генеративного искусственного интеллекта» (КНР), действуют с 15 августа 2023 г.; Меры «По маркировке контента, созданного и синтезированного с помощью искусственного интеллекта» (КНР), действуют с 1 сентября 2025 г.; а также обязательный национальный стандарт с тем же названием, утвержденный Государственным управлением по регулированию рынка и Администрацией стандартизации в 2025 г.

[140] Рамочный закон «О развитии искусственного интеллекта и формировании основы для доверия» (Республика Корея), обнародован в январе 2025 г., действует с 22 января 2026 г.

[141] Законопроект № 2338/2023 (Бразилия), статья 62 (в редакции, направленной в Палату депутатов); о статусе рассмотрения законопроекта см. часть II, сноска 100 настоящего исследования.

соответствующего исключения; практический же вопрос сводится к тому, окажется ли резюме достаточно конкретным для проведения подобной идентификации.

2. Атаки на данные, фильтрация и распространение стандартов отказа от использования

Вторая проблема охватывает технические методы, с помощью которых правообладатели, разработчики и посредники пытаются реагировать на технические особенности обучения и логического вывода, имеющие значение для обеспечения соблюдения прав. Можно выделить четыре самостоятельные категории методов.

Первая — атаки с целью извлечения обучающих данных, разработанные в научной литературе по информационной безопасности и ныне применяемые в судебных спорах. Атаки с целью извлечения обучающих данных представляют собой методы побуждения обученной модели к воспроизведению контента из ее обучающих данных посредством тщательно составленных запросов. Массив исследований в этой области, рассмотренный в части I настоящего исследования, приводится в литературе по безопасности как доказательство того, что возможность извлечения может выступать структурным свойством больших генеративных моделей, а не исключительно следствием специфической стратегии составления запросов[142]. С точки зрения правообладателей такие атаки служат методом сбора доказательств: они способны подтвердить, что конкретные произведения вошли в процесс обучения или могут быть восстановлены из модели. С точки зрения разработчиков аналогичные атаки представляют собой уязвимость, подлежащую устранению с помощью дедупликации обучающих данных, фильтрации выходных данных и методов машинного отучения. Доктринальный вопрос о том, должен ли способ получения воспроизведенного из памяти результата влиять на материально-правовой анализ авторских прав, в настоящее время находится на рассмотрении судов в ряде юрисдикций, как поясняется в разделе C.2 части I настоящего исследования.

Вторая категория — группа фильтров входных и выходных данных. Разработчики все чаще внедряют фильтры, не позволяющие модели воспроизводить определенный идентифицированный контент (фильтры выходных данных) либо исключающие определенные материалы из обучающего корпуса на этапе подготовки данных (фильтры входных данных)[143]. Значение таких фильтров для материально-правовой оценки остается неопределенным. В Соединенных Штатах решения по делам Bartz и Kadrey подошли к роли фильтров по-разному: в частности, судья Чабрия связал вопрос о замещении на рынке отчасти с эффективностью фильтров в предотвращении создания моделью результатов, замещающих конкретные исходные произведения. В Европейском союзе обязательства Регламента ЕС об искусственном интеллекте по обеспечению прозрачности содержат отсылку к внутренним политикам соблюдения авторского права, внедряемым разработчиками; исполнение этих требований может влиять на правовую квалификацию практики фильтрации как составной части общих мер комплаенса разработчика.

---

[142] Об атаках с целью извлечения обучающих данных см. часть I, сноска 18 (Cooper, Gokaslan, Ahmed, Cyphert, De Sa, Lemley, Ho and Liang) и сноска 22 (Carlini et al. (2021 г.) и Nasr et al. (2025)).

[143] Фильтрация на этапе ввода данных может включать исключение произведений, метаданные которых указывают на резервирование прав, дедупликацию корпуса для снижения риска запоминания данных моделью, а также исключение контента из заведомо нелицензированных источников. Фильтрация на этапе вывода может предполагать выявление и блокирование результатов, обнаруживающих близкое сходство с идентифицированными охраняемыми произведениями. Соотношение мер фильтрации на этапе ввода и вывода данных различается у разных разработчиков и для различных типов моделей.

Третья категория — методы отравления данных, с помощью которых правообладатели или третьи лица вносят вредоносные, ошибочные или искаженные данные в наборы данных, стремясь нарушить работу любой модели, обучаемой на их основе. В Китае Государственная канцелярия интернет-информации расценила отравление данных как угрозу безопасности и запустила специальную кампанию («Цинлан: наведение порядка в сфере применения ИИ») для борьбы с ним; рекомендательный национальный стандарт «Базовые требования безопасности для сервисов генеративного искусственного интеллекта» предписывает проводить фильтрацию обучающих данных для удаления незаконной и вредоносной информации[144]. В других исследованных юрисдикциях аналогичного регуляторного подхода к проблеме отравления данных пока не сформировалось. Практическое значение отравления данных в качестве способа самозащиты прав правообладателей пока не получило авторитетной правовой оценки ни в одной юрисдикции; порождаемые им технические и правовые вопросы остаются по существу открытыми: образует ли применение методов отравления данных против разработчиков ИИ юридически значимое нарушение каких-либо прав разработчика либо представляет собой правомерное осуществление собственных прав правообладателя.

Четвертая категория имеет особое значение для европейской практики защиты прав. Речь идет о стандартах машиночитаемого заявления о резервировании прав (отказа от использования)[145]. В решении Высшего земельного суда Гамбурга по делу Kneschke v. LAION e.V., рассмотренном выше в разделе B.5 части II, указывается, что действительный отказ от использования в соответствии со статьей 4(3) Директивы CDSM должен быть одновременно машиночитаемым и пригодным для автоматизированного исполнения. Техническая реализация этого требования к настоящему времени окончательно не определена. Разработка единого словаря предпочтений в отношении использования материалов системами ИИ ведется в Рабочей группе по предпочтениям в области ИИ (AI Preferences Working Group) органа Internet Engineering Task Force (IETF) (Инженерный совет Интернета) и продолжалась по состоянию на середину 2026 года; рабочие документы все еще носят предварительный характер и определяют лишь небольшое число категорий использования (по существу, обучение моделей ИИ и виды использования, связанные с поиском, при этом категория генерации с расширенным поиском находится на стадии обсуждения), и ни один стандарт пока не принят[146]. Смежный вопрос, заслуживающий рассмотрения, состоит в том, как обязанность заявлять об отказе от использования согласуется с запретом формальностей для возникновения или осуществления прав, охраняемых Бернской конвенцией. Ответ может зависеть от фактической структуры и юридической силы конкретного механизма отказа от использования.

Параллельно появился целый ряд конкурирующих стандартов, которые распадаются на две группы, нередко ошибочно смешиваемые между собой. Первая выражает

---

[144] Государственная канцелярия интернет-информации КНР, специальная кампания «Цинлан: наведение порядка в сфере применения ИИ»; *Технологии кибербезопасности — Базовые требования безопасности для сервисов генеративного искусственного интеллекта*, рекомендательный национальный стандарт, действует с 1 ноября 2025 г., статья 4.2.1 (фильтрация содержания обучающих данных).

[145] О судебном решении по делу LAION II и распространении стандартов отказа от использования см. часть II, раздел B.5 и приведенные там ссылки на комментарий Келлера.

[146] О Рабочей группе по предпочтениям в области ИИ Инженерного совета Интернета (Internet Engineering Task Force, IETF) см. https://datatracker.ietf.org/wg/aipref/about/. Протокол резервирования прав для интеллектуального анализа текста и данных (TDM Reservation Protocol, TDMRep), TDM·AI, протокол объединения Really Simple Licensing (RSL), механизм contentsignals.org компании Cloudflare, а также стандарты C2PA и IPTC PLUS относятся к числу ведущих инициатив, действующих в этом направлении.

исключительно заявление о резервировании прав, то есть отказ от использования. Она включает протокол TDM Reservation Protocol (TDMRep) и TDM·AI, специально созданные для реализации отказа от использования в соответствии со статьей 4(3) Директивы CDSM; механизм сигналов об условиях использования контента компании Cloudflare, выступающий проприетарным расширением протокола исключения поисковых роботов (robots.txt); а также более широкие инициативы, такие как C2PA и IPTC PLUS, предлагающие функционал отказа от использования через собственные словари терминов. Сигналы такого рода уже применяются на практике против веб-краулеров диалоговых поисковых систем ИИ: по имеющимся данным, значительная доля ведущих новостных сайтов блокирует программу-краулер, используемую одной из подобных поисковых систем[147]. Вторая группа идет дальше, позволяя правообладателю внедрять позитивные машиночитаемые условия лицензирования непосредственно в ту же веб-инфраструктуру. Соответственно, вопрос сводится уже не только к тому, разрешено ли использовать произведение, но и на каких условиях, включая стоимость и выбор между оплатой за каждое сканирование или за каждый запрос к модели. Ведущим примером выступает стандарт Really Simple Licensing (RSL) — инициатива, поддерживаемая интернет-издателями и сетями доставки контента США; другой пример представляет собой модель оплаты за каждое сканирование компании Cloudflare, функционирующая непосредственно на уровне запроса доступа путем отправки нелицензированному краулеру ответа о необходимости оплаты. В этой второй группе сигнал, с которым сталкивается автоматизированный агент, служит не только отказом от использования, но и предложением заключить лицензионное соглашение. Появление ИИ-агентов способно упростить автоматизированное лицензирование в этой сфере.

Соблюдение норм и судебное разбирательство осложняются сосуществованием этих двух типов заявлений о резервировании прав в рамках единой и все еще не завершенной дискуссии по стандартам. Один и тот же поставщик может действовать в обоих направлениях, как делает Cloudflare с помощью своих механизмов сигналов об условиях использования контента и оплаты за каждое сканирование, вследствие чего статус конкретного сигнала — простое резервирование или предложение заключить лицензионное соглашение — далеко не всегда очевиден при его получении. Европейская комиссия запустила консультации по протоколам резервирования прав в отношении интеллектуального анализа текста и данных в рамках Регламента ЕС об искусственном интеллекте и Кодекса практики для моделей ИИ общего назначения с целью составления перечня общепринятых машиночитаемых протоколов, которые поставщики моделей ИИ общего назначения обязаны соблюдать для выполнения своих обязательств по Регламенту ЕС об искусственном интеллекте[148].

Фрагментация системы стандартов сама по себе создает проблему для обеспечения соблюдения прав. Постановление Высшего земельного суда Гамбурга указывает на то, что машиночитаемость отказа от использования подлежит оценке с учетом технологических возможностей на момент спорного использования; однако в условиях

---

[147] В исследовании, опубликованном в 2026 году, было установлено, что многие ведущие новостные сайты Соединенных Штатов и Соединенного Королевства блокируют программу-краулер Perplexity — PerplexityBot — в своих файлах robots.txt. См. BuzzStream, «Most Major News Publishers Block AI Training & Retrieval Bots» (17 декабря 2025 г., опубликовано в январе 2026 г.) <https://www.buzzstream.com/blog/publishers-block-ai-study/>, дата обращения: 3 июля 2026 г.

[148] Европейская комиссия, Консультации по протоколам резервирования прав в отношении интеллектуального анализа текста и данных в рамках Регламента ЕС об искусственном интеллекте и Кодекса практики для моделей ИИ общего назначения (European Commission, Consultation on protocols for reserving rights from text and data mining under the AI Act and the GPAI Code of Practice) (2025 г.); см. https://digital-strategy.ec.europa.eu.

сосуществования множества несовместимых стандартов практический вопрос о том, какие именно стандарты обязан соблюдать разработчик для следования закону, не имеет однозначного ответа. В комментариях к судебному постановлению отмечается риск того, что при отсутствии единых стандартов с открытым управлением машиночитаемые отказы от использования способны послужить основой для нового слоя посредников, превращающего доступ к открытой части Интернета на основе оплаты за каждое сканирование или оплаты за каждый клик в обычную норму. Реализуется ли этот риск, зависит от решений, которые предстоит принять органам стандартизации и разработчикам государственной политики в ближайшие годы.

Механизм отказа от использования входит в состав многоуровневой инфраструктуры управления правами. Первый уровень образуют идентификаторы и метаданные. Он охватывает признанные международные стандартные идентификаторы произведений и авторов, а также машиночитаемые метаданные об управлении правами и метаданные о происхождении, способные связать произведение со сведениями о статусе прав на него. Ведется работа по обобщению передового опыта формирования функционально совместимых, надежных и машиночитаемых данных об авторских правах, адаптированных к среде ИИ, в частности в рамках Целевой группы по инфраструктуре авторского права (Copyright Infrastructure Task Force), созванной Министерством образования и культуры Финляндии[149], исходя из предпосылки, что достоверная идентификация служит предварительным условием как для выражения предпочтений в отношении использования, так и для последующего отслеживания фактов использования после совершения соответствующего действия. Этот уровень опирается на рассмотренные выше положения ДАП, ДИФ и Пекинского договора, касающиеся информации об управлении правами.

Второй уровень образуют средства контроля доступа. Он охватывает интерфейсы прикладного программирования (API) и меры по предотвращению автоматизированного сбора данных, посредством которых правообладатели и платформы стремятся ограничить доступ к своему контенту только авторизованными агентами, наряду с техническими средствами защиты, применяемыми в сфере стриминга и издательской деятельности против несанкционированного извлечения данных. Этот уровень подкреплен приведенными выше нормами ДАП, ДИФ и Пекинского договора о средствах защиты от обхода технических средств.

Третий уровень представляет собой установление вклада произведений, под которым понимается оценка роли конкретных произведений в полученном результате работы модели с целью распределения вознаграждения там, где это применимо. Установление вклада произведений в этом значении следует отличать от выявления созданного контента. Технологии выявления контента, созданного ИИ, направлены на определение с определенной долей вероятности того, создан ли файл человеком или машиной; они не идентифицируют произведения-источники и сами по себе не способны обосновать авторско-правовое требование. Напротив, технологии установления вклада произведений ориентированы на оценку доли идентифицированных произведений в сгенерированном результате и по этой причине требуют определенного сотрудничества и прозрачности со стороны разработчика. Инвестиции в этот потенциал уже начались: один поставщик технологии установления вклада произведений, уже задействованный в упомянутом в части II шведском механизме коллективного лицензирования, был приобретен крупной

[149] Целевая группа по инфраструктуре авторского права (CITF), созванная под эгидой Министерства образования и культуры Финляндии, в 2025 году опубликовала первый отчет «Функционально совместимые, надежные и машиночитаемые данные об авторских правах в эпоху ИИ» (Interoperable, trustworthy and machine-readable copyright data in the AI era).

звукозаписывающей компанией в 2026 году[150]. Уровень установления вклада произведений напрямую связан с моделями вознаграждения на уровне результатов работы модели, упомянутыми в разделе B.6 части II. Некоторые из этих моделей опираются не на факт нарушения на этапе выдачи результатов, а на эффект замещения, вызываемый результатами работы определенных моделей.

Самостоятельное семейство инструментов установления происхождения и средств нанесения цифровых водяных знаков наряду с инициативами по маркировке контента, созданного человеком, направлено на проверку подлинности и происхождения результатов работы модели, а не на использование охраняемых произведений при обучении; эти инструменты относятся главным образом к вопросам результатов генерации, не рассматриваемым в данном исследовании, и упомянуты здесь лишь для определения места уровня установления вклада произведений в более широкой инфраструктуре.

3. Судебная практика в целом

Рассмотренная совокупность судебной практики позволяет сделать несколько сквозных выводов относительно современного обеспечения соблюдения авторских прав в контексте обучения моделей генеративного ИИ.

Первый состоит в том, что суды в целом действуют с определенной осторожностью и обычно воздерживаются от вынесения решений, претендующих на всеобъемлющее урегулирование доктринальной сферы. Принятые к настоящему времени судебные решения касались конкретных технологий (латентная диффузия в деле Getty, языковые модели в деле GEMA, правомерно и незаконно приобретенные книги в деле Bartz), конкретных процессуальных ситуаций (ходатайство об отклонении иска в разбирательстве по иску New York Times, решение в порядке упрощенного судопроизводства в делах Bartz и Kadrey) и отдельных вопросов (вопрос о признании модели копией в делах GEMA и Getty, добросовестное использование в делах Bartz и Kadrey и др.). Они пока не сформировали комплексной системы правового регулирования ни в одной отдельно взятой правовой семье. Правовая позиция в Германии, хотя и нашла отражение уже в трех решениях, сама по себе не отличается внутренним единством: линия решений по делу LAION допускает применение исключения в отношении интеллектуального анализа текста и данных к обучению моделей, тогда как позиция по делу GEMA строится на ином подходе. Положение дел в Соединенных Штатах имеет аналогичный многослойный характер и ожидает пересмотра в апелляционном порядке, а в конечном итоге — вынесения одного или нескольких постановлений Верховного суда США.

Второй вывод заключается в том, что совокупность судебной практики тем не менее выработала определенные точки сближения. Наблюдается согласие относительно закрепленного в Бернской конвенции требования о наличии автора-человека для создания оригинальности, необходимой для предоставления правовой охраны результатам работы ИИ. Это проявляется в Соединенных Штатах, Кении[151], Германии на уровне судов первой инстанции и в Европейском союзе, где критерий

---

[150] Компания Warner Music Group приобрела поставщика технологии установления вклада произведений Sureel в июне 2026 г. Технология этого же поставщика применяется в рамках лицензионной договоренности с участием STIM, рассмотренной в разделе B.3 части II.

[151] Aryeh Movement Limited v. Cynthia Beldina Akoth Okello, COPTA/E001/2025 (Copyright Tribunal of Kenya (Трибунал по авторскому праву Кении)) (где установлено, что статья 22(3) закона «Об авторском праве» (Cap 130) предполагает участие человека в приложении «достаточных усилий», вследствие чего произведение, созданное системой ИИ автономно, не способно отвечать установленному законом порогу, если только достаточное вмешательство человека не придает произведению оригинальность).

оригинальности, подтвержденный Судом Европейского союза (CJEU) по делу Mio относительно авторско-правовой охраны мебели, предполагает наличие автора-человека[152]. Наблюдается также единство мнений относительно применения права на воспроизведение к процессам формирования обучающих наборов данных (в Германии, Китае и Канаде). Наконец, прослеживается определенное сближение подходов в отношении учета замещения произведений результатами работы модели в рамках анализа (в Соединенных Штатах и потенциально в Европейском союзе через механизм применения трехступенчатого теста). К вопросам, по которым судебная практика остается разделенной, относятся признание того, содержат ли обученные модели копии произведений, использованных для обучения, переобучения и тонкой настройки, распространяется ли и каким образом исключение в отношении интеллектуального анализа текста и данных на обучение моделей ИИ, а также пределы действия трансграничной юрисдикции.

Третье замечание состоит в том, что практические проблемы, обозначенные выше в разделах C.1 и C.2 — а именно прозрачность, распространение стандартов отказа от использования, а также неравенство ресурсов и информационная асимметрия сторон, — скорее всего, будут определять развитие судебной практики в предстоящие месяцы и годы. Даже в тех случаях, когда суды выработают ответы на материально-правовые вопросы авторского права, реальная возможность принудительного осуществления требований правообладателей будет зависеть от механизмов, с помощью которых правообладатели смогут устанавливать факты использования своих произведений, выражать отказ от использования и представлять доказательства в пригодной для судебной оценки форме. В этой области техническая архитектура и правовой анализ тесно переплетены, и обеспечение соблюдения прав будет определяться решениями, принимаемыми в обеих плоскостях.

ЗАКЛЮЧЕНИЕ

В настоящем исследовании зафиксировано стремительное и неравномерное развитие по двум направлениям: экспериментирование юрисдикций с различными комбинациями исключений, механизмов лицензирования и систем вознаграждения, а также формирование совокупности судебной практики, которая пролила свет на ключевые проблемы, но пока не выработала устойчивой системы координат. Государства-члены, рассматривающие в настоящее время законодательные или нормативные меры реагирования, действуют в условиях более развитой нормативной базы, чем та, что существовала еще двенадцать месяцев назад; аспекты, по которым практический опыт и судебная практика пришли к согласию, закладывают частичную основу, тогда как спорные вопросы потребуют выбора политических решений, которые настоящее исследование стремилось обозначить, но не предрешать. Такие решения

---

[152] *Mio and Others*, C-580/23, объединенное с делом *USM v konektra*, C-795/23 (CJEU, судебное решение от 4 декабря 2025 г.). Объединенные дела касались авторско-правовой охраны произведений прикладного искусства, в частности предметов мебели (в деле C-580/23 — обеденных столов «Palais Royal»). Суд подтвердил, что объект признается охраняемым произведением лишь в том случае, если он оригинален в том смысле, что представляет собой собственное интеллектуальное творение автора, отражающее свободный творческий выбор, выражающий личность автора, и что произведения прикладного искусства подчиняются тому же требованию оригинальности, что и прочие произведения. Судебное решение не касалось искусственного интеллекта; требование наличия автора-человека для охраны результатов работы ИИ вытекает из этого критерия оригинальности, а не из какого-либо вывода Суда, направленного непосредственно на регулирование ИИ. Авторы могут в дополнение столкнуться с новым препятствием в рамках процедур защиты прав, а именно с необходимостью доказывания того, что их произведение не было создано ИИ или как минимум что они внесли достаточный творческий вклад, оправдывающий предоставление правовой охраны авторским правом.

будут по-прежнему приниматься на национальном уровне и, при соответствующем решении государств-членов, обсуждаться на международном уровне в рамках Постоянного комитета по авторскому праву и смежным правам и более широкой системы ВОИС.

Исследование показывает, что баланс между интересами правообладателей и разработчиков ИИ может определяться различными государствами-членами по-разному. По итогам анализа выделяются пять более частных вопросов.

Первый касается того, должны ли действия по воспроизведению, совершаемые при формировании корпуса и обучении модели, требовать разрешения либо подпадать под действие исключения или ограничения (специального для систем ИИ, нормы об интеллектуальном анализе текста и данных или общей нормы) и на каких условиях, включая законный доступ к материалам, цель обучения и коммерческий либо некоммерческий характер использования.

Второй охватывает построение механизма резервирования прав там, где он применяется. Сюда относится форма отказа от использования с учетом того, что машиночитаемые заявления о резервировании прав требуют технических возможностей, которыми отдельные авторы и исполнители могут не обладать. Напротив, режим резервирования, не пригодный для масштабного автоматизированного исполнения, не дает разработчикам ясного пути соблюдения установленных требований. Следует также учитывать совместимость режима отказа от использования со статьей 5(2) Бернской конвенции.

Третий состоит в том, должна ли компенсация связываться с загрузкой материала для обучения, с замещением произведений результатами работы модели или с обоими факторами; должна ли она принимать форму рыночного лицензирования на основе исключительных прав либо установленного законом требования о вознаграждении; и через какие институты она должна выплачиваться в каждом из этих случаев. Сюда входят вопросы о том, распространяются ли полномочия по коллективному управлению на использование материалов при обучении моделей, требуются ли расширенные или новые формы коллективного лицензирования, и как выстроить цепочки договоров, чтобы создаваемая стоимость доходила до авторов и исполнителей, чьи произведения используются.

Четвертый вопрос — прозрачность: какие меры по учету и раскрытию сведений об обучающих материалах требуются для обеспечения осуществления прав и распределения вознаграждения, и как согласовать эти обязательства с законной коммерческой тайной и соображениями конкуренции.

Пятый и заключительный вопрос заключается в том, сохраняют ли средства правовой защиты эффективность и соразмерность в ситуациях, когда удаление произведений из обученной модели технически неосуществимо, и как должна определяться юрисдикция в случаях, когда обучение модели проводится за рубежом, а полученные сервисы предлагаются внутри страны.

[Конец документа]