Observatoire ARGA

МЕЖДУНАРОДНЫЙ АНАЛИТИЧЕСКИЙ ДОКЛАД

Исследование №23

Авторское право и обучение генеративных моделей: прозрачность происхождения контента и независимый аудит

Правовые основания использования произведений, лицензирование, TDM и проверяемая ответственность

Организация: Observatoire ARGA

Русская версия

21 rue de l’Aviation, 64600 Anglet, France

info@argaobservatory.org | www.argaobservatory.org

23 октября 2026 года

Содержание

Резюме для руководства — 4
1. Предмет исследования, вопросы и границы выводов — 4
2. Юридические объекты: произведение, исполнение, база данных и информация — 5
3. Цепочка жизненного цикла: где именно возникают правовые риски — 6
4. Исключительные права в праве ЕС: воспроизведение и сообщение публике — 7
5. Статья 3 Директивы DSM: исследовательский TDM — 7
6. Статья 4 Директивы DSM: общее исключение и право резервирования — 8
7. Французская транспозиция: статья L122-5-3 CPI — 8
8. Базы данных, издательские каталоги и массовое извлечение — 9
9. Статья 53 AI Act: политика соблюдения авторского права — 10
10. Статья 53 AI Act: публичная сводка обучающего контента — 10
11. Добровольный GPAI Code of Practice и границы его доказательности — 11
12. Сроки, переходные правила и институциональные полномочия — 11
13. Машиночитаемые запреты и техническая архитектура opt-out — 12
14. Chain of title: как устанавливается право на конкретное произведение — 13
15. Лицензирование обучающих массивов: границы и компенсация — 13

16. Посредники, готовые датасеты и распределённая ответственность — 14
17. Обучение, память модели и выходное воспроизведение — 14
18. Техническая проверка memorization: воспроизводимость и пределы — 15
19. Удаление контента и модели: что можно обещать технически — 15
20. Цифровые доказательства: сохранение, цепочка хранения и экспертиза — 16
21. США: fair use без универсальной индульгенции обучению — 17
22. Великобритания: итоги консультации марта 2026 года — 17
23. Сингапур и сравнительный подход к computational data analysis — 18
24. GDPR, личные фотографии и двойная правовая проверка — 18
25. Метод независимого аудита: мандат, критерии и выборка — 19
26. Инвентаризация активов и граф происхождения данных — 20
27. Классификация рисков в корпусе: не одна категория «из интернета» — 20
28. Статистический аудит и честное выражение неопределённости — 21
29. Контроль согласованности независимых оценщиков — 21
30. Проверка claims поставщика: от маркетинга к юридическому доказательству — 22
31. Исправления, запросы правообладателей и право на ответ — 22
32. Оценка воздействия на рынок и доходы авторов — 23
33. Независимость Observatoire ARGA: финансы и процессуальные границы — 24
34. Применение к юридической аналитике, розыску и санкциям — 24
35. Рекомендации к консультации European Commission Technology and Copyright — 25
36. План внедрения и показатели качества независимого контроля — 25
37. Практические выводы: что подтверждено, а что остаётся спорным — 26
38. Доказательство ущерба рынку: причинность, замещение, альтернативы — 26
39. Роялти и коллективное распределение: проверяемая арифметика — 27
40. Точность юридического резервирования: время и юридический эффект — 28
41. Торговая тайна и закрытый аудит: пределы прозрачности — 28
42. Формат доказательства происхождения: совместимость и переносимость — 29
43. Выходные произведения: авторство, производность, лицензии — 29
44. Применение AI Act к модификациям GPAI: граница поставщика — 30
45. Проверяемость договоров без публикации личных и коммерческих данных — 30
46. Сравнение раскрываемых и реально проверяемых показателей — 30
47. Открытая публикация и право на возражение против экспертного вывода — 31
48. Двухконтурная проверка статьи 53 AI Act и материального авторского права — 31
49. Проверка мультимодального обучения: изображение, звук, видео — 32
50. Архитектура безопасного хранения доказательств и litigation hold — 33
51. Договорные гарантии, права на аудит и пределы заверений — 33
52. Правовое измерение открытых репозиториев и программного кода — 34
53. Происхождение фотографий, метаданные и контроль ложных атрибуций — 34
54. Музыка и смежные права: разные слои разрешений — 35
55. Издательский архив, цифровые библиотеки и лицензии доступа — 35
56. Репрезентативность тестов: зачем нужен отрицательный контроль — 36
57. Точки принятия решений: остановка, исправление, публикация — 36
58. Предметная проверка консультационных предложений ЕС — 37
59. Итоговая модель оценки доказательств без ложной точности — 37
60. Отдельные исключения: цитирование, временное копирование и TDM — 38
61. Коллективное управление правами: доступность репертуара и мандат — 38
62. Право издателей прессы и редакционные базы данных — 39
63. Институциональные и научные библиотеки: научный статус не является общим разрешением — 39
64. Достоверность метаданных, provenance-маркеров и машинных подписей — 39
65. Дедупликация и агрегирование: доказательная цена потери исходного следа — 40
66. Международные версии произведения и право перевода — 40
67. Обработка исправлений и удаления без ложных технических обещаний — 41
68. Экономика соблюдения правил для небольших поставщиков и авторов — 41
69. Оспаривание экспертного заключения и процедурная справедливость — 42
70. Доказательная шкала для юридической консультации и судебного материала — 42
71. Сводный набор минимальных мер для ARGA Observatory — 42
Приложение 1. Паспорт корпуса обучающих данных — 43
Приложение 2. Акт приёма нового источника — 43
Приложение 3. Журнал резервирования прав и обработки opt-out — 44
Приложение 4. Матрица лицензии и способов использования — 44
Приложение 5. Протокол доказательства правомерного доступа — 45

Приложение 6. Due diligence поставщика обучающих данных — 45
Приложение 7. Чек-лист публичной сводки по статье 53(1)(d) — 45
Приложение 8. Технический граф происхождения и контроль преобразований — 46
Приложение 9. Протокол дообучения и обновления версии модели — 46
Приложение 10. Испытание риска воспроизведения охраняемого текста — 47
Приложение 11. Форма регистрации авторско-правовой претензии — 47
Приложение 12. Chain of custody для цифровых материалов — 48
Приложение 13. План статистической выборки закрытого корпуса — 48
Приложение 14. Проверка согласованности двух экспертов — 48
Приложение 15. Процедура временной остановки использования набора — 49
Приложение 16. Проверка корректирующих мер и остаточного риска — 49
Приложение 17. Реестр независимости и конфликтов интересов — 50
Приложение 18. Совмещённая карта интеллектуальных прав и GDPR — 50
Приложение 19. Условный кейс: фотографический каталог — 50
Приложение 20. Условный кейс: тексты песен и генеративная модель — 51
Приложение 21. Условный кейс: учебные материалы и перевод — 51
Приложение 22. Условный кейс: программный код и открытая лицензия — 51
Приложение 23. Условный кейс: новости и издательский архив — 52
Приложение 24. Условный кейс: музейная коллекция и научные исследования — 52
Приложение 25. Условный кейс: материалы суда и коммерческая база — 53
Приложение 26. Поставщик API как источник данных для дообучения — 53
Приложение 27. Модель переговоров о лицензии на тематический каталог — 53
Приложение 28. Проверка полномочий коллективной организации — 54
Приложение 29. Трансграничная матрица юридических действий — 54
Приложение 30. Аудит заявлений о «100% законном» обучении — 55
Приложение 31. Матрица оценки стоимости соблюдения требований — 55
Приложение 32. Политика исправления публичных ошибок Observatoire — 55
Приложение 33. Проект ответов ARGA для консультации Еврокомиссии — 56
Приложение 34. Протокол защищённой комнаты данных для аудитора — 56
Приложение 35. Девяностодневный учебный пилот ARGA — 57
Приложение 36. Шаблон итогового аудиторского заключения — 57
Приложение 37. Процедура запроса уточнений у разработчика модели — 58
Приложение 38. Особый режим сведений о санкциях и международном розыске — 58
Приложение 39. Контрольная карта рисков до публикации отчёта — 58
Приложение 40. Матрица мониторинга изменений после октября 2026 года — 59
Приложение 41. Протокол сопоставления права на фотографию с версиями файла — 59
Приложение 42. Тест машиночитаемого резервирования прав при смене URL — 60
Приложение 43. Рабочий пример проверки смешанного архивного каталога — 60
Приложение 44. Сопоставление версий GPAI и публичной сводки обучающего контента — 60
Приложение 45. Тест авторских текстов без воспроизведения полного произведения — 61
Приложение 46. Договорная карта цифровой музыкальной библиотеки — 61
Приложение 47. Пример реконструкции лицензии по цепочке посредников — 62
Приложение 48. Инструкция по защищённому ответу на запрос правообладателя — 62
Приложение 49. Контроль публикации экспертного вывода Observatoire ARGA — 62
Приложение 50. Контроль трансграничного судебного сценария — 63
Приложение 51. Условный расчёт роялти без произвольной презумпции ущерба — 63
Приложение 52. Реестр версий и критических дат правовой проверки — 64
Приложение 53. Таблица разграничения TDM и выдачи модели — 64
Приложение 54. Матрица возражений правообладателя по категориям контента — 64
Приложение 55. Схема доступа независимого аудитора к закрытому корпусу — 65
Приложение 56. Протокол стратифицированной выборки источников — 65
Приложение 57. Регламент проверки версии документа во время консультации ЕС — 66
Приложение 58. Каталог запретов на использование методики аудита не по назначению — 66
Приложение 59. Проверка совместимости лицензий и сублицензий на код — 66
Приложение 60. Форма сравнительного заключения по судебной практике США и ЕС — 67
Приложение 61. Проверка издательского набора после смены лицензии — 67
Приложение 62. Протокол проверки персональных данных в лицензируемом архиве — 68
Приложение 63. Протокол внутреннего рецензирования перед утверждением — 68
Приложение 64. Итоговый редакционный реестр нерешённых вопросов — 68
Источники и нормативно-методические основания — 69

Резюме для руководства

Главный результат исследования: правомерность обучения генеративной модели на контенте не определяется одним универсальным разрешением «для искусственного интеллекта». Она складывается из проверяемых прав на произведение и базу данных, законности доступа, применимости конкретных исключений TDM, полномочий поставщика, соблюдения резервирования прав, правил охраны персональных данных и поведения модели после выпуска. Ни открытый URL, ни декларация об использовании «публичных источников», ни публикация сводки по статье 53 AI Act не подтверждают законность каждого включённого произведения. [1][2][3][6]

По состоянию на 23 октября 2026 года для ЕС действуют статьи 3 и 4 Директивы DSM, а на поставщиков применимых GPAI-моделей распространяются обязанности статьи 53 AI Act, включая политику copyright compliance и публичную сводку обучающего контента. Европейская комиссия опубликовала обязательный шаблон такой сводки в июле 2025 года. Добровольный GPAI Code of Practice помогает исполнению требований, но сам по себе не выступает судебным доказательством отсутствия нарушений. Для моделей, размещённых до 2 августа 2025 года, предусмотрены переходные положения. [1][2][3][4]

Сравнительная судебная практика требует точного разграничения действий и процессуальных стадий. В Мюнхене в ноябре 2025 года суд первой инстанции рассмотрел воспроизведение песенных текстов в модели и её выходах. В США решения по делам Bartz и Kadrey от июня 2025 года разграничили вопросы способа получения копий, использования при обучении и доказательства рыночного вреда. При этом 20 июля 2026 года спор Bartz v. Anthropic завершился утверждением группового мирового соглашения и окончательным судебным решением по соглашению. Последнее не создаёт универсального решения по существу для всех случаев обучения моделей. [11][12][13][22][23]

Для Observatoire ARGA предлагается независимая проверка по пяти уровням: юридическая карта прав и исключений; реестр происхождения исходных объектов и производных файлов; испытания обработки машиночитаемых ограничений; тесты риска воспроизведения в результатах; структурированная публичная отчётность с сохранением границ неопределённости. Рекомендуется учитывать коммерческую тайну через защищённый доступ, но не заменять им полный отказ от доказательств. В исследовании представлены 64 приложения с практическими формами, протоколами и учебными кейсами. Они не являются результатом проведённого внешнего аудита.

Особое институциональное значение имеет открытая до 3 ноября 2026 года консультация Комиссии Technology and Copyright. Позиция ARGA может быть направлена на доступные малым правообладателям технические сигналы, прозрачность цепочки лицензионных полномочий, раздельное раскрытие публичной сводки и закрытых доказательств, а также низкозатратные процедуры рассмотрения споров. Настоящий доклад формирует аналитическую основу возможного участия, но сам по себе не подтверждает подачу консультационного ответа. [5]

Итоговый принцип	Минимальное доказательство
Законный источник	Условия доступа на дату получения и правовое основание
Уважение opt-out	Действующий технический контроль и исторические журналы
Лицензирование	Объект, правообладатель, объём, сублицензия и территория
Обязанности GPAI	Политика соблюдения прав, сводка и привязка к версии модели
Достоверный аудит	Воспроизводимая выборка, ограничения, независимость

Основная аналитическая часть

1. Предмет исследования, вопросы и границы выводов

Настоящий доклад исследует правовую и доказательственную цепочку использования охраняемого контента для разработки генеративных моделей. Предметом является не абстрактное противопоставление

творчества и технологии, а определение того, какие действия совершаются на отдельных этапах: поиск источника, получение доступа, копирование, нормализация, создание обучающего набора, обучение и дообучение, хранение промежуточных копий, распространение модели и получение результатов. Каждое действие может затрагивать разные исключительные права, исключения из них, договорные обязательства и режимы контроля. Юридическое заключение о правомерности одного этапа не является разрешением для остальных. [1][2][6]

Исследование опирается на право ЕС по состоянию на 23 октября 2026 года, действующую редакцию AI Act, Директиву DSM 2019/790, InfoSoc 2001/29, право о базах данных, французскую транспозицию, официальные материалы Европейской комиссии и ограниченную сравнительную практику Великобритании, США и Сингапура. Рекомендации об аудите являются предложением Observatoire ARGA; они не представляют собой нормативно установленную сертификацию и не подтверждают проведение проверки конкретной компании. Судебные дела цитируются с указанием инстанции, даты и пределов их значения. [1][2][9][10]

Метод различает четыре статуса: обязательная норма, официальное разъяснение или административная практика, судебная оценка в конкретном деле и исследовательское предложение. Такой способ изложения особенно важен в быстро меняющейся сфере, где добровольный отраслевой кодекс иногда ошибочно выдают за закон, а решение суда первой инстанции — за окончательно установленное универсальное правило. Выводы о соблюдении прав могут делаться только по документируемой фактической цепочке, а не из коммерческой репутации разработчика или самого факта публикации декларации о данных.

Исследовательский вопрос	Что требуется установить	Чего доказательство не позволяет заключить
Можно ли использовать произведение для TDM?	Правомерность доступа, основание исключения, статус резервирования прав, вид действия	Что разрешено публично воспроизводить произведение в ответе модели
Есть ли надлежащая лицензия?	Объект, правообладатель, территория, способы использования, срок, передача прав	Что лицензия охватывает все версии, деривативы и сторонние наборы
Исполнена ли статья 53 AI Act?	Политика соблюдения авторского права и опубликованная сводка обучающего контента	Что каждое отдельное произведение использовалось законно
Можно ли доверять независимому аудиту?	Объём выборки, доступ к доказательствам, пределы проверки и независимость эксперта	Что отсутствие найденного нарушения доказывает его невозможность

2. Юридические объекты: произведение, исполнение, база данных и информация

Первая задача аудитора — классифицировать не файл, а содержащиеся в нём охраняемые объекты. Один веб-документ способен содержать авторский текст, фотографию, иллюстрацию, оформление базы данных, фонограмму, товарный знак и персональные сведения. Интерес к «публичным данным» не устраняет интеллектуальные права на творческую форму их представления. Факты сами по себе в общем случае не охраняются авторским правом как литературное произведение, но отбор, расположение и оригинальное описание могут охраняться. Отдельный sui generis режим изготовителя базы данных защищает инвестиции в получение, проверку и представление содержимого. [6][7]

Особенно важно различать право на исходную фотографию и права лиц, чьи образы на ней запечатлены; право на музыкальную композицию и на конкретную запись; право автора статьи и права издателя прессы; право на программный код и условия лицензии репозитория. Наличие открытого URL подтверждает техническую доступность, но не удостоверяет ни переход прав, ни согласие на любое использование. Договор о доступе к архиву также не обязательно включает право на извлечение тысяч материалов для разработки коммерческой модели. Юридическая карта набора данных должна отражать множественность прав и возможных управомоченных лиц.

В спорах о генеративных моделях полезно различать промежуточное копирование и воспроизведение существенных элементов в выходном продукте. Для первого важны основания доступа и исключения TDM, для второго — сходство охраняемой формы, способы предоставления результата и возможная применимость иных исключений. Отсутствие имени автора в выдаче не исключает нарушения, а наличие узнаваемой тематики само по себе его не доказывает. Экспертиза должна фиксировать подлежащие сравнению элементы, а не оценивать сходство по субъективному впечатлению.

Вид контента	Дополнительный правообладатель или режим	Специальный контроль
Книга и перевод	Права автора, переводчика, издательские лицензии	Раздельные версии, языки, объём разрешений
Фотография с людьми	Авторское право и возможные права личности/данных	Источник, согласия и пределы публикации
Текст песни и звукозапись	Права на текст, музыку, исполнение и фонограмму	Раздельный анализ текста и аудио
База судебной практики	База данных, отдельные редакционные материалы	Права изготовителя и условия доступа
Программный репозиторий	Право на код, лицензии зависимостей	Совместимость условий и сохранение уведомлений

3. Цепочка жизненного цикла: где именно возникают правовые риски

Типовая цепочка обучения начинается с идентификации URL, каталога или посредника. Далее следуют загрузка, распаковка, преобразование форматов, фильтрация и дедупликация; затем материал распределяется между обучением, оценкой и настройкой. В процессе могут создаваться временные и устойчивые копии, эмбеддинги, контрольные выборки, индексы и журналы. Ошибка в правовом основании на раннем этапе наследуется последующими производными массивами и не исчезает лишь потому, что исходный файл удалён из открытого хранилища. Напротив, отсутствие исходного файла осложняет доказательство законности. [1][6]

Рекомендуется привязать каждый переход к проверяемому артефакту: манифесту источников, дате и способу доступа, снимку условий лицензии, версии робота-сборщика, решению о применимом исключении, статусу opt-out и перечню продуктов, получивших данные. Контроль должен охватывать и поставщиков готовых датасетов. Передача ответственности по договору не делает незаконное содержимое законным; гарантия поставщика только перераспределяет часть договорного риска и может оказаться экономически бесполезной при отсутствии доказательств и страхового покрытия.

В распределённой инфраструктуре одна и та же сущность может появляться как исходный файл, извлечённый текст, запись после очистки и фрагмент для дообучения. Поэтому обычный URL не служит устойчивым идентификатором. Следует хранить криптографический хеш конкретного файла, идентификатор произведения при наличии, родительский идентификатор преобразования, сведения об источнике и связь с версией набора. Такая модель обеспечивает прослеживаемость без обязательной публикации полного списка произведений и учитывает коммерческую тайну, но не позволяет скрывать объекты от уполномоченного проверяющего без эквивалентного доказательства.

Этап	Типичный риск	Обязательный след проверки
Обнаружение	Смешение общедоступности и разрешения	URL, оператор, время, версия правил доступа
Получение	Нарушение ограничений или прав третьих лиц	Основание доступа и технический журнал
Подготовка	Потеря атрибуции и opt-out	Граф преобразований, хеши, неизменяемый реестр
Обучение	Использование неразрешённых категорий	Версия набора, принятое решение, исключения
Релиз и обновление	Отсутствие сведения о новых источниках	Связь версии модели и публичной сводки
Выдача	Воспроизведение существенной части	Протокол безопасного теста и

Этап	Типичный риск	Обязательный след проверки
		реагирования

4. Исключительные права в праве ЕС: воспроизведение и сообщение публике

Наряду с национальными исключениями необходимо учитывать международный «трёхступенчатый тест» статьи 13 TRIPS: специальные случаи, отсутствие конфликта с обычной эксплуатацией и отсутствие необоснованного ущерба законным интересам правообладателя. Он не подменяет конкретные условия DSM, но служит рамкой их интерпретации. [19]

Директива 2001/29 закрепляет широкое право разрешать или запрещать прямое и косвенное, временное и постоянное воспроизведение произведения целиком или частично. Для обучения это означает, что даже технически промежуточная копия не должна автоматически рассматриваться как юридически нейтральная. Требуется либо разрешение правообладателя, либо конкретное применимое исключение, при соблюдении его условий. Право сообщения публике и доведения до всеобщего сведения оценивается отдельно, если модель или сервис позволяет воспроизводить охраняемые фрагменты. [6]

Необходимо установить, где совершались копирование и иные действия, какое национальное право применяется, кто контролировал процесс и какие права могли быть нарушены. В трансграничной системе разработчик может быть зарегистрирован в одной стране, хранить данные в другой, обучать модель в третьей и предоставлять её пользователям во множестве государств. Это не сводится к утверждению, что «право страны сервера» решает весь спор. Территориальность исключительных прав, правила о применимом праве и судебной юрисдикции требуют отдельного заключения по фактам.

Наконец, правообладатель должен доказать соответствующий охраняемый интерес, а истец — связать спорное использование с конкретным действием ответчика. Аудитор не подменяет суд, но способен организовать материал: сопоставить экземпляры, зафиксировать условия доступа, построить техническую хронологию и указать неопределённости. Корректное заключение предусматривает как доказательства потенциального нарушения, так и доказательства разрешённого использования, включая лицензии, добросовестно сделанные резервирования или применимые исключения.

Юридический элемент	Проверочный вопрос	Предел вывода
Оригинальность	Есть ли охраняемая творческая форма?	Не охранять сами факты и идеи
Воспроизведение	Создана ли копия охраняемого элемента?	Различать исходник, производный фрагмент, параметры модели
Исключение	Подпадает ли действие под конкретное правило?	Не переносить исключение на иной способ использования
Сообщение публике	Воспроизводится ли материал внешним пользователям?	Анализировать конкретный выход и механизм доступа

5. Статья 3 Директивы DSM: исследовательский TDM

Статья 3 Директивы 2019/790 предусматривает обязательное для государств-членов исключение для воспроизведений и извлечений при text and data mining в целях научных исследований исследовательскими организациями и учреждениями культурного наследия. Необходим законный доступ; копии должны храниться с соответствующим уровнем безопасности и могут сохраняться для научных исследований, включая проверку результатов. Для этой конструкции важен не лозунг о пользе технологии, а статус субъекта, цель, правомерность доступа и рамки деятельности. [1]

Коммерческая компания не превращается в исследовательскую организацию лишь потому, что нанимает учёных или публикует статью. Аналогично наличие университетского партнёра само по себе не обеспечивает всей технологической группе иммунитет. В каждом проекте следует оценивать действительные роли, контроль над результатами и условия, при которых коммерческий участник получает доступ к исследовательским копиям. При смешанных инициативах рекомендуется разделять исследовательскую среду и коммерческий цикл обучения технически и договорно.

Это исключение нельзя описывать как универсальное разрешение публично распространять копии корпуса или вводить обученные модели в коммерческий оборот без дополнительных проверок. Результаты исследования и сами копии подпадают под разные режимы. Для аудита нужен отдельный журнал сохранения копий, обеспечивающий проверяемость результатов и одновременно ограничивающий доступ лицам, не участвующим в научной задаче.

Условие	Подтверждение	Красный флаг
Статус выгодоприобретателя	Устав, фактическая деятельность, управление	Компания названа университетом только в рекламе
Научная цель	Утверждённый протокол и вопросы исследования	Разработка коммерческого API без исследовательских границ
Законный доступ	Лицензии, подписка, открытый источник	Скачивание с пиратского зеркала
Хранение	Ролевая модель, сроки и защита	Неограниченный доступ подрядчиков

6. Статья 4 Директивы DSM: общее исключение и право резервирования

Статья 4 допускает воспроизведения и извлечения законно доступных произведений для TDM не только в научных целях. Однако правообладатель вправе прямо зарезервировать соответствующее использование надлежащим способом. Для общедоступного онлайн-контента статья прямо приводит машиночитаемые средства как характерный механизм резервирования. Нельзя заменить проверку закономерности фразой «мы скачали материалы, находившиеся в Google»: законный доступ и отсутствие действенного резервирования должны подтверждаться применительно к моменту и маршруту использования. [1]

Нельзя также утверждать, будто любой набор правил robots.txt автоматически разрешает либо запрещает все виды обучения. Файл роботов создавался прежде всего для управления доступом автоматизированных агентов; юридическая оценка конкретного сигнала зависит от его содержания, адресата, времени, национальной реализации и других сведений о воле правообладателя. Правильный технический подход — сохранять исходную версию машиночитаемого сигнала и результат его интерпретации crawler, фиксируя расхождения и проверяя ограничения доступа на уровне произведения и домена.

Резервирование может быть опубликовано после первоначального сбора, либо произведение может находиться на нескольких площадках с различными условиями. В таких случаях требуется хронология и юридический анализ последующих операций; нельзя обещать ни автоматическое обратное действие всякого нового opt-out, ни его полное отсутствие для любых последующих копирований и дообучений. Разработчик должен как минимум предотвращать будущий неправомерный сбор и иметь процесс для спорных активов, а обладатель прав — понятный канал заявления и подтверждения своих полномочий.

Ситуация	Процедура	Ошибочный вывод
Машиночитаемый запрет до сбора	Зафиксировать сигнал, не включать в общий сбор без правового основания	«Сайт общедоступен, значит запрета нет»
Сигнал противоречив	Передать на юридическую проверку, сохранить снимки	«Любой robots.txt полностью заменяет закон»
Opt-out поступил после сбора	Привязать к датам, версиям и будущим действиям	«Всегда задним числом» или «никогда не влияет»
Несколько правообладателей	Проверить объём полномочий заявителя	«У редактора есть все права на материал»

7. Французская транспозиция: статья L122-5-3 CPI

Во Франции статья L122-5-3 Кодекса интеллектуальной собственности разграничивает научный TDM и другие способы автоматизированной обработки текстов и данных. Для общего режима закон допускает копии произведений при законном доступе за исключением ситуации надлежащего возражения автора, в частности в машиночитаемой форме в отношении онлайн-контента. Копии должны храниться безопасно и уничтожаться по завершении соответствующего анализа согласно предусмотренным нормой условиям.

Для научного режима предусмотрены специальные правила сохранения результатов в целях проверки. [21]

Для французского субъекта принципиально важно не просто сослаться на директиву, но сверить её с применимым национальным правом и фактическими условиями. Значение имеют квалификация актора, характер произведения, категория охраняемых прав, техническое местонахождение действий и применимая лицензия. Например, лицензия фотобанка может разрешать редакционное использование изображения и одновременно запрещать его использование для обучения машинных моделей. Исключение нельзя анализировать в отрыве от способа получения контента.

Аудитор Observatoire ARGA, оценивающий французскую практику, должен документировать: дату копирования, доказательство доступа, содержание opt-out, копии правил использования, запись о правовом основании и дату удаления промежуточных копий. Отсутствие этих сведений означает дефицит доказательств, а не автоматически установленное нарушение. Такой подход позволяет готовить достоверные выводы в консультациях, не приписывая ARGA полномочий суда или национального надзорного органа.

Действие	Критерий по французскому праву	Документ
Научный анализ	Статус исследовательского субъекта и научная цель	Устав, проектная документация, правила доступа
Общий TDM	Законный доступ и отсутствие надлежащего возражения	Снимок условий и журнал signals
Сохранение копий	Различать научный и общий режим	Политика хранения и контроль удаления
Передача поставщику	Проверить права, договор, территорию	Лицензионный реестр и DPA при персональных данных

8. Базы данных, издательские каталоги и массовое извлечение

Даже когда отдельные факты не обладают творческой оригинальностью, извлечение большого количества структурированных записей способно затрагивать sui generis право изготовителя базы данных. Директива 96/9 защищает значительные инвестиции в получение, проверку и представление содержимого базы. Объектом анализа становятся масштабы извлечения, качественная существенность выбранной части, регулярность повторных операций и эффект для обычной эксплуатации базы. [7]

Технологическая команда иногда считает, что копирование метаданных — названий дел, аннотаций, аннотационных категорий или библиотечных карточек — по определению свободно. На деле различные элементы могут одновременно подпадать под режим базы данных, авторское право на редакционную аннотацию и договорный контроль над API. Важна также разница между официальными государственными источниками, редакционными сервисами и коммерческими агрегаторами: право на исходный публичный акт не всегда покрывает чужую коммерческую систематизацию.

При оценке поставщика следует требовать карту прав на сам набор, а не только справку о праве на отдельные записи. Для каждого коллектора нужны критерии извлечения, суммарный объём, скорость, периодичность и наличие независимой охраняемой структуры. Если один и тот же набор перепродан по цепочке посредников, необходимо восстановить как права предоставления, так и техническое происхождение. Коммерческая сделка с посредником без права сублицензирования не ликвидирует дефект.

Признак	Что собирается	Юридический риск
Массовая выгрузка каталога	Записи, аннотации, структура	Существенное извлечение из базы
Периодический delta-crawl	Новые или изменённые записи	Повторяемость операций, совокупный объём
Обогащение метаданных	Дополнительные описания редактора	Самостоятельное авторское право
Перепродажа	Лицензия поставщика, права на сублицензию	Разрыв цепочки прав

9. Статья 53 AI Act: политика соблюдения авторского права

Согласно статье 53(1)(c) Регламента (ЕС) 2024/1689 поставщики моделей общего назначения должны внедрить политику соблюдения права ЕС об авторском праве и смежных правах, в частности выявлять и соблюдать резервирование прав в соответствии со статьёй 4(3) DSM с применением современных технологий. Речь идёт не о праве автоматически использовать любые произведения, а об организационной обязанности поставщика в дополнение к существующим нормам авторского права. Эта обязанность охватывает разработчика в зависимости от его роли и статуса поставщика модели GPAI на рынке Союза. [2][8]

Исполнимая политика не ограничивается обещанием уважать интеллектуальную собственность. Она должна перечислять источники, правила доступа, поддержку машиночитаемых opt-out, контроль посредников, исключение нелегальных источников, порядок обращения правообладателей, механизм реагирования на выходное воспроизведение и назначенных ответственных. Проверить её можно через операционные доказательства: решения об остановке сбора, отчёты о тестах crawler, статистику спорных обращений и журнал обновлений. Если политика существует только как общедоступный PDF, но не отражена в коде и договорах, её доказательная ценность минимальна.

Обязанности по п. 1(c) и 1(d) сохраняются для открытых моделей: исключение, предусмотренное пунктом 2 статьи 53 для некоторых открытых моделей, касается перечисленных там документов по пунктам 1(a) и 1(b), а не общей обязанности иметь copyright policy и публичную сводку обучающего контента. Для моделей с системным риском действует отдельная рамка, включая дополнительные обязанности. Следует также разграничивать дату применения требований GPAI для новых моделей и переходный период в отношении ранее размещённых моделей. [2][3][8]

Проверяемый элемент политики	Сильное доказательство	Недостаточно
Обработка opt-out	Логи принятия сигнала и исключения URL/asset	Ненаполненная веб-форма
Проверка датасета	Перечень источников, статусы, выборка	Гарантия «всё из открытого интернета»
Обновление моделей	Соответствие version ID и copyright policy	Политика без версии и владельца
Жалобы правообладателей	Время реакции и результат по каждому делу	Универсальный отказ без анализа

10. Статья 53 AI Act: публичная сводка обучающего контента

Статья 53(1)(d) AI Act требует составить и опубликовать достаточно подробную сводку использованного для обучения контента по шаблону, предоставленному AI Office. Европейская комиссия опубликовала шаблон 24 июля 2025 года и пояснила его структуру: общие сведения о модели, перечень видов источников и существенные аспекты обработки. Эта публикация служит повышению прозрачности и возможности реализации прав, но не требует автоматически выкладывать весь корпус, включая каждое защищённое произведение и коммерческие секреты, в общедоступный доступ. [2][3][17]

В официальном FAQ Комиссия требует публикации сводки не позднее размещения соответствующей модели на рынке Союза; при дальнейшей тренировке предусмотрено обновление как минимум каждые шесть месяцев либо раньше при существенном изменении сведений. Для отдельных моделей, выпущенных до 2 августа 2025 года, действует переходный срок до 2 августа 2027 года. Эти условия следует связывать с точным юридическим статусом поставщика и версии. Публичная сводка может содержать обоснованные пробелы в отношении старых моделей, но пробелы не следует маскировать как полное знание корпуса. [3]

При аудите полезно проверять сводку на соответствие внутреннему реестру происхождения: отражены ли источники лицензируемого контента, публично доступные наборы, веб-сбор, пользовательские данные, синтетический материал и изменение долей после fine-tuning. Высокая точность цифр невозможна, если их исходная оценка не объяснена; вместо ложной десятичной точности необходимы диапазоны, формула

отбора и документированная неопределённость. Сама сводка не заменяет доказательство законного доступа и отсутствия запрета TDM на уровне конкретного произведения.

Вопрос аудитора	Публичный слой	Конфиденциальный слой
Какие типы источников использованы?	Категории, основные наборы, доли/диапазоны	Asset-level реестр, лицензии, хеши
Кто ответственен за сбор?	Общая схема провайдеров и ролей	Контракты и протоколы контроля подрядчиков
Как учитывается opt-out?	Описание процедуры и каналов	Полные журналы сигналов и исключений
Что изменилось после обновления?	Дата и характер существенного изменения	Delta-манифест и решение о новом цикле

11. Добровольный GPAI Code of Practice и границы его доказательности

Кодекс практики для моделей общего назначения содержит главы по прозрачности и авторскому праву, а также отдельную главу по безопасности для моделей с системным риском. Европейская комиссия описывает его как добровольный механизм содействия исполнению требований AI Act. Присоединение к кодексу либо заявление о соответствии полезно для оценки зрелости процедур, но само по себе не является судебным установлением правомерности всей истории обучения модели. В отличие от него обязательный шаблон публичной сводки следует из нормативной конструкции статьи 53(1)(d). [4][17][18]

Независимый оценщик должен разделять три вопроса: подписал ли поставщик кодекс, внедрены ли фактически его обещанные меры и законны ли конкретные операции с объектами авторского права. Невыполнение определённой меры кодекса не всегда означает прямое нарушение статьи AI Act, если есть альтернативный способ исполнения; аналогично формальная подпись не делает нарушения невозможными. Рекомендуется использовать «матрицу соответствия», связывающую каждое добровольное обязательство с конкретным доказательством и отдельной обязательной нормой, если такая имеется.

Особенно важно указывать наличие существенных ограничений доступа аудитора. Если компания предоставляет только публичную сводку и маркетинговые заявления, заключение должно называться обзором раскрытой информации, а не полноценным аудитом происхождения. Независимость заключается не в громком названии, а в возможности получить отрицательный результат, опубликовать оговорки и не зависеть от коммерческого исхода проверки.

Утверждение	Что реально подтверждает	Что не подтверждает
«Подписан Кодекс GPAI»	Взяты добровольные обязательства	Легальность каждого обучающего файла
«Сводка опубликована»	Выполнено требование публичного раскрытия при надлежащем объёме	Полноту и подлинность скрытого реестра
«Аудитор не нашёл нарушения»	Нарушений не выявлено в определённой выборке	Абсолютное отсутствие нарушений
«Корпус имеет open licence»	Наличие некоторых лицензионных условий	Что все объекты свободны от иных ограничений

12. Сроки, переходные правила и институциональные полномочия

Для поставщиков GPAI обязанности по статье 53 начали применяться с 2 августа 2025 года, а полномочия Комиссии по их обеспечению — с 2 августа 2026 года; для моделей, размещённых на рынке до первой даты, установлены специальные переходные сроки, отражённые в статье 111 и разъяснениях Комиссии. Вопрос о конкретной версии модели и моменте её размещения на рынке обязателен: нельзя автоматически применять один и тот же дедлайн ко всем продуктам или считать последующее дообучение всегда юридически несущественным. [2][3][8]

AI Office обеспечивает надзор за соответствующими обязанностями поставщиков GPAI в рамках AI Act, тогда как авторско-правовые споры, лицензирование, компенсация вреда и конкретные способы защиты

реализуются по применимому авторскому, процессуальному и национальному праву. Из факта наличия полномочия затребовать документацию по AI Act не следует, что гражданский истец автоматически получает любой закрытый набор данных в полном объёме. И наоборот, публикация публичной сводки не отменяет права суда и иных компетентных органов требовать сведения при наличии процессуального основания.

Для внутреннего контроля рекомендуется вести отдельные календари: один по регуляторным требованиям AI Act, второй — по срокам лицензий, отказов и судебных претензий, третий — по циклам технического пересмотра корпуса. Смешивание этих календарей ведёт к ошибкам, например подмене обязанности обновлять summary заявлением о «повторном получении согласия каждого автора раз в полгода». Такая обязанность в приведённой формулировке не следует из шаблона Комиссии.

Контрольная дата	Значение	Контрольное действие
2 августа 2025	Начало применения обязанностей GPAI для соответствующих новых моделей	Проверка статуса и версии модели
2 августа 2026	Полномочия Комиссии по обеспечению GPAI-требований	Проверить готовность доказательств и процедуры
2 августа 2027	Переходный ориентир для ранее размещённых GPAI моделей	Документировать применимость и пробелы
Каждые шесть месяцев или ранее	Обновление публичной сводки при новых обучающих данных по разъяснению Комиссии	Сопоставить версию модели с delta-манифестом

13. Машиночитаемые запреты и техническая архитектура opt-out

Резервирование прав должно поступать к технической системе не как необъяснимая галочка, а как событие с источником, временем, объёмом и степенью доверия. Возможны сигналы на уровне сайта, URL, коллекции, конкретного файла и стандартизированных метаданных. Технический формат необходимо оценивать в правовом контексте статьи 4(3) DSM и французской реализации, а не предполагать, что существует единственный юридически достаточный тип сигнала для всех видов контента. В частности, сигналы поисковой индексации не тождественны автоматически сигналам, адресованным обучению моделей. [1][21]

Процесс должен сохранять запрос робота, полученный ответ, хеш снимка правил, правила обработки переадресаций и повторных проверок, а также причину допуска или блокировки. Отдельной проблемой являются зеркала сайтов, CDN, архивные копии и перепубликации: правообладатель может заявить opt-out на первоисточнике, а копия произведения продолжит циркулировать у посредника без того же сигнала. Следовательно, реестр должен поддерживать не только URL, но и идентичность объекта, включая визуальные и текстовые отпечатки, с осторожностью в отношении ложных совпадений.

Для публичной консультации предлагается проверяемый минимальный стандарт: общедоступная спецификация сигналов, возможность выразить возражение без платного посредника, журналы обнаружения и разрешения коллизий, тесты на соблюдение запретов при повторном сборе. При этом стандартизация должна допускать доступность для небольших издателей и авторов, не имеющих технических команд. Нельзя строить режим, в котором техническая сложность фактически лишает правообладателя права выбора.

Канал сигнала	Технический риск	Контроль аудитора
Правила сайта	Робот использует устаревший кэш	Временная метка, TTL, повторная проверка
Метаданные файла	Они теряются при конвертации	Проверка до и после преобразования
Реестр произведений	Ошибка привязки правообладателя	Процедура верификации полномочий
Уведомление напрямую	Неполный охват производных наборов	Asset lineage и подтверждение маршрутизации

14. Chain of title: как устанавливается право на конкретное произведение

Поставщик корпуса может гарантировать, что все материалы им созданы или законно лицензированы, однако это договорное утверждение не заменяет проверку цепочки прав. Для художественного изображения требуется установить автора и вероятные договоры уступки или лицензирования; для музыкального трека — права на композицию и запись; для книги — отношения автора, переводчика, издателя и цифрового дистрибьютора. Важно разграничивать право выдавать контент пользователям и право разрешать его копирование для машинного обучения. [6][7]

Аудит начинается с выбора уровня гранулярности. Для юридического анализа желательно иметь идентификатор произведения, конкретной версии, категории прав, титульного правообладателя и лица, предоставившего лицензию. Если правовой цепочки нет, неопределённость должна маркироваться, а объект выводиться из высокорисковых потоков до подтверждения. Выборочное тестирование не должно ограничиваться крупными известными издателями: небольшие авторы и архивные материалы часто имеют более сложные или хуже оцифрованные договоры.

Особая проблема — лицензии с правом сублицензирования. Поставщик может владеть лишь правом на показ произведения в собственном интерфейсе и не иметь полномочий передавать массив иному разработчику. Рекомендуемый контроль: запрос заверенного описания прав, проверка выборки исходных договоров, сверка исключений и ограничений по территории, сроку, виду использования и моделям. Сомнение в праве не устраняется высоким коммерческим рейтингом поставщика.

Элемент цепочки	Подтверждение	Последствие пробела
Авторство и первоначальные права	Договоры, реестры, метаданные создания	Неопределённость статуса актива
Передача/лицензия	Объект и конкретные способы использования	Невозможность доказать право на обучение
Сублицензия	Прямое полномочие передавать права	Риск незаконности downstream использования
Территория и срок	Юрисдикции и даты	Расхождение с географией обучения и размещения
Право на обновления	Условия новых версий произведения	Недопустимость автоматической преемственности лицензии

15. Лицензирование обучающих массивов: границы и компенсация

Коммерческая лицензия даёт более определённую основу использования при корректном объёме и полномочиях сторон. Однако само слово «лицензия» ничего не сообщает о разрешённых способах обработки: необходимо явно оценивать копирование, предобработку, обучение, fine-tuning, хранение промежуточных копий, передачу подрядчику, использование для оценки и возможное повторное обучение. Некоторые лицензии требуют атрибуции, ограничивают категории моделей либо запрещают использование в продуктах, конкурирующих с правообладателем. Эти ограничения нельзя игнорировать при формировании смешанных корпусов.

Для переговоров следует формировать реестр разрешений на уровне asset, а не единую строку «корпус лицензирован». Оплата может зависеть от объёма, срока, количества продуктов, территории или подтверждённых случаев применения. Экономическое сравнение разных моделей вознаграждения возможно лишь при указании того, кто оплачивает административные издержки, как распределяются выплаты между соавторами и что происходит с ранее использованным контентом после расторжения.

Observatoire ARGA предлагает отдельно оценивать доступность компенсации малым авторам. Коллективная лицензия уменьшает транзакционные издержки, но требует ясности о репертуаре, полномочиях представителя, праве выхода и распределении сумм. Не следует объявлять любую модель коллективного управления обязательной для всех без законодательного основания. Практическая рекомендация — сохранять конкуренцию добровольных лицензий и доказательное право автора проверить факт включения в охраняемый набор, не требуя публичного раскрытия всего закрытого корпуса. [5][9]

Условие лицензии	Вопрос договору	Проверка исполнения
Объём копирования	Разрешены ли технические и устойчивые копии?	Реестр операций и производных форм
Обучение и оценка	Какие модели, этапы и задачи охвачены?	Манифест версий
Сублицензирование	Допущен ли подрядчик/партнёр?	Перечень downstream субъектов
Вознаграждение	Модель расчёта и контрольные данные	Аудиторская формула, отчётность
Прекращение	Что делать с копиями и существующими весами?	План перехода и оценка технической осуществимости

16. Посредники, готовые датасеты и распределённая ответственность

В цепочке обучения часто участвуют независимые сборщики, владельцы облачных хранилищ, площадки продаж датасетов, разработчики моделей и поставщики приложений. Важно не переносить единую характеристику на всех: один участник определяет источники и юридические основания, второй только хранит копии, третий модифицирует модель для нового рынка. Ответственность зависит от конкретного действия, договора и применимого права, а не от выбранного торгового названия роли. [2][8]

Аудитор должен построить граф поставок, выделив узлы принятия решений: кто разрешил сбор, кто принимал opt-out, кто имел возможность удалить спорный материал, кто запустил обучение и кто публикует summary. Нередко поставщик заявляет наличие «чистого» датасета, опираясь на собственный стандарт без раскрытия критериев. Разумная проверка должна включать репрезентативную выборку первичных источников, документы о лицензиях и возможность оспаривать ошибочную классификацию.

Договоры должны разграничивать обязанность предоставить доказательства и обязанность возместить убытки. Даже широкое обязательство возместить претензии не помогает, если отсутствуют журнал происхождения или исходные метаданные. Обязательство о сотрудничестве при регуляторной проверке и сохранении документов нередко важнее абстрактной декларации абсолютной юридической чистоты. Важен также план, если поставщик прекращает деятельность, а модель продолжает распространяться.

Субъект	Возможный контроль	Запрашиваемое доказательство
Сборщик	Источники, правила обхода, момент копирования	Crawl-логи, скриншоты условий
Агрегатор	Сборка и перепродажа	Перечень upstream лицензий
Разработчик модели	Обучение и релиз	Манифест данных и версия модели
Интегратор	Fine-tuning, RAG, API	Дополнительные наборы и пределы разрешений
Внешний аудитор	Метод проверки	Выборка, ограничения, конфликты интересов

17. Обучение, память модели и выходное воспроизведение

Модель может воспроизвести существенный охраняемый фрагмент, несмотря на то что её архитектура не является обычной базой последовательных файлов. Технически это связывают с memorization и regurgitation, но юридическая квалификация зависит от конкретного права и доказательств. В постановлении суда Мюнхена I от 11 ноября 2025 года по делу GEMA против компаний группы OpenAI суд счёл в исследованных обстоятельствах воспроизводимыми в модели и её ответах части песенных текстов и удовлетворил существенную часть требований. Суд отдельно указал, что решение не вступило в законную силу на дату сообщения. Этот результат нельзя безоговорочно переносить на любые модели или иные жанры. [11]

Для проверки нужны контролируемые тестовые запросы, фиксация версии модели и параметров генерации, сравнение выхода с документированным оригиналом, а также исключение случайной идентичности с широко известной неохраняемой формулой. Следует различать обычный тематический отклик, пересказ фактов, стилистическое сходство и протяжённое близкое воспроизведение охраняемой

формы. Простое указание процента совпадения без описания базы сравнения не годится для правового заключения.

Вопрос о роли поставщика и пользователя в создании спорного ответа также нельзя решить единым правилом. Имеет значение характер запросов, настройки защит, предсказуемость выдачи и возможность поставщика предотвращать воспроизведение. Для независимой процедуры необходимо хранить журнал тестов без открытой публикации длинных защищённых фрагментов, которые сами могут стать источником нового нарушения.

Тип результата	Метод тестирования	Возможный вывод
Фактический пересказ	Проверка охраняемой формы	Как правило, не достаточен для вывода о копировании
Короткая типовая фраза	Анализ оригинальности и распространённости	Низкая доказательная сила совпадения
Длинный уникальный пассаж	Побуквенное сопоставление, версия модели	Основание для предметной юридической оценки
Воспроизведение по простому запросу	Несколько независимых повторов и настройки	Повышенная предсказуемость риска

18. Техническая проверка memorization: воспроизводимость и пределы

Нельзя проводить аудит memorization одной удачной или неудачной подсказкой. Для достоверного результата выборку произведений нужно формировать до тестирования, разделяя обычные, редкие и высокорисковые фрагменты, а количество попыток и стратегию запросов фиксировать заранее. В противном случае эксперт способен непреднамеренно подобрать prompts, дающие желаемый результат, и создать впечатление высокой или нулевой склонности к воспроизведению. В отчёте отдельно указываются частота точных фрагментов, существенность совпадения и вариабельность ответа.

Для текстовых моделей полезно применять нормализованное сопоставление последовательностей, но автоматическая метрика не заменяет анализ оригинальности. Для изображения требуются критерии композиционного сходства и оценка характерных охраняемых деталей, для аудио — отдельные эксперты по записи и музыкальному произведению. Следует исключать тесты, при которых сам проверяющий вставляет весь защищённый оригинал в prompt, а затем считает его появление в output доказательством того, что модель запомнила его при обучении.

Результаты следует представлять с доверительными интервалами и оговорками. Если при 100 независимых попытках не обнаружено ни одного близкого воспроизведения, статистическая верхняя граница для вероятности события при упрощённой биномиальной модели составляет примерно 3% на уровне 95% (правило трёх), а не ноль. Это учебный пример, не реальная измеренная частота конкретной модели; независимость попыток и качество выборки требуют отдельного подтверждения.

Компонент протокола	Правило	Почему важно
Выборка	Зафиксировать до запуска	Предотвращает cherry-picking
Prompts	Отделить простые и целевые атаки	Различает обычное и намеренно вызванное поведение
Параметры	Версия, температура, контекст, дата	Обеспечивает воспроизводимость
Метрики	Точные фрагменты, качественная значимость	Избегает ложных выводов из одного процента
Вывод	Риск и пределы, а не абсолютная гарантия	Не превращает статистику в юридическую презумпцию

19. Удаление контента и модели: что можно обещать технически

После обнаружения неразрешённого произведения разработчик может удалить его из будущих обучающих датасетов и индексов, заблокировать конкретные выходные фрагменты, исключить доступные копии и начать исследования по machine unlearning. Но нельзя без испытаний утверждать, что из ранее обученных весов гарантированно устранено любое влияние произведения. Эффективность удаления

зависит от архитектуры, характера данных, стадии обучения и используемого метода. Следовательно, проект договорного положения «полное удаление из памяти модели за 24 часа» может обещать технически недостижимый результат.

Практическая политика должна различать категории: исходные файлы, промежуточные копии, датасеты будущих запусков, RAG-базы, кэши, опубликованные веса, эксплуатационные журналы и производные модели, созданные сторонними участниками. Для каждой категории указать фактическую возможность удаления, подтверждения и необходимые меры снижения остаточного риска. Нельзя путать удаление записи о запросе правообладателя с устранением самих данных.

В спорных случаях допускается приостановка использования, дополнительное дообучение, фильтрация выдачи и прекращение поддержки версии. Выбор меры должен быть пропорционален риску, доказательствам и доступным полномочиям, а не коммерческому удобству. Внешнему аудитору необходимо получить протокол тестирования после меры и право указать, что остаточная неопределённость не устранена. [11][12]

Объект	Относительная управляемость	Обязательное доказательство меры
Исходный файл	Обычно высокая при контроле хранилища	Логи удаления и резервных копий
RAG-индекс	Обычно высокая при управляемом re-index	Индекс версии до/после
Обучающий набор следующего цикла	Высокая при рабочем asset registry	Manifest exclusion
Параметры уже выпущенной модели	Не гарантирована	Результаты специальных испытаний и оговорка
Открытые копии третьих лиц	Ограниченная	Уведомления партнёрам и оценка остаточного риска

20. Цифровые доказательства: сохранение, цепочка хранения и экспертиза

В авторско-правовом споре о данных для ИИ особенно легко уничтожить важное доказательство, исправляя систему. Поэтому после правдоподобной претензии необходимо законно сохранить относящиеся к делу журналы доступа, сведения о версии корпуса, исходные условия лицензии, конфигурации crawler, набор тестовых запросов и версии выхода. Это не означает бессрочного хранения всех персональных данных: обязательства о доказательствах следует согласовывать с защитой данных и юридическим режимом удержания. [20]

Независимый специалист фиксирует время по согласованной временной зоне, происхождение файла, криптографический хеш, лицо, выполнявшее действие, способ получения, неизменяемый контрольный журнал и ограничения на доступ. Если скриншот страницы сделан после изменения её содержания, нужно указать, что это последующее состояние, а не подлинный снимок на дату обучения. Нельзя выдавать копию репозитория, полученную от заявителя, за доказательство того, что именно эта версия содержалась в исходном наборе.

В судебных и регуляторных процедурах необходимо разграничивать свидетельство эксперта, техническую экспертизу по поручению компетентного органа и внутреннее корпоративное расследование. ARGA может готовить независимые аналитические материалы в пределах компетенции и законного доступа, но не присваивать процессуальный статус эксперта суда и не заявлять о полном доступе к закрытым весам модели, если такого доступа нет. Доказательная сила зависит от раскрытых методов и воспроизводимости, а не от длины документа.

Доказательство	Фиксируемые поля	Возможное искажение
Скриншот правил сайта	URL, дата, браузер, архив	Снято после спорного сбора
Crawl-журнал	Агент, IP, временная метка, объект	Неполный охват хранилища
Manifest корпуса	Hash, version, dataset ID	Потеря связи при преобразованиях
Output теста	Prompt, версия, настройки, лог	Подмена выборки после наблюдения
Лицензия	Полный текст и приложения	Неизвестные ограничения сублицензии

21. США: fair use без универсальной индульгенции обучению

В США анализ fair use требует учёта конкретного использования, характера произведения, объёма и существенности копирования, а также влияния на реальный или потенциальный рынок. В определении от 23 июня 2025 года по делу Bartz v. Anthropic федеральный суд разделил три вида действий: использование книг для обучения модели, внутреннее оцифрование законно приобретённых печатных экземпляров и создание постоянной библиотеки пиратских копий. Суд признал fair use для первых двух категорий на представленных фактах, но не распространил такую защиту на постоянную пиратскую библиотеку. Это было промежуточное решение по конкретным вопросам, а не общая лицензия на обучение. [12]
20 июля 2026 года другой судья того же федерального суда окончательно утвердил групповое мировое соглашение в этом деле, вынес решение на его условиях и прекратил производство с отказом от повторного предъявления охваченных соглашением требований. Это процессуальное завершение по соглашению, не пересмотр универсального теста fair use и не решение о законности всякой модели. Поэтому при цитировании Bartz необходимо различать июньское определение 2025 года и окончательное утверждение соглашения в июле 2026 года. [22]

В деле Kadrey v. Meta 25 июня 2025 года суд рассмотрел встречные ходатайства о частичном суммарном решении, применив fair use к определённому копированию в рамках доказательств, представленных сторонами. Суд указал на недостаточную доказанность предполагаемого рыночного ущерба в этом процессуальном контексте, в частности возможного замещения или размывания рынка. Нельзя цитировать это определение как окончательный иммунитет от всех требований, связанных с обучением, либо как универсальное правило для других корпусов и рынков. Отдельные вопросы и притязания могут иметь другую процессуальную судьбу. [13][23]
Доклад U.S. Copyright Office 2025 года содержит аналитическую оценку и рекомендации, но не является законом, судебным решением либо общим разрешением на обучение без проверки правового основания. [10]

Для транснациональной модели это означает невозможность применять американскую доктрину fair use как прямую замену европейским исключениям TDM. Практическая матрица должна проверять отдельно страну действия, применимое право, способ получения контента, рынок правообладателя и наличие действующего лицензионного механизма. Решение коммерческой команды «обучение можно делать в США, значит законно в ЕС» юридически недостаточно и требует консультации профильных юристов.

Американский материал	Установленный узкий смысл	Недопустимая экстраполяция
Bartz v. Anthropic (2025)	Различная квалификация обучения и пиратской библиотеки	«Пиратские источники не имеют значения»
Kadrey v. Meta (2025)	Существенность доказанного рыночного вреда	«Любое коммерческое обучение fair use»
US Copyright Office Part 3	Экспертный анализ политики и права	«Новый обязательный закон США»
EU DSM Articles 3/4	Особые европейские условия TDM	«Полный эквивалент американского fair use»

22. Великобритания: итоги консультации марта 2026 года

Правительство Великобритании 18 марта 2026 года опубликовало доклад по авторскому праву и искусственному интеллекту, предусмотренный национальным законодательством. В нём зафиксировано, что изначально предпочтительная идея широкой исключительной нормы с opt-out больше не является предпочтительным курсом правительства после сильного сопротивления, прежде всего со стороны творческих индустрий. Вместо объявления принятого закона правительство указало на необходимость дополнительного изучения доказательств, рынков лицензирования и технологической прозрачности. [9]

Поэтому по состоянию на дату исследования нельзя подавать обсуждавшийся британский opt-out как уже введённое универсальное исключение для коммерческого обучения. Следует анализировать действующее национальное право, договоры и существующие исключения, отдельно отмечая возможные

законопроекты и консультации. Британский доклад полезен и в методологическом отношении: он описывает экономические неопределённости, трудности небольших создателей контента и возможную роль прозрачности для защиты прав.

Для ARGA важен вывод о том, что эффективность резервирования прав нельзя оценить без измерения реальной доли произведений, по которым автор способен выразить волю, и процента технического соблюдения сигналов. Правовой режим, выгодный крупнейшим обладателям каталогов, может оставлять малым создателям недоступные затраты. Этот риск должен быть включён в оценку будущей политики ЕС и международных консультационных позиций.

Вопрос	Подтверждённый статус в марте 2026	Что требуется изучить дальше
Широкое исключение с opt-out	Больше не предпочтительный вариант правительства	Эффект на авторов и разработчиков
Прозрачность входных данных	Направление дальнейшей работы	Уровень детализации и стоимость соблюдения
Лицензирование	Существующий и развивающийся рынок	Доступность малым участникам
Территория действия	Национальный анализ отдельно от ЕС	Трансграничные потоки контента

23. Сингапур и сравнительный подход к computational data analysis

Сингапурский Copyright Act 2021 содержит отдельное регулирование computational data analysis: статья 243 описывает анализ как идентификацию и извлечение информации или использование произведений в качестве примеров для совершенствования компьютерной программы, а статья 244 устанавливает условия допустимого копирования. Конструкция не должна пересказываться как безусловная свобода скачивать любые произведения из любого источника: закон имеет требования, которые должны проверяться по тексту и фактам. [15]

Сравнительный анализ целесообразен только по одинаковым правовым действиям. Например, допустимость создания аналитических копий не автоматически означает право распространять оригинальные произведения либо игнорировать обязательства персональной конфиденциальности. При использовании иностранного набора разработчик должен определить, в какой юрисдикции были созданы копии, где затем происходило обучение и какие права заявляются на рынке целевого государства. Различия национальных исключений способны породить конфликт между законностью upstream обработки и downstream распространением модели.

В международной экспертной позиции ARGA важно избегать «выбора наиболее удобной страны» как синонима соблюдения права. Более устойчивой практикой является унифицированная карта происхождения и разрешений, к которой добавляются конкретные национальные квалификации. Это обеспечивает возможность пройти аудит даже при изменении юрисдикции и снижает затраты на повторную проверку всей цепочки.

Критерий сравнения	ЕС	Сингапур
Основная конструкция TDM	Исключения статей 3 и 4 DSM	Computational data analysis, Copyright Act 2021
Публичный opt-out	Значим для режима статьи 4	Требуется анализ отдельных условий национального закона
Доказательство доступа	Правомерный доступ	Проверить предусмотренные национальным законом предпосылки
Трансграничность	Учитывать право иных государств	Не приравнивать внутреннее исключение к мировой лицензии

24. GDPR, личные фотографии и двойная правовая проверка

Обучающий набор может одновременно содержать охраняемые авторским правом произведения и персональные данные. Успешная проверка TDM не отменяет обязанности соблюдать GDPR, если он

применяется. Например, журналистская фотография может быть лицензирована издателем для определённой цели, но содержать идентифицируемых людей; запись о судебном деле может быть общедоступной по национальному правилу публикации, однако включать сведения об уголовном преследовании и иных чувствительных обстоятельствах. В таких случаях нужны самостоятельные основания обработки, прозрачность, точность, минимизация и применимые специальные ограничения. [20][16]

Риск особенно высок при обучении генеративных юридических помощников на судебных архивах, расследованиях и публикациях о предполагаемых преступлениях. Авторско-правовая лицензия на текст справки не разрешает автоматически повторную обработку персональных сведений о подозреваемом. В случае граждан, которые не являются публичными деятелями, применение сведений в глобальном масштабе может создать несоразмерное вторичное распространение. Нужен раздельный правовой реестр IP и data protection, связанный единым asset ID.

Технический аудитор должен проверять назначение массива и характер данных до их загрузки в внешнюю модель. Даже если для авторского права достаточно установленной лицензии, передача текстов иностранному API может порождать отдельные требования к персональным данным, тайне, конфиденциальности и международной передаче. Заключение должно содержать два разных результата: оценку интеллектуальных прав и оценку допустимости обработки персональных сведений, не сводя одну к другой. [16][20]

Ситуация	Авторско-правовой вопрос	Вопрос защиты данных
Лицензированная фотография лица	Разрешено ли обучение?	Законна ли обработка изображения человека?
Судебный материал	Права на редакционный текст/базу	Статья 10 GDPR и публичность производства
Письмо клиента	Права автора текста	Конфиденциальность и правовое основание обработки
Публикация в открытой сети	TDM и opt-out	Принципы GDPR сохраняются при применимости

25. Метод независимого аудита: мандат, критерии и выборка

Независимый аудит происхождения контента начинается не с тестовых запросов модели, а с договора о мандате и критериях. В мандате необходимо указать юридические лица, модели и версии, временной период обучения, географию, категории источников, применимые нормы, объём доступа к закрытым сведениям и право аудитора выдавать отрицательное заключение. Если клиент заранее ограничивает предмет проверки только публичной сводкой, это допустимо как desk review, но недопустимо именовать полным аудитом лицензий и asset-level provenance. [2][3]

Методика должна быть риск-ориентированной. Полная проверка каждого элемента многомиллиардного корпуса нередко технически невозможна; выборка же должна позволять оценить документируемые риски, не скрывая вероятности пропуска. Для формирования страт рекомендуется использовать тип права, поставщика, временной период сбора, правовое основание, язык, контент с высокой коммерческой ценностью, историю претензий и наличие signals opt-out. Элементы, выбранные после обнаружения нарушения, следует выделять как целевые, а не смешивать со случайной выборкой.

Отчёт обязан раздельно показывать дизайн выборки, наблюдаемую частоту дефектов, неустранённые пробелы и юридическую квалификацию. Процент объектов без полного лицензионного паспорта — это показатель документального пробела, а не автоматически процент незаконно использованных произведений. Вместе с тем отсутствие доказательств правомерности не позволяет формулировать позитивный вывод о полном соответствии. Рекомендуемая шкала: подтверждено, условно подтверждено, неполные сведения, выявленное несоответствие и вне объёма доступа.

Элемент программы	Минимальное решение	Ошибка, которую предотвращает
Границы	Модели, версии, наборы, сроки, нормы	Ложное ощущение проверки всей компании

Элемент программы	Минимальное решение	Ошибка, которую предотвращает
Отбор	Случайная и целевая страты раздельно	Скрытый cherry-picking
Доступ	Публичные и закрытые источники доказательств	Отождествление summary с реестром
Оценка	Право, факт, техническая неопределённость	Подмена судебного вывода мнением аудитора
Публикация	Степень уверенности и ограничения	«Сертификат» без понятного содержания

26. Инвентаризация активов и граф происхождения данных

Рекомендуемая единица наблюдения — не URL как таковой, а «объект-версия-право-источник». Каждый актив должен иметь внутренний идентификатор, тип контента, hash, источник и дату получения, сведения о лицензии/исключении, историю правовых сигналов, отношения к исходной и преобразованным версиям и перечень моделей, в которых он использовался. Такие поля позволяют восстановить путь спорного произведения от веб-сайта до конкретного обучающего запуска. Без графа происхождения удаление и проверка downstream использования превращаются в ручное расследование по памяти сотрудников.

Архитектура должна различать правовые и технические идентификаторы. ISBN указывает на издание, но не всегда на конкретный цифровой файл; URL изменяется; визуальный отпечаток может ошибочно объединять разные фотографии; hash улавливает точные байты, но не распознаёт изменения формата. Поэтому для значимых объектов рекомендуется комбинация идентификаторов и журнал преобразований. Каждая операция deduplication, tokenization, OCR или нарезки на блоки должна сохранять связь с родительским объектом и применимыми ограничениями.

Для GDPR-совместимости реестр должен минимизировать персональные сведения, особенно при работе с авторскими заявлениями и судебными кейсами. Полный текст охраняемого произведения не нужно дублировать в аудиторском отчёте: достаточно контрольного хеша, безопасного доказательства существования и доступа проверяющего по протоколу. Независимый аудитор вправе сообщить об отсутствии доступа, но не должен раскрывать коммерчески чувствительные каталоги без основания. [16] [20]

Поле	Содержание	Проверка целостности
asset_id/version	Устойчивый ID и версия файла	Отсутствие коллизий и пропусков
parent_id	Источники преобразования	Возможность обратной трассировки
legal_basis	Лицензия/исключение и статус	Соответствие действию и времени
reservation	Источник и версия opt-out	Снимок состояния на дату сбора
model_lineage	Версия/запуск использования	Независимое согласование с обучающим манифестом

27. Классификация рисков в корпусе: не одна категория «из интернета»

Корпус следует разделять по проверяемым признакам: лицензированные каталоги, авторские открытые источники с ясными условиями, данные, полученные по исключениям TDM, наборы от агрегаторов и активы с неясным происхождением. Отдельно маркируются пиратские репозитории, нелегальные зеркала, материалы из закрытых сообществ и источники с техническими ограничениями. Такая классификация нужна не для впечатляющего дашборда, а для выбора правового и технического контроля.

Риск определяется не только вероятностью отсутствия лицензии, но и потенциальным эффектом: исключительная экономическая ценность, трудность компенсации правообладателя, наличие персональных сведений, вероятность длинного выходного воспроизведения и распространение по многочисленным downstream моделям. Условная оценка риска может представляться качественной матрицей, если статистические оценки не имеют надлежащей эмпирической основы. Нельзя выводить «точно 82 балла юридической безопасности» из произвольных весов признаков.

Высокий риск следует компенсировать документальной проверкой, исключением спорного источника, специальной лицензией либо остановкой использования. Для аудита полезно учитывать механизмы обхода: crawler может блокировать первичный домен, но продолжать читать копию в общедоступном архиве. Следовательно, оценка должна захватывать устойчивую идентичность произведения и маршруты перепубликации, а не только соблюдение конкретного списка доменов.

Категория	Рекомендуемый статус	Необходимая процедура
Подтверждённая лицензия	Допуск в рамках объёма	Проверка срока и версии
Законный TDM без резервирования	Условный допуск	Доказательство доступа и сигналов
Неизвестная цепочка прав	Карантин	Дообследование или исключение
Заведомо пиратский источник	Запрет на использование	Блокировка и расследование ранее созданных копий
Установленная претензия	Индивидуальное решение	Сохранение доказательств и оценка мер

28. Статистический аудит и честное выражение неопределённости

Для статистического вывода следует заранее определить популяцию: все записи корпуса, уникальные произведения, документы одного поставщика или конкретные версии произведений. Эти величины не взаимозаменяемы. Например, один художественный текст, встречающийся 500 раз в сети, может составлять 500 файлов, но одно охраняемое произведение. Для оценки правовых дефектов рекомендуется указывать показатели как по файлам, так и по уникальным активам, если доступны обе идентификации.

Допустим, независимая случайная выборка из 200 объектов обнаружила 12 объектов с отсутствующим доказательством правомерного доступа: наблюдаемая доля равна 6%. Это не означает, что 6% корпуса содержат правонарушения. Во-первых, категория обозначает документальный пробел; во-вторых, интервал статистической неопределённости ненулевой; в-третьих, результаты нерепрезентативны, если случайность отбора нарушена или значимая часть корпуса была недоступна. Следует показывать отдельный показатель отказа предоставить сведения.

При сложном дизайне выборки применяются стратифицированные веса и проверка кластеризации. Для редких тяжёлых нарушений целевое тестирование может быть рациональнее случайного поиска, однако его результаты нельзя экстраполировать на весь корпус обычной пропорцией. В докладе отделяются системные недостатки процедуры от эпизодических дефектов. Это обеспечивает честный предел выводов и позволяет сравнивать аудит во времени без искусственного завышения успеха. [10]

Учебный пример	Расчёт	Допустимый вывод
12 из 200 не имеют подтверждения доступа	12 / 200 = 6%	Доля пробелов в выборке, не нарушений в целом корпусе
0 из 100 воспроизведений	Правило трёх: ~3% верхняя граница	Нулевой результат не равен нулевому риску
8 из 80 прицельных тестов	10% среди выбранных риск-объектов	Нельзя переносить 10% на случайную популяцию
40% корпуса недоступно аудитору	Охват 60%	Позитивный вывод ограничен известной частью

29. Контроль согласованности независимых оценщиков

Юридическая классификация лицензий и техническая классификация совпадений требуют человеческого решения. Если два аудитора независимо проверяют один актив и один считает копирование допустимым по лицензии, а второй указывает на её ограниченный предмет, система должна не скрывать разногласие, а документировать его основания и передавать на юридическую рецензию. Простая доля совпадений недостаточна, когда почти все объекты относятся к очевидной категории.

Практический протокол включает инструктаж оценщиков, общие критерии, слепую оценку заранее выбранной контрольной выборки, расчёт согласованности и разбор расхождений с фиксацией

окончательной категории. Для номинальных категорий может использоваться коэффициент каппа Коэна, но только при корректных предпосылках и прозрачной матрице разметки. Высокая согласованность не гарантирует юридической истинности, если все проверяющие использовали одинаково ошибочную инструкцию.

Отдельная роль должна отвечать за обновление правовой рубрики при изменении закона или поступлении судебного решения. Не допускается менять исторические протоколы задним числом без сохранения старой версии. Иначе последующая «высокая согласованность» может отражать ретуширование истории, а не зрелость экспертной процедуры.

Контроль	Метод	Документируемый результат
Слепая разметка	Два оценщика без доступа к ответу коллеги	Исходные категории
Согласованность	Матрица пересечения и коэффициент при уместности	Количественная и качественная оценка
Споры	Мотивированная юридическая рецензия	Согласованное решение и несогласие
Изменение нормы	Версия методики и дата	Историческая сопоставимость

30. Проверка claims поставщика: от маркетинга к юридическому доказательству

Распространённые заявления «мы не используем данные без разрешения», «наш корпус безопасен» или «мы обучались только на открытых ресурсах» не обладают одинаковой доказательной ценностью. Аудит должен преобразовывать их в проверяемые утверждения: существуют ли источники со спорным доступом, как проверялись opt-out, у каких поставщиков запрошены лицензии, какие исключения применялись, кто утвердил спорные категории. Простая публикация политики без отчёта об исполнении не доказывает соблюдения. [2][3]

Рекомендуется требовать конкретные артефакты под каждое заявление: signed contract, юридическую матрицу исключений, отказ crawler, манифест обучения, протокол реакции на претензию. Если доказательство не раскрывается из-за коммерческой тайны, возможна процедура confidential clean room и публикация агрегированного вывода. Но нельзя заменять независимую проверку устным заверением представителя компании. Аудитор указывает, каким доказательствам он не имел доступа, и какой вывод от этого невозможен.

Отдельно стоит оценить конфликт интересов при финансировании проверки поставщиком. Если вознаграждение аудитора зависит от выдачи положительного заключения, доверие к выводу подрывается. Процедура должна предусматривать фиксированную оплату за работу, право публиковать отрицательный результат, документирование консультационных отношений и проверку финансовых зависимостей сотрудников.

Claim	Надлежащий вопрос	Пример доказательства
«Все права соблюдены»	На каком участке и по каким нормам?	Legal-basis asset sample
«Опубликована прозрачная сводка»	Какие версии покрывает?	Public summary + manifests
«Opt-out работает»	Проверена ли техническая реализация?	Negative tests crawler
«Аудит независимый»	Кто оплачивает, утверждает и публикует?	Engagement letter и conflicts register

31. Исправления, запросы правообладателей и право на ответ

Надлежащая система должна иметь процедуру приёма обоснованных запросов от авторов, наследников, издателей и организаций коллективного управления. Не следует требовать от обычного фотографа предоставить закрытую копию всех весов модели, чтобы заявить о возможном нарушении. Достаточно канала, позволяющего идентифицировать произведение и интерес заявителя, с соблюдением принципа

минимизации данных. Компания вправе проверить полномочия представителя и наличие действительного права, но эта проверка не должна превращаться в бесконечный барьер. [5][9]

После получения запроса следует зафиксировать дату и предмет, временно сохранить относящиеся к делу доказательства, проверить source lineage и основания использования, дать мотивированный ответ и документировать предложенную меру. Возможны отказ с объяснением законной основы, ограничение будущего сбора, лицензирование, исключение из индекса, договорное урегулирование либо обращение к компетентному органу. Форма ответа должна отличать подтверждённый факт нахождения произведения в наборе от предположения, основанного только на сходстве выходного текста.

Независимый Observatoire может оценивать общую доступность процедуры и качество реагирования, но не обещает правообладателю принудительную компенсацию и не объявляет виновность компании вместо суда. Следует сохранять возможность правообладателя оспорить недостаточно мотивированное решение, а компании — исправить ошибочное публичное утверждение. Публикация имён сторон допускается только при отдельном правовом основании.

Этап	Контроль качества	Выходной документ
Получение	Подтверждение, номер, предмет	Карточка обращения
Верификация	Полномочия, идентичность объекта	Проверочный протокол
Исследование	Lineage, условия, тест воспроизведения	Мотивированная оценка
Решение	Мера и пределы исполнения	Ответ заявителю
Обжалование	Независимый пересмотр	Реестр исправлений

32. Оценка воздействия на рынок и доходы авторов

Один из центральных вопросов консультации Комиссии — может ли обучение ИИ замещать спрос на оригинальные произведения и каким образом могла бы работать справедливая компенсация. Надёжного универсального коэффициента «один использованный файл равен такой-то потере дохода» не существует. Рынки фотографии, музыкальных каталогов, перевода, специализированных новостей, лицензирования данных и программного кода существенно различаются. Анализ должен учитывать произведения, качество заменителей, структуру договоров и контрфактическое развитие спроса. [5][9][10]

Рекомендуется измерять не только объём обучающих данных, но и наблюдаемые условия лицензирования: доступность лицензий малым авторам, динамику тарифов, требования эксклюзивности, наличие посредников, расходы на юридическую защиту и роль коллективного управления. При этом данные самих сторон являются потенциально заинтересованными и нуждаются в независимой проверке. Статистика числа жалоб может увеличиваться вследствие лучшей доступности каналов защиты, а не роста реальных нарушений.

В исследовательском предложении ARGA рекомендуется использовать несколько сценариев: базовый рынок без нового механизма, рынок добровольных лицензий с минимальной прозрачностью, усиленное документирование обучающих корпусов и стандартизированное opt-out с эффективным контролем. Каждый сценарий должен сопровождаться допущениями и интервалами неопределённости, не превращаясь в голословный прогноз многомиллиардных компенсаций.

Показатель	Что характеризует	Возможная ловушка
Выручка от лицензий	Совершённые сделки	Не показывает всех неполученных возможностей
Цена каталога	Рыночные условия	Искажается эксклюзивностью и пакетом прав
Жалобы на обучение	Доступность и активность защиты	Не равны установленным нарушениям
Выходные совпадения	Технический риск воспроизведения	Не измеряют автоматически ущерб рынку

33. Независимость Observatoire ARGA: финансы и процессуальные границы

Observatoire ARGA может формировать экспертные исследования об авторском праве и ИИ в рамках своей аналитической программы, но нельзя выдавать его за аккредитованный орган сертификации, государственный регулятор или судебного эксперта, если соответствующее качество не подтверждено. Особенно важно разграничить общественный мониторинг и платные консультационные проекты: если компания, проходящая аудит, одновременно оплачивает иные услуги структуре или её партнёрам, конфликт интересов должен быть зарегистрирован и оценён независимым лицом.

Предлагаемая модель предусматривает научного руководителя, юридического рецензента, технического эксперта, секретаря доказательств и комитет по конфликтам интересов. Нельзя допускать, чтобы коммерческий менеджер единолично редактировал вывод в пользу оплачивающей стороны. Для каждого проекта определяются права доступа, правила публикации методики, несоответствий и существенных оговорок. Если открытый отчёт невозможен по режиму тайны, опубликованное заявление должно ясно обозначать глубину проверки и природу закрытой информации.

Ключевой критерий независимости — возможность честного отрицательного результата без санкций к исследователю. Отчёт не должен подразумевать, что ARGA имеет прямой доступ к непубличным обучающим наборам компаний без договора или законного полномочия. Предварительный публичный мониторинг возможен по открытым данным; в этом случае все выводы ограничиваются доступным материалом.

Риск независимости	Минимальная защита	След для проверки
Оплата поставщиком	Фиксированная оплата независимо от результата	Договор и тарифная политика
Параллельное консультирование	Disclosure и recusal	Журнал интересов
Давление на публикацию	Политика независимой редакции	История версий и голосования
Отсутствие закрытого доступа	Ясный статус desk review	Ограничения в титуле и выводах

34. Применение к юридической аналитике, розыску и санкциям

Для ARGA авторско-правовой аудит возникает при использовании обучающих массивов судебных публикаций, справочников, журналистских расследований и коммерческих баз санкционного комплаенса. Надо различать законность использования публичного государственного акта как источника фактов и права на редакционную базу, в которой акт снабжён аннотациями, классификациями и платной поисковой инфраструктурой. Лицензия на доступ юриста к базе не обязательно покрывает массовое копирование для обучения собственной модели. [7][21]

Отдельную категорию составляют доказательства по конкретным клиентским делам. Даже когда адвокат или эксперт законно получил документы, их загрузка в глобальный обучающий контур без отдельного анализа конфиденциальности, персональных данных, профессиональных обязанностей и прав на документы может быть недопустимой. Необходимо разделять закрытый retrieval для конкретного поручения и использование клиента как источника постоянного обучения общего сервиса. Для коммерческих санкционных баз аналогично проверяются договорные запреты на массовое извлечение, право переработки и публикации сводной аналитики.

Предложение ARGA — запускать pilot only с обезличенными/лицензированными учебными данными и доказательным реестром. Никаких фактических утверждений о том, что сторонние модели уже используются INTERPOL или иными организациями для конкретных материалов ARGA, из данного исследования не следует. Аналитический проект не заменяет официальные разрешения и требования допуска к соответствующей информации.

Материал	Основной риск	Допустимый пилотный подход
Официальный судебный акт	Условия публикации и персональные данные	Анализ публичных фактов и отдельный GDPR review
Коммерческая правовая база	База данных и договорный доступ	Лицензия для анализа и согласование API

Материал	Основной риск	Допустимый пилотный подход
Документы клиента	Тайна, персональные данные и IP	Исключить из общего обучения без отдельной оценки
Санкционный коммерческий каталог	Права агрегатора и качество записей	Договор, аудит источников, исправление ошибок

35. Рекомендации к консультации European Commission Technology and Copyright

Целевая консультация Европейской комиссии, начатая 29 сентября 2026 года и открытая до 3 ноября, предоставляет подходящую площадку для предложений о прозрачности авторско-правового происхождения обучающих данных. При подготовке ответов необходимо придерживаться рамки вопросов формы, а не заявлять о подаче уже в самом докладе. Предлагаемая позиция ARGA состоит в сочетании проверяемой прозрачности, эффективного резервирования прав, доступных средств защиты для малых авторов и управляемых условий доступа независимых аудиторов. [5]

Первое предложение — установить совместимый минимальный набор полей provenance: категория источника, период сбора, правовая основа, подтверждение обработки правовых сигналов, идентификатор модели и изменения при дообучении. Второе — развивать машиночитаемые механизмы резервирования с бесплатным и недискриминационным доступом. Третье — обеспечить защищённый доступ к достаточным доказательствам для правомочного контроля, не требуя раскрывать весь лицензированный корпус публично. Четвёртое — развивать процедуры недорогого оспаривания для отдельных правообладателей и независимые тестовые протоколы.

Предложения должны сопровождаться оценкой затрат разработчиков и обладателей прав. Необоснованно строгая публичная детализация может раскрыть коммерческие договоры и уязвимости инфраструктуры; слишком общий отчёт становится бесполезным для реализации прав. Компромисс: публичный слой с достаточной категоризацией, закрытый подтверждаемый реестр и законные процедуры независимой проверки. Участие в консультации возможно только после официальной подачи и получения подтверждения; настоящее исследование само по себе не является отправленной позицией ARGA.

Предложение	Цель	Условие пропорциональности
Стандарт provenance	Прослеживаемость происхождения	Многоуровневое раскрытие без полного dump
Машиночитаемый opt-out	Реальная осуществимость прав	Бесплатный интерфейс для малых авторов
Доказательный доступ	Возможность независимой проверки	Защищённая среда, журнал доступа
Внесудебное рассмотрение	Снижение расходов на защиту	Право обжаловать и доступ к суду сохраняются

36. План внедрения и показатели качества независимого контроля

Практическая программа внедрения состоит из трёх последовательных блоков: нормативная карта, испытательный реестр provenance и независимый обзор пилотных наборов. На первом этапе юридическая группа создаёт классификатор лицензий, TDM, правовой статус моделей и процедурные правила. На втором технический специалист строит граф происхождения небольшого контролируемого набора и тестирует обработку opt-out. На третьем отдельная команда проводит слепую проверку выборки и оформляет отчёт с доказательствами, неопределённостями и правом на корректировку результатов.

Рекомендуемые индикаторы не должны включать искусственную цель «ноль жалоб». Полезнее измерять долю записей с прослеживаемым источником, долю подтверждённых правовых оснований, частоту неверной классификации opt-out, время реагирования на правдоподобное обращение, полноту manifest после обновления и долю заключений с независимой рецензией. Для каждого индикатора надо фиксировать популяцию, метод измерения, границы интерпретации и критерий эскалации.

Пилот нельзя запускать на закрытых клиентских данных только для наращивания портфолио. Стоп-условиями должны быть отсутствие правовой основы, доказанный обход запрета, невозможность изолировать спорный набор, существенные разрывы provenance и неуправляемый конфликт интересов. Финальное решение о публикации не должно зависеть от количества страниц отчёта: качественная работа может закончиться мотивированным отказом от использования данных.

Этап	Проверяемый результат	Основание остановки
Нормативная карта	Матрица обязанностей и статусов	Неустановленное применимое право
Пилот данных	Граф происхождения и журнал opt-out	Существенные неразрешённые активы
Рецензия	Воспроизводимая выборка и вывод	Конфликт интересов без защиты
Публикация	Оговорки, источники, программа пересмотра	Неподтверждённые заявления о сертификации

37. Практические выводы: что подтверждено, а что остаётся спорным

Международные материалы WIPO дополнительно подтверждают, что сравнительная оценка ИИ и интеллектуальной собственности должна сохранять различие между национальными исключениями, лицензиями и условиями вывода продукта на рынок. [14]

В праве ЕС имеются ясные обязательные элементы: исключения TDM при определённых условиях, право на резервирование в общем режиме, право на воспроизведение и обязанности GPAI-поставщика по политике copyright compliance и публичной сводке. Одновременно многие вопросы судебной квалификации сложны: как применять конкретное исключение к различным этапам обучения, когда правовой сигнал считается надлежащим, как доказать связь произведения с моделью и какие факторы определяют гражданскую ответственность за выходные совпадения. Окончательные ответы нельзя подменять правилами внутреннего аудита. [1][2][6]

Институциональный вывод заключается в том, что прозрачность без доказательной прослеживаемости слабее, чем механизм защиты прав, а лицензия без цепочки полномочий слабее, чем подлинное разрешение правообладателя. При этом нельзя требовать невозможной абсолютной определённости: добросовестный аудит характеризует пределы проверенного и строит управляемые меры снижения риска. Для ARGA это даёт основу профессиональных исследований и участия в консультациях без имитации регуляторных полномочий.

Настоящий доклад содержит правовую методику, практические формы и гипотетические сценарии; он не доказывает нарушение конкретной компанией авторского права и не является актом оценки соответствия её закрытых моделей. Всякий переход к фактическому заключению требует отдельного поручения, защищённого доступа и воспроизводимого анализа по документированному протоколу. Сведения о состоянии спорных судебных разбирательств следует перепроверять на дату официальной публикации.

Итог	Статус	Предлагаемое действие ARGA
TDM исключения и opt-out	Действующая нормативная основа ЕС	Методические разъяснения и сравнительные заключения
Публичная сводка GPAI	Обязательная регуляторная обязанность при применимости	Проверка уровня раскрытия и версий
Всеобщая незаконность/законность обучения	Не установлена универсальным правилом	Отказ от категорических выводов
Аудит provenance	Предлагаемая независимая методика	Пилот с контролируемым корпусом
Консультация Technology and Copyright	Текущая публичная процедура до 3 ноября 2026	Согласованная документированная позиция

38. Доказательство ущерба рынку: причинность, замещение, альтернативы

При обсуждении ущерба от обучения модели необходимо разделять несколько возможных событий: создание копии без разрешения, потерю платы за лицензию, снижение продаж оригинального произведения, уменьшение стоимости будущей лицензии и возникновение выдачи, конкурирующей с

оригиналом. Эти события не равнозначны по характеру доказательств и возможным мерам защиты. Например, отсутствие лицензионного договора не позволяет автоматически умножить размер корпуса на максимальный коммерческий тариф и назвать результат установленным ущербом. Нужны предмет договора, реальный рынок, время, допустимость лицензирования и применимое национальное право. [10] [13]

При исследовании конкуренции полезно строить контрфактический сценарий: что произошло бы без спорной технологии, но с учётом иных факторов, влияющих на спрос. Для фотографий это может быть изменение платформенного распространения и спроса на стоковые изображения; для юридических аннотаций — развитие бесплатных официальных баз; для музыкальных текстов — рост или снижение лицензионных сервисов. Оценка требует данных о ценах, транзакциях и качестве заменителей, а не одного опроса заинтересованных сторон. Нельзя выдать корреляцию снижения продаж и роста ИИ за доказанную причинность.

Предлагается трёхслойная экспертиза: юридическая квалификация заявленного права; экономическая модель предполагаемого вреда; независимая чувствительность результата к альтернативным объяснениям. Особо нужно отделять стоимость самих произведений от стоимости коммерческих коллекций, интерфейсов, разметки и сервисной инфраструктуры. Для прозрачности раскрываются ограничения доступа к коммерческим данным и то, кем они предоставлены. Даже корректная экономическая модель не заменяет судебную оценку соответствующих обстоятельств.

Гипотеза вреда	Требуемые данные	Риск ошибки
Неполученная лицензия	Сравнимые договоры и права	Не существовало рынка данной лицензии
Потеря продаж	Продажи, спрос, заменители, время	Смешение с отраслевым трендом
Подмена оригинального сервиса	Функциональное сходство и поведение пользователей	Сходство задачи не означает подмену
Издержки защиты	Реальные расходы и причинная связь	Отнесение обычных расходов к инциденту

39. Роялти и коллективное распределение: проверяемая арифметика

Модели компенсации могут строиться на фиксированных платежах, объёме произведений, интенсивности использования, оценке экономического вклада или коллективном распределении. Каждая несёт риск манипуляции. В частности, число токенов или копий файла может сильно зависеть от технической сегментации и дублирования и не отражает творческой ценности; просмотры выходов модели не всегда позволяют надёжно установить причинный вклад конкретного произведения. Поэтому предлагаемая формула должна быть проверяема и сопровождаться оговорками о невозможности точной атрибуции каждой единицы генерации.

Учебный пример: лицензионный пул 120 000 евро, после согласованного административного удержания 10% остаётся 108 000 евро. Если 60% суммы распределяется пропорционально подтверждённому репертуару, а 40% — по объективно установленным классам использования, соответствующие части равны 64 800 и 43 200 евро. Это математическая иллюстрация, не существующий тариф и не предложение обязательного удержания. Реальные критерии должны пройти проверку на дискриминацию малых правообладателей, прозрачность затрат и права на возражение.

Для аудита важна не только проверка умножения процентов, но и правомерность того, какие произведения включены в числитель и знаменатель. Дублированные копии, непроверенные полномочия общества и каталоги вне территориального охвата могут систематически искажать выплаты. Желательно предусмотреть публичные правила, закрытую проверяемую ведомость начислений и независимый механизм корректировки ошибочной атрибуции.

Шаг	Учебный расчёт	Ограничение
Исходный пул	120 000 евро	Условная сумма
Удержание 10%	12 000 евро	Только если согласовано
Чистый пул	108 000 евро	Проверить реестр правообладателей
Класс A (60%)	64 800 евро	Вес репертуара требует обоснования

Шаг	Учебный расчёт	Ограничение
Класс Б (40%)	43 200 евро	Вес использования требует доказательств

40. Точность юридического резервирования: время и юридический эффект

Конфликт часто возникает не из-за полного отсутствия сигнала, а из-за временного несовпадения. Правообладатель публикует машинный запрет после того, как материал был ранее скопирован и использован в одной версии модели; позднее разработчик создаёт новый обучающий набор из старой копии. Для правовой оценки нужны четыре даты: первоначальный доступ, создание копии, появление резервирования и новое использование. Нельзя рассматривать весь процесс как одно вечное действие, поскольку разные операции могут иметь самостоятельную правовую квалификацию. [1][21]

Аудиторская система должна хранить версию условий сайта на момент каждой загрузки и юридическое решение по последующему использованию. Если сигнал был скрыт из-за технической ошибки crawler, требуется определить, можно ли было его разумно обнаружить и соответствует ли механизм современным технологиям. Если автор ошибочно указал чужое произведение, нужна проверка полномочий, а не автоматическое блокирование любой интеллектуальной деятельности навсегда. При множественных копиях обязательна матрица того, какие именно экземпляры и производные наборы затронуты.

В рекомендациях ARGA следует предусмотреть процедуру добровольного урегулирования исторических конфликтов без предрешения ретроактивности по всем правовым системам. Она может включать добровольную лицензию, ограничение будущего сбора, технические меры предотвращения выдачи или соглашение об использовании спорного массива. Но юридическая позиция по прошлому поведению должна строиться на праве и доказательствах, а не на удобной политике компании.

Событие	Фиксируемая дата	Юридический вопрос
Первый законный доступ	T1	Какое основание имелось?
Создание копии	T2	Было ли допустимо копирование?
Появление opt-out	T3	Надлежащая ли форма и адресат?
Повторное обучение	T4	Требуется ли новая правовая оценка?

41. Торговая тайна и закрытый аудит: пределы прозрачности

Полный публичный список произведений, исходных URL, стоимость каждого лицензионного договора и структура датасета могут затрагивать охраняемую коммерческую тайну, безопасность и права третьих лиц. Однако этот аргумент не равен праву отказаться от всякого подтверждения происхождения. Регламент AI Act уже устанавливает определённый публичный минимум сводки, одновременно предусматривая защищённые маршруты доступа к иной документации для компетентных органов. Исследовательское предложение может развивать сходную многоуровневую архитектуру, не представляя её как уже обязательный детальный реестр каждого файла. [2][3]

В публичном слое должны быть понятны типы источников, временной охват, изменения и порядок соблюдения ограничений правообладателей. В конфиденциальном слое независимый проверяющий, имеющий законное основание, может исследовать конкретные договоры и записи. Для судебного уровня применяются процессуальные гарантии соответствующей страны. Задача состоит не в том, чтобы любой желающий скачал весь корпус, а в возможности проверить существенное утверждение посредством контролируемой и воспроизводимой процедуры.

Для реализации требуется политика доступа, регистрация запросов, запрет несанкционированного вывоза материалов, ограничение кругов экспертов и разделение ролей. При этом сама закрытая среда не делает заключение независимым: проверяющий должен иметь возможность выбрать элементы выборки и проверить полноту популяции, иначе клиент может показать только отобранные заранее «хорошие» записи.

Уровень раскрытия	Материал	Предохранитель
Публичный	Категории и процесс	Не раскрывать чужие договоры
Конфиденциальный аудиторский	Asset registry и выборка лицензий	Роли, логи, независимый отбор
Регуляторный	Документация при законном запросе	Компетенция и защита тайны

Уровень раскрытия	Материал	Предохранитель
Судебный	Материалы в рамках процесса	Процессуальные гарантии сторон

42. Формат доказательства происхождения: совместимость и переносимость

История происхождения имеет практическую ценность, только если переживает изменение технологий и поставщиков. Реестр, целиком привязанный к закрытому внутреннему названию платформы, усложнит переход между облачными средами и независимую проверку. Рекомендуется фиксировать универсальные атрибуты: ID объекта, формат, хеш, источник, время, правовую основу, статус согласия/возражения, parent-child зависимости, оператора изменения и идентификатор использовавшей модельной версии. При этом сами стандарты метаданных должны развиваться с участием авторов, разработчиков и аудиторов.

Не следует путать provenance обучающих данных и provenance сгенерированного результата. Метаданные, подтверждающие, какая камера создала фотографию, могут помогать в установлении происхождения исходника, но не доказывают автоматически, что именно эта фотография присутствовала в обучающем корпусе. Аналогично цифровая подпись публичной сводки подтверждает целостность документа, а не истинность каждого заявления. Для сильного вывода требуется связь подтверждённых технических записей с юридическими разрешениями.

Система должна корректно учитывать преобразования, при которых hash изменяется, но охраняемое произведение остаётся узнаваемым. Для этих случаев нужны алгоритмы приблизительного сопоставления, проверка ложных совпадений и сохранение исходных доказательств. Если компания заявляет о высокой точности распознавания идентичности произведений, аудитор должен потребовать данные о validation set и независимой оценке ошибок.

Утверждение provenance	Доказательство	Чего не подтверждает
«Файл получен в дату T»	Подписанный log и источник	Правомерность использования
«Это то же произведение»	Hash + content matching	Все права на иные версии
«Есть лицензия»	Договор и полномочия	Отсутствие иных ограничений
«Сводка не изменена»	Цифровая подпись	Правдивость содержания

43. Выходные произведения: авторство, производность, лицензии

Доклад преимущественно касается входных обучающих материалов, однако экономический риск нередко проявляется в генерации. Необходимо различать вопрос авторства полностью автоматически созданного результата, вопрос защиты человеческих творческих дополнений и вопрос возможного воспроизведения охраняемых элементов исходных произведений. Отсутствие авторского права на отдельный чисто машинный результат по определённому праву не означает, что копирование чужой охраняемой формы этим результатом становится разрешённым. [6][10]

Поэтому политики продукта должны разделять условия владения/использования выходов, уведомления о возможных чужих правах и тесты совпадения. Генератор изображений способен создать результат, напоминающий общую художественную манеру, но это не тождественно воспроизведению конкретного защищённого рисунка. В другом случае выход может содержать отчётливо узнаваемые элементы отдельного произведения. Разница устанавливается предметной экспертизой, а не маркетинговой фразой «style transfer» или «креативность модели».

Поставщик и интегратор должны заранее определить, кто рассматривает претензии на результат, кто сохраняет доказательства и какие ограничения доступа допускаются до решения спора. В отчёте отдельной графой отмечается невозможность установить происхождение на основании одного выходного совпадения: такая информация может быть сильным сигналом, но требует проверки цепочки training inputs и альтернативных причин совпадения.

Юридический вопрос	Объект проверки	Ошибочное отождествление
Охраноспособность вывода	Человеческий вклад и применимое право	«ИИ всегда автор»
Производность	Существенные охраняемые элементы	«Сходство темы = нарушение»
История обучения	Provenance и журналы	«Похожий output доказывает URL»
Ответственность	Действия поставщика/пользователя	«Всегда отвечает только пользователь»

44. Применение AI Act к модификациям GPAI: граница поставщика

Статус лица, модифицирующего модель, имеет значение для распределения обязанностей по статье 53 AI Act. Некоторые операции fine-tuning могут приводить к тому, что downstream организация становится поставщиком результирующей модели общего назначения, но не любая пользовательская настройка интерфейса либо промпт делает её таковой. Комиссия предусматривает критерии охвата и разъясняет порядок документирования новых тренировочных данных при изменении модели. Поэтому аудит должен начинаться с технического и юридического описания модификации. [2][3][8]

Для документов по лицензиям это означает необходимость проверить, распространяется ли разрешение на дообучение за пределами исходного провайдера. Downstream участник может иметь право использовать API, но не право копировать training inputs первичной модели. При создании собственного нового корпуса он отвечает за подтверждение прав на дополнительные источники; полагаться на старую публичную сводку базовой модели недостаточно для новых данных.

Рекомендуется вести split provenance: набор и версия базовой модели, собственная delta, новый юридический статус, источники заявленных исключений и актуальная сводка. При публикации нужно избегать двойного счёта общих источников и ясным образом идентифицировать референтную базовую модель. Отдельные обязанности системного риска и документации оцениваются при применимости, а не автоматически выводятся из факта использования большого API.

Вид изменения	Нужная проверка	Документ
Prompt-политика	Не меняет ли правовой статус?	Архитектурное описание
Fine-tuning	Собственный корпус и роль	Delta manifest
Новая GPAI-модель	Article 53 и сроки	Legal status memo
Новая публичная сводка	Покрытие изменений	Ссылка на базовую и новую версии

45. Проверяемость договоров без публикации личных и коммерческих данных

Для проверки лицензионного основания аудитор нередко получает договор, содержащий адреса, банковские условия, сведения об авторах, персональные подписи и бизнес-условия. Необходимо сохранить достаточную часть правовой информации и одновременно исключить ненужное раскрытие. Практическая модель допускает несколько представлений одного контракта: полный оригинал в защищённом архиве, аналитическое извлечение существенных условий и публичную запись об объёме права без коммерчески чувствительных сумм. [20]

Правовой минимум извлечения: управомоченные стороны, предмет лицензии, территория, срок, разрешённые операции, сублицензия, условия прекращения и ограничения по типам моделей. Исключение данных о цене из публичного документа не должно приводить к сокрытию существенного запрета на машинное обучение. Редактор обязан определить, какие сведения влияют на юридический вывод и какие можно обобщить без потери проверяемости.

Проверка доступа оформляется письменно: кто видел полный договор, в какой версии, по какой цели и где хранится доказательство. Заключение указывает, была ли проведена выборка оригинальных подписанных документов или использовались только предоставленные компанией резюме. Такое разграничение важно для честной оценки доверия и для предотвращения недостоверного заявления о внешней сертификации.

Слой	Вид сведений	Контроль
Полный оригинал	Все условия и подписи	Защищённый доступ
Аудиторское извлечение	Существенные права и запреты	Проверка по оригиналу
Публичная запись	Категория и диапазон права	Нет утечки цены/персональных данных
История проверки	Кто, когда, какую версию видел	Audit trail

46. Сравнение раскрываемых и реально проверяемых показателей

Публичная сводка может показывать, что 30% источников относятся к лицензированным каталогам, но такой показатель мало полезен без определения единицы измерения: доля файлов, токенов, объёма байтов,

уникальных произведений или стоимости лицензий. Для каждой категории должны существовать словарь и метод расчёта. Аудитор обязан проверять, не маскируют ли несколько крупных наборов значительную группу материалов с неопределённым происхождением. [3]

Например, если 1000 уникальных произведений представлены 10 000 файлов, а лицензия подтверждена для 900 произведений, доля лицензированного репертуара составляет 90% произведений, но никак не обязана равняться доле файлов. Массовое дублирование разрешённых материалов может искусственно поднять «процент соответствия» при файловом подсчёте. Следует показывать обе цифры либо объяснять невозможность одного из измерений и проводить проверку на устойчивость результата к deduplication.

Отчёт также должен раскрывать доступность доказательств. Если 50% источников имеют документированное правовое основание, а для оставшихся 50% документы не предоставлены, нельзя описывать показатель как «50% законных, 50% незаконных». Правомернее фиксировать «50% доказательно подтверждены» и отдельную категорию непроверенных, затем определять юридические и организационные меры для уменьшения неопределённости.

Показатель	Единица	Методологическое ограничение
Доля лицензированного репертуара	Уникальные произведения	Нужна дедупликация
Доля лицензированных файлов	Файлы	Чувствительна к копиям
Доля provenance-complete	Записи с полями	Полнота метаданных не равна законности
Доля audit-covered	Части популяции	Не распространять на недоступные слои

47. Открытая публикация и право на возражение против экспертного вывода

Публикация результатов независимого аудита способна влиять на репутацию компании и на положение автора. Поэтому процедура должна обеспечивать точное разграничение фактов, выводов и гипотез, а также механизм корректировки доказуемых ошибок. Перед публикацией существенно негативных утверждений допустимо дать затрагиваемой стороне разумную возможность ответить на проверяемые факты при условии сохранения редакционной независимости и соблюдения сроков безопасности. Этот механизм не должен превращаться в право компании запретить неудобное исследование.

В тексте нужно ясно указывать, какие данные получены из публичных источников, какие от проверяемого поставщика и какие проверены третьей стороной. Если компания не согласна с юридической квалификацией, отчёт может содержать краткую позицию компании, но окончательная ответственность за методический вывод остаётся у авторов. Неподтверждённые обвинения в «массовом воровстве» неуместны, как и обещания «абсолютного соответствия», если исследователь не изучил весь корпус.

Для Observatoire ARGA предлагается отделить фактическое исправление от изменения научной оценки. Новые доказательства должны приводить к датированному дополнению или новой версии с changelog, а не к бесследной редактуре ранее опубликованного материала. В конце каждого отчёта приводится контакт для корректировок, версия методики, дата отсечения источников и пределы доступа.

Событие	Редакционная мера	Сохранение независимости
Фактологическое опровержение	Проверка и корректировка	Публикация изменений
Юридическое несогласие	Краткая позиция стороны	Вывод остаётся мотивированным
Новый источник	Обновление версии	Сохранить первоначальный архив
Давление спонсора	Протокол конфликта	Отвод/независимое решение

48. Двухконтурная проверка статьи 53 AI Act и материального авторского права

При подготовке аудиторского задания необходимо открыть две самостоятельные линии проверки. Первая отвечает на вопрос, выполняет ли поставщик GPAI формальные обязанности по статье 53(1)(c)–(d): существует ли применяемая политика соблюдения авторского права, как устроено выявление резервирования прав, опубликована ли сводка обучающего контента и соотнесена ли она с идентифицируемой версией модели. Вторая линия касается правомерности конкретных действий с конкретными объектами: можно ли доказать правомерный доступ, кто является обладателем соответствующих прав, применимо ли исключение TDM, было ли действительное резервирование, каков

территориальный и временной охват лицензии и не произошло ли отдельного воспроизведения при генерации. Первое не заменяет второго. Статья 53 не создаёт общей авторско-правовой лицензии на весь Интернет, а заключение о правомерном TDM не освобождает поставщика от прозрачности, если применим AI Act. [1][2][3][8]

Полезно выстроить доказательственную матрицу, в которой каждому нормативному требованию соответствует один или несколько источников, ответственный владелец и критерий достаточности. Для регуляторной линии это утверждённая версия политики, журнал исполнения, публичная страница со сводкой, history of changes. Для материальной линии — договоры, лицензионные условия на дату доступа, журналы краулера, снимки сигналов opt-out, идентификаторы произведений и сведения о цепочке прав. Если проверяющему доступна только публичная сводка, вывод обязан быть ограничен её соответствием по полноте и форме: прозрачность уровня категорий не равна подтверждению законности каждого элемента многомиллиардного корпуса.

Методологическая ошибка возникает, когда организация вводит один показатель «compliance score 95%», механически усредняя наличие процедур и выявленные нарушения исключительных прав. Нельзя компенсировать подтверждённое незаконное копирование хорошо оформленной политикой. Поэтому результаты представляются раздельно: процедурные соответствия, доказательства прав на выборке, существенные неопределённости и решения о продолжении или остановке применения спорных источников. Порог принятия решения определяется заранее, а не выбирается после выявления неблагоприятных результатов.

Контур	Контрольное утверждение	Доказательство	Запрет на вывод
AI Act, политика	Механизм opt-out действует, а не описан	Versioned policy, журналы проверок, инциденты	«Следовательно, все произведения использованы законно»
AI Act, summary	Сводка охватывает указанную модель/версию	Публичная ссылка, дата, категории источников	«В сводке раскрыт каждый отдельный объект»
Авторское право	Есть основание конкретного воспроизведения	Лицензия или обоснование исключения и факты доступа	«Эта лицензия покрывает все версии корпуса»
Выход модели	Нет выявленного недозволенного существенного воспроизведения в тестах	Протокол запросов, критерии сравнения, ограничения выборки	«Любая будущая генерация безопасна»

49. Проверка мультимодального обучения: изображение, звук, видео

Нельзя переносить критерии анализа книжных текстов на изображения, фонограммы и видеоматериалы без дополнительных доказательств. Фотоархив содержит права автора снимка и возможные права на изображение человека; музыкальная запись связывает произведение, исполнение и фонограмму; видеоролик совмещает изображение, звук, монтаж и зачастую фрагменты произведений третьих лиц. Одно разрешение от владельца сервера не может автоматически закрыть все эти слои. Поэтому аудит мультимодального корпуса начинается с карты типов прав и сторон, которые потенциально должны дать разрешение. Публичный файл в сети не становится общественным достоянием вследствие утраты EXIF-метаданных при перепубликации. [6][7][20]

Технический протокол содержит не только криптографический хеш файла, но и осторожно настроенные перцептивные признаки. Точные хеши надёжны при побитовом совпадении, однако не связывают исходный JPEG с изменённым кропом, а перцептивные хеши способны выдавать ложные совпадения. Аналогичная проблема возникает с аудио-отпечатками, ускоренным воспроизведением или шумоподавлением. Совпадение отпечатка не устанавливает правообладателя; оно лишь указывает на объект для юридической проверки. При сомнении применяется ручная верификация с фиксацией степени уверенности, а не автоматическое обвинение.

Для воспроизведения на выходе следует тестировать конкретные режимы модели: генерацию текста, image-to-image, аудиопродолжение, голосовую имитацию и видео с референсными кадрами. Условия предъявления оригинала модели документируются особенно тщательно: если исследователь сам приложил защищённое изображение или фрагмент музыки в качестве референса, похожий вывод имеет иной доказательственный смысл, чем воспроизведение без предоставления исходного материала.

Публикуемые иллюстрации должны быть достаточно малы и законно используемы либо заменяться описанием результата, чтобы аналитический доклад сам не стал каналом повторного нарушения.

Объект	Что фиксирует технический эксперт	Что отдельно устанавливает юрист
Фотография	Преобразования, сходство кропов, источник копии	Авторство, цепочка лицензий, пределы использования
Звукозапись	Fingerprint, темп, преобразования, наличие референса	Права на произведение, исполнение и фонограмму
Видео	Сегменты, длительность, синхронизация аудио	Совокупность прав на составные объекты
Синтетический голос	Технический признак подражания	Смежные, личные и договорные права по применимому праву

50. Архитектура безопасного хранения доказательств и litigation hold

Собранная в ходе аудита информация может впоследствии стать доказательством в суде или объектом доступа регулятора. Из этого не следует, что её допустимо хранить бесконечно. Политика retention устанавливает срок в зависимости от вида объекта, основания хранения, ожидаемого спора и режима персональных данных. В обычном цикле контроля временные копии удаляются или минимизируются после достижения цели; при реально возникшем или разумно предвидимом судебном разбирательстве допустимость и объём временного сохранения оцениваются отдельно с юристами. Необходимо отличать сохранение минимального доказательства от продолжения использования спорного произведения для обучения.

Для критически важных элементов формируется пакет: источник файла, защищённый идентификатор, контрольная сумма, извлечение заголовков и договорных метаданных, доказательство состояния веб-страницы на момент доступа, журналы запросов, время и среда получения, цепочка передачи, основания каждого копирования. Часы исследовательской среды синхронизируются и фиксируются с временной зоной. Исправление текста протокола производится отдельной записью, а не перезаписью исходного журнала. Если журнал был сформирован единственным участником без внешней верификации, степень независимости такого доказательства следует честно обозначить.

Поскольку исходные произведения могут быть коммерчески ценными или содержать персональные сведения, публичному отчёту нужен отдельный слой обезличенных доказательств: агрегированные числовые таблицы, безопасные ссылки на записи в закрытом реестре, параметры повторяемого теста и перечень непроверенных гипотез. Защищённый депозитарий с журналом доступа и сроками уничтожения обеспечивает возможность последующей проверки без неконтролируемого распространения полного корпуса. Протокол доступа отделяет сотрудников заказчика от независимых рецензентов и регистрирует раскрытие потенциальных конфликтов интересов.

51. Договорные гарантии, права на аудит и пределы заверений

При закупке учебного набора либо коммерческой лицензии следует проверять не красоту гарантийной формулы, а её юридическую применимость. Заявление поставщика о наличии «all necessary rights» может быть ограничено договорным перечнем объектов, территорией, максимальным объёмом использования, датой получения контента или правом распространения производных массивов. Существенны право на machine learning, право на копии и преобразования, использование через субподрядчика, дообучение, сохранение весов, повторные версии модели, использование фрагментов в выходных результатах и передача доказательств аудитору. Ни одно из этих разрешений не нужно механически подразумевать при отсутствии соответствующей нормы или ясного договора.

Структура соглашения, ориентированная на последующий аудит, содержит манифест лицензируемых ресурсов, процедуру обновления перечня, представления о происхождении, сроки предоставления уведомлений о предъявленных претензиях, формат документов, право на ограниченную проверку и порядок оспаривания находок. Возмещение убытков и предел ответственности рассчитываются отдельно от обязательства предоставить доступ к доказательствам. Запрет предоставлять третьей стороне необоснованные исходники не мешает использовать защищённую data room, где аудитору разрешено просмотреть контрольную выборку без копирования всего массива.

Нельзя обещать, что лицензия одного посредника защищает от требований всех неизвестных авторов. Для сложных реестров разработчик требует документы о полномочиях посредника и способах распределения прав. Если источник фактически получен через агрегатор, обладающий только правом публичного показа, а не разрешением на воспроизведение для обучения, аудитор фиксирует конкретный пробел и возможные меры: дополнительная лицензия, исключение спорного каталога из будущих циклов, уточнение цепочки прав или правовая оценка применимого исключения. Оплата поставщику сама по себе не решает авторско-правовой вопрос.

Договорный вопрос	Минимальная доказательная форма	Последствие отсутствия
Кто предоставил права?	Полномочия сторон и карта правообладателей	Нельзя подтвердить chain of title
На какие версии и территории?	Перечень, лицензия, temporal scope	Вывод только для зафиксированной версии
Можно ли независимое тестирование?	Ограниченный доступ, NDA и журнал	Существенное ограничение заключения
Как реагировать на претензии?	Процедура уведомления и исправлений	Повышенный операционный риск

52. Правовое измерение открытых репозиториев и программного кода

Публикация программного кода в открытом репозитории не равнозначна отказу автора от прав. Открытая лицензия определяет объём разрешений и связанные обязанности, например сохранение уведомлений, раскрытие модификаций либо распространение производного результата на установленных условиях. В обучающих наборах отдельная проблема — вложенные зависимости, файлы без лицензии, копии сторонних библиотек и код, добавленный участником без необходимых полномочий. Политика типа «репозиторий помечен как open source, следовательно, весь код свободен» не выдерживает проверки даже на уровне исходных предпосылок.

Аудит проводится по слоям: лицензия проекта, лицензии подмодулей, содержимое vendored dependencies, временная версия, условия доступа и юридическая квалификация действий. Не следует автоматически утверждать, что всякое обучение на коде создаёт производное программное произведение или, наоборот, что никакое воспроизведение в модели невозможно. Это разные вопросы, для которых требуются техническая проверка и применимое право. Коммерческое использование открытого кода может быть допустимо при соблюдении условий лицензии, но сами условия нельзя игнорировать только из-за появления кода в наборе данных.

Для проверки генератора кода используется контрольная выборка редких функций, сигнатур, комментариев и лицензированных текстов уведомлений. Сходство коротких стандартных конструкций имеет иную доказательственную ценность, чем воспроизведение длинного уникального фрагмента; сравнительная методика учитывает функциональную необходимость, стандартность решений и доступный предшествующий код. В заключении нужно отдельно оценивать риск выдачи проприетарного кода, риск невыполнения условий лицензии и риск уязвимостей в сгенерированном коде. Последний не является автоматически авторско-правовым нарушением.

53. Происхождение фотографий, метаданные и контроль ложных атрибуций

Для фотоархивов и профессиональных иллюстраций типовая проблема заключается не только в отсутствии лицензии, но и в ошибочной атрибуции. Изображение может появиться на агрегаторе без подписи автора, с поддельными метаданными или под изменённым именем файла. Прямая ссылка на агрегатор не устанавливает цепочку прав. Исследователь фиксирует самый ранний доступный проверяемый источник и сопоставляет его с лицензиями, актами передачи, контрактами агентств и архивными публикациями; при конфликте показаний вывод остаётся предварительным. [6][20]

Техническое сравнение объединяет точные хеши, перцептивное сходство, сопоставление кадров, встречающиеся подписи и историю публикаций, но не выводит «вероятность авторства 98%» из одной метрики визуального сходства. Для ретушированных, обрезанных и рекомпрессированных изображений нужно показать, какие признаки сохранились. Условия платформ, на которой фотография появилась впервые, проверяются на историческую дату. Законная демонстрация снимка правообладателем не предполагает разрешение роботизированной выгрузки всей библиотеки.

Отдельно ведётся реестр изображений, которые нельзя приводить целиком в публичной версии доклада. При необходимости эксперту предоставляется защищённый идентификатор и краткое описание совпадения с минимальной иллюстрацией в пределах законно допустимого. Для портфолио и концертной съёмки особую роль играет авторская атрибуция: ошибочная подпись способна создавать не только имущественные, но и репутационные последствия. Поэтому корректирующая мера охватывает права, согласование кредитов и исправление опубликованных метаданных, а не ограничивается удалением одной копии файла.

54. Музыка и смежные права: разные слои разрешений

Обучение модели на песне может включать одновременно охраняемое музыкальное произведение, текст, исполнение и конкретную фонограмму. Права могут принадлежать разным субъектам; управление частью репертуара через коллективную организацию не доказывает полномочие разрешить машинное обучение всех его слоёв. Следовательно, юридический запрос к музыкальному каталогу должен раздельно перечислять объекты, территорию, срок, разрешённые способы использования, возможность изменения и передачи производной информации. Проверка лицензии на потоковое прослушивание сама по себе не устанавливает разрешение делать учебные копии либо использовать запись для настройки генератора. [6] [19]

Отдельный технический сценарий касается обнаружения длинных дословных фрагментов текста песни, а другой — сходства новой аудиозаписи с конкретной фонограммой. Правовые критерии для них различаются. Короткая последовательность распространённых слов не обладает той же доказательственной силой, что характерный защищённый текстовый фрагмент, а стилистическое сходство музыки не равно тождеству охраняемой записи. Эксперт описывает объективные признаки и метод сопоставления; юридическая квалификация учитывает оригинальность, содержание и обстоятельства. Решение мюнхенского суда по спору GEMA от ноября 2025 года имеет значение для обсуждения memorization и воспроизведения, но не является универсальным окончательным правилом для всех европейских судов. [11]

Отчёт должен различать авторско-правовой риск, право на голос и персональные сведения исполнителя, требования маркировки синтетического материала и возможные договорные запреты имитации. Попытка объединить все эти вопросы одним термином «copyright violation» делает заключение непригодным для суда. Рекомендуется оформлять отдельные оценки по слоям прав, с понятной цепочкой доказательств и конкретными адресатами возможных исправительных мер.

55. Издательский архив, цифровые библиотеки и лицензии доступа

В отношении издательских архивов принципиально разделять право читать содержание, воспроизводить его в исследовательском процессе, извлекать значительную часть базы данных и использовать материал для коммерческой генеративной модели. Условие подписки или открытая выдача поисковой системы не равнозначны передаче исключительных прав. В издательских каталогах возможны как права отдельных авторов и фотографов, так и права изготовителя базы и специальные права издателя прессы. Для юридического заключения требуется карта объектов с привязкой к конкретным статьям, версиям и датам извлечения. [6][7]

Аудит фиксирует тип доступа: публичная страница, подписка, институциональная лицензия, правомерная покупка, агрегатор либо зеркало. Наличие академического статуса у исследовательской организации не превращает любую коммерческую учебную задачу в исследовательский TDM по статье 3 DSM. При оценке статьи 4 необходимо проверить юридические требования и возможное резервирование прав. Предположение, что всякий robots.txt одинаково равен opt-out, слишком грубо: значения имеют содержание сигнала, условия использования и технологические обстоятельства. Решение принимается по конкретным фактам, а не по бинарному тегу.

Предлагается использовать защищённую пилотную выборку из ограниченного числа издателей и типов доступа, заранее зафиксировав показатели доли отсутствующих лицензий, конфликтов атрибуции, блокировок источников и времени рассмотрения претензий. При пересечении с персональными данными

авторское право и GDPR проверяются параллельно. Нельзя выводить законность обработки персональных данных только из наличия авторско-правовой лицензии и нельзя делать противоположный вывод без анализа основания обработки.

56. Репрезентативность тестов: зачем нужен отрицательный контроль

Протокол проверки memorization легко исказить, если выбирать только самые узнаваемые отрывки или многократно подбирать prompts до получения совпадения. Такой эксперимент показывает возможность воспроизведения в условиях поиска, но не распространённость явления для обычного использования. Для разных вопросов нужны разные генеральные совокупности: все выходные ответы, все произведения определённого каталога, все запросы с заданным контекстом либо все версии одной модели. Деноминатор объявляется до теста. Результаты намеренного извлечения, естественного использования и повторных запросов учитываются отдельно.

Отрицательный контроль включает тексты и формулировки, отсутствующие в анализируемом корпусе по известным аудитору основаниям, а также стандартные неоригинальные фразы. Он помогает оценить ложное совпадение метода. Положительный контроль может содержать заранее согласованные и законно используемые тестовые произведения, однако его наличие не доказывает качеств модели на чужих объектах. Несколько исследователей раздельно размечают существенность совпадений; несогласия рассматриваются по журналу, а не скрываются усреднённым итогом.

Результат «0 найденных совпадений из 300 попыток» не означает нулевой вероятности утечки. При допущениях независимости можно сообщить приблизительную верхнюю 95%-границу 3/300, то есть около 1%, но если запросы коррелированы, такая оценка не имеет указанной интерпретации. Столь же недопустимо интерпретировать «12 совпадений из 300» как долю всех незаконно использованных произведений в обучении. Технический тест отвечает на узкий вопрос о наблюдаемой выдаче; юридический аудит корпуса требует иной выборки.

Наблюдение	Что можно заключить	Что заключить нельзя
Совпадение после референсного input	Модель способна трансформировать предоставленный объект	Что она хранила этот объект в обучении
Совпадение без референса	Обнаружен воспроизводимый риск в условиях теста	Что все модели этого класса нарушают права
0/300 совпадений	В данном протоколе совпадений нет	Что риск равен нулю
Сходный стиль	Выявлено стилистическое сходство	Что обязательно воспроизведена охраняемая форма

57. Точки принятия решений: остановка, исправление, публикация

Хороший независимый аудит заканчивается не индексом риска, а ограниченным набором решений с ответственными лицами. Ситуация может требовать дополнительного документального запроса, ограничения использования спорного датасета, получения лицензии, исправления публичной сводки, настройки фильтров выдачи либо передачи материалов юристу. Для каждого решения указываются юридическое основание, непроверенные предпосылки, срок и критерий закрытия. Приоритет определяется не только вероятностью спора, но и масштабом возможного ущерба, характером произведения, возможностью повторного копирования и наличием срочных запретительных требований.

Эскалация не должна строиться на презумпции виновности. Если источник не предоставил лицензию аудитору, это ограничивает доказательства, но не равно признанию нарушения. Если договор обнаружен, но охватывает только демонстрацию материала, возникает документируемый пробел в обосновании обучения. Если резервирование прав получено после определённой даты, необходимо отдельно оценить последствия для нового сбора и уже созданных копий; нельзя автоматически переносить поздний сигнал на все предыдущие действия без юридического анализа. [1][21]

Для независимого Observatoire вводится разделение ролей: исследователь устанавливает факты, юридический эксперт квалифицирует риски, редактор оценивает безопасную публичную форму, а ответственное лицо разрешает спор об интересах и ресурсах. Лицо, ведущее коммерческий спор с проверяемой компанией, не должно единолично утверждать публичный вывод о ней. Публичное

сообщение содержит версию и дату методики, понятное описание ограничений, право проверяемой стороны представить документы и механизм исправления существенной ошибки.

58. Предметная проверка консультационных предложений ЕС

В консультации Technology and Copyright 2026 года целесообразно разделять меры, которые можно реализовать добровольно, и предложения, требующие изменения нормативной конструкции. Первые включают стандартизированный формат журналов разрешений, унифицированный ограниченный data room для независимого аудитора, машиночитаемые способы регистрации opt-out, описание происхождения источников в объёме, допускающем защиту коммерческой тайны. Вторые могут касаться процессуального доступа к доказательствам, механизмов компенсации и общеевропейских правил, которые невозможно навязать компаниям одним методическим докладом. Консультационный ответ не следует представлять как окончательный нормативный акт. [5]

Любое предложение о прозрачности нужно проверять на соразмерность: раскрытие полного списка URL иногда может облегчить неправомерное скачивание архивов, раскрыть личные данные или коммерческую тайну, тогда как публикация только общего слова «Internet» лишает правообладателей инструмента проверки. Рабочий компромисс — публичные категории и агрегированные доли плюс обоснованный закрытый доступ компетентным органам и независимому эксперту при строгих ограничениях на распространение. Метод подтверждения должен быть воспроизводим для небольших правообладателей, иначе инновация превращается в платный пропуск для крупных каталогов.

Предлагаемые показатели для будущей оценки: доля источников с подтверждённым правовым основанием, медианное время исполнения opt-out, число исправленных ошибок атрибуции, доля претензий с рассмотренным доказательством, ошибки автоматического классификатора прав и доля моделей с актуальными summary. При публикации следует фиксировать деноминаторы и период наблюдения, иначе сравнение участников будет вводить в заблуждение. Перед направлением позиции ARGA важно согласовать мандат, институциональный статус, степень представительности выводов и недопустимость выдачи предложенной методики за уже проведённое эмпирическое исследование.

59. Итоговая модель оценки доказательств без ложной точности

Для практической организации работы Observatoire полезна шкала доказательной определённости с четырьмя уровнями. На первом уровне имеются только утверждения поставщика и публичные маркетинговые материалы. На втором представлены проверяемые документы и выборочные журналы, но аудит не имел независимого доступа к корпусу. На третьем доступна защищённая контрольная выборка, воспроизводимые испытания и независимое сопоставление документов. На четвёртом возможна расширенная внешняя проверка релевантных стадий жизненного цикла с достаточной полнотой, временной привязкой и независимыми рецензентами. Номера уровней описывают полноту доказательств, но не заменяют юридическую квалификацию.

Публикация не должна превращать отсутствие ошибки в доступной выборке в «сертификат законности» всей модели. Любая найденная проблема должна быть привязана к конкретному объекту, юридическому режиму, стадии использования, периоду и степени доказанности. Фактический вопрос «использован ли файл» не совпадает с юридическим «было ли это запрещено». Отсутствие коммерческой лицензии ещё не исключает применения законного исключения; наличие лицензии одной юрисдикции не гарантирует её универсальной пригодности за рубежом. [1][2][6]

Рекомендации о выплатах и корректирующих мерах могут включаться только как предложения с указанием основания и компетенции, а не как обязательные решения частной экспертной организации. Финальная структура заключения должна позволять третьему лицу повторить ход проверки: перечень материалов, решение о выборке, результаты, альтернативные гипотезы, ограничения, особые мнения и итоговые действия. Именно такая доказательная дисциплина формирует доверие к Observatoire лучше, чем искусственная оценка «98,7% legal AI» без определённого смысла.

Уровень	Характер доказательств	Формулировка публичного вывода
1	Самодекларация и открытые страницы	«Поставщик заявил; независимой проверки не было»
2	Договоры и частичные журналы	«Документально подтверждена часть процедур; есть ограничения»
3	Контрольная выборка и повторяемые тесты	«Проверены указанные выборка, версии и критерии»
4	Расширенный независимый доступ и рецензия	«Проверены конкретные элементы в установленном объёме, не весь будущий выпуск»

60. Отдельные исключения: цитирование, временное копирование и TDM

В дискуссии об обучении моделей часто смешивают несколько юридически различных ситуаций: техническое копирование при передаче данных, извлечение массива для автоматизированного анализа, воспроизведение ограниченной цитаты в тексте научного исследования и выдачу пользователю протяжённого фрагмента охраняемого произведения. Для каждого случая действуют собственные условия. В частности, исключения TDM из статей 3 и 4 Директивы 2019/790 относятся к воспроизведению или извлечению в целях анализа текста и данных и не являются универсальным разрешением на все будущие формы распространения исходного произведения. [1][6]

При оценке исключения необходимо последовательно проверить законный доступ, вид действия, категорию пользователя и цели, наличие резервирования по статье 4, требования хранения копий и связь спорного использования с применимым национальным правом. Общее право цитирования, где оно действует, имеет собственные условия и не заменяет такую проверку. Технический факт размещения исходника в поисковом индексе не доказывает, что создание полной тренировочной копии входило в соответствующее исключение. И наоборот, наличие иной законной лицензии может сделать анализ условий TDM не единственным основанием.

В независимом заключении рекомендации должны быть операционными: для каждого спорного действия выбрать одну из категорий «разрешено договором», «потенциально охвачено исключением при доказанности условий», «основание не подтверждено» либо «установлено противоречие конкретному требованию». Такие состояния не следует переводить в математический скоринг. Если законность результата зависит от спорного толкования, рекомендуется привести аргументы обеих сторон и пределы судебной практики, а не имитировать окончательный ответ.

61. Коллективное управление правами: доступность репертуара и мандат

Коллективное лицензирование потенциально снижает транзакционные издержки при работе с большими творческими каталогами. Но само существование организации коллективного управления не означает, что она вправе разрешить обучение по всем находящимся в обороте произведениям, во всех странах и во всех способах использования. Практическая проверка начинается с её юридического статуса, объёма полномочий правообладателей, соглашений о взаимном представительстве, предметного репертуара и правил распределения доходов. Предложение нового вида лицензии должно отличаться от утверждения, что такая лицензия уже существует и обязана признаваться третьими лицами.

Для генеративной модели особенно важны границы опосредованного управления: составные произведения, музыкальные тексты и звукозаписи, произведения с несколькими авторами, издательские каталоги, трансграничные права и отозванные мандаты. Если лицензия применима лишь к одному слою прав, это указывается отдельно. Необходимо также проверить возможности правообладателя заявлять резервирование использования, условия выхода из каталога и доступность эффективного механизма оспаривания распределения. Исследовательская модель роялти должна объяснять, почему определённый способ расчёта выбран, и признавать случаи отсутствия наблюдаемой причинной связи с доходом модели.

В предложениях для консультации ЕС можно описать технически нейтральный реестр мандатов и проверяемых платежей, но нельзя исходить из автоматического создания общеевропейской обязательной схемы без законодательного основания. Малые авторы и исполнители должны иметь сопоставимую возможность проверять учёт своего репертуара; иначе общий денежный поток будет видимым, а

фактическое распределение — непрозрачным. Необходимо обособить стоимость аудита от прав на произведения и предусмотреть право на исправление неверной записи.

62. Право издателей прессы и редакционные базы данных

Права издателя прессы в рамках права ЕС и правовая охрана базы данных образуют самостоятельные вопросы наряду с правами автора конкретной статьи. Даже когда факт, изложенный в новости, не защищён авторским правом, отбор, редакционное изложение, фотография и значительный объём базы могут охраняться. Разработчик не должен классифицировать архив только по признаку «тексты доступны в поиске»; необходимо отделять источники, приобретённые по коммерческому договору, подписные библиотеки, открытые официальные документы и материалы с лицензией Creative Commons, включая их возможные ограничения.

Практический кейс касается модели, обучаемой на агрегированных журналистских архивах. Для каждой коллекции требуется установить поставщика, отношения с издателем, источник прав на фотографии и авторские тексты, а также то, подпадает ли конкретное извлечение под установленное исключение. В запросе к архиву не нужно требовать раскрыть всех журналистов и редакторов, если достаточно проверенного мандата на конкретные категории. Однако общие заверения агрегатора без возможности проверить хотя бы выборку недостаточны для позитивного независимого заключения.

Рынок лицензий на новостные архивы нельзя оценивать только по цене подписки для читателя: продукт обучения может создавать иной спрос и иной рынок. Для доказательства ущерба необходимы договоры, динамика использования и альтернативные объяснения. При публикации исследования следует избегать цитирования длинных охраняемых материалов: названия, даты, идентификаторы и агрегированные оценки обычно позволяют изложить метод без воспроизведения всего корпуса. [6][7]

63. Институциональные и научные библиотеки: научный статус не является общим разрешением

Исключение исследовательского TDM по статье 3 DSM тесно связано с квалификацией пользователя и цели. Участие университета, библиотеки или научной организации само по себе не означает, что любое действие подрядчика, коммерческого партнёра или отдельного разработчика автоматически попадает в исключение. Исследователь должен документировать институциональную роль, характер проекта, законный доступ, условия хранения и возможность защищённого воспроизведения. Если научный центр передаёт созданный набор данных коммерческому партнёру, это отдельный юридически значимый переход, а не техническая мелочь.

Для внутренней проверки готовится таблица источников: приобретённые электронные книги, научные статьи, архивы из открытого доступа, материалы с разрешительными лицензиями и библиотечные сканы. У каждого вида могут быть собственные ограничения, а право читать экземпляр не всегда тождественно праву передать файлы сторонней модели. Неполный договор не следует дополнять выдуманным разрешением на массовый анализ. При этом защита научных исследований требует не превращать каждое законное техническое копирование в обязательную коммерческую лицензию без проверки применимого исключения.

Методическая рекомендация заключается в прозрачном разделении академической проверки и заказной коммерческой разработки. Когда используется смешанная инфраструктура, журнал фиксирует, кто инициировал извлечение, на чьи средства, для какой задачи, кто получил копию, мог ли партнёр использовать данные повторно и кто отвечает за последующее удаление. Отрицательный вывод должен быть привязан к отсутствующему правовому условию, а не к субъективной оценке коммерческой природы ИИ.

64. Достоверность метаданных, provenance-маркеров и машинных подписей

Метаданные о происхождении контента полезны, но по своей природе являются утверждениями, подлежащими проверке. Подпись, прикреплённая к файлу, может удостоверять факт изменения файла с момента подписания определённым ключом; она сама по себе не устанавливает, что подписант обладал

всеми исключительными правами или имел разрешение на использование чужого произведения. Аналогично стандартные метаданные о создателе могут быть неполными, копированными или устаревшими. Поэтому аудит объединяет технические доказательства целостности с договорными и организационными документами.

При построении графа происхождения система должна регистрировать как позитивные, так и негативные сведения: кто подписал, какой объект, на каком этапе цепочки, в какой версии, какие поля подтверждены независимым источником и какое событие стало основанием для изменения. Если изображение лишилось части метаданных после конвертации, отсутствие полей не доказывает отсутствия автора. Необходимо сохранять альтернативные пути подтверждения — договора, первичные публикации, архивные квитанции, реестры и свидетельства передачи.

Обязательная публикация полного графа не всегда соразмерна: часть записи может содержать персональные данные, конфиденциальные коммерческие условия и внутренние идентификаторы. Для общественной проверки достаточно стандартизированного описания классов доказательств и доступного механизма независимой проверки с разграничением прав доступа. Метод описания provenance не должен превращаться в технологическую монополию одного поставщика решений.

Метка	Доказательное значение	Ограничение
Криптографическая подпись	Целостность относительно подписанного состояния	Не подтверждает все исключительные права
EXIF/авторская подпись	Возможный источник атрибуции	Может быть изменена или отсутствовать
Контракт/акт передачи	Объём согласованных прав	Нужно проверить полномочия и пределы
Журнал загрузки	Время и путь появления файла	Не гарантирует законность копирования

65. Дедупликация и агрегирование: доказательная цена потери исходного следа

Подготовка обучающих наборов нередко включает дедупликацию, фильтрацию и удаление исходных URL. Эти процессы полезны для качества моделей, но могут разрушать возможность последующей правовой проверки. Если система хранит только один «представитель» группы похожих объектов, auditor должен знать, какие оригиналы, лицензии и opt-out сигналы вошли в группу. Иначе правомерная копия способна скрыть вторую, полученную из источника с запрещением, либо наоборот: спорный источник ошибочно окрасит в красный всю группу законно лицензированных версий.

Техническая схема включает индекс исходных идентификаторов, дату и метод группировки, правило выбора репрезентативного файла и механизм восстановления provenance без сохранения всех исходных данных в открытой зоне. Каждый merge, split и delete должен иметь версионное событие. При рекомпрессии или нормализации текстов фиксируются трансформации, которые могут повлиять на проверку тождества и авторства. Сопоставление по приблизительным отпечаткам сопровождается журналом ложных позитивных и ложных негативных совпадений.

В аналитическом заключении не следует писать «в корпусе нет указанного произведения», если на самом деле проведён поиск только по одному хешу, а все преобразованные варианты не проверены. Равно нельзя говорить «произведение незаконно использовалось», когда найден лишь похожий файл без установленного права. Для каждого результата указывается полнота индексирования и проверенные преобразования, чтобы юридический эксперт понимал доказательственный предел технической процедуры.

66. Международные версии произведения и право перевода

Переводы, адаптации, сокращённые публикации и локализованные издания могут иметь собственные объекты охраны и цепочки прав. Лицензия на оригинальный роман не всегда охватывает перевод третьего лица, а разрешение на один язык и рынок не обязательно распространяется на весь многоязычный тренировочный корпус. Для автоматического анализа необходимо отделять авторство исходного произведения и вклад переводчика, существование авторско-правовых исключений и конкретные договорные ограничения.

Пример: учебный набор содержит английский оригинал статьи, французский перевод издателя и машинный русский перевод, созданный сотрудником подрядчика. У каждого элемента фиксируются источник, дата, способ создания и правовая роль участника. Нельзя автоматически предполагать, что машинный перевод свободен от прав на оригинал; равно некорректно без анализа считать любую автоматическую переработку самостоятельным охраняемым произведением. В итоговом графе отдельно показываются происхождение содержания, творческий вклад, лицензионные разрешения и зависимость от исходной версии.

При аудите результатов модели следует заранее определить, что тестируется: дословное воспроизведение оригинала, перевод защищённого фрагмента, пересказ факта или отдельная творческая переработка. Метод сходства на одном языке не применяется механически к другому; критерии существенности устанавливаются специалистами, знакомыми с соответствующим правовым и языковым контекстом.

67. Обработка исправлений и удаления без ложных технических обещаний

В ответах правообладателям часто встречаются взаимоисключающие заявления: «исходный файл удалён из хранилища» и «модель больше не содержит сведений из произведения». Первое может быть подтверждено операционным журналом, второе требует отдельного анализа и зачастую вообще не может быть доказано универсально. Веса обученной модели не являются каталогом файлов, поэтому простое удаление исходника из будущего набора не равнозначно обратному применению обучения. Возможны фильтры вывода, изменение набора для будущего обучения, технические процедуры unlearning или переобучение, но каждую меру следует описывать по действительному результату.

Аудитор фиксирует дату претензии, релевантную версию модели, перечень прекращённых загрузок, применённые защитные фильтры, проведённые тесты, оставшиеся риски и план повторной проверки. Если исключение сведений из будущих версий выполнено, это следует указать без обещания полной невозможности любой генерации. Модель может воспроизводить сходную форму и вследствие иных источников или общего характера выражения; такой результат необходимо отличать от доказанного сохранения конкретного исходного файла.

Документ для правообладателя должен перечислять выполнимые меры, а также доступные способы дальнейшего правового оспаривания. Реакция на авторско-правовую претензию и рассмотрение запроса субъекта персональных данных по GDPR имеют различные основания и сроки; объединение их в одну неразличимую «процедуру удаления» создаёт риск неправильного ответа. [20]

68. Экономика соблюдения правил для небольших поставщиков и авторов

Не все участники способны содержать внутреннюю юридическую службу, команду аудита и собственный хранилище лицензий. Это имеет значение для оценки будущих механизмов прозрачности: слишком дорогой режим может формально одинаково требовать действий от международной модели и небольшого разработчика, но фактически исключать последних из рынка. Вместе с тем низкая стоимость работы не создаёт общего права использовать чужие произведения без соблюдения закона. Поэтому исследовательская модель должна учитывать пропорциональность инструментов, не снижая необходимый уровень защиты правообладателей.

Предлагаемый минимальный пакет для небольшого поставщика: версионный реестр источников, стандартный парсер machine-readable резервирований, журнал изменений лицензий, публичное описание категорий обучения и процедура обращения правообладателя. Для независимой оценки допустим стратифицированный ограниченный аудит с честным указанием интервала неопределённости; дорогостоящая проверка всего корпуса не должна объявляться обязательной законодательной нормой, если такого требования нет. При дефиците доказательств публикация выводов должна становиться осторожнее, а не смелее.

Для малых авторов важны доступный бесплатный канал opt-out, понятный ответ и возможность проверить судьбу своего конкретного произведения без оплаты коммерческому посреднику. В числе обсуждаемых в консультации ЕС мер может быть стандартный формат претензий, но его использование не должно

блокировать иные юридически допустимые формы обращения. Индикаторы качества должны включать и стоимость реакции, и число обоснованно удовлетворённых претензий.

69. Оспаривание экспертного заключения и процедурная справедливость

Независимое исследование, утверждающее возможные нарушения определённой компании, может повлиять на её репутацию и судебные позиции. Поэтому до публичного выхода должны существовать право на исправление фактической ошибки, разумный порядок предоставления релевантных документов и независимый механизм разрешения конфликта интересов. Право представить возражения не означает право проверяемой стороны наложить вето на исследовательский вывод. Но отсутствие ответа не может преобразовываться в признание вины; формулировка должна обозначать, какие обращения направлялись и какие документы фактически получены.

Для редакционной проверки составляется список утверждений с уровнем доказанности. Отдельно помечаются факты, выводы на выборке, мнение эксперта и нормативные предложения. К каждому существенному факту прикладывается ссылка на доказательство, а к правовому выводу — используемое право и дата среза. Если один эксперт считает признаки нарушения достаточными, а другой не согласен, причины разногласия отражаются в рецензии, не скрываются под «коллективным единогласием».

Публичная поправка должна оставлять историю исправления: какая версия была опубликована, что исправлено и почему. Тихая замена PDF без changelog разрушает доверие к независимой организации, особенно в дискуссии о правах и технологических корпорациях. Эти процедуры применимы и к случаям, когда заказчиком исследования выступает потенциальный истец; финансирование и роль заказчика необходимо раскрывать в допустимом объёме.

70. Доказательная шкала для юридической консультации и судебного материала

Рабочий отчёт для научной аудитории и процессуальное доказательство не тождественны. В научном тексте допустимо сформулировать гипотезу, отметить пределы выборки и предложить проверку; судебное дело требует следовать нормам допустимости и представления доказательств конкретной юрисдикции. Эксперт Observatoire не должен обещать, что его отчёт автоматически будет принят судом или регулятором, и не может подменить адвоката, имеющего полномочия представлять сторону. Наличие нотариального заверения снимка экрана также не решает вопрос материальной законности контента.

Для возможного судебного использования требуется пакет происхождения доказательства: кто получил, когда, на какой системе, какие изменения вносились, каким образом подтверждена целостность, какие лица имели доступ. Методы эксперимента документируются до тестов, включая критерии отбора запросов и виды контрольных наблюдений. Должна существовать возможность воспроизвести значимые вычисления без нарушения исключительных прав на сами тестовые объекты.

Результат может включать четыре варианта: достаточность для исследовательского утверждения, необходимость дополнительной проверки, невозможность надёжного вывода и прямое опровержение первоначальной гипотезы. Ни один из них не является сам по себе процессуальным решением. При трансграничном деле квалификация доказательства и способ его получения требуют отдельной локальной юридической оценки.

71. Сводный набор минимальных мер для ARGA Observatory

Практическое внедрение доклада предполагает три независимых функции: юридическое исследование, технический аудит и редакционную публикацию. Юридическая функция проверяет нормы, лицензии, применимость исключений и компетенцию; техническая — происхождение источников, версии модели, reproducibility и журналы; редакционная — соразмерность раскрытия, отсутствие недоказанных обвинений, репутационные риски и возможность исправления. Подмена всех функций одним специалистом без peer review особенно рискованна, если организация одновременно оказывает платные консультации заинтересованным сторонам.

Для пилота рекомендуется ограничить предмет несколькими видами данных и источников. Например, одна категория издательских материалов, одна группа фотоархивов и одна коллекция открытого программного кода позволяют проверять разные цепочки прав при управляемом объёме. До старта фиксируется допустимая форма доступа к данным, условия NDA, критерии статистического отбора, формат evidence ledger и пределы публичного вывода. Неисследованные категории прямо обозначаются в отчёте; пилот не должен изображать сертификацию всей модели.

Итоговый документ должен давать руководству решение по существу: какие процедуры можно запустить уже сейчас на основе существующего закона и добровольных соглашений; какие меры требуют отдельного договора или ресурса; какие предложения следует вынести на публичную консультацию; и какие идеи пока остаются исследовательской гипотезой. Подотчётность в области авторского права состоит не в максимальном числе таблиц, а в возможности проверить каждый существенный вывод и вовремя исправить ошибку.

Практические приложения и исследовательские протоколы

Приложение 1. Паспорт корпуса обучающих данных

Паспорт корпуса является верхнеуровневой карточкой, по которой можно установить ответственность за набор и понять, какие версии модели получали материалы. Заполняется до первого использования и обновляется при существенном изменении состава или разрешённого назначения. Один паспорт на все корпуса компании непригоден, если смешивает лицензированные данные, TDM-источники и закрытые клиентские сведения. Минимальная единица учёта — управляемый набор со своей историей происхождения и назначением.

Карточка должна содержать владельца, юридическое лицо, контакт для запросов правообладателей, дату формирования, диапазон дат сбора, модель и версию, задачи использования, реестр upstream поставщиков, классификацию рисков и статус утверждения. В отдельном поле фиксируются неизвестные источники и причины неполноты сведений. Подписант подтверждает полноту предоставленных ему доказательств, а не абсолютную свободу всех прав во всём мире. При расширении применения набора для новой модели проводится повторная юридическая оценка.

Поле паспорта	Формат записи	Проверочное правило
Dataset ID и версия	Уникальный код, дата, ответственный	Связь с manifest и резервной копией
Назначение	Pre-training / fine-tuning / evaluation / RAG	Запрет иных целей без review
Источники	Лицензия / DSM3 / DSM4 / иное основание	Обязательная ссылка на реестр
Риски	По категориям права и поставщика	Оставить отдельный флаг «неизвестно»
Решение	Допуск / условный / карантин	Подпись и дата, версия регламента

Приложение 2. Акт приёма нового источника

Приём источника начинается с определения фактического владельца процесса получения. Нельзя принимать миллион записей по одной заявке «источник открытый», не проверив юридическое основание, машиночитаемые запреты и права на повторное использование. В акте должны фиксироваться точные домены, конечные URL после переадресаций, период обхода, категория контента, способ аутентификации, правила лицензии и технические средства сбора.

Приёмка проводится в два независимых этапа. Технический специалист подтверждает, что журнал crawl соответствует реальным объектам и корректно обрабатывает сигналы. Юридический специалист решает, можно ли применять конкретную лицензию или исключение. При споре доступ ограничивается до разрешения. Данные, для которых невозможно установить происхождение либо явно показаны запрещённые источники, не должны попадать в обучение только потому, что уже скачаны.

Проверка	Условие прохождения	Документ
Идентичность	Определён источник и поставщик	Регистрационная карточка
Доступ	Подтверждена допустимость доступа	Условия на дату сбора
Права	Основание связано с видом операции	Меморандум IP
Robots/opt-out	Отработаны правила на дату действия	Машинный журнал и его хеш
Конфиденциальность	Исключены запрещённые категории	Двойная оценка IP/GDPR

Приложение 3. Журнал резервирования прав и обработки opt-out

Журнал нужен для ответов на четыре вопроса: кто выразил волю, в какой форме, когда сигнал стал доступен и какие активы исключены либо дополнительно проверены. Для каждого события создаётся неизменяемая запись со снимком исходного сигнала, версией интерпретатора, статусом верификации полномочий и результатом. Система должна работать как с автоматическими файлами сайта, так и с индивидуальными мотивированными обращениями, не смешивая их юридические последствия.

Если сигнал обнаружен после первоначального копирования, журнал не должен механически переписывать прошлую историю. Сохраняется дата первоначального доступа и отдельно дата изменения статуса для будущих операций. При изменении условий проводится юридический анализ соответствующих последующих действий. Отдельно регистрируются коллизии между зеркальными копиями, псевдонимами автора, синдикацией и правами нескольких соавторов. Уведомление о результате позволяет заявителю проверить, что его обращение не растворилось в автоматической системе.

Событие	Обязательные данные	Решение
Машиночитаемый сигнал	URL, timestamp, snapshot, parser version	Принят / требует разбора
Уведомление автора	Объект, полномочия, контакт	Проверка прав и масштаба
Конфликт прав	Несколько заявителей	Заморозка спорной категории
Downstream маршрут	Наборы/модели-получатели	Исключение из будущих сборок
Закрытие	Мотивировка, дата, ответственный	Отправить уведомление и сохранить лог

Приложение 4. Матрица лицензии и способов использования

Матрица лицензии связывает формулировки договора с реальными техническими операциями. Она не заменяет договор, а служит инструментом контроля исполнения и обнаружения пробелов. Например, право «отображать материалы внутри сервиса» не должно автоматически переводиться в разрешение для последующего обучения универсальной модели. Каждая строка матрицы должна содержать текст соответствующего разрешения, юридическую интерпретацию, ограничения, период действия и подтверждающий пункт контракта.

Особого внимания требуют объекты с несколькими правообладателями, аудиовизуальные произведения, переводы, фотографии и независимые лицензии на данные в составе базы. Юрист согласует матрицу с владельцем продукта: условия могут зависеть от конкретной архитектуры, территории, численности конечных пользователей и наличия публичных весов. При неясности компания должна получить дополнение к договору или не выполнять спорное действие; экономическая необходимость не восполняет отсутствующее право.

Способ использования	Статус	Основание и ограничение
Загрузка и техническая копия	Да / нет / неясно	Пункт договора или исключение
Предобработка и индекс	Да / нет / неясно	Место, срок, доступ
Обучение базовой модели	Да / нет / неясно	Модели и назначение
Fine-tuning новой версии	Да / нет / неясно	Дополнительные разрешения
Downstream передача	Да / нет / неясно	Сублицензия и территория

Приложение 5. Протокол доказательства правомерного доступа

Доступ к веб-странице должен доказываться обстоятельствами, существовавшими на момент релевантного копирования. Аудитор фиксирует точный URL, источник редиректа, состояние подписки, технический способ получения и текст условий на соответствующую дату. Архивная копия может иметь доказательственное значение, но обязана маркироваться как внешний исторический источник с оговоркой о полноте. Более поздний скриншот не подтверждает автоматически условия годичной давности.

Для подписных изданий дополнительно проверяется, что полномочие пользователя распространялось на автоматизированный анализ, а не только индивидуальное чтение. Для открытых сайтов — наличие ограничений и opt-out. При отсутствии надёжной информации вывод формулируется как «правомерный доступ не подтверждён», а не «доказана кража». Фактический обход аутентификации и получение материала из пиратского архива фиксируются отдельно и передаются юридическому эксперту для квалификации.

Артефакт	Что подтверждает	Основной предел
Лог HTTP и код ответа	Технический маршрут получения	Не показывает право на копирование
Подписка и её условия	Вид законного доступа	Может ограничивать автоматизацию
Снимок licensing terms	Режим на фиксированную дату	Требует проверки подлинности
Правила opt-out	Воля правообладателя/сервиса	Нужна правовая интерпретация
Hash скачанного файла	Идентичность копии	Не доказывает законность источника

Приложение 6. Due diligence поставщика обучающих данных

Перед закупкой датасета следует запросить карту upstream источников, сведения о полученных правах, условия сублицензии, историю претензий и доказательства исполнения ограничений правообладателей. Заявление о соблюдении закона полезно как договорное обязательство, но без доступа хотя бы к выборке первичных документов не снимает риск. Особое внимание уделяется данным, собранным сторонними роботами, потому что подрядчик мог потерять либо проигнорировать сигналы резервирования при подготовке массива.

Проверка подрядчика должна учитывать финансовую устойчивость, возможность сохранения документов после прекращения договора и обязанности сотрудничать при претензиях. Независимая выборка проверяет небольшие и крупные источники; следует включить активы с повышенным риском незаконного происхождения. Условие о возмещении убытков без процедуры передачи provenance является неполноценным. Внутреннее заключение обязательно указывает, какие доказательства предоставлял сам поставщик, а какие были независимо подтверждены.

Запрос	Риск непредоставления	Возможный контроль
Список категорий upstream источников	Невозможно определить происхождение	Независимая выборка
Право сублицензии	Разрыв цепочки разрешений	Проверка договоров
Opt-out log	Недоказанное соблюдение статьи 4 DSM	Тест на выбранных объектах
Правила удаления	Неисполняемая ремедиация	Технический walkthrough
План выхода подрядчика	Потеря доказательств	Обязательство передачи журнала

Приложение 7. Чек-лист публичной сводки по статье 53(1)(d)

Форма публичной сводки должна охватывать именно те модели и версии, которые указаны на странице поставщика. Аудитор сопоставляет её с обязательным шаблоном AI Office, опубликованным в 2025 году, и внутренним реестром происхождения, не подменяя этот нормативный минимум собственными требованиями. Важен правильный охват категорий: публичные датасеты, веб-сбор, частные лицензированные наборы, пользовательский и синтетический контент, существенные последующие обновления. [3][17]

Когда определённые сведения о старых моделях объективно недоступны, документ должен содержать причину и предел пробела, если такое объяснение допускает соответствующий режим. Отказ раскрывать коммерческую тайну не освобождает автоматически от соблюдения структуры шаблона. Публичная форма и закрытая доказательственная ведомость должны быть сопоставимы по идентификаторам, но не обязаны содержать одинаковую глубину. Аудитор не утверждает законность всех произведений на основании того, что сводка выглядит законченной.

Раздел	Проверочный вопрос	Красный флаг
Идентичность модели	Версия и поставщик указаны?	Сводка относится к неясному продукту
Категории источников	Все существенные группы отражены?	Скрыт весь web-crawl как «прочее»
Обновления	Даты и изменения документированы?	Нет delta после fine-tuning
Пробелы	Ограничения объяснены?	Утверждение абсолютной полноты без манифеста

Приложение 8. Технический граф происхождения и контроль преобразований

Граф provenance должен связывать исходный объект с каждой производной записью: OCR-извлечением, нормализацией, разбивкой на блоки, удалением дублей, эмбеддингом и обучающим batch. При переработке файла его hash меняется, поэтому хранение одного контрольного значения без parent relations не обеспечивает прослеживаемости. Аудитор должен иметь возможность отобрать случайный элемент в training manifest и восстановить путь к исходному источнику и связанному правовому основанию.

Для защиты персональных данных журнал не должен превращаться в публичный каталог лиц и произведений. Публичному отчёту достаточно агрегатов и проверяемых процедур, а закрытый доступ к конкретным объектам предоставляется по необходимости и под контролем. Особенно важен механизм отзыва: изменение юридического статуса исходного актива должно инициировать проверку всех downstream производных. Это можно тестировать на синтетическом контрольном наборе без реальных охраняемых работ.

Узел	Обязательные связи	Тест независимого аудитора
Source object	URL/договор/hash/дата	Восстановить юридическую карточку
Extracted text	Родитель и инструмент извлечения	Проверить неизменность связи
Dedup group	Все исходные объекты	Не потерять условия разных лицензий
Training batch	Версия датасета и активы	Обратная трассировка случайной записи
Model version	Использованные batches	Определить область затронутого риска

Приложение 9. Протокол дообучения и обновления версии модели

Дообучение способно изменить применимые документы и публичную прозрачность. В протоколе фиксируется базовая модель, юридический поставщик, назначение дополнительного корпуса, его категории прав, сумма новых активов, исключённые источники, результаты тестов memorization и дата будущего обновления сводки. Нужно юридически оценить, превращает ли модификация конкретного downstream участника в поставщика новой модели общего назначения по применимым критериям AI Act. Не всякая настройка интерфейса означает такую смену статуса. [2][3]

Обязательной частью является delta registry: какие активы добавлены, заменены или удалены относительно предыдущей версии. Нельзя просто перенести старое заключение об аудите при значительном изменении данных. Проверка может ограничиваться новыми участками при наличии доказательства, что остальная цепочка не изменилась; но это доказательство должно быть положительным и документированным, а не основанным на заверении разработчика.

Поле обновления	Значение	Решение
Base model ID	Юридическое лицо и версия	Проверить применимость

Поле обновления	Значение	Решение
		обязанностей
Added data	Категории и provenance	Разрешение/исключение/opt-out
Removed data	Основание и подтверждение	Остаточная оценка весов
Public summary	Что и когда пересмотрено	Публикация по режиму шаблона
Independent retest	Объём выборки и результаты	Допуск, ограничения, эскалация

Приложение 10. Испытание риска воспроизведения охраняемого текста

Протокол определяет фиксированный репертуар произведений, статус прав, выборку фрагментов, допустимые запросы, число повторов и критерии оценки до запуска эксперимента. Нельзя включать длинный оригинальный текст в prompt, а затем считать его появление в ответе доказательством memorization. Отдельно выполняются обычные запросы и направленные попытки восстановления, с раздельной статистикой и качественным анализом оригинальных элементов.

В тестовых логах сохраняются версия модели, время, настройки генерации, параметры фильтров, точная формулировка запроса и метрики совпадения. Запрещается публиковать длинные охраняемые пассажи в открытом докладе лишь ради демонстрации результата. Юридический эксперт оценивает существенность охраняемой формы, технический — воспроизводимость, а редактор — необходимость раскрытия и риск вторичного нарушения. При отрицательном результате указывается конечный охват тестов, а не гарантия невозможности воспроизведения.

Критерий	Метод	Допустимый результат
Репертуар	Предварительно зафиксированная выборка	Исключить cherry-picking
Обычный запрос	Контекст без длинного оригинала	Частота ненамеренного воспроизведения
Целевая проверка	Ограниченное число заранее описанных запросов	Риск извлечения при атаке
Сопоставление	Уникальность, длина, идентичность	Качественный правовой анализ
Отчёт	Интервалы, ограничения, исходные логи	Воспроизводимость без утечки контента

Приложение 11. Форма регистрации авторско-правовой претензии

Карточка претензии должна обеспечить проверку права, предмета и технического маршрута, не требуя от заявителя доказать невозможные внутренние сведения. Принимаются данные об авторстве, ссылке на оригинал, предполагаемом использовании, характере вреда и предпочтительном способе ответа. Заявителю сообщаются номер, дата регистрации, запрошенные недостающие сведения и контакт ответственного. Лишние идентификационные документы не собираются без необходимости.

Внутри организации претензия сопоставляется с asset registry и версией модели. Если исходный объект не находится, результат обозначается «не обнаружен по доступным доказательствам», а не «его никогда не было». При обнаружении активного правового конфликта сохраняются относящиеся к делу журналы, технически ограничивается дальнейшее распространение доказанного воспроизведения и проводится правовая оценка разрешённых мер. Ответ содержит основания и процедуру пересмотра, не раскрывая чужие данные.

Поле обращения	Заполняется	Проверка
Заявитель	Автор / представитель / CMO	Полномочия на заявленные права
Произведение	Ссылка, идентификатор, доказательство	Различить произведение и копию
Поведение модели	Публичный вывод или сбор	Привязка к версии и времени
Запрашиваемая мера	Ответ, блокировка, переговоры	Юридическая реализуемость
Закрытие	Мотивированное решение	Возможность апелляции

Приложение 12. Chain of custody для цифровых материалов

Каждый переданный аудитором файл получает неизменяемый идентификатор, исходный hash, временную метку, описание маршрута получения и перечень лиц с доступом. При копировании создаётся производная копия для анализа, исходная сохраняется отдельно. Если материал получен через экранную запись или архивную страницу, фиксируются технические ограничения формата, чтобы не выдать вторичный образ за исходный документ.

Когда доказательство включает спорные произведения, доступ ограничивается необходимым кругом проверяющих. Перед публикацией отчёта используются краткие описания и хеши, если полного фрагмента не требует законная процедура. При значимых процессуальных действиях привлекается профильный юрист для оценки допустимости доказательств в конкретной юрисдикции. Нарушение цепочки хранения должно быть раскрыто, а не замаскировано восстановленной позже датой.

Событие	Обязательная запись	Ответственность
Получение	Источник, канал, дата, хеш	Custodian
Передача	Отправитель, получатель, цель	Оба участника
Преобразование	Версия инструмента, новый хеш	Технический эксперт
Публикация	Объём раскрытия и основание	Редактор + юрист
Архивация	Срок, доступ, уничтожение	Контролёр доказательств

Приложение 13. План статистической выборки закрытого корпуса

До проверки составляется карта популяции: количество файлов, уникальных произведений, поставщиков, языков, типов контента и периодов сбора. Стратификация учитывает как долю соответствующей категории, так и её риск. Например, небольшой лицензированный музыкальный каталог может требовать отдельной страты из-за высоких потенциальных последствий даже при малой доле в общем корпусе. При отсутствии полного списка исходной популяции репрезентативность не подтверждается.

Случайная выборка формируется воспроизводимым методом; источник случайности и seed регистрируются. Прицельные объекты, выбранные из жалоб, тестируются отдельно. Результат разделяет отсутствие доказательства, нарушение договорного условия и установленное техническое несоответствие. Статистические интервалы указываются лишь тогда, когда выполнены предпосылки соответствующей модели. Вывод не переносится на закрытые 40% корпуса без дополнительных данных.

Параметр	Техническое решение	Предупреждение
Популяция	Дедуплицированный asset registry	Повторяющиеся копии искажают долю
Страты	Право, источник, период, риск	Не заменять рисковые группы средней частотой
Seed	Документируемый генератор	Нельзя менять после результатов
Пропуски	Отдельная категория	Не считать «неизвестно» соответствием
Отчёт	Доля, интервал, ограничение	Не переводить дефект метаданных в «пиратство»

Приложение 14. Проверка согласованности двух экспертов

Для юридически неоднозначных объектов два подготовленных оценщика получают один и тот же минимальный пакет сведений, но не видят заключения друг друга. Инструкция заранее различает действующую лицензию, действующее исключение, неизвестный статус, установленное несоответствие и внерамочный объект. Затем сопоставляются первичные метки и мотивировки. Противоречие рассматривается третьим экспертом, а причины исправления заносятся в реестр обучения методики.

Следует не допускать формального повышения согласованности путём объединения всех сложных случаев в неопределённую категорию. Для каждой версии рубрики сохраняются примеры типичных и

пограничных случаев, а новое судебное решение приводит к датированному пересмотру. Если одну половину дел оценивали по старой, а другую по новой инструкции, итоговый показатель нельзя интерпретировать как единый тест без методической оговорки.

Показатель	Документ	Управленческое решение
Исходные метки	Отдельные формы экспертов	Проверить слепой режим
Матрица расхождений	Пересечение категорий	Уточнить неоднозначные правила
Согласованный вывод	Мотивированное решение	Зафиксировать dissent
Новая редакция рубрики	Номер версии и дата	Не перезаписывать прошлую разметку

Приложение 15. Процедура временной остановки использования набора

Временная остановка применяется не как наказание до суда, а как мера контроля вероятного юридического и репутационного риска. Триггерами могут быть подтверждённый пиратский источник, систематически проигнорированные запреты, отсутствие прав на значительную категорию или техническая невозможность локализовать распространение спорного материала. Решение фиксирует конкретные версии, операции и срочные меры сохранения доказательств.

Остановка должна быть адресной: блокировка конкретного источника, training batch, fine-tuning процесса или выдачи определённого класса результатов в зависимости от риска. Разработчик вправе представить документы, снимающие сомнение; аудитор проверяет их без заранее обещанного исхода. Возобновление возможно после документированного решения, устранения причины и тестирования. Без этого простое переименование корпуса не является ремедиацией.

Триггер	Немедленная мера	Условие возобновления
Пиратский архив	Блокировать источник и производные наборы	Новая правовая основа либо исключение
Нарушенный opt-out	Остановить будущий сбор	Исправленный парсер и подтверждённый тест
Разрыв provenance	Карантин неидентифицируемой части	Восстановление цепочки доказательств
Массовый output-риск	Ограничение функции/версии	Независимый повторный тест

Приложение 16. Проверка корректирующих мер и остаточного риска

После решения об устранении дефекта проверяется не только факт удаления записи из табличного реестра, но и влияние на копии, индексы, будущие training batches и обслуживаемые модели. Для каждой меры указывается техническая область действия, оператор и свидетельство результата. Если нельзя подтвердить устранение из параметров ранее обученной модели, формулируется остаточная неопределённость и дополнительные защиты на этапе выдачи.

Полезен негативный тест: искусственно включённый контрольный объект не должен попасть в будущую сборку после выставления запрета. Он демонстрирует работу конвейера, но не доказывает, что все прошлые спорные элементы устранены. Итоговое заключение отмечает, что именно было подтверждено (например, блокировка конкретного training bucket) и что остаётся без доступа (например, независимые модели сторонних разработчиков).

Мера	Подтверждение	Что не подтверждается автоматически
Удалён исходный файл	Hash и storage log	Отсутствие производных копий
Исключён training batch	Manifest новой версии	Изменение старых весов
Добавлен output-фильтр	Регрессионные тесты	Невозможность иных обходов
Направлено уведомление партнёру	Подтверждение получения	Реальное удаление у всех партнёров

Приложение 17. Реестр независимости и конфликтов интересов

Независимый отчёт должен раскрывать, кто оплачивает работу, существуют ли консультационные отношения сторон и кто может повлиять на публикацию. Форма включает личные и институциональные интересы проверяющих, прежнюю работу на проверяемого поставщика, связи с правообладателями и участие в переговорах по спорным лицензиям. Не всякая связь делает эксперта автоматически непригодным, но значимое обстоятельство требует защитной меры, замены проверяющего или мотивированного ограничения заключения.

Договор на проверку должен отделять вознаграждение за работу от позитивного исхода. Руководство не вправе изымать неудобные выводы без документированного редакционного решения. При невозможности обеспечить независимость проект может быть оформлен как клиентская консультация, но не как независимая экспертиза. Необходимо регулярно обновлять реестр, особенно при появлении нового параллельного поручения.

Конфликт	Степень риска	Реакция
Вознаграждение за позитивный вывод	Неприемлемая	Изменение договора или отказ
Недавняя работа у проверяемого	Повышенная	Disclosure и независимая рецензия
Личная связь с правообладателем	Контекстная	Recusal по объектам
Отсутствие доступа к доказательствам	Методическая	Ограничить название и вывод отчёта

Приложение 18. Совмещённая карта интеллектуальных прав и GDPR

Карта нужна для тех корпусов, где произведение содержит персональные данные. Литературное право автора и конфиденциальность личности на фотографии являются отдельными основаниями. Даже юридически разрешённое копирование текста для TDM не отвечает на вопрос о допустимости обработки в нём сведений о здоровье, политических взглядах или уголовном преследовании. Для каждого набора следует назначить отдельные поля IP_basis и DP_basis, а также результат проверки специальных ограничений. [16][20]

Публичная публикация сведений о человеке не означает автоматически право на их неограниченное использование в генеративной системе. Проверяется назначение, минимизация, требуемая прозрачность, права субъектов и международные передачи. Для закрытых клиентских материалов рекомендуется исключение из общего обучения, если нет отдельного надёжного правового решения. Аудитор фиксирует оба результата раздельно, чтобы отсутствие copyright-риска не скрывало нарушение режима данных.

Объект	IP-проверка	Data protection проверка
Авторская фотография	Автор/лицензия/TDM	Изображённые лица и цель
Судебная статья	Право издателя и база данных	Данные о подозреваемых
Биографический архив	Права на текст и изображения	Пропорциональность обработки
Письмо клиента	Права на документ	Конфиденциальность и законность передачи

Приложение 19. Условный кейс: фотографический каталог

Фотограф предоставляет редакции лицензию на размещение снимков на сайте. Агрегатор скачивает изображения и продаёт технической фирме «открытый» набор. Наличие лицензии редакции не подтверждает право агрегатора обучать коммерческую модель и сублицензировать исходные файлы. Аудит обязан реконструировать условия исходного договора, проверить статус opt-out на момент сбора и отделить права фотографа от возможных персональных прав изображённых людей.

Технический тест включает идентификацию дублированных файлов и изменённых форматов, проверку метаданных авторства и сопоставление с downstream реестром. Рекомендуемое решение до устранения пробела — карантин спорной категории, обращение за подтверждением цепочки прав и документирование того, какие версии модели уже получили контент. Кейс учебный и не касается реального каталога, издателя или модели.

Факт кейса	Правовая проблема	Доказательство
Редакционная лицензия	Разрешено ли обучение?	Полный текст разрешения
Перепродажа агрегатором	Есть ли сублицензия?	Контракт upstream
Метаданные удалены	Установим ли автора?	Снимок оригинала и идентичность изображения
Несколько версий снимка	Применим ли opt-out к копиям?	Перцептивный отпечаток и provenance

Приложение 20. Условный кейс: тексты песен и генеративная модель

Учебная ситуация: модель регулярно выдаёт длинные узнаваемые фрагменты песенного текста по обычным вопросам. По аналогии с подходом, обсуждавшимся в деле GEMA, расследование проверяет воспроизводимость, качество литературной формы, источники обучения и фактический контроль поставщика. Но решение мюнхенского суда первой инстанции не должно автоматически трактоваться как итог любого нового дела; национальное право, условия лицензии и технические детали могут различаться. [11]

Процедура предусматривает выбор независимых тестировщиков, отсутствие длинного оригинала в prompt, документирование версии модели и точное сравнение защищённых фрагментов. Результат разделяет потенциальное воспроизведение в модели и публичный output, потому что правовая квалификация этих операций отличается. Публичный отчёт использует описания результатов вместо многократного воспроизведения самих текстов песен.

Шаг	Данные	Критерий
Правообладатель	Права на текст и полномочия	Объём притязания
Обучающий источник	Лицензия или исключение	Применимость к конкретному действию
Output	Простые prompts и логи	Охраняемая узнаваемая форма
Ремедиация	Защита выдачи и спорного корпуса	Повторный тест

Приложение 21. Условный кейс: учебные материалы и перевод

Издательство выпускает учебник и авторизованный перевод, а университет предоставляет студенческий цифровой доступ. Стартап копирует оба языка, считая подписку университета достаточным разрешением на обучение своей коммерческой модели. Аудит должен разделить права на оригинал, перевод и цифровое издание, проверить условия университетского доступа и возможные применимые исключения TDM. Наличие научного партнёра не превращает все действия стартапа в исследовательский TDM по статье 3 DSM. [1]

Отдельно проверяется, формирует ли модель ответы с длинными фрагментами учебника, а не только использует усвоенные идеи. Подготовка доказательств должна исключать распространение полной нелицензированной электронной версии. Рекомендуемая мера — ограничение доступа к спорным файлам до документальной проверки, переговоры по правам, отдельное решение о последующих моделях и обновление публичной сводки при необходимости.

Узел	Уточнение	Документ
Исходник	Права автора и издателя	Издательский договор
Перевод	Самостоятельные права	Договор переводчика
Университет	Только доступ или обучение тоже?	Подписные условия
Стартап	Статус научной/коммерческой роли	Протокол фактической деятельности

Приложение 22. Условный кейс: программный код и открытая лицензия

Компания использует публичный репозиторий с кодом разных авторов и лицензиями разных типов. Разработчики называют весь набор open source и поэтому игнорируют требования о сохранении copyright

notices и специфические условия использования. На практике нужно проследить лицензии репозитория, зависимостей, отдельных файлов и изменения условий между версиями. Не каждое требование лицензии применяется одинаково к обучению и к распространению сгенерированного кода, и этот вопрос требует отдельной юридической оценки.

Протокол предусматривает выявление точных и близких совпадений выходных фрагментов, но автоматический детектор лицензии не даёт окончательного вывода без просмотра фактов. Следует регистрировать случаи, где модель выдаёт длинные узнаваемые блоки кода с удалёнными уведомлениями об авторстве. При отсутствии прямого сходства вывод не должен описывать лицензию как не имеющую значения для upstream копирования.

Проверка	Вход	Выход
Лицензии	Файлы и версии репозитория	Перечень различных условий
Provenance	Источник и hash	Связь с набором обучения
Output	Длинные сходные участки	Технический протокол совпадений
Документация	Notice, attribution, distribution	Юридическая оценка обязанностей

Приложение 23. Условный кейс: новости и издательский архив

Агрегатор собирает статьи, фотографии и редакционные аннотации с новостных порталов, затем продаёт набор по общим условиям «для аналитики». Правовая оценка не может сводиться к наличию общественно значимых фактов. В наборе могут быть права журналистов, фотографов, издателей прессы, изготовителя коммерческой базы и условия доступа к платному архиву. Требуется разделить извлечение фактов, воспроизведение текста и дальнейшую выдачу пользователям. [6][7]

Аудитор проверяет объём скопированных материалов, регулярность обновлений, наличие opt-out, права агрегатора и конфигурацию выдачи. Особое внимание уделяется замещению доступа к первоисточнику: сервис, возвращающий пользователю полную платную статью, юридически и экономически отличается от системы поиска публичных фактов. В выводе отдельно отмечаются рыночные последствия как предмет будущих исследований, а не как автоматически рассчитанный ущерб каждого издателя.

Объект	Возможное право	Проверка
Текст новости	Автор и издатель	Лицензия и TDM
Фотография	Фотограф и иные лица	Цепочка прав
Архивная база	Sui generis/условия	Масштаб извлечения
Полный ответ модели	Воспроизведение/доступ	Поведение конкретной версии

Приложение 24. Условный кейс: музейная коллекция и научные исследования

Музей предоставляет исследовательской организации цифровые материалы для изучения техники каталогизации. Параллельно коммерческий участник проекта предлагает использовать весь набор в универсальной модели изображений. Анализ должен проверить, какие объекты охраняются, кем распоряжаются соответствующие права, имеются ли особые договорные условия и отделено ли научное исследование от коммерческого использования. Режим статьи 3 DSM не переносится автоматически на каждого партнёра. [1]

Рекомендуется иметь два независимых рабочих пространства и отдельные data manifests. Научная среда хранит копии в установленном защищённом режиме для верификации результатов; коммерческий путь требует собственного юридического основания и решения о допуске. Аудитор проверяет, что доступ подрядчика к исследовательским копиям не фактически равен их свободной выгрузке в продуктовую инфраструктуру.

Параметр	Научная среда	Коммерческая среда
Цель	Проверяемый исследовательский вопрос	Разработка продукта

Параметр	Научная среда	Коммерческая среда
Основание	Проверка условий статьи 3	Лицензия/статья 4 и иные нормы
Доступ	Ограниченный и защищённый	Отдельное юридическое решение
Выход	Публикация результатов	Проверка прав на выпуск модели

Приложение 25. Условный кейс: материалы суда и коммерческая база

Независимая исследовательская команда собирает опубликованные судебные решения и хочет обучить систему правового поиска. Частный агрегатор предлагает добавить платные аннотации и классификаторы без отдельной лицензии, мотивируя это тем, что судебные акты общедоступны. Проблема заключается в различии официальных текстов, охраняемых редакционных материалов, прав на базу данных и конфиденциальности лиц, упомянутых в делах. [7][20]

Для пилота необходимо использовать только материалы, на которые имеются документированные права или надёжное основание, и исключить скрытые клиентские досье. Для уголовно-процессуальных материалов проводится отдельная GDPR-оценка; публикация суда не всегда означает свободу вторичного обучения. Публичный отчёт должен использовать учебные гипотетические данные и проверенную библиографию, а не воспроизводить идентифицирующие детали реального дела в иллюстрации алгоритма.

Слой	Правовая проблема	Мера
Официальный акт	Национальный режим повторного использования	Отдельная квалификация
Аннотация издателя	Авторское право	Лицензионная проверка
Коммерческий поиск	Права на базу и договор	Условия API
Персональные сведения	GDPR и специальные режимы	Минимизация и доступ

Приложение 26. Поставщик API как источник данных для дообучения

Интегратор получает ответы внешней модели по API, собирает их в собственную базу и использует в дальнейшем для обучения конкурирующего продукта. Необходимо анализировать условия API, права на выходные материалы, возможные ограничения на обучение на ответах, происхождение защищённых фрагментов и правила обработки персональных данных. Тезис «выход модели свободен, поскольку написан ИИ» юридически неполон: в ответе могут содержаться охраняемые человеческие элементы или данные пользователя.

Аудитор должен установить, какой поставщик фактически формировал ответ, какой prompt использовался, кому принадлежат права на включённые в него исходные документы и какие условия принял интегратор. Если модель случайно воспроизвела охраняемый пассаж, дальнейшее обучение на таком ответе не очищает его юридическую историю. Рекомендуется запретить передачу спорных выходных текстов в новую обучающую среду до оценки.

Звено	Документ	Контроль
API terms	Версия условий на дату обращения	Разрешено ли повторное обучение
Prompt	Источник пользовательских вложений	Конфиденциальность и IP
Output	Hash и журнал	Вероятное воспроизведение
Следующая модель	Manifest дообучения	Не допустить наследование дефекта

Приложение 27. Модель переговоров о лицензии на тематический каталог

Для тематического каталога предлагается переговорная матрица, в которой правообладатель предоставляет не абстрактные «все права», а конкретные разрешения с проверяемыми параметрами: виды работ, обучающие цели, временной горизонт, территория, права на воспроизведение и технические копии,

условия атрибуции, аудит использованного объёма и порядок обработки претензий. Вознаграждение обсуждается отдельно от организационных затрат на администрирование.

Стороны могут определить смешанную модель: фиксированная стоимость доступа и переменный компонент по согласованной методике. Однако формулу нельзя утверждать без предварительной оценки качества данных и договорной структуры прав. Особый вопрос — судьба уже обученной модели после расторжения: договор должен предусматривать реалистичные технические меры и не обещать автоматическое стирание любых следов в весах. Для малых авторов необходимо предусмотреть доступный способ подтверждения включённых произведений.

Предмет переговоров	Возможный вариант	Риск неопределённости
Репертуар	Список asset IDs и версии	Невозможно подтвердить охват
Разрешённые этапы	Training / fine-tuning / evaluation	Разная трактовка «использования»
Вознаграждение	Фиксированное / по объёму / комбинированное	Непроверяемая база расчёта
Расторжение	Запрет будущего сбора и процедура ремедиации	Неисполнимое обещание удалять веса

Приложение 28. Проверка полномочий коллективной организации

Организация коллективного управления может обладать значимым репертуаром, но её полномочия нужно установить применительно к конкретным способам использования и территории. Договор на публичное исполнение не всегда покрывает запись, цифровую репродукцию или обучение модели. Для нескольких правообладателей важны распределение выплат, правила представительства и возможность противоречащих заявлений. На основе одной печати или письма нельзя считать решённым весь массив прав.

Аудитор запрашивает сведения о правовом статусе представителя, репертуаре, переданных полномочиях, географии и соответствующей категории использования. Соглашение о неразглашении может защищать персональные детали каталога, но итоговое заключение должно ясно указать границу проверенных прав. Наличие коллективного механизма полезно для уменьшения издержек, однако нельзя выдавать его за единственный обязательный маршрут получения разрешения.

Документ	Проверяется	Потенциальный пробел
Устав/мандат	Полномочия организации	Категория прав не охвачена
Репертуар	Объекты и исключения	Авторы вышли из системы
Территория	Страны и трансграничные соглашения	Не охвачено место действия
Распределение	Принципы и отчётность	Непроверяемая компенсация

Приложение 29. Трансграничная матрица юридических действий

Для каждого этапа необходимо учитывать техническое место действия, место коммерческого использования, права заявляемых правообладателей и вероятную юрисдикцию спора. Не рекомендуется выводить законность всей операции из страны регистрации одного подрядчика. Даже если копия была создана за пределами ЕС на основании местного исключения, дальнейшее предоставление сервиса в Союзе может вызвать отдельные обязательства AI Act и авторско-правовые вопросы. Точный конфликт законов решается на основании конкретной сделки и фактов.

Матрица помогает избежать несовместимых упрощений: американский fair use не является европейской статьёй 4 DSM, британская консультация не создала автоматического opt-out в 2026 году, а сингапурская норма computational data analysis имеет собственные условия. Для официального заключения следует привлекать профильного юриста соответствующей юрисдикции и фиксировать дату актуальности нормативной карты.

Этап	Юрисдикционная привязка	Что оценивается
Сбор в интернете	Фактическое место и право источника	Доступ, лицензии, TDM

Этап	Юрисдикционная привязка	Что оценивается
Обучение	Место выполнения и применимое право	Копирование и исключения
Релиз GPAI в ЕС	Роль поставщика и рынок Союза	Статья 53 AI Act
Спор о выводе	Суд, территория и характер прав	Воспроизведение/ответственность

Приложение 30. Аудит заявлений о «100% законном» обучении

Исследователь получает рекламное заявление поставщика, что все тренировочные данные полностью легальны. Его нельзя подтверждать без проверки определения слова «законно», юридической географии, объёма и периода. Рекомендуется составить перечень подутверждений: источники имеют подтверждённое основание доступа; исключения применялись корректно; машиночитаемые возражения обрабатывались; лицензии покрывали нужные операции; политика и сводка соответствуют AI Act при его применимости.

Для каждого подутверждения фиксируется наблюдаемое доказательство, недоступная часть и остаточный риск. Отчёт может подтвердить определённую процедуру, например «для проверенной выборки из источника X запреты crawler обрабатывались корректно в период Y», но не должен превращать это в универсальное заверение о миллиардах неизвестных объектов. Формула вывода должна быть проверяемой и доступной для внешнего критического чтения.

Claim	Проверка	Корректная формулировка
Полная законность	Доступ к полному источниковому корпусу?	Ограничить объём и критерии
Отсутствие пиратства	Независимый источник-контроль	«Не выявлено в проверенной выборке»
Уважение opt-out	Протокол негативных испытаний	Процент прохождения и пробелы
Независимое подтверждение	Финансовая и редакционная автономия	Раскрыть существенные связи

Приложение 31. Матрица оценки стоимости соблюдения требований

Экономическая оценка должна включать как расходы поставщика модели на документирование, лицензирование и аудит, так и затраты авторов на резервирование прав, проверку включения произведений и предъявление претензий. Сравнение, в котором затраты одной стороны учтены, а другой игнорируются, методически неустойчиво. Для малых участников важно выделить фиксированные расходы: проверка стандартов, консультации, регистрация каталогов, пересмотр договоров и обучение персонала. [5][9]

Вместо недоказанного прогноза «регулирование стоит десятки миллиардов» предлагается пилот на ограниченной популяции с документированными ценами операций, затем сценарный анализ. Вывод должен указывать чувствительность к охвату данных, масштабу модели, автоматизации сигналов и частоте споров. Отдельно учитывается рост доверия и снижение расходов на будущие разбирательства как потенциальная выгода, но не как гарантированная экономия.

Статья расходов	Разработчик	Правообладатель
Реестр активов	Проведение инвентаризации	Сопоставление репертуара
Opt-out	Проверка сигналов	Выражение и контроль запрета
Разрешения	Переговоры/выплаты	Представительство/получение выплат
Споры	Доказательства и экспертиза	Юридическое обращение и мониторинг

Приложение 32. Политика исправления публичных ошибок Observatoire

Если опубликованный аналитический вывод содержит фактическую неточность, организация должна обеспечить публичное и трассируемое исправление без тихой замены исторической версии. Исправление

указывает, какой фрагмент изменён, почему, когда и на какие источники опирается обновлённая редакция. Для спорных материалов о конкретной компании следует предоставлять возможность представить замечания по фактам, не наделяя ответчика правом предварительной цензуры независимого исследования.

Технически сохраняются исходная редакция, новый документ, changelog, дата публикации и ссылки на исправление. Существенная ошибка о правовом статусе лицензии или о ходе судебного разбирательства требует явной заметки в начале текста. Простая правка опечатки может отражаться в общем журнале. Защита коммерческой тайны и персональных данных оценивается отдельно от принципа прозрачности. Такая политика необходима для доверия к независимым исследованиям.

Событие	Реакция	Публичное сообщение
Фактическая ошибка	Верификация и документальное исправление	Дата и описание
Новый судебный акт	Дополнение без ретуши прежнего статуса	Обновлённый контекст
Оспаривание стороной	Независимое рассмотрение	Итог и мотивировка
Опасное раскрытие	Срочное ограничение доступа	Разумно достаточное объяснение

Приложение 33. Проект ответов ARGA для консультации Еврокомиссии

Рекомендуемая позиция по Technology and Copyright: повышение прозрачности обучения должно обеспечивать возможность правообладателя проверить законность использования произведения, но не создавать открытый реестр коммерчески конфиденциальных контрактов и технических уязвимостей. Предлагается двухуровневая архитектура: публичная сводка существенных категорий и частных источников по установленному шаблону и закрытый проверяемый реестр asset-level для законных регуляторных и судебных процедур. Следует стимулировать совместимые форматы provenance и сохранение истории opt-out. [5]

В отношении права резервирования предлагается минимальный машиночитаемый стандарт, доступный небольшим авторам без платы, с проверяемыми ответами разработчиков. В отношении лицензирования — прозрачные условия распространения прав по цепочке поставщиков и доступные договорные конструкции для коллективного или индивидуального вознаграждения. В отношении аудита — независимость проверяющего, публикация границ выборки и защитный порядок доступа к закрытым данным. Это заготовка содержательной позиции, а не подтверждение подачи от имени ARGA; текст необходимо адаптировать к конкретным вопросам официальной формы и согласовать с руководством.

Тема консультации	Предлагаемый ответ	Доказательная потребность
Training transparency	Многоуровневое раскрытие	Провайдерские manifests и проверки
Rights reservation	Бесплатные машиночитаемые сигналы	Тесты соблюдения на разных платформах
Licensing	Прослеживаемые права по цепочке	Договорная выборка и компенсация
Enforcement	Доступный канал споров	Исследование затрат малых авторов

Приложение 34. Протокол защищённой комнаты данных для аудитора

Чтобы проверять закрытый корпус без публикации коммерческих каталогов, можно организовать защищённую среду с доступом к выборке реестра, соответствующим лицензиям и журналам технической обработки. Доступ должен быть роль-ориентированным, журналируемым, ограниченным по сроку и с контролем экспортируемых результатов. Наличие clean room не является само по себе юридическим основанием обработки или передачи персональных данных: до открытия доступа проверяются договоры, законность, необходимость и категории сведений.

Независимому аудитору требуется право проверять полноту популяции, иначе поставщик может показать только удобные объекты. Полезны случайные запросы по утверждённой процедуре, внешняя сверка выборки и доказательство неизменности manifest. Все ограничения, препятствующие проверке полноты,

должны раскрыться в итоговом отчёте. Не следует выдавать закрытую среду с заранее выбранными поставщиком десятью примерами за аудит многомиллиардного корпуса.

Контроль	Реализация	Проверка
Полнота реестра	Версионированный manifest	Сопоставление с pipeline
Ролевой доступ	Разделение функций	Review логов
Экспорт	Агрегаты, разрешённые доказательства	Контроль IP/GDPR
Независимость	Случайный отбор аудитора	Не ограничивать удобной выборкой

Приложение 35. Девяностодневный учебный пилот ARGA

Пилот не предполагает доступа к реальным закрытым системам без отдельного договора. В первые тридцать дней разрабатываются юридические критерии, образец лицензии, synthetic reference corpus и методика обработки opt-out. Следующие тридцать дней отводятся на настройку журнала источников, графа преобразований, документацию кейсов и негативные тесты. Последние тридцать дней — на независимую выборочную проверку, отчёт о пробелах и корректировку методики.

План должен иметь stop gates: отсутствие правообладательского основания для учебного материала, разрыв цепочки доступа или конфликт интересов, исключающий независимый вывод. Пилот измеряет выполнимость самой методики, а не долю незаконно обученных моделей на рынке. Его успешный результат — воспроизводимые протоколы и честные ограничения; количество таблиц и страниц не служит метрикой успеха.

Период	Работа	Критерий завершения
Дни 1–30	Право, формы, учебные данные	Утверждённая методика
Дни 31–60	Provenance pipeline, opt-out tests	Повторяемые технические записи
Дни 61–90	Слепая проверка и рецензия	Отчёт с ограничениями
Stop gate	IP, GDPR, независимость	Приостановка до устранения

Приложение 36. Шаблон итогового аудиторского заключения

Заключение начинается с идентификации проверяемой модели, организации, версии данных, периода, юрисдикций, заказчика и роли эксперта. Вслед за этим приводятся критерии нормативной оценки, степень доступа, методы выборки, источники доказательств, выявленные несоответствия, корректирующие меры и остаточная неопределённость. Термины «проверено», «не выявлено» и «подтверждено соответствие» не взаимозаменяемы. Последний применим только к определённому набору критериев и достаточным доказательствам.

В заключении нельзя умалчивать о непросмотренной части корпуса или предоставленных только самим заказчиком заверениях. Для каждого существенного утверждения должен существовать audit trail. Коммерческая тайна позволяет обобщить некоторые сведения в публичной версии, но не убрать сам факт ограничений. Отчёт обязательно заканчивается датой, полномочиями подписанта, процедурой исправлений и статусом рецензии.

Раздел заключения	Требование	Недопустимая замена
Scope	Точные модели и источники	«Вся компания» без проверки
Criteria	Нормы и стандарты с версиями	Общая декларация этики
Evidence	Выборка, хеши, документы	Рекламные заверения
Findings	Факт, право, неопределённость	Автоматическое обвинение
Limitations	Непроверенные части и причины	Умолчание

Приложение 37. Процедура запроса уточнений у разработчика модели

Запрос должен быть узким и юридически мотивированным. Вместо требования выдать весь обучающий корпус без основания задаются вопросы об идентичности модели, происхождении конкретной категории источников, применимой лицензии или исключении, обработке резервирования и доступности ограниченного аудита. Включение коммерческих секретов требует отдельного режима защиты и оценки необходимости. Контактный адрес правообладателя и данные представителя обрабатываются по цели запроса.

Проект письма фиксирует нейтральный статус исследовательской проверки и прямо исключает предположение о доказанном нарушении до изучения ответа. При отсутствии ответа используется формулировка «информация не предоставлена по нашему запросу в пределах указанного периода», а не «компания признала нарушение». Для повторного запроса необходима новая материальная причина или запрошенный срок, а не автоматическая бесконечная переписка.

Блок запроса	Предмет	Ожидаемый ответ
Идентификатор	Модель и версия	Подтверждение
Source type	Категория/поставщик	Provenance summary
Правовое основание	Лицензия/DSM	Мотивированное подтверждение
Opt-out	Реализация и тесты	Процедуры или ограничения

Приложение 38. Особый режим сведений о санкциях и международном розыске

Учебный сценарий касается модели, обучаемой на журналистских расследованиях, записях международных баз и судебных документах о предполагаемых преступлениях. Нельзя смешивать законность цитирования публичного факта с правом массового копирования редакционного каталога и возможностью дальнейшей обработки сведений о конкретном лице. Проверка авторских прав идёт параллельно GDPR, возможной конфиденциальности и риску диффамации либо ошибочной атрибуции. [7][20]

Рекомендуемая мера для ARGA — не передавать клиентские досье в общий тренировочный сервис, хранить отдельную доказательственную историю происхождения каждой публикации и документировать исправления при изменении процессуального статуса лица. Если модель генерирует ложный «статус розыска», это вопрос фактической достоверности и защиты личности; сам по себе он не доказывает авторско-правовое нарушение. Независимое заключение удерживает эти категории раздельно.

Риск	Отдельный правовой маршрут	Контроль
Копирование коммерческой базы	Авторское право/права на базу	Лицензия и извлечение
Данные уголовного преследования	GDPR, статья 10 и национальное право	Доступ, цель и точность
Ошибочный статус лица	Исправление и защита прав личности	Traceable correction
Клиентская тайна	Конфиденциальность	Не использовать в общем обучении

Приложение 39. Контрольная карта рисков до публикации отчёта

Перед передачей доклада внешней аудитории редактор проверяет статус каждого правового акта, актуальность судебных решений, связку библиографии с текстом и отсутствие информации, способной раскрыть закрытый источник или чужое произведение сверх допустимого объёма. Особое внимание — формулировкам судебной практики: решение суда первой инстанции не представляется окончательной общеевропейской нормой; консультация не называется принятым законом; добровольный кодекс не получает статус обязательной сертификации.

Техническая проверка включает документальный формат, активность гиперссылок, отсутствие шрифтовых квадратов, корректную нумерацию, читаемость таблиц, наличие логотипа и запрет на скрытые

персональные метаданные в свойствах файла. Юридическая рецензия и технический preflight являются разными процедурами; успешное открытие PDF не подтверждает точности нормативного анализа.

Контроль	Ответственный	Условие допуска
Нормы и даты	Юридический эксперт	Подтверждены первоисточниками
Фактические утверждения	Аналитик	Проверяемая ссылка и оговорка
Цитаты и IP	Редактор	Не нарушены чужие права
Word/PDF	Технический редактор	Открываются, без квадратов
Утверждение ARGA	Руководство	Отдельное согласование

Приложение 40. Матрица мониторинга изменений после октября 2026 года

Право обучения моделей находится в стадии развития. Поэтому итоговая публикация должна сопровождаться перечнем событий, требующих пересмотра: принятие новых обязательных норм, финализация консультаций Комиссии, изменения официального шаблона GPAI, значимые судебные решения и новые стандарты машиночитаемых сигналов. Мониторинг не означает автоматическую смену выводов: каждый триггер оценивается на юридическую релевантность для конкретного раздела.

Обновления должны сохранять версию исходного исследования и содержать датированный changelog. Если поступает решение по конкретному спору, редактор проверяет инстанцию, возможность обжалования, предмет и правовую систему. Если публикуется проект руководства, его статус указывается отдельно. Такой дисциплинированный процесс предотвращает распространённую ошибку, когда новостной заголовок используется как будто окончательно изменившаяся норма права.

Будущее событие	Проверяемый факт	Решение редактора
Закрытие консультации	Есть ли итоговый документ?	Обновить статус, не выдумывать решение
Судебный акт	Инстанция, предмет, обжалование	Пересмотреть конкретный тезис
Новая версия AI Act	Дата применения и охват	Исправить календарь
Технический стандарт opt-out	Статус и внедрение	Обновить методические тесты

Приложение 41. Протокол сопоставления права на фотографию с версиями файла

Цель протокола — установить происхождение конкретного изображения без ложного вывода о том, что похожий файл обязательно принадлежит найденному первым пользователю. Исследователь регистрирует уникальный ID дела, размер, формат, контрольную сумму оригинальной копии, найденный URL, дату и способ получения, возможные EXIF-поля, номер версии, историю кропов и сведения об авторской подписи. Полученные от поставщика данные отделяются от самостоятельно подтверждённых. При попытке сопоставить исходный RAW, экспорт JPEG и рекламный ресайз записываются параметры сходства и альтернативные объяснения. Перцептивное сходство — повод для документального запроса, но не замена установления авторства.

Если заявлены права фотографа, проверяются заказ на съёмку, исключительная или неисключительная лицензия, полномочия агентства, пределы передачи третьим лицам, способы использования при обучении и действия по дообучению. Поле «источник неизвестен» нельзя заменять предположением о свободной лицензии. Если авторская атрибуция спорна, аудитор запрашивает письменные подтверждения и временно отмечает объект «требуется проверка», не объявляя одну из сторон нарушителем. Отдельная отметка относится к портретам с персональными данными: наличие авторско-правового разрешения не снимает вопросов по GDPR.

Итог протокола содержит вывод трёх типов: «цепочка прав подтверждена в исследованном объёме», «цепочка не подтверждена из-за конкретного отсутствующего документа» или «выявлены противоречащие друг другу доказательства». Третье состояние не должно принудительно сводиться к

первому или второму. Фотография не приводится полностью в открытом отчёте без самостоятельного основания её воспроизведения. [6][20]

Приложение 42. Тест машиночитаемого резервирования прав при смене URL

Тестовая среда имитирует цепочку «первоначальная страница — 301/302-перенаправление — CDN — архивная копия — файл изображения». Для каждой стадии предварительно фиксируются URL, правила доступа, ответы сервера, настройки робота, момент извлечения и сохранённый сигнал о резервировании прав. Если правообладатель обновил заявление, проверяется, получил ли его сборщик при следующем обходе и сохраняет ли версию решения. Нельзя считать отсутствующий сигнал разрешением без подтверждения других необходимых условий правомерного доступа.

Испытание включает три заранее заданные проверки. Первая: запрет присутствует на основном URL и не исчезает при редиректе. Вторая: при обновлении правила в середине сеанса система повторно оценивает пригодность источника к включению. Третья: система не принимает зеркало за независимый новый правообладательский ресурс только потому, что изменилась доменная строка. Для каждого теста указываются ожидаемый и фактический результаты; сбои классифицируются по происхождению (краулер, обработчик прав, поставщик данных, ручное исключение).

Юридический комментарий должен напоминать, что техническая спецификация сигналов и правовой смысл их использования — разные вопросы. Машиночитаемость актуальна для онлайн-контента по статье 4(3) DSM, но конфликт конкретных сигналов, сторон и юрисдикций подлежит отдельной оценке. Тестовый набор должен содержать лишь данные, на которые исследователь имеет право доступа и воспроизведения. [1][21]

Приложение 43. Рабочий пример проверки смешанного архивного каталога

Условная организация приобретает доступ к каталогу из 20 000 элементов: 12 000 газетных статей, 4 000 фотографий, 2 000 аудиозаписей и 2 000 редакционных справок. Числа являются учебными, не описывают реальную компанию. Архивный договор разрешает чтение и внутренний поиск сотрудниками, но молчит об извлечении для коммерческого обучения. Команда ошибочно присваивает всему каталогу статус «licensed». Аудитор разбивает массив на четырёхтипную стратифицированную выборку с выделением редких категорий и фиксирует неопределённость разрешений отдельно для каждого типа. Проблемы фотографий и аудиофайлов не маскируются высоким качеством документов по статьям.

Проверка начинается с 40 случайных объектов каждого типа, всего 160, и не предполагает автоматически распространение результатов на всю генеральную совокупность без весов, правил отбора и учёта кластеризации. Для каждого объекта заполняются дата доступа, существенные права, лицензия, резервирование и подтверждение полномочий. В условной группе из 40 аудиозаписей 9 имеют неполную цепочку смежных прав. Это показатель 22,5% на исследованной подвыборке, а не 22,5% законно доказанных нарушений и тем более не оценка для всех песен Интернета.

Рекомендация — до получения разъяснений приостановить включение несогласованных типов данных в новый цикл обучения, сохранить документы аудита и запросить у поставщика договорное подтверждение. Если выяснится, что речь идёт о допустимом исследовательском TDM при соблюдении установленных условий, вывод пересматривается. Отчёт должен пояснять, что нехватка доказательства лицензии сама по себе не исключает существование правового исключения.

Приложение 44. Сопоставление версий GPAI и публичной сводки обучающего контента

Для модели условного разработчика составляется таблица версий: дата размещения на рынке ЕС, исходная базовая модель, дата существенного дополнительного обучения, категории вновь использованных данных, новая опубликованная summary и ответственное лицо. Проверка проводится на уровне идентифицируемого продукта, а не абстрактного бренда. Если несколько версий имеют одинаковое содержание сводок, допустимость такого раскрытия оценивается по официальным

пояснениям Комиссии; если дообучение добавляет новые категории источников, требуется отдельно оценить необходимость обновления. [2][3][17]

Пример: базовая модель M1 обучалась на трёх категориях; версия M1.1 получила дополнительный юридический корпус от коммерческого поставщика. Аудитор не требует безусловного раскрытия всех клиентских договоров в публичном документе, но должен убедиться, что новый класс источников отражён на надлежащем уровне и что имеются защищённые доказательства происхождения. В журнале фиксируется срок следующей проверки и причины досрочного обновления. Официальное FAQ Комиссии описывает периодическое обновление с шестимесячным ориентиром или раньше при материально значимых новых данных; этот ориентир нельзя превращать в самостоятельное положение об автоматическом истечении всех авторско-правовых разрешений.

Контрольный результат оформляется как «сводка опубликована и оценена по шаблону», «сводка требует уточнения» либо «проверка невозможна по недостатку данных». Публичная summary не подтверждает правомерность каждого элемента корпуса, а непрерывность версий модели должна быть воспроизводима по документации.

Приложение 45. Тест авторских текстов без воспроизведения полного произведения

Аудитор формирует небольшую законно доступную контрольную выборку творческих текстов с заранее определёнными основаниями использования для теста. В протоколе хранятся название, версия, источник, тип лицензии, причины включения, признаки оригинальности и запрет на публикацию полного содержания в отчёте. До запросов к модели устанавливаются границы допустимых промптов, число итераций, настройки температуры и фильтров, временной интервал. Протокол разделяет обычные пользовательские вопросы, попытки восстановить редкий фрагмент и задачи с предоставленным моделью самим исследователем референсом.

Результат оценивается не только по символьному сходству. Факт совпадения общеупотребительной фразы юридически отличается от воспроизведения длинной индивидуализированной структуры, а сходство идей или жанровых клише не равно заимствованию охраняемой формы. Сопоставляются оригинальные элементы, длина фрагмента, возможность независимого совпадения и роль предоставленных входных данных. Эксперты сохраняют разногласия между технической и правовой оценкой вместо усреднения в один необъяснимый балл.

Для публикации применяется безопасная форма: статистика, обезличенный идентификатор, короткое описание признаков, метрики интервалов и метод воспроизводимости с минимальным раскрытием. Даже выявленное воспроизведение в тесте не доказывает без дополнительных материалов юридический статус исходного обучения. Отдельная правовая оценка возможна после установления конкретных действий и применимых исключений. [6][11]

Приложение 46. Договорная карта цифровой музыкальной библиотеки

Для условной библиотеки из 800 фонограмм аудитор создаёт четыре связанные таблицы: произведение, исполнители, мастер-запись, лицензии на машинное использование. Одно название композиции не обозначает всех управомоченных лиц. Каждой записи присваиваются внутренний ID, сведения об авторе текста и музыки, идентификатор фонограммы, исполнитель, география, срок и договорное основание. Наличие лицензии на публичное воспроизведение не заполняет автоматически поле «использование для дообучения». [6][19]

При обнаружении документа посредника проверяется не только его подлинность, но и мандат. Вопросы включают территориальную компетенцию, виды произведений, способ коммерческого использования, права передачи технологическому подрядчику и процедуру корректировки каталога. Отсутствие одной части цепочки прав фиксируется как конкретная проблема доказательств. Чтобы избежать ложных претензий, результат может содержать пометку «в доступном пакете нет документа, подтверждающего конкретное право», а не безусловное «право не существует».

В контрольной проверке выхода модели отдельно рассматриваются дословный текст песни, узнаваемые фрагменты фонограммы, общее стилистическое сходство и голосоподобие. Каждая категория использует самостоятельный тест и правовой анализ. Метод сохраняет дату, модель, конкретные запросы и условия; вывод не переносится на иные версии модели без соответствующих доказательств.

Приложение 47. Пример реконструкции лицензии по цепочке посредников

Учебная схема включает автора A, издателя B, агрегатора C, поставщика набора D и разработчика E. Каждое звено должно показать конкретные права и способы использования, а не только факт оплаты предыдущему участнику. В частности, право B размещать произведение на платформе не всегда охватывает право C извлекать его массово, а договор C–D о предоставлении данных не создаёт больший объём исключительных прав, чем имел C. Субподряд и сублицензирование проверяются отдельно. Даты перехода прав сопоставляются с фактической датой загрузки.

Для каждого звена указывают договорный пункт, территорию, срок, объект, право дальнейшей передачи и подтверждающий документ. Проверяющий обозначает степень доступности доказательств: оригинал, заверенная копия, извлечение, письменное заявление либо неподтверждённая запись в CRM. Отсутствующий договор не восполняется общей фразой о «партнёрском правообладательском каталоге». Если сведения содержат коммерческую тайну, доступ к полному документу можно ограничить защищённой комнатой данных с проверкой релевантных положений.

Итог содержит путь прав с отмеченными пробелами, а при необходимости альтернативное правовое основание, например применимость установленного исключения TDM. Нельзя автоматически утверждать, что всякое обучение возможно только по договору, так же как нельзя исходить из того, что исключение всегда применяется к любому публичному контенту. [1][6]

Приложение 48. Инструкция по защищённому ответу на запрос правообладателя

Входящая претензия может содержать фамилию автора, описание произведения, ссылку на спорный источник, основания прав и требование об исключении. Секретариат регистрирует дату, канал, отправителя и уникальный ID, не распространяя персональные данные в общедоступном рабочем чате. На первом этапе проверяется, относится ли объект к проверяемой модели и версии. У заявителя запрашиваются ровно те документы, которые нужны для идентификации объекта и его права, без необоснованного требования передать паспорт или все коммерческие договоры.

Затем команда раздельно оценивает перспективу прекращения будущего сбора, юридическую позицию по уже созданным копиям и возможное воспроизведение в выходах модели. Эти вопросы не сводятся к одной кнопке «удалить». Для любого утверждения о полном удалении из весов требуется техническое доказательство, которого у многих систем нет; поэтому обещание формулируется строго по фактически выполненным действиям. Если компания не располагала исходником в заявленном корпусе, ответ должен подтверждать объём проведённой проверки и её ограничения.

В заключение заявителю направляется понятный ответ с перечнем рассмотренных объектов, выполненных мер и неразрешённых вопросов. Журнал фиксирует пересмотр решения при новых документах и право на обращение к компетентному органу. Если претензия касается одновременно авторства фотографии и персонального изображения, передаются двум компетентным направлениям, не смешивая применимые правовые основания. [1][20]

Приложение 49. Контроль публикации экспертного вывода Observatoire ARGA

Перед публикацией заключения назначается редактор, не участвовавший в сборе фактов, который проверяет соответствие текста внутреннему evidence ledger. Каждое юридически значимое утверждение получает ID документа, дату проверки, юрисдикцию, статус нормы и владельца проверки. Рассматриваются возможные конфликты интересов: платное консультирование одной из сторон,

финансирование исследования, участие сотрудников в связанном споре и прежние публичные оценки конкретной компании. Конфликт не всегда исключает публикацию, но должен вести к предусмотренному заранее независимому разрешению вопроса о рецензии или отводе.

В тексте отдельно маркируются подтверждённый факт, юридическая оценка, аналитический риск, условный сценарий и гипотеза. Если проверяемая компания не ответила на запрос, не допускается формулировка «она признала нарушение». Соответственно, ответы представителей учитываются с указанием их доказательного статуса и возможной проверки. Нельзя скрывать неблагоприятные для предварительной гипотезы факты или задним числом менять правила выборки ради желаемого результата.

Публичная версия не раскрывает целиком чужие охраняемые произведения, документы с коммерческой тайной, персональные сведения без основания или секретные идентификаторы доступа. Приложение для внутренних рецензентов может содержать защищённые доказательства. Обе версии должны иметь общую нумерацию выводов и таблицу соответствия для последующей проверки.

Приложение 50. Контроль трансграничного судебного сценария

Условный разработчик расположен в государстве X, обучает модель при участии поставщика данных из Y и делает продукт доступным в ЕС. Правовой анализ начинается с карты действий и территорий, а не с автоматического переноса любой национальной судебной позиции на всю цепочку. Нужно отделить право, регулирующее соответствующее исключительное право и возможный деликт, от права договора о поставке данных и процессуальных правил доступа к доказательствам. Наличие представительства в ЕС не означает, что любое требование о копировании за рубежом разрешается одними и теми же материальными нормами.

Судебные решения США о fair use по отдельным обстоятельствам нельзя механически использовать вместо анализа статей 3 и 4 DSM в ЕС. Равно ошибочно представлять немецкое решение первой инстанции по memorization как универсальный ответ о законности копирования всех книг, песен и изображений. Для каждой позиции юрист фиксирует юрисдикцию, стадию решения, фактическую ситуацию и пределы прецедентного значения. [1][11][12][13]

Практический итог включает карту потенциальных компетентных органов, предполагаемые виды требований, риски сохранности доказательств, правила конфиденциальности, план запросов к контрагентам и предупреждение о необходимых локальных консультациях. Эта карта является аналитическим инструментом, а не юридическим заключением по конкретному делу без фактического пакета.

Приложение 51. Условный расчёт роялти без произвольной презумпции ущерба

Исходные значения примера условны: каталог содержит 2 000 произведений, из которых использование для оговорённого продукта согласовано для 1 500. Для остальных 500 нет подтверждённого договорного основания, но это ещё не устанавливает нарушение из-за возможных исключений и других правовых оснований. В переговорной модели стороны рассматривают цену доступа на определённый период к конкретному типу использования. Если согласованная ставка составила бы 8 евро на объект, иллюстративная сумма за 500 объектов равнялась бы 4 000 евро, но из этого расчёта не следует фактический размер ущерба или рыночная справедливость ставки.

Для полноценной финансовой оценки нужно установить релевантный лицензируемый рынок, сопоставимые сделки, объём фактического использования, заменяемость продукта, возможность распределения между обладателями разных прав и условия прекращения. Условная ставка может служить отправной точкой переговоров, но не доказательством того, что суд должен назначить именно эту компенсацию. Одно произведение иногда участвует через несколько прав, и повторное начисление без основания приводит к завышению суммы.

Документ рекомендует вести три раздельных числа: измеренная фактическая выручка, наблюдаемый объём контента и предполагаемый переговорный платёж. Нельзя смешивать их как будто это одна

переменная. При публикации раскрывается метод расчёта и влияние ключевых допущений на итог; чувствительная коммерческая информация может быть показана в диапазоне или закрытом приложении.

Приложение 52. Реестр версий и критических дат правовой проверки

Для каждого публикуемого исследования создаётся master register: версия доклада, дата среза законодательства, дата просмотра официальных FAQ, версия исследованной модели, набор исследовательских данных, действующие судебные решения и нерешённые вопросы. Изменение нормативного текста порождает запись об оценке влияния на разделы, а не автоматическую перезапись исторической версии исследования. Консультация Европейской комиссии Technology and Copyright на дату 23 октября 2026 года ещё открыта до 3 ноября 2026 года; её нельзя представлять как уже окончательно принятое изменение директив. [5]

В реестре для каждого значимого события указаны ссылка на источник, точная дата, статус «закон/акт/судебное решение/официальное разъяснение/консультация/исследовательское предложение» и действие ответственного редактора. Особая проверка требуется при публикации апелляционных актов, обновлении template GPAI либо принятии новых технических спецификаций по резервированию прав. Архивированная копия исследовательской версии сохраняет ранее сделанные выводы и позволяет понять, какие из них устарели.

Последний шаг — контроль публикационного пакета. Word и PDF получают идентичные реквизиты, логотип и название версии; в источниках присутствуют активные гиперссылки; содержание ссылается на правильные страницы; все количественные примеры обозначены как учебные. Подпись или институциональное утверждение не добавляются без фактического решения руководства. Итоговый редакционный лист разделяет технический preflight, юридическую рецензию и принятие решения о публикации.

Приложение 53. Таблица разграничения TDM и выдачи модели

Аудитор создаёт две независимые записи для одного произведения. Первая отражает жизненный цикл входа: дату законного доступа, создание временных копий, применимую статью 3 либо 4 DSM, наличие резервирования и результат проверки. Вторая отражает поведение вывода: были ли запросы с референсом, воспроизведены ли существенные оригинальные элементы, какие исключения и договорные разрешения могут иметь значение. Одно действие не оправдывает автоматически другое. [1][6]

Практический пример: архивный текст был законно использован для исследовательского анализа, но модель затем выдала протяжённый характерный фрагмент без предоставления оригинала пользователем. Ссылка на исследовательский TDM сама по себе не завершает анализ второго действия. Обратная ситуация: похожий общий пересказ фактов в ответе модели не доказывает, что обучающий корпус содержал защищённую статью. В журнале фиксируются отдельные стандарты доказательства и не допускается единая галочка «copyright cleared».

Стадия	Документ	Самостоятельный вопрос
Получение доступа	Договор или условия законного доступа	Можно ли исследовать объект?
Создание учебной копии	Фактические операции и исключение	Допустимо ли воспроизведение для TDM?
Передача партнёру	Лицензия/условия взаимодействия	Есть ли право дальнейшего использования?
Выдача модели	Протокол, совпадения, режим входа	Есть ли отдельное использование охраняемой формы?

Приложение 54. Матрица возражений правообладателя по категориям контента

Заявления правообладателей разделяются на текст, изображение, аудио, видео, код и базы данных. Для каждой категории создаётся обязательный минимум: идентификатор объекта, дата первоначального получения, заявитель, тип права, уровень подтверждения полномочий, конкретный требуемый результат и контакт ответственного. Запрашивать лишние биометрические или паспортные сведения для

подтверждения прав на произведение не следует без отдельного обоснования. Если заявление подано агентом, проверяется его мандат.

Для каждого объекта доступен отдельный статус: зарегистрировано; необходимо уточнение; проверено основание; временно исключено из нового сбора; мера выполнена; отказ с мотивировкой; открыто на пересмотр. Эти статусы не равны судебному признанию прав. Техническая команда подтверждает фактическое изменение датасета или блокировки, юридическая — смысл действия, редактор — корректность ответа заявителю. Внешний ответ не должен обещать безусловное «стирание из весов», если такая мера технически не выполнена и не доказана.

Контрольная выборка претензий раз в квартал позволяет оценить среднее время ответа, долю исправленных ошибок классификации и повторяемость обращений к одному источнику. Такая статистика должна содержать числа оснований, а не исключительно проценты, чтобы малый объём выборки не выглядел как устойчивый тренд. При обнаружении массовой ошибки механизм эскалации затрагивает категорию источников, а не только отдельную заявку.

Приложение 55. Схема доступа независимого аудитора к закрытому корпусу

Стандарт data room определяется договором о целях, пределах доступа, ответственности за сведения и процедуре фиксации действий. Аудитор заранее заявляет набор проверок, но может сохранить право случайного отбора по заранее согласованной методике. Заказчик не должен подменять выборку исключительно удобными примерами. Просмотр документов, их экстракты и сам факт проверки фиксируются журналом; полные файлы не выгружаются в открытые инструменты и не используются для посторонних моделей.

Разграничение доступа имеет три слоя: технический специалист проверяет версионность и происхождение; юрист оценивает релевантные договорные фрагменты; внешний рецензент получает обезличенный evidence ledger и воспроизводимые параметры тестов. Для каждого слоя обосновывается, нужна ли видимость самого охраняемого произведения. По окончании сеанса фиксируется, какие сведения сохранены, какие уничтожены, какие обязаны сохраняться из-за спора, и кто подтвердил выполнение процедур.

Необходимо заранее описать, как ведёт себя аудит при отказе предоставить ключевой документ. Допустимые выводы включают «проверка ограничена», «поставщик предоставил только заверение», «для данного вывода доказательство отсутствует»; запрет на доступ не подменяет доказательство нарушения. Закрытый режим может защищать коммерческую тайну, но не оправдывает позитивное независимое заключение при фактическом отсутствии проверяемых материалов.

Приложение 56. Протокол стратифицированной выборки источников

Перед извлечением выборки аудитор определяет генеральную совокупность, временной период, категории источников, известные кластеры, основные гипотезы и ожидаемую долю ошибок. Технически удобные файлы, которые поставщик заранее отобрал для презентации, не образуют случайной выборки корпуса. Для смешанного каталога допустимо деление на страты: книги, новости, изображения, музыка, код, специализированные архивы. Размер выборки может различаться по категориям, но затем итоговая оценка должна учитывать веса и вероятности включения.

Учебный пример: 10 000 элементов разбиты на 8 000 текстов и 2 000 фотографий. Если аудитор проверит по 100 из каждой группы и обнаружит 10 неопределённых правовых оснований для текстов и 20 для фото, объединять 30/200 = 15% в оценку всего корпуса без весов некорректно. При случайных репрезентативных выборках грубая взвешенная оценка составила бы 0,8×10% + 0,2×20% = 12%, не учитывая доверительные интервалы и корреляцию. При этом «неопределённое основание» не равно доказанному нарушению.

Результат обязан отдельно сообщать численность и долю непроверенных объектов, пределы доступности данных, критерии исключения и метод отбора. Если случайность не обеспечена, вывод называется проверкой целевой выборки, а не статистической оценкой всего корпуса. Разделять фактическое

отсутствие лицензии и вероятность действительного правового исключения особенно важно в итоговых таблицах.

Приложение 57. Регламент проверки версии документа во время консультации ЕС

Для каждой публичной консультации фиксируются источник, URL, дата публикации, дата завершения, список официальных материалов и рабочая версия экспертной позиции. Доклад №23 анализирует targeted consultation Technology and Copyright от 29 сентября 2026 года; по состоянию на 22 октября срок представления позиции истекает 3 ноября 2026 года. До официального завершения и последующего нормативного действия предложения участников не являются принятым законодательством. [5]

Перед подачей проверяются мандат ARGA на направление позиции, согласованное название организации, язык, отсутствие разглашения клиентских сведений, доказательственная состоятельность статистики и корректность ссылки на конкретный вопрос анкеты. Условные кейсы обозначаются как условные, а реальная статистика приводится только с проверенным методом сбора. Следует различать предложения к праву, замечания к технической реализации и просьбы о дополнительных исследованиях; смешивание их создаёт у адресата впечатление, что организация требует немедленного изменения закона ради одной технической проблемы.

Контроль на день подачи включает проверку обновлённого текста консультации, наличие подтверждения регистрации и архивирование самой направленной позиции в неизменяемой версии. Внутренний отчёт фиксирует фактический статус «подготовлено», «направлено» или «получено подтверждение». Нельзя заявлять о направлении или принятии позиции без соответствующего документа.

Приложение 58. Каталог запретов на использование методики аудита не по назначению

Список ограничений публикуется вместе с методикой. Она не устанавливает авторство фотографии по одному перцептивному хешу, не доказывает происхождение конкретного текста из обучающего корпуса только по сходству ответа, не даёт универсальную лицензию на обучение по статье 53 AI Act и не подтверждает законность каждой транзакции на основании общего vendor score. Тесты на выходы одной модели не переносимы автоматически на все будущие обновления.

Отдельно запрещаются некорректные формулировки в рекламных материалах: «сертифицировано Observatoire ARGA» без настоящей процедуры и установленного стандарта, «100% copyright compliant» по малой выборке, «установлен ущерб» без экономического анализа и «суд подтвердил разрешение на любое обучение» при цитировании решения по конкретному спору. Для каждого публичного показателя необходимо указывать предмет и предел тестирования, дату, лиц с конфликтами интересов и степень доступа аудитора.

Нарушение этих условий приводит к обязательной редакционной проверке публичного материала и при необходимости к исправлению. Если контрагент продолжает использовать исследовательский отчёт как доказательство полномочий, которых он не предоставляет, Observatoire вправе опубликованно уточнить его действительную сферу и обратиться за правовой оценкой ситуации. Такая мера защищает и исследовательскую репутацию, и интересы авторов.

Приложение 59. Проверка совместимости лицензий и сублицензий на код

Форма проверки начинается с идентификации версии репозитория, commit hash, путей модулей, зависимостей, текста лицензий и даты выгрузки. Приоритет отдаётся первичным лицензиям и условиям конкретного кода, а не агрегированному индикатору «лицензия MIT» на карточке репозитория. Отдельно учитываются файлы с иной лицензией, фрагменты сторонних библиотек и изменения авторских уведомлений. [6]

Для учебного продукта аудитор составляет матрицу: объект; правообладатель; тип лицензии; есть ли право копирования; какие уведомления обязательны; возможные ограничения распространения; право

передать материал подрядчику; потенциальный случай существенного воспроизведения в выходном коде. Если условия требуют сохранения текста лицензии, факт удаления уведомления при подготовке данных фиксируется как конкретный риск. Но нельзя автоматически делать из этого вывод о способе, которым модель впоследствии выдаст код.

Заключение содержит отдельные категории: договорная неопределённость, лицензия с особыми условиями, возможное воспроизведение охраняемого фрагмента и информационная безопасность генерируемого кода. Последняя категория требует другого стандарта проверки. Аудитор отмечает случаи, где применимое право может допускать исключение TDM, и не утверждает, что все публикации открытого кода обязательно требуют индивидуального договора.

Приложение 60. Форма сравнительного заключения по судебной практике США и ЕС

Для каждого дела фиксируются: суд, юрисдикция, дата, процессуальная стадия, предмет копирования, доказанные факты, вопрос для решения и границы правового вывода. Включаются решение по частичному вопросу fair use по Bartz v. Anthropic от 23 июня 2025 года, решение по ходатайствам в Kadrey v. Meta от 25 июня 2025 года и отдельное окончательное определение от 20 июля 2026 года об утверждении мирового соглашения по Bartz. Они показывают разные стадии и последствия, поэтому в реестре не должны объединяться под одним статусом «дело решено в пользу ИИ». [12][13][22][23]

Отдельной строкой фиксируется решение мюнхенского суда по делу GEMA v OpenAI от ноября 2025 года с его первоинстанционным статусом, предметом memorization/output и неопределённостью последующего пересмотра. Национальный судебный акт не становится общеевропейской нормой лишь потому, что касается известного разработчика. Сравнение с Директивой DSM проводится с указанием различий материально-правовых тестов и национальных режимов. [11]

Итоговая матрица не содержит единой колонки «победа ИИ / победа правообладателя». Вместо этого выводит, какое конкретное юридическое утверждение поддерживает решение, какой факт отличает его от рассматриваемого кейса и какие вопросы остаются открытыми. Такая форма предотвращает использование неполных пресс-релизов вместо судебной мотивировки.

Приложение 61. Проверка издательского набора после смены лицензии

В учебном сценарии разработчик загрузил редакционные материалы в январе; лицензионные условия изменились в марте; новая версия модели готовится в сентябре. Первая задача — установить юридическое основание января и область действия последующих изменений. Нельзя автоматически делать вывод, что любой новый запрет ретроспективно отменил все прошлые действия, или что разрешение января продолжает покрывать любой новый сбор и коммерческий способ использования. Важно сохранить исторические версии условий, подтверждения согласования и реальные сроки доступа.

Реестр должен показать, какие файлы уже включены в исходный обучающий корпус, какие поступали в новые циклы и какие впервые были извлечены после изменения условий. Вопросы дальнейшего использования и технической памяти модели рассматриваются отдельно. Если речь идёт об opt-out, проверяются статья 4 DSM, надлежащая форма, время резервирования и применимые национальные особенности. Решение не принимается одним лишь сравнением дат без анализа договора и характера действий. [1]

Финальная таблица предлагает меры: прекращение нового сбора, сохранение подтверждений прошлой правомерности, переговоры о дополнительной лицензии, исправление публичной сводки при необходимости и ограниченное тестирование выходного воспроизведения. Ни один шаг не обозначается как всеобщее требование закона без самостоятельного основания.

Приложение 62. Протокол проверки персональных данных в лицензируемом архиве

Авторская лицензия на журналистский архив и правовое основание обработки персональных данных — два независимых вопроса. Условный каталог содержит биографические сведения, фото, упоминания судебных разбирательств и имена третьих лиц. Аудитор выявляет персональные данные, возможные специальные категории и сведения об уголовных правонарушениях, а также назначение дальнейшего обучения. При этом не следует считать любое упоминание суда автоматически доказанной судимостью. Для данных, подпадающих под статьи 9 и 10 GDPR, необходим отдельный анализ специальных условий. [20]

Для каждого поля проводится оценка минимизации, точности, актуальности, риска объединения с другими источниками, законного основания, ограничений международной передачи и обработки запросов субъектов. При использовании данных в AI-модели проверяется возможность позднего воспроизведения ложного или устаревшего статуса лица. Авторско-правовой договор с издателем не снимает обязанности исправлять ошибочные персональные сведения или надлежащим образом рассматривать запросы.

Финальное решение может рекомендовать исключение отдельных полей из нового обучения, псевдонимизацию в аудитном хранилище, усиление контроля качества и сохранение доказательств в закрытом режиме. Публичный доклад по такому кейсу должен показывать типовую проблему без раскрытия персональных деталей конкретного клиента либо дела.

Приложение 63. Протокол внутреннего рецензирования перед утверждением

Рецензент получает окончательную версию Word, PDF и реестр нормативных источников. Он проверяет, соответствуют ли ссылки действующей редакции актов на дату доклада, не представлены ли первоинстанционные решения как окончательные, подтверждены ли расчёты, ясно ли указаны масштабы выборки и используются ли авторско-правовые термины одинаково во всех разделах. Каждое замечание имеет категорию «юридическая точность», «доказательство», «метод», «редактура», «верстка» и приоритет.

Изменения вносятся с фиксацией затронутого раздела и краткого обоснования. Если исправление вводит новый источник, его следует связать с текстом, а если удаляет существенный вывод, нужно проверить, не остались ли отсылки в резюме и практических приложениях. После завершения проводится новый PDF-экспорт, пересчёт оглавления и визуальный осмотр титульной страницы, таблиц, колонтитулов и конца библиографии. Успешная техническая проверка не означает автоматического юридического одобрения.

Рецензент указывает статус: можно представить руководству; нужны уточнения; содержательно существенные риски; публикация пока не рекомендована. Дату и подпись ответственного руководителя нельзя добавлять от его имени без отдельного подтверждения. Внутреннее использование проекта и официальная публикация различаются.

Приложение 64. Итоговый редакционный реестр нерешённых вопросов

К моменту выпуска исследования фиксируются вопросы, на которые имеющаяся база не даёт окончательного ответа: квалификация некоторых видов обучения в национальном праве, юридический эффект конкретных машинных резервирований, полнота транзакционных лицензий, доступность доказательств закрытых корпусов, правовой статус отдельных результатов генерации и последствия дообучения. Для каждого вопроса указываются существенные альтернативные интерпретации, необходимое новое доказательство и событие, требующее пересмотра.

Пример записи: «Подтверждено ли право посредника лицензировать архив для коммерческого обучения?»; статус «не подтверждено документально»; требуемый материал «договор о полномочиях»; ответственный «юридическая группа»; критический срок «до следующего цикла загрузки». Другая запись: «Есть ли риск воспроизведения фрагмента при работе модели версии V?»; статус «не оценено»; требуемый материал «контрольный протокол без референсного ввода»; ответственный «технический эксперт». Эти вопросы не смешиваются в одну безличную общую оценку.

Реестр публикуется в сокращённом виде вместе с исследовательским отчётом для честного обозначения границ вывода. Он является инструментом подготовки новой версии доклада, но не свидетельствует о проведении технических испытаний или получении документов, которых фактически не было. При закрытии пункта сохраняются дата, доказательство и причина изменения статуса.

Источники и нормативно-методические основания

Дата проверки нормативной и методологической базы: 23 октября 2026 года. Номера в квадратных скобках внутри доклада соответствуют следующим документам. Нормативные акты, судебные решения, проекты и добровольные рекомендации различаются по юридическому статусу.

[1] Директива (ЕС) 2019/790, статьи 3 и 4: исключения text and data mining. Европейский парламент и Совет; действующее право ЕС. Открыть первоисточник

[2] Регламент (ЕС) 2024/1689, редакция 27 июля 2026 года: статьи 53 и 111. EUR-Lex; AI Act; действующее право ЕС. Открыть первоисточник

[3] FAQ о публичном шаблоне сводки обучающего контента GPAI. Европейская комиссия; официальные разъяснения. Открыть первоисточник

[4] Questions and answers on GPAI Code of Practice. Европейская комиссия; добровольный инструмент и официальные разъяснения. Открыть первоисточник

[5] Technology and Copyright: целевая консультация от 29 сентября 2026 года. Европейская комиссия; консультация открыта до 3 ноября 2026 года. Открыть первоисточник

[6] Директива 2001/29/ЕС об авторском праве в информационном обществе. EUR-Lex; право воспроизведения и сообщения публике. Открыть первоисточник

[7] Директива 96/9/ЕС о правовой охране баз данных. EUR-Lex; sui generis право изготовителя базы. Открыть первоисточник

[8] Guidelines on obligations for General-Purpose AI providers. Европейская комиссия; официальное толкование охвата GPAI. Открыть первоисточник

[9] Report on Copyright and Artificial Intelligence от 18 марта 2026 года. Правительство Великобритании; официальный доклад о консультации. Открыть первоисточник

[10] U.S. Copyright Office, Copyright and Artificial Intelligence, Part 3 (pre-publication), май 2025. Copyright Office; аналитический доклад, не закон. Открыть первоисточник

[11] LG München I, GEMA v OpenAI, 42 O 14139/24, сообщение суда 11 ноября 2025. Bayerisches Staatsministerium der Justiz; решение первой инстанции, не окончательно на указанную дату. Открыть первоисточник

[12] Bartz v. Anthropic PBC, Order on Fair Use, 23 июня 2025 года. U.S. District Court, N.D. California; решение по отдельному вопросу, позднейшее завершение дела см. [22]. Открыть первоисточник

[13] Kadrey v Meta Platforms, 25 июня 2025: WIPO Lex case summary. WIPO Lex; судебная практика США и пояснение объёма решения. Открыть первоисточник

[14] WIPO Learning Machines: An Introduction to AI and IP for SMEs, 2026. WIPO; сравнительное методическое исследование. Открыть первоисточник

[15] Singapore Copyright Act 2021, sections 243–244. Singapore Statutes Online; computational data analysis. Открыть первоисточник

[16] EDPB Opinion 28/2024 on AI models and personal data. European Data Protection Board; 18 декабря 2024 года. Открыть первоисточник

[17] Explanatory Notice and Template for Public Summary of Training Content, 24 июля 2025. Европейская комиссия; шаблон и пояснения. Открыть первоисточник

[18] General-Purpose AI Code of Practice. Европейская комиссия; добровольный кодекс. Открыть первоисточник

[19] TRIPS Agreement, Article 13: limitations and exceptions. World Trade Organization; международный трёхступенчатый тест. Открыть первоисточник

[20] GDPR, Регламент (ЕС) 2016/679. EUR-Lex; законность и подотчётность обработки персональных данных. Открыть первоисточник

[21] Code de la propriété intellectuelle, article L122-5-3. Légifrance; французские исключения TDM. Открыть первоисточник

[22] Bartz v. Anthropic PBC, Order Granting Final Approval of Class Action Settlement and Judgment, 20 июля 2026 года. U.S. District Court, N.D. California, Docket 680; официальный процессуальный итог по мировому соглашению, не общее решение по существу всех правовых вопросов. Открыть первоисточник

[23] Kadrey v. Meta Platforms, Inc., Order on Cross-Motions for Partial Summary Judgment, 25 июня 2025 года. U.S. District Court, N.D. California, Docket 598; текст судебного решения (в дополнение к обзору WIPO [13]). Открыть первоисточник

Редакционная оговорка: текст содержит предложения исследовательского характера. Отсутствие фактического доступа к закрытому корпусу, исходным моделям либо договорам конкретного разработчика не позволяет делать вывод о его индивидуальном правонарушении или о полной законности практики. Редакция от 23 октября 2026 года включает проверку правовой хронологии, выборки ключевых официальных источников, арифметических примеров и отображения Word/PDF. Она не удостоверяет законность конкретных закрытых обучающих массивов. Перед официальным утверждением требуется отдельная профильная юридическая и научная рецензия.