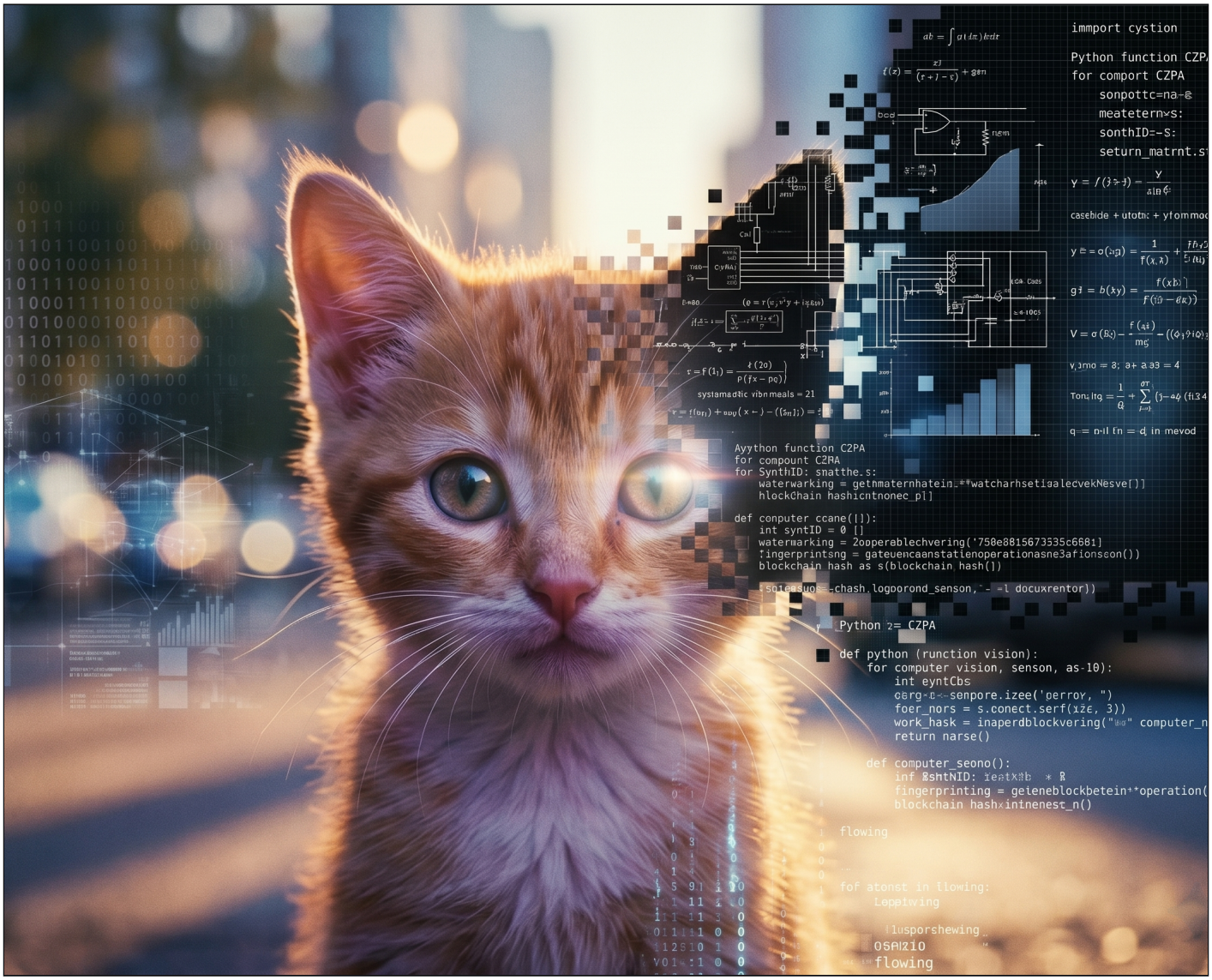

Маркировка ИИ-изображений

Правовой статус, требования EU AI Act и технические средства

Степан Леонтьев · Евдокия Шматина

ПРЕПРИНТ

2026-06-04

// СОДЕРЖАНИЕ

Правовой статус и регулирование

ВВЕДЕНИЕ

Нейросети генерируют контент с такой скоростью, что люди больше не могут доверять собственным глазам. Появляется логичный вопрос: как достоверно подтвердить, кем был создан цифровой объект?

Глобального консенсуса пока не существует, и разные юрисдикции решают эту проблему по-своему. Евдокия Шматина разберёт существующие правовые подходы: от России, где специальное регулирование отсутствует, до строгих европейских правил, где информация о происхождении контента превратилась в техническую обязанность.

Однако законы не работают без технологий. Степан Леонтьев расскажет о технических аспектах: от открытых «паспортов контента» (C2PA) до невидимых меток SynthID. В финале авторы совместно дадут прогнозы о двойственном потенциале новой цифровой прозрачности.

О статье. Когда писательский зуд побеждает академическую точность, рождаются подобные статьи, которые по существу являются научно-популярными очерками для первичного погружения в тематику. В короткий срок авторы изучили огромное количество материалов в абсолютно незнакомой для себя теме. Стыдливая надпись «ПРЕПРИНТ» означает, что некоторые аспекты были исследованы не до конца (наличие искомых метаданных в технических устройствах, общие охваты применения технологии), поэтому призываем к критике.

// 1 ЧТО ГОВОРИТ ПРАВО?

Россия и большинство правопорядков мира придерживаются **человекоцентричного подхода**, то есть считают единственным субъектом авторского права человека и признаком его творческий труд.

В российском праве[1] эта конструкция опирается на три нормы:

1. Ст. 1228 ГК РФ признаёт автором гражданина, творческим трудом которого создан результат. Те, кто оказал лишь техническое, консультационное или организационное содействие, авторами не являются.
2. Ст. 1257 ГК РФ воспроизводит тот же принцип для произведений науки, литературы и искусства.

3. п. 5 ст. 1259 ГК РФ уточняет: идеи, методы, концепции, факты и алгоритмы авторским правом не охраняются.

Из этого следует прямой вывод: контент, сгенерированный искусственным интеллектом, охраноспособен только при наличии творческого вклада человека. Исключительно машинная генерация авторского права не порождает.

Ту же позицию поддерживает и Верховный суд в Постановлении Пленума от 23.04.2019 № 10 «О применении части четвёртой Гражданского кодекса Российской Федерации»[2]. Постановление подтверждает, что творческий характер создания произведения не зависит от того, создано произведение автором собственноручно или с использованием технических средств. Вместе с тем результаты, созданные с помощью технических средств в отсутствие творческого характера деятельности человека (например, фото- и видеосъёмка работающей в автоматическом режиме камерой видеонаблюдения, применяемой для фиксации административных правонарушений), объектами авторского права не являются (абз. 5 п. 80).

При этом Верховный суд отмечает, что пока не доказано иное, результаты интеллектуальной деятельности предполагаются созданными творческим трудом (абз. 2 п. 80).

— СЛУЖЕБНЫЕ ПРОИЗВЕДЕНИЯ И СГЕНЕРИРОВАННЫЙ КОНТЕНТ

Сказанное применимо и к служебным произведениям. Отмечаю это для представителей бизнеса. Потому что не всегда созданное работником принадлежит ему, а значит, не станет принадлежать бизнесу. Это важно знать и при покупке креативов у агентства.

По смыслу ст. 1295 ГК РФ исключительное право на служебное произведение принадлежит работодателю. Однако право авторства остаётся за работником и не может быть передано работодателю или иному лицу. Это применяется и к сгенерированному контенту, созданному сотрудником при исполнении трудовых обязанностей, если творческий вклад человека присутствует. В ситуации со служебным произведением требуется доказать не только, что оно было создано в рамках трудовых обязанностей с использованием инструментов работодателя в рабочее время, но и что оно создано человеком и его творческим трудом.

— ПРОБЛЕМА «ДОСТАТОЧНОСТИ» ТВОРЧЕСКОГО ВКЛАДА

Вопрос о том, насколько большой должна быть доля творческого труда в любом произведении, теперь ставится всё чаще в рамках судебных споров. Если раньше любые технические инструменты облегчали творческую работу человека, но не могли полностью создать произведение с нуля только по общему описанию, то теперь это будничный запрос.

А значит, вопросы о том:

- насколько точным и подробным должен быть промпт;
- насколько существенно он должен редактироваться после первой и последующей генерации;
- должно ли изображение после генерации дополнительно обрабатываться человеком —

должны быть решены судом. Не забудем и о том, что по одному и тому же запросу и разные модели искусственного интеллекта, и даже одна и та же модель могут выдать (и, скорее всего, выдадут) разные результаты. Этот факт усложняет создание прозрачной классификации или безусловного подхода к решению проблемы творческого вклада в создание произведения с помощью искусственного интеллекта.

// 2 СУДЕБНАЯ ПРАКТИКА РОССИИ

— DEEPFAKE КАК ИНСТРУМЕНТ МОНТАЖА

Deepfake (от англ. *deep learning* + *fake*) — синтетический видео-, фото- или аудиоматериал, в котором нейросеть правдоподобно воссоздаёт или подменяет внешность, голос либо движения реального человека.

Арбитражный суд города Москвы, Девятый арбитражный апелляционный суд[3] и Суд по интеллектуальным правам[4] пришли к единому мнению по делу № А40-200471/2023 о взыскании компенсации за нарушение исключительного права на произведение.

Суд признал видеоролик с технологией deepfake объектом авторского права и удовлетворил иск о защите исключительных прав. Указал, что deepfake — это технология, основанная на использовании генеративно-состязательных нейросетей (GAN), которые позволяют частично преобразовывать исходный видеоряд с помощью алгоритмов, созданных человеком. Такое преобразование не исключает того факта, что исходный видеоряд создан творческим трудом коллектива авторов.

Суд отклонил довод ответчика о том, что видеоряд полностью сгенерирован нейросетью, а также о том, что созданный видеоролик не является объектом авторского права по причине использования технологии deepfake, — потому что такая технология является дополнительным инструментом обработки (технического монтажа) видеоматериалов, а не способом их создания.

△ **ВНИМАНИЕ**

Доказательством принадлежности прав послужил договор об отчуждении исключительного права. Суд применил сугубо легалистский подход: наличие права у конечного правообладателя подтверждается самим договором, наличие права у передающей стороны — фактом создания ролика группой авторов по заказу.

Показательно, что ни одна из сторон не вышла за пределы договорной цепочки. Сторонами не предоставлено экспертного заключения о наличии и достаточности творческих элементов в исходном ролике. Истец не представил log-файлы или иные технические материалы, подтверждающие характер и объём промптов. Ответчик, в свою очередь, не привёл технических доказательств. Например, данных о цифровой подписи или метаданных файла, которые прямо указывали бы на генерацию произведения искусственным интеллектом.

Иными словами, вопрос о творческом вкладе человека в создание спорного контента остался за рамками судебного исследования. Стороны и суд действовали в логике традиционного авторско-правового спора, где ИИ как инструмент создания фактически не был предметом доказывания.

— БРЕМЯ ДОКАЗЫВАНИЯ ОТСУТСТВИЯ АВТОРСТВА

Решение Арбитражного суда Краснодарского края и Пятнадцатого арбитражного апелляционного суда по делу № А32-40085/2024 [5] о взыскании денежных средств за нарушение исключительных авторских прав на произведения изобразительного искусства.

Суд признал подтверждённой принадлежность авторского права на изображения (рисунки), поскольку ответчиком иное не доказано, презумпция авторства не опровергнута.

Ответчик заявлял, что спорные изображения не являются результатом творчества конкретного человека, а продуктом искусственного интеллекта. Доказывал, что продукт искусственного интеллекта, созданный на основе чужих изображений, не является авторским. Суд указал, что истцом представлены копии актов приёма-передачи исключительных прав на рисунки, трудовые договоры с работниками, которыми подтверждается, что обладателем исключительных прав на рисунки является истец.

△ **ВНИМАНИЕ**

Доказательствами принадлежности прав послужили акты приёма-передачи исключительных прав и трудовые договоры с работниками. Суд снова применил легалистский подход, не выходя за пределы документарной цепочки. Суд прямо обозначил границы презумпции: необходимость дополнительных доказательств возникает лишь тогда, когда авторство оспаривается надлежащим образом. Пока доказательств нет, презумпция творческого труда работает в пользу правообладателя.

ВЫВОДЫ

Суды пока не создали формулу для признания доли творческого труда достаточной для отнесения объекта авторских прав к охраняемым. Чаще всего достаточным доказательством принадлежности прав правообладателю признаётся факт договора и акта по отчуждению исключительного права. В части оспаривания наличия авторского права и, как следствие, исключительного права аргументов у сторон нет — только ссылка на эмпирический опыт оспаривающих и их визуальное знакомство со сгенерированными искусственным интеллектом объектами.

Однако есть два примера другого подхода к использованию объектов, созданных при помощи искусственного интеллекта. Это юридическая рамка в ЕС и технический инструмент от Google. К их доказательственной роли мы вернёмся в § 16.

// 3 ПРАВОВОЙ КОНТЕКСТ В ЕС

Почему я буду говорить об ЕС? В России недавно опубликован законопроект об использовании искусственного интеллекта, но он был раскритикован разными представителями и будет дорабатываться[6]. Кроме того, он носит рамочный характер и не даёт полноценного механизма работы с технологией. В США законодатели пока ограничиваются отраслевыми рекомендациями. В Китае тоже нет единого акта, только разрозненные регламенты и положения[7]. Европа же близка к России в правовой защите авторского права и имеет очень полный с технической и правовой точек зрения правовой акт — **EU AI Act**[8].

Начну с общих подходов и дальше углублюсь в сам EU AI Act.

В основе европейской системы защиты авторского права стоит Бернская конвенция, Женевская конвенция и другие международные соглашения ВОИС, которые ратифицированы практически всеми европейскими странами и обязательны для членов ЕС. Право ЕС гармонизирует эти нормы через директивы, в том числе Директиву 2001/29/EC (InfoSoc Directive)[9] и Директиву (ЕС) 2019/790 о цифровом едином рынке (DSM Directive)[10]. Как и в России, авторские права возникают автоматически с момента создания произведения, без какой-либо регистрации, при условии оригинальности результата. Законодательство не обеспечивает охрану идей и

концепций. Оно предоставляет охрану только способу их выражения в конкретном произведении. Наглядны в части защиты два прецедента:

- В деле № C-5/08 Infopaq International A/S против Danske Dagblades Forening [11] Европейский суд постановил, что хранение и распечатка 11-словных выдержек из газетных статей приравниваются к воспроизведению в соответствии со статьёй 2(a) Директивы об авторском праве (2001/29/EC), если воспроизведённые элементы являются выражением интеллектуального творчества автора;
- В деле № C-145/10 Painer против Standard Verlags GmbH [12] Европейский суд постановил, что портретные фотографии могут пользоваться защитой, если они оригинальны в том смысле, что являются собственным интеллектуальным творением автора, например, если они демонстрируют личный стиль фотографии.

Европейский союз достаточно ревностно защищает авторов произведений не только в традиционных формах искусства, но и в цифровой среде. Активно развивает требования к разработчикам и поставщикам искусственного интеллекта с этической точки зрения.

В июне 2024 г. Европейский парламент и Совет Европейского союза совместно издали крупнейший единый акт, регулирующий правила по использованию искусственного интеллекта на территории Европейского союза. Сам документ описывает цели так:

> «Улучшение функционирования внутреннего рынка и содействие внедрению человекоцентричного и заслуживающего доверия искусственного интеллекта (ИИ), при этом обеспечивается высокий уровень защиты здоровья, безопасности, основных прав, закреплённых в Хартии, включая демократию, верховенство права и охрану окружающей среды, от вредного воздействия систем ИИ в Союзе и поддерживается инновация» (п. 1, ст. 1).

Акт устанавливает обязательные требования к поставщикам, импортёрам и дистрибьюторам системы искусственного интеллекта вообще и высокого риска в частности. При несоблюдении требований такая система не может быть выпущена на рынок для использования, а если выпущена с нарушениями — понесёт ответственность.

Размер штрафов: от 7,5 млн до 35 млн евро *или* от 1% до 7% от общего годового оборота по всему миру за предыдущий финансовый год — в зависимости от того, какая сумма больше. Есть исключения для малых и средних предприятий.

// 4 ТРЕБОВАНИЕ ПРОЗРАЧНОСТИ

Особый интерес представляет обязательство по обеспечению прозрачности для поставщиков определённых систем искусственного интеллекта. Вот как эти требования определяет EU AI Act:

Первое требование (ст. 50) касается понятности для пользователя того факта, что он взаимодействует с искусственным интеллектом. Не должно скрываться, что программа или система содержит или полностью состоит из программного кода и моделей нейросетей.

Второе требование (п. 2, ст. 50) — поставщики систем искусственного интеллекта, включая системы ИИ общего назначения, генерирующие синтетический аудио-, графический, видео- или текстовый контент, должны обеспечить, чтобы выходные данные системы ИИ были помечены в машиночитаемом формате и могли быть идентифицированы как искусственно сгенерированные или изменённые.

Третье требование (п. 4, ст. 50) — любой пользователь системы искусственного интеллекта, генерирующей или изменяющей изображения, аудио- или видеоконтент, представляющий собой deepfake, обязан раскрывать информацию о том, что контент был создан или изменён искусственно.

Целью является прозрачность и этичность: отделение искусственного контента от естественного, созданного человеком. Чем больше развивается технология искусственного интеллекта, тем в большей мере сгенерированный контент становится неотличимым для человеческого восприятия. Без дополнительных визуальных и технических (программных) требований идентификация будет усложняться.

> **› ЗАМЕЧАНИЕ**
>
> Регуляторные требования ускоряют внедрение технических инструментов идентификации. Статья 50 *EU AI Act* вступает в полную силу 2 августа 2026 г. *Cybersecurity Information Sheet* от 29 января 2025 г. (NSA, ASD's ACSC, CCCS, NCSC-UK) рекомендует Content Credentials как меру для государственных систем и критической инфраструктуры. Спецификация C2PA при этом проходит fast-track к статусу *ISO 22144*.

Технические инструменты

// 5 C2PA КАК «ПАСПОРТ КОНТЕНТА»

Установить **происхождение** (*provenance*) контента с момента его создания невозможно без технических средств. С этой целью был создан **C2PA** — *Coalition for Content Provenance and Authenticity* — открытый технический стандарт, прикрепляющий к медиафайлам сведения об источнике и истории редактирования.

Стандарт разрабатывается под эгидой Joint Development Foundation (часть Linux Foundation) и публикуется как открытая спецификация. На момент написания статьи последняя версия — **C2PA v2.4** (апрель 2026 г.). Официальные ресурсы коалиции:

- c2pa.org — сайт коалиции: управление стандартом, список участников (руководящий комитет и рядовые члены), программа сертификации продуктов (*conformance*), новости и база знаний с разделом FAQ. Точка входа для общего знакомства со стандартом.
- spec.c2pa.org — все технические документы стандарта с архивом версий, основная спецификация Content Credentials и сопутствующие материалы — *Explainer*, руководства по внедрению и пользовательскому опыту, *Security Considerations*, *Harms Modelling*, рекомендации по ИИ и по идентичности автора. Ресурс для разработчиков.
- contentauthenticity.org — портал инициативы *Content Authenticity Initiative* (CAI), которая отвечает за внедрение стандарта, а не за его спецификацию (§ 6): открытые инструменты и SDK для встраивания и проверки Content Credentials, обучающие материалы и вступление в сообщество CAI.

> **[i] ПРИМЕЧАНИЕ**
> C2PA удостоверяет только происхождение и историю изменений, а не истинность содержимого. Например, манифест, подписанный камерой, не гарантирует, что изображённая сцена не была сфабрикована при фотосъёмке.

К началу 2026 г. к стандарту присоединилось более 6 000 организаций; он внедрён в Adobe Creative Cloud, Microsoft Edge/Bing, Google Pixel 10 и поисковое «About this image», OpenAI (DALL·E 3, Sora), Meta (*признана экстремистской организацией, деятельность запрещена в РФ*), TikTok, LinkedIn, Cloudflare, в камеры Leica, Sony, Nikon, Canon, Samsung и в редакционные конвейеры BBC, AFP, Reuters, AP, CBC.

// 6 ИСТОРИЯ ВОЗНИКНОВЕНИЯ

Content Authenticity Initiative. В ноябре 2019 г. компания Adobe совместно с The New York Times и Twitter запустила Content Authenticity Initiative (CAI) — инициативу, направленную на создание открытого стандарта атрибуции цифрового контента. CAI формулировала задачу с двух сторон: со стороны *создателя* — предоставить инструмент криптографической фиксации авторства и истории правок; со стороны *потребителя* — дать возможность проверить, кем и в какой момент изображение было создано или отредактировано. К CAI быстро присоединились новостные медиа, производители камер, платформы стоковых фото и поставщики облачных сервисов.

Project Origin. Параллельно с CAI коалиция в составе *BBC*, *Microsoft*, *CBC/Radio-Canada* и *The New York Times* с сентября 2019 г. начала работу над Project Origin — инициативой, ориентированной на атрибуцию новостного контента. Если CAI исходила из проблематики авторского творчества (фотографы, иллюстраторы), то Project Origin сосредоточился на новостной экосистеме: как гарантировать, что под брендом крупного вещателя выходит именно его публикация — не подделка и не вырванный из контекста фрагмент. Project Origin разрабатывал собственный технический пилот и публично оформился в 2020 г.; позднее к четвёрке присоединялись другие участники — в том числе IPTC, вошедший в руководящий комитет, и новостные организации для пилотных испытаний. Накопленные технические наработки затем легли в основу единой спецификации.

Образование C2PA. 22 февраля 2021 г. обе инициативы — *Content Authenticity Initiative* и *Project Origin* — объединились в Coalition for Content Provenance and Authenticity (C2PA), формальный консорциум разработки стандартов. Учредителями выступили Adobe, Arm, BBC, Intel, Microsoft и Truepic. Позднее в руководящий комитет (steering committee) вошли Sony, Google, Meta, OpenAI, Amazon и другие отраслевые игроки. Производители камер Canon и Nikon участвуют в коалиции как рядовые члены (general members), а не в составе руководящего комитета.

Декларируемая цель коалиции сформулирована в учредительном пресс-релизе C2PA от 22 февраля 2021 г.: *разработать технические стандарты сертификации источника и истории происхождения медиаконтента для противодействия дезинформации и манипуляциям*. CAI как организационный «верхний» слой при этом не исчез: коалиция C2PA занимается *спецификацией*, а CAI — *внедрением* (информационная работа, документация, инструменты разработчика, отношения с медиа). При этом более 6 000 организаций, собравшихся вокруг стандарта (§ 5), относятся прежде всего к CAI; собственно коалиция C2PA значительно меньше — около десяти членов руководящего комитета и несколько десятков рядовых членов.

// 7 СТРУКТУРА МАНИФЕСТА

Манифест (*manifest*) — носитель сведений о происхождении и истории файла, прикрепляемый к нему по стандарту C2PA. Технически это набор структурированных данных (в текстовом формате *JSON* или его компактной двоичной версии *CBOR*); он хранится либо внутри самого файла — в служебном контейнере для метаданных *JUMBF* (*JPEG Universal Metadata Box Format*, ISO 19566-5), — либо в отдельном сопутствующем файле `.c2pa`. Все манифесты одного файла вместе образуют его «хранилище» (*manifest store*), активным считается последний добавленный.

Манифест состоит из трёх обязательных слоёв:

1. **Утверждения** (*assertions*) — отдельные факты о контенте: модель устройства, GPS-координаты, время съёмки, факт ИИ-генерации, действия редактирования, миниатюра, идентичность автора, ссылки на родительские манифесты. Спецификация описывает около 20 типов таких фактов; обязательны как минимум привязка к содержимому файла и перечень выполненных действий.
2. **Заявление** (*claim*) — компактная опись: перечисляет все утверждения и хранит их «цифровые отпечатки» (хэши). Подписывается именно она, а не сами утверждения по отдельности.
3. **Подпись** (*claim signature*) — цифровая подпись заявления закрытым ключом подписанта. К ней приложены сертификат, цепочку которого можно проследить до доверенного корня (*C2PA Trust List*), и, как правило, доверенная метка времени. Это и позволяет проверить, кто и когда подписал файл и не менялся ли он после.

// 8 CONTENT CREDENTIALS

Content Credentials — это отражение стандарта C2PA для конечного пользователя, условный «бренд» C2PA (который по существу является технической спецификацией), свидетельствующий о наличии валидного манифеста.

Content Credentials состоит из трёх элементов:

- **Иконка «CR»** — квадратный знак-пиктограмма, размещаемый поверх изображения в правом верхнем углу или рядом с медиа в интерфейсе. По клику на иконку открывается панель, отображающая разобранный манифест: подписанта, дату, цепочку редактирований, наличие ИИ-генерации, ingredient-историю.
- **Унифицированная панель Inspect** — стандартизированный UI для отображения манифеста, реализованный в Adobe Inspect, в расширении Adobe Content Authenticity для Chrome/Edge, в contentcredentials.org/verify и в `openai.com/verify`. Цель — единообразный пользовательский опыт независимо от платформы.

- **Кампания узнаваемости** — *Content Authenticity Initiative* ведёт информационную работу с медиа, образовательными учреждениями и регуляторами, продвигая «CR» как универсальный знак «у этого файла есть проверяемое происхождение».

— МАСШТАБЫ ПРИМЕНЕНИЯ

К 2026 г. Content Credentials используется:

КАТЕГОРИЯ	ПРИМЕРЫ
Создатели и редакторы	Adobe Photoshop, Lightroom, Premiere Pro, Firefly; Adobe Content Authenticity (beta)
Генеративные ИИ-сервисы	OpenAI (DALL·E 3, Sora); Microsoft Bing Image Creator; Google Imagen, Nano Banana Pro (Gemini 3 Pro Image), Nano Banana 2 (Gemini 3.1 Flash Image)
Камеры	Leica M11-P, SL3-S; Sony Alpha 1 II, Alpha 9 III, FX3, FX30, PXW-Z300; Canon EOS R1, R5 Mark II; Nikon Z6 III
Смартфоны	Google Pixel 10 (подпись каждого снимка); Samsung Galaxy S25 (только ИИ-обработанные)
Социальные платформы	TikTok (с мая 2024); LinkedIn; Meta (Facebook, Instagram, Threads); YouTube (расширенные ИИ-ярлыки с мая 2026 — см. § 10)
Инфраструктура	Cloudflare Images (с февраля 2025; покрытие ≈ 20% веба)
Новостные медиа	BBC News, AFP, Reuters, AP, CBC/Radio-Canada, The New York Times

При этом реальное внедрение остаётся низким — по данным Reuters Institute (*Journalism, Media, and Technology Trends and Predictions 2026*, январь 2026 г.), менее 1% новостных изображений и видео имеют C2PA-метаданные.

// 9 ЦЕПОЧКА ПОДПИСЕЙ C2PA

Условный жизненный цикл подписанного контента выглядит следующим образом:

1. **Создание.** Камера, поддерживающая C2PA (*Leica M11-P*, *Sony Alpha 1 II*, *Nikon Z6 III*, *Google Pixel 10*), подписывает каждый кадр аппаратным ключом из защищённого элемента — например, *Titan M2* у *Google Pixel*. В манифест включаются модель устройства, время, при включённом GPS — координаты. (*Samsung Galaxy S25* реализует более узкий сценарий: подписывает не каждый кадр, а только изображения, обработанные ИИ-инструментами.)
2. **Редактирование.** C2PA-совместимый редактор (Adobe Photoshop, Lightroom, Premiere Pro) добавляет в *manifest store* новый манифест. Этот манифест ссылается на родительский через `c2pa.ingredient`-assertion и описывает выполненные действия (`c2pa.actions`: `c2pa.cropped`, `c2pa.color_adjustments`, `c2pa.edited` и др.).

3. **Публикация.** Новостное агентство (Reuters, AFP) или редакция подписывает финальный манифест доверенным сертификатом издателя, выдаваемым IPTC в рамках программы Origin (наследницы *Project Origin*, § 6).
4. **Верификация.** Конечный пользователь видит значок Content Credentials и через валидатор (c2paviewer.com, Adobe Inspect, openai.com/verify или расширение Adobe Content Authenticity для Chrome/Edge) получает полную историю файла.

Базовая верификация может выполняться офлайн: валидатор разбирает JUMBF, проверяет цепочку сертификатов против Trust List, пересчитывает хэши утверждений и hard binding, рекурсивно проверяет ingredient-манифесты. Проверка отзыва сертификатов (OCSP/CRL) и обновление Trust List при этом требуют сетевого доступа либо опираются на закэшированные данные (OCSP-stapling).

△ **ВНИМАНИЕ**

Манифест — это метаданные внутри файла, а не часть самого изображения, поэтому при обычных операциях он легко теряется: при пересохранении в формат без поддержки JUMBF или при загрузке на платформы, которые перекодируют изображения (WhatsApp, iMessage, часть механизмов TikTok). Помешать этому стандарт не может.

// 10 АВТОМАТИЧЕСКОЕ ОБНАРУЖЕНИЕ YOUTUBE

В мае 2026 г. YouTube объявил[13] о расширении системы маркировки ИИ-контента, перейдя от декларативной модели (ранее раскрытие было отдано на волю автора в *YouTube Studio*) к автоматическому обнаружению с помощью чтения C2PA:

- контент создан в собственных ИИ-инструментах YouTube (*Veo*, *Dream Screen*);
- контент несёт C2PA-метаданные с пометкой *fully generative AI*.

▷ **ПРИМЕР**

Изображение или видео, сгенерированное *Nano Banana Pro*, *Imagen* или *Sora* со встроенным C2PA-манифестом (§ 7), при загрузке на YouTube получает ИИ-ярлык, который невозможно снять. Содержимое манифеста интерпретируется платформой как утверждение о «полностью генеративном» характере контента. Это иллюстрация того, как Content Credentials выходят за рамки роли «индикатора прозрачности» и превращаются в «триггер» применения политики платформы в отношении сгенерированного контента.

// 11 HARD BINDING И SOFT BINDING

В спецификации C2PA различаются два режима привязки манифеста к содержимому файла:

- **Hard binding** — криптографическая привязка через хэш покрытых байт файла (как правило, SHA-256, но алгоритм берётся из реестра хэш-функций C2PA и жёстко не

зафиксирован). Любое изменение бит, включая повторное сжатие JPEG, ломает hard binding и делает манифест невалидным относительно текущего файла.

- **Soft binding** — мягкая привязка через перцептивный хэш (цифровой «отпечаток» по внешнему виду картинки, устойчивый к сжатию, перекодированию и изменению размера) или невидимый цифровой водяной знак. Главный такой водяной знак — *SynthID* от Google, которому посвящён следующий раздел. Soft binding позволяет восстановить связь с манифестом после трансформаций, разрушивших hard-привязку (перекодирование соцсетью, скриншот, изменение размера). На базе soft binding строятся **Durable Content Credentials** — концепция «выживающего» происхождения.

// 12 SYNTHID ОТ GOOGLE DEEPMIND

Как раз такой механизм — **SynthID-Image** от Google DeepMind, описанный в техническом отчёте Gowal et al. (октябрь 2025 г.)[14]. В отличие от манифеста, который можно удалить из файла[15], водяной знак встроен в сами пиксели. Авторы прямо формулируют взаимодополняемость двух подходов:

> «Хотя C2PA устанавливает стандарт кодирования и проверки происхождения с помощью метаданных, прикреплённых к контенту, такой подход уязвим к удалению этих метаданных. Нанесение водяных знаков, напротив, — перспективный механизм ответственного раскрытия, позволяющий встраивать сведения о происхождении непосредственно в сам сгенерированный контент».
>
> *While C2PA establishes a standard for encoding and verifying provenance through metadata that is attached to the content, this approach is vulnerable to removing said metadata. Watermarking is a promising responsible disclosure mechanism that allows to embed provenance information within the generated content directly.*

— АРХИТЕКТУРА

SynthID-Image наносит водяной знак уже после генерации, поверх готового изображения. В основе — две совместно обученные нейросети: одна встраивает в картинку метку ниже порога заметности для глаза, другая по проверяемому файлу определяет, есть ли в нём эта метка. Чтобы метка не разрушалась при обычных операциях, обе сети с самого начала обучают на наборе искажений — сжатии, кадрировании, изменении размера, поворотах, шуме, цветокоррекции, скриншоте.

У такого подхода три ключевых свойства:

- **наносится после генерации** (*post-hoc*) — метка добавляется к уже готовому изображению и не зависит от устройства самой генеративной модели;

- **не зависит от модели** (*model-independent*) — один и тот же механизм применяется ко всем генеративным моделям Google: *Imagen*, *Veo*, семейство *Nano Banana*. Маркировка единообразна, какая бы модель ни породила контент;
- **работает без оригинала** (*blind*) — чтобы найти метку, детектору не нужна исходная, неотмеченная версия картинки; решение принимается только по проверяемому файлу.

От метки требуют четырёх свойств сразу: быть незаметной, уверенно обнаруживаться даже после искажений, сопротивляться попыткам её удалить или подделать и быстро наноситься и считываться. Эти требования тянут в разные стороны — например, чем заметнее метка, тем она устойчивее, — поэтому полностью снять компромисс невозможно.

— МАСШТАБЫ ПРИМЕНЕНИЯ

К октябрю 2025 г. SynthID-Image использовался для маркировки более 10 млрд изображений и видеокадров в продуктах Google. Внешний вариант **SynthID-O**, доступный через партнёрства (в т.ч. NVIDIA Cosmos), в публичном бенчмарке (рис. 1 препринта Gowal et al.) демонстрирует SOTA-компромисс «качество × устойчивость» среди post-hoc-методов: превосходит *StegaStamp*, *TrustMark-Q/P*, *VideoSeal 0.0/1.0*, *WAM*, *InvisMark*.

— ПРЕДЕЛЫ УСТОЙЧИВОСТИ SYNTHID

SynthID очень устойчив: метка переживает большинство типовых операций с файлом — кадрирование, многократное пересохранение, добавление шума. Но не неуязвим — это признают и сами авторы препринта[14:1]:

> «достичь идеальной защищённости невозможно, поэтому мы сосредоточились на том, чтобы сделать ключевые атаки настолько трудными и дорогими, насколько возможно»
>
> *achieving perfect security is impossible; thus, we focused our efforts on making key attacks as difficult and expensive as possible*

Метку всё же можно удалить или подделать, но для этого нужно серьёзное усилие. Например, изображение повторно прогоняют через нейросеть, чтобы метка растворилась, либо пытаются восстановить и обойти её алгоритм. Такие атаки уже описаны в исследованиях; каждая требует отдельной работы — то есть метка осложняет подделку, но не делает её невозможной.

// 13 СВЯЗКА SYNTHID + C2PA

SynthID — это и есть реализация soft binding (§ 11). Манифест может быть срезан при пересохранении файла на платформе-перекодировщике, а пиксельный сигнал SynthID

— нет: он буквально «живёт на уровне пикселей» и выдерживает сжатие, изменение размера и фильтрацию. Через **C2PA Soft Binding API** (введён в сентябре 2024 г.) эта метка позволяет восстановить связь с манифестом, даже когда сам манифест из файла удалён.

При этом стандарт C2PA не предписывает конкретный механизм водяного знака, а лишь определяет, *как* на него ссылаться. Поэтому SynthID формально не входит в реестр алгоритмов C2PA и конкурирует с другими решениями (*TrustMark*, *Digimarc*) — по устойчивости сигнала и доступности детектора. SynthID — проприетарная разработка Google.

Связка SynthID + C2PA не случайна — это сознательная архитектурная пара:

- **C2PA-манифест** даёт *криптографически проверяемую и человекочитаемую* историю: кто подписал, что именно, когда, какие правки внесены.
- **SynthID-метка** даёт *неудаляемый при штатных операциях* сигнал «контент сгенерирован моделью Google», даже если манифест потерян.

Именно эта дуальность реализована во всех изображениях семейства *Nano Banana* (с ноября 2025 г.). Google формулирует это так:

> *coupling our state-of-the-art SynthID technology with interoperable C2PA Content Credentials*
>
> «связывая нашу передовую технологию SynthID с интероперабельными C2PA Content Credentials»

Доступ к детектору SynthID открыт через несколько сервисов Google:

- **SynthID Detector portal** — проверка изображений, аудио, видео и текста; для журналистов, медиа и исследователей по предварительной заявке;
- **Gemini app** — для всех: загрузка файла и вопрос «*Was this created with Google AI?*»;
- **Vertex AI Media Studio** — корпоративный интерфейс в Google Cloud;
- **About this image** (Google Поиск, Lens, Circle to Search) — сведения о происхождении изображения, включая флаг SynthID.

Зачем нужны C2PA и SynthID?

// 14 ЧЕТЫРЕ СЦЕНАРИЯ ПРИМЕНЕНИЯ

Мы постепенно подошли к вопросу о том, зачем же технологические гиганты развивают технические средства, позволяющие идентифицировать сгенерированный контент лучше, чем этого требует право.

Мы предлагаем четыре сценария, условно обозначенные нами как регуляторный минимум, инструмент идентификации и давления, принудительная инфраструктура, потенциальная монополия.

Сделаем предположение, что требование, предъявляемое Европейским союзом, может стать базовым в рамках правоприменения других стран. Или же технические метки, созданные для соблюдения требований EU AI Act, будут использоваться как самостоятельный инструмент в рамках других юрисдикций, пока в них нет иных требований.

// 15 РЕГУЛЯТОРНЫЙ МИНИМУМ

Самым явным вариантом создания и использования технологии C2PA будет формальное соблюдение требований EU AI Act — то есть возможность быть представленным на рынке Европейского союза и дальнейшее развитие искусственного интеллекта в рамках его правовой зоны. Технические метки в таком случае будут служить доказательством генерации объектов авторского права с помощью систем искусственного интеллекта. Это закрывает сразу два больших блока:

- разграничение естественного труда от искусственного;
- появление дополнительного доказательства отсутствия авторства при спорах в судах.

Кроме того, EU AI Act является примером наиболее оберегающего подхода к персональным данным и охране естественных прав человека. Это означает, что технологическим компаниям можно считать требования EU AI Act самыми строгими в отношении создаваемой ими технологии — а значит, использовать эти наработки для подтверждения этичности продукта в других юрисдикциях.

Но именно здесь возникает первый вопрос. Для регуляторного минимума хватило бы одного C2PA: открытая машиночитаемая метка о происхождении полностью закрывает требование прозрачности EU AI Act. Google же идёт дальше — поверх C2PA-манифеста он встраивает в каждое изображение собственный неудаляемый водяной

знак SynthID, которого закон не требует. Зачем добровольно делать больше, чем предписывает регулятор?

Первое объяснение — стремление стать условным «законодателем мод». Внедрив самую строгую и заведомо избыточную маркировку раньше других, Google задаёт фактический стандарт, под который со временем придётся подстраиваться и конкурентам, и платформам, и регуляторам других юрисдикций. Тот, кто первым выстроил инфраструктуру, потом диктует её правила. Впрочем, у этого шага могут быть и иные причины — к ним мы переходим в следующих сценариях.

// 16 ИНСТРУМЕНТ ИДЕНТИФИКАЦИИ И ДАВЛЕНИЯ

Технические метки могут использоваться технологическими компаниями для доказывания принадлежности сгенерированного объекта их моделям. Это может быть использовано для подтверждения в судах отсутствия авторского вклада в создание — что является неоспоримым плюсом для обывателя или компании, поскольку расширяет возможности защиты и создаёт ситуацию ещё большей ценности творческого труда человека.

Показательно дело № А32-40085/2024 (§ 2): ответчик настаивал, что спорные изображения сгенерированы ИИ, но проиграл, поскольку доказать это было нечем, кроме «визуального знакомства». Верифицируемая метка могла перевести спор в состязательное русло.

Но те же метки технологические компании могут использовать и для дополнительного давления на пользователей своих продуктов. Особенно это может быть чувствительно для пользователей «вселенной продуктов» — таких как Google или Meta.

Так, Google обладает одной из ведущих систем искусственного интеллекта, генерирующих различные объекты, попадающие под действие авторского права: изображения, видео, аудио. Согласно условиям Пользовательского соглашения[16], каждый пользователь должен соблюдать требования и условия ненарушения[17]. Например, к нарушениям относится использование сгенерированного объекта без автоматического водяного знака, принадлежащего Google. Этот водяной знак является визуальным подтверждением создания объекта искусственным интеллектом, а также принадлежности к конкретной модели генерации Google. Удаление водяного знака влечёт различные последствия для пользователя — от штрафов и возмещения убытков Google до отказа в доступе к продуктам сервиса[18]. В широком смысле можно лишиться доступа ко всем сервисам Google и потерять находящиеся на его серверах данные.

А что, если такое требование будет расширено и до технических меток? Ведь C2PA можно удалить. Как же сервис сможет найти таких нарушителей? Именно здесь на

сцену выходит SynthID — метка, устойчивая к штатным операциям публикации (подробнее см. § 13): Google легко отличить «свои» объекты от других.

Возможно, Google будет обязывать указывать визуальный знак, а в противном случае — ограничивать людей в доступе к сервисам. Такой подход продолжал бы продвигать результаты генерации и повышать узнаваемость среди конкурентов.

Кстати, Google уже добавил возможность[19] проверить любые данные на наличие SynthID внутри своей системы искусственного интеллекта — Gemini. Проверка происходит в свободной форме: достаточно спросить в обычном чате «Не создано ли это искусственным интеллектом?».

// 17 ПРИНУДИТЕЛЬНАЯ ИНФРАСТРУКТУРА

Реальность, в которой реклама в Интернете распространяется только с маркировкой, ещё несколько лет назад казалась нереалистичной — а сегодня это насущная обязанность каждого рекламодателя. Случай с YouTube (§ 10) — первый прецедент того, что маркировка сгенерированного контента способна повторить этот путь. C2PA-метаданные впервые работают как юридически значимый сигнал: при наличии манифеста *fully generative AI* статус контента фиксируется платформой и не снимается по воле загрузившего.

Следующий шаг можно спрогнозировать исходя из поведения регулятора (ISO 22144, EU AI Act ст. 50): контент без «надлежащего цифрового паспорта» будет невозможно использовать, а сама инфраструктура маркировки даст официальным органам дополнительные инструменты, которые вполне могут оказаться фискальными.

// 18 ПОТЕНЦИАЛЬНАЯ МОНОПОЛИЯ

Пока мы обсуждали с соавтором, почему технологические гиганты создают всё более сложные и не требуемые по законам метки, возникла одна мысль. А что, если использовать эти метки не только в рамках подтверждения или опровержения создания объекта искусственным интеллектом? Может быть, это новый инструмент продвижения.

Традиционные варианты поиска и продвижения не учитывают новую реальность: люди всё чаще не только ищут информацию через нейросети, но и получают в результате сам сгенерированный контент — изображения, тексты, видео. Встроить в такую выдачу рекламу или приоритет конкретных источников пока сложно. Но машиночитаемая метка происхождения здесь становится готовым рычагом: она позволяет заранее отличить «свой» ИИ-контент от чужого.

Только Google обладает практической технологией SynthID. Что, если Google будет показывать в приоритетном порядке изображения и другой контент, сгенерированные

именно его моделями? В таком случае получим новый виток давления от технологического гиганта и монополиста — ещё больше данных, интересных Google, и ещё менее конкурентную выдачу.

ЗАКЛЮЧЕНИЕ

Проведённый анализ показывает, что правовое регулирование объектов, сгенерированных искусственным интеллектом, неравномерно: технологии развиваются быстрее, чем общество успевает создавать рамки, а технические инструменты — быстрее, чем суды вырабатывают стандарты доказывания.

На сегодняшний день в российском и мировом праве сохраняется некое единство: автором может быть только человек, а охраноспособность определяется степенью его творческого вклада. Судебная практика показывает, что эта позиция устойчива, однако её применение остаётся ситуативным: критерий «достаточности» творческого труда до сих пор не формализован ни законодательно, ни доктринально.

EU AI Act стал первым актом, который перевёл требование прозрачности в плоскость технической обязанности. Однако наше исследование показывает, что технические инструменты идентификации несут в себе двойственный потенциал. С одной стороны, они могут стать доказательственным инструментом в судах и упростить разграничение человеческого и машинного труда. С другой — они создают инфраструктуру, которую технологические компании способны использовать для ограничения конкуренции, усиления контроля над пользователями и построения преференциальных алгоритмов выдачи.

Описанное представляется будущей проблемой, не осмысленной ни обществом, ни регуляторами. Может сложиться ситуация, при которой регуляторное требование работает в интересах тех, кого оно формально регулирует.

АВТОРЫ

Леонтьев Степан — частнопрактикующий юрист и вольный legaltech стрелок. Автор Telegram-канала «LAW_EXE» (https://t.me/law_exe).

Шматина Евдокия — инхаус-юрист в российской фармацевтической компании, в области рекламы и маркетинга и внедрения ИИ. Регулярно пишу про свои победы и поражения в Telegram-канале «Евдокия | Legal AI» (https://t.me/evdokia_arch).

1. Гражданский кодекс Российской Федерации часть 4 (ГК РФ ч.4). URL: https://www.consultant.ru/document/cons_doc_LAW_64629/. ↩

2. Постановление Пленума Верховного Суда РФ от 23.04.2019 № 10 «О применении части четвёртой Гражданского кодекса Российской Федерации». URL: https://www.consultant.ru/document/cons_doc_LAW_323470/ee07ed989eace60044215bd4b9b812f99a2f6919/. ↩

3. Постановление Девятого арбитражного апелляционного суда от 08.04.2024 № 09АП-642/2024 по делу № А40-200471/2023. URL: https://www.consultant.ru/cons/cgi/online.cgi?req=doc&base=MARB&n=2651955#8stN7LVB6uk8IeqC. ↩

4. Постановление Суда по интеллектуальным правам от 19.08.2024 № С01-1330/2024 по делу № А40-200471/2023. URL: https://www.consultant.ru/cons/cgi/online.cgi?req=doc&cacheid=410966A1D422E1E4A148FCCDBED04D06&mode=backrefs&base=SIP&n=132617&BASENODE=IiIiMSwzMjg4MCwzMjg4MCIiLCIiMiwyLEFSQiIiIg&rnd=UKMRA#QVylBLVoegAtin0t. ↩

5. Постановление Пятнадцатого арбитражного апелляционного суда от 28.11.2024 № 15АП-15937/2024 по делу № А32-40085/2024. URL: https://www.consultant.ru/cons/cgi/online.cgi?req=doc&base=RAPS015&n=279555&cacheid=C9183BD2CD9C101224721A69CB761815&mode=splus&rnd=1T20Sg#3yzT7LVCGbHVWJEg. ↩

6. Проект закона «Об основах государственного регулирования сфер применения технологий искусственного интеллекта в РФ» № 166424. URL: https://regulation.gov.ru/projects/166424/. ↩

7. Опыт Китая в регулировании ИИ: между Сциллой и Харибдой, автор Евгений Тонких, 17.12.2024 г. URL: https://russiancouncil.ru/analytics-and-comments/analytics/opyt-kitaya-v-regulirovanii-ii-mezhdu-stsilloy-i-kharibdoy/. ↩

8. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689. ↩

9. Директива 2001/29/ЕС Европейского парламента и Совета от 22 мая 2001 г. о гармонизации некоторых аспектов авторского права и смежных прав в информационном обществе (InfoSoc Directive). URL: https://www.wipo.int/wipolex/ru/legislation/details/23502. ↩

10. Директива (ЕС) 2019/790 Европейского парламента и Совета от 17 апреля 2019 г. об авторском праве и смежных правах на едином цифровом рынке (DSM Directive). URL: https://www.wipo.int/wipolex/ru/legislation/details/18927. ↩

11. Решение Суда ЕС от 16 июля 2009 г. по делу № C-5/08 Infopaq International A/S против Danske Dagblades Forening. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:62008CJ0005. ↩

12. Решение Суда ЕС от 1 декабря 2011 г. по делу № C-145/10 Eva-Maria Painer против Standard Verlags GmbH и др. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:62010CJ0145. ↩

13. YouTube Team. *Improving AI labels for viewers and creators*. blog.youtube, май 2026. URL: https://blog.youtube/news-and-events/improving-ai-labels-viewers-creators/. ↩

14. Gowal S., Bunel R., Stimberg F., Stutz D., Ortiz-Jimenez G., Kouridi C., Hayes J., Rebuffi S.-A., Bernard P., Gamble C., Horváth M.Z., Kaczmarczyck F., Kaskasoli A., Petrov A., Shumailov I., Thotakuri M., Wiles O., Yung J., Ahmed Z., Martin V., Rosen S., Savčak C., Senoner A., Vyas N., Kohli P. *SynthID-Image: Image watermarking at internet scale*. Google DeepMind, arXiv:2510.09263v1, 10 октября 2025 г. URL: https://arxiv.org/abs/2510.09263. ↩ ↩

15. Gowal et al. отмечают в сноске оригинала: «*metadata is often stripped accidentally and can also be trivially removed*». См. также § 9 настоящей статьи. ↩

16. Политика конфиденциальности и условия использования Google. URL: https://policies.google.com/terms?hl=ru#toc-content. ↩

17. Правила в отношении запрещённого использования генеративного искусственного интеллекта. URL: https://policies.google.com/terms/generative-ai/use-policy. ↩

18. Список сервисов, для которых действуют Условия использования Google, а также их дополнительные условия и правила. URL: https://policies.google.com/terms/service-specific?hl=ru. ↩

19. Справочный центр. Как проверить, были ли изображение, видео или аудио сгенерированы ИИ. URL: https://support.google.com/gemini/answer/16722517?visit_id=639157369303191306-94683198&p=verify_ai&rd=1. ↩