ZAMONAVIY HUQUQSHUNOSLIKNING AKTUAL MUAMMOLARI

XXIII-RESPUBLIKA ILMIY-AMALIY KONFERENSIYA

YAKUNLARI BO'YICHA ILMIY ISHLAR TO'PLAMI

Issue - 1(2026) Available at www.uznauka.uz

«ЗАМАСКИРОВАННОЕ НАРУШЕНИЕ АВТОРСКИХ ПРАВ» И АДВЕРСАРИАЛЬНЫЕ АТАКИ НА ОБУЧАЮЩИЕ ДАННЫЕ

Шухратжон Ёкубов
Преподаватель Сектора права
интеллектуальной собственности
Ташкентского государственного
юридического университета
E-mail: shukhrat.y2000@gmail.com

Введение

Развитие генеративного искусственного интеллекта (ИИ) поставило острые юридические вопросы, связанные с использованием охраняемых авторским правом данных при обучении моделей. Обычно под **нарушением авторских прав** в контексте обучения понимают ситуацию, когда модель получила прямой доступ к защищённому произведению и впоследствии способна создавать выходные данные, существенно похожие на оригинал. Однако новое техническое явление – **«замаскированное» (disguised) нарушение** – усложняет выявление таких случаев. Недавнее исследование Lu и др. (2024) показало, что возможно **преобразовать («замаскировать») обучающие данные**, сделав их внешне неузнаваемыми, но при этом добиться того же эффекта обучения, как если бы модель обучалась на исходном защищённом материале. Иными словами, в датасет вводятся преднамеренные искажения (adversarial disguises), благодаря которым данные выглядят совершенно иначе, но **латентные паттерны** модели формируются так же, как при обучении на оригинальном материале. Визуально или при ручном аудите такие замаскированные образцы неотличимы от обычных данных и не указывают на исходное произведение, что позволяет

злоумышленнику обойти существующие инструменты обнаружения нарушения. Это создает серьезный вызов для правоприменения: очевидные признаки копирования отсутствуют, хотя нарушение сути авторских прав налицо.

Одновременно с развитием данной проблемы растёт и значимость ИИ для государств, в том числе для Узбекистана. В 2024 году в Узбекистане утверждена стратегия развития ИИ до 2030 года, и уже реализовано более 20 проектов с применением ИИ. В условиях государственной поддержки ИИ-технологий важно обеспечить, чтобы их развитие происходило с соблюдением правовых норм, включая защиту интеллектуальной собственности. В данном тезисе, написанном с точки зрения исследователя из Узбекистана, рассматриваются правовые рамки и вызовы, связанные с **технически изощрёнными методами обхода обнаружения** авторского контента в обучающих данных. Особое внимание уделяется случаю «замаскированного» нарушения авторских прав и другим адверсариальным атакам на обучающие данные, а также анализируется, какие правовые механизмы могут противостоять таким угрозам.

Технические методы маскировки данных и атаки на датасеты

Прежде чем обсуждать правовую сторону, необходимо понять суть технических атак. **Адверсариальные атаки на обучающие данные** – это преднамеренные изменения данных, рассчитанные на то, чтобы обмануть алгоритмы машинного обучения. В нашем контексте цель злоумышленника – включить в обучающий набор сведений, несущие информацию защищённого

произведения, **таким образом, чтобы их было трудно распознать или связать с оригиналом**.

Пример 1: Замаскированные изображения. Исследование Lu и соавт. (ICML 2024) продемонстрировало возможность **генерировать специальные «маскировки» для изображений**. Берётся оригинальное изображение (например, картина или фотография, защищённая авторским правом) и автоматически преобразуется так, что итоговое изображение визуально не похоже на оригинал. Тем не менее, диффузионная модель при обучении воспринимает скрытые особенности этого изображения и **учится так, словно «видела» оригинал**. Авторы показывают, что такие замаскированные данные **невозможно отличить визуально** от безобидных, поэтому традиционный аудит обучающих выборок их не выявит. Однако внутренняя репрезентация (latent representation) модели содержит информацию оригинала. В результате генеративная модель может впоследствии выдавать результаты, существенно схожие с защищённым контентом, хотя в явном виде этот контент в обучающем наборе не фигурировал. Такой прием позволяет обойти проверки правообладателей: даже если последние просматривают датасеты или используют алгоритмы поиска похожих изображений, замаскированный образ не будет сопоставлен с оригиналом. Это **скрытое нарушение авторских прав** по существу, поскольку модель получила «косвенный» доступ к произведению. В исследовании Lu и др. предложены подходы к детектированию подобных маскировок (алгоритмы выявления и “раскрытия” скрытых образов), но пока такие инструменты только начинают разрабатываться.

Пример 2: «Ядовитый попугай» для текстов. Адверсариальные атаки возможны не только с изображениями, но и с текстовыми данными. Например, в 2025 году представлен метод **PoisonedParrot («Отравленный попугай») – первая скрытная атака отравлением данных**, которая заставляет языковую модель (LLM) выдавать фрагменты защищённого текстового контента, даже если напрямую на полном тексте она не обучалась. Суть атаки: в обучающие тексты незаметно внедряются небольшие фрагменты оригинального произведения (например, несколько предложений из книги), сгенерированные авторам через другую модель. Эти фрагменты малы и разбросаны, потому **неразличимы при обычной проверке корпуса**, а их включение выглядит как естественный текст. Тем не менее, при обучении они «приманивают» модель к запоминанию целого произведения. В итоге такая модель склонна генерировать крупные куски исходного текста по запросу – фактически нарушая авторское право. Исследования показывают, что **даже простые и малозаметные вставки способны эффективно “заразить” модель**, вызывая воспроизведение защищённого контента без видимых побочных эффектов. Более того, существующие методы защиты от подобных атак оказались малоэффективны. Авторы PoisonedParrot предложили защитный механизм ParrotTrap, но в целом данная угроза только начинает осознаваться академическим сообществом.

Таким образом, сегодня имеются различные способы **скрыть присутствие защищённых данных** в обучающем наборе – от глубоко изменённых изображений до точечных «ядом» для языковых моделей.

Общим для них является **обход человеческого восприятия и многих автоматических фильтров**: атаки намеренно эксплуатируют разрыв между тем, как данные выглядят для человека, и тем, как они интерпретируются моделью. **Модель "видит" оригинал, человек – нет.** Это создает беспрецедентную ситуацию для системы защиты авторских прав.

Правовые вызовы и анализ существующих рамок

Традиционные критерии нарушения и проблема доказательств. В авторском праве ключевое значение имеют понятия **доступа** к произведению и **существенного сходства** между оригиналом и предполагаемой копией. В случаях с ИИ-моделями уже давно обсуждается: считается ли копированием само использование произведения в обучении, и если да, то может ли оно быть оправдано исключениями (fair use, "честная практика" и т.п.). Однако при **замаскированном нарушении** вопрос еще сложнее – доступ был не прямой, а через трансформированные данные, и явного похожиего фрагмента не предъявишь. Например, если художник заподозрил, что его картина могла повлиять на результаты нейросети, он не найдет её копию в датасете – там будут лишь абстрактные узоры. **Доказать факт использования его произведения затруднительно**, нужна экспертиза поведения модели или анализ весов нейросети. По сути, правообладателю пришлось бы доказывать, что латентные признаки его работы присутствуют в модели, что выходит за рамки обычного сравнения изображений или текстов. Такое положение дел может требовать новых подходов к понятию «существенного сходства» – возможно, учитывающего схожесть воспроизведённых моделью результатов или совпадение статистических паттернов, а не только прямое сравнение

контента. Уже отмечалось, что генеративная модель способна воспроизвести охраняемые элементы даже без прямой копии в датасете, за счет общего обучения на больших данных. Тем более, если такие элементы целенаправленно внедрялись скрытно, можно говорить о **новой форме нарушения, не очевидной на уровне данных, но проявляющейся на уровне возможностей модели**.

Существующие правовые исключения: применимы ли? Второй важный аспект – даже если использование чужих данных признать копированием, защищено ли оно законом как допустимое? В мире сейчас нет единого подхода. В США многие компании ссылаются на доктрину **fair use (справедливого использования)**, утверждая, что обучение моделей – трансформативное и несет общественную пользу. Действительно, два недавних решения в Калифорнии (Bartz v. Anthropic и Kadrey v. Meta, 2025) подтвердили, что обучение крупных языковых моделей на чужих книгах может считаться «трансформативным» использованием в целях ИИ. Судьи назвали такое использование «высоко трансформативным», поскольку модель не выдает просто копии книг, а создает новое на их основе. Однако важно подчеркнуть: эти дела касались открытого, пусть и неразрешенного копирования (в одном случае с пиратских библиотек) – **без попыток спрятать следы**. Более того, один из судей (Чабрия) выразил озабоченность, что массовое использование книг для ИИ может подорвать стимулы к творчеству. Если же представить, что разработчик **умышленно маскирует данные, чтобы их не обнаружили**, такая ситуация может быть рассмотрена иначе. **Добросовестность** использования – важный неформальный фактор:

злонамеренное сокрытие улик способно сыграть против компании. Даже по американскому праву, где критерии гибкие, суд может усмотреть в подобных действиях явное нарушение (willful infringement), не прикрываемое добросовестной целью. Кроме того, отдельно незаконным актом может быть само извлечение контента с обходом мер защиты. В случае Anthropic судья заметил, что скачивание миллионов книг с пиратских сайтов не оправдано даже ради обучения модели. Аналогично, генерирование замаскированных данных можно трактовать как обход технических мер и обман аудита – что вряд ли получит снисхождение суда. В целом, **американская доктрина fair use не предусматривала таких уловок**, и их появление может побудить суды и законодателя уточнить границы допустимого обучения.

В Европейском союзе подход иной: там действуют **изъятия для текстового и дата-майнинга (TDM)**, установленные Директивой 2019/790 (DSM). Статья 3 разрешает исследовательским организациям использовать данные для научных целей, а Статья 4 – коммерческим организациям для любых целей майнинга, но с существенными условиями. К таким условиям относятся **законный доступ** к контенту (то есть данные должны быть получены без обхода защиты или нарушений) и уважение запрета правообладателя (право **opt-out**). Правообладатели могут явно запретить майнинг своих материалов, указав это в машинно-читаемой форме (например, в файле robots.txt сайта). Таким образом, в ЕС подразумевается: если контент взят из открытого источника и правообладатель не возражал, его можно временно скопировать для обучения; иначе – нет. Замаскированное включение чужих данных фактически **обходит эти нормы**.

Во-первых, если автор запретил использовать его сайт или изображения (opt-out), а разработчик всё равно их применил, просто исказив до неузнаваемости – это прямое нарушение условий TDM-изъятия. Во-вторых, **законный доступ** тоже под вопросом: маскировка часто предполагает снятие или искажение элементов, могущих быть защитными (водяных знаков, уведомлений и т.п.), что может рассматриваться как обход техсредств защиты. В целом, в ЕС подобный скрытый майнинг не будет покрываться исключениями и потребует разрешения правообладателя. Поскольку явного разрешения не получено, возникает ответственность за нарушение. Кроме того, **Digital Services Act и грядущий AI Act** в Европе уделяют внимание прозрачности и ответственности при использовании данных. Хотя прямых норм про авторское право в AI Act 2024 нет (он больше про безопасность продуктов), сама атмосфера регуляторики в ЕС такова, что скрытое включение чужого контента воспринимается негативно. Уже сейчас коммерческим майнерам в Европе часто приходится очищать данные или покупать лицензии, и выявление факта маскировки наверняка повлечет юридические санкции.

Другие юрисдикции. В ряде стран действуют аналоги fair use или TDM-оговорок (например, «fair dealing» в Канаде, узкое толкование которого пока не даёт ясности для ИИ-тренировок; отсутствие специальных норм в законе Китая при наличии требований соблюдать общие авторские права; и т.д.). Для Узбекистана, как страны, стремящейся войти в число лидеров по развитию ИИ, актуально изучение этих международных подходов. В национальном законодательстве (Закон РУз «Об авторском праве и смежных

правах» № 42 2006 г.) пока нет норм, прямо относящихся к ИИ. Авторское право охраняет произведения и их части, определяя нарушения через **копирование формы выражения** произведения без согласия правообладателя. При замаскированном использовании возникает вопрос: считается ли искажённое до неузнаваемости произведение всё еще «формой выражения» оригинала? Аргументировано можно утверждать, что да, если результирующая модель фактически содержит элементы оригинального творчества. Узбекское право, следуя Бернской конвенции, защищает произведения независимо от способа их воспроизведения, поэтому даже скрытое воспроизведение (через код, латентные образы и т.п.) может подпадать под охрану. Более того, Узбекистан поддерживает международные инициативы по защите интеллектуальной собственности и цифровой трансформации, а его новая стратегия развития ИИ подразумевает создание правовой базы для таких технологий. Следовательно, национальный регулятор рано или поздно столкнется с необходимостью адаптировать законы к вызовам, вроде адверсариальных атак на данные.

Пробелы в механизмах правоприменения. Помимо вопроса «является ли это нарушением», стоит проблема **как выявить и доказать нарушение**. Здесь право сталкивается с техническим барьером. Если замаскированное нарушение не обнаружено, правообладатель даже не узнает о факте использования его контента. Стандартные методы контроля (например, поиск своих изображений в открытых датасетах, отслеживание прямых копий текста) не сработают. Потребуются новые инструменты аудита и, возможно, обязательства по раскрытию данных. Один из путей –

нормативно обязать разработчиков ИИ **документировать источники данных** и предоставлять регуляторам или правообладателям возможность проверки. В ряде сфер такая прозрачность уже обсуждается: например, **регистры обучающих датасетов** или водяные знаки в моделях. Если бы компания была вынуждена указать, что использовала «производные» от защищённых работ, это бы упростило защиту прав. К сожалению, введение таких требований сталкивается с сопротивлением индустрии и пока не реализовано широко. Кроме того, замаскированные данные могут поступать стихийно (например, размещены третьими лицами в интернете с целью отравления публично обучаемых моделей) – в таких случаях на кого возлагать ответственность тоже вопрос. Ни одно действующее законодательство прямо не регулирует **атаки на обучающие данные**, что создает пространство для злоупотреблений. По удачному выражению исследователей, законодатель заметно **отстаёт от стремительного развития** генеративного ИИ.

Предложения по совершенствованию правовых механизмов

Учитывая вышеизложенное, назрела потребность в обновлении правовых рамок и разработке новых подходов, сочетающих технологические и юридические меры:

- **Введение понятия скрытого (латентного) копирования.** Законодателям следует рассмотреть расширение определений нарушения авторского права, включив в них случаи, когда произведение воспроизводится в изменённом математическом представлении, но сохраняет свои творческие элементы. По аналогии с понятием производного

произведения можно предложить термин вроде «производное обучающее представление» – данные, полученные на основе оригинала и способные заменить его в обучении модели. Такое понятие позволило бы явно квалифицировать замаскированные данные как объект авторско-правового регулирования.

• **Антиобходные нормы.** В законодательстве многих стран (включая Узбекистан) уже есть положения, запрещающие обход технических средств защиты авторских прав (например, DRM). Можно аргументировать, что замаскированное включение контента – это **обход мер обнаружения**, которые правообладатель разумно предпринимает. Если, к примеру, художник размещает в интернете изображения с пометкой «нельзя использовать для обучения» (robots.txt или видимый знак), а разработчик их скачал и замаскировал, – это заведомое снятие защиты. Стоит уточнить нормы, чтобы подобные действия прямо считались отягчающим фактором или отдельным нарушением. Это бы не только усилило ответственность за замаскированное копирование, но и имело превентивный эффект.

• **Обязательства по раскрытию и аудит данных.** На уровне политики предлагается требовать от разработчиков ИИ-систем **большей прозрачности** в отношении обучающих наборов. Один из вариантов – обязательная регистрация крупными моделями метаданных об источниках данных и используемых методах аугментации. Регуляторы или независимые аудиторские органы могли бы проверять модели на наличие в них следов известных защищённых произведений (благо исследования, как у Lu и др., дают и технические средства обнаружения маскировок). Такая проверка

может стать условием допуска модели на рынок, особенно для коммерческих систем. Например, будущие сертификации ИИ-моделей (по аналогии с сертификатами безопасности) могут включать и **проверку на "чистоту" обучающих данных**. Узбекистан, формируя свою нормативную базу ИИ, мог бы учесть этот момент и заложить механизмы сотрудничества между разработчиками, правообладателями и госорганами в части контроля данных.

• **Международное сотрудничество и стандарты.** Поскольку проблема носит глобальный характер, необходимо участие на уровне международных организаций (ВТО, ВОИС). Возможна разработка рекомендаций или соглашений, устанавливающих **минимальные стандарты прозрачности** и **методы противодействия адверсариальным атакам**. Например, создание единой базы данных известных случаев маскировок, обмен информацией о выявленных атаках (как это делается в кибербезопасности) и согласование терминологии (что считать нарушением в контексте ИИ). Узбекистан, участвуя в международных дискуссиях по ИИ, может внести вклад и в эту сферу, защищая интересы своих авторов и стимулируя ответственное развитие ИИ-индустрии.

• **Повышение осведомленности и обучение специалистов.** Наконец, важно, чтобы юристы, судьи, эксперты по ИИ получили представление о новых техниках. Необходимы семинары, учебные курсы по вопросам ИИ и права, где рассматривались бы случаи вроде замаскированного нарушения. Подготовка экспертизы по таким делам потребует междисциплинарных знаний – сотрудничества программистов и юристов. Узбекистан уже делает

шаги к внедрению ИИ в судебную систему и образованию, что можно расширить на сферу защиты интеллектуальной собственности в эпоху ИИ.

ЗАКЛЮЧЕНИЕ

Замаскированное нарушение авторских прав и адверсариальные атаки на данные – это новый рубеж в противостоянии между технологиями ИИ и правом интеллектуальной собственности. Если ранее споры фокусировались на том, является ли само по себе обучение на чужих данных нарушением, то теперь возникла более тонкая проблема: нарушение, **скрытое под технологическим камуфляжем**. Исследования 2024–2025 гг. однозначно подтверждают реальность этой угрозы: модели могут учиться на защищённых материалах, которые были намеренно преобразованы, сохраняя «идентичный обучающий эффект» с оригиналом. Это требует переосмысления подходов к охране авторских прав.

Для Узбекистана, активно интегрирующего ИИ в экономику и госсектор, данные вопросы имеют не только академический, но и практический интерес. Обеспечение правовых гарантий для авторов и одновременно поддержка инноваций – деликатный баланс, которого сложно достичь. Тем не менее, игнорировать проблему нельзя: риски двояки. Без надлежащей защиты пострадают правообладатели (художники, писатели, и др.), чьи работы могут тайно использоваться и обесцениваться. С другой стороны, **неясность правового статуса** таких действий создает и неопределенность для добросовестных разработчиков ИИ, которые хотят соблюдать закон.

Необходимо движение навстречу: юристы должны глубже понять технологии ИИ, а технологи – учитывать требования права при разработке методов обучения. **Юридические рамки должны быть адаптированы** – как на национальном уровне (внести изменения в законодательство об авторском праве, ИИ, цифровых технологиях), так и на международном (обмен лучшими практиками, обновление конвенций). Технические же меры (детектирование маскировок, пометки данных) станут эффективными только при поддержке нормативных требований к их использованию.

Подводя итог, **«adversarial» атаки на данные – вызов, но и стимул для развития законодательства**. Столкнувшись с новой формой скрытого копирования, правовая система обязана отреагировать инновационно. Это подтверждает известную истину: эволюция технологий всегда влечет эволюцию права. Современные исследования, подобные работе Lu et al. (2024) и другим, дают нам понимание проблемы и первые решения. Теперь задача за юридическим сообществом – встроить эти решения в регулятивную ткань. Узбекистан, стремящийся к технологическому прогрессу, имеет шанс учесть эти уроки заранее и сформировать прогрессивную правовую политику, адресующую замаскированное нарушение авторских прав в эру ИИ. Это обеспечит гармоничное развитие ИИ-отрасли, где инновации идут рука об руку с уважением к закону и творческому труду.

ИСТОЧНИКИ:

1. Lu Y. et al. Disguised Copyright Infringement of Latent Diffusion Models (ICML 2024)

2. Panaitescu-Liess M.A. et al. PoisonedParrot: Data Poisoning Attacks to Elicit Copyright-Infringing Content from LLMs (NAACL 2025)

3. Dornis T., Stober S. Generative AI Training and Copyright Law (2025)

4. Copyright Lately – Analysis of Bartz v. Anthropic & Kadrey v. Meta decisions (2025)

5. Arxiv.org – Generative AI and TDM exceptions (EU DSM Directive)

6. Grimmelmann J. et al. – On machine unlearning and copyright (2024)

7. News Central Asia – AI development strategies in Central Asia (2025)