

АЛГОРИТМИЧЕСКАЯ МЕМОРИЗАЦИЯ И ДОКТРИНА СУЩЕСТВЕННОГО СХОДСТВА В ЛАТЕНТНОМ ПРОСТРАНСТВЕ ИИ

Шухратжон Ёкубов

Преподаватель Сектора права интеллектуальной собственности
Ташкентского государственного юридического университета

E-mail: shukhrat.y2000@gmail.com

<https://doi.org/10.5281/zenodo.18333287>

Современные системы генеративного искусственного интеллекта создают беспрецедентные вызовы для традиционных правовых конструкций авторского права, особенно в контексте доктрины существенного сходства и феномена алгоритмической меморизации. **Фундаментальная проблема заключается в том, что системы ИИ хранят авторские произведения в виде математических абстракций в латентном пространстве, что принципиально не укладывается в существующие правовые рамки**, разработанные для традиционного копирования между людьми. Исследование данной проблематики с позиций узбекского правоведения выявляет критические пробелы в национальном законодательстве об авторском праве и необходимость разработки новых подходов к регулированию ИИ-технологий.

Актуальность данного исследования обусловлена растущим влиянием генеративных систем ИИ на творческие индустрии и необходимостью адаптации правовых механизмов защиты авторских прав к новым технологическим реалиям. Традиционные тесты на существенное сходство, основанные на прямых причинно-следственных связях между оригинальным произведением и предполагаемым нарушением, оказываются неэффективными при работе с системами, которые извлекают статистические паттерны из миллионов произведений и генерируют новый контент через вероятностную рекомбинацию латентных представлений.

Теоретические основы: концепция Cooper & Grimmelmann

Фундаментальный вклад в понимание проблемы алгоритмической меморизации внесли **A. Feder Cooper** и **James Grimmelmann** в своей работе 2024 года «**The Files are in the Computer: Copyright, Memorization, and Generative AI**». Их четырехкомпонентное технико-правовое определение меморизации устанавливает, что модель «запомнила» часть обучающих данных, когда из неё возможно реконструировать почти точную копию существенной части этих данных. Критически важным является понимание того, что **меморизация происходит во время обучения и представляет собой латентную способность, встроенную в параметры модели**, а не поведение, управляемое пользователем.

Концептуальная рамка Cooper & Grimmelmann разграничивает четыре связанных, но различных понятия: меморизацию как латентную способность, экстракцию как намеренное действие пользователя, регургитацию как генерацию точных копий независимо от намерений пользователя, и реконструкцию как возможность получения копий любыми средствами. Их ключевой принцип «**файлы находятся в компьютере**» утверждает, что **запомненные обучающие данные буквально существуют в модели в виде математических представлений, что делает их «копиями» в смысле авторского права.**

Данная концепция кардинально меняет понимание процесса обучения ИИ, показывая, что модели не просто изучают статистические паттерны, но и сохраняют конкретные авторские выражения в распределённом виде через параметры нейронной сети. Это создаёт фундаментальную проблему для правовых систем, которые традиционно различают идеи (не подлежащие защите) и выражения (подлежащие защите), поскольку латентные представления ИИ занимают промежуточное положение между этими категориями.

Несостоятельность традиционных правовых тестов в контексте ИИ

Анализ судебной практики в различных юрисдикциях демонстрирует **фундаментальную несовместимость между традиционными тестами на существенное сходство и техническими особенностями работы систем ИИ**. Дело Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc. (2025) стало первым крупным судебным решением, отвергающим защиту добросовестного использования в контексте обучения ИИ. Судья Stephanos Bibas установил, что использование компанией Ross Intelligence авторских аннотаций Westlaw для обучения конкурирующего ИИ-инструмента правового поиска не является добросовестным использованием, что сигнализирует о значительных рисках для компаний, использующих авторские данные без разрешения.

Дело Andersen v. Stability AI Ltd. иллюстрирует сложности применения традиционных тестов к ИИ-генерированному контенту. **Судья William Orrick отметил, что «обычные дела об авторском праве» не подходят из-за уникальной природы ИИ**. Традиционные тесты предполагают прямые отношения копирования, но обучение ИИ включает сжатие в латентные пространственные представления, что создаёт принципиально иной механизм использования авторских произведений.

Тест абстракции-фильтрации-сравнения, изначально разработанный для программного обеспечения, терпит неудачу при применении к ИИ, поскольку латентные представления являются математическими абстракциями, а не программными компонентами. Этап фильтрации удаляет именно то, что изучает ИИ – паттерны, стили и концепции. Тест обычного наблюдателя оказывается проблематичным, поскольку ИИ-системы могут воспроизводить «стиль» или «ощущение» без копирования конкретных выразительных элементов, что создаёт путаницу между подражанием стилю и нарушением авторских прав.

Правовед Benjamin J. Ard в своей влиятельной статье «Copyright's Latent Space» выявил основную проблему: **традиционные тесты предполагают отслеживаемую причинность между оригиналом и предполагаемым нарушением, но ИИ разрывает эту цепочку** через декомпозицию обучающих данных в абстрактные математические представления, извлечение статистических паттернов и вероятностную рекомбинацию при генерации новых результатов.

Технические примеры меморизации в системах ИИ

Эмпирические исследования конкретных систем ИИ предоставляют убедительные доказательства алгоритмической меморизации. **Исследование Carlini et al. (2023) «Extracting Training Data from Diffusion Models» успешно извлекло более 1000 обучающих примеров из современных диффузионных моделей, включая Stable**

Diffusion, DALL-E 2 и Imagen. Используя конвейер генерации и фильтрации, исследователи идентифицировали запомненный контент от фотографий отдельных лиц до логотипов торговых марок, доказав, что диффузионные модели «гораздо менее приватны, чем предыдущие генеративные модели».

Техническая архитектура Stable Diffusion использует модель латентной диффузии с тремя компонентами: VAE энкодер/декодер, U-Net и текстовый энкодер. VAE энкодер сжимает изображения из пиксельного пространства в латентное пространство с коэффициентом сжатия 8x8. **Меморизация происходит во время процесса деношения в латентном пространстве, где модели учатся реконструировать конкретные обучающие изображения.** Исследования показали, что модели могут восстанавливать серьёзно повреждённые изображения с аддитивным гауссовым шумом, что указывает на глубокую меморизацию.

В области языковых моделей исследование Chang et al. (2023) «Speak, Memory: An Archaeology of Books Known to ChatGPT/GPT-4» использовало тест на заполнение имён для идентификации запомненных книг. **GPT-3 воспроизвёл приблизительно 240 слов из «Гарри Поттера и философского камня» перед первой ошибкой,** а GPT-2 сгенерировал 264 строки исходного кода Bitcoin-клиента только с шестью незначительными ошибками. Эти примеры демонстрируют способность языковых моделей к дословной меморизации обширных фрагментов авторских текстов.

Huang et al. (2024) в работе «Demystifying Verbatim Memorization in Large Language Models» показали, что **меморизация «переплетена с общими способностями языковых моделей»** и полагается на распределённые состояния модели, которые кодируют высокоуровневые признаки. Абстрактные признаки в средних слоях запускают меморизацию связанного контента, что указывает на глубокую интеграцию запомненного материала в архитектуру модели.

Анализ узбекского законодательства об авторском праве

Действующее узбекское законодательство об авторском праве, основанное на Законе Республики Узбекистан «Об авторском праве и смежных правах» (ЗРУ-42 от 20 июля 2006 года), **не содержит специальных положений для регулирования ИИ-генерированного контента или использования ИИ-систем на основе авторских материалов.** Значительные поправки августа 2021 года продлили срок действия авторского права с 50 до 70 лет после смерти автора и ввели более строгие механизмы правоприменения, но не затронули вопросы искусственного интеллекта.

Фундаментальный принцип узбекского авторского права заключается в том, что **правообладателями могут быть только физические лица,** за исключением фильмов, где правообладателями могут быть юридические лица (продюсеры). Защищённые произведения должны быть «духовными творениями автора», что принципиально несовместимо с ИИ-генерированным контентом. Отсутствие явных положений создаёт значительную правовую неопределённость для разработчиков ИИ, создателей контента и коммерческих пользователей ИИ-генерированных материалов.

Критический регуляторный пробел заключается в том, что **законодательство Узбекистана не содержит явных обязательств для разработчиков ИИ избегать нарушений авторских прав при использовании данных для обучения.** Это создаёт потенциальные вопросы ответственности в рамках существующих общих положений об

авторском праве без чётких руководящих принципов для соблюдения требований. Текущий закон не определяет распределение ответственности между разработчиками ИИ, конечными пользователями систем ИИ, поставщиками платформ и поставщиками источников данных.

Сравнительный анализ с международными подходами показывает отставание узбекского законодательства. **Европейский союз принял Закон об ИИ (2024), включающий требования прозрачности для обучающих данных генеративного ИИ**, в то время как Соединённые Штаты разрабатывают доктрину добросовестного использования через судебные разбирательства. Азиатские страны, такие как Япония и Сингапур, ввели либеральные исключения для интеллектуального анализа текста и данных, позволяющие коммерческое обучение ИИ.

Рекомендации по совершенствованию правового регулирования

Для устранения выявленных пробелов узбекское законодательство требует комплексного обновления с учётом специфики ИИ-технологий. **Немедленные регуляторные действия должны включать разъяснение статуса ИИ-генерированного контента** при сохранении антропоцентрического подхода к авторскому праву, определение критериев для произведений, созданных с помощью ИИ, по сравнению с произведениями, созданными ИИ, и установление механизмов регистрации для творений, созданных с помощью ИИ.

Введение исключений для интеллектуального анализа текста и данных представляется критически важным для развития ИИ-индустрии в Узбекистане. **Необходимо создать исключения для исследований при некоммерческом обучении ИИ, установить механизмы отказа от коммерческого обучения ИИ и определить параметры добросовестного использования для алгоритмической обработки.** Эти меры позволят сбалансировать права авторов и потребности инновационного развития.

Всеобъемлющая законодательная база должна включать специальные положения об авторском праве для ИИ с определением систем ИИ и процессов обучения в контексте авторского права, установлением структур ответственности с чётким распределением обязанностей, созданием требований прозрачности для источников данных обучения ИИ и внедрением требований маркировки для ИИ-генерированного контента.

Долгосрочное стратегическое развитие должно включать регулирование латентного пространства через разработку специальной защиты для моделей и архитектур ИИ, создание лицензионных схем для передачи технологий ИИ, установление стандартов для документирования и раскрытия информации о моделях ИИ и решение вопросов производных произведений от обучения ИИ.

Заключение

Анализ проблемы алгоритмической меморизации и доктрины существенного сходства в контексте ИИ-систем выявляет фундаментальные вызовы для традиционных правовых конструкций авторского права. **Техническая реальность хранения авторских произведений в виде математических абстракций в латентном пространстве не укладывается в существующие правовые рамки**, разработанные для традиционного копирования между людьми. Концепция Cooper & Grimmelmann предоставляет важную теоретическую основу для понимания

меморизации как латентной способности, встроенной в параметры модели, что требует пересмотра традиционных подходов к авторскому праву.

Узбекское законодательство, несмотря на соответствие международным стандартам через соблюдение принципов Бернской конвенции, нуждается в специфических положениях для эры ИИ. Правовая неопределённость в отношении ИИ-генерированного контента, использования обучающих данных и распределения ответственности создаёт значительные вызовы как для инноваторов, так и для правообладателей.

Страна обладает сильным академическим лидерством через такие институты, как Ташкентский государственный юридический университет, и растущим региональным сотрудничеством через инициативы вроде CAT Law. Однако немедленные законодательные действия необходимы для устранения критических пробелов, особенно в отношении латентных пространственных представлений ИИ и исключений для обучающих данных. Узбекистан должен принять сбалансированный подход, который поддерживает сильную защиту авторских прав, одновременно обеспечивая инновации в области ИИ через тщательно проработанные исключения и чёткие структуры ответственности.

Adabiyotlar, References, Литературы:

1. Ard B.J. Copyright's Latent Space: Generative AI and the Limits of Fair Use // Cornell Law Review. - 2025. - Vol. 110. - P. 1-45.
2. Carlini N., Hayes J., Nasr M., Jagielski M., Sehwag V., Tramèr F., Borgeaud S., Ippolito D., Wallace E. Extracting Training Data from Diffusion Models // arXiv preprint arXiv:2301.13188. - 2023.
3. Chang K.K., Cramer M., Soni S., Bamman D. Speak, Memory: An Archaeology of Books Known to ChatGPT/GPT-4 // Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. - 2023. - P. 18841-18852.
4. Cooper A.F., Grimmelmann J. The Files are in the Computer: Copyright, Memorization, and Generative AI // arXiv preprint arXiv:2404.12590. - 2024.
5. Huang Z., Zhang Y., Chen Y., Zhang R. Demystifying Verbatim Memorization in Large Language Models // arXiv preprint arXiv:2407.17817. - 2024.
6. Sobel B. Elements of Style: Copyright, Similarity, and Generative AI // Houston Law Review. - 2024. - Vol. 61. - P. 1-52.
7. Вергилес Э.В. Авторское право в эпоху искусственного интеллекта: анализ работ, созданных ИИ, и проблемы авторского права в Австралии // International Property & Technology Law Journal. - 2025. - Vol. 37. - № 1. - P. 45-72.
8. Закон Республики Узбекистан «Об авторском праве и смежных правах» от 20 июля 2006 года № ЗРУ-42 (с изменениями и дополнениями от 24 августа 2021 года).
9. AI vs. Copyright: Navigating the Legal Maze of Generative AI Content Creation // Uzbekistan Law Blog. - 2024. URL: <https://uzbekistanlawblog.com/ai-vs-copyright-navigating-the-legal-maze-of-generative-ai-content-creation/>
10. Andersen v. Stability AI Ltd., No. 3:23-cv-00201 (N.D. Cal. 2023).
11. Copyright Law in Uzbekistan: Legal Framework and Application // Mondaq. - 2024. URL: <https://www.mondaq.com/copyright/1404528/copyright-law-in-uzbekistan-legal->

framework-and-application

12. Copyright Office Releases Part 2 of Artificial Intelligence Report // U.S. Copyright Office. - 2025. URL: <https://newsroom.loc.gov/news/copyright-office-releases-part-2-of-artificial-intelligence-report/>
13. EU AI Act: How Far Will EU Copyright Principles Extend? // Morgan Lewis. - 2024. URL: <https://www.morganlewis.com/pubs/2024/02/eu-ai-act-how-far-will-eu-copyright-principles-extend>
14. Legal Risks and Opportunities of Using Artificial Intelligence in Web Development: Issues of Intellectual Property and Liability // American Journal of Education and Evaluation Studies. - 2024. - Vol. 3. - № 2. - P. 45-58.
15. Scanner Darkly: Copyright Liability and Exceptions in Artificial Intelligence Inputs and Outputs // GRUR International. - 2024. - Vol. 73. - № 2. - P. 111-125.
16. Substantial Similarity // Wikipedia. URL: https://en.wikipedia.org/wiki/Substantial_similarity
17. Thomson Reuters Enterprise Centre GMBH v. Ross Intelligence Inc., No. 20-cv-613 (D. Del. 2025).
18. The Files are in the Computer: Copyright, Memorization, and Generative-AI Systems // GenLaw Blog. - 2024. URL: <https://blog.genlaw.org/explainers/files.html>