12+

НАУЧНАЯ ОБЩЕСТВЕННАЯ ОРГАНИЗАЦИЯ
ПРОФЕССИОНАЛЬНАЯ НАУКА

СБОРНИК ЭССЕ СТУДЕНТОВ И УЧАЩИХСЯ ШКОЛ

По итогам международных конкурсов эссе
от 05.08.2025, 15.08.2025

www.interclover.ru

УДК 00
ББК 00
С17

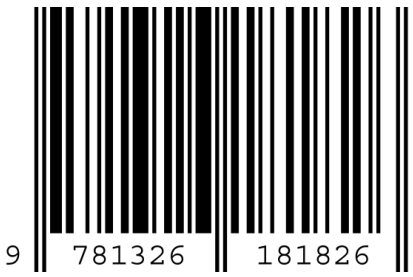
Сборник эссе студентов и учащихся школ по итогам международных конкурсов эссе от 05.08.2025, 15.08.2025/
Профессиональная наука, 2025 – 33 с.

ISBN 978-1-326-18182-6

Данная книга является сборником эссе по результатам конкурсов, проводимых НОО «Профессиональная наука» в рамках проекта Interclover.

Эта книга будет наиболее полезна для учащихся школ, студентов, магистрантов и аспирантов.

УДК 00
ББК 00



СОДЕРЖАНИЕ

Небольсина Е.М. ИИ-ассистированные атаки на приватность: Как генеративные модели (LLM, Deepfakes) переопределяют ландшафт угроз защите персональных данных и методы противодействия

Небольсина Е.М.
Студент. Факультет креативных индустрий
НИУ ВШЭ

Введение

В последние годы развитие технологий искусственного интеллекта (ИИ) и, в частности, генеративных моделей, таких как большие языковые модели (LLM) и deepfake-технологии, стало важной темой обсуждения в контексте кибербезопасности. Эти инновации открыли новые горизонты для создания контента, но одновременно с этим породили новые угрозы для приватности и безопасности персональных данных. В 2025 году эти угрозы достигли нового уровня, что требует переосмысления подходов к защите информации. Генеративные модели, благодаря своим возможностям создавать гиперреалистичный контент и персонализированные атаки, значительно усложнили задачу защиты данных. Это привело к необходимости изучения новых вызовов, которые они создают, и разработки эффективных методов противодействия. В условиях, когда технологии защиты отстают от темпов развития угроз, актуальность исследования данной темы не вызывает сомнений.

Одним из ключевых вызовов, связанных с использованием генеративных моделей, является их способность создавать синтетический контент, который практически невозможно отличить от реального. Это позволяет злоумышленникам использовать такие технологии для компрометации пользователей и организаций, создавая персонализированный фишинг, синтетические идентичности и даже поддельные доказательства.

Целью данной работы является анализ влияния генеративных моделей на ландшафт угроз приватности и разработка рекомендаций по эффективному противодействию этим угрозам. Исследование направлено на выявление ключевых направлений для улучшения защитных технологий и создания новых подходов, способных противостоять современным вызовам.

Генеративный ИИ как катализатор качественной трансформации угроз приватности (2023-2025)

Эволюционный скачок: От специализированных моделей к массово доступным мощным генеративным ИИ

В 2014 году Иэн Гудфеллоу и его коллеги представили концепцию генеративно-состязательных сетей (GAN), что стало важным шагом в области создания

синтетического контента. Эти сети функционируют на основе взаимодействия двух нейронных сетей: генератора, который создает данные, и дискриминатора, оценивающего их правдоподобие. Такая структура позволяет моделям эффективно обучаться на больших объемах данных и генерировать изображения, которые трудно отличить от реальных. Изначально GAN использовались в узкоспециализированных областях, таких как улучшение качества изображений и создание искусственных объектов для исследований. Тем не менее, их потенциал быстро привлек внимание более широкого круга разработчиков, что стало основой для перехода к универсальным моделям. Как отмечается в статье «Генеративно-состязательные сети для получения изображений высокого качества», в которой представлены модели, архитектура и сравнение работы сетей (Аверченков, Андросов, Малахов, 2020. 4 с.), GAN открывают новые возможности для создания высококачественного контента.

К 2023 году технологии генеративного ИИ, такие как GPT-3 и Stable Diffusion, стали общедоступными благодаря усилиям компаний, стремящихся демократизировать доступ к мощным инструментам. Эти модели, способные генерировать текст, изображения и даже видео, начали активно использоваться как в легальных, так и в нелегальных целях. Например, зафиксированы случаи создания фейковых новостей и изображений, которые распространялись в социальных сетях, ставя под угрозу приватность и репутацию отдельных личностей. Широкая доступность таких инструментов значительно снизила порог вхождения для злоумышленников, что привело к увеличению числа инцидентов, связанных с их использованием.Тем не менее, генеративные нейронные сети находят применение и в положительных сферах, таких как автоматический перевод и адаптация учебных материалов. Они также используются для создания альтернативных форматов представления информации, включая субтитры, аудиодескрипции и тифлокомментарии (Кутликова, Верезубова, Кишкинова, 2025. 3 с.). Таким образом, несмотря на существующие риски, связанные с использованием генеративного ИИ, его потенциал в различных областях остается значительным.

Ключевые характеристики генеративных моделей, определяющие их опасность для приватности: 1.2.1 Гиперперсонализация атак на основе анализа цифрового следа 1.2.2 Беспрецедентная масштабируемость и автоматизация злонамеренных кампаний 1.2.3 Способность преодолевать барьеры "человеческого фактора" в защите 1.2.4 Создание синтетических идентичностей и компрометирующего контента

С развитием цифровых технологий и социальных сетей пользователи оставляют значительное количество данных, известных как цифровой след. Эти данные включают информацию о предпочтениях, местоположении, контактах и личных интересах. В 2023 году исследование компании Norton показало, что 85% пользователей активно создают цифровой след, что делает их уязвимыми для атак, основанных на анализе этой информации. Генеративные модели, такие как современные языковые модели, способны анализировать огромные массивы

данных, выявлять закономерности и создавать гиперперсонализированные атаки. Злоумышленники могут использовать эти технологии для формирования убедительных сообщений, нацеленных на конкретных пользователей, что значительно увеличивает вероятность успеха таких атак.В условиях растущей угрозы Иванова подчеркивает необходимость создания Совета по данным в компаниях для разработки политики обработки данных и рассмотрения потенциальных вариантов их использования (Иванова, 2024, с. 13). Это не только поможет защитить пользователей, но и оптимизирует использование собранной информации.

Одной из ключевых характеристик генеративных моделей является их способность к автоматизации задач, которые ранее требовали значительных временных и человеческих ресурсов. Согласно отчету IBM 2023 года, использование автоматизированных инструментов на базе ИИ позволило сократить время подготовки кибератак на 60%. Это означает, что злоумышленники могут запускать масштабные кампании с минимальными затратами. Автоматизация фишинговых атак, например, позволяет генерировать тысячи уникальных сообщений, адаптированных под различные целевые группы, что значительно повышает их эффективность и охват. Такая масштабируемость представляет собой серьезную угрозу для организаций и пользователей, так как традиционные методы защиты часто не успевают адаптироваться к быстро меняющимся угрозам. В условиях растущего числа кибератак «киберстрахование становится необходимым инструментом для компаний всех отраслей» (Михеенко, 2025, с. 7).

Технологии генеративного ИИ, такие как deepfake, открывают новые возможности для создания синтетического контента, включая фальшивые изображения и видео. В 2023 году компания Sensity AI зафиксировала 30%-ный рост случаев компрометации личной и корпоративной репутации, связанный с использованием этих технологий. Злоумышленники могут создавать убедительные поддельные материалы, которые используются для шантажа, дискредитации или манипуляции общественным мнением. Генеративные модели позволяют, например, создать видео, где человек говорит или делает то, чего он на самом деле не совершал, что создает серьезные риски для приватности и безопасности.Проблема соблюдения приватности и защиты данных заключается в способности систем ИИ анализировать огромные объемы информации о пользователях. Это вызывает обеспокоенность по поводу возможного нарушения неприкосновенности частной жизни и использования персональных данных без согласия пользователей.

Качественное превосходство ИИ-ассистированных угроз над традиционными: Сравнительный анализ по ключевым параметрам (скрытность, адаптивность, ущерб, скорость)

ИИ-ассистированные угрозы, такие как фишинг с использованием языковых моделей, демонстрируют высокий уровень скрытности и адаптивности по сравнению с традиционными методами. Эти атаки, анализируя цифровой след жертвы, способны генерировать персонализированные сообщения, имитирующие стиль общения,

предпочтения и контекст взаимодействий. Это делает их сложными для обнаружения, в отличие от шаблонных атак, которые имеют предсказуемую структуру и легко распознаются защитными системами.При этом утечка корпоративной информации может происходить не только через фишинг, но и через электронные письма, копирование файлов сотрудниками и несанкционированные внедрения злоумышленников в систему. Этого можно избежать с помощью шифрования данных (Чеченец, [б. г.], 171 с.). Таким образом, комплексный подход к безопасности, включающий шифрование и обучение сотрудников, становится необходимым для защиты от современных угроз.

Скорость и масштаб атак, проводимых с помощью искусственного интеллекта, значительно превосходят традиционные методы. По данным IBM Security (2023), автоматизация атак с использованием ИИ позволила сократить время их реализации на 70%, что ведет к увеличению потенциального ущерба. Злоумышленники могут одновременно нацеливаться на множество объектов, используя минимальные ресурсы, что делает такие атаки особенно опасными. Тем не менее, с развитием киберугроз традиционные антивирусы часто оказываются недостаточными, поскольку они не способны эффективно противостоять более сложным методам, таким как атаки с использованием новых видов вирусов или программ, использующих «нулевые уязвимости» (Ермак, б. г. 2 с.).

Примером эффективности ИИ-ассистированных угроз является инцидент 2023 года, когда злоумышленники использовали технологию deepfake для создания поддельного голоса руководителя компании. Это позволило им убедить сотрудников перевести более 200 тысяч долларов на мошеннический счёт. Данный случай демонстрирует, как ИИ может быть использован для создания убедительных и сложных атак, которые трудно предотвратить традиционными методами. С другой стороны, «на третьем месте в 2024 году оказались ИТ-компании. Злоумышленники используют их в качестве «трамплина» для доступа в инфраструктуры более крупных организаций» (Источник, 2025. 7 с.). Следовательно, ИТ-компании становятся целями для атак, что создает дополнительные риски для более крупных организаций.

Реалии 2025 года: Анализ резонансных инцидентов с использованием генеративного ИИ против приватности 1.4.1 Компрометация биометрической аутентификации (Deepfake face/voice bypass) 1.4.2 Целевой бизнес-фишинг и CEO-мошенничество на базе LLM 1.4.3 Генерация синтетических личностей и компромата для шантажа 1.4.4 Автоматизированный анализ и структурирование утекших персональных данных (LLM как "аналитик зла")

В последние годы технологии Deepfake достигли такого уровня развития, что их использование стало угрозой для биометрических систем аутентификации. В 2024 году исследователи из компании Sensity AI выявили более 85 000 случаев применения deepfake-технологий для обхода систем распознавания лиц и голоса. Эти инциденты показали, что злоумышленники способны эффективно подделывать биометрические данные, создавая реалистичные изображения и аудиозаписи, которые обманывают

системы безопасности. Ярокова отмечает, что «дипфейк был признан новой угрозой для общества». Это открывает перед киберпреступниками возможности для доступа к защищенным устройствам и банковским счетам, что ставит под угрозу конфиденциальность и финансовую безопасность пользователей.

Целевой фишинг и CEO-мошенничество стали значительно более изощренными благодаря использованию крупных языковых моделей (LLM). В 2023 году компания Symantec зафиксировала рост числа атак с применением LLM на 70%. Эти модели позволили злоумышленникам создавать персонализированные фишинговые сообщения, которые выглядели настолько правдоподобно, что компании потеряли более 2 миллиардов долларов из-за таких манипуляций. Инциденты подобного рода подчеркивают, что LLM могут генерировать контент, который обманывает сотрудников и наносит значительный финансовый ущерб, подрывая доверие к внутренним коммуникациям.В контексте изменений в промышленной среде «трансформация производственной информацизации может быть представлена на рисунке 2, и сейчас можно говорить не о двух (старый и новый, киберфизический), а о трех этапах изменения промышленной среды» (Зегжда и др., 2018, с. 3). Таким образом, развитие технологий, включая LLM, оказывает влияние не только на киберугрозы, но и на более широкие аспекты производственной информацизации.

В 2025 году использование генеративного ИИ для работы с утекшими данными стало одной из ключевых угроз для приватности. Согласно отчету IBM X-Force, более 40% утечек данных сопровождались применением ИИ для анализа и структурирования информации, что значительно ускорило процесс обработки данных и их продажи на черном рынке. При этом генеративный ИИ активно использовался для создания синтетических личностей, которые применялись в различных мошеннических схемах, включая шантаж и кредитные махинации. Эти случаи подчеркивают необходимость совершенствования методов защиты данных в условиях растущего применения ИИ в киберпреступности. «Таким образом, стремительный рост использования больших данных и развитие технологий их обработки требуют более широкого обсуждения того, как компании собирают, используют и монетизируют персональные данные, и как закон идет в ногу с технологическими достижениями» (Иванова, 2024. 14 с.).

Кризис адаптации: Оценка эффективности современных систем защиты данных перед лицом генеративного ИИ

Арсенал защитника: Обзор актуальных технологий и практик защиты персональных данных (2025) 2.1.1 Традиционные решения: DLP, шифрование, IAM, осведомленность пользователей 2.1.2 Современные ИИ/ML-инструменты в защите: UEBA, NGAV, ML-based антифишинг, песочницы

Традиционные методы защиты данных продолжают оставаться важным элементом в обеспечении безопасности информации в 2025 году. Одним из таких методов являются системы предотвращения утечек данных (DLP), которые позволяют отслеживать и предотвращать несанкционированное копирование или передачу

данных. Растущая популярность DLP-систем свидетельствует об их эффективности и востребованности в современных условиях.Шифрование данных также играет ключевую роль в защите информации, обеспечивая конфиденциальность даже в случае компрометации систем. В 2022 году 71% организаций применяли шифрование для защиты своих данных, что подчеркивает его значимость в контексте киберугроз.Системы управления доступом (IAM) стали неотъемлемой частью защиты данных, предоставляя контроль доступа на основе ролей и привилегий, что минимизирует риски несанкционированного доступа. При этом контроль цепочек поставок данных в системах искусственного интеллекта и выявление искажающих данных в схемах обучения систем ИИ также играют важную роль в обеспечении безопасности.Обучение сотрудников основам кибербезопасности остается важной практикой. Исследования IBM показывают, что компании, инвестирующие в обучение, значительно снижают риск успешных атак, что подтверждает важность осведомленности пользователей в вопросах безопасности.

Современные технологии защиты данных, основанные на искусственном интеллекте и машинном обучении, открывают новые возможности для противодействия сложным угрозам. Аналитика поведения пользователей и сущностей (UEBA) активно применяется для выявления аномалий в действиях пользователей, что способствует предотвращению внутренних угроз. К 2025 году такие системы стали ключевым инструментом в арсенале кибербезопасности.Антивирусы нового поколения (NGAV), использующие машинное обучение, способны анализировать поведение программ в реальном времени, что позволяет предотвращать активацию вредоносного ПО. Кроме того, антифишинговые системы на основе машинного обучения показывают высокую точность в обнаружении фишинговых атак, что делает их незаменимыми в борьбе с этой угрозой. Песочницы, такие как Cuckoo Sandbox, обеспечивают безопасный анализ подозрительных файлов, изолируя их от основной системы и минимизируя риски заражения.Акилов и Есмаханова подчеркивают, что использование искусственного интеллекта в кибербезопасности открывает новые горизонты для прогресса и инноваций, особенно в области защиты от кибератак. Эти технологии, основанные на ИИ, представляют собой значительный шаг вперед в защите данных, обеспечивая проактивный подход к киберугрозам.

Системные ограничения существующих подходов в контексте генеративных ИИ-угроз: 2.2.1 Реактивность vs Проактивность: Отставание в обнаружении уникальных, эволюционирующих атак 2.2.2 Неэффективность против гиперперсонализированного контента и синтетических медиа 2.2.3 Уязвимость биометрических систем перед Deepfakes: Проблемы liveness-детекции 2.2.4 Истощение ресурсов и "слепота" пользователей к совершенному фишингу/Deepfakes

Современные системы защиты данных часто страдают от излишне реактивного подхода, что приводит к значительным задержкам в обнаружении и предотвращении

атак. Согласно отчету IBM X-Force 2023, среднее время обнаружения кибератаки составляет 287 дней, что указывает на недостаточную оперативность существующих решений. Эта задержка позволяет злоумышленникам не только проникнуть в систему, но и длительное время оставаться незамеченными, усугубляя последствия атак. Проблема усугубляется неспособностью систем проактивно выявлять новые угрозы. Исследование Gartner показывает, что 60% успешных атак в 2023 году произошло из-за отсутствия механизмов раннего предупреждения о новых, эволюционирующих угрозах. Это подчеркивает необходимость перехода от реактивного подхода к проактивному, который способен предвидеть и предотвращать атаки до их реализации.Системы обнаружения вторжений (IDS) предназначены для мониторинга и анализа сетевого трафика с целью выявления подозрительной активности или потенциальных атак. При этом они не блокируют атаки непосредственно (Ермак, б. г. 3 с.). Таким образом, для эффективной защиты необходимо не только выявление угроз, но и внедрение проактивных мер, которые помогут предотвратить атаки до их реализации.

Гиперперсонализированные атаки и синтетические медиа представляют собой новую волну угроз, с которой традиционные системы защиты не справляются. В 2023 году компания Deeptrace обнаружила, что 96% всех Deepfake-видео в сети были связаны с мошенничеством или киберпреступностью, что свидетельствует об их активном использовании злоумышленниками. Кроме того, согласно данным Symantec, гиперперсонализированные фишинговые атаки увеличились на 67% в 2022 году, что демонстрирует их эффективность. Эти атаки используют персональные данные жертв для создания высокоубедительного контента, что затрудняет их обнаружение и блокировку. Современные защитные инструменты часто не способны адекватно анализировать и идентифицировать такие угрозы, что делает пользователей более уязвимыми.Важно отметить, что «атака мошенников рассчитана на то, что лицо, обнаружившее накопитель, подсоединит его к компьютеру, тогда произойдет кража персональных данных» (Старостенко, Старостенко. [б. г.]. 3 с.). Это подчеркивает необходимость комплексного подхода к защите, который учитывает все аспекты киберугроз.

Биометрические системы, ранее считавшиеся надежным способом аутентификации, оказались уязвимы перед атаками с использованием Deepfake-технологий. Согласно отчету компании Sensity, в 2023 году 85% биометрических систем, основанных на распознавании лица, подверглись успешным атакам с использованием Deepfakes, что указывает на недостаточную защиту пользователей от таких угроз. Исследование MIT также выявило, что точность современных систем liveness-детекции, предназначенных для определения подлинности биометрических данных, составляет менее 75% при противодействии Deepfake-атакам. Спуфинг и дипфейки представляют собой серьезную угрозу для бизнеса. Результаты исследования показывают, что более 60% участников опроса занимают руководящие должности, что подчеркивает необходимость усовершенствования как алгоритмов

защиты, так и методов проверки подлинности биометрических данных. Это позволит более эффективно противостоять растущим угрозам в этой области.

Потенциал и барьеры использования ИИ в защитных системах (2025): 2.3.1 Текущие возможности защитного ИИ: Анализ аномалий, прогнозирование, автоматизация рутин 2.3.2 Ключевые проблемы: Нехватка данных для обучения, вычислительная стоимость, ложные срабатывания, сложность интерпретации (XAI) 2.3.3 Почему "ИИ vs ИИ" пока не панацея? Анализ фундаментальных и практических ограничений

Современные технологии искусственного интеллекта (ИИ) продемонстрировали значительный потенциал в области кибербезопасности. Одной из ключевых возможностей является анализ аномалий, который позволяет выявлять подозрительные действия в сетях и системах в режиме реального времени. Исследование, проведённое компанией IBM в 2023 году, показало, что использование ИИ может сократить время обнаружения угроз на 96%, что существенно снижает вероятность успешных атак. Кроме того, ИИ способен прогнозировать потенциальные угрозы, анализируя исторические данные и текущие тенденции, что помогает организациям принимать проактивные меры для предотвращения инцидентов.Автоматизация рутинных задач, таких как обработка журналов событий или управление доступами, позволяет специалистам сосредоточиться на более сложных аспектах безопасности, что в свою очередь увеличивает общую эффективность защитных систем. Акилов и Есмаханова отмечают, что «искусственный интеллект (ИИ) уже нашел применение и помогает решать проблемы в области кибербезопасности, делая наш мир немного безопаснее» (2024. 1 с.).

Несмотря на очевидные преимущества, использование ИИ в защитных системах сталкивается с рядом серьёзных проблем. Одной из наиболее заметных является нехватка качественных данных для обучения моделей. Без достаточного количества разнообразных данных ИИ-системы могут быть склонны к ошибкам или демонстрировать предвзятость. Ещё одной проблемой является высокая вычислительная стоимость, связанная с обучением и эксплуатацией сложных моделей ИИ, что делает их недоступными для многих организаций. Ложные срабатывания также представляют собой значительную проблему. Согласно отчёту Gartner за 2023 год, до 40% всех внедрений ИИ в области кибербезопасности сталкиваются с этой сложностью, что снижает доверие к таким системам. Наконец, сложность интерпретации решений ИИ, известная как проблема объяснимости (XAI), затрудняет понимание и принятие решений на основе рекомендаций ИИ, особенно в критически важных ситуациях.

Существуют веские причины, по которым концепция "ИИ против ИИ" не может считаться универсальным решением для защиты данных. Во-первых, системы ИИ, предназначенные для защиты, могут стать мишенью для атак, осуществляемых другими ИИ. Исследование DARPA в 2022 году показало, что гонка вооружений между атакующими и защитниками приводит к постоянному совершенствованию

методов атак, что затрудняет создание надежных защитных систем. При этом Иванов и Долгова подчеркивают, что использование ИИ в кибербезопасности может порождать новые угрозы, так как злоумышленники способны автоматизировать атаки, что значительно увеличивает их эффективность. Во-вторых, разработка и внедрение таких систем требуют значительных ресурсов, включая высококвалифицированных специалистов и мощные вычислительные мощности. Наконец, фундаментальные ограничения, такие как невозможность полного предсказания действий атакующего ИИ, делают защитные системы уязвимыми перед неожиданными угрозами. Таким образом, хотя ИИ демонстрирует потенциал, его использование в качестве универсального защитного инструмента нуждается в значительных доработках.

Дисбаланс сил: Глубинный анализ асимметрии между атакующими и защитниками в эпоху генеративного ИИ

Тактические и стратегические преимущества злоумышленников: 3.1.1 Низкий порог входа: Доступность мощных моделей и сервисов ("MaaS с ИИ") 3.1.2 Скорость и гибкость: Быстрое создание и адаптация атакующих методик 3.1.3 Эффект "первого удара": Использование уязвимостей до их обнаружения и патчинга

Развитие технологий генеративного искусственного интеллекта привело к появлению на теневых рынках сервисов "Модели как услуга" (MaaS), которые предоставляют доступ к мощным моделям, таким как ChatGPT, за минимальную плату. Это создает низкий порог входа для злоумышленников, позволяя даже лицам без технических навыков создавать сложные фишинговые письма, поддельные документы и другие инструменты для атак. В 2023 году исследователи отметили увеличение числа таких сервисов, что значительно упростило доступ к технологиям, ранее доступным только профессиональным разработчикам.Для защиты от подобных угроз необходима комплексная стратегия. Комбинация технологий — фильтрация спама, межсетевые экраны и антивирусы — в сочетании с организационными мерами, такими как обучение и тестирование персонала, эффективно защищает внешний информационный периметр компании от spear phishing (Журин, Комарков, 2018. 2 с.).

Использование генеративных моделей предоставляет злоумышленникам возможность быстро адаптироваться к изменениям в защитных системах и разрабатывать новые виды атак. В 2022 году зафиксированы случаи, когда киберпреступные группы применяли эти модели для автоматизации фишинговых кампаний, что значительно сократило время подготовки атак и повысило их эффективность. Такая скорость и гибкость дают преступникам значительное преимущество, особенно в контексте эффекта "первого удара", когда уязвимости эксплуатируются до их обнаружения и исправления защитниками.Вместе с тем стоит подчеркнуть, что «машинное обучение и нейронные сети, в частности, могут использоваться для выявления аномалий в поведении пользователей и атакующих,

обнаружения вирусов и вредоносного ПО, а также для предотвращения фишинговых атак и кражи данных» (Иванов, Долгова, б. г. 2 с.). Таким образом, несмотря на угрозы, связанные с использованием генеративных моделей, существуют и технологии, способные эффективно противостоять киберугрозам.

Системные сложности на стороне защитников: 3.2.1 Ресурсные ограничения: Высокая стоимость разработки, внедрения и поддержки защитных ИИ-систем 3.2.2 Регуляторные и этические барьеры: Ограничения на сбор данных, требования к прозрачности (XAI) 3.2.3 Организационная инерция: Сложности интеграции новых решений в существующую ИБ-инфраструктуру

Ресурсные ограничения представляют собой значительное препятствие для обеспечения эффективной защиты данных в условиях современных угроз. Исследование компании Gartner, проведенное в 2023 году, показывает, что более половины организаций сталкиваются с нехваткой квалифицированных специалистов по кибербезопасности, что ограничивает их возможности в разработке и поддержании защитных систем на основе искусственного интеллекта. Высокая стоимость утечки данных, которая в среднем составляет 4,45 миллиона долларов США по данным IBM, подчеркивает важность инвестиций в защитные технологии. Несмотря на это, компании с ограниченными бюджетами не всегда могут позволить себе внедрение таких решений, что делает их более уязвимыми к атакам. Это особенно актуально для организаций, управляющих активами, представляющими наибольший интерес для злоумышленников, поскольку они с большей вероятностью заплатят за расшифровку данных, чтобы минимизировать простои в бизнесе.

Регуляторные и этические барьеры существенно влияют на способность организаций защищать данные. С введением Общего регламента по защите данных (GDPR) в 2018 году компании столкнулись с необходимостью нести значительные затраты на адаптацию своих процессов обработки данных. Это создало особые трудности для малого и среднего бизнеса, не имеющего достаточных ресурсов для выполнения всех требований. По данным отчета Accenture, 68% компаний отметили, что необходимость соблюдения регуляторных требований замедляет внедрение инновационных технологий, включая системы защиты данных. Эти обстоятельства, наряду с организационной инерцией, затрудняют интеграцию новых решений в существующую инфраструктуру информационной безопасности, что, в свою очередь, снижает общую эффективность защиты.Современные угрозы требуют от работников наличия необходимых знаний и навыков для защиты личных и корпоративных данных. Вместе с тем, далеко не все обладают такими компетенциями (Кошелева, Зуфарова, 2025. 5 с.).

Временной фактор: Анализ лагов в цикле "атака-обнаружение-реакция-адаптация защиты"

Временные лаги в цикле "атака-обнаружение-реакция" остаются одним из ключевых факторов, определяющих эффективность защиты информационных систем. Исследование IBM показало, что среднее время обнаружения и сдерживания

кибератак в 2022 году составило 277 дней. Этот значительный временной интервал предоставляет злоумышленникам возможность нанести серьезный ущерб до того, как будут приняты защитные меры. Причины таких задержек разнообразны: это и сложность современных атак, и недостаточная интеграция систем мониторинга и анализа угроз, и ограниченные ресурсы для их оперативного обновления. В этом контексте важно отметить, что «пиковое отношение сигнал/шум (PSNR) — это обычно используемая объективная метрика для измерения качества восстановления преобразования с потерями» (Аверченков, Андросов, Малахов, 2020. 8 с.). Таким образом, временные лаги становятся критическим аспектом, который необходимо учитывать при разработке защитных мер.

Для сокращения временных лагов в цикле обнаружения и реагирования на атаки необходимо внедрение современных технологий, таких как искусственный интеллект и машинное обучение. Компании, использующие ИИ в своих системах кибербезопасности, сокращают время обнаружения угроз в среднем на 96%, согласно отчету компании Capgemini за 2021 год. Эти системы способны анализировать большие объемы данных в реальном времени, выявлять аномалии и предлагать действия для устранения угроз. Автоматизация процессов реагирования минимизирует влияние человеческого фактора, что, в свою очередь, ускоряет адаптацию защитных механизмов к новым атакам.С другой стороны, в рамках исследования были проанализированы возможности применения генеративных нейронных сетей в системе инклюзивного образования, включая создание персонализированных учебных материалов и адаптивных интерфейсов (Кутликова, Верезубова, Кишкинова, 2025. 1 с.). Таким образом, интеграция ИИ и автоматизации становится ключевым элементом в повышении эффективности киберзащиты.

Прогноз развития асимметрии: Сценарии на ближайшие 3-5 лет с учетом эволюции ИИ (AGI риски, автономные агенты)

Генеративный искусственный интеллект (ИИ) продолжает стремительно развиваться, открывая новые возможности для применения в различных отраслях. С появлением таких моделей, как GPT-4 от OpenAI, наблюдается значительное улучшение в обработке естественного языка и генерации текста. Эти достижения позволяют использовать ИИ не только для создания качественного контента, но и для автоматизации сложных задач. Тем не менее, такая функциональность может быть использована и в злонамеренных целях, например, для кибератак. Современные модели способны генерировать убедительные тексты для фишинговых писем или поддельных документов, что увеличивает их эффективность и затрудняет обнаружение.В промышленности генеративные модели находят применение для различных целей, таких как проектирование деталей с уменьшением веса конструкции и увеличением прочности (Лепешко, Марков, [б. г.]. 498 с.). Это демонстрирует, что потенциал генеративного ИИ охватывает не только области, связанные с текстом, но и инженерные и производственные процессы, подчеркивая его универсальность и значимость в современном мире.

Сценарии развития угроз с использованием генеративного ИИ предсказывают дальнейшее увеличение асимметрии между атакующими и защитниками. Согласно отчету Europol 2023, к 2026 году ожидается активное применение ИИ для создания персонализированных атак, таких как фишинг. Это усложнит обнаружение таких угроз, поскольку они будут адаптированы под конкретные цели. Защитники, в свою очередь, вынуждены разрабатывать все более сложные системы для обнаружения и предотвращения атак, что требует значительных ресурсов и времени. Увеличение количества кибератак в 3-3,5 раза с 2021 по 2023 год, а также продолжающийся тренд на рост числа атак и инцидентов в 2024 году подчеркивают необходимость усиленной защиты информации и ИТ-инфраструктуры. Будущее кибербезопасности, таким образом, будет характеризоваться постоянной гонкой технологий между атакующими и защитниками, где генеративный ИИ занимает ключевую роль.

Стратегии преодоления асимметрии: Инновационные направления защиты приватности в 2025 году и далее

Переосмысление фундаментальных принципов верификации: 4.1.1 Разработка устойчивых к Deepfakes методов биометрической аутентификации (мультимодальность, поведенческие биометрия, продвинутый liveness detection) 4.1.2 Внедрение инфраструктур доверенной аутентификации и децентрализованных идентификаторов (SSI) 4.1.3 Методы цифрового водяного знака и детектирования синтетического контента

Современные технологии Deepfake представляют собой серьезную угрозу для безопасности биометрической аутентификации, поскольку они позволяют создавать крайне реалистичные подделки изображений и видео. В ответ на эту угрозу исследователи разрабатывают новые методы, направленные на повышение устойчивости систем биометрической аутентификации. Например, в 2023 году специалисты Microsoft представили методику анализа микровыражений лица, которая позволяет определять подлинность записей на основе тонких мимических изменений, неуловимых для генеративных моделей. Эти инновации подчеркивают важность применения мультимодальных подходов, включая поведенческую биометрию и продвинутые технологии liveness detection, что делает системы аутентификации более надежными.В то же время рост спроса на облачные продукты в области информационной безопасности и развитие концепций «нулевого доверия» (Zero Trust) становятся важными трендами в сфере кибербезопасности.

Централизованные системы хранения данных часто становятся целью атак, что приводит к утечкам персональной информации. В этом контексте децентрализованные идентификаторы (SSI) представляют собой перспективное решение. Согласно отчету IBM за 2022 год, внедрение SSI позволяет сократить риск утечек данных на 35% за счет минимизации необходимости центрального хранения. Такие системы предоставляют пользователям больший контроль над своими данными, позволяя безопасно управлять идентификацией без передачи полного доступа третьим сторонам. Это создает более защищенную цифровую среду,

особенно в условиях, когда злоумышленники применяют различные методы, такие как «вишинг — телефонный фишинг, при котором звонят жертве с целью получения конфиденциальной информации, выдавая себя за сотрудника банка или службы поддержки» (Кошелева, Зуфарова, 2025, с. 5).

Синтетический контент, созданный с использованием генеративных моделей, представляет угрозу для достоверности информации в цифровой среде. Эти модели способны имитировать человеческий интеллект, что позволяет им эффективно анализировать и генерировать данные, применяясь в различных сферах, от услуг до проектирования деталей космических кораблей (Лепешко, Марков, [б. г.]. 497 с.). Для решения этой проблемы разрабатываются технологии цифрового водяного знака, которые позволяют идентифицировать искусственно созданные материалы. В 2021 году Google Research представила инструмент, интегрирующий водяные знаки в изображения, сгенерированные генеративными моделями. Это значительно упрощает процесс детектирования синтетического контента, помогая различать подлинные и поддельные данные. Такие методы становятся важной частью борьбы с распространением дезинформации.

Создание специализированных защитных ИИ-систем нового поколения: 4.2.1 Генеративные ИИ для защитников: Создание реалистичных тренировочных сред и симуляция атак 4.2.2 Системы реального времени для обнаружения и блокировки ИИ-ассистированных атак (на основе LLM для анализа контекста) 4.2.3 Развитие "криптографии приватности" (FHE, ZKP) для безопасной обработки данных ИИ

Современные технологии генеративного ИИ открывают новые возможности для создания тренировочных сред и моделирования атак, что значительно повышает эффективность подготовки специалистов в области кибербезопасности. В 2023 году компания OpenAI представила GPT-4, который используется для симуляции кибератак. Такие симуляции позволяют специалистам изучать поведение потенциальных угроз в контролируемых условиях, что способствует разработке более эффективных методов защиты. Тренировочные среды выявляют слабые места в существующих системах и проверяют устойчивость защитных мер в условиях, максимально приближенных к реальным.При этом важно учитывать, что «каждая из технологий, используемых отдельно, не позволяет эффективно защищаться от spear phishing» (Журин, Комарков, 2018. 2 с.). Это подчеркивает необходимость комплексного подхода к защите, который учитывает различные аспекты и факторы, влияющие на безопасность информационных систем.

Криптографические методы играют ключевую роль в обеспечении конфиденциальности данных при их обработке. Одним из таких методов является Fully Homomorphic Encryption (FHE), который позволяет выполнять вычисления над зашифрованными данными без необходимости их расшифровки. В 2022 году эта технология была успешно применена в облачных вычислениях, открыв новые возможности для безопасной обработки информации. FHE обеспечивает защиту

данных даже в случае компрометации вычислительной инфраструктуры, что минимизирует риски утечек. Развитие таких технологий представляет собой важный шаг к созданию безопасных систем обработки данных в эпоху активного использования искусственного интеллекта. Мещеряков, Мельников, Пересыпкин и Хорев отмечают, что «методы защиты информации систематизированы с точки зрения применения технологий искусственного интеллекта и систем в отношении задачи защиты информации» (2024. 2 с.). Это подчеркивает значимость интеграции криптографических методов в современные системы безопасности.

Совершенствование нормативно-правового и организационного контура: 4.3.1 Разработка стандартов и регуляторных требований к безопасности ИИ-моделей, используемых в обработке ПДн 4.3.2 Стимулирование ответственного раскрытия информации об уязвимостях в генеративных моделях 4.3.3 Обязательные "красные команды" (Red Teaming) с использованием ИИ для тестирования защит

Разработка стандартов и регуляторных требований к безопасности ИИ-моделей, используемых в обработке персональных данных, представляет собой ключевой шаг в обеспечении защиты приватности. Прогнозы Gartner указывают на то, что к 2025 году более 60% организаций будут применять стандартизированные подходы к управлению рисками ИИ, что свидетельствует о растущей значимости этой области. Такие стандарты устанавливают четкие критерии для оценки безопасности ИИ, что, в свою очередь, минимизирует риски утечек данных и злоупотреблений. Внедрение регуляторных требований способствует повышению доверия к технологиям, обеспечивая прозрачность их работы и ответственность разработчиков. Синча отмечает, что «комплексный подход к защите больших данных позволит минимизировать риски, повысить доверие к цифровым технологиям и создать устойчивую к угрозам информационную среду».

Обязательные меры тестирования безопасности, такие как использование "красных команд" (Red Teaming), становятся важной частью стратегии защиты ИИ-систем. Эти команды, применяя методы атакующих, помогают выявить уязвимости и слабые места в системах. В 2023 году Microsoft сообщила об успешном использовании "красных команд" для тестирования своих ИИ-систем, что позволило обнаружить и устранить более 50 потенциальных уязвимостей. Такие практики не только улучшают безопасность технологий, но и повышают их устойчивость к реальным угрозам.Сокращение «враждебного пространства данных» для систем ИИ, предотвращение мобильности атак на вычислительные модели и затруднение действий нарушителей являются ключевыми аспектами защиты. Комплексный подход к тестированию и защите ИИ-систем позволяет значительно повысить их надежность в условиях постоянно меняющихся угроз.

Международное сотрудничество и этические принципы: Необходимость глобальных инициатив по контролю за "двойным использованием" генеративного ИИ

Этические принципы играют ключевую роль в обеспечении ответственного использования генеративных технологий. Исследования 2022 года продемонстрировали, что около 60% компаний, работающих с ИИ, сталкиваются с моральными дилеммами при внедрении технологий. Это подчеркивает необходимость разработки стандартов, которые будут направлять действия организаций и предотвращать злоупотребления. В 2021 году ООН инициировала обсуждение глобальных принципов этичного использования ИИ, что свидетельствует о растущей потребности в международной координации. Многие международные организации активно занимаются вопросами регулирования ИИ. Например, ООН и ЮНЕСКО выдвинули рекомендации и принципы этического использования ИИ, направленные на защиту прав человека и предотвращение негативных социальных последствий (Саидова, Байрамова, [б. г.]. 2 с.). Эти усилия должны сосредоточиться на создании платформы для обмена опытом и лучшими практиками между странами и организациями.

Заключение

В ходе исследования было установлено, что генеративный ИИ, включая модели LLM и технологии Deepfake, значительно трансформировал природу угроз для приватности. Эти технологии позволяют злоумышленникам осуществлять атаки с высокой степенью персонализации, автоматизации и масштабируемости, что делает их особенно опасными в контексте защиты персональных данных. Текущие защитные технологии, хотя и продолжают развиваться, сталкиваются с существенными ограничениями, не позволяющими эффективно противостоять этим новым вызовам. Анализ выявил значительную асимметрию между возможностями атакующих и защитников. Злоумышленники обладают преимуществами в доступности технологий, скорости разработки атак и использовании уязвимостей до их обнаружения. В то же время, защитники сталкиваются с ограничениями в ресурсах, нормативно-правовыми барьерами и сложностями интеграции новых решений. Это подчеркивает необходимость фундаментального пересмотра подходов к защите данных.

Для преодоления выявленных проблем необходимо сосредоточиться на разработке устойчивых к Deepfake методов биометрической аутентификации, внедрении инфраструктур доверенной идентификации и создании специализированных защитных ИИ-систем. Также важно стимулировать международное сотрудничество и разработку стандартов безопасности для генеративных моделей. Эти меры помогут создать более устойчивую систему защиты персональных данных.

Дальнейшие исследования могут быть направлены на разработку новых технологий для обнаружения и предотвращения ИИ-ассистированных атак, изучение методов повышения устойчивости систем защиты к гиперперсонализированным угрозам и анализ долгосрочных последствий использования генеративного ИИ в различных сферах. Также важно оценить эффективность международных инициатив по регулированию использования ИИ.

Список литературы

1. ZHURIN S. I., KOMARKOV D. E. Protection of external information perimeter of organization from spear phishing // IT Security (Russia). — [S.l.], 2018. — v. 25, n. 4. — p. 95-107. — ISSN 2074-7136. — DOI: http://dx.doi.org/10.26583/bit.2018.4.09.

2. Аверченков А.В., Андросов А.А., Малахов Ю.А. Анализ и применение генеративно-состязательных сетей для получения изображений высокого качества // Управление в социальных и экономических системах. — 2020. — № 4. — С. 167–176. DOI: 10.30987/2658-4026-2020-4-167-176.

3. Акилов Е.К., Есмаханова Л.Н. Искусственный интеллект в мире кибербезопасности [Текст] / Е.К. Акилов, Л.Н. Есмаханова // Механика и технологии / Научный журнал. — 2024. — №3(85). — С. 465-471. — URL: https://doi.org/10.55956/LRFR4621.

4. Ермак К.К. Методы защиты от кибератак: современные подходы и технологии // [б. м.]. — [б. г.]. — [б. и.].

5. Зегжда Д.П., Васильев Ю.С., Полтавцева М.А., Кефели И.Ф., Боровков А.И. Кибербезопасность прогрессивных производственных технологий в эпоху цифровой трансформации // Вопросы кибербезопасности. — 2018. — № 2(26). — С. 2–3. DOI: 10.21681/2311-3456-2018-2-2-15.

6. Иванова А.П. Большие данные и право на неприкосновенность частной жизни (Обзорная статья) // Социальные и гуманитарные науки. Отечественная и зарубежная литература: ИАЖ. Сер. Государство и право. – 2024. – № 4. – С. 149–163. – DOI: 10.31249/iajpravo/2024.04.12.

7. Иванов А.В., Долгова О.С. Искусственный интеллект в кибербезопасности: новые угрозы и методы защиты // Научный журнал «Наука и мировоззрение». — г. Москва, Россия. — [б. г.]. — [б. м.]. — [б. и.].

8. https://ruscrypto.ru/resource/archive/rc2025/files/16_poltavtseva.pdf

9. https://jetcsirt.su/upload/%D0%90%D0%BD%D0%B0%D0%BB%D0%B8%D0%B7_%D0%BB%D0%B0%D0%BD%D0%B4%D1%88%D0%B0%D1%84%D1%82%D0%B0_%D1%83%D0%B3%D1%80%D0%BE%D0%B7_%D0%BA%D0%B8%D0%B1%D0%B5%D1%80%D0%B1%D0%B5%D0%B7%D0%BE%D0%BF%D0%B0%D1%81%D0%BD%D0%BE%D1%81%D1%82%D0%B8_2024(2).pdf

10. https://b1.ru/local/assets/surveys/russian-information-security-market-survey-2025.pdf

11. https://mts.ai/wp-content/uploads/Spufing-issledovanie.pdf

12. Кошелева А.Д., Зуфарова А.С. Обучение сотрудников как ключевой фактор защиты от киберугроз: современные вызовы и решения // ЦИТИСЭ. — 2025. — № 1. — С. 124-138.

13. Кутликова И.В., Верезубова Н.А., Кишкинова О.А. Генеративные нейронные сети и их использование в системе инклюзивного образования (на примере стран ЕС, США и Канады) // Серия: Гуманитарные науки. — 2025. — № 3-3. — С. 82–83. DOI 10.37882/2223-2982.2025.3-3.16.

14. Лепешко Р.О., Марков А.Н. Сферы использования генеративных моделей // 60-я юбилейная научная конференция аспирантов, магистрантов и студентов БГУИР. — Минск, Республика Беларусь, [б. г.]. — С. 497–498.

15. Мещеряков Р. В., Мельников С. Ю., Пересыпкин В. А., Хорев А. А. Перспективные направления применения технологий искусственного интеллекта при защите информации // Вопросы кибербезопасности. — 2024. — № 4(62). — С. 2–12. — DOI: 10.21681/2311-3456-2024-4-02-12.

16. Михеенко А. В. В поле защиты от киберрисков // КОММЕРСАНТЪ guide. — 2025. — № 6. — С. 7.

17. Саидова А.Б., Байрамова Б. Искусственный интеллект как объект международного регулирования // Научный журнал «Наука и мировоззрение». — [б. г.]. — [б. м.]. — [б. и.].

18. Синча З.И. Анализ методов устранения проблем уязвимости больших данных в информационных системах // [б. и.]. — [б. м.], [б. г.]. — [б. с.].

19. Соболь В.А., Дождикова Р.Н. Искусственный интеллект: новое решение старых проблем // [б. м.]. — [б. г.]. — [б. и.].

20. Старостенко Н.И., Старостенко О.А. Криминалистическая характеристика способов мошенничества, совершенного с использованием методов социальной инженерии // Краснодарский университет МВД России. — [б. г.]. — [б. м.]. — [б. и.].

21. Чеченец В.А. Шифрование как способ защиты корпоративной информации // 59-я научная конференция аспирантов, магистрантов и студентов. — Минск, Республика Беларусь: Учреждение образования «Белорусский государственный университет информатики и радиоэлектроники», [б. г.]. — С. 171–172.

22. Ярокова А. О. Дипфейк как угроза индивидуального медиапотребления // Курсантские исследования. — 2023. — Вып. 10. — С. 266–267.

ЭССЕ СТУДЕНТОВ И УЧАЩИХСЯ «ЧЕЛОВЕЧЕСКАЯ ЭТИКА И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: НА СТЫКЕ ТЕХНОЛОГИЙ И МОРАЛИ»

Юрченко А.А. Трансформации искусства в связи с совершенствованием искусственного интеллекта

Юрченко Анна Александровна
ГБОУ СОШ «Петришуле» в Петербурге (10 класс)

Экономическая и технологическая экспансия общества имеет влияние на «возвышенную» эстетическую сферу, как и на другие, «низменные» сферы человеческой производственной деятельности. Значение фресочной живописи для истории культуры обусловлено тем, что фреска представляет собой особый способ нанесения краски на холст. С изобретением звукового фильма жесты выразительности, характерные для актёров театра-пантомимы, исчезли из кино. Доступные нам способы создания искусства зависят от производственных возможностей общества, а любая технологическая революция провоцирует эстетические революции. Однако оптимизм или пессимизм в их отношении не может проистекать из технологически-детерминистского подхода, который бы обращался к дискретной математике или науке о построении киберфизических систем, чтобы предсказать будущее ИИ-искусства. Рассуждая о колонизации Марса, оружии массового поражения и медицинских инновациях, Ханна Арендт утверждала, что применение технических достижений в интересах человечества не может быть объектом прескрипций со стороны профессионалов точной науки (физиков, математиков), а определяется политически — то есть, в поле публичных дискурсов о морали философии. Разговор об ИИ-искусстве и его перспективах в точно таком же смысле политичен и по этой причине не может быть передан в безраздельное владение точных наук, а должен проблематизироваться гуманитарной и социальной теорией.

Появление ИИ расширило наши представления о воспроизводимости искусства. Раньше для создания репродукции, так или иначе, требовался оригинал. Критический дискурс о технологиях тиражной печати, фотографии, киносъёмки (к примеру, в работах Вальтера Беньямина) был организован вокруг оппозиции между копией и оригиналом. Состояние копии представляет проблему для теории культуры лишь в определённом отношении к аутентичности оригинала, в контексте потери присущих оригиналу свойств или трансформаций содержания оригинала, перенесённого на копию (а почти полное их отсутствие заставляло Ролана Барта выделять фотографию среди других видов искусства). Однако искусственный интеллект, продуцирующий музыку, живопись и художественные тексты, ставит

перед нами проблему существования копии без оригинала. Это соответствует философскому определению симулякра но если бодрийяровский концепт культурной единицы постмодерна является метафорическим, то «нейросеточные» звуковые дорожки, рисунки или стихи это копии без оригинала в самом буквальном смысле. Машинное обучение закладывает в алгоритм ИИ импровизационный репертуар, набор паттернов и ходов, из которых ИИ при общении с пользователем может составлять новые комбинации. Искусственный интеллект способен выдавать уникальные зримые и другие чувственно воспринимаемые формы, однако он не стремится произвести эстетическую новизну — в том смысле, в котором illusio творчества принуждает художника-авангардиста к поиску новаторских образов, — а лишь видоизменяет и рекомбинирует известные ему канонические образцы.

Эстетический режим, идентифицирующий ИИ-искусство в качестве легитимного искусства и делающий возможным критические дискурсы о нём, должен радикально пересмотреть роль того, кто созидает художественное произведение. Фигура творца в современном мире переживает метаморфозы, которые только усугубятся в результате распространения творчества ИИ. В эпоху коммодификации и гетерогенности искусства деятельность артиста не может быть безоговорочно отнесена к области сакрального, так как в производство культурного продукта вовлекаются такие профанные агенты, как коммерческий продюсер, менеджер по продажам при автономном художнике, технический редактор звука, издатель, владелец сервиса, массовая публика. Часто можно услышать, и в профессиональных культурных сообществах, и от исследователей современного искусства что истинные творцы это они: те, кто приводят в товарный вид, выпускают, продвигают, интерпретируют et cetera. Происходит повсеместное профанирование артиста. Возможно, его заключительным этапом станет превращение творца в пользователя ИИ, извлекающего нужное изображение из дигитальной модели путём введения алгоритмически верной команды.

Небольсина Е.М. Цифровые двойники и реальные 'Я': Кризис аутентичности в 2025 году

Небольсина Е.М.
Студент. Факультет креативных индустрий
НИУ ВШЭ

Введение

В последние десятилетия цифровые технологии стали неотъемлемой частью человеческой жизни, трансформируя не только способы взаимодействия, но и саму суть идентичности. В 2025 году концепция «цифровых двойников» перестала ограничиваться ролью инструментов, превратившись в активных участников формирования социальной и личной идентичности. Это привело к новой фазе кризиса аутентичности, где границы между реальным и цифровым «Я» становятся все более размытыми.

Феномен цифровых двойников: технологические и социальные аспекты

Эволюция цифровых двойников: от инструментов к агентам

Цифровые двойники, как концепция, были впервые представлены в 2002 году Майклом Гривзом и его командой. Их идея заключалась в создании цифровых моделей, отражающих физические объекты и способствующих их мониторингу и управлению. С тех пор технология претерпела значительные изменения. Изначально цифровые двойники использовались преимущественно в промышленности для повышения эффективности производства, но со временем их функции значительно расширились. К 2021 году, согласно исследованию Gartner, 13% организаций внедрили цифровых двойников для оптимизации своих процессов. Это свидетельствует о том, что цифровые двойники стали не только инструментами, но и активными участниками взаимодействий в различных сферах.В 2018 году General Electric продемонстрировала успешное применение цифровых двойников на своих промышленных объектах, что позволило снизить затраты на обслуживание на 20%. Эти примеры показывают, как цифровые двойники эволюционировали от вспомогательных инструментов до самостоятельных агентов, способных не только анализировать данные, но и принимать решения. При этом важно учитывать и этические аспекты использования цифровых двойников, особенно в контексте искусственного интеллекта. Ву Тхиен Тхюи Хиен отмечает, что «в рамках первого этического измерения можно назвать такой риск ИИ в области цифрового искусства как нарушение авторских прав» (2020, с. 2). Это подчеркивает необходимость комплексного подхода к внедрению технологий цифровых двойников, который учитывает как их преимущества, так и потенциальные риски.

Роль цифровых двойников в формировании идентичности

Цифровые двойники, такие как аватары в социальных сетях или виртуальных мирах, существенно влияют на формирование личного самовосприятия. Исследование компании McKinsey, проведенное в 2023 году, показывает, что 65% пользователей социальных сетей считают, что их цифровой образ оказывает значительное влияние на самооценку и восприятие себя. Это объясняется тем, что в цифровой среде человек может конструировать и представлять себя так, как он хочет быть воспринятым. Такой процесс может как повышать уверенность в себе, так и вызывать внутренний конфликт из-за несоответствия между реальным и желаемым образом.Психологи отмечают, что использование цифровых аватаров в метавселенных оказывает разнообразное влияние на самооценку. Если аватар отражает желаемый образ пользователя, это может привести к положительным изменениям в восприятии себя. В противном случае, несовпадение с реальностью может усугубить чувство неудовлетворенности. Один из экспертов подчеркивает, что «цифровой двойник всегда должен иметь своего реально существующего и работающего физического “родственника”. На то он и двойник» (Источник, 2025. 1 с.). Это утверждение акцентирует внимание на важности связи между цифровым образом и реальной личностью, что играет решающую роль в формировании самооценки.

Социальная идентичность значительно зависит от цифровых двойников. Исследование Pew Research Center, проведенное в 2024 году, показывает, что 74% пользователей интернета считают, что их цифровая активность влияет на восприятие их личности окружающими. Важным аспектом является феномен "социального доказательства" в цифровой среде, где мнение о человеке часто формируется на основе его цифрового присутствия, включая профили в социальных сетях и аватары. Это приводит к тому, что цифровые двойники становятся ключевым инструментом формирования социальных связей и статуса. При этом возникают вызовы для аутентичности, так как пользователи могут стремиться к созданию идеализированного образа, который не всегда отражает реальность. Цифровизация, как отмечают Иванов и Асочаков, «стала одной из популярнейших тем отечественной социологии» (2023. 1 с.), что подчеркивает необходимость глубокого исследования влияния цифровых двойников на социальную идентичность.

Алгоритмическая конструкция идентичности: вызовы и возможности

Современные алгоритмы, такие как модели искусственного интеллекта, разработанные OpenAI, включая GPT-4, демонстрируют значительный прогресс в способности имитировать человеческую речь и поведение. Эти технологии используют сложные нейронные сети и методы машинного обучения для анализа огромных объемов данных, что позволяет создавать тексты, которые кажутся аутентичными и индивидуальными. Однако этот прогресс поднимает вопросы о том, где проходит граница между искусственным интеллектом и человеческой идентичностью. Алгоритмы, способные адаптироваться к стилю и манере общения,

становятся не просто инструментами, но и активными участниками формирования цифрового образа человека.

С развитием технологий и распространением социальных сетей пользователи все чаще сталкиваются с проблемой подлинности своего образа в онлайн-среде. Исследование Pew Research Center в 2022 году показало, что 64% пользователей считают, что социальные сети способствуют созданию ложного образа личности. Это связано с тем, что алгоритмы, нацеленные на оптимизацию взаимодействия, часто подталкивают пользователей к созданию более привлекательных, но не всегда аутентичных образов. Таким образом, возникает вопрос, как сохранить подлинность в условиях, когда алгоритмы активно вмешиваются в процесс самовыражения.

Алгоритмы открывают новые возможности для самовыражения, несмотря на существующие вызовы. Платформа TikTok в 2023 году наглядно продемонстрировала, как алгоритмы способствуют созданию уникального контента и нахождению аудитории. Пользователи, опираясь на алгоритмические рекомендации, могут делиться своими увлечениями, талантами и идеями, находя единомышленников и создавая новые формы взаимодействия. Это подчеркивает, что алгоритмы, при правильном использовании, могут стать мощным инструментом для раскрытия индивидуальности и укрепления социальной связи.Вместе с тем, широкое применение ИИ в повседневной жизни может изменить образы взаимодействия между людьми, вызывать проблемы в общении и создавать новые формы зависимости от технологий. Как отмечают Голенчук и Чуешов (2023), важно учитывать как позитивные, так и негативные аспекты влияния алгоритмов на общество.

Кризис аутентичности в эпоху цифровизации

Переживание подлинности в цифровой среде

Исследование Университета Южной Калифорнии в 2022 году показало, что 34% участников испытывали кризис идентичности из-за разрыва между их реальной жизнью и цифровым образом. Это явление подчеркивает значительное влияние цифровой среды на восприятие подлинности. Виртуальные образы, созданные для социальных сетей, часто оказываются более значимыми для окружающих, чем реальные черты личности. Таким образом, цифровая среда не только изменяет представление о подлинности, но и создает новые вызовы для психологического благополучия.Синтез экзистенциальной философии и современной теории цифровых технологий формирует прочную основу для глубокого исследования философских аспектов, связанных с понятием аутентичности в контексте виртуального бытия (Зайцева, Вякин, Холодович, 2023, с. 3).

Социальные и индивидуальные последствия кризиса аутентичности

Кризис аутентичности, вызванный влиянием цифровых двойников, значительно влияет на социальные взаимодействия и индивидуальное восприятие. Исследование, проведенное в 2023 году Университетом Южной Калифорнии, выявило, что 62% пользователей социальных сетей испытывают давление

соответствовать идеализированным образом, созданным в цифровой среде. Это давление может приводить к изменениям в поведении, когда люди стремятся соответствовать ожиданиям, а не выражать свои подлинные чувства и мысли. В результате формируется культура поверхностных связей, где важность подлинности уступает место стремлению к одобрению и признанию. Данилов отмечает, что «технический прогресс оказывает большое влияние на многие сферы общественной жизни», что подчеркивает необходимость изучения воздействия цифровых технологий на наше восприятие и социальные отношения.

Стратегии адаптации и цифрового самоопределения

В условиях цифровизации, когда личность сталкивается с необходимостью адаптации к новым реалиям, вызванным влиянием цифровых двойников, важно рассмотреть подходы, способствующие сохранению аутентичности. Согласно исследованию, проведенному в 2023 году Институтом цифрового общества, более 70% пользователей интернета отмечают, что испытывают давление соответствовать своему цифровому образу, что становится источником стресса и тревоги. Это подчеркивает необходимость разработки стратегий, направленных на уменьшение этого давления. Одним из таких подходов является повышение цифровой грамотности. Европейская комиссия в 2022 году представила стратегию, ориентированную на обучение граждан критическому восприятию информации в интернете и защите своей цифровой идентичности. Это включает в себя навыки анализа контента и осознанного взаимодействия с цифровой средой, что помогает пользователям сохранять чувство подлинности в виртуальном пространстве. Кроме того, исследование Гарвардского университета 2024 года выявило, что практика осознанности и цифрового детокса значительно снижает негативное воздействие социальных сетей на самооценку и восприятие собственной идентичности. Эти данные указывают на важность внедрения методов, способствующих осознанию и дистанцированию от навязанных цифровых стандартов. В результате, такие подходы могут стать основой для формирования устойчивой и аутентичной цифровой идентичности.

Заключение

В ходе исследования было выявлено, что цифровые двойники, изначально рассматриваемые как вспомогательные инструменты, стали значимыми агентами, активно влияющими на формирование идентичности. Этот процесс сопровождается кризисом аутентичности, выражающимся в замещении внутренних убеждений и опыта внешними алгоритмическими конструкциями. Данный кризис подчеркивает необходимость переосмысления понятия подлинности и выработки новых подходов к цифровому самоопределению. Разработка стратегий адаптации к цифровому миру является ключевым направлением для преодоления кризиса аутентичности. Важным шагом в этом направлении является внедрение образовательных программ, способствующих развитию критического мышления и цифровой грамотности, а также создание платформ для обсуждения и осмысления влияния цифровых

двойников. Эти меры помогут людям лучше понимать и управлять своим взаимодействием с цифровой средой, сохраняя при этом ценности подлинности. Перспективы дальнейших исследований включают изучение возможностей интеграции цифровых технологий в общественные и личные практики без утраты человеческих ценностей. Также важно исследовать, как новые технологии могут способствовать не только адаптации, но и развитию личности в цифровом пространстве. Таким образом, комплексный подход к анализу и решению проблемы кризиса аутентичности станет основой для гармоничного сосуществования человека и цифровой среды.

Список литературы

1. Белинская Е. П., Тихомандрицкая О. А. Глобальные риски цифрового мира как актуальная проблема социальной психологии // Московский государственный университет им. М. В. Ломоносова. — 2022. — С. 25–26.
2. Велиев Д.Д. Цифровое сознание и идентичность // Studia Humanitatis. — 2024. — № 2. — [Электронный ресурс]. — URL: www.st-hum.ru.
3. Ву Тхиен Тхюи Хиен. Этические риски искусственного интеллекта в рамках цифрового искусства // [б. и.]. — 2020. — [б. с.].
4. Голенчук М.В., Чуешов В.И. О социальных последствиях использования искусственного интеллекта // 60-я юбилейная научная конференция аспирантов, магистрантов и студентов БГУИР. — Минск, Республика Беларусь, 2023. — С. 656–657.
5. Данилов Ю. Д. Цифровизация и идентичность человека // [б. м.]. — [б. г.]. — [б. и.].
6. Зайцева Н. В., Вякин Н. М., Холодович Д. А. Цифровой экзистенциализм: распутывание философских нитей виртуального бытия // Научно-практический электронный журнал Аллея Науки. — 2023. — № 10(85). — [Электронный ресурс]. — URL: Alley-science.ru.
7. Иванов Д.В., Асочаков Ю.В. Цифровизация и критическая теория общества // Социологические исследования. — 2023. — № 1. — С. 17–30. DOI: 10.31857/S013216250024389-0.
8. http://remmag.ru/upload_data/files/2019-0102/ANSYS.pdf
9. Левченко И.Е., Шуталева А.В., Керимов А.А., Путилова Е.А. Цифровые ландшафты: особенности проявления идентичности в онлайн-мире // Контекст и рефлексия: философия о мире и человеке. — 2024. — Том 13. — № 7А. — С. 96-103.
10. Лисенкова А.А. Трансформация идентичности в цифровую эпоху // Вопросы философии. — 2020. — № 3. — С. 65–74.
11. Лопатинская Т.Д. Трансформация культурной идентичности в условиях цифровой эпохи // Контекст и рефлексия: философия о мире и человеке. — 2019. — Том 8. — № 1А. — С. 126-132.

12. Поппер К. Р. Предположения и опровержения / К. Р. Поппер. — М.: Изд-во «АСТ»; «Ермак», 2004. — С. 372.

13. Рыльская А.А. Социальные сети как один из инструментов формирования и продвижения политического имиджа регионального лидера на примере губернатора Омской области А.Л. Буркова: исследовательская работа / А. А. Рыльская. — Омск, 2023. — 24 с.

14. Сорокина М.Д., Мичурова А.С., Дворянинова С.А., Емец Е.А. Влияние алгоритмов персонализации на аудиторию и приватность в медиакоммуникациях // Теории и проблемы политических исследований. — 2024. — Том 13. — № 7А. — С. 73-80.

15. Ференец Ю.А., Лойко А.И. Этика искусственного интеллекта: моральные аспекты создания и использования искусственного интеллекта и ответственность за действия автономных систем // [б. и.]. — [б. м.], [б. г.]. — [б. с.].

16. Черникова И.В., Якунина И.В. Идентичность человека в условиях цифровой трансформации // Вестник Томского государственного университета. Философия. Социология. Политология. — 2025. — № 83. — С. 103–113. — doi: 10.17223/1998863X/83/10.

Ялбаков Э. А. Алгоритмизированное правоприменение как инструмент регулирования пихофизиологического влияния искусственного интеллекта

Ялбаков Эмиль Андреевич
Студент 1 курса Юридической школы
Дальневосточного Федерального университета,
Россия, г Владивосток.

Инвазивные нейроинтерфейсы представляют собой технологические устройства, которые имплантируются непосредственно в мозговую ткань и устанавливают двустороннюю связь с нервной системой человека. Эти системы обладают уникальной способностью не только считывать и интерпретировать электрическую активность мозга, отражающую когнитивные процессы и эмоциональные состояния, но и целенаправленно стимулировать определенные нейронные сети посредством точно дозированных электрических импульсов, вызывая различные эффекты. Современные разработки в этой области объединяют последние достижения нейробиологии, нанотехнологий, искусственного интеллекта и материаловедения и обещают совершить прорыв в главной сфере человеческой жизни – мышлении, связав мозг человека и искусственный интеллект в единое целое. Эта система позволит обеспечить максимально тесную и непрерывную связь человека с искусственным интеллектом – одним из самых выдающихся и опасных изобретений XXI века, и откроет перед человечеством новые перспективы и идущие вместе с ними вызовы.

Футуролог Юваль Ной Харари называет этот симбиоз эволюцией вида Homo Sapiens в новый виток развития человека – Homo Deus [deus с лат. «Бог»]. «Сверхчеловек» или «богочеловек», подключенный своим мышлением и психикой напрямую к искусственному интеллекту и мировой базе данных, не просто будет улучшенной версией Homo Sapiens, но антропологически превзойдет человека на порядок, что в результате может разделить человечество на два разных биологических вида. Одни силой мысли будут коммуницировать со всем миром и управлять окружающими технологиями, а так же своим телом, подавляя страхи, эмоции, слабости и инстинкты; будут обладать абсолютной памятью и сверхинтеллектом, способным мгновенно решать сложнейшие задачи; смогут погружаться в мир грез и фантазий и самых заветных и несбыточных мечт, которые для людей далеких от технологий так и останутся сказкой – пока и они не решат стать частью этого нового вида «существ». Впоследствии процесс увеличения степени видового различия может привести к новому социальному устройству, где, с одной стороны, превосходством и, следовательно, властью будут обладать пользователи системы «мозг-компьютер», а с другой – к новому политическому устройству, при

котором данные эволюционные изменения и различия будут контролировать те, кто создает эти технологии.

Нейроинтерфейсы, несмотря на их значительный потенциал для расширения человеческих возможностей и лечения неврологических заболеваний, порождают комплекс серьёзных междисциплинарных проблем. В общественном сознании процесс интеграции нейротехнологий с человеческим организмом вызывает закономерные опасения, связанные с возможностью внешнего воздействия на психические процессы. Современные исследования подтверждают реальность влияния на психофизиологическое состояние человека, в то время как прямое воздействие на мыслительные процессы пока остаётся в области гипотетических предположений.

Особую тревогу вызывает возможность формирования зависимости от искусственной стимуляции мозговых центров удовольствия - своеобразного "цифрового допинга" для нервной системы. В условиях наблюдаемого роста психических расстройств, нейростимуляционные технологии могут привести к появлению новых форм аддиктивного поведения. Кроме того, нейробиологические исследования указывают на потенциальный риск атрофии тех участков мозга, функции которых постепенно перекладываются на импланты, что ставит фундаментальные вопросы о долгосрочных последствиях такой био-технической интеграции для когнитивных способностей человека.

Эти вызовы требуют комплексного подхода, включающего разработку стандартов безопасности нейроинтерфейсов, создание правовых механизмов защиты психической автономии личности, а также всестороннюю оценку социальных последствий возможного технологического расслоения общества. Важно подчеркнуть, что остановка технологического прогресса в этой области невозможна и нецелесообразна - вместо этого международное сообщество должно сосредоточиться на выработке сбалансированных регуляторных рамок. Такая работа требует консолидации усилий экспертов из различных областей: нейробиологов, специалистов по искусственному интеллекту, юристов, философов и социологов, чтобы обеспечить ответственное развитие этих трансформационных технологий.

Несмотря на комплексный характер биосоциальных, политических и правовых вызовов, связанных с развитием нейроинтерфейсов, существует лаконичное и технологически обоснованное решение — внедрение программно-правовых механизмов регулирования.

Ключевым инструментом может стать инновационное алгоритмизированное правоприменение, способное радикально оптимизировать соблюдение правовых норм в условиях массового использования нейротехнологий и ИИ. Традиционные судебные и правоохранительные системы не справятся с миллиардами потенциальных правоотношений, возникающих при взаимодействии психики человека и ИИ. Однако главное ограничение алгоритмизированного права — необходимость в машиночитаемых и объективно измеримых юридических составах

— в данном случае может преодолеться за счет уникальных свойств нейроинтерфейсов.

Электромагнитная активность мозга, регистрируемая инвазивными системами, предоставляет точные и воспроизводимые психофизиологические показатели. Несмотря на технические погрешности (например, обрастание электродов тканями), эти данные обладают достаточной объективностью, чтобы служить юридическими фактами. Их совокупность формирует полноценный юридический состав, делая алгоритмизированное правоприменение не только возможным, но и исключительно эффективным.

Предлагаемый алгоритм функционирует по принципу автоматизированной обратной связи: при обнаружении устойчивого ухудшения психофизиологического состояния пользователя во время взаимодействия с ИИ, система либо корректирует параметры стимуляции в соответствии с установленными нормами, либо инициирует защитный режим («восстановление», «покой»), а так же напоминает о чрезмерно пониженной активности зон мозга за определенный период, для недопущения потери естественного процесса мышления. Это создает прецедент превентивного права, где нормы соблюдаются не постфактум, а в реальном времени, минимизируя потенциальный вред.

Такой подход не только решает технические и правовые проблемы, но и закладывает основу для международных стандартов «этичного нейровзаимодействия», сочетающих технологические инновации с защитой фундаментальных прав человека. Данный правовой алгоритм разом решает ряд проблем и относительно прост в интеграции.

Ключевым принципом предлагаемой системы остается ориентация на автономию пользователя, что принципиально отличает данный подход от тотального контроля. Современные социокультурные тенденции демонстрируют сложную динамику взаимодействия человека с технологиями: с одной стороны — растущий спрос на цифровую детоксикацию и осознанное потребление контента, с другой — коммерциализация технологий, эксплуатирующих психологические уязвимости. В этом контексте система предлагает не жесткие ограничения, а гибкие механизмы саморегуляции, позволяющие пользователям находить баланс между технологической вовлеченностью и психологическим благополучием.

Важнейшее преимущество системы — ее адаптируемость. Хотя изначально алгоритм рассматривался для инвазивных нейроинтерфейсов, его архитектура позволяет применять принципы этичного взаимодействия и к неинвазивным технологиям. Будущее решение технических проблем, связанных с точностью считывания биоэлектрической активности у неинвазивных нейроинтерфейсов, открывает возможности для интеграции этих принципов в повседневные цифровые устройства через неинвазивные нейроинтерфейсы. Таким образом, концепция психологически безопасного взаимодействия может стать универсальным стандартом для различных технологических платформ.

Предлагаемый международный стандарт регулирования нейроинтерфейсов и ИИ создает сбалансированные рамки для их развития, обеспечивая защиту психического здоровья пользователей при сохранении технологического прогресса. Основу подхода составляет превентивная система алгоритмического контроля, отслеживающая психофизиологические показатели в реальном времени и автоматически корректирующая параметры психофизиологического влияния ИИ через нейроинтерфйес и иные устройства на основе машиночитаемого юридического состава. Эта модель сочетает технологические решения с правовыми гарантиями, применяясь как с инвазивными, так и неинвазивными нейроинтерфейсами. Важный принцип - сохранение автономии пользователя через гибкие механизмы саморегуляции, а не жесткие запреты. Реализация стандарта требует международной координации между нейробиологами, разработчиками ИИ, юристами и политиками, чтобы обеспечить ответственное развитие технологий, минимизирующее риски при максимальном использовании их потенциала для улучшения человеческих возможностей. Такой междисциплинарный подход позволяет создать устойчивые этико-правовые рамки для нейрокогнитивных технологий, ведущих нас в будущее.

Формат 60x84/16. Усл. печ. л. 1,9. Тираж 100 экз.
Издательство НОО Профессиональная наука
Нижний Новгород, ул. Горького, 4/2,
4 этаж, офис №1
Издательство Smashwords, Inc.