Классификация пользовательских цифровых рисков в условиях массового применения генеративного искусственного интеллекта

Автор: Величко Екатерина

Аффилиация: независимый исследователь в области цифровой грамотности, искусственного интеллекта и цифровой безопасности, Россия

Аннотация

В статье рассматривается классификация пользовательских цифровых рисков, возникающих в условиях массового применения генеративного искусственного интеллекта, цифровых коммуникаций, онлайн-сервисов и электронных документов. Актуальность темы обусловлена тем, что нейросетевые инструменты стали частью повседневной и профессиональной практики пользователей, однако уровень навыков проверки информации, источников и последствий цифрового действия формируется медленнее, чем скорость распространения технологий. По данным ВЦИОМ, каждый второй российский интернет-пользователь обращался к нейросетям в течение года, а наиболее распространённой целью их применения является поиск информации. Это повышает значимость рисков, связанных с некритичным доверием к ИИ-ответам и цифровым объектам.

На основе анализа открытых статистических и аналитических данных МВД России, Банка России, ВЦИОМ, ИСИЭЗ НИУ ВШЭ, Stanford AI Index, NIST AI Risk Management Framework, OWASP Top 10 for LLM Applications и International AI Safety Report 2026 предложена типология пользовательских цифровых рисков. Выделены информационные, финансовые, коммуникационные, документные, профессиональные, конфиденциальностные, репутационные и поведенческие риски. Отдельно рассматривается понятие риска цифрового действия: негативные последствия возникают не только из-за самого цифрового объекта — ссылки, сообщения, документа или ИИ-ответа, — но и из-за действия, которое пользователь совершает после взаимодействия с ним.

В статье показано, что пользовательский риск в эпоху генеративного ИИ имеет смешанную природу: техническую, информационную, психологическую и поведенческую. Делается вывод о необходимости развития механизмов предварительной проверки цифровых объектов, оценки источников, контекста, достоверности и последствий действия. Предложенная классификация может быть использована для дальнейших исследований цифровой грамотности, профилактики мошенничества, разработки обучающих методик и оценки пользовательской уязвимости в цифровой среде.

Введение

Массовое распространение генеративного искусственного интеллекта изменило не только способы создания текстов, изображений, аудио, видео и аналитических материалов, но и саму структуру пользовательского доверия в цифровой среде. Если раньше пользователь чаще взаимодействовал с цифровыми объектами, созданными человеком или организацией, то сейчас он всё чаще сталкивается с результатами, полностью или частично созданными алгоритмическими системами.

Это касается не только развлекательного контента. Нейросети используются для поиска информации, подготовки текстов, анализа документов, перевода, консультаций, деловой переписки, учебных задач и рабочих процессов. По данным ВЦИОМ, 51% российских интернет-пользователей обращались к нейросетям в течение последнего года, а наиболее популярная цель использования — поиск информации, 63%. В другом исследовании ВЦИОМ указал, что 92% интернет-активных россиян интересуются темой ИИ, а среди распространённых задач применения нейросетей названы консультации, поиск советов, создание текстов, переводы, изображения и анализ информации.

Параллельно сохраняются риски цифрового мошенничества. По данным МВД России, после роста преступлений с использованием информационно-телекоммуникационных технологий в 2023–2024 годах в 2025 году было зафиксировано снижение: таких преступлений стало на 11,8% меньше, дистанционных мошенничеств на 9% меньше, преступлений в сфере компьютерной информации на 42,2% меньше. Однако данные Банка России показывают, что в 2025 году объём операций без добровольного согласия клиентов увеличился на 6,4%, а количество таких операций на 31,2%. Следовательно, снижение отдельных уголовно-правовых показателей не отменяет сохранения пользовательских цифровых рисков.

В этих условиях возникает необходимость не только фиксировать сам факт цифрового риска, но и классифицировать его. Пользователь может столкнуться с разными типами угроз: недостоверным ИИ-ответом, фишинговой ссылкой, поддельным документом, манипулятивным сообщением, вредоносным файлом, ошибочной аналитикой, утечкой данных или синтетическим контентом. Эти угрозы различаются по источнику, механизму воздействия, типу пользовательского действия и последствиям.

Цель настоящей статьи: предложить классификацию пользовательских цифровых рисков в условиях массового применения генеративного искусственного интеллекта и цифровых коммуникаций.

Материалы и методы

Исследование выполнено в формате обзорно-аналитической работы. В качестве источников использованы открытые статистические, аналитические и методические материалы:

- данные ВЦИОМ об использовании нейросетей и восприятии ИИ-контента российскими интернет-пользователями;

- статистика МВД России по преступлениям с использованием информационно-телекоммуникационных технологий;
- обзоры Банка России по операциям без добровольного согласия клиентов финансовых организаций;
- данные ИСИЭЗ НИУ ВШЭ о распространении ИИ в российских организациях;
- Stanford AI Index 2025;
- NIST AI Risk Management Framework: Generative AI Profile;
- OWASP Top 10 for Large Language Model Applications;
- International AI Safety Report 2026;
- отчёт EBU/BBC News Integrity in AI Assistants;
- материалы FBI IC3 2025 по cyber-enabled fraud и AI-related complaints.

Методы исследования:

1. сравнительный анализ открытых статистических данных;
2. содержательный анализ пользовательских цифровых угроз;
3. классификация рисков по объекту, механизму, действию и последствиям;
4. интерпретация данных с учётом ограничений источников;
5. обобщение факторов пользовательской уязвимости.

Исследование не включает сбор персональных данных, проведение опросов, интервью, психологических экспериментов или обработку сведений об индивидуальных пользователях. Поэтому одобрение комитета по научной этике и получение информированного согласия участников не требуется.

Эмпирические основания классификации

1. Массовое использование ИИ как источник нового класса пользовательских рисков

Массовое использование нейросетей означает, что пользователь всё чаще принимает решения на основе данных, которые он получил не напрямую из первичного источника, а через генеративную систему. Это особенно важно, когда ИИ применяется для поиска информации, советов, текстов, анализа или подготовки рабочих материалов.

Данные ВЦИОМ подтверждают, что нейросети уже стали частью пользовательской практики в России: каждый второй интернет-пользователь обращался к ним в течение года, а основной целью был поиск информации. Такое использование отличается от развлекательного: пользователь начинает воспринимать ИИ как справочный и аналитический инструмент.

Отдельно важен вопрос распознавания синтетического контента. ВЦИОМ в 2026 году исследовал способность пользователей отличать ИИ-контент от созданного человеком и пришёл к выводу, что в среднем интернет-пользователи корректно распознают происхождение

контента не во всех ситуациях. Это указывает на ограниченность бытового распознавания ИИ-материалов без дополнительных навыков и инструментов проверки.

Таким образом, первый эмпирический вывод состоит в том, что пользовательский риск возникает не только в ситуациях мошенничества. Он появляется уже в момент, когда человек начинает доверять цифровому результату без понимания источника, ограничений и возможных ошибок.

2. Цифровое мошенничество и пользователь как точка воздействия

Данные МВД России показывают, что цифровые каналы стали значимой частью преступности. В 2023 году каждое третье преступление было совершено с использованием информационно-телекоммуникационных технологий, а в 2024 году доля таких преступлений достигла 40%. В 2025 году МВД зафиксировало снижение ИКТ-преступлений на 11,8%, а в начале 2026 года снижение продолжилось по ряду показателей.

Однако данные Банка России показывают иную, дополняющую картину: в 2025 году объём операций без добровольного согласия клиентов вырос на 6,4%, а количество таких операций — на 31,2%. Банк России отдельно отмечает, что злоумышленники часто используют социальную инженерию, убеждая человека перейти по фишинговой ссылке или скачать вредоносное программное обеспечение.

Эта статистика показывает важный сдвиг: пользователь становится не только жертвой технической атаки, но и участником сценария, в котором его убеждают совершить действие. С этой точки зрения цифровой риск нельзя сводить только к вредоносному сайту, вирусу или уязвимости. Существенная часть риска возникает на стыке цифрового объекта и поведения человека.

3. Использование ИИ в организациях и профессиональные риски

По данным ИСИЭЗ НИУ ВШЭ, основанным на обследованиях Росстата, технологии ИИ уже применяются в российских организациях, однако внедрение распределено неравномерно. В 2026 году ИСИЭЗ НИУ ВШЭ указал, что средняя доля организаций, использующих технологии ИИ, составляет 4,8%; среди организаций с численностью более 500 человек показатель достигает 14,9%, а среди организаций с численностью 100 и менее человек 4,1%.

В другом исследовании ИСИЭЗ НИУ ВШЭ отмечалось, что численность работников с компетенциями в области ИИ по итогам 2024 года превысила 242 тыс. человек, но составляла менее 1% численности работников списочного состава обследованных организаций. Это означает, что ИИ входит в рабочие процессы быстрее, чем формируется широкий слой специалистов, способных профессионально оценивать качество, ограничения и последствия применения ИИ-результатов.

Следовательно, пользовательские цифровые риски имеют не только бытовой, но и профессиональный характер. Ошибка ИИ-ответа или некритичное использование сгенерированного текста может попасть в отчёт, деловое письмо, коммерческое предложение, юридически значимый документ, публичную публикацию или управленческое решение.

4. Международный контекст ИИ-рисков

Международные источники подтверждают, что риски генеративного ИИ рассматриваются не как локальная проблема, а как отдельное направление управления безопасностью. Stanford AI Index 2025 фиксирует рост AI-related incidents до 233 случаев в 2024 году, что является рекордным показателем и на 56,4% больше, чем в 2023 году.

NIST в профиле управления рисками генеративного ИИ указывает, что организациям необходимо выявлять уникальные риски, создаваемые генеративными системами, и выстраивать управление такими рисками с учётом целей и приоритетов организации.

OWASP Top 10 for LLM Applications выделяет среди ключевых рисков LLM-приложений prompt injection, раскрытие чувствительной информации, уязвимости цепочки поставки, отравление данных и моделей, небезопасную обработку вывода, чрезмерную автономность, утечку системных инструкций, слабые места в векторных хранилищах и misinformation.

International AI Safety Report 2026 рассматривает риски систем общего назначения через категории возможностей, рисков и способов управления ими; в отчёте подчёркивается, что часть рисков уже материализуется, а часть остаётся неопределённой, но потенциально серьёзной.

Эти источники важны для настоящей статьи, потому что показывают: цифровой риск в эпоху ИИ не сводится только к мошенничеству. Он включает ошибки, злоупотребления, недостоверность, утечки данных, уязвимости ИИ-приложений и поведенческие последствия неправильного доверия.

Понятие пользовательского цифрового риска

В настоящей статье под пользовательским цифровым риском понимается вероятность негативных последствий, возникающих при взаимодействии пользователя с цифровым объектом, если пользователь не способен корректно оценить его источник, достоверность, безопасность, контекст и последствия дальнейшего действия.

К цифровым объектам относятся:

- ответы генеративного ИИ;
- сообщения в мессенджерах;
- электронные письма;
- ссылки;
- сайты;

- документы;
- файлы;
- изображения;
- аудио- и видеоматериалы;
- финансовые предложения;
- деловые уведомления;
- цифровые инструкции;
- формы ввода персональных или платёжных данных.

Особенность пользовательского риска состоит в том, что вред возникает не только из-за свойств самого объекта. Например, сообщение может не содержать вредоносного кода, но провоцировать пользователя на перевод денег. Ссылка может выглядеть технически безопасной на первом этапе, но вести к сценарию сбора данных. Ответ ИИ может быть грамотно написан, но содержать неподтверждённые факты. Документ может быть формально оформлен, но содержать рискованные условия или несоответствие реквизитов.

Поэтому пользовательский цифровой риск необходимо анализировать через четыре элемента:

1. **объект** — с чем взаимодействует пользователь;
2. **механизм воздействия** — как объект влияет на восприятие и решение;
3. **действие** — что пользователь делает после контакта с объектом;
4. **последствие** — какой ущерб может возникнуть.

Базовая классификация пользовательских цифровых рисков

Таблица 1. Классификация пользовательских цифровых рисков

Тип риска	Основные объекты	Механизм возникновения	Возможные последствия
Информационный	ИИ-ответ, статья, справка, аналитика	Недостоверный факт, устаревшие данные, галлюцинация, отсутствие источника	Ошибочное решение, распространение ложной информации
Финансовый	Ссылка, сообщение, счёт, сайт, инвестиционное предложение	Фишинг, социальная инженерия, поддельные реквизиты, вредоносное ПО	Потеря денег, перевод средств мошенникам

Коммуникационный	Мессенджер, письмо, звонок, голосовое сообщение	Давление, срочность, авторитет, имитация доверенного лица	Передача данных, подтверждение операции, подчинение сценарию
Документный	Договор, счёт, акт, КП, оферта	Подмена условий, реквизитов, сторон, юридически рискованные формулировки	Финансовый или юридический ущерб
Профессиональный	Отчёт, деловой текст, презентация, аналитика	Некритичное использование ИИ-результата	Ошибка в работе, управленческий ущерб, снижение качества решений
Конфиденциальностный	Документы, переписка, персональные данные	Передача чувствительных данных в цифровую систему или третьим лицам	Утечка данных, нарушение конфиденциальности
Репутационный	Публикация, экспертный текст, визуальный контент	Использование недостоверного или синтетического материала без проверки	Потеря доверия, публичная ошибка
Поведенческий	Любой цифровой объект	Решение под давлением страха, срочности, выгоды или авторитета	Ошибочное действие, потеря контроля над ситуацией

Информационные риски

Информационные риски связаны с тем, что пользователь воспринимает цифровой текст как достоверный без проверки источника. В эпоху генеративного ИИ этот риск усиливается, потому что ИИ-ответ может быть логичным, структурированным и убедительным, но при этом содержать ошибки, устаревшие данные или вымышленные источники.

Исследование EBU/BBC News Integrity in AI Assistants 2025, проведённое с участием 22 общественных медиаорганизаций в 18 странах и 14 языках, показало, что ИИ-ассистенты могут системно ошибаться при ответах на вопросы о новостях и текущих событиях. Reuters,

описывая это исследование, указывает, что 45% ответов содержали значимые ошибки, а серьёзные проблемы с источниками встречались примерно в трети ответов.

Для пользовательского риска это означает следующее: даже если ИИ-ответ выглядит качественно, он требует проверки, особенно если используется в контексте права, финансов, медицины, безопасности, образования, бизнеса или публичной коммуникации.

Информационный риск можно разделить на пять подтипов:

1. **фактический риск** — неверные даты, цифры, имена, нормативные сведения;
2. **источниковый риск** — отсутствие ссылок, ложные ссылки, неподтверждённые источники;
3. **контекстный риск** — верная информация применена к неверной ситуации;
4. **актуализационный риск** — данные устарели или изменились;
5. **интерпретационный риск** — ИИ делает вывод, который не следует из источников.

Финансовые риски

Финансовые риски связаны с ситуациями, в которых цифровой объект приводит пользователя к денежному действию: переводу, оплате, вводу данных карты, установке приложения, подтверждению операции или участию в сомнительной инвестиционной схеме.

Данные Банка России показывают, что финансовый цифровой риск сохраняет значимость. В 2025 году объём операций без добровольного согласия клиентов вырос на 6,4%, а количество таких операций на 31,2%. Банк России также отметил, что злоумышленники убеждали пользователей переходить по фишинговым ссылкам или скачивать вредоносное программное обеспечение.

Финансовые риски включают:

- фишинговые ссылки;
- поддельные сайты банков, маркетплейсов, госорганов или служб доставки;
- инвестиционные предложения с обещанием доходности;
- поддельные счета и реквизиты;
- просьбы срочно перевести деньги;
- сообщения от имени руководителя, банка, родственника или государственного органа;
- вредоносные приложения;
- схемы с кредитами, займами и псевдоброкерами.

Особенность финансового риска заключается в том, что вред часто возникает не сразу при получении сообщения, а после действия пользователя. Поэтому классификация должна учитывать не только объект, но и то, к какому действию он подталкивает.

Коммуникационные риски и социальная инженерия

Коммуникационные риски связаны с манипулятивным воздействием на пользователя через цифровую переписку, звонки, голосовые сообщения, электронные письма или уведомления. В отличие от технической атаки, социальная инженерия направлена на восприятие, эмоции и решение человека.

Банк России в обзоре за 2025 год указывает, что злоумышленники используют приёмы и методы социальной инженерии, убеждая человека совершить операцию, перейти по фишинговой ссылке или скачать вредоносное ПО.

Основные механизмы коммуникационного воздействия:

1. **срочность** — «сделайте сейчас», «иначе счёт заблокируют»;
2. **страх** — угроза штрафа, потери денег, уголовной ответственности, блокировки;
3. **авторитет** — имитация банка, полиции, руководителя, госоргана, службы безопасности;
4. **выгода** — обещание выплаты, доходности, бонуса, компенсации;
5. **секретность** — просьба никому не рассказывать;
6. **перегрузка** — много деталей, сложные инструкции, быстрый темп общения;
7. **персонализация** — обращение по имени, использование контекста, имитация знакомого стиля.

Эта группа рисков особенно важна, потому что она объясняет, почему даже технически грамотный пользователь может совершить ошибку. Решение принимается не в спокойном аналитическом режиме, а под давлением эмоций, срочности и доверия к источнику.

Документные риски

Документные риски возникают при работе с договорами, счетами, актами, коммерческими предложениями, офертами, инвестиционными материалами, деловыми письмами и файлами, которые требуют финансового или юридически значимого действия.

К документным рискам относятся:

- несоответствие реквизитов и заявленной организации;
- подмена стороны договора;
- отсутствие существенных условий;
- опасные условия предоплаты;
- гарантии доходности без обоснования;
- ссылки на несуществующие документы или лицензии;
- противоречия между письмом и вложением;
- использование внешне убедительного оформления без подтверждения источника.

Документный риск отличается от информационного тем, что он чаще связан с последующим действием: оплатой, подписанием, отправкой данных, подтверждением условий или вступлением в финансовые обязательства.

Для предпринимателей, самозанятых и малого бизнеса эта группа рисков особенно значима, поскольку у них не всегда есть постоянное юридическое, финансовое или информационно-безопасностное сопровождение. При этом ИСИЭЗ НИУ ВШЭ показывает, что использование ИИ и цифровых технологий в организациях зависит от размера организации: крупные компании чаще используют ИИ и обладают большими ресурсами для контроля, чем малые организации.

Профессиональные риски использования ИИ

Профессиональные риски связаны с использованием ИИ-результатов в рабочих процессах без достаточной проверки. Пользователь может применять ИИ для подготовки отчётов, коммерческих предложений, статей, презентаций, аналитики, договорных формулировок, клиентских сообщений или управленческих выводов.

Проблема состоит в том, что ИИ способен ускорить подготовку материала, но не снимает ответственность с человека. Если результат содержит фактическую ошибку, неподтверждённое утверждение или неверный вывод, последствия возникают уже в профессиональной среде.

Профессиональные риски включают:

- использование неподтверждённых цифр;
- некорректные ссылки на законы, исследования или статистику;
- ошибочную аналитику;
- автоматическое копирование ИИ-выводов;
- публикацию недостоверных материалов;
- передачу конфиденциальных данных в ИИ-инструменты;
- снижение качества экспертной проверки.

NIST рассматривает управление рисками генеративного ИИ как отдельную задачу для организаций, подчёркивая необходимость выявлять, оценивать и снижать риски в зависимости от целей и контекста применения.

Риски, связанные с конфиденциальностью

Возникают, когда пользователь передаёт в цифровую систему сведения, которые не должны становиться доступными третьим лицам. Это могут быть персональные данные, коммерческая информация, договоры, клиентские базы, внутренние документы, медицинские сведения, финансовые данные, пароли, токены, ключи доступа или фрагменты переписки.

OWASP Top 10 for LLM Applications выделяет sensitive information disclosure как один из ключевых рисков LLM-приложений. Такой риск может возникать как из-за неправильной настройки системы, так и из-за некорректного поведения пользователя или приложения, которое передаёт чувствительную информацию в модель или раскрывает её в ответе.

Для пользователя этот риск особенно сложен, потому что ущерб может быть отложенным. Человек не всегда видит немедленную проблему в том, что загрузил в нейросеть договор, клиентскую переписку или внутренний документ. Однако такая передача может нарушать режим конфиденциальности, коммерческую тайну, договорные обязательства или правила обработки персональных данных.

Репутационные риски

Репутационные риски связаны с публичным использованием непроверенных цифровых материалов. Это может быть публикация текста с ошибочными данными, распространение фейкового изображения, использование неподтверждённой статистики, цитирование несуществующего источника или выдача ИИ-сгенерированного материала за экспертный анализ.

Международные исследования показывают, что ИИ-ассистенты могут ошибаться в новостных и фактических вопросах. В сфере публичной коммуникации такая ошибка может привести к потере доверия, критике, юридическим претензиям или снижению экспертного статуса.

Репутационный риск особенно значим для:

- экспертов;
- предпринимателей;
- журналистов;
- блогеров;
- преподавателей;
- консультантов;
- компаний;
- образовательных проектов;
- публичных лиц.

Поведенческие риски и риск цифрового действия

Поведенческий риск является сквозным. Он возникает не из-за конкретного типа объекта, а из-за того, как пользователь принимает решение.

В цифровой среде пользователь часто сталкивается с ограниченным временем, большим объёмом информации, авторитетной формой подачи, эмоциональным давлением и

невозможностью быстро проверить источник. Именно поэтому центральным понятием предлагаемой классификации является риск цифрового действия.

Под риском цифрового действия понимается вероятность негативных последствий, возникающих из-за действия, совершённого пользователем после контакта с цифровым объектом.

Таблица 2. Риск цифрового действия

Цифровой объект	Возможное действие пользователя	Возможный риск
Ссылка	Перейти и ввести данные	Фишинг, кража данных, финансовые потери
Сообщение	Выполнить инструкцию	Перевод денег, передача кода, установка вредоносного ПО
ИИ-ответ	Использовать как факт	Ошибочное решение, недостоверная публикация
Документ	Подписать или оплатить	Юридический или финансовый ущерб
Файл	Открыть на устройстве	Вредоносное ПО, компрометация данных
Инвестиционное предложение	Внести средства	Потеря денег, участие в нелегальной схеме
Голосовое или видео	Поверить личности отправителя	Ошибка доверия, перевод средств, раскрытие данных
Деловое письмо	Изменить реквизиты или оплатить счёт	Business email compromise, перевод средств мошенникам

Риск цифрового действия важен потому, что многие угрозы не работают без участия человека. Мошеннический сценарий становится успешным только тогда, когда пользователь совершает нужное действие: кликает, вводит, отправляет, подписывает, переводит, скачивает, публикует или доверяет.

FBI в отчёте IC3 2025 также указывает, что cyber-enabled fraud составил 85% всех потерь, заявленных в IC3 в 2025 году, а среди затратных категорий были инвестиционные мошенничества, компрометация деловой переписки, техподдержка и имитация государственных органов. В пресс-релизе FBI отдельно отмечено, что AI-related complaints впервые выделены в отчёте IC3 как отдельный раздел: 22 364 жалобы и почти 893 млн долларов потерь.

Классификация по уровню последствий

Пользовательские цифровые риски можно классифицировать не только по типу объекта, но и по тяжести последствий.

Таблица 3. Уровни последствий пользовательских цифровых рисков

Уровень	Характер последствий	Примеры
Низкий	Ошибка без значимого ущерба	Неточная справка, бытовая ошибка, неверная рекомендация без применения
Средний	Потеря времени, исправимая ошибка, локальный ущерб	Неверный текст, некорректная публикация, переход по подозрительной ссылке без ввода данных
Высокий	Финансовый, профессиональный или репутационный ущерб	Перевод денег, подписание рискованного документа, публикация ложной аналитики
Критический	Существенные потери, утечка данных, юридические последствия	Передача кодов, потеря крупной суммы, компрометация бизнес-данных, вовлечение в мошенническую схему

Такая классификация позволяет отделять бытовые ошибки от ситуаций, где требуется повышенная осторожность и дополнительная проверка.

Классификация по группе пользователей

Пользовательские цифровые риски распределены неравномерно. Разные группы сталкиваются с разными объектами, действиями и последствиями.

Таблица 4. Группы пользователей и характерные риски

Группа пользователей	Характерные цифровые объекты	Основные риски
Пользователи 25–55 лет	Сообщения, ссылки, ИИ-ответы, документы, онлайн-банкинг	Финансовые, коммуникационные, информационные

Предприниматели и самозанятые	Договоры, счета, КП, деловые письма, ИИ-тексты	Документные, финансовые, профессиональные
Сотрудники организаций	Рабочие файлы, отчёты, ИИ-ответы, корпоративная переписка	Профессиональные, репутационные
Родственники пожилых пользователей	Сообщения, звонки, ссылки, банковские сценарии	Коммуникационные, финансовые, поведенческие
Студенты и учащиеся	ИИ-ответы, учебные тексты, источники	Информационные, образовательные, репутационные
Публичные эксперты и авторы	Публикации, аналитика, изображения, цитаты	Репутационные, информационные, правовые
Малый бизнес	Документы, счета, сайты, переписка, реквизиты	Документные, финансовые

Эта классификация показывает, что универсального пользовательского риска не существует. Один и тот же цифровой объект может быть низкорисковым для бытового использования и высокорисковым для бизнеса, финансов или публичной коммуникации.

Классификация по источнику риска

Цифровой риск может исходить из разных источников.

Таблица 5. Источники пользовательского цифрового риска

Источник риска	Содержание риска	Пример
Ошибка ИИ	Система выдаёт недостоверный или неподтверждённый ответ	ИИ ссылается на несуществующий источник
Мошенник	Злоумышленник манипулирует пользователем	Сообщение от имени банка или руководителя
Небезопасный ресурс	Сайт или ссылка используются для сбора данных	Фишинговая форма оплаты
Ошибка пользователя	Пользователь не проверяет источник или контекст	Вводит код из SMS по просьбе «службы безопасности»

Организационный пробел	Нет регламентов проверки ИИ-результатов	Сотрудник вставляет в отчёт неподтверждённые данные
Низкая цифровая грамотность	Пользователь не знает признаков риска	Скачивает приложение по ссылке из сообщения
Доверие к форме	Пользователь верит визуальному оформлению	Поддельный документ выглядит официально

Предложенная классификация показывает, что пользовательские цифровые риски в эпоху генеративного ИИ имеют смешанную основу.

Во-первых, они являются информационными, потому что связаны с достоверностью сведений, источников и контекста.

Во-вторых, они являются техническими, потому что могут включать ссылки, сайты, файлы, вредоносное ПО, уязвимости ИИ-приложений и небезопасную обработку данных.

В-третьих, они являются поведенческими, потому что значительная часть вреда возникает после действия пользователя. Именно это подтверждает статистика Банка России, где социальная инженерия остаётся важным элементом мошеннических сценариев.

В-четвёртых, они являются профессиональными, поскольку ИИ всё чаще используется не только в быту, но и в организациях. Данные ИСИЭЗ НИУ ВШЭ показывают, что ИИ уже применяется в российских организациях, но масштаб использования и уровень компетенций распределены неравномерно.

Следовательно, профилактика пользовательских цифровых рисков не может сводиться только к предупреждению о мошенниках или технической блокировке вредоносных сайтов. Необходима более широкая рамка: проверка информации, источника, контекста и действия.

Ограничения исследования

Настоящее исследование имеет ряд ограничений.

Во-первых, статья основана на вторичных открытых данных и не включает собственное эмпирическое исследование пользователей. Для дальнейшей проверки классификации целесообразны опросы, глубинные интервью, фокус-группы и тестирование пользовательских сценариев.

Во-вторых, официальная российская статистика пока не всегда выделяет случаи, напрямую связанные с генеративным ИИ. Поэтому некорректно утверждать, что весь цифровой риск вызван искусственным интеллектом. Более точная формулировка: генеративный ИИ усиливает отдельные категории риска, связанные с масштабированием, убедительностью, автоматизацией и персонализацией цифрового контента.

В-третьих, статистика МВД России и Банка России отражает разные объекты учёта. МВД фиксирует преступления в уголовно-правовой логике, а Банк России — операции без

добровольного согласия клиентов финансовых организаций. Эти данные дополняют друг друга, но не являются единой статистической серией.

В-четвёртых, международные источники не могут напрямую переноситься на российский рынок без учёта правового, культурного, технологического и институционального контекста. Они используются как сравнительный фон и подтверждение глобального характера проблемы.

Заключение

Массовое применение генеративного искусственного интеллекта и цифровых коммуникаций изменило структуру пользовательских рисков. Пользователь всё чаще взаимодействует с объектами, происхождение, достоверность и безопасность которых невозможно оценить только по внешнему виду: ИИ-ответами, ссылками, сообщениями, документами, файлами, сайтами, изображениями, аудио- и видеоматериалами.

Проведённый анализ позволяет выделить восемь групп пользовательских цифровых рисков: информационные, финансовые, коммуникационные, документные, профессиональные, репутационные, поведенческие и риски конфиденциальности. Эти риски различаются по объекту, механизму воздействия, типу действия и последствиям.

Ключевой вывод статьи состоит в том, что риск цифрового объекта проявляется в действии, которое он инициирует у пользователя. Поэтому в эпоху генеративного ИИ цифровая грамотность должна включать не только умение пользоваться технологиями, но и навык предварительной проверки: источника, контекста, достоверности, безопасности и последствий действия.

Предложенная классификация может быть использована как основа для дальнейших исследований пользовательской уязвимости, разработки образовательных программ, методик профилактики цифрового мошенничества и оценки рисков при взаимодействии с ИИ-системами и цифровыми коммуникациями.

Этическое заявление

Автор заявляет, что несёт полную ответственность за содержание настоящей рукописи. Исследование выполнено на основе открытых статистических, аналитических и методических источников и не включает сбор, обработку или публикацию персональных данных участников. Эксперименты с участием людей или животных не проводились. Информированное согласие участников исследования не требовалось, поскольку эмпирический материал представлен вторичными открытыми данными. Конфликт интересов отсутствует.