DOI 10.35775/PSI.2026.127.3.041
УДК 32.327

И.А. ФИНАГИН
магистрант факультета управления и политики,
МГИМО МИД России, Россия, г. Москва
E-mail: finagin505@yandex.ru

ГЕНЕРАТИВНЫЙ ИИ КАК ФАКТОР РУТИНИЗАЦИИ УТЕЧКИ ЧУВСТВИТЕЛЬНЫХ ДАННЫХ: ВЫЗОВЫ ДЛЯ МЕЖДУНАРОДНОЙ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ И ЦИФРОВОГО СУВЕРЕНИТЕТА

Статья посвящена влиянию генеративного искусственного интеллекта (ИИ) на трансформацию киберрисков в сфере международной информационной безопасности (МИБ). Актуальность темы обусловлена тем, что в последние годы ИИ активно встраивается в государственные и корпоративные информационные системы, вследствие чего возрастает необходимость анализа связанных с этим рисков раскрытия чувствительной информации. Гипотеза исследования заключается в том, что генеративный ИИ рутинизирует передачу чувствительных данных за пределы контролируемого контура и смещает риск утечек с технических уязвимостей на повседневное поведение пользователей. Теоретическую основу исследования составляют концепции международной информационной безопасности и цифрового суверенитета, позволяющие анализировать трансграничные риски обработки данных на внешних ИИ-платформах. В работе используются метод вторичного анализа данных и кейс-стади. Ключевые результаты свидетельствуют о том, что добровольная передача данных через ИИ приобретает признаки массовой практики и трансформирует механизм возникновения киберугроз в контексте МИБ.

Ключевые слова: *генеративный искусственный интеллект, киберугрозы, международная информационная безопасность, цифровой суверенитет, трансграничная утечка данных, технологическая зависимость.*

Введение. Генеративный ИИ за последние годы стал значимым элементом цифровой инфраструктуры бизнеса, государственного управления и повседневной жизни. Вместе с ростом эффективности и доступности ИИ-сервисов усиливается и спектр угроз, связанных с возможной утечкой чувствительной информации за пределы национальной юрисдикции. Если ранее подобные риски ассоциировались преимущественно с кибератаками или инсайдерской деятельностью, то сегодня значимая часть угроз возникает в процессе обычного использования внешних цифровых платформ сотрудниками компаний и государственных структур. Научная новизна статьи заключается в том, что генеративный ИИ рассматривается в качестве фактора формирования новых киберрисков, связанных с утечкой данных и изменением повседневных практик

работы с информацией. Целью статьи является выявление того, как генеративный ИИ меняет механизм киберугроз, превращая добровольную передачу данных во внешний контур в массовое и рутинное явление. Гипотеза исследования заключается в том, что генеративный ИИ рутинизирует передачу чувствительных данных во внешний контур и смещает риск утечек с технических уязвимостей на повседневное поведение пользователей.

Следует подчеркнуть, что в работах российских и зарубежных авторов, опубликованных в последние годы, освещается широкий спектр вопросов близких к данной предметной области [2; 3; 7; 8; 11; 13; 15; 16; 17; 18; 19].

Однако проблему влияния генеративного искусственного интеллекта на трансформацию киберрисков в сфере международной информационной безопасности нельзя назвать однозначно исчерпанной. В силу многих объективных обстоятельств изучение обозначенной темы продолжает сохранять высокий уровень актуальности.

В рамках настоящего исследования под внешним контуром понимается цифровая среда обработки и хранения данных, которая находится вне прямого организационного, технологического и юрисдикционного контроля субъекта, передающего информацию. Иными словами, речь идет о таких внешних платформах и инфраструктурах, при использовании которых организация или государство не способны в полной мере определять режим хранения, последующего использования, удаления и трансграничного перемещения данных. Таким образом, передача информации во внешний контур рассматривается не только как корпоративный риск, но и как фактор МИБ.

Смещение риска в статье понимается как смещение преобладающего механизма возникновения утечки от инцидентов, обусловленных внешним взломом, вредоносным проникновением или инсайдерским хищением, к инцидентам, возникающим в результате штатного использования сотрудниками внешних генеративных ИИ-платформ.

Рутинизация в статье понимается как превращение передачи чувствительной информации во внешний ИИ-контур в повторяющуюся, охватывающую разные сектора и встроенную в повседневный рабочий процесс практику, которая воспринимается пользователями как обычный способ решения профессиональных задач.

Теоретическую основу исследования составляют концепции МИБ и цифрового суверенитета, позволяющие анализировать трансграничные риски обработки данных на внешних ИИ-платформах. Объектом исследования выступают риски использования внешних генеративных ИИ-платформ в корпоративной и государственной практике. Предметом исследования являются механизмы рутинизации передачи чувствительной информации во внешний ИИ-контур и их значение для МИБ и цифрового суверенитета. Единицей анализа являются статистически зафиксированные практики передачи конфиденциальных данных, а также институционально значимые кейсы, демонстрирующие утрату контроля над информацией при использовании внешних генеративных ИИ-сервисов. Метод вторичного анализа данных используется для обобщения исследований,

докладов и статистических материалов о рисках утечки данных через генеративные ИИ-сервисы. Метод кейс-стади используется для рассмотрения конкретных корпоративных и государственных эпизодов, показывающих трансформацию механизма киберугроз при использовании внешних ИИ-платформ. Статистические материалы используются для выявления массовости и повторяемости практики передачи чувствительных данных во внешний ИИ-контур, тогда как кейсы позволяют выявить институциональные последствия этой практики и показать ее проявление в разных организационных средах. Эмпирический материал охватывает период 2023-2026 гг., то есть этап наиболее интенсивного массового распространения внешних генеративных ИИ-платформ в корпоративной и административной практике. В исследование включаются кейсы, которые одновременно соответствуют следующим критериям. Во-первых, кейс должен фиксировать использование внешней генеративной ИИ-платформы, то есть сервиса, находящегося за пределами полного организационного или государственного контроля субъекта передачи данных. Во-вторых, кейс должен быть связан с передачей чувствительной информации либо с документированным риском такой передачи. В-третьих, кейс должен обладать институциональной значимостью, то есть сопровождаться запретом, ограничением, официальной реакцией, изменением внутренних правил или выявлением риска на корпоративном или государственном уровне.

Основная часть. Международная информационная безопасность определяется как «состояние международных отношений, исключающее нарушение мировой стабильности и создание угрозы безопасности государств и мирового сообщества в информационном пространстве». К числу ключевых угроз МИБ относятся создание условий технологической зависимости в ущерб другим государствам, приобретение контроля над национальными информационными ресурсами и инфраструктурами, а также неконтролируемое трансграничное распространение информации [9. С. 193].

Исторические прецеденты, такие как WikiLeaks (масштабная трансграничная утечка государственных документов) и PRISM (институционализированная массовая эксфильтрация данных через инфраструктуру интернет-компаний), показали, что контроль над информационными потоками и инфраструктурой становится одним из ключевых элементов международной безопасности [9. С. 201].

В этих условиях все большую роль приобретает концепция цифрового суверенитета, которая акцентирует внимание на способности государства контролировать критически важные данные, инфраструктуру и платформы в условиях трансграничных потоков информации и технологической зависимости. Многие авторы также связывают проблематику цифрового суверенитета с вопросами обеспечения информационной безопасности и технологической независимости государства [5. С. 40].

Связь между МИБ и цифровым суверенитетом заключается в том, что угрозы в информационной сфере возрастают при отсутствии у государства достаточной технологической самостоятельности и контроля над ключевыми цифровыми компонентами, тогда как цифровой суверенитет выступает рамкой их защищенного и независимого функционирования [1. С. 79-81].

Важным аспектом МИБ становится все более широкое проникновение машинного обучения и ИИ в разные сферы применения информационных технологий, что приводит к новым пересечениям с кибербезопасностью и уязвимостям ИИ-систем [14. С. 135-140].

В случае генеративного ИИ принципиально важно не только то, что появляются новые уязвимости самих ИИ-систем, но и то, что меняется механизм возникновения инцидента. В классической модели информационной безопасности основное внимание уделяется взлому, вредоносному проникновению или инсайдерскому хищению данных, однако с распространением внешних ИИ-платформ часть риска переносится в сферу повседневных рабочих практик. Сотрудники используют ИИ-помощников для подготовки текстов, редактирования изображений, анализа документов и планирования, а для повышения качества ответа загружают в сервисы клиентские данные, сведения о сотрудниках, юридические и финансовые документы, учетные данные и конфиденциальный программный код [4].

По данным исследования Harmonic Security 8,5% запросов сотрудников ИИ-сервисам в 2025 г. содержат конфиденциальные сведения и были связаны с утечкой чувствительной информации, включая персональные данные, идентификаторы доступа к сервисам, информацию о сделках, отчеты об инцидентах, ключи доступа и т.д. [21]. Этот показатель позволяет сделать вывод, что передача чувствительных данных во внешние ИИ-сервисы перестает быть исключительным нарушением и приобретает признаки повседневной пользовательской практики. Схожую динамику фиксируют и российские данные. За последний год количество конфиденциальной информации, отправляемой сотрудниками российских компаний в популярные ИИ-сервисы, такие как ChatGPT и Gemini выросло в 30 раз [6].

Ключевой риск состоит в том, что после попадания в открытую ИИ-систему такая информация утрачивает прежнюю степень контролируемости. Она может сохраняться в журналах обработки, использоваться при дообучении модели или быть раскрыта третьим лицам, вследствие чего утечка приобретает практически необратимый характер [20].

Одним из наиболее известных примеров является кейс Samsung. В 2023 году сотрудники компании загружали в ChatGPT внутренний исходный код и иные чувствительные материалы. После серии таких эпизодов компания пришла к выводу, что данные, передаваемые во внешние генеративные платформы, хранятся на внешних серверах, их трудно вернуть или удалить, и они потенциально могут быть раскрыты другим пользователям. В ответ Samsung сначала ввела ограничения, а затем полностью запретила использование генеративных ИИ-систем на корпоративных устройствах и во внутренних сетях [23].

Реакция Samsung не осталась единичной. В 2023 году ограничения или запреты на использование подобных инструментов ввели и крупные банки, в том числе Bank of America, Commonwealth Bank of Australia, Citigroup, Deutsche Bank, Goldman Sachs Group, Wells Fargo и JPMorgan Chase [12. С. 137].

Еще один важный международный пример связан с государственным сектором США. Исполняющий обязанности директора Агентства по кибербезопасности и защите инфраструктуры США Мадху Готтумуккала загружал в общедоступную версию ChatGPT документы под грифом «для служебного пользования», что вызвало многократное срабатывание системы предотвращения утечек в сетях федерального правительства [24].

Таким образом, использование внешних генеративных ИИ-платформ уже перестало быть частным эпизодом и превратилось в системный источник риска как для корпоративного, так и для государственного сектора.

Показателен и пример проникновения генеративного ИИ в публичную сферу. В Португалии внимание вызвал проект закона парламентской группы Социал-демократической партии (PSD) о мерах защиты детей в цифровой среде, в преамбуле которого была обнаружена гиперссылка, заканчивающаяся на «?utm_source=chatgpt.com» [22].

Схожий случай произошел в российской практике. В пояснительной записке к законопроекту № 1126815-8 также была найдена ссылка на ChatGPT. В комитете ГД РФ объяснили это тем, что сотрудник аппарата комитета «использовал ИИ как инструмент для быстрого поиска только источника самого исследования» [10].

Таким образом, даже если внешняя ИИ-система не использовалась для подготовки самого законопроекта или пояснительной записки, обращение к ней при поиске источников уже означает передачу во внешний контур поисковых формулировок, служебного контекста и связанных метаданных. В результате по совокупности запросов внешняя платформа потенциально может реконструировать предмет интереса ведомства, круг анализируемых проблем и отдельные элементы регуляторного замысла. При работе государственных структур этот фактор приобретает значение риска для информационной безопасности, контроля над данными и цифрового суверенитета.

Рассмотренные кейсы показывают, что на уровне повседневной практики генеративный ИИ нормализует передачу чувствительной информации во внешний контур. На институциональном уровне это ведет к утрате контроля над хранением, удалением, повторным использованием и последующим перемещением данных, а на государственном и международном уровнях ИИ усиливает зависимость от внешних платформ и расширяет потенциальные каналы доступа к национально значимой информации. Тем самым рутинное пользовательское действие приобретает значение проблемы МИБ, поскольку локальная практика обработки запросов масштабируется в проблему юрисдикции, инфраструктурного контроля и цифрового суверенитета.

Заключение. Генеративный ИИ меняет механизм возникновения киберинцидентов. Теперь риск утечек смещается с технических уязвимостей (взлом, вредоносное проникновение, инсайдерское хищение) на повседневное поведение пользователей. С каждым годом значительно увеличивается масштаб добровольной передачи чувствительных данных во внешние ИИ-системы. При этом последствия подобной практики оказываются труднообратимыми.

После попадания информации во внешнюю ИИ-систему возможности контроля над ее хранением, удалением и последующим использованием существенно сокращаются. В частности, данные могут сохраняться в журналах обработки, использоваться при дообучении модели или быть раскрыты третьим лицам. Существенно, что риск утечки значимой информации возникает даже при косвенном использовании внешних ИИ-платформ, например, при поиске источников и справочных материалов.

Подводя итог сказанному выше, следует подчеркнуть, что распространение генеративного ИИ создает предпосылки для актуализации трех угроз МИБ. ИИ усиливает технологическую зависимость, расширяет каналы для внешнего доступа к национальном значимой информации и обеспечивает ее трансграничное перемещение вне контроля государства. Именно в этом контексте описанные угрозы непосредственно связаны с проблематикой цифрового суверенитета.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК:

1. **Бухарин В.В.** Компоненты цифрового суверенитета российской федерации как техническая основа информационной безопасности // Вестник МГИМО-Университета. 2016. № 6 (51).
2. **Гавров С.Н., Еремкин М.П.** Использование искусственного интеллекта в контексте информационной войны // Вопросы политологии. 2025. № 2.
3. **Дресвянин Д.В.** Управление искусственным интеллектом в мировой политике: вызовы национального суверенитета и перспективы глобального регулирования // Вопросы национальных и федеративных отношений. 2025. № 12.
4. Дружелюбный инсайдер: как ИИ становится новым каналом утечек // Forbes.ru. 22.05.2025 // https://www.forbes.ru/tekhnologii/537555-druzelubnyj-insajder-kak-ii-stanovitsa-novym-kanalom-utecek.
5. **Зиновьева Е.С., Шитьков С.В.** Цифровой суверенитет в практике международных отношений // Международная жизнь. 2023. № 3.
6. Из российских компаний утекло через ИИ-сервисы в 30 раз больше данных, чем в прошлом году // CNews. 04.02.2026 // https://www.cnews.ru/news/top/2026-02-04_sotrudniki_rossijskih_kompanij.
7. **Инхеев А.Б.** Искусственный интеллект в политических технологиях США: от стратегических коммуникаций к когнитивному воздействию // Вопросы национальных и федеративных отношений. 2025. № 10.
8. **Квиндт А.В.** Феномен цифрового аватара и его применение в государственной политике в условиях цифровой трансформации // Вопросы политологии. 2026. № 1.
9. **Крутских А.В., Бирюков А.В., Бойко С.М.** [и др.] Международная информационная безопасность: теория и практика: в 3 томах. Т. 1 / под общ. ред. А.В. Крутских. М.: Аспект Пресс, 2019. 384 с.
10. Комитет ГД ответил на слухи об использовании ChatGPT при подготовке проекта закона // Forbes.ru. 24.01.2026 // https://www.forbes.ru/

society/554152-komitet-gd-otvetil-na-sluhi-ob-ispol-zovanii-chatgpt-pri-podgotovke-proekta-zakona.
11. **Кубанцева Е.В.** Индекс цифровой независимости как инструмент оценки суверенитета государства в условиях цифрового неоколониализма // Вопросы политологии. 2026. № 1.
12. Международная безопасность в эпоху искусственного интеллекта: в томах. Т. 1 / под ред. М.В. Захаровой, А.И. Смирнова. М.: Аспект Пресс, 2024. 401 с.
13. **Москвичев В.Р., Ершова Л.В.** Цифровые медиа как фактор трансформации политического лидерства в XXI веке // Вопросы политологии. 2026. № 1.
14. **Намиот Д.Е., Ильюшин Е.А., Чижов И.В.** Искусственный интеллект и кибербезопасность // International Journal of Open Information Technologies. 2022. Т. 10. № 9.
15. **Сурма И.В.** «Общество 5.0» и квантовый искусственный интеллект: глобальная социально-экономическая трансформация через призму новых технологий // Евразийский Союз: вопросы международных отношений. 2026. № 1.
16. **Сюй Чан.** The Reconstruction of World Order in the Digital Age: Technological Power, Rule Game, and New Geopolitical Landscape/Трансформация мирового порядка в цифровую эпоху: технологическая мощь, правила игры и новый геополитический ландшафт // Вопросы политологии. 2025. № 9.
17. **Талагаева Д.А., Кочеринский Д.Ю.** Этические аспекты искусственного интеллекта и технологический суверенитет: национальные модели и глобальное управление // Евразийский Союз: вопросы международных отношений. 2025. № 11.
18. **Шерматов А.Г.** Перспективы развития политики искусственного интеллекта в Узбекистане // Вопросы политологии. 2026. № 1.
19. **Шкурина С.С.** Влияние информационно-коммуникационных технологий на современные общественно-политические процессы // Региональное и муниципальное управление: вопросы политики, экономики и права. 2025. № 2.
20. Эксперт Content AI предупредил о рисках утечки данных при использовании публичных нейросетей // Content AI. 02.12.2025 // https://contentai.ru/news/b8i7nuugm1-ekspert-content-ai-predupredil-o-riskah.
21. Harmonic: использование GenAI в работе организаций подвергает риску ее конфиденциальные данные // CISOCLUB. 20.01.2025 // https://cisoclub.ru/harmonic-ispolzovanie-genai-v-rabote-organizacij-podvergaet-risku-ejo-konfidencialnye-dannye/.
22. Link com referência ao ChatGPT em projeto de lei do PSD é real? // SIC Notícias. 23.02.2026 // https://sicnoticias.pt/sicverifica/2026-02-23-video-link-com-referencia-ao-chatgpt-em-projeto-de-lei-do-psd-e-real--76b5a96.
23. Samsung Bans Staff's AI Use After Spotting ChatGPT Data Leak // Bloomberg. 02.05.2023 // https://www.bloomberg.com/news/articles/2023-05-02/samsung-bans-chatgpt-and-other-generative-ai-use-by-staff-after-leak.
24. Trump's acting cyber chief uploaded sensitive files into a public version of ChatGPT // Politico. 27.01.2026 // https://www.politico.com/news/2026/01/27/cisa-madhu-gottumukkala-chatgpt-00749361.

I.A. FINAGIN
Master's student, MGIMO-University,
Moscow, Russia

GENERATIVE AI AS A DRIVER OF THE ROUTINIZATION OF SENSITIVE DATA LEAKAGE: CHALLENGES FOR INTERNATIONAL INFORMATION SECURITY AND DIGITAL SOVEREIGNTY

The rapid integration of generative AI into business operations, public administration, and everyday information practices is reshaping the logic of cyber risk. Rather than functioning only as a new technical attack surface, external AI platforms normalize the routine transfer of sensitive information beyond the organizational, technological, and jurisdictional perimeter. This article argues that generative AI shifts the dominant mechanism of data leakage from external intrusion, malicious compromise, or insider theft to ordinary user behavior embedded in daily work with AI assistants. The novelty of the study lies in examining generative AI not merely as a source of technical vulnerabilities, but as a socio-technical factor that transforms established practices of handling information and, consequently, the structure of cyber threats. The analysis is based on the concepts of international information security and digital sovereignty, which make it possible to interpret AI-related risks not only at the corporate level, but also as a problem of cross-border governance and state control over critical data flows. Methodologically, the article combines secondary data analysis, document analysis, and case studies. The empirical material includes reports on the growing disclosure of confidential information to widely used AI services, corporate restrictions introduced after incidents involving source code and internal documentation, and public-sector examples showing that even indirect interaction with external AI tools may expose sensitive contextual information. This is especially significant for state institutions, where prompts, metadata, and thematic search patterns may allow external platforms to infer policy priorities or reconstruct elements of regulatory design. The findings demonstrate that generative AI routinizes the transfer of valuable data into an external environment where storage, reuse, deletion, and onward transfer are not fully controlled by the original holder. As a result, generative AI reinforces technological dependence, facilitates external access to nationally significant data, and intensifies uncontrolled cross-border information flows, thereby increasing the strategic importance of digital sovereignty.

Key words: *AI, cybersecurity, international information security, digital sovereignty, data leakage, cross-border data flows, AI governance.*