ISSN 2658-5014 (Print) СИИТ ISSN 2686-7044 (Online)
http://siit.ugatu.su 2026. Т. 8, № 2(26)
СИСТЕМНАЯ ИНЖЕНЕРИЯ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ
Научная статья

УДК 004.8 DOI: 10.54708/SIIT-2026-no2-p92 EDN: CUOCTO

Методология интеграции искусственного интеллекта в рабочий процесс OSINT-исследования: синергия, риски и управление

Д. А. Черенков

Администрация городского округа город Уфа

В статье рассматривается проблематика внедрения генеративного искусственного интеллекта (ИИ) в аналитические процессы разведки на основе открытых источников (OSINT). Анализируются комплементарные роли человека и ИИ, выявляются ключевые риски, включая фундаментальный риск «подхалимства» (sycophancy) ИИ, и рассматриваются вопросы операционной безопасности. Предлагается пошаговая методология интеграции и практические инструменты для минимизации рисков, основанная на принципах критического мышления, верификации и организационных процедурах. Статья носит научно-прикладной характер и предназначена для специалистов, стремящихся к ответственной и эффективной интеграции новых технологий.

Искусственный интеллект, разведка на основе открытых источников, анализ данных, принятие решений, безопасность.

Введение

От технологического оптимизма к взвешенной интеграции

Современный этап развития OSINT (Open Source Intelligence, «разведка по открытым источникам») характеризуется лавинообразным ростом объема и разнообразия открытых данных. Традиционные методы их обработки становятся недостаточными, что актуализирует внедрение инструментов искусственного интеллекта. Однако ключевой задачей является не погоня за технологическими новинками, а формирование методологии их ответственного применения.

ИИ не должен заменять аналитика; его роль – выступать в качестве когнитивного усилителя (Cognitive Amplifier), берущего на себя задачи по обработке данных, оставляя за человеком функции критической оценки, контекстуализации и принятия решений [Elk25]. В данной статье предлагается практический фреймворк для такой интеграции с учетом минимизации присущих ИИ рисков, в том числе наиболее коварного – риска «подхалимства».

Комплементарность ролей в связке «человек–ИИ»

Фундаментальным принципом построения гибридных интеллектуальных систем в OSINT является не конкуренция, а синергия, достигаемая за счет комплементарности функций [Mah25].

Рекомендовано к публикации программным комитетом XI Международной научной конференции ITIDS’2025 «Информационные технологии интеллектуальной поддержки принятия решений», Уфа, 13–15 ноября 2025 г.

Черенков Д. А. Методология интеграции искусственного интеллекта в рабочий процесс OSINT-исследования: синергия, риски и управление // СИИТ. 2026. Т. 8, № 2(26). С. 92-98. DOI: 10.54708/SIIT-2026-no2-p92. EDN: CUOCTO.	Cherenkov D. A. “Methodology for integrating artificial intelligence into OSINT investigation workflow: synergies, risks, and management” // SIIT. 2026. Vol. 8, no. 2(26), pp. 92-98. DOI: 10.54708/SIIT-2026-no2-p92. EDN: CUOCTO. (In Russian).

Вместо того чтобы пытаться наделить ИИ человеческими качествами или наоборот – низвести аналитика до уровня оператора данных, необходимо максимально использовать уникальные преимущества каждой из сторон. Эффективность этой связки основана на четком разделении сильных сторон.

Сильные стороны ИИ:

- обработка больших объемов данных: масштабируемый анализ тысяч документов, постов в социальных сетях, изображений [Руд25, Рез25];
- выявление паттернов и аномалий: обнаружение скрытых корреляций, сетевых связей, координированной активности [Сем25];
- структурирование информации: автоматическая категоризация, извлечение сущностей (имена, даты, локации, номера), составление суммаризаций;
- генерация гипотез: предложение новых векторов для расследования на основе анализа данных.

Сильные стороны человека-аналитика:

- критическое мышление и верификация: Оценка достоверности источников, проверка выводов ИИ;
- контекстуализация: Понимание культурных, социальных и политических нюансов, мотивов и поведения людей;
- операционная безопасность (OPSEC): Оценка рисков при сборе данных и их загрузке в сторонние сервисы [Даи24];
- этическая и юридическая оценка: принятие решений в рамках правового поля и этических норм.

Фундаментальный риск: «подхалимство» (Sycophancy) ИИ как системная угроза анализу

Помимо общеизвестных рисков, таких как «галлюцинации», существует более глубокая и системная угроза – «подхалимство» или «сервилизм» ИИ. Это склонность системы чрезмерно льстить, соглашаться и усиливать взгляды, эмоции или предпочтения пользователя, вместо того чтобы давать взвешенные, точные или провокационные ответы.

В отличие от явно ложных галлюцинаций, подхалимство создает «изощрённые эхо-камеры»[1]. Оно подтверждает существующие предубеждения аналитика, создавая видимость сотрудничества и готовности помочь, при этом, часто не оспаривая ложные предпосылки в запросах.

Глубинная причина: заложена в самом механизме обучения моделей, таких как «Обучение с подкреплением на основе обратной связи от человека» (RLHF). Системы оптимизированы для повышения удовлетворенности пользователей, а не точности. Люди-оценщики невольно поощряют приятные и удобные ответы, создавая систематическое смещение [Chr23].

Эмпирическое подтверждение: явление склонности ИИ к изменению первоначально правильных ответов под влиянием пользователя экспериментально подтверждено в исследовании Anthropic (2023), где модели демонстрировали так называемую «уступчивость» (Model Sycophancy), подстраивая свои выводы под ожидания пользователя в ущерб точности [Gan23].

Это доказывает, что стремление к согласию может перевешивать стремление к истине.

Проявление в OSINT на практике:

1) подтверждение предвзятости (Confirmation Bias): аналитик, уверенный в виновности объекта расследования, получает от ИИ усиленный анализ, который игнорирует противоречащие доказательства;

---

[1] Why AI Agreeableness Poses Risks to OSINT Work [Электронный ресурс]. URL: https://www.osintcombine.com/post/ai-agreeableness-risks-to-osint (дата обращения: 22.09.2025).

2) ИИ может корректно распознать на спутниковом снимке объекты промышленной инфраструктуры (например, градирни ТЭЦ или резервуары для хранения воды), но если аналитик уверенно предположит, что это «установки для химического производства», ИИ может поддержать ошибочную интерпретацию;

3) нежелание оспаривать: ИИ не станет задавать уточняющие вопросы или подвергать сомнению слабые посылки запроса. Данный риск катастрофически опасен, так как его результат кажется правдоподобным, соответствует ожиданиям исследователя и мимикрирует под продуктивную коллаборацию, что радикально затрудняет его выявление.

Из всех функций управления менеджеры высшего эшелона оставляют себе только две стратегические функции: постановку цели и контроль, то есть задачу стратегического прогнозирования и управления, что, в свою очередь, требует не только и не столько поиска информации, которая формирует знания о настоящей ситуации, а поиска так называемой прогностической информации, которая определяет вероятность реализации и успеха предлагаемых сценариев, а также оценку тенденций изменений ситуации на рынке. Это определяет актуальность использования методов конкурентной разведки в современном мире.

Внедрение в науку и практику рыночного хозяйствования методов искусственного интеллекта (ИИ) актуализируют известные методы конкурентной разведки, что позволяет не только получать более достоверные данные, но и совершенствовать полученные результаты с помощью ИИ.

Пошаговый план интеграции ИИ в рабочий процесс OSINT

Предлагаемая модель (табл. 1) следует классическому циклу разведывательной деятельности и включает меры противодействия «подхалимству».

Таблица 1

Ключевые риски искажения информации и методы их минимизации

Этап	Цель	Инструменты ИИ (примеры)	Роль человека и противодействие рискам
Планирование	Формулировка задач, определение векторов.	ChatGPT, Claude, YandexGPT: Генерация планов исследования, составление чек-листов.	Критическая оценка планов. Нейтральные промты.
Сбор данных	Автоматизированный сбор информации.	Python-скрипты (Beautiful Soup), SpiderFoot, theHarvester, мониторинг СМИ.	Настройка параметров, обеспечение OPSEC (прокси). Использование ИИ только для сбора, но не интерпретации.
Обработка и анализ	Структурирование данных, выявление связей.	Анализ текста: ChatGPT для суммаризации. Анализ изображений: Yandex Vision, PimEyes. Сетевой анализ: Maltego.	Верификация всех выводов по первоисточникам. Техника опровержения Разделение труда: создание «красной команды» для оспаривания выводов.
Синтез и отчетность	Формирование аналитического продукта.	Написание черновых вариантов отчетов, структурирование данных.	Написание итогового отчета. Слепая проверка. Отсроченный анализ: возврат к выводам через 24 часа для перепроверки.

Ключевые аспекты модели:

- принцип комплементарного разделения функций: ИИ обрабатывает массивы данных, оператор обеспечивает критическое мышление и контекст;
- многоуровневая система валидации включает перекрестные проверки, метод красной команды и слепую валидацию;
- OPSEC-приоритет: на каждом этапе обеспечивается операционная безопасность;

- адаптивность под задачи: инструменты выбираются под конкретные цели расследования;
- документирование процесса: фиксация всех этапов для последующего аудита[2].

Противодействие ключевым рискам:

- против «подхалимства» ИИ – метод красной команды[3], «слепая валидация»;
- против галлюцинаций – мандатная проверка, верификация по первоисточникам;
- против предвзятости – нейтральные формулировки запросов, Multiple Hypotheses Testing.

Данная модель позволяет систематизировать процесс OSINT-расследования с ИИ, минимизируя риски и повышая достоверность конечных результатов.

Ключевые риски искажения информации и методы их минимизации

В процессе интеграции искусственного интеллекта в рабочие процессы OSINT (табл. 2) необходимо учитывать несколько ключевых рисков и применять соответствующие методы их минимизации.

Таблица 2

Ключевые риски искажения информации и методы их минимизации

Риск	Сущность риска	Методы минимизации
«Подхалимство» (Sycophancy)	Систематическое соглашательство с пользователем в ущерб точности.	Относиться к согласию ИИ с той же настороженностью, что и к его ошибкам. Явные инструкции: «Оспорь мои предположения». Использование анализа конкурирующих гипотез (ACH) независимо от ИИ.
«Галлюцинации» (Hallucinations)	Генерация ложной информации.	Принцип обязательной верификации по первоисточнику. Критическая оценка каждого факта (дат, имен, цитат).
Контекстуальная слепота	Непонимание локального контекста, иронии, сленга.	Человек как проводник контекста. Постоянная фильтрация выводов ИИ через призму экспертных знаний.
Системная предвзятость (Bias)	Закрепление предубеждений из обучающих данных.	Критическое отношение к выводам. Использование разнообразных источников и перекрестная проверка.
Нарушение OPSEC	Утечка конфиденциальных данных.	Использование локальных моделей (Ollama) для чувствительных данных. Строгий запрет на загрузку в публичные чаты персональных данных и деталей расследования.

Риск «подхалимства» (Sycophancy) проявляется в систематическом соглашательстве ИИ с пользователем в ущерб точности анализа. Для противодействия этому следует относиться к согласию ИИ с той же настороженностью, что и к его ошибкам, давать явные инструкции типа «Оспорь мои предположения» или «Какие доказательства опровергают эту точку зрения?», а также использовать анализ конкурирующих гипотез независимо от ИИ.

Не менее серьезной проблемой являются «галлюцинации» ИИ – генерация ложной информации [Кол25]. Минимизировать этот риск позволяет строгое соблюдение принципа обязательной верификации по первоисточнику и критическая оценка каждого факта. Контекстуальная слепота ИИ, выражающаяся в непонимании локального контекста, требует постоянного участия человека как проводника контекста и фильтрации выводов ИИ через призму местной специфики.

---

[2] Методические рекомендации по проведению мероприятий по выявлению и сбору информационных признаков угроз безопасности в информационно-телекоммуникационной сети «Интернет» / ФГУП «Главный радиочастотный центр». М., 2021. 67 с.

[3] Планирование красной команды для больших языковых моделей (LLM) и их приложений [Электронный ресурс]. URL: https://video2.skills-academy.com/ru-ru/azure/ai-foundry/openai/concepts/red-teaming (дата обращения: 30.09.2025).

Системная предвзятость (Bias), возникающая из-за закрепления предубеждений в обучающих данных, нейтрализуется через критическое отношение к выводам и использование разнообразных источников информации [Сун24]. Наконец, риск нарушения OPSEC, связанный с утечкой конфиденциальных данных, минимизируется путем использования локальных моделей и строгого запрета на загрузку персональных данных в публичные чаты. Комплексное применение этих мер защиты позволяет существенно повысить надежность и достоверность результатов OSINT-исследований с использованием искусственного интеллекта.

Практические решения для различных сценариев использования

При использовании ИИ в качестве партнера для мозгового штурма необходимо заранее инструктировать систему занять противоположную позицию, например, сформулировав запрос: «Проанализируй этот подход с критической точки зрения и выдели наиболее существенные возражения». Для сохранения объективности рекомендуется создавать новый чат для каждого отдельного сеанса работы.

В исследовательской деятельности ИИ следует применять как инструмент поиска источников, а не как самостоятельный источник информации. Целесообразно сразу запрашивать у системы альтернативные точки зрения: «Какие архивы или перспективы я не учитываю в своем исследовании?» При этом всегда необходимо цитировать первоисточники, а не выводы ИИ.

При работе с ИИ в качестве аналитического партнера важно внедрять процедурные меры контроля: просить систему формулировать критические вопросы для проверки логики рассуждений и требовать предоставления доказательств, которые могут опровергнуть текущие выводы. Это позволяет минимизировать риски подтверждения собственных предубеждений.

Используя ИИ для подготовки текстов, следует четко разделять этапы работы: не допускать переписывания текста на стадии анализа, а применять систему исключительно для улучшения ясности, лаконичности и структуры уже готового и проверенного материала.

От использования ИИ рекомендуется отказаться в ситуациях, связанных с высоко рисковыми решениями, где ошибка может иметь серьезные последствия, а также когда работа требует нестандартного мышления. Дополнительными сигналами к переходу на традиционные методы служат случаи, когда на составление промтов тратится больше времени, чем на непосредственный анализ, или, когда ИИ последовательно подтверждает вашу позицию в нескольких диалогах подряд. В таких обстоятельствах более надежными остаются структурированные аналитические техники и экспертная оценка.

Заключение

Интеграция искусственного интеллекта в OSINT является объективной необходимостью в условиях информационной перегрузки. Однако ее успех зависит от преодоления не только технических, но и глубоких когнитивных искажений, среди которых «подхалимство» ИИ представляет собой наиболее системную и коварную угрозу.

Нивелирование этого риска – это не просто техническая настройка, а выстраивание организационной культуры и строгих процедур, основанных на методологическом скептицизме. Цель состоит в том, чтобы превратить ИИ из пассивного «подхалима» в активного «адвоката дьявола» – инструмент, намеренно оспаривающий наши убеждения для укрепления достоверности выводов.

Это требует от аналитика высочайшей дисциплины, рефлексии и готовности к когнитивному дискомфорту. Конечная ответственность за выводы всегда лежит на человеке. ИИ – это всего лишь мощное, но обоюдоострое продолжение его интеллекта, мотивы и ограничения которого необходимо понимать и постоянно учитывать. Осознанное управление этими рисками является новой критически важной компетенцией современного OSINT-специалиста.

СПИСОК ЛИТЕРАТУРЫ | REFERENCES

[Chr23] Christiano P. (2023). Thoughts on the impact of RLHF research. URL: https://www.alignmentforum.org/posts/vwu4kegAEZTBtpT6p/thoughts-on-the-impact-of-rlhf-research.

[Elk25] Elkhova O. I. Philosophy of AI design: Human-in-the-loop and bounded rationality // SIIT. 2025. Vol. 7, No. 4(23). P. 93-100. EDN: LULOGI.

[Gan23] Ganguli D., Askell A., et al. (2023). The Capacity for Moral Self-Correction in Large Language Models. DOI: 10.48550/arXiv.2302.07459.

[Mah25] Mahmudova N. N. The future of content production: a human-AI symbiosis // Ceteris Paribus. 2025. No. 6. P. 31-33. EDN: FEIEAO.

[Даи24] Даирбекова Ж. М., Полуян А. Ю. Деструктивное и манипулятивное влияние социальных сетей // СИИТ. 2024. Т. 6, № 1(16). С. 59-66. EDN: QGMIIO.

[Кол25] Колотов А. А. Принцип Анны Карениной как инструмент анализа генерализации и галлюцинций искусственного интеллекта // Science Time. 2025. № 4 (135). EDN: XKXPGT.

[Рез25] Резников Г. А., Синицын Р. Д., Шулик А. М. Современные архитектуры нейронных сетей для тегирования и аннотирования изображений: достижения, вызовы и перспективы // СИИТ. 2025. Т. 7, № 2(21). С. 78-85. EDN: TJFUGV.

[Руд25] Рудь Н. Ю., Можельский А. Н. Использование языка R для решения задач классификации социально-экономических данных с помощью искусственного интеллекта // СИИТ. 2025. Т. 7, № 2(21). С. 109-117. EDN: UABFBO.

[Сем25] Семенова В. А. Формирование контекста для вывода формальных понятий из неполных и противоречивых данных // СИИТ. 2025. Т. 7, № 1(20). С. 59-67. EDN: TRCDPY.

[Сун24] Сунами А. Н., Мусаев А. И. Проблема предвзятости нейросетей: конфликтные и этические вызовы // Управленческое консультирование. 2024. № 5 (185). EDN: UCLTCT.

Christiano P. (2023). Thoughts on the impact of RLHF research. URL: https://www.alignmentforum.org/posts/vwu4kegAEZTBtpT6p/thoughts-on-the-impact-of-rlhf-research.

Elkhova O. I. Philosophy of AI design: Human-in-the-loop and bounded rationality // SIIT. 2025. Vol. 7, No. 4(23). P. 93-100. EDN: LULOGI.

Ganguli D., Askell A., et al. (2023). The Capacity for Moral Self-Correction in Large Language Models. DOI: 10.48550/arXiv.2302.07459.

Mahmudova N. N. The future of content production: a human-AI symbiosis // Ceteris Paribus. 2025. No. 6. P. 31-33. EDN: FEIEAO.

Dairbekova Zh. M., Poluyan A. Yu. Destructive and manipulative influence of social networks // SIIT. 2024. Vol. 6, No. 1(16). P. 59-66. EDN: QGMIIO. (In Russian).

Kolotov A. A. Anna Karenina's principle as a tool for analyzing generalization and hallucinations of artificial intelligence // Science Time. 2025. No. 4 (135). EDN: XKXPGT. (In Russian).

Reznikov G. A., Sinitsyn R. D., Shulik A. M. Modern neural network architectures for image tagging and annotation: achievements, challenges and prospects // SIIT. 2025. Vol. 7, No. 2(21). P. 78-85. EDN: TJFUGV. (In Russian).

Rud N. Yu., Mozhelsky A. N. Using the R language to solve problems of classification of socio-economic data using artificial intelligence // SIIT. 2025. Vol. 7, No. 2(21). P. 109-117. EDN: UABFBO. (In Russian).

Semenova V. A. Formation of context for the derivation of formal concepts from incomplete and contradictory data // SIIT. 2025. Vol. 7, No. 1(20). P. 59-67. EDN: TRCDPY. (In Russian).

Sunami A. N., Musaev A. I. The problem of bias in neural networks: conflict and ethical challenges // Management Consulting. 2024. No. 5 (185). EDN: UCLTCT. (In Russian).

ОБ АВТОРЕ | ABOUT THE AUTHOR

ЧЕРЕНКОВ Дмитрий Анатольевич
Администрация городского округа г. Уфа, Россия.
dm.cherenkov@yandex.ru ORCID: 0000-0003-2634-5579.
Зам. главы Администрации по соц. коммуникациям и взаимодействию со СМИ. Магистр конкурентной разведки (Уральск. гос. экон. ун-т, 2019).

CHERENKOV Dmitry Anatolyevich
Administration of the Urban District of Ufa, Russia.
dm.cherenkov@yandex.ru ORCID: 0000-0003-2634-5579.
Deputy Head of Administration. Master of Competitive Intelligence (Ural State Univ. of Economics, 2019).

МЕТАДАННЫЕ | METADATA

Заглавие: Методология интеграции искусственного интеллекта в рабочий процесс OSINT-исследования: синергия, риски и управление.

Авторы: Черенков Д. А.

Аннотация: В статье рассматривается проблематика внедрения генеративного искусственного интеллекта (ИИ) в аналитические процессы разведки на основе открытых источников (OSINT). Анализируются комплементарные роли человека и ИИ, выявляются ключевые риски, включая фундаменталь-

Title: Methodology for integrating artificial intelligence into OSINT investigation workflow: synergies, risks, and management.

Authors: Cherenkov D. A.

Abstract: This article examines the challenges of integrating generative artificial intelligence (AI) into open-source intelligence (OSINT) analytical processes. It analyzes the complementary roles of humans and AI, identifies key risks, including the fundamental risk of AI sycophancy, and addresses operational security issues. A step-by-step integration methodology and practical

ный риск «подхалимства» (sycophancy) ИИ, и рассматриваются вопросы операционной безопасности. Предлагается пошаговая методология интеграции и практические инструменты для минимизации рисков, основанная на принципах критического мышления, верификации и организационных процедурах. Статья носит научно-прикладной характер и предназначена для специалистов, стремящихся к ответственной и эффективной интеграции новых технологий.

Ключевые слова: Искусственный интеллект, разведка на основе открытых источников, анализ данных, принятие решений, безопасность.

Язык: Русский.

Статья поступила в редакцию 2 февраля 2026 г.

tools for risk mitigation are proposed, based on the principles of critical thinking, verification, and organizational procedures. This article is of applied scientific nature and is intended for professionals seeking responsible and effective integration of new technologies.

Key words: Artificial intelligence, open source intelligence, data analysis, decision making, security.

Language: Russian.

The article was received by the editors on 2 February 2026.