ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И ТЕЛЕКОММУНИКАЦИИ
INFORMATION TECHNOLOGY AND TELECOMMUNICATIONS

УДК 004.094

DOI: 10.21822/2073-6185-2026-53-1-108-115 Оригинальная статья/ Original article

Оценка влияния рисков на безопасность систем искусственного интеллекта

И.И. Лившиц
Национальный исследовательский университет ИТМО,
197101, г. Санкт-Петербург, Кронверкский пр., д. 49, Россия

Резюме. **Цель**. Целью исследования является оценка влияния рисков на безопасность систем искусственного интеллекта (ИИ). Новая методика учитывает существующую научную и научно-практическую базу в области управления рисков, обеспечения информационной безопасности (ИБ) и стандартов в области ИИ. **Метод**. Представлена методика оценки влияния рисков ИБ на процессы безопасности систем ИИ, основанная на известных и новых аналитических методах исследования безопасности СлПО (стандарты ISO, IEC, ГОСТ). **Результат**. Сформулировано математическое описание исследования, суть которого состоит в оценке рисков ИБ для обеспечения заданного уровня безопасности рассматриваемой системы ИИ как функциональной подсистемы в ограничениях финальной полной системы СлПО (КИИ). Представлены результаты применения предложенного подхода к оценке рисков ИБ для современных и перспективных систем ИИ на базе международных стандартов ISO/IEC. **Вывод**. Перспективы практической реализации предложенной методики связаны с обеспечением объективности, точности и полноты количественной оценки рисков ИБ, что обеспечивает принятие взвешенного и экономически обоснованного решения по обеспечению безопасности систем ИИ. Полученные результаты могут быть применены экспертами при проектировании, оценке соответствия и оптимизации систем ИИ в составе полных систем СлПО (КИИ) в аспекте обеспечения ИБ.

Ключевые слова: искусственный интеллект, система, риск, оценка, аудит, стандарт, информационная безопасность, требования, сложный промышленный объект, критическая информационная инфраструктура

Для цитирования: И.И. Лившиц. Оценка влияния рисков на безопасность систем искусственного интеллекта. Вестник Дагестанского государственного технического университета. Технические науки. 2026; 53(1):108-115. DOI:10.21822/2073-6185-2026-53-1-108-115.

Assessing the impact of risks on the security of Artificial Intelligence systems

I.I. Livshits
National Research University ITMO,
49 Kronverksky Ave., St. Petersburg 197101, Russia

Abstract. Objective. The purpose of the study is to assess the impact of risks on the security of artificial intelligence (AI) systems. The new methodology takes into account the existing scientific and practical basis in the field of risk management, IT-security and standards in the AI fields. **Method.** This paper presents a methodology for assessing the impact of information security risks on AI system security processes, based on established and new analytical methods for studying software security (ISO, IEC, and GOST standards). **Result.** A mathematical description of the information security risk assessment is formulated to ensure a given level of security for the AI system under consideration as a functional subsystem within the constraints of the final complete software system (CII). The results of applying the proposed approach to assessing information security risks for AI systems based on international ISO/IEC standards are presented. **Conclusion.** The potential for implementing this methodology lies in ensuring the objectivity, accuracy, and completeness of quantitative information security risk assessments, which enables informed decision-making on ensuring the security of AI systems. The results can be applied

by experts in the design, compliance assessment, and optimization of AI systems within critical information infrastructure (CII) systems to ensure information security.

Keywords: artificial intelligence, system, risk, assessment, audit, standard, information security, requirements, complex industrial facilities, critical information infrastructure

For citation: I.I. Livshits. Assessing the impact of risks on the security of Artificial Intelligence systems. Herald of Daghestan State Technical University. Technical Sciences. 2026;53(1):108-115. (In Russ) DOI:10.21822/2073-6185-2026-53-1-108-115.

Введение. В настоящее время все оценки систем ИИ, активно предлагаемые на рынке, выполнены только по собственным методикам разработчиков, не имеют общепринятых и объективных критериев. Существующее положение не позволяет службам, ответственным за безопасное функционирование сложных промышленных объектов (СлПО) получить все необходимые данные для корректной оценки рисков ИБ. Это обстоятельство, в свою очередь, существенно затрудняет создание результативной и эффективной системы защиты различных СлПО, часть которых может относиться к объектам критической инфраструктуры (КИИ).

Оценка рисков ИБ не является тривиальной проблемой даже применительно к хорошо известным техническим системам. Можно отметить несколько методик оценки рисков (ISO 31000, ISO 31010, ISO 27005, NIST и пр.), каждая из которых дает различные наборы практических инструментов, алгоритмов и систем интерпретации результатов. Следует отметить, что даже методики идентификации, обработки и оценивания остаточных рисков существенно различаются. При рассмотрении современных СлПО, тем более с включением модулей ИИ различных разработчиков, данная проблема становится сложнее, поскольку новая программная компонента ИИ сложно формализуется.

Актуальность исследования обусловлена необходимостью обеспечения безопасности систем ИИ, разрабатываемых различными компаниями, как компонентов СлПО (КИИ); несовершенством существующих методов оценки рисков ИБ; отсутствием единых универсальных методов оценки соответствия для систем ИИ в составе СлПО (КИИ).

Постановка задачи. В данной публикации автор полагает обоснованным применение более широкого термина СлПО (сложный программный объект), поскольку для относительно простых программных компонентов разработаны и применяются системы верификации с различной степенью доверия, но для компонентов ИИ таких методов верификации пока не представлено. С ростом сложности требований к СлПО в аспекте обеспечения ИБ, применения принципиально новых типов программных компонентов (например, систем мультиагентов на базе ИИ), появления новых типов угроз безопасности информации (УБИ), актуальность разработки и реализации объективной и достоверной системы управления рисками ИБ будет возрастать.

Известно несколько подходов к оценке рисков ИБ в СлПО. В частности, можно отметить подход «Общих критериев» (представленный в стандартах ISO/IEC серии 15408), подход по процессам программной инженерии (представленный в стандартах ISO/IEC серии 12207) и подход систем менеджмента информационной безопасности (представленный в стандартах ISO/IEC серии 27001). Все указанные международные стандарты в РФ приняты в качестве национальных стандартов ГОСТ Р ИСО/МЭК и известны в программной индустрии. Однако для решения современных проблем оценки уровня безопасности для систем ИИ эти стандарты могут быть неэффективны по причине неформализованных компонентов ИИ, появления цепочек «поставщиков», неопределенности предмета рисков для ИИ и, реализации новых типов УБИ (например «галлюцинации» систем ИИ).

Методы исследования. С учетом известных требований к обеспечению безопасности объектов СлПО (КИИ), требуется определить математическое отображение, которое обеспечивает: наблюдение за множеством собственных параметров системы ИИ как объекта оценки; принятие во внимание множество базовых требований ИБ (требований

регуляторов); формирование выходных параметров для обеспечения заданного уровня ИБ через оптимизацию рисков, присущих конкретной реализации системы ИИ.

Для краткого описания данной проблемы целесообразно отметить публикации по тематике оценки влияния технологий ИИ на безопасность промышленных систем автоматизации [1 - 4], важности качественного образования в области ИБ с учетом современных факторов риска [5,6], методике расчета уровня полноты безопасности для сложных промышленных объектов [7- 9] и верификации данных для процессов цифровой трансформации [10].

Среди международных публикаций можно отметить примеры работ в области оценки рисков и влияния ИИ на процессы юриспруденции [11], финансов [12], бизнес-процессов [13], в сфере высшего образования [14] и в области машиностроения [15]. Определенно, можно говорить о всеобъемлющем влиянии современных систем ИИ практически на все отрасли науки, производства, транспорта, образования и пр. Рассмотрим несколько наиболее характерных примеров.

1. Испытание веб-инструментов на основе ИИ, применяемых для аудита смарт-контрактов [16]. Результаты показали, что ни одна из систем не продемонстрировала одновременно высокую точность и полноту выявления уязвимостей. Наилучший показатель точности не превысил 42%. Определено, что системы ИИ лучше справлялись с поверхностными и структурными ошибками (например, нарушениями авторизации или простыми логическими сбоями). Но сценарии, где нужно учитывать сложные взаимосвязи между функциями, оказались особенно проблемными. Эксперты полагаю, что высок риск ошибок систем ИИ в тех задачах, где требуются аналитические выводы, а не шаблонное сопоставление паттернов.
2. Тестирования минимального порога «отравления» систем ИИ (GPT-3.5 Turbo и Llama 3.1). Исследование Anthropic [16] показало, что достаточно всего 250 вредоносных документов, чтобы вызвать нарушения работы модели ИИ с 13 миллиардами параметров, т.е. всего 0,00016% от общего объёма обучающего корпуса.

Далее рассмотрим несколько доступных ресурсов, содержащих примеры оценки влияния рисков ИБ на безопасность систем ИИ:

1. Ресурс «ATLAS (Adversarial Threat Landscape for Artificial-Intelligence Systems)» [17] представляет обобщенные примеры оценивания некоторых функций для систем ИИ. Для GPT-2 Model Replication известно несколько УБИ, реализуемых злоумышленниками, обладающими достаточными техническими навыками. Допускается, что атакующие могут использовать систему GPT для вредоносных целей до того, как сообщество по безопасности ИИ будет готово к этому.
2. Ресурс NIST представляет документ «NIST AI Risk Management Framework (AI RMF 1.0) [18], в котором изложены примеры управления рисками в системах ИИ. Данный документ NIST определяет задачи по оценки воздействия ИИ, среди которых – изучение влияния систем ИИ и защищенность от несанкционированного доступа. В приложении «Appendix B» указано, что для систем ИИ следует принять во внимание требования «триады безопасности», в частности доступности и конфиденциальности. Для целей данной публикации важно, что в разделе терминологии рассматриваемый документ NIST примечателен подходом к учету рисков, в частности, в п.1.2.3 (Residual risk) указывается, что «документирование остаточных рисков потребует от поставщика системы полного учета рисков, связанных с внедрением продукта ИИ…». В п. 3.2 (Safety) указано, что «подходы к управлению рисками безопасности ИИ должны … соответствовать существующим отраслевым или прикладным рекомендациям или стандартам».

Таким образом, оценка влияния рисков ИБ на безопасность систем ИИ является актуальной и важной, поскольку затрагивает практически всю совокупность СлПО, в состав которых входят различные компоненты ИИ.

Обсуждение результатов. Рассмотрим характерные примеры реализации рисков ИБ, которые привели к компрометации различных систем ИИ:

- Опубликован факт компрометации многоуровневой системы защиты технологии eSIM [19], причем конкретная система прошла сертификацию по жестким требованиям ISO/IEC 15408 по классу EAL4+. Независимая лаборатория реализовала множество уязвимостей для технологии eSIM, что позволило реализовать риск полного доступа через критические уязвимости в конкретной системе.
- Опубликован факт компрометации специализированных ИИ-детекторов в издательстве Dispatch [20]. Базовый сценарий риска показал, что несколько текстов написаны человеком, однако при тщательном анализе выяснилось, что это все были вымышленные истории. Примечательно, что фальшивки (deep fake) прошли в несколько издательств (Dispatch, Wired и Press Gazette).
- Известный эксперимент Массачусетского Университета (MIT)[21] показал: даже при строгих инструкциях модели продолжают поддерживать искажённое восприятие, вплоть до одобрения самоубийственных идей.
- Пример «полевых» испытаний ИИ-системы распознавания лиц в Лондоне показал, что только 8 из 42 совпадений оказались достоверными [22], однако эта система распознавания все равно внедряется в аэропортах;
- Известны общие противоречия между «комфортными» лабораторными показателями (достигающими 99,95% точности) и реальными данными «полевой» практики. В отчете «Face Analysis Technology Evaluation (FATE)» [23] даются пояснения, что лабораторные тесты без валидации не подходят для оценки работы алгоритмов ИИ в реальной среде (на шумной улице, в условиях плохой освещённости и пр.).

Выполним сопоставление стандартов, применимых для оценки рисков ИБ в системах ИИ. Результаты краткого сопоставления представлены в табл.1

Таблица 1. Сопоставление стандартов для оценки рисков ИБ в системах ИИ
Table 1. Comparison of standards for assessing IT-security risks in AI systems

Международный стандарт	ISO 26000:2010	ISO/IEC TS 5723:2022	ISO/IEC 25010:2011	ISO/IEC 22989:2022	ISO/IEC TR 24368:2022	ISO/IEC 42001:2023	ISO/IEC 42005:2025	ISO/IEC 42006:2025
Наименование	Guidance on social responsibility	Trustworthiness — Vocabulary	Systems and software engineering — Systems and software Quality Requirements and Evaluation — System and software quality odels	Information technology – Artificial intelligence - Artificial intelligence concepts and terminology	Information technology — Artificial intelligence — Overview of ethical and societal concerns	Information technology — Artificial intelligence — Management system	Information technology — Artificial intelligence (AI) — AI system impact assessment	Information technology — Artificial intelligence — Requirements for bodies providing audit and certification of artificial intelligence management systems
Национальный стандарт РФ	ГОСТ Р ИСО 26000-2012	н/д	ГОСТ Р 59898-2021	ГОСТ Р 71476-2024	ПНСТ 840-2023	ГОСТ Р ИСО/МЭК 42001-2024	н/д	н/д
Наименование	Руководство по социальной ответственности		Оценка качества систем искусственного интеллекта. Общие положения	Искусственный интеллект. Концепции и терминология искусственного интеллекта	Искусственный интеллект. Обзор этических и общественных аспектов	Искусственный интеллект. Система менеджмента		
Примечание	Формируют образ надежного поставщика, заботящегося о благополучии общества и	Коррелирует с 27000 по «триаде»		Позволяет измерять уровень зрелости в вопросах кибербезопасности и проводить улучшение	Помогает компаниям избежать нарушений морали и права при применении ИИ, предлагая		Предоставляют возможность убедиться в правильности технических решений и снижении	Способствует снижению правовых рисков, предоставляя рекомендации по соблюдению норм и

Международный стандарт	ISO 26000:2010	ISO/IEC TS 5723:2022	ISO/IEC 25010:2011	ISO/IEC 22989:2022	ISO/IEC TR 24368:2022	ISO/IEC 42001:2023	ISO/IEC 42005:2025	ISO/IEC 42006:2025
	соблюдении законов.			системы защиты	критерии оценки этичности принимаемых решений		рисков систем ИИ	обеспечению безопасности
Преимущества	Хорошо сопрягается с ISO/IEC TR 24368:2022 для определения границ безопасности	Хорошо сопрягается со всеми стандартами ISO/IEC	Представляет показатели качества систем ИИ как метрики безопасности	Применимо для валидации данных и сравнения различных моделей ИИ	Применимо для оценивания профессиональной ответственности и понимания различного вреда от систем ИИ	Хорошо сопрягается с сертификационными стандартами ISO/IEC для интегрированных систем менеджмента	Применимо для оценивания ущерба от разумно предзаказанного неправильного использования систем ИИ	Применимо для выполнения аудита систем менеджмента ИИ

Произвольный программный компонент ИИ может быть включен в состав СлПО, который, в свою очередь, может пройти независимую оценку соответствия по различным стандартизованным подходам, указанным выше: ISO/IEC серии 15408, ISO/IEC серии 12207 и ISO/IEC серии 27001.

Представляется целесообразным принять во внимание систему требований на базе стандарта ГОСТ Р МЭК серии 61508 и цели безопасности, применимые для оценки рисков ИБ в системах ИИ, сгруппированные по основным признакам в табл.2

Таблица 2. Определение базовых требований и целей безопасности
Table 2. Defining Basic Safety Requirements and Objectives

Нормативная база Regulatory framework	**Пункт Paragraph**	**Цели безопасности Security goals**
ГОСТ Р МЭК 61508-1-2012	Таблица1; п.3	Определение последовательностей событий, приводящих к определенным опасным событиям. Определение рисков, связанных с определенными опасными событиями
	Таблица1; п.13	Подтвердить, что системы, связанные с безопасностью, отвечают спецификации требований к безопасности всей системы в терминах требований к функциям безопасности всей системы
	п.7.6.2.7	Системы управления, связанные с безопасностью, и другие меры по снижению риска должны рассматриваться как независимые: должны основываться на различных технологиях (должно использоваться оборудование различных видов для достижения одних и тех же результатов); не должны иметь общих процедур эксплуатации, технического обслуживания или тестирования.
	п. 7.18.2	Для каждой стадии ЖЦ всей системы безопасности, одновременно с разработкой плана этой стадии должен быть установлен план верификации. В плане верификации должны содержаться критерии, методы и средства, используемые при верификации
ГОСТ Р МЭК 61508-2- 2012	Таблица 1; п. 10.6	Тестировать и оценивать выходные результаты конкретной стадии, чтобы гарантировать их правильность и соответствие требованиям разделов стандартов, предусмотренных для этой стадии. Провести исследование и получить заключение по функциональной безопасности, достигнутой с системы, связанной с безопасностью
	п. 7.3	Планирование подтверждения соответствия безопасности системы: процедуры оценочных испытаний (с обоснованиями); процедуры испытаний и критерии, применяемые для подтверждения соответствия заданным пределам
	п. 7.4.11.2	Канал связи должен быть полностью разработан, реализован и для него проведена процедура подтверждения соответствия (так называемый «белый канал»).
	Таблица В.6	Проведение «Полевых испытаний». Уровень низкой эффективности: 10 тыс. часов эксплуатации; статистическая точность 95 %; отсутствие каких-либо критических отказов безопасности Уровень высокой эффективности: 10 млн. часов эксплуатации; статистическая точность 99,9 %; подробная документация всех изменений
ГОСТ Р МЭК 61508-3-2018	Таблица1, п.10.2	Планирование подтверждения соответствия безопасности системы. Соответствующий план верификации (зависит от стадии): самоконтроль ПО; периодическое тестирование функций безопасности во время выполнения программы; функции ПО для выполнения контрольных испытаний и всех диагностических тестов
	п. 7.4.7	Требования к тестированию ПО: разбиение системы на уровни интеграции; тестовые примеры и тестовые данные; типы выполняемых проверок; условия тестирования,

Нормативная база Regulatory framework	**Пункт Paragraph**	**Цели безопасности Security goals**
		включая инструменты, программы поддержи и описание конфигурации; условия, при которых проверка считается выполненной.
ГОСТ Р МЭК 61508-4-2012	п. 3.8.1	Верификация (verification)
	п. 3.8.2	Подтверждение соответствия (validation)
	п. 3.8.6	Охват диагностикой (diagnostic coverage)
ГОСТ Р МЭК 61508-5-2012	Таблица Е.2	Пример калибровки графа рисков общего назначения. Методы графа риска могут оказаться не лучшим решением задачи, если объект работает в непрерывном режиме.
ГОСТ Р МЭК 61508-6-2007	Таблица В.1	Параметры и диапазоны их значений (применяется к разным архитектурам), Диагностическое покрытие (DC) 0; 60; 90; 99
	Таблица В.2	Средняя вероятность отказа по запросу в течение 6 мес. интервала между контрольными проверками при среднем времени ремонта 8 ч
	Таблица В.9	Неидеальные контрольные испытания
	Таблица С.2	Уровни и диапазоны диагностического охвата различных компонентов (40 % - 60 %, Низкий диагностический охват). В настоящее время для подсистем, схемы высокого диагностического охвата которых отсутствуют, средства и методы высокой достоверности диагностики неизвестны
	Таблица Е.18	Строго рекомендуемые методы верификации ПО: Анализ влияния Повторная верификация измененных программных модулей Повторная верификация программных модулей, на которые оказывают влияние изменения в других модулях Повторная верификация системы в целом Управление конфигурацией программного обеспечения
ГОСТ Р МЭК 61508-7-2007	А.9.3	Логический контроль последовательности выполнения программ. Обеспечивается правильная последовательность выполнения отдельных частей программы с помощью встроенных (процедур учета, ключевых процедур) или внешних средств контроля.
	В.1.3	Предотвращение влияния систем, связанных с безопасностью, на системы, не связанные с безопасностью, в непредвиденных ситуациях.
	В.2.2	Формальные методы являются математическими моделями функции и/или структуры системы и обеспечивают однозначное описание системы.
	В.2.5	Таблицы контрольных проверок применяют на всех этапах полного ЖЦ безопасности ПО для оценки функциональной безопасности. Использование таблицы контрольных проверок зависит от экспертной оценки и суждения специалиста, который выбирает и применяет таблицу контрольных проверок. Принятые им решения относительно выбранных(ой) таблиц(ы) контрольных проверок должны быть документально оформлены и обоснованы. Если таблицы контрольных проверок пересматриваются, то гарантируется получение одних и тех же результатов, если только не используются различные критерии. Для документирования результатов каждого вопроса должен использоваться ответ «успешно», «неуспешно» или «не завершено», либо аналогичный набор ответов.

С учетом постановки задачи и всех аналитических материалов, сформируем математическое отображение:

$$\mathbf{R = [(X); (I); (R); (N); (G); (O)]}$$

где: (X) – множество параметров безопасности системы безопасности (система ИИ);
(I) – исходное множество параметров объекта оценки (система ИИ);
(R) – множество рисков, присущих системе безопасности;
(N) – множество базовых требований ИБ (стандарты, методики);
(G) – множество требований регуляторов (органов по оценке соответствия);
(O) – множество выходных параметров для обеспечения уровня ИБ.

Представленное математическое отображение обеспечивает:

1. Устойчивое наблюдение за множеством собственных параметров системы ИИ как объекта оценки (например, по базе ISO/IEC 15408 «Общие критерии»), что позволяет оптимально описать все требования к объекту оценки (политики безопасности, границы безопасности, функции безопасности и пр.);
2. Принятие во внимание совокупность входных параметров (базовый «движок», например GigaChat, Chat GPT, DeepSeek, и пр.);
3. Учет множества базовых требований ИБ (например, on-prem сервера или возможность «частного» корпоративного облака и пр.);

4. Учет требований регуляторов (ФСТЭК, Центральный банк, ФСБ и пр.);
5. Формирование выходных параметров для обеспечения заданного уровня ИБ через оптимизацию рисков, присущих конкретной реализации системы ИБ (например, количественная модель по базе ISO/IEC серии 27005 или 31000);
6. Формирование объективных и прослеживаемых требований для верификации и валидации систем ИИ (например, по базе ISO/IEC серии 61508);
7. Возможность независимой оценки соответствия систем ИИ по международным признанным стандартам (например, по базе ISO/IEC серии 42001, 42005 и 42006).

Вывод. Для обеспечения безопасности системы ИИ необходимо идентифицировать, оценивать и управлять рисками ИБ (в частности – доступности и качества всех данных, которые обрабатываются в конкретной системе ИИ).

Представлен подход к оценке рисков ИБ для систем ИИ, основанный на методах объективной оценки безопасности СлПО с учетом совокупности параметров (объекта оценки, требований ИБ, требований регуляторов и пр.).

Представленный подход также позволяет объективно оценить и другие риски ИБ (например, соблюдение политик безопасности в отношении чувствительных данных в конкретной системе ИИ).

Библиографический список:

1. Бурланков С.П., Семёнов К.О., Галактионова Е.В., Комаров В.А. Управление рисками как элемент экономической безопасности в компаниях, работающих в сфере цифровых технологий//Вестник Российского экономического университета имени Г.В. Плеханова. 2024. Т.21. № 6 (138). С. 123-130.
2. Демба С. Роль искусственного интеллекта в современной банковской системе // Вестник Российского экономического университета имени Г.В. Плеханова. 2024. Т. 21. № 3 (135). С. 164-172.
3. Сушкова И.А., Мамаева Л.Н. Искусственный интеллект в экономике и системе экономической безопасности // Вестник Российского экономического университета имени Г.В. Плеханова. 2023. Т. 20. № 4 (130). С. 44-53.
4. Лившиц И.И. Влияние современных технологий искусственного интеллекта на безопасность промышленных систем автоматизации // Автоматизация в промышленности. 2025. № 6. С. 34-37.
5. Лившиц И.И. Оценка необходимости совершенствования действующего порядка подготовки квалифицированных кадров в области информационной безопасности // Газовая промышленность. 2024. № 9 (871). С. 200-205
6. Лившиц И.И. Анализ процесса подготовки специалистов в области информационной безопасности // Автоматизация в промышленности. 2023. № 9. С. 56-60.
7. Конаков А.М., Лившиц И.И. Поиск оптимального пути построения системы защиты информации на основе марковских цепей // Вестник Дагестанского государственного технического университета. Технические науки. 2024. Т. 51. № 3. С. 86-92.
8. Лившиц И.И., Сунцова Д.И. Методика расчета уровня полноты безопасности для сложных промышленных объектов топливно-энергетического комплекса // Энергобезопасность и энергосбережение. 2024. № 1. С. 5-12.
9. Лившиц И.И., Понаморева К.А. Формирование требований к методике оценки рисков для компонентов АСУТП // Энергобезопасность и энергосбережение. 2024. № 2. С. 5-13.
10. Лившиц И.И. Верификация данных для процессов цифровой трансформации // Информационно-экономические аспекты стандартизации и технического регулирования. 2024. № 6 (81). С. 240-245.
11. Fine A., Le S., Miller M.K. Content analysis of judges' sentiments toward artificial intelligence risk assessment tools. Criminology, Criminal Justice, Law and Society. 2023. Т. 24. № 2. С. 31-46.
12. Tao C., Liu Y. Application and development of artificial intelligence risk control in internet finance. Frontiers in Business, Economics and Management. 2024. Т. 14. № 2. С. 10-12.
13. Muria-Tarazón Ju.C., Oltra-Gutiérrez Ju.V., Oltra-Badenes R., Escobar-Román S. Uncovering research trends on artificial intelligence risk assessment in businesses: a state-of-the-art perspective using bibliometric analysis. Applied Sciences (Switzerland). 2025. Т. 15. № 3. С. 1412.
14. Schaeffer D., Coombs L., Luckett J., Marin M., Olson P. Risks of AI applications used in higher education. Electronic Journal of e-Learning. 2024. Т. 22. № 6. С. 60-65.
15. Shin H.S., Choi Su.B., Kim J.W. Harnessing highly efficient triboelectric sensors and machine learning for self-powered intelligent security applications. Materials Today Advances. 2023. Т. 20. С. 100426.
16. https://www.securitylab.ru/news/564486.php
17. https://atlas.mitre.org/
18. https://www.nist.gov/itl/ai-risk-management-framework
19. https://www.securitylab.ru/news/561256.php

20. https://www.cnews.ru/news/top/2025-08-25_wired_i_business_insider_popalis_na
21. https://arxiv.org/abs/2504.18412
22. https://www.securitylab.ru/news/562549.php
23. https://pages.nist.gov/frvt/reports/morph/fate_morph_4B_NISTIR_8584.pdf

References:

1. Burlankov S.P., Semenov K.O., Galaktionova E.V., Komarov V.A. Risk management as an element of economic security in companies operating in the field of digital technologies. *Bulletin of the Plekhanov Russian University of Economics.* 2024;21(6)(138):123-130. (In Russ)
2. Demba S. The role of artificial intelligence in the modern banking system. *Bulletin of the Plekhanov Russian University of Economics*. 2024; 21(3 (135):164-172. (In Russ)
3. Sushkova I.A., Mamaeva L.N. Artificial intelligence in the economy and the system of economic security. *Bulletin of the Plekhanov Russian University of Economics.* 2023; 20(4) (130):44-53. (In Russ)
4. Livshits I.I. The Impact of Modern Artificial Intelligence Technologies on the Security of Industrial Automation Systems. *Automation in Industry*. 2025;6:34-37. (In Russ)
5. Livshits I.I. Assessment of the Need to Improve the Current Procedure for Training Qualified Personnel in the Field of Information Securit. *Gas Industr*y. 2024;9 (871): 200-205. (In Russ)
6. Livshits I.I. Analysis of the Process of Training Specialists in the Field of Information Security. *Automation in Industry.* 2023; 9:56-60. (In Russ)
7. Konakov A.M., Livshits I.I. Search for an Optimal Way to Construct an Information Security System Based on Markov Chains. *Herald of Dagestan State Technical University. Technical Sciences*. 2024;51(3): 86-92. (In Russ)
8. Livshits I.I., Suntsova D.I. Methodology for Calculating the Safety Integrity Level for Complex Industrial Facilities in the Fuel and Energy Complex. *Energy Safety and Energy Saving*. 2024;1:5-12. (In Russ)
9. Livshits I.I., Ponomoreva K.A. Formation of Requirements for the Risk Assessment Methodology for APCS Components. *Energy Safety and Energy Saving.* 2024;2:5-13. (In Russ)
10. Livshits I.I. Data Verification for Digital Transformation Processes. *Information and Economic Aspects of Standardization and Technical Regulation*. 2024;6 (81):240-245. (In Russ)
11. Fine A., Le S., Miller M.K. Content analysis of judges' sentiments toward artificial intelligence risk assessment tools. *Criminology, Criminal Justice, Law and Society*. 2023;24(2):31-46.
12. Tao C., Liu Y. Application and development of artificial intelligence risk control in internet finance. *Frontiers in Business, Economics and Management.* 2024;14(2):10-12.
13. Muria-Tarazón Ju.C., Oltra-Gutiérrez Ju.V., Oltra-Badenes R., Escobar-Román S. Uncovering research trends on artificial intelligence risk assessment in businesses: a state-of-the-art perspective using bibliometric analysis. Applied Sciences (Switzerland). 2025;15(3):1412.
14. Schaeffer D., Coombs L., Luckett J., Marin M., Olson P. Risks of AI applications used in higher education. Electronic Journal of e-Learning. 2024; 22(6):60-65.
15. Shin H.S., Choi Su.B., Kim J.W. Harnessing highly efficient triboelectric sensors and machine learning for self-powered intelligent security applications. *Materials Today Advances*. 2023;20:100426.
16. https://www.securitylab.ru/news/564486.php
17. https://atlas.mitre.org/
18. https://www.nist.gov/itl/ai-risk-management-framework
19. https://www.securitylab.ru/news/561256.php
20. https://www.cnews.ru/news/top/2025-08-25_wired_i_business_insider_popalis_na
21. https://arxiv.org/abs/2504.18412
22. https://www.securitylab.ru/news/562549.php
23. https://pages.nist.gov/frvt/reports/morph/fate_morph_4B_NISTIR_8584.pdf

Сведения об авторе:
Лившиц Илья Иосифович, доктор технических наук, профессор практики, Livshitz.i@yandex.ru
Information about author:
Ilya I. Livshits, Dr. Sci. (Eng.), Prof. of Practice; Livshitz.i@yandex.ru

Конфликт интересов/Conflict of interest.
Автор заявляет об отсутствии конфликта интересов/The author declare no conflict of interest.
Поступила в редакцию/Received 16.10.2025**.**
Одобрена после рецензирования/Reviced 30.11.2025.
Принята в печать/Accepted for publication 17.01.2026.