УДК 004.8
doi: 10.21685/2307-4205-2025-1-3

ГЕНЕЗИС РИСКОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И ИХ РЕАЛИЗАЦИЯ В АВТОНОМНЫХ СИСТЕМАХ ОТВЕТСТВЕННОГО НАЗНАЧЕНИЯ

О. В. Прокофьев

Пензенский государственный технологический университет, Пенза, Россия
Prokof_ow@mail.ru

Аннотация. *Актуальность и цели*. Многие исследования искусственного интеллекта с момента его создания посвящены изучению множества проблем и подходов для автономной работы в областях применения, связанных со здоровьем и жизнью человека. Представлено описание происхождения рисков, систематизации способов их проявления для построения систем ответственного назначения. *Материалы и методы.* Поскольку исследования в этой области пока не опираются на большой опыт внедрения, обсуждаемые источники рисков являются обобщением экспертных оценок авторитетных разработчиков и приведены на примере систем автономных вооружений. *Результаты.* Выявлены основные источники возникновения рисков, траектории их развития для различных вариантов причинно-следственной трансформации. *Выводы.* Сформулированы направления совершенствования процессов тестирования и отладки автономных систем ответственного назначения с точки зрения развития и формирования рисков применения.

Ключевые слова: риски искусственного интеллекта, генезис рисков, автономные системы, системы ответственного назначения

Для цитирования: Прокофьев О. В. Генезис рисков искусственного интеллекта и их реализация в автономных системах ответственного назначения // Надежность и качество сложных систем. 2025. № 1. С. 20–27. doi: 10.21685/2307-4205-2025-1-3

GENESIS OF RISKS OF ARTIFICIAL INTELLIGENCE AND THEIR IMPLEMENTATION IN AUTONOMOUS SYSTEMS FOR RESPONSIBLE PURPOSE

O.V. Prokofiev

Penza State Technological University, Penza, Russia
Prokof_ow@mail.ru

Abstract. *Background.* Much research on artificial intelligence (AI) since its inception has been devoted to exploring a variety of problems and approaches for autonomous operation in application areas related to human health and life. The article is devoted to describing the origin of risks, systematizing the ways they manifest themselves in order to build systems for responsible purposes. *Materials and methods.* Since research in this area is not yet based on extensive implementation experience, the sources of risks discussed are a generalization of expert assessments of authoritative developers and are given as an example of autonomous weapons systems. *Results.* The main sources of risks and the trajectories of their development for various variants of cause-and-effect transformation have been identified. *Conclusions.* Directions for improving the processes of testing and debugging autonomous systems for critical purposes are formulated from the point of view of development and formation of application risks.

Keywords: risks of artificial intelligence, genesis of risks, autonomous systems, critical systems

For citation: Prokofiev O.V. Genesis of risks of artificial intelligence and their implementation in autonomous systems for responsible purpose. *Nadezhnost' i kachestvo slozhnykh sistem = Reliability and quality of complex systems.* 2025;(1):20–27. (In Russ.). doi: 10.21685/2307-4205-2025-1-3

Введение

Значительные достижения в области искусственного интеллекта (ИИ), наблюдаемые за последние годы, одновременно сопровождаются обоснованным ростом обеспокоенности о последствиях решений, принимаемых ИИ. Высокая скорость принятия решений, удешевление автономных систем по мере масштабирования производства, возможность собирать и обрабатывать большие объемы

информации привлекают будущих пользователей и бенефициаров внедрения систем. В ряде случаев предполагаемые пользователи высказывают мнение о том, что ИИ позволяет принять справедливые, объективные решения, свободные от влияния эмоций и неуверенности. В то же время управление автономными системами в транспорте, медицине, вооружениях несет в себе потенциальные риски, полный перечень которых невозможно предвосхитить. Вопросы этики, права, границ допустимых решений, принимаемых ИИ в автономных системах ответственного назначения, сегодня переходят из области теории в практическую область организации машинного обучения. Системное изложение происхождения рисков имеет важное применение с точки зрения тестирования и отладки автономных систем, особенно имеющих прямое отношение к здоровью и жизни человека. В данной статье в качестве примера выбрана предметная область автономных систем вооружений, наиболее значимая с точки зрения возможного ущерба. Быстрое развитие автономных систем с ИИ, закрытый статус многих разработок и непрозрачность процесса формирования критически важного решения в самих системах – это далеко не полный перечень причин, из-за которых данные в области исследования собираются на основе опроса экспертов. В статье использованы числовые и порядковые результаты опроса респондентов, качественные оценки. Необходимо междисциплинарное исследование безопасности ИИ в диапазоне от экономики, права и философии до компьютерной безопасности, формальных методов логики и, конечно же, прикладных дисциплин для реализации ИИ. Целью работы стало исследование генезиса рисков ИИ на примере систем вооружения, систематизация и описание траекторий формирования потенциальных угроз.

Методология

Обобщая качественные оценки экспертами преимуществ автономных систем вооружений [1–12], были выявлены следующие конкретные и ожидаемые результаты:

1. Скорость, внезапность и живучесть. Например, автономный самолет может быть более искусным в выявлении и предотвращении угроз ПВО, например, лучше прогнозировать и побеждать противников в воздушном бою, что делает его более способным выполнить свою миссию.

2. Повышенная точность и уменьшенный побочный ущерб. Благодаря большей вычислительной мощности, лучшему доступу к информации и большей скорости, чем у людей, автономное оружие может более точно и быстро рассчитывать и предпринимать действия, которые повышают точность и минимизируют сопутствующий ущерб.

3. Машинное соблюдение правил и закона. В отличие от людей, ИИ может быть обучен для точного выполнения инструкций по ведению боя, а также он избавлен от действий, выполняемых на основе человеческих эмоций.

4. Уменьшение ущерба гражданскому населению. Будущие автономные системы могут превзойти людей в различении комбатантов от гражданских лиц, тем самым уменьшая жертвы среди гражданского населения на войне.

5. Удаление людей с поля боя. Будущие войны с участием необитаемых машин сберегут много человеческих жизней. Например, роботы по разминированию уже спасают человеческие жизни. Опасная работа делегируется агентам, не являющимся людьми.

Кроме того, в исследовании были использованы экспертные оценки института SIPRI [1–11], а также результаты статистического исследования корпорации вооружений RAND в форме опроса, проведенного среди 29 специалистов по военному применению ИИ [12]. Примеры высказанных экспертами озабоченностей о последствиях применения автономных вооружений представлены на рис. 1, 2.

Рис. 1. Автономное оружие чаще совершает ошибки, чем люди (% опрошенных)

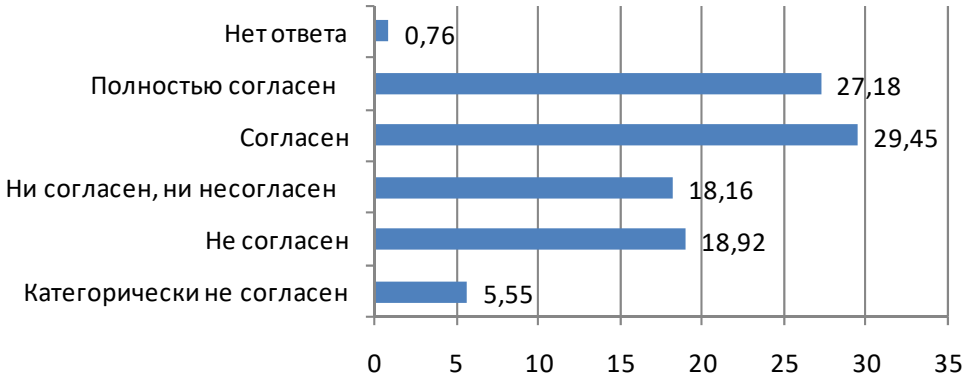

Рис. 2. Автономное оружие должно быть запрещено с этической точки зрения, поскольку его нельзя привлечь к ответственности за неправомерные действия (% опрошенных)

Результаты

Автономные системы вооружения действуют как «мультипликатор угрозы», активируя пути для различных рисков, как показано на рис. 3. Эти пути не исчерпывают весь спектр рисков, но помогают проиллюстрировать различные способы, которыми ученые и эксперты по этому вопросу в настоящее время обоснованно показывают, что автономное оружие может представлять серьезную угрозу для человечества.

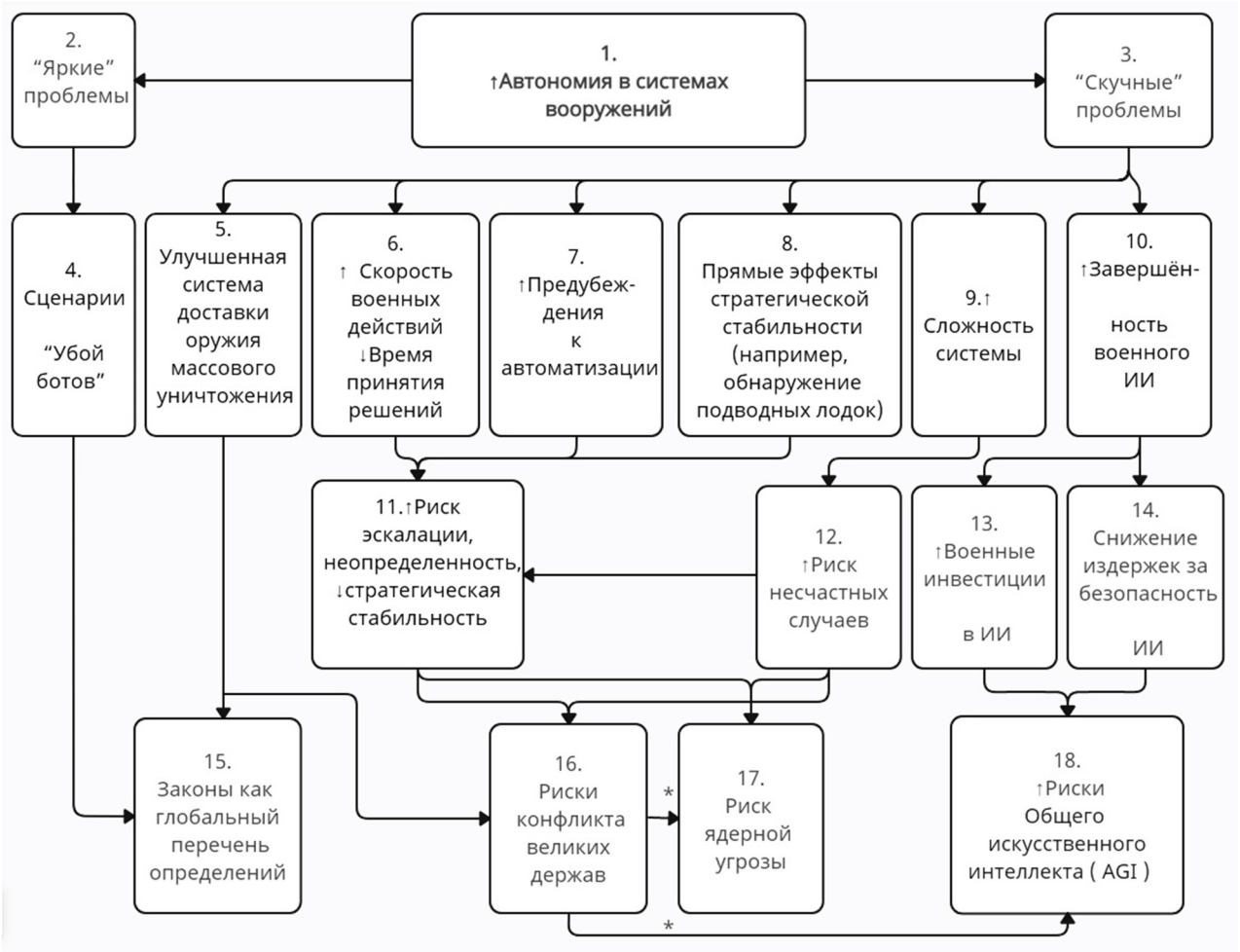

Рис. 3. Сценарии (траектории) происхождения рисков

Траектория I: «Боевой робот». Относится к «ярким» проблемам, которым уделяется гораздо больше внимания в средствах массовой информации. Соответствует путь обхода иерархической

структуры на рис. 3, по блокам с номерами 1, 2, 4, 15. В работе [13] утверждается, что боевые роботы с роевым алгоритмом поведения являются наиболее экономичным оружием массового поражения. Соотношение «Затраты/Эффективность» при массовом производстве приближенно автономных устройств оценивается как стоимость одного погибшего $400 (2020 г.), стоимость гибели 100 000 человек (сопоставимая с ядерным ударом по городу среднего размера) составит 40 миллионов долларов и, как ожидается, в будущем снизится. Покупка относительно дешевых подвижных устройств и программного обеспечения (например, алгоритмов роения и программного обеспечения для распознавания лиц) может позволить некоторым субъектам компенсировать военную слабость, дестабилизируя международную систему. Эта угроза, как и далее перечисленные, заслуживает не только разработки технических мер противодействия (например, блокирования коммуникаций роя), но и правовой, этической экспертизы.

Траектория II: «Системы доставки оружия массового поражения (ОМП)». Сценарий описывается блоками с номерами 1, 3, 5, 15, 16, 17, 18. Автономные транспортные средства – это эффективный способ доставки химического, радиологического и биологического оружия. Они могут более точно нацеливаться на врагов, корректируя траектории в зависимости от местных условий ветра и влажности. Кроме того, небольшие устройства могут подниматься в воздух, летать внутри помещений и совместно совершать атаки. Системы доставки могут давать дестабилизирующие последствия применительно к доставке ядерного оружия. Они более склонны к потере контроля со стороны человека из-за неисправности, взлома или подделки, что может привести к случайному использованию ядерного оружия. Автономные системы более склонны к непредсказуемому поведению или сбоям, взлому или спуфинг-атакам, тогда может возникнуть риск несчастных случаев и непреднамеренной эскалации ситуации. Программируемые автономные системы коммерчески доступны и, таким образом, представляют собой легкодоступный метод, который не подвергает опасности собственно боевиков террористической группы при проведении радиологической, биологической или химической атаки. Для полного понимания рисков, связанных с доставкой оружия массового уничтожения, необходимы дополнительные исследования, следует изучить стратегические риски и проявления потенциального вмешательства.

Траектория III: «Война на скорости машины». Сценарий описывается блоками с номерами 1, 3, 7, 11, 16, 17, 18. Термин «скорость машины» используется потому, что автономные системы могут обрабатывать информацию и принимать решения быстрее, чем люди. Это может сократить время принятия решений, что, в свою очередь, может увеличить риски несчастных случаев и непреднамеренной эскалации как на поле боя, так и при принятии стратегических решений. Преимущества скорости стимулируют принятие решений великими державами до такой степени, что военные теоретики предсказывают новый вид войны, характеризующийся экстремальными скоростями, неподвластными человеческому контролю. Это известно как «гипервойна» или «сингулярность поля боя». Известны исторические факты, когда поспешные решения, подсказываемые автоматизированной системой, могли послужить началом ядерной войны. Если автоматизация и системы раннего предупреждения с поддержкой искусственного интеллекта могут сократить время, необходимое для обнаружения входящей атаки (до принятия решения), и сократить время, необходимое для начала атаки (после принятия решения), то время принятия решения, обращения к здравому смыслу может быть увеличено.

Траектория IV: «Предвзятость к автоматизации». Сценарий описывается блоками с номерами 1, 3, 7, 11, 16, 17, 18. В парах человек-машина люди, скорее всего, будут предвзято относиться к суждениям машины, даже при наличии противоречивых доказательств. Другими словами, предвзятость автоматизации определяется как тенденция «чрезмерно воспринимать компьютерную информацию как эвристическую замену бдительного поиска и обработки информации». При ненадежной системе некоторые люди склонны одобрять рекомендации, сгенерированные компьютером. В мире с более автономными системами люди могут быть предвзяты в отношении обнаружения ложных тревог и ошибок в смертоносных системах с поддержкой ИИ, включая «более умные» системы раннего предупреждения. К сожалению, создание более сложных и менее подверженных ошибкам систем искусственного интеллекта может не решить проблему предвзятости автоматизации, а может даже усугубить ее. Высокий уровень точности системы может непреднамеренно способствовать систематической ошибке автоматизации. Это может быть связано с тем, что точность порождает доверие, и было показано, что пользователи, которые больше доверяют автоматизации, с меньшей вероятностью обнаруживают сбои в автоматизации. По этим причинам предвзятость к автоматизации может оказаться одним из самых коварных «скучных» рисков увеличения автономности систем вооружения и автоматизации в целом.

Совершают ли пары человек-машина меньше ошибок в критических ситуациях, чем люди в одиночку? Ответы на такие риторические вопросы будут иметь большую информационную ценность.

Траектория V: «Прямые последствия стратегической стабильности». Сценарий описывается блоками с номерами 1, 3, 8, 11, 16, 17, 18 и характеризуется опасностью непреднамеренной эскалации. Существует потенциальная возможность того, что по результатам автономного дистанционного зондирования и распределенного поиска будет использовано автономное вооружение. Достижения в области искусственного интеллекта могут вызвать изменения в ядерных доктринах государств, допускающие автоматизацию запусков ядерного оружия.

Траектория VI: «Риски, связанные со сложностью систем». Сценарий описывается блоками с номерами 1, 3, 9, 12, 11, 16, 17, 18. Теория обычных аварий системы предполагает, что по мере увеличения сложности системы увеличивается и риск аварий и что некоторый уровень аварий неизбежен в сложных системах, кроме того, риск аварий в автономном оружии, поэтому системы могут быть выше. Алгоритмический хаос и неожиданная эскалация будут иметь катастрофические последствия для автономных систем на войне. Например, один рой подвижных устройств может неожиданно ошибочно истолковать микродвижения роя противника как «возможно враждебные», принять немного более оборонительную последовательность, которую второй рой может затем интерпретировать как «возможно враждебные», занять более враждебную позицию, в свою очередь, только подтверждая интерпретацию первого роя как «враждебную» и т.д., развиваясь по спирали в сторону эскалации. Похожая динамика может наблюдаться и на других уровнях конфликта, включая принятие стратегических решений государствами с помощью систем поддержки принятия решений на базе искусственного интеллекта.

Траектория VII: «Конкурс военного ИИ». Сценарий описывается блоками с номерами 1, 3, 10, 13, 14, 18. У государств есть стимулы, например, желание показаться технологически продвинутыми как противникам, так и внутренней аудитории – делать вид, что система имеет более высокую степень автономии, чем она есть на самом деле. Государство может интерпретировать «фальшивую» автономную систему противника как свидетельство реального превосходства и в результате принять решение инвестировать больше средств в сам военный ИИ. Таким образом, государство может спровоцировать динамику, подобную гонке вооружений. Можно столкнуться с конкурентом, который (по предположениям) с большей готовностью делегирует полномочия машинам, и по мере того, как эта конкуренция развивается, придется принимать соответствующие решения.

Обсуждение

Общественное мнение склоняется к тому, что автономные системы с ИИ могут принести беспрецедентную пользу, и поэтому стоит изучить, как максимизировать эти выгоды, избегая при этом потенциальных ловушек. Перечисленные закономерности в процессе генезиса рисков вовсе не означают того, что общественность однозначно будет добиваться всеобщего запрета этих систем. Приведем ряд доводов о практической нереализуемости формального полного запрета.

1. Запреты на тот или иной вид оружия соблюдаются при наличии сомнений в эффективности или присутствии вредных последствий, побочных эффектов. Остальные запреты в странах мира зачастую нарушаются.

2. Запрет на программное обеспечение для автономных систем сложно проверить. Принципиально, что отличает автономное оружие от дистанционно пилотируемого устройства, так это программное обеспечение; обе системы «выглядят» одинаково и могут вести себя (вне боевой задачи) одинаково, что создает большие трудности для проверки на основе договоров.

3. Граница между автономными и автоматизированными системами размыта, терминология в этой области вызывает дискуссии [13] и постановка задачи недостаточно четко сформулирована.

4. ИИ используется в разных областях. В отличие от некоторых запрещенных военных технологий, таких как наземные мины, ИИ представляет собой «технологию общего назначения», которую коммерческие и другие частные субъекты интегрируют в современную жизнь. Это усложняет контроль за его использованием.

5. Автономные системы обеспечивают военные преимущества. Автономия и применение ИИ на поле боя могут представляться выгодными в военном отношении. Исторически сложилось, что запреты лучше всего работали в отношении оружия, которое не считалось решающим [13].

6. Страны мира, располагающие наиболее мощными армиями, не выступали на международных площадках с идеей полного запрета автономных систем вооружений.

Заключение

На оперативном уровне компьютерные исследования для создания надежного ИИ должны продолжаться по направлению верификации: как доказать, что система удовлетворяет определенным желаемым формальным свойствам («Правильно ли я построил систему?») [14–17]. На уровне стратегического управления стоит задача построения сценария (прогноза) для случая комплексного применения таких систем. Имитационное моделирование (квазиэкспериментальный метод, «варгеймминг») используется для статистического оценивания сценариев уже сегодня и имеет перспективу применения для моделирования поведения автономных систем ответственного назначения. Методы количественного анализа позволяют перейти к оценке качества принимаемых решений, поможет ответить на ряд вопросов. Насколько серьезным будет влияние автоматизации в боевых ситуациях с автономным оружием? Как скорость моделируемой войны влияет на качество решений при использовании автономных систем? Является ли эскалация более или менее вероятной, когда задействованные системы являются автономными? Насколько доверяют участники необоснованным заявлениям о том, что противник развернул летальные автономные системы? Например, в исследовании [13] выявили 53 аналитических центра и другие организации, чья работа, вероятно, имеет отношение к исследованиям автономного оружия, но только девять активно работают над стратегическими рисками, связанными с автономией. В данном случае экономические затраты на исследования в области снижения рисков означают высокую отдачу в виде спасенных жизней людей в будущем, что подтверждается независимыми исследователями и международными организациями [18–27].

Список литературы

1. Boulanin V., Saalman L., Topychkanov P. [et al.]. Artificial Intelligence, Strategic Stability and Nuclear Risk. 2020. URL: https://www.sipri.org/sites/default/files/2020-06/artificial_intelligence_strategic_stability_and_nuclear_risk.pdf
2. Integrating Cybersecurity and Critical Infrastructure. National, Regional and International Approaches / ed. by L. Saalman. 2018. URL: https://www.sipri.org/sites/default/files/2018-04/integrating_cybersecurity_0.pdf
3. The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk. Vol. I. Euro-Atlantic Perspectives / ed. by V. Boulanin. 2020. URL: https://www.sipri.org/publications/2020/other-publications/artificial-intelligence-strategic-stability-and-nuclear-risk
4. Boulanin V., Verbruggen M. Mapping the Development of Autonomy in Weapon Systems. 2020. URL: https://www.sipri.org/sites/default/files/2018-04/integrating_cybersecurity_0.pdf
5. Boulanin V., Bruun L., Goussac N. Autonomous Weapon Systems And International Humanitarian Law. Identifying Limits and the Required Type and Degree of Human–Machine Interaction. 2021. URL: https://www.sipri.org/sites/default/files/2021-06/2106_aws_and_ihl_0.pdf
6. Saalman L., Su F., Saveleva Dovgal L. Cyber Posture Trends in China, Russia, the United States and the European Union. 2022. URL: https://www.sipri.org/sites/default/files/2022-12/2212_cyber_postures_0.pdf
7. Boulanin V. Mapping the development of autonomy in weapon systems. A primer on autonomy. 2017. URL: https://www.sipri.org/sites/default/files/Mapping-development-autonomy-in-weapon-systems.pdf
8. Boulanin V., Goussac N., Bruun L., Richards L. Responsible Military Use of Artificial Intelligence. Can the European Union Lead the Way in Developing Best Practice? 2020. URL: https://www.sipri.org/publications/2020/other-publications/responsible-military-use-artificial-intelligence-can-european-union-lead-way-developing-best
9. Boulanin V., Brockmann K., Richards L. Responsible Artificial Intelligence Research And Innovation For International Peace And Security. 2020. URL: https://www.sipri.org/sites/default/files/2020-11/sipri_report_responsible_artificial_intelligence_research_and_innovation_for_international_peace_and_security_2011.pdf
10. Bromley M., Maletta G. The Challenge of Software and Technology Transfers to Non-Proliferation Efforts. Implementing and Complying with Export Controls. 2018. URL: https://www.sipri.org/publications/2018/other-publications/challenge-software-and-technology-transfers-non-proliferation-efforts-implementing-and-complying
11. Su F., Boulanin V., Turell J. Cyber-incident Management Identifying and Dealing with the Risk of Escalation. 2020. IPRI Policy Paper No. 55. URL: https://www.sipri.org/publications/2020/sipri-policy-papers/cyber-incident-management-identifying-and-dealing-risk-escalation
12. Morgan F. E., Boudreaux B., Lohn A. J. [et al.]. Military Applications of Artificial Intelligence Ethical Concerns in an Uncertain World. RAND Corporation, 2020. 202 p. URL: https://www.rand.org/pubs/research_reports/RR3139-1.html
13. Ruhl Ch. Autonomous weapon systems and military artificial intelligence (AI) applications report. 2022. URL: https://www.founderspledge.com/research/autonomous-weapon-systems-and-military-artificial-intelligence-ai
14. Research Priorities for Robust and Beneficial Artificial Intelligence: An Open Letter. Future of Life Institute. URL: https://futureoflife.org/data/documents/research_priorities.pdf (дата обращения: 09.03.2024).

15. Marr B. The 15 Biggest Risks Of Artificial Intelligence. URL: https://www.forbes.com/sites/bernardmarr/2023/06/02/the-15-biggest-risks-of-artificial-intelligence/?sh=20c095a92706
16. Pause Giant AI Experiments: An Open Letter. URL: https://futureoflife.org/open-letter/pause-giant-ai-experiments/
17. AI Ethics Code. URL: https://ethics.a-ai.ru/ (дата обращения: 09.03.2024).
18. Delaborde A. Risk assessment of artificial intelligence in autonomous machines. 1st international workshop on Evaluating Progress in Artificial Intelligence (EPAI 2020) in conjunction with ECAI. URL: 2020. https://hal.science/hal-03009978
19. ICRC position on autonomous weapon systems. International Committee of the Red Cross. URL: https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems
20. Macrae C. Learning from the Failure of Autonomous and Intelligent Systems: Accidents, Safety and Sociotechnical Sources of Risk // SSRN Electronic Journal. 2021. doi: 10.2139/ssrn.3832621
21. Hindriks F., Veluwenkamp H. The risks of autonomous machines: from responsibility gaps to control gaps // Synthese. 2023. Vol. 201. doi: 10.1007/s11229-022-04001-5
22. Radanliev P., De Roure D., Maple C. [et al.]. Super forecasting the technological singularity risks from artificial intelligence // Evolving Systems. 2022. Vol. 13. P. 747–757. doi: 10.1007/s12530-022-09431-7
23. The risks of Lethal Autonomous Weapons. The Future of Life Institute. URL: https://autonomousweapons.org/the-risks
24. Asilomar AI Principles. The Future of Life Institute. URL: https://futureoflife.org/open-letter/ai-principles/
25. Artificial Intelligence Risk & Governance. By Artificial Intelligence/Machine Learning Risk & Security Working Group (AIRS). The Wharton School, The University of Pennsylvania, 2024. URL: https://ai.wharton.upenn.edu/white-paper/artificial-intelligence-risk-governance/
26. Skelton S. K. AI experts question tech industry's ethical commitments. TechTarget, 2024. https://www.computerweekly.com/feature/AI-experts-question-tech-industrys-ethical-commitments
27. Иванов А. И., Иванов А. П., Савинов К. Н., Еременко Р. В. Виртуальное усиление эффекта распараллеливания вычислений при переходе от бинарных нейронов к использованию q-арных искусственных нейронов // Надежность и качество сложных систем. 2022. № 4. С. 89–97.
28. Ширинкина Е. В. Механизм применения искусственного интеллекта в обучении // Надежность и качество сложных систем. 2022. № 4. С. 24–30.

References

1. Boulanin V., Saalman L., Topychkanov P. et al. *Artificial Intelligence, Strategic Stability and Nuclear Risk*. 2020. Available at: https://www.sipri.org/sites/default/files/2020-06/artificial_intelligence_strategic_stability_and_nuclear_risk.pdf
2. Saalman L. (ed.). *Integrating Cybersecurity and Critical Infrastructure. National, Regional and International Approaches*. 2018. Available at: https://www.sipri.org/sites/default/files/2018-04/integrating_cybersecurity_0.pdf
3. Boulanin V. (ed.). *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk. Vol. I. Euro-Atlantic Perspectives*. 2020. Available at: https://www.sipri.org/publications/2020/other-publications/artificial-intelligence-strategic-stability-and-nuclear-risk
4. Boulanin V., Verbruggen M. *Mapping the Development of Autonomy in Weapon Systems*. 2020. Available at: https://www.sipri.org/sites/default/files/2018-04/integrating_cybersecurity_0.pdf
5. Boulanin V., Bruun L., Goussac N. *Autonomous Weapon Systems and International Humanitarian Law. Identifying Limits and the Required Type and Degree of Human–Machine Interaction*. 2021. Available at: https://www.sipri.org/sites/default/files/2021-06/2106_aws_and_ihl_0.pdf
6. Saalman L., Su F., Saveleva Dovgal L. *Cyber Posture Trends in China, Russia, the United States and the European Union*. 2022. Available at: https://www.sipri.org/sites/default/files/2022-12/2212_cyber_postures_0.pdf
7. Boulanin V. *Mapping the development of autonomy in weapon systems. A primer on autonomy*. 2017. Available at: https://www.sipri.org/sites/default/files/Mapping-development-autonomy-in-weapon-systems.pdf
8. Boulanin V., Goussac N., Bruun L., Richards L. *Responsible Military Use of Artificial Intelligence. Can the European Union Lead the Way in Developing Best Practice?* 2020. Available at: https://www.sipri.org/publications/2020/other-publications/responsible-military-use-artificial-intelligence-can-european-union-lead-way-developing-best
9. Boulanin V., Brockmann K., Richards L. *Responsible Artificial Intelligence Research And Innovation For International Peace And Security*. 2020. Available at: https://www.sipri.org/sites/default/files/2020-11/sipri_report_responsible_artificial_intelligence_research_and_innovation_for_international_peace_and_security_2011.pdf
10. Bromley M., Maletta G. *The Challenge of Software and Technology Transfers to Non-Proliferation Efforts. Implementing and Complying with Export Controls*. 2018. Available at: https://www.sipri.org/publications/2018/other-publications/challenge-software-and-technology-transfers-non-proliferation-efforts-implementing-and-complying

11. Su F., Boulanin V., Turell J. *Cyber-incident Management Identifying and Dealing with the Risk of Escalation*. 2020. IPRI Policy Paper No. 55. Available at: https://www.sipri.org/publications/2020/sipri-policy-papers/cyber-incident-management-identifying-and-dealing-risk-escalation
12. Morgan F.E., Boudreaux B., Lohn A.J. et al. *Military Applications of Artificial Intelligence Ethical Concerns in an Uncertain World*. RAND Corporation, 2020:202. Available at: https://www.rand.org/pubs/research_reports/RR3139-1.html
13. Ruhl Ch. *Autonomous weapon systems and military artificial intelligence (AI) applications report*. 2022. Available at: https://www.founderspledge.com/research/autonomous-weapon-systems-and-military-artificial-intelligence-ai
14. *Research Priorities for Robust and Beneficial Artificial Intelligence: An Open Letter. Future of Life Institute*. Available at: https://futureoflife.org/data/documents/research_priorities.pdf (data obrashcheniya: 09.03.2024).
15. Marr B. *The 15 Biggest Risks Of Artificial Intelligence*. Available at: https://www.forbes.com/sites/bernardmarr/2023/06/02/the-15-biggest-risks-of-artificial-intelligence/?sh=20c095a92706
16. *Pause Giant AI Experiments: An Open Letter*. Available at: https://futureoflife.org/open-letter/pause-giant-ai-experiments/
17. *AI Ethics Code*. Available at: https://ethics.a-ai.ru/ (accessed 09.03.2024).
18. Delaborde A. *Risk assessment of artificial intelligence in autonomous machines. 1st international workshop on Evaluating Progress in Artificial Intelligence (EPAI 2020) in conjunction with ECAI*. Available at: 2020. https://hal.science/hal-03009978
19. *ICRC position on autonomous weapon systems. International Committee of the Red Cross*. Available at: https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems
20. Macrae C. Learning from the Failure of Autonomous and Intelligent Systems: Accidents, Safety and Sociotechnical Sources of Risk. *SSRN Electronic Journal*. 2021. doi: 10.2139/ssrn.3832621
21. Hindriks F., Veluwenkamp H. The risks of autonomous machines: from responsibility gaps to control gaps. *Synthese*. 2023;201. doi: 10.1007/s11229-022-04001-5
22. Radanliev P., De Roure D., Maple C. et al. Super forecasting the technological singularity risks from artificial intelligence. *Evolving Systems*. 2022;13:747–757. doi: 10.1007/s12530-022-09431-7
23. *The risks of Lethal Autonomous Weapons. The Future of Life Institute*. Available at: https://autonomousweapons.org/the-risks
24. *Asilomar AI Principles. The Future of Life Institute*. Available at: https://futureoflife.org/open-letter/ai-principles/
25. *Artificial Intelligence Risk & Governance. By Artificial Intelligence/Machine Learning Risk & Security Working Group (AIRS)*. The Wharton School, The University of Pennsylvania, 2024. Available at: https://ai.wharton.upenn.edu/white-paper/artificial-intelligence-risk-governance/
26. Skelton S.K. *AI experts question tech industry's ethical commitments*. TechTarget, 2024. Available at: https://www.computerweekly.com/feature/AI-experts-question-tech-industrys-ethical-commitments
27. Ivanov A.I., Ivanov A.P., Savinov K.N., Eremenko R.V. Virtual enhancement of the effect of parallelization of computing during the transition from binary neurons to the use of q-ary artificial neurons. *Nadezhnost' i kachestvo slozhnykh system = Reliability and quality of complex systems.* 2022;(4):89–97. (In Russ.)
28. Shirinkina E.V. The mechanism of application of artificial intelligence in teaching. *Nadezhnost' i kachestvo slozhnykh system = Reliability and quality of complex systems.* 2022. № 4. S. 24–30. (In Russ.)

Информация об авторах / Information about the authors

Олег Владимирович Прокофьев
кандидат технических наук, доцент,
доцент кафедры информационных
технологий и систем,
Пензенский государственный технологический
университет
(Россия, г. Пенза, пр-д Байдукова/
ул. Гагарина, 1а/11)
E-mail: prokof_ow@mail.ru

Oleg V. Prokofiev
Candidate of technical sciences, associate professor,
associate professor of the sub-department
of informational technologies and systems,
Penza State Technological University
(1a/11 Baydukov passage/Gagarin street,
Penza, Russia)

Автор заявляет об отсутствии конфликта интересов /
The author declares no conflicts of interests.
Поступила в редакцию/Received 15.12.2024
Поступила после рецензирования/Revised 10.01.2025
Принята к публикации/Accepted 10.02.2025