УДК 004.832.2
doi: 10.21685/2307-4205-2025-1-2

РИСКИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ ИСПОЛЬЗОВАНИИ В СОЦИОТЕХНИЧЕСКИХ СИСТЕМАХ

М. Ю. Михеев[1], О. В. Прокофьев[2], И. Ю. Семочкина[3]

[1, 2, 3] Пензенский государственный технологический университет, Пенза, Россия
[1] mix1959@gmail.com, [2] prokof_ow@mail.ru, [3] ius1961@gmail.com

Аннотация. *Актуальность и цели.* Применение автономных устройств с искусственным интеллектом в социотехнических системах привело к появлению новых задач, решение которых значительно сложнее по сравнению с задачами предыдущего этапа совершенствования человекомашинного интерфейса. Авторами работы было проведено исследование по выявлению источников риска в социотехнических системах и способов их снижения на этапах жизненного цикла автономных устройств. *Материалы и методы.* В исследовании использованы данные открытых отчетов к статистическим опросам разработчиков и пользователей устройств ответственного назначения, отчетов об исследовании аварий автономного транспорта Uber AV. *Результаты.* Даны формулировки источников риска, возникающих при использовании автономных устройств с искусственным интеллектом в социотехнических системах. Представлен пример жизненного цикла устройства, предусматривающий контроль человека на этапах разработки и эксплуатации. *Выводы.* Управление рисками в случаях ответственного применения устройства возможно при соответствии реальных этапов жизненного цикла перечисленному в статье ряду критериев безопасности.

Ключевые слова: риски искусственного интеллекта, социотехнические системы, автономные устройства, устройство ответственного назначения

Для цитирования: Михеев М. Ю., Прокофьев О. В., Семочкина И. Ю. Риски искусственного интеллекта при использовании в социотехнических системах // Надежность и качество сложных систем. 2025. № 1. С. 12–19. doi: 10.21685/2307-4205-2025-1-2

THE ARTIFICIAL INTELLIGENCE RISKS WHEN USED IN SOCIOTECHNICAL SYSTEMS

M.Yu. Mikheev[1], O.V. Prokofiev[2], I.Yu. Semochkina[3]

[1, 2, 3] Penza State Technological University, Penza, Russia
[1] mix1959@gmail.com, [2] prokof_ow@mail.ru, [3] ius1961@gmail.com

Abstract. *Background*. The use of autonomous devices with artificial intelligence in socio-technical systems has led to the emergence of new problems, the solution of which is much more complex compared to the tasks of the previous stage of improving the human-machine interface. The authors of the work conducted a study to identify risk sources in socio-technical systems and ways to reduce them at the stages of the life cycle of autonomous devices. *Materials and methods.* The study used data from open reports on statistical surveys of developers and users of critical devices, reports on the study of accidents of autonomous transport Uber AV. *Results.* The formulations of risk sources arising from the use of autonomous devices with artificial intelligence in socio-technical systems are given. An example of the life cycle of a device is presented, providing for human control at the stages of development and operation. *Conclusions*. Risk management in cases of critical use of a device is possible if the actual stages of the life cycle correspond to a number of safety criteria listed in the article.

Keywords: artificial intelligence risks, socio-technical systems, autonomous devices, critical device

For citation: Mikheev M.Yu., Prokofiev O.V., Semochkina I.Yu. The artificial intelligence risks when used in sociotechnical systems. *Nadezhnost' i kachestvo slozhnykh sistem = Reliability and quality of complex systems.* 2025;(1):12–19. (In Russ.). doi: 10.21685/2307-4205-2025-1-2

Введение

Социотехнические системы (СТС), включающие людей, технологии, способы взаимодействия и зависимости между ними, предназначены для достижения наибольших результатов конечным

пользователям, повышения эффективности управления производством и решения социальных проблем [1]. В предыдущие 60 лет развитие концепции этих систем было сконцентрировано в относительно узкой области человекомашинных систем на производстве, и мотивацией развития являлось решение проблем рабочей среды в индустрии, в том числе реализации новых технологий, промышленного дизайна и эргономики. Затем расширение сферы применения социотехнических систем было связано не только с более высокими требованиями к программно-аппаратному обеспечению, но и с появлением личных и общественных аспектов для конечных пользователей. Социальные сети, мессенджеры и блогерство привели к необходимости оценки качества информации, соблюдения правил цифровой гигиены при взаимодействии людей и технологий. Сегодняшняя реальность, связанная с появлением автономных устройств с искусственным интеллектом (ИИ) как части социотехнической системы, дала новый источник рисков, трудно поддающихся оценке, моделированию и прогнозированию.

Быстрое развитие и внедрение ИИ в области применения от здравоохранения до обороны, от финансов до автономного транспорта не может не затронуть общественную жизнь и права отдельных людей. Существует обоснованное мнение, отличающееся от мнения разработчиков ИИ о том, что принятие решений автономными устройствами основано на принципах беспристрастности, объективности и свободно от человеческих предубеждений. Необходимость обеспечения ускоренного развития искусственного интеллекта, с одной стороны, а также опасения в отношении трудовой занятости, возможность злоупотребления общественным контролем, ненулевая вероятность принятия недостаточно обоснованных решений в отношении здоровья и жизни людей, с другой стороны, привели к появлению законодательных актов, деклараций, открытых писем, руководств по этическим принципам ИИ [2–16].

Указом Президента № 124 от 15.02.2024 были внесены изменения в Национальную стратегию развития ИИ до 2030 г. Обновлены основные принципы развития и использования технологий ИИ, которые должны обязательно соблюдаться, в т.ч. заявлена прозрачность и объяснимость работы ИИ, недискриминационный доступ пользователей к информации об алгоритмах [17].

Сдержанные оценки в отношении безопасности технологий ИИ высказываются экспертами, обладающими опытом разработки автономных устройств в сферах транспорта, медицины и других чувствительных для человека и общества областях [18, 19]. Поэтому целью работы авторами выбрана разработка методологии снижения рисков искусственного интеллекта для безопасного внедрения автономных устройств в социотехнические системы [20].

Материалы и методы

В исследовании использованы результаты опроса экспертов в области ИИ, проводимые корпорациями-разработчиками систем ответственного назначения с ИИ. На рис. 1 изображен пример результатов опросов корпорацией RAND 2500 респондентов, имеющих отношение к области разработок ИИ и к смежным областям [21]. В продолжение темы, затронутой в предыдущей диаграмме, приведен рис. 2, затрагивающий потенциальную возможность провоцирующего поведения автономного оружия.

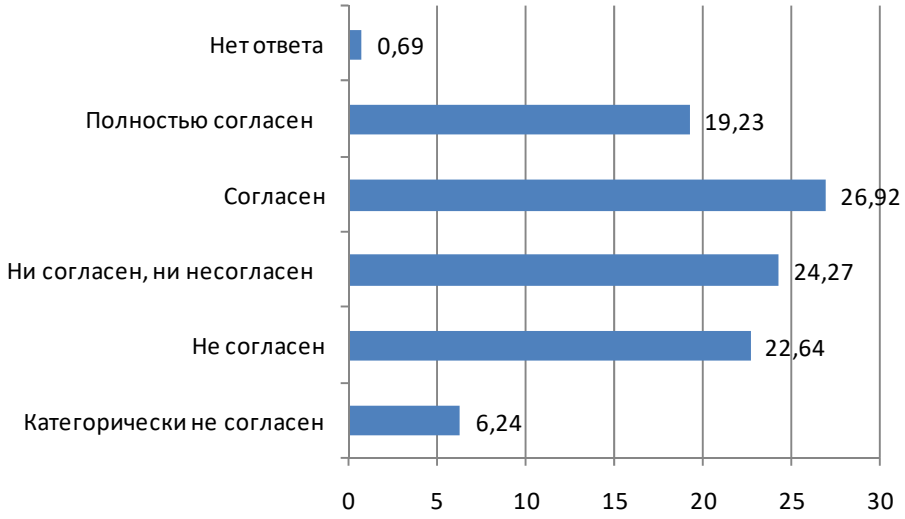

Рис. 1. Оценка высказывания «Автономное оружие должно быть запрещено с этической точки зрения, поскольку оно снижает ценность человеческой жизни (% опрошенных)»

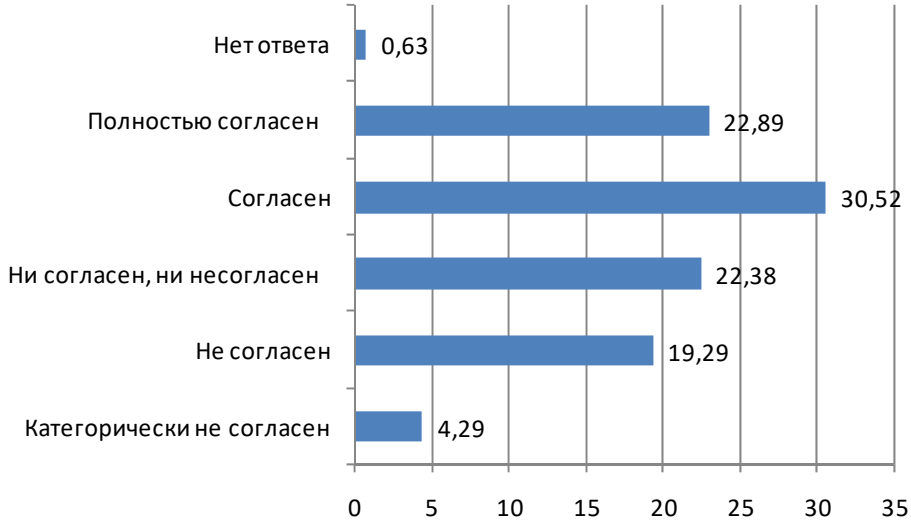

Рис. 2. Оценка высказывания «Развитие автономного оружия повысит вероятность возникновения войны (% опрошенных)»

Ниже представлены примеры результатов опроса экспертов Рабочей группы по рискам и безопасности в области ИИ и машинного обучения (МО) [22]: заданные вопросы и распределение ответов.

– Есть ли у вашего предприятия согласованное определение ИИ /МО? 40 % ответили «да».

– Обновило ли ваше предприятие существующие политики и стандарты для ИИ / МО? 50 % ответили «да».

– Определены ли на вашем предприятии роли и ответственность за системы ИИ / МО? 30 % ответили «да».

Как видно из этих результатов, в настоящее время адекватное понимание проблемы снижения рисков возникновения массовых вооруженных конфликтов полностью не достигается даже в наиболее передовых экономиках. Поэтому задачами исследования становится выявление мер противодействия рискам и способов их применения в социотехнических системах.

Результаты

Анализ открытых отчетов [19, 20] позволяет охарактеризовать пять фундаментальных источников социотехнического риска в автономных устройствах с ИИ, каждый из которых характеризуется рядом конкретных моделей социотехнических отказов и ошибок. В совокупности области структурного, организационного, технологического, эпистемологического и культурного рисков образуют структуру социотехнического риска в системах искусственного интеллекта (*Artificial Intelligence System*) (рис. 3).

Рис. 3. Социотехнические источники риска

Источники структурного риска возникают из-за взаимозависимостей и взаимодействия между различными частями технических и социальных структур. Структура СТС определяет, как различные части системы взаимодействуют друг с другом, и формирует то, как развиваются сбои. Структурные механизмы могут выступать в качестве источников риска, усиливая или передавая локальные сбои способами, которые выводят из строя всю систему. Например, аварии автономного транспорта Uber AV возникали из-за структурно взаимосвязанных сбоев, которые взаимодействовали между различными частями и в разных масштабах системы, охватывали конструкцию системы автономного вождения, деятельность операторов транспортных средств, решения инженеров и менеджеров, процессы тестирования транспортных средств и действия других участников дорожного движения и регулирующих органов. Структурные характеристики этой СТС позволяли сбоям в одной области быстро разрастаться или влиять на другие части системы (зачастую непредсказуемым образом).

Организационная деятельность может выступать источником риска в системах искусственного интеллекта, особенно когда организационный контекст нечувствителен к характеристикам деятельности человека, а организационные процессы неэффективны при обнаружении и управлении ошибками и сбоями. Авариям с Uber AV способствовала совокупность организационных, контекстуальных и человеческих факторов, в том числе недостатки в системах контроля, пробелы в знаниях в области безопасности, плохое взаимодействие человека и машины.

Разработка и эксплуатация автономных устройств с ИИ зависят от накопления и поддержания знаний о том, как оно работает, а также о том, как и почему оно может выйти из строя. Эпистемологические проблемы могут выступать в качестве источников риска, когда эти знания являются частичными, неверными или устаревшими, позволяя под недостаточной информированностью скрывать неожиданные угрозы. Аварии с Uber сформировали различные эпистемологические процессы, в которых участвовал экспериментальный автомобиль, который использовался для тестирования и исследования производительности системы. При этом допускались задержки в рассмотрении неожиданных инцидентов, были пробелы в изучении дорожных ситуаций. Эпистемологические проблемы усугубляются сложными, инновационными технологиями, которые создают возможности для неожиданных эффектов.

Социокультурные особенности могут выступать в качестве источников риска, поддерживая поведение и убеждения, которые приближают автономное устройство к пределам безопасной эксплуатации и приводят к тому, что предупреждающие сигналы неправильно интерпретируются, сводятся к минимуму, пропускаются или игнорируются. Сообщается, что с аварией Uber был связан целый ряд культурных факторов, в том числе акцент на показателях производительности, давление управляющего персонала, игнорирование мнения оперативного персонала и ошибочные предположения относительно эффективности человеческой бдительности и рисков испытания на дороге. Такие риски существуют в СТС как коллективные практики, нормы, ценности и предположения, которые глубоко формируют организационное поведение, расставляют приоритеты внимания. Коллективные ценности и убеждения могут сосредоточить внимание организации на конкретных данных и показателях, одновременно игнорируя другие.

Технологические свойства ИИ могут выступать в качестве источников риска, когда они приводят к тому, что автономное устройство может не работать должным образом, ведет себя неожиданным образом или его становится трудно контролировать. Как в случае с авариями Uber AV, которые стали результатом сложной системы технических сбоев и ограничений, включая проблемы со способностью автомобиля воспринимать объекты и прогнозировать путь пешеходов, правила подавления действий, которые задерживали реакцию автомобиля на предполагаемые опасности, а также ограничения на функции экстренного торможения. Технологические недостатки могут возникнуть во время проектирования и разработки: ошибки кодирования или ошибочные предположения могут быть заложены в технические объекты, снижая технологически обусловленную надежность. Конструктивные особенности могут усложнить или исключить эффективное взаимодействие человека и машины. Сбои могут возникнуть в мониторинге и контроле технологий: технологические стандарты могут быть плохо определены или неправильно применяться. Таким образом, источники технологических рисков проявляются на фазах жизненного цикла устройства с ИИ, эти риски в наибольшей степени доступны управлению за счет инженерных решений.

Обсуждение

На рис. 4 представлен пример жизненного цикла по версии, представленной *RAND Corporation* [21] для автономного вооружения с ИИ, в основу которого положено снижение рисков в столь чувствительной области применения. Как видно из рис. 4, этот подход направлен на то, что снижение рисков под контролем человека начинается еще до того, как система будет разработана посредством

правовой регламентации, и распространяется на весь жизненный цикл, вплоть до анализа боевых повреждений и других событий после внедрения.

Рис. 4. Жизненный цикл автономного вооружения с поэтапным контролем

Этот подход показывает, что участие человека реализуется в разных ролях в разное время. Необходимо рассмотреть различные точки соприкосновения со специалистами на протяжении жизненного цикла устройства, в том числе в рамках проектирования и разработки, создания правил взаимодействия и формирования процесса определения целей. Показанная формальная модель жизненного цикла позволяет, несмотря на упрощение реальных процессов, сделать ряд рекомендательных выводов.

Заключение

На основании примера можно сформулировать несколько ключевых критериев, имеющих отношение к обеспечению участия человека в ответственном применении ИИ для снижения рисков. Эти критерии не являются всеобъемлющими, но они указывают тип сопутствующих проблем проектирования устройства, которые необходимо решить.

Надежность и предсказуемость. Необходимо установить стандарты надежности и предсказуемости, чтобы гарантировать, что система будет работать в соответствии с назначением проекта. Такие стандарты должны быть установлены на этапе проектирования и разработки при производстве автономных устройств и они должны соответствовать требованиям, прежде чем они будут объявлены функциональными. После того, как устройства введены в эксплуатацию, необходимо постоянно собирать данные и периодически проводить оценки, чтобы убедиться, что они продолжают соответствовать стандартам.

Варианты вмешательства. Учитывая риски, связанные с авариями, непредвиденными последствиями или другими проблемами, важным критерием проектирования системы является возможность вмешательства для перенаправления или остановки системы. Насколько это возможно с параметрами требований миссии, человек-оператор должен иметь возможность своевременно вмешиваться в действия системы, чтобы перенаправить ее по мере необходимости. Насколько своевременными должны быть такие варианты вмешательства, будет зависеть от контекста. Операторы должны иметь возможность максимально оперативно вмешиваться в действия систем, критически важных для жизни и здоровья людей.

Доступность информации. Критерий касается требуемой степени прозрачности решений ИИ. Чтобы снизить риски ИИ, оператор должен знать, для чего предназначена система в конкретном рабочем контексте, и должен быть в состоянии определить в приемлемые сроки, как система пришла к критически важным решениям или почему она предприняла определенные действия. Передовая практика взаимодействия между людьми и машинами должна заключаться в том, чтобы оперативная ситуация была легко понятной для обученных операторов, обеспечивалась прослеживаемая обратная связь о состоянии системы; а также возможность предоставить обученным операторам четкие процедуры для активации и деактивации системных функций.

Ограничения на типы задач. Хотя большинство современных автономных устройств, как правило, имеют лишь узкий диапазон функциональных возможностей, будущие технологические разработки могут создать системы, способные выполнять широкий круг задач. Это условие будет ограничивать количество и типы задач, которые может выполнять система. Системы, которые совершают ответственные действия и образуют сопутствующие риски, могут включать дополнительные средства контроля.

Ограничения по географии. Некоторые автономные устройства имеют широкие навигационные возможности для перемещения в пространстве. Чтобы снизить риск, связанный с неограниченным движением, операторы должны иметь возможность устанавливать параметры, в пределах которых должно оставаться устройство. Если события или сбои приводят к тому, что система выходит за свои географические ограничения, оно может быть запрограммировано на прекращение задания и возвращение на исходную позицию.

Руководствуясь перечисленными критериями, разработка автономных устройств с ИИ позволит в полной мере реализовать их потенциальные возможности для многочисленных пользователей СТС при поддержании уровня рисков на приемлемо низком уровне [20, 23].

Список литературы

1. Sociotechnical system. URL: https://en.wikipedia.org/wiki/Sociotechnical_system
2. Pause Giant AI Experiments: An Open Letter. URL: https://futureoflife.org/open-letter/pause-giant-ai-experiments/
3. 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
4. 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
5. The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk. Volume 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
6. 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
7. 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
8. 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
9. 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
10. 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
11. 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
12. 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
13. 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
14. Национальная стратегия развития искусственного интеллекта на период до 2030 года. URL: http://pravo.gov.ru/proxy/ips/?docbody=&firstDoc=1&lastDoc=1&nd=102608394
15. Альянс в сфере искусственного интеллекта. URL: https://a-ai.ru/
16. Кодекс этики в сфере искусственного интеллекта. URL: https://ethics.a-ai.ru/assets/ethics_files/2023/05/12/%D0 %9A %D0 %BE %D0 %B4 %D0 %B5 %D0 %BA %D1 %81_ %D1 %8D %D1 %82 %D0 %B8 %D0 %BA %D0 %B8_20_10_1.pdf
17. Применение искусственного интеллекта на финансовом рынке : докл. для общественных консультаций. М. : Банк России, 2023. 51 с. URL: https://www.cbr.ru/Content/Document/File/156061/Consultation_Paper_03112023.pdf
18. 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
19. 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
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. 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
22. 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/
23. 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
24. Михеев М. Ю., Прокофьев О. В., Савочкин А. Е., Семочкина И. Ю. Обеспечение надежности в жизненном цикле систем искусственного интеллекта ответственного назначения // Надежность и качество сложных систем. 2023. № 3. С. 12–20.
25. 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.

References

1. *Sociotechnical system*. Available at: https://en.wikipedia.org/wiki/Sociotechnical_system
2. *Pause Giant AI Experiments: An Open Letter*. Available at: https://futureoflife.org/open-letter/pause-giant-ai-experiments/
3. 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
4. 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
5. Boulanin V. (ed.). *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk. Volume I. Euro-Atlantic Perspectives*. 2020. Available at: https://www.sipri.org/publications/2020/other-publications/artificial-intelligence-strategic-stability-and-nuclear-risk
6. 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
7. 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
8. 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
9. 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
10. 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
11. 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
12. 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
13. 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
14. *Natsional'naya strategiya razvitiya iskusstvennogo intellekta na period do 2030 goda = National Strategy for the development of artificial intelligence for the period up to 2030*. (In Russ.). Available at: http://pravo.gov.ru/proxy/ips/?docbody=&firstDoc=1&lastDoc=1&nd=102608394
15. *Al'yans v sfere iskusstvennogo intellekta = Alliance in the field of artificial intelligence*. (In Russ.). Available at: https://a-ai.ru/
16. *Kodeks etiki v sfere iskusstvennogo intellekta = Code of Ethics in the field of artificial intelligence*. (In Russ.). Available at: https://ethics.a-ai.ru/assets/ethics_files/2023/05/12/ %D0 %9A %D0 %BE %D0 %B4 %D0 %B5 %D0 %BA %D1 %81_ %D1 %8D %D1 %82 %D0 %B8 %D0 %BA %D0 %B8_20_10_1.pdf
17. *Primenenie iskusstvennogo intellekta na finansovom rynke: dokl. dlya obshchestvennykh konsul'tatsiy = Application of artificial intelligence in the financial market : a report for public consultations*. Moscow: Bank Rossii, 2023:51. (In Russ.). Available at: https://www.cbr.ru/Content/Document/File/156061/Consultation_Paper_03112023.pdf

18. 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
19. 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
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. 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
22. *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/
23. 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
24. Mikheev M.Yu., Prokof'ev O.V., Savochkin A.E., Semochkina I.Yu. Ensuring reliability in the life cycle of artificial intelligence systems for responsible purposes. *Nadezhnost' i kachestvo slozhnykh system = Reliability and quality of complex systems.* 2023;(3):12–20. (In Russ.)
25. 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.)

Информация об авторах / Information about the authors

Михаил Юрьевич Михеев
доктор технических наук, профессор,
заведующий кафедрой информационных технологий и систем,
Пензенский государственный технологический университет
(Россия, г. Пенза, пр-д Байдукова/ул. Гагарина, 1а/11)
E-mail: mix1959@gmail.com

Mikhail Yu. Mikheev
Doctor of technical sciences, professor,
head of the sub-department of informational technologies and systems,
Penza State Technological University
(1a/11 Baydukov passage/Gagarin street, Penza, Russia)

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)

Ирина Юриевна Семочкина
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: ius1961@gmail.com

Irina Yu. Semochkina
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)

Авторы заявляют об отсутствии конфликта интересов /
The authors declare no conflicts of interests.
Поступила в редакцию/Received 15.01.2025
Поступила после рецензирования/Revised 28.02.2025
Принята к публикации/Accepted 10.03.2025