

Искусственный интеллект для миростроительства

Обещания и подводные камни

Резюме для политиков 2025.3

СИНОПСИС

Искусственный интеллект (ИИ) быстро становится частью конфликтных ситуаций в мире. Он используется для мониторинга рисков конфликтов, поддержки гуманитарной координации и содействия мирным диалогам, но также и для расширения репрессий, слежки за гражданским населением и распространения дезинформации. В этом «Резюме для политиков» освещаются наиболее перспективные варианты применения ИИ в миростроительстве и описываются основные меры по предотвращению ущерба.

Сегодня ИИ используется для:

1. укрепления систем раннего предупреждения и прогнозирования
2. обеспечения инклюзивного цифрового диалога и гражданской активности
3. поддержки картографирования кризисных ситуаций и гуманитарного реагирования
4. содействия просоциальному взаимодействию и смягчению поляризации в сети Интернет

Однако успешное использование ИИ в миростроительстве зависит от обеспечения ответственного управления ИИ в нестабильных конфликтных ситуациях.

Чтобы использовать миростроительный потенциал ИИ и уменьшить его опасность, политики должны:

- Учитывать основные права человека и принципы «не навреди» на всех этапах проектирования и внедрения ИИ.
- Финансировать инициативы, основанные на фактических данных, и обеспечивать широкий доступ к данным об успехах и неудачах.
- Поддерживать небольшие, локально релевантные языковые модели, отражающие различные контексты.
- Продвигать прозрачность и подотчетность в отношении инструментов ИИ двойного назначения.
- Налаживать партнерские отношения между секторами миростроительства, развития и технологий.

Данный отчет основан на *Техническом документе 2025.3 «Искусственный интеллект и миростроительство: возможности и вызовы»*. В нем обобщены выводы, сделанные на основе обширного обзора более 600 источников по вопросам миростроительства, гуманитарного реагирования, этики ИИ и управления технологиями. В этом разнообразном корпусе

собраны тематические исследования из стран и конфликтных ситуаций по всему миру с использованием широкого спектра средств искусственного интеллекта, машинного обучения и смежных инструментов.

ВВЕДЕНИЕ

Искусственный интеллект занимает все более важное место в миростроительной деятельности. В лучших случаях ИИ может заблаговременно предупреждать о насилии, обеспечивать всесторонний общественный диалог и помогать миротворцам защищать гражданское население. Вместе с тем он также может усиливать цифровую слежку, искажать информационную среду и ускорять принятие военных решений без участия человека. Эти противоречия двойного назначения ставят перед политиками вопросы, требующие безотлагательного решения.

Результаты, обобщенные здесь, отражают работу исследователей, практиков и консультантов на пяти континентах. Этот анализ основан на обзоре 600 рецензируемых статей, технических документов, тематических исследований и отчетов по материалам «серой» литературы. В сборник вошли работы международных миротворческих организаций, ученых, разработчиков технологий и местных лидеров гражданского общества. Этот процесс обеспечил широкий спектр точек зрения и углубил наше понимание как потенциала, так и подводных камней ИИ в контексте миростроительства.

Искусственный интеллект уже играет разнообразную роль в сфере миростроительства. От реагирования на кризисные ситуации до гражданской активности — целый ряд приложений начинает влиять на то, как специалисты предотвращают насилие и строят более устойчивые сообщества.

Пять выводов, приведенных ниже, подчеркивают, где ИИ оказывает наибольшее влияние и где новые проблемы требуют срочного внимания со стороны политики.

Дополнительные сведения и региональные примеры см. в *Техническом документе 2025.3*.

ВЫВОД 1: ПРОГНОЗИРОВАНИЕ С ОСТОРОЖНОСТЬЮ

ОСНОВНЫЕ

ИИ может улучшить системы раннего предупреждения, но чрезмерное увлечение прогностическим и моделями рискованно.

Инструменты искусственного интеллекта все чаще используются для прогнозирования вероятности вспышек насилия. Такие системы, как система раннего предупреждения насилия (ViEWS) и проект Armed Conflict Location & Event Data (ACLED), показывают, что модели машинного обучения могут выявлять предвестники насилия — например, тенденции к разжиганию ненависти, необычные перемещения войск или всплески экономического неблагополучия — и прогнозировать конфликт с точностью до 95% в краткосрочной перспективе. Эти инструменты могут быть особенно полезны в условиях, когда мониторинг в режиме реального времени ограничен [1], [2].

Однако прогностические модели по-прежнему подвержены ошибкам и предвзятости, особенно в условиях дефицита данных.

Гуманитарные работники предупреждают, что ложноположительные результаты могут привести к ненужным вмешательствам, а ложноотрицательные — к риску для жизни. ИИ также не справляется с долгосрочными прогнозами и политическими нюансами. Политики должны рассматривать прогнозы как дополнительные, а не окончательные. Гибридный подход, сочетающий сигналы ИИ с местным опытом и анализом на основе широкого участия, дает наибольшие перспективы для создания действенных, учитывающих контекст систем раннего предупреждения.

ВЫВОД 2: МАСШТАБИРОВАНИЕ ЦИФРОВОГО ДИАЛОГА

ОСНОВНЫЕ

Инструменты цифрового диалога расширяют возможности гражданского участия в зонах конфликтов.

Платформы на базе ИИ помогают расширить гражданское участие в регионах, затронутых конфликтами, анализируя общественный вклад и усиливая голоса маргиналов. Такие инструменты, как Remesh и Talk to the City, позволяют тысячам участников обмениваться мнениями, находить точки соприкосновения и вносить свой вклад в разработку политики [3], [4], [5], [6]. В Ливии, Сирии и Йемене ООН и другие миростроительные организации использовали эти платформы для проведения многоязычных, инклюзивных консультаций, отражающих различные социальные и политические контексты.

Эти технологии могут расширить доступ и повысить легитимность, но они требуют тщательного внедрения. Доверие, цифровая грамотность и безопасность имеют огромное значение.

Неправильно организованные процессы могут исключить уязвимые группы или создать риск неправомерного использования данных.

Инициативы по проведению диалога должны поддерживаться личным участием и партнерством с сообществом. При правильном подходе цифровые консультации помогают расширить масштабы мирного диалога так, как это часто не удастся сделать на традиционных форумах, что делает их жизненно важным компонентом современных стратегий миростроительства. ИИ создает новые захватывающие возможности для диалога, но глобальное цифровое неравенство сохранится.

ВЫВОД 3: КАРТОГРАФИРОВАНИЕ С УЧЕТОМ ГАРАНТИЙ

ОСНОВНЫЕ

Картографирование с помощью ИИ ускоряет процесс гуманитарного реагирования, но может быть использовано не по назначению.

В сфере гуманитарного реагирования ИИ помогает создавать карты зон конфликтов, ущерба от стихийных бедствий и перемещения населения в режиме реального времени. Такие инструменты, как «Анализ данных для социальных и гуманитарных действий» (Data Insights for Social & Humanitarian Action (DISHA), многосторонняя инициатива под руководством UN Global Pulse), а также платформы, используемые для объединения видеозаписей с беспилотников, спутниковых снимков и краудсорсинговых отчетов, позволяют быстрее и в большем масштабе проводить детальные оценки [7]. В ходе пилотных исследований системы ИИ позволили сократить время анализа ущерба в шесть раз. Если такие системы будут применяться в реальных условиях, они помогут специалистам по реагированию более эффективно добираться до пострадавших сообществ.

Однако картографические технологии также могут представлять серьезную опасность, если их использовать без надзора. В некоторых случаях геопространственные данные использовались для преследования гражданских лиц или подавления инакомыслия. Алгоритмы могут неправильно классифицировать важные социальные тенденции в момент кризиса, особенно в условиях недостатка данных. Эти инструменты должны быть оснащены надежными гарантиями конфиденциальности, этическими протоколами и механизмами вовлечения сообщества. Цель — повысить ситуационную осведомленность без ущерба для безопасности и автономности людей на земле.

ВЫВОД 4: ПРОЕКТИРОВАНИЕ ДЛЯ ДИАЛОГА

ОСНОВНЫЕ

Миростроительство выигрывает от просоциального проектирования технологических платформ.

Основная бизнес-модель социальных медиаплатформ — максимизация вовлеченности — часто поощряет возмущение и поляризующий контент. Но ИИ можно использовать и для создания более конструктивных онлайн-пространств. Новые итерации Perspective API помогают изменить этот стимул, отдавая предпочтение вдумчивым комментариям и личным историям, а не сенсациям. Такие инструменты, как Phoenix, позволяют миростроителям проводить прослушивание социальных сетей, которое служит основой для миротворческой работы [8], [9]. Эти изменения способствуют созданию более здоровой информационной среды, что крайне важно в нестабильных ситуациях.

Миростроители начинают сотрудничать с технологами по внедрению в цифровые инфраструктуры таких ценностей, как эмпатия, взаимопонимание и укрепление доверия. Хотя эти усилия еще только формируются, первые пилотные проекты показывают, что просоциальные алгоритмы могут расширить возможности ознакомления с различными точками зрения и уменьшить количество вредной дезинформации. Такой проактивный подход к разработке технологий, основанный на проектировании, может помочь восстановить социальную сплоченность в глубоко разделенных обществах.

ВЫВОД 5: ОТВЕТСТВЕННОЕ УПРАВЛЕНИЕ ИИ

ОСНОВНЫЕ

Ответственное управление ИИ имеет решающее значение в нестабильных конфликтных ситуациях.

Эффективное миростроительство требует, чтобы разработка и внедрение ИИ осуществлялись под строгим этическим и правовым надзором. Однако во многих зонах конфликтов системы регулирования слабы или вовсе отсутствуют. Без оградительных мер инструменты ИИ могут способствовать слежке, предиктивной работе полиции и алгоритмической дискриминации. Генеративные модели также могут галлюцинировать или искажать факты, что может иметь реальные последствия для человеческих жизней [10].

Чтобы устранить эти риски, управление ИИ должно основываться на правах человека и принципах широкого участия. Это включает в себя консультации с местным населением, открытую документацию, прозрачность данных об обучении и механизмы подотчетности. Доноры и исполнители должны отдавать приоритет инвестициям в маломасштабные, учитывающие конкретные условия модели и брать на себя обязательство открыто отчитываться как об успехах, так и о

неудачах. Ответственное управление — это не только моральный долг, но и стратегическая необходимость, чтобы ИИ поддерживал мир, а не углублял нестабильность.

ЗАКЛЮЧЕНИЕ

Искусственный интеллект не заменит дипломатию, примирение и укрепление доверия. Но он может поддерживать эти усилия — если будет спроектирован с соблюдением этических норм, развернут с участием всех заинтересованных сторон и будет управляться прозрачно. Миростроители должны относиться к ИИ как к инструменту дополнения, а не как к замене человеческого общения. Для достижения реального прогресса необходимы усилия всего общества по ответственному использованию ИИ, особенно в условиях конфликта и нестабильности.

Даже самые совершенные инструменты ИИ могут усугубить, а не предотвратить конфликт, если их развернуть в нездоровой информационной среде. Чтобы укрепить доверие к системам ИИ, особенно в нестабильных и постконфликтных ситуациях, политикам необходимо обеспечить внедрение этих технологий в рамки, в которых приоритетными являются справедливость, равенство и инклюзивность. Одной технической сложности недостаточно. Для обеспечения легитимности и прочного мира в процессах принятия решений необходимо реальное участие местных сообществ. ИИ должен применяться таким образом, чтобы обеспечить уважение человеческого достоинства и расширить права и возможности людей, находящихся на передовой.

Многосторонние организации, гражданское общество и технологи должны совместно работать над установлением глобальных норм и инвестировать в долгосрочный мониторинг воздействия ИИ. Регулярная адаптация политики, прозрачность при проектировании и внедрении, а также внимание к непредвиденным последствиям помогут предотвратить ущерб и повысить устойчивость институтов. Такой проактивный подход необходим для того, чтобы не допустить превращения ИИ в растущий источник нестабильности и сохранить его в качестве инструмента мира.

ИСТОЧНИКИ

- [1] H. Hegre, J. Karlsen, H. Mokleiv, H. Strand, and H. Urdal, “Forecasting Fatalities in Armed Conflict: VIEWS Forecasts for April 2022–March 2025,” Uppsala University, May 2022. [Online]. Available: <https://uu.diva-portal.org/smash/get/diva2:1665945/FULLTEXT01.pdf>
- [2] E. Albrecht, “Predictive Technologies in Conflict Prevention: Practical and Policy Considerations for the Multilateral System,” United Nations University Centre for Policy Research, Jun. 2023. [Online]. Available: https://unu.edu/sites/default/files/2023-09/predictive_technologies_conflict_prevention_.pdf
- [3] D. Masood Alavi, M. Wahlisch, C. Irwin, and A. Konya, “Using artificial intelligence for peacebuilding,” *J. Peacebuilding Dev.*, vol. 17, no. 2, pp. 239–243, May 2022, doi: 10.1177/15423166221102757.
- [4] B. Marnette and C. McKenzie, “Talk to the city: An open-source AI tool for scaling deliberation,” [Online]. Available: <https://ai.objectives.institute/blog/talk-to-the-city-an-open-source-ai-tool-to-scale-deliberation>
- [5] T. Bernard, “Can tech promote social cohesion?,” *Tech Policy Press*, Apr. 03, 2023. [Online]. Available: <https://techpolicy.press/can-tech-promote-social-cohesion>
- [6] S. Thompson and A. Piirtola, “Artificial Intelligence and Peace - The Case of Digital Dialogues in Sudan,” CMI, Feb. 2024. [Online]. Available: https://cmi.fi/wp-content/uploads/2024/02/CMI_INSIGHT_2024_sudan.pdf
- [7] L. Bromley, K. Jauer, and Y. Matias, “AI from Google Research and UN boosts humanitarian disaster response: Wider coverage, faster damage assessments,” *disha - UN Global Pulse*. [Online]. Available: <https://disha.unglobalpulse.org/ai-from-google-research-and-un-boosts-humanitarian-disaster-response-wider-coverage-faster-damage-assessments/>
- [8] J. Hawke, H. Puig Larrauri, A. Sutjahjo, and B. Cerigo, “Understanding to intervene: The codesign of text classifiers with peace practitioners,” *Data Policy*, vol. 6, 2024, doi: 10.1017/dap.2024.44.
- [9] Build Up, “We transform conflict in the digital age.” [Online]. Available: <https://howtobuildup.org/>
- [10] International Panel on the Information Environment, “Towards A Global AI Auditing Framework: Assessment and Recommendations,” Feb. 2025. [Online]. Available: <https://www.ipie.info/research/sr2024-3>

БЛАГОДАРНОСТИ

Соавторы

Авторы текста: Крейг Зелизер (ученый-консультант, Колумбия), Фредрик Огенга (председатель группы, Кения), Лиза Ширч (заместитель председателя группы, США) и Эвелин Таухниц (заместитель председателя группы, Швейцария), Себастьян Валенсуэла (главный научный сотрудник IPIE и председатель Научно-методического комитета, Чили), Филипп Ховард (президент и генеральный директор IPIE, Канада/Великобритания).

Независимые рецензенты: Джозеф Айлетт-Буллок, Микеле Джованарди и Бранка Панич. Проверка фактов: Хайди Шульц. Дизайн: Доменико Ди Донна. Копирайтинг: Беверли Сайкс и Ромилли Голдинг. Мы с благодарностью признаем поддержку со стороны Секретариата IPIE: Лола Гимферрер, Эгертон Нето, Виктория Шульц, Донна Сеймур, Анна Штаендер и Алекс Янг.

Финансирование

Международная группа по информационной среде (IPIE) с благодарностью отмечает партнеров, оказавших финансовую поддержку. Полный список партнеров по финансированию можно найти на сайте www.IPIE.info. Любые мнения, выводы, заключения или рекомендации, изложенные в данном материале, принадлежат IPIE и не обязательно отражают точку зрения партнеров по финансированию.

Декларация интересов

Отчеты IPIE разрабатываются и рецензируются глобальной сетью исследовательских филиалов и ученых-консультантов, которые образуют целенаправленные научные группы и группы соавторов. Все соавторы и рецензенты заполняют декларации интересов, которые рассматриваются IPIE на соответствующих этапах работы.

Предпочтительное цитирование

Резюме IPIE для политиков представляет собой краткий обзор состояния знаний на высоком уровне и предназначено для широкой аудитории. *Сводный отчет IPIE* использует методы научного мета-анализа, систематического обзора и другие инструменты для обобщения доказательств, обобщения знаний и формирования научного консенсуса, и написан для экспертной аудитории. В *Техническом документе IPIE* рассматриваются вопросы методологии или проводится анализ политики по конкретной проблеме нормативно-правового регулирования. Все отчеты доступны на сайте IPIE (www.IPIE.info).

Этот документ следует цитировать как:

International Panel on the Information Environment [C. Zelizer, F. Ogenga, L. Schirch, E. Tauchnitz, P. N. Howard, S. Valenzuela (eds.)], “AI for Peacebuilding: Promises and Pitfalls,” Zurich, Switzerland: IPIE, 2025. Summary for Policymakers, SFP2025.3, doi: 10.61452/VKJF9318.

Информация об авторских правах



Эта работа лицензирована в соответствии с Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

ОБ IPIE

Международная группа по информационной среде (IPIE) является независимой глобальной научной организацией, стремящейся предоставлять наиболее действенные научные знания об угрозах информационной среде в мире. Миссия IPIE, базирующегося в Швейцарии, заключается в предоставлении политикам, промышленникам и гражданскому обществу независимых научных оценок глобальной информационной среды путем организации, оценки и повышения уровня исследований с широкой целью улучшения глобальной информационной среды. Сотни исследователей со всего мира участвуют в подготовке отчетов IPIE.

За дополнительной информацией обращайтесь в Международную группу по информационной среде (IPIE) по адресу secretariat@IPIE.info. Seefeldstrasse 123, Почтовый ящик, 8034 Цюрих, Швейцария.



International Panel on
the Information
Environment

Seefeldstrasse 123
P.O. Box 8034 Zurich
Switzerland

