

Бойкова Анна Викторовна
Тверской государственный технический университет
Никольская Вера Александровна
Тверской государственный технический университет

Применение технологий ИИ в системах принятия решения военного назначения: за и против

Аннотация. Внедрение систем поддержки принятия решений (СППР) на основе искусственного интеллекта (ИИ) в процессы, связанные с применением военной силы, представляет собой один из наиболее значимых и противоречивых технологических сдвигов в современном обществе. В статье анализируются ключевые положения экспертного отчета Международного Комитета Красного Креста и Женевской академии международного гуманитарного права и прав человека, посвященного этой проблеме. Ее цель – структурировать дискуссию вокруг концептуальных различий между ИИ-СППР и автономными системами вооружения, выявить основные технические, правовые и этические риски, связанные с использованием ИИ в критических решениях, а также оценить предлагаемые меры по сохранению осмысленного человеческого контроля. На основе анализа делается вывод о том, что ключевой гуманитарной императивой является не запрет технологий, а разработка строгих ограничений, протоколов тестирования в человеко-машинном взаимодействии, обеспечивающих примат права и человеческого достоинства в условиях цифровизации войны.

Ключевые слова: искусственный интеллект, системы поддержки принятия решений, военное дело, международное гуманитарное право, автономное оружие, человеческий контроль, этика, конфликт

Boykova Anna Viktorovna
Tver State Technical University
Nikolskaya Vera Alexandrovna
Tver State Technical University

The Pros and Cons of Using AI in Military Decision-Making Systems

Abstract. The integration of artificial intelligence (AI)-based decision support systems (DSS) into military processes related to the use of force represents one of the most significant and controversial technological shifts in modern conflict. This article analyzes the key provisions of the 2024 expert report by the International Committee of the Red Cross (ICRC) and the Geneva Academy on this issue. The article aims to structure the discussion around the conceptual differences between AI-based DSS and autonomous weapons systems, identify the main technical, legal, and ethical risks associated with the use of AI in critical decisions, and evaluate proposed measures to preserve meaningful human control. Based on the analysis, it is concluded that the key humanitarian imperative is not a ban on technologies, but the development of strict restrictions, testing protocols, and a “constraints architecture” in human-machine interactions that ensure the primacy of law and human dignity in the context of the digitalization of war.

Keywords: artificial intelligence, decision support systems, military affairs, international humanitarian law, autonomous weapons, human control, ethics, conflict

Эволюция военных технологий исторически определяла характер вооруженных конфликтов. В последнее десятилетие дискуссии о будущем войны были сфокусированы

на автономных системах вооружения, способных самостоятельно выбирать и поражать цели. Однако, как отмечается в отчете Международный Комитет Красного Креста (МККК) и Женевской академии международного гуманитарного права и прав человека, сфера военного применения искусственного интеллекта (ИИ) гораздо шире и потенциально более проникающая. Особое внимание привлекают системы поддержки принятия решений (СППР), усиленные алгоритмами машинного обучения, которые предназначены для помощи человеку на всех этапах принятия решения о применении силы – от стратегического планирования до тактического исполнения.

Актуальность исследования обусловлена тем, что внедрение СППР на основе ИИ (ИИ-СППР)) происходит на фоне фундаментальной трансформации информационной составляющей военного конфликта. Объем и скорость обработки данных превышают когнитивные возможности человека, что вызывает необходимость автоматизации аналитических процессов. При этом, возникают сложные вопросы на стыке технологии, международного гуманитарного права и этики: как сохранить когнитивные способности человека в условиях жестких временных ограничений на принятие решений? Каковы риски делегирования аналитических, а фактически – оценочных функций алгоритмам в ситуациях, связанных с жизнью и смертью?

Цель данной статьи – на основе анализа экспертного отчета «Искусственный интеллект и смежные технологии при принятии решений о применении силы в вооруженных конфликтах» (2024 г.) систематизировать вызовы, связанные с использованием ИИ-СППР, и проанализировать возможные пути смягчения связанных с ними гуманитарных рисков [1]. Для достижения цели решаются следующие задачи:

- определить концептуальные рамки и отличия ИИ-СППР от автономных систем вооружения;

- проанализировать существующие варианты разработок в данной области;

- выявить ключевые технические, правовые и этические риски;

- оценить предлагаемые в отчете технические и организационные меры по сохранению человеческого контроля.

Первым шагом в осмыслении проблемы является четкое терминологическое разделение. В отчете подчеркивается, что, несмотря на схожесть базовых технологий, ИИ-СППР и АСВ представляют собой концептуально и юридически различные категории.

Принятие решений о применении Военные силы – это сложный, многоуровневый процесс, вовлекающий политическое руководство страны, стратегический, оперативный и тактический уровни командования. Решения варьируются в зависимости от политических целей и правил ведения боевых действий, конкретных задач по выбору цели, оценке сопутствующего ущерба и применению оружия. ИИ-СППР потенциально могут использоваться на любом из этих этапов.

ИИ-СППР определяются как компьютеризированные инструменты, призванные информировать и поддерживать, но не заменять человека-оператора. Они обрабатывают большие объемы данных для описания ситуации, предсказания развития событий или выработки рекомендаций. В отличие от ранних экспертных систем, современные ИИ-СППР, основанные на машинном обучении, «учатся» на данных, самостоятельно выявляя паттерны и формируя правила, что снижает прозрачность их работы.

Ключевое отличие от автономных систем вооружения заключается в месте и характере человеческого вмешательства. Они самостоятельно выбирают и применяют силу к целям без дальнейшего вмешательства человека. В случае с ИИ-СППР финальное решение – «нажать на курок» – остается за человеком. Таким образом, основной вызов для ИИ-СППР заключается не в «пробеле в ответственности» (который характерен для автономных систем вооружения), а в том, как гарантировать, что система действительно поддерживает, а не подрывает способность человека принимать обоснованные решения в соответствии с нормами международного гуманитарного права.

Однако в отчете отмечается, доверяющий рекомендациям ИИ-СППР, человек становится лишь пассивным, утверждающим органом, что чревато серьезными проблемами.

Интерес военных к ИИ-СППР обусловлен несколькими взаимосвязанными факторами, центральным из которых является стремление ускорить цикл принятия решений. В отчете используется модель петли OODA (Наблюдение-Ориентация-Решение-Действие), согласно которой сторона, способная быстрее обработать информацию и действовать, получает решающее тактическое преимущество. Современные сенсорные системы генерируют большие массивы данных («Большие данные»), анализ которых человеком становится непосильной задачей. ИИ-СППР могут быть использованы этапе «ориентации» – превращения данных в осмысленную тактическую картину.

Кроме того, применение систем поддержки принятия решения способствует повышению точности проводимых расчетов и снижению сопутствующего ущерба (особенно в условиях городских боевых действий), а также разработке новых эффективных тактик в условиях меняющегося характера конфликтов.

В отчете приводит далеко не весь список примеров разработок и применения ИИ-СППР на разных уровнях:

- на стратегическом уровне применяются системы анализа поведения государств, военное прогнозирование и моделирование. Например, проект Maven (США) по автоматическому анализу видеоданных с беспилотников для обнаружения и классификации объектов;

- на оперативном уровне используются системы автоматического распознавания целей (ATR), анализа метаданных связи для выявления паттернов активности, алгоритмы оценки сопутствующего ущерба и выбора средств поражения;

- на тактическом уровне используются виртуальные системы управления боем позволяющие в режиме реального времени выбрать оптимальное оружие и минимизировать потерь среди гражданского населения; системы прогнозирования перемещений противника и повышения осведомленности о ситуации на поле боя (например, платформа Palantir Gotham).

Использование ИИ-СППР в контексте применения силы создает специфические риски, многие из которых имеют катастрофические последствия.

Во-первых, возможен семантический разрыв и неспособность к контекстуальной оценке. ИИ-СППР способны технически идентифицировать объект, но не понимают его значения или контекста. Ошибочная классификация гражданского человека (например, как противника) может привести к фатальному исходу. Более того, ключевые понятия международного гуманитарного права, такие как «сомнение» в гражданском статусе или «чрезмерность» сопутствующего ущерба, являются оценочными и контекстуальными. Их нельзя свести исключительно к вычислительным алгоритмам без риска утраты самой сути гуманитарной нормы.

Во-вторых, проблема качества и достоверности исходных данных. Алгоритмы обучаются на данных, которые могут отражать и усиливать существующие человеческие предубеждения или порождать новые. В условиях вооруженного конфликта получение репрезентативных и качественных данных для обучения крайне затруднено из-за уникальности каждого конфликта, стремления сторон к скрытности. Прогнозы, основанные на устаревших или нерепрезентативных данных, ненадежны.

В-третьих, непредсказуемость и необъяснимость («черный ящик»). Сложные модели машинного обучения, особенно глубокие нейронные сети, часто не позволяют понять, почему система выдала тот или иной результат. Это подрывает способность человека-оператора критически оценивать рекомендации ИИ, что противоречит требованию МГП о принятии всех практически возможных мер предосторожности для верификации целей.

В-четвертых, уязвимость к «сопоставительным атакам» (adversarial attacks). Впервые они были «зафиксированы в 2013–2014 годах исследователями Google Brain – Кристианом

Середы (Christian Szegedy, англ.) и Иэном Гудфеллоу (Ian Goodfellow, англ., США)» [7]. Adversarial attacks — «это злонамеренное манипулирование входными данными модели машинного обучения с целью заставить ее выдать неправильные предсказания. Эти атаки проектируются, чтобы использовать уязвимости алгоритмов машинного обучения и часто могут быть выполнены незаметно для системы или пользователя» [8].

Противник может намеренно искажать входные данные (например, модифицировать изображение), чтобы обмануть алгоритм классификации. В условиях конфликта, где обман – легитимная тактика, такие атаки становятся высоковероятными.

В-пятых, чрезмерное доверие и автоматизационная предвзятость (automation bias). Человек склонен некритически полагаться на выводы автоматизированных систем, особенно в условиях стресса, нехватки времени и высокой когнитивной нагрузки – типичных для тактического уровня. Это может привести к тому, что оператор станет «кнопкой подтверждения», фактически делегируя решение алгоритму. Риск усугубляется при использовании систем с непрерывным обучением, поведение которых после развертывания может меняться непредсказуемым образом.

В отчете приводится набор взаимодополняющих мер, направленных на смягчение рисков и сохранение осмысленного человеческого суждения. Их можно разделить на технические и организационно-процедурные.

К техническим и проектным мерам следует отнести:

контекстуальное ограничение – использование ИИ-СППР должно быть строго ограничено теми задачами и условиями (театр военных действий, время суток, погода), для которых они были специально и всесторонне протестированы. Система, обученная распознавать танки днем, не должна применяться ночью;

требования к тестированию и сертификации - необходимы строгие, повторяемые и реалистичные протоколы тестирования, учитывающие условия конфликта. Для систем с непрерывным обучением требуется постоянный мониторинг производительности и периодическая переаттестация;

повышение объяснимости – несмотря на сложность, необходимо продолжать исследования в области объяснимого ИИ (XAI) для повышения прозрачности решений. Однако отчет скептически относится к простым решениям, таким как показатели уверенности, которые могут создать ложное ощущение объективности и усилить автоматизационную предвзятость, особенно если оператор не понимает их методологию.

В свою очередь, организационные и процедурные меры включают в себя:

обучение и подготовка операторов – ключевая мера – обучение военных не просто пользованию интерфейсом, а пониманию фундаментальных ограничений ИИ, природы смещений и принципов работы систем. Разработчики должны предоставлять ясные и практические руководства по эксплуатации, описывающие границы возможностей системы;

архитектура принятия решений и «ограничение скорости» – наиболее радикальное предложение – намеренное замедление процесса принятия решений в критических точках. Хотя основной мотив внедрения ИИ – ускорение цикла принятия решения, для выполнения качественных оценок, требуемых нормами международного гражданского права, человеку необходимо время. Процессы должны быть спроектированы так, чтобы обеспечивать «скорость человека», а не «скорость машины», исключая осмысленное вмешательство;

нормативные ограничения – учитывая устойчивость многих рисков, могут потребоваться прямые ограничения на использование ИИ-СППР для определенных типов решений (например, окончательной оценки законности цели), особенно на тактическом уровне, или полный отказ от систем с непрерывным обучением в критически важных приложениях.

Внедрение систем поддержки принятия решений на основе искусственного интеллекта в военную сферу представляет собой парадоксальный вызов. С одной стороны,

эти технологии обладают потенциалом для повышения точности и скорости принятия решения. С другой стороны, их внутренние ограничения – необъяснимость, зависимость от данных, склонность к смещениям – в сочетании с психологией человеко-машинного взаимодействия в стрессовых условиях создают серьезные риски подрыва самого человеческого суждения, являющегося основой законности любого применения силы в соответствии с нормами международного гражданского права.

Как показал анализ положений совместного отчета Международный Комитет Красного Креста (МККК) и Женевской академии международного гуманитарного права и прав человека, различие между системой, которая информирует человека, и системой, которая управляет им, может стать крайне размытым. Поэтому дискуссия не должна сводиться лишь к противопоставлению «хороших» ИИ-СППР «плохим» автономным системам вооружения. Центральным вопросом становится проектирование всей экосистемы процесса принятия решений: технических стандартов, протоколов тестирования, доктрин применения, учебных программ и, в конечном итоге, нормативных рамок, которые гарантировали бы, что технология остается инструментом в руках ответственного человека, а не наоборот.

Гуманитарный императив заключается в том, чтобы опережать технологическое развитие жесткими требованиями к контролю. Следующими шагами, как отмечено в отчете, должны стать углубленные междисциплинарные исследования, конкретизация мер по проверке и сертификации, а также международный диалог о возможных нормативных ограничениях. Цель – не остановить прогресс, а направить его в русло, обеспечивающее сохранение человеческого достоинства и верховенства права даже на поле боя будущего.

Список источников

1. ICRC and Geneva Academy. Expert Consultation Report on AI and Related Technologies in Military Decision-Making on the Use of Force in Armed Conflicts. Geneva: ICRC, March 2024.
2. International Committee of the Red Cross. “ICRC Position Paper: Artificial intelligence and machine learning in armed conflict: A human-centred approach.” International Review of the Red Cross, March 2021.
3. International Committee of the Red Cross. “ICRC position on autonomous weapon systems.” 12 May 2021.
4. Holland Michel, Arthur. Decisions, Decisions, Decisions: Computation and Artificial Intelligence in Military Decision-Making. ICRC, March 2024.
5. Ekelhof, Merel, and Giacomo Persi Paoli. “The human element in decisions about the use of force.” UNIDIR, March 2020.
6. Boulanin, V., and M. Verbruggen. “Mapping the Development of Autonomy in Weapon Systems.” SIPRI, 2017.
7. Богданова Ж. Состязательные атаки (Adversarial Attacks) – что это такое и как они обманывают модели ИИ – Режим доступа: <https://dtf.ru/id2687299/4082105-sostyazatelnye-ataki-kak-obmanyvayut-modeli-ii> (дата обращения 05.12.2025)
8. Введение в Adversarial attacks: как защититься от атак в модели глубокого обучения на транзакционных данных – Режим доступа: <https://habr.com/ru/companies/vtb/articles/718024/> (дата обращения 01.12.2025)

Сведения об авторах

Бойкова Анна Викторовна, доцент, Тверской государственный технический университет, Тверь, Россия.

Никольская Вера Александровна, доцент, Тверской государственный технический университет, Тверь, Россия.

Information about the authors

Boykova Anna Viktorovna, Associate Professor, Tver State Technical University, Tver, Russia.
Nikolskaya Vera Alexandrovna, Associate Professor, Tver State Technical University, Tver, Russia