DOI: 10.17976/jpps/2022.06.06

СМЕРТОНОСНЫЕ АВТОНОМНЫЕ СИСТЕМЫ ВООРУЖЕНИЯ: ЛАНДШАФТ И ПЕРСПЕКТИВЫ

К.В. Богданов, М.Г. Евтодьева

БОГДАНОВ Константин Вадимович, кандидат технических наук, старший научный сотрудник, Центр международной безопасности, ИМЭМО им. Е.М. Примакова РАН, Москва, email: cbogdanov@imemo.ru;
ЕВТОДЬЕВА Марианна Георгиевна, кандидат политических наук, старший научный сотрудник, Центр международной безопасности, ИМЭМО им. Е.М. Примакова РАН, Москва, email: marianna133@mail.ru

Богданов К.В., Евтодьева М.Г. 2022. Смертоносные автономные системы вооружения: ландшафт и перспективы. *Полис. Политические исследования*. № 6. С. 67-80. https://doi.org/10.17976/jpps/2022.06.06

Статья опубликована в рамках проекта "Посткризисное мироустройство: вызовы и технологии, конкуренция и сотрудничество" по гранту Министерства науки и высшего образования РФ на проведение крупных научных проектов по приоритетным направлениям научно-технологического развития (Соглашение № 075-15-2020-783).

Статья поступила в редакцию: 22.04.2022. Принята к публикации: 01.09.2022

Аннотация. Активное развитие военных технологий искусственного интеллекта приближает момент массового развертывания смертоносных автономных систем вооружения (САС). Точное международно признанное определение САС до сих пор отсутствует, что приводит к терминологическим спорам, за которыми скрываются как военно-технические, так и этико-философские проблемы. Задача контроля над автономными системами оружия сложна и должна рассматриваться междисциплинарно, одновременно в нескольких плоскостях, включая соображения оперативного применения, морально-этические дилеммы и правовые коллизии. Все три типа соображений фокусируются вокруг проблем значимого человеческого контроля и атрибуции ответственности за применение САС. Показана проблема фиктивности номинального человеческого контроля над сложными техническими системами, использующими непрозрачные алгоритмы функционирования. Значительное сопротивление разработчиков новых видов оружия, доказывающих достаточность существующих норм международного гуманитарного права для регулирования новой области, затрудняет введение дополнительных ограничений на САС. Существующие соглашения по контролю над распространением технологий лишь частично способны служить инструментом для контроля над САС. Отмечено, что в случае с САС возможен мощный импульс к развитию особого направления в контроле над вооружениями, связанного с юридическими аспектами атрибуции ответственности за неправомерное применение систем оружия, чего не наблюдалось прежде в корпусе международно-правовой базы по контролю над вооружениями. В условиях обрушения системы контроля над вооружениями, сложившейся после окончания холодной войны, в кратко- и среднесрочном периоде могут быть предложены определенные частные меры предотвращения наиболее дестабилизирующих проявлений САС.

Ключевые слова: беспилотные системы, искусственный интеллект, контроль над вооружениями, международное гуманитарное право, оговорка Мартенса, смертоносные автономные системы, человеческий контроль.

Тема разработок, принятия на вооружение, регулирования и применения в военной сфере вооружений, способных за счет элементов искусственного интеллекта (ИИ) к самостоятельному действию, становится все актуальнее. Согласно определению, утвердившемуся в дискуссиях на площадках ООН, их именуют смертоносными автономными системами вооружения (САС).

Активно продвигаются разработки, необходимые для создания эффективных САС. Уже есть базовые пакеты технологий и платформы (беспилотные/необитаемые аппараты, наземные дистанционно-управляемые роботы), которые при соответствующей модернизации бортовых систем превратятся в САС. Есть примеры систем с простейшими алгоритмами боевой работы (в частности, "интеллектуальные" средства охраны военных объектов, способные самостоятельно применять оружие). При этом накапливаются разнородные проблемы, связанные с САС.

Наиболее ярко они отразились в дискуссиях в Группе правительственных экспертов ООН по САС (ГПЭ), созданной в рамках Конвенции 1981 г. о негуманном оружии (КНО). За несколько лет работы (с 2014 г. фактически, с 2016 г. с мандатом) ГПЭ так и не согласовала общепризнанную дефиницию САС и не установила, какие именно вооружения к ним относятся. Нет консенсуса и в части этических принципов применения САС, путей разрешения правовых коллизий, а также контроля над такими вооружениями.

Настоящая работа призвана рассмотреть весь многомерный ландшафт проблематики САС, проанализировать взаимодействие ключевых проблемных блоков в дискуссиях по САС, а также обсудить вопросы контроля над ними. Предлагаемая междисциплинарная теоретико-методологическая модель рассмотрения САС на основе системного подхода соединяет проблемное поле, действующие факторы и ключевые понятия, относящиеся к оперативным, правовым и этико-гуманитарным проблемам разработки и применения САС. В рамках этих трех концептуальных направлений анализируются понятия автономности и различные ее аспекты, человеческого контроля, а также боевого применения САС. Такая модель проявляет картину в целом, способствует выявлению технических проблем САС, которые не актуализируются в работах экспертов по правовым и гуманитарным аспектам САС, и наоборот. Это позволяет вскрыть пласт фундаментальных проблем, препятствующих развитию контроля над автономными системами, и определить их взаимозависимости.

Определение и ключевые проблемы развития САС

Дискуссии об определении САС показывают, что споры о терминах, часто воспринимаемые как крючкотворство и намеренное затягивание процесса, способны выявить ключевые проблемы и значительно влияют на выработку подходов к контролю над вооружениями.

Согласно одному из широко признанных рабочих определений, данному в 2013 г. специальным докладчиком ООН К. Хейнсом[1] и в дорожной карте по беспилотным системам Минобороны США 2013 г.[2], САС – это "роботизированные системы вооружения, которые после приведения в действие могут выбирать и поражать цели без последующего вмешательства со стороны оператора".

Эксперт ГПЭ ООН Л. Сачмен уточнила, что под САС следует понимать лишь те системы оружия, которые без участия человека-оператора способны распознавать и брать на сопровождение в качестве целей людей (комбатантов

[1] Доклад Специального докладчика по вопросу о внесудебных казнях, казнях без надлежащего судебного разбирательства или произвольных казнях Кристофа Хейнса. Совет по правам человека ООН, 23-я сессия. Документ A/HRC/23/47. 09.04.2013. https://undocs.org/A/HRC/23/47 (accessed 10.04.2022).

[2] US Department of Defense. *Unmanned Systems Integrated Roadmap. FY 2013-2038*. https://info.publicintelligence.net/DoD-UnmannedRoadmap-2013.pdf (accessed 10.04.2022).

либо гражданских лиц), а также инициировать применение против них силы[3]. Это дает так называемое *широкое* определение САС, которое вынужденно включает в сферу охвата беспилотные летательные аппараты (БПЛА), необитаемые (безэкипажные) наземные и морские аппараты. В формальные рамки определения также попадают самонаводящиеся управляемые ракеты, барражирующие боеприпасы ("дроны-камикадзе") и прочие виды оружия с высокоавтоматизированными процессами поиска и наведения на цель, способные автономно маневрировать и атаковать выбранную цель без внешнего источника целевой информации благодаря своим бортовым сенсорам.

На ход практических дискуссий о САС существенно влияют две методологические неопределенности [Ivanov, Korzhenyak, Lapikhina 2021: 12-13]. Это, во-первых, вопрос о различении "автономного" и "высокоавтоматизированного" поведения. Во-вторых, чем "высокоавтоматизированная" работа систем управления вооружением отличается от искусственного интеллекта? Их практическая значимость хорошо видна на примере систем с самонаведением, которые подпадают под "широкое" определение, но САС не являются. Есть и ударные дроны, использующие алгоритмы автономного поведения для запрограммированных действий при потере связи с оператором или просто запускающиеся в район применения с включенными системами самонаведения.

Военное применение автономной техники подразумевает широкий спектр операций: это роботы, включая подводные, осуществляющие видео- и фотосъемку и разминирование; автоматизированные системы управления полетом в современной авиации; перспективные автономные средства медико-санитарной эвакуации с переднего края или из вражеского тыла и пр. Главное отличие САС в том, что они имеют автономные или полуавтономные (находящиеся под операторским надзором) системы нацеливания [Boulanin, Veerbruggen 2017: 5-7] для применения вооружения – другими словами, способны самостоятельно селектировать цели в окружающей обстановке. Поэтому ограничения – например, требование подконтрольности САС человеку-оператору, – касаются именно этих систем с селекцией целей, а не всех автономных систем в целом.

В то же время был выработан более узкий подход к определению САС как оружия, "запрограммированного на обучение или адаптацию своего функционирования к изменению условий среды" [Энтони, Холланд 2015: 457-458]. Он более точно соотносится с современным пониманием автономии систем вооружений и ее связи с самообучением, а также дает свой ответ на развернувшуюся в ГПЭ ООН дискуссию об опасности САС и о том, какие их виды следует запрещать или ограничивать. "Широкий" подход выглядит неконкретным и значительно наращивает число систем, подлежащих регулированию, что снижает вероятность достижения консенсуса о режимах контроля над САС. Поэтому авторы данной работы придерживаются описанного выше "узкого" определения этих систем, дополняя его акцентами на достижимых степенях автономности.

Так, если исходить из классификаций четырех поколений роботизированных (беспилотных) систем по степени автономности как *автоматизированных, адаптивных, интеллектуальных* (обладающих системой управления с элементами ИИ) и *автономных*[4], то в настоящее время на вооружении развитых

[3] Suchman L. Situational awareness and adherence to the principle of distinction as a necessary condition for lawful autonomy. CCW informal meeting of experts on lethal autonomous weapons, panel 'towards a working definition of LAWS'. Geneva, 12.04.2016. P. 3. https://www.research.lancs.ac.uk/portal/services/downloadRegister/167586397/CCW_Autonomy_Suchman.pdf (accessed 10.04.2022)

[4] Кондратьев А.Е. Боевые роботы США – под водой, в небесах и на суше. *Независимое военное обозрение*, 14.05.2010. https://nvo.ng.ru/armament/2010-05-14/8_robots.html (accessed 10.04.2022); Сизов В.Ю. Какие роботизированные комплексы нужны России? Тезисы выступления на круглом столе "Боевые

стран находятся в основном системы 1-го и 2-го поколения. К ним относится значительная доля разработанных и принятых на вооружение БПЛА и необитаемых наземных и морских аппаратов. Руководствуясь "широким" определением, к системам 1-го и 2-го поколения следовало бы также отнести и все управляемые ракеты с автономным самонаведением, а также барражирующие боеприпасы. Но если исходить из того, что САС должны уметь адаптировать поведение при изменении условий среды, то ни БПЛА, ни управляемые ракеты и барражирующие боеприпасы как таковые в перечень САС включены быть не могут (хотя они могут выступать в роли платформ).

Другими словами, *подлинно автономными системами вооружения* могут считаться лишь те, которые относятся к 3-му (и потенциально 4-му) поколению. Речь идет о системах с элементами ИИ, способных обучаться не в ходе, а еще до начала выполнения боевой задачи, при ее постановке. Такие типы систем могут соединять в себе возможности передовых технологий машинного обучения, автономной навигации, визуального распознавания объектов, высокоточного ударного вооружения и др. Для эффективной работы им необходимы сложное программное обеспечение, требующее мощной вычислительной базы [Макаренко 2016: 75], а также многоканальные средства сбора данных об окружающей обстановке. К боевым автономным системам с элементами ИИ можно отнести, например, микро-беспилотники, работающие в режиме "роя", разработки по которым активно идут в США и Китае [Контроль над вооружениями... 2020: 64].

По прогнозам специалистов, широкое распространение полностью автономные боевые системы смогут получить не раньше, чем через 20 лет, – из-за недостаточной зрелости технологий и сложности решаемых задач [Макаренко 2016]. Так, для боевых систем важны высокая точность, чтобы не нарушать законодательство, в том числе нормы международного гуманитарного права (МГП). Максимальная эффективность требует их включения в единую систему боевого управления, связи, разведки и наблюдения на театре военных действий (C4ISR).

При этом САС может и не представлять собой единого физического объекта в виде оборудования и оружия, комплексированных на одной платформе (например, в виде БПЛА). Она может быть *пространственно распределенной* – в том числе глобально, за счет спутниковых каналов связи – собирая воедино управляющий центр с мощной вычислительной базой (ИИ) и действующие на театре военных действий необитаемые аппараты с бортовыми сенсорами и вооружением, управляемые по командам из центра. Такая система, за исключением блока с ИИ, архитектурно уже выстроена в США и используется для дистанционного управления БПЛА человеком-оператором, находящимся в соседнем полушарии (ударные дроны, действующие на Ближнем Востоке, например, управляются операторами из командного центра авиабазы Крич в штате Невада). Подобная архитектура дополнительно осложняет точную дефиницию САС и определение юрисдикции при расследовании неправомерного применения.

В числе ключевых дискутируемых аспектов "институционализации" САС находится вопрос о *"значимом человеческом контроле"* (*meaningful human control*). Одно из наиболее емких и кратких его определений звучит так: "способность [человека-оператора] принимать своевременные и осознанные решения относительно работы систем на основе ИИ, обеспечивающие лучший

роботы в войне будущего: выводы для России", Москва, 11 февраля 2016 г. *Независимое военное обозрение*, 04.03.2016. https://nvo.ng.ru/armament/2016-03-04/5_robots2.html (accessed 10.04.2022).

операционный результат из возможных”[5]. Иногда определение раскрывают, включая в него наличие у операторов достаточной информации и соответствующего уровня подготовки, а также требование, чтобы оружие успешно прошло полный цикл испытаний согласно установленным нормам.

Сама идея значимого контроля интуитивно очевидна, но его теория по-прежнему не проработана на уровне, способном дать универсальные критерии, достаточные для применения при создании автономных систем вооружений. Философские проблемы, восходящие к абстрактным диспутам в отношении противоречий между детерминизмом и свободной волей (релевантные задаче создания автономно действующих искусственных объектов), нашли отражение в работах Дж. Фишера и М. Равиццы [Fisher, Ravizza 1998]. Они создали теорию “направляющего контроля” – способности к самостоятельному поведению с полной моральной ответственностью за результат даже в условиях фактического отсутствия альтернатив. Это направление получило широкое применение при анализе проблем автономных систем оружия; развитие его шло в работах Ф. де Сио и Й. ван ден Ховена [De Sio, van den Hoven 2018].

Различие по уровню контроля дает три возможные формы отношения человека к автономной системе [Marra, McNeil 2012: 1144-1148], использующие в своей основе разделение ответственности в рамках “петли Бойда” (“наблюдение – ориентация – решение – действие”) [Osinga 2007]. Первая – “человек в управлении” (*human in the loop*) – подразумевает автономность системы только в пределах сбора информации и выработки интерпретаций обстановки; все решения принимаются оператором. Вторая – “человек над управлением” (*human on the loop*): автономная система действует самостоятельно как в сборе и интерпретации данных об обстановке, так и в нацеливании оружия и решении на его применение, а человек выполняет надзорную функцию, сохраняя возможность вмешательства в любой момент цикла. Третья – “человек вне управления” (*human out of the loop*): подлинно автономный цикл работы САС на основании алгоритмов ИИ, без вмешательства человека.

Оперативные, гуманитарные и правовые проблемы применения САС

Необитаемые системы играют все более важную роль в разведке, постановке помех и обеспечении связи, выборе целей и нанесении ударов по ним. В операциях определенного типа наблюдается тенденция к замещению пилотируемых аппаратов беспилотными. Разрабатываются схемы совместного применения БПЛА с пилотируемыми самолетами тактической авиации и боевыми кораблями. В скором времени, как ожидается, такие разведывательно-ударные сетевые комплексы станут обыденным элементом войны [Контроль над вооружениями... 2020: 63-64].

Проблематику боевого применения САС можно разбить на следующие категории: оперативные (военно-технические) проблемы, морально-этические (гуманитарные) и правовые. Весь цикл разработки, применения и ограничения САС с системной точки зрения следует рассматривать как единое проблемное поле, в котором взаимодействуют данные факторы. Все они связаны с реализацией значимого человеческого контроля, и в той или иной степени адресуются к прозрачности, надежности и предсказуемости технологий [Autonomy, Artificial Intelligence... 2019: 2-3, 10-13]. Прозрачность подразумевает четкость

[5] Boardman M., Butcher F. An exploration of maintaining human control in AI enabled systems and the challenges of achieving it. MP-IST-178-07 NATO Report, 2019. P. 7-1. https://www.sto.nato.int/publications/STO%20Meeting%20Proceedings/STO-MP-IST-178/MP-IST-178-07.pdf (accessed 10.04.2022).

и понятность для человека архитектуры системы, ее алгоритмов и внутренних критериев целеполагания (в том числе по итогам самообучения), а также возможность их извлечения и корректной интерпретации человеком на любом этапе жизненного цикла. Надежность – это сохранение ключевых принципов работы при технических отказах или вследствие частичного поражения противником. Предсказуемость связана с устойчивостью исполнения поставленной задачи и стабильностью реакции САС на внешние раздражители.

К числу важнейших *оперативных проблем* использования САС в первую очередь следует отнести вопросы (а) обеспечения точности применения по боевой технике и (б) включенности человека-оператора в контур принятия решений на применение оружия.

Проблемы с точностью применения возникают не только из-за недостатков систем отображения сведений, когда без подтверждения информации по независимому каналу невозможно точно сказать, о каком именно объекте подаются данные. Операторы, находящиеся на удалении от места действия, при выборе вариантов дальнейших действий или оценке обстановки в большинстве случаев соглашаются с рекомендациями компьютера. Большая психофизическая нагрузка, нехватка времени, значительный объем и специфичность информации ведут к тому, что операторы склонны пренебрегать ее анализом и игнорировать "двусмысленность" интерпретаций обстановки автоматизированной системой [Bode, Watts 2021: 63].

Даже добросовестный оператор может не располагать возможностями проверить адекватность оценки ИИ. Например, если та ведется не путем распознавания и синтеза образов из данных, полученных через несколько информационных каналов (оптический, радиолокационный, тепловой, акустический и пр.), а через набор специфических сочетаний вторичных признаков, которые экспертная система ИИ считает достаточным для идентификации цели (скажем, "основной боевой танк" с вероятностью 90%). Ситуация кратно усложняется в системах с самообучением в процессе работы: в какой-то момент в них могут возникнуть правила, которые не транслируются в человеческий язык и логику и потому не могут быть верифицированы даже при наличии желания и возможности [McGregor, Murray, Ng 2019: 319]. Оператор системы, строго говоря, не обязан иметь квалификации ее разработчика и понимать архитектуру на уровне внутренних алгоритмов и специфики их взаимодействия, а в целом ряде случаев это и невозможно по соображениям военной и коммерческой тайны. Таким образом, формально предусмотренная для оператора возможность остановить САС или дать ей разрешение на применение оружия не решает проблему избирательности и правомерности этого применения. Более того, она фактически маскирует эту проблему за ширмой номинального человеческого контроля, по сути, предоставляя машину самой себе.

Развитие военного ИИ за пределами проблематики САС тоже оставляет вопросы фиктивности вовлечения человека-оператора в значимый контроль. Например, по мнению ряда американских экспертов [Kania 2017: 14], в Китае выбран особый путь развития военного ИИ, подразумевающий максимально допустимое исключение человека из цепочки обнаружения, нацеливания и управления огнем – в интересах сокращения цикла боевого управления (времени от момента обнаружения цели до ее фактического уничтожения). Речь идет о развитии автоматизированной штабной экспертной системы, определяющей приоритеты целераспределения и оптимизирующей наряд сил и средств на поражение целей – с минимальным участием человека или даже без него.

Применение САС выдвигает на передний план и *гуманитарные (морально-этические) проблемы*. Техническое разъединение человека-оператора и его оружия создает принципиально новый класс этических дилемм. Проблема фиктивности человеческого контроля – в той же степени гуманитарная проблема, в коей и оперативная. "Ползучий" вывод машины с летальным оружием из-под реального управления человеком становится не просто военно-технической, а философской проблемой. Многие противники использования таких боевых систем считают, что делегирование подобных решений машине – "оскорбление человеческого достоинства"[6].

Существует точка зрения, согласно которой САС является спасением от эксцессов войны: машина лишена эмоций и не способна к психологическим срывам и ошибочным решениям из-за усталости. Точность заложенного алгоритма (закона войны) обеспечивает точность следования ему, что на первый взгляд можно рассматривать как прогрессивное достижение. Однако, оставив юридические вопросы пока в стороне, зададимся этическим: может ли международное гуманитарное право быть превращено в формальный алгоритм без потери заложенного в него смысла?

Этой проблеме посвящались длительные дискуссии, в том числе по вопросу об уточнении собственно положений МГП для его прояснения. На современном этапе главный вывод не благоприятствует идее "автоматизации права". МГП в значительной степени основывается на здравом смысле, имплицитно подразумеваемых цивилизационных ценностях и сложившемся обычае ведения войны – как в случае с "оговоркой Мартенса", лежащей в фундаменте этого права. Недвусмысленная кодификация этих норм в формальный алгоритм не представляется возможной, пока ИИ не обретет способность к обучению и накоплению знаний, сравнимую с человеческой.

На данный момент, однако, есть и более простые и насущные гуманитарные проблемы, которые уже видны в практике применения ударных БПЛА. Так, отсутствует минимальная транспарентность при анализе решений на боевое использование дронов, приведших к гибели гражданских лиц, – включая и намеренное введение в заблуждение. С правовой точки зрения это можно квалифицировать как "*внесудебные казни*", которые являются незаконными (и это неоднократно подчеркивалось в докладах специальных докладчиков ООН по внесудебным казням – например, в последней версии такого доклада 2020 г.[7]).

"Распределенное" применение оружия, когда оператор не участвует непосредственно в боевых действиях, создает разрыв между местом принятия решения и местом применения, с соответствующими изменениями в оценке ситуации и саморефлексии оператора. Это в корне меняет морально-этический микроклимат в коллективах и создает поверхностное отношение к войне. Лично оператор находится в безопасности, ему не грозит быть убитым или искалеченным по случайности или в результате удара возмездия, он не видит всех ужасов войны [Benjamin 2013]. Война "геймифицируется": воспринимается как компьютерная игра, что поощряет бесконтрольность и безответственность

[6] Шарки Н., Канунникова Т. Автономное оружие представляет угрозу для международной безопасности. *РСМД*. https://Russiancouncil.Ru/Analytics-And-Comments/Columns/Military-And-Security/Avtonomnoe-Oruzhie-Predstavlyaet-Ugrozu-Dlya-Mezhdunarodnoy-Bezopasnosti/ (accessed 10.04.2022).

[7] Использование ударных беспилотников для целенаправленных убийств. Доклад Специального докладчика по вопросу о внесудебных казнях, казнях без надлежащего судебного разбирательства или произвольных казнях. Совет по правам человека ООН, 44-я сессия. Документ A/HRC/44/38, 17.08.2020. https://undocs.org/ru/A/HRC/44/38 (accessed 10.04.2022).

поведения операторов, особенно на фоне фиктивности человеческого контроля. При этом оборотной формой "геймификации" является тот факт, что операторы все-таки испытывают тяжелый стресс и даже получают посттравматические стрессовые расстройства, как показывает опыт операторов дронов с упомянутой выше авиабазы Крич. Однако механизмы и динамика такого рода психологических травм отличаются от тех, что "зарабатываются" бойцами на передовой[8], и в силу сложившегося легковесного отношения начальства к микроклимату в тыловых подразделениях этот фактор часто упускается из виду. Это ведет к накоплению психологических проблем личного состава и, как результат, – неизбирательному и неправомерному применению оружия [Lee 2018].

Правовые проблемы САС начинаются с того, что их развертывание не представляется юридически допустимым без четкого соотнесения со статьей 36 "Новые виды оружия" Дополнительного протокола I к Женевским конвенциям 1949 г. Она в явном виде запрещает развертывать новое оружие без проверки на возможность нарушения им МГП. Но как отмечено выше, проблема с кодификацией МГП неизбежно ведет и к проблемам с его исполнением. Нельзя гарантировать ни точность применения САС, ни то, что оно соответствует правилам ведения войны. (Например, потому, что у независимых экспертов нет возможности официально оценить работу алгоритмов в условиях неопределенности боевой обстановки – в особенности в случае с самообучающимися системами.)

С точки зрения МГП, приложимого к ситуациям или действиям, квалифицирующимся как международный вооруженный конфликт, важнейшими критериями в отношении САС становятся обеспечение *избирательности* их применения (исключение атак против некомбатантов) и его *соразмерности* (предотвращения случайной гибели/ранений гражданского населения, разрушения жизненно важной инфраструктуры и т.п.). В ситуации вооруженного конфликта "*целенаправленное убийство"* (*targeted killing*) является законным лишь тогда, когда цель – "комбатант", т.е. человек, который непосредственно участвует в военных действиях [Benjamin 2013; Голубенко 2017]. В этом смысле любые ошибки в применении САС (случайное либо непреднамеренное использование против гражданских лиц) будут нарушать МГП.

Ответственность за нарушение МГП в случае с САС выглядит размыто. Можно назвать сразу несколько потенциально ответственных: операторы, введшие систему в действие; лица, отдавшие приказ о нанесении удара; и разработчики алгоритмов, на основании которых совершено нарушение [Davison 2017: 17]. Значимый человеческий контроль подразумевает наличие в цикле боевого управления осознанной воли оператора. Это влечет за собой его ответственность – но, как мы уже показали, значимость такого контроля легко может быть поставлена под сомнение. В том числе и юридически, если в момент разбирательства будет доказано, что технические особенности САС, заложенные разработчиком, не позволяли оператору на практике осуществлять такой контроль и выносить информированные суждения о последствиях своих действий.

Кроме того, системы с ИИ упрощают для личного состава вооруженных сил снятие с себя правовой ответственности за военные преступления. Причем речь идет не только о непреднамеренных ошибках, приведших к гибели некомбатантов или разрушению гражданской инфраструктуры, но и о воз-

[8] Press E. The Wounds of the Drone Warrior. *The New York Times*, 13.06.2018. https://www.nytimes.com/2018/06/13/magazine/veterans-ptsd-drone-warrior-wounds.html (accessed 10.04.2022).

можном целенаправленном использовании САС в этом качестве – с последующей попыткой создать юридически непрозрачную ситуацию с атрибуцией ответственности за совершенное. То есть факт наличия в войсках боевых роботов может поощрять к военным преступлениям.

ПРОМЕЖУТОЧНОЕ ОБСУЖДЕНИЕ ПРОБЛЕМАТИКИ САС

Как видно из рассмотренного, основным препятствием к регулированию САС является *высокая комплексность проблемы*, сочетающей в себе разнородные факторы. Помимо уже рассмотренной модели из трех взаимозависимых пространств проблемного поля, речь идет также о переплетении гражданских и военных технологий, а также распределенном характере архитектуры интегрированных разведывательно-ударных комплексов, в чью идеологию отлично вписываются САС. При выборе подхода к регулированию следует учитывать, что исключение любого из этих факторов из фокуса рассмотрения способно обесценить принимаемые решения в целом.

Сочетание очевидной заинтересованности многих стран в повышении эффективности своего высокоточного оружия с многогранным характером проблемы и сложностью подготовки методологической базы ограничений и запретов задает сложную динамику гонки вооружений в области САС. Заключение многостороннего соглашения о запрете либо сильном ограничении САС (по аналогии, например, с Договором по ПРО 1972 г.) не представляется возможным. Скорее, речь пойдет об ограничениях явочным порядком по результатам первой волны практических внедрений САС и наработки опыта войсковой эксплуатации. При этом если проблемы, сопровождающие практическую отработку САС в войсках, окажутся серьезнее, чем предполагается сейчас, а их боевая эффективность, с учетом возможных мер борьбы со стороны противника и механизма "действие–противодействие", ниже, – то шансы на быстрое заключение рамочных соглашений довольно велики. В этом случае стороны позволят себе сфокусироваться на долгосрочных НИОКР по отработке САС, сравнительно легко согласившись на ограничения таких систем оружия.

Переплетение разнородных факторов создает слабо разрешимые правовые коллизии контроля над САС как на национальном, так и на международном уровне. Возможно перекладывание ответственности – когда, например, при неизбирательном поражении гражданских лиц операторы будут обвинять самообучающуюся систему с ее алгоритмами, а разработчики – плохую подготовку и безответственность операторов. В отличие от применения кибервооружений, где почти невозможно атрибутировать злоумышленника, при использовании САС ключевая проблема иная: факт атаки зафиксирован, известен круг вовлеченных лиц, но ответственность за нее крайне сложно квалифицировать юридически. Число таких коллизий и споров может вырасти, в том числе из-за широких обсуждений характеристик и особенностей применения САС на международном уровне, а также активной деятельности НПО и экспертов, добивающихся максимальной транспарентности и регулирования (запрета) использования.

В конечном счете это неизбежно приведет к развитию отдельного значимого направления, связанного с чисто юридическими проблемами контроля над применением военных технологий. Любые перспективные методы и формы регулирования или контроля над САС должны будут учитывать, во-первых, растущие требования к транспарентности данных о применении САС. Во-вторых, требуется разрешать юридические коллизии при атрибуции ответственности за использование, обеспечивать контроль над технологиями и их распространением.

Принимая во внимание изложенное в этом разделе, рассмотрим прикладной аспект проблемы: подходы к контролю над САС на международном уровне.

МЕРЫ КОНТРОЛЯ: СОСТОЯНИЕ И ПЕРСПЕКТИВЫ

Экспертные дискуссии о контроле над САС ведутся давно. Смежные с ними вопросы регулирования БПЛА рассматривались в 2010-2014 гг. в рамках Совета ООН по правам человека [Энтони, Холланд 2015: 461-462], а с 2014 г. особенности САС, развитие их "автономности", а также вопросы правового регулирования и ограничения обсуждаются в рамках КНО и Группы правительственных экспертов КНО по САС.

В силу сложности и противоречивости проблематики результат пока что скромен. К нему можно отнести консенсус, достигнутый в 2019 г. государствами-участниками КНО относительно 11 руководящих принципов регулирования САС (соблюдение норм МГП, ответственность человека за применение оружия и др.) [Пелдан Карлсон, Буланен 2021: 584-586]; согласие участников ГПЭ с разноплановым пониманием определений САС; а также согласованную в 2019 г. в рамках ГПЭ по САС замену термина "человеческий контроль" на "взаимодействие (этапы взаимодействия) человека и машины"[9].

Не проработаны многие *правовые моменты*. Во-первых, если исходить из довольно популярной идеи о том, что для регулирования САС достаточно уже имеющихся норм МГП, то остается неясным, какие субъекты в цикле боевого управления несут ответственность за неправомерное или ошибочное применение. Даже в наиболее простом примере обособленной высокоавтономной системы такими субъектами могут быть не только операторы, но также командиры, поставившие боевую задачу, высшие представители военной организации или государство-участник конфликта само по себе. В случае с пространственно распределенной боевой системой с элементами в виде САС в этот список можно добавить и операторов всей сети боевого управления. Если же речь идет о подлинно автономных, непрерывно самообучающихся системах с ИИ, то в силу сложности алгоритмов могут быть также поставлены вопросы об ответственности разработчиков за точность предоставляемых технических данных и адекватность методики испытаний – а, следовательно, за точность и избирательность боевого применения САС.

В последние годы с подачи НПО, а также многих стран, не являющихся ключевыми разработчиками и производителями САС, наблюдались попытки ввести мораторий или полный запрет на разработку и использование такого оружия. Аргументами выступали упомянутые выше оперативные, морально-этические и правовые проблемы в различных сочетаниях. Еще в 2014 г. Австрией на неофициальной встрече экспертов КНО было предложено ввести мораторий на САС, что отверг ряд стран-разработчиков и эксплуатантов таких систем, включая США, Израиль и (на тот момент) Францию [Энтони и др. 2016: 635-636]. Предложения об ограничении использования САС активно продвигались Международным комитетом Красного Креста, Кампанией "Остановить роботов-убийц" и значительным числом других НПО[10].

[9] Смысл замены в том, что контроль человека над САС не должен трактоваться противоречиво. Требуется всеобъемлющая норма, подразумевающая гарантии того, что военные возможности САС контролируются людьми *на всех этапах* проектирования, разработки, обучения и использования и соответствуют международному и национальному законодательству. Речь, таким образом, идет о существенном расширении круга лиц, потенциально ответственных за неправомерное применение САС [Пелдан Карлсон, Буланен 2021: 590-591].

[10] Кампания "Остановить роботов-убийц" после неудачных попыток добиться ограничений САС в рамках КНО предполагала перенос обсуждения на уровень Генеральной Ассамблеи ООН или в отдельный процесс с последующим выходом на подписание самостоятельной конвенции.

К концу 2019 г. многие государства-участники конвенции (Австрия, Китай с его "особой позицией" по САС, Алжир, Египет и ряд других стран Движения неприсоединения) призывали к согласованию – в форме протокола к КНО или независимого договора, – юридически обязывающего инструмента, который запрещал бы САС или, в более мягком варианте, налагал обязательства при их использовании. Другая группа стран (включая Францию и Германию) предлагала согласовать политически обязывающую декларацию, которая легла бы в основу кодекса ответственного поведения в отношении САС, опирающегося на руководящие принципы.

Однако обе концепции неизменно наталкивались на сопротивление ряда ведущих военных держав – например, США, Израиля, России и Австралии, – которые продолжают заявлять на официальном уровне, что "существующее международное гуманитарное право полностью отвечает на вызовы, связанные с САС" [Пелдан Карлсон, Буланен 2021: 593]. Этими же странами приводятся аргументы о том, что ограничения в отношении САС подразумевают и ограничения составных технологий (ИИ, сенсоров, визуализации данных, машинного обучения и т.п.), включая гражданское их измерение, а развитие этих технологий, как и автономных систем в целом, имеет большое положительное значение.

Обострившаяся с начала 2022 г. конфронтация России и Запада (НАТО) представляется иррелевантной проблеме САС. Как было показано выше, разделение позиций связано не с военно-политической ориентацией, а с интересом к соответствующим НИОКР. Это парадоксальным образом упрощает взаимодействие стран по поэтапной разработке режимов контроля над САС. Наличие общих интересов теоретически позволяет исключить или снизить вероятность эмоциональной политизации вопроса, потенциально способно помочь формированию единой позиции и послужить одним из каналов возобновления прагматического стратегического взаимодействия. Для этого, впрочем, сначала необходимо согласовать универсальную методологию регулирования САС, что потребует серьезного вовлечения в диалог по существу, притом многосторонний.

Налицо сложившийся пул стран, которые создают и будут продолжать создавать автономные боевые системы (включая и Китай, несмотря на его официальную позицию по юридически обязывающему соглашению), а также высокий спрос на закупки и разработки САС. Прямой запрет не имеет перспективы, идеи о нем носят преимущественно политико-пропагандистский характер. Такая позиция, как и в случае с ядерным аболиционизмом стран-разработчиков и подписантов Договора о запрете ядерного оружия (ДЗЯО), не позволяет консолидированно отделить в ней важное от второстепенного и уводит в сторону от предметного обсуждения технических и юридических рамок регулирования.

В качестве таких рамок, на наш взгляд, следует рассматривать правовые и верификационные вопросы, включая определение того, в какой форме и за какими субъектами должна закрепляться юридическая ответственность за применение САС. Работу алгоритмов САС верифицировать в этом смысле крайне сложно, если вообще возможно. Однако может быть определен набор соответствующих правил "разрешенного" или "запрещенного" поведения. Может быть введен и верифицирован (в рамках моратория, кодекса поведения или конвенции) запрет на оснащение САС оружием массового уничтожения, что тесно соотносится с задачами режимов нераспространения, разоружения и мер по снижению ядерных рисков. Такого рода предложения, на наш взгляд, могли бы получить поддержку многих стран, в том числе при рассмотрении в рамках КНО.

Еще одним возможным вариантом подхода к верификациии САС может быть обмен видеозаписями, подтверждающими факт контроля со стороны оператора. Это своего рода аналог передачи данных телеметрии, применявшихся для целей контроля в рамках соглашений о стратегических наступательных вооружениях. Такой обмен данными, фактически фиксирующими ответственность оператора, мог бы потенциально лечь в основу политических или юридических обязательств по контролю над применением САС. Существенно препятствует введению этих мер то, что страны, использующие автономные системы, считают обмен такой информацией угрозой своей безопасности[11].

Определенные возможности ограничения САС дают режимы экспортного контроля и ряд сохраняющихся направлений контроля над вооружениями. Однако де-факто они ограничены либо сферой охвата (когда им подотчетны не все системы или лишь некоторые страны), либо низкой эффективностью контрольных и верификационных механизмов.

Режим контроля за ракетными технологиями (РКРТ) с участием 35 стран мира налагает ограничения на экспорт только ракет и БПЛА с дальностью полета не менее 300 км, но не включает небольшие тактические ракеты и БПЛА, которые играют все более важную роль в боевых действиях. Например, под РКРТ по своим параметрам не подпадают ударные БПЛА “Байрактар”, успешно применявшиеся в последние годы в конфликтах на территории Сирии, Ливии, Нагорного Карабаха и Украины. В рамках Вассенаарских договоренностей (ВД), участниками которых являются 42 страны, а также Договора о торговле оружием (ДТО) с участием 110 стран, в контрольные списки внесены только БПЛА самолетного типа и некоторые морские системы [Евтодьева,Ознобищев 2019]. Ни РКРТ, ни ВД не включают Израиль и Китай – крупнейших производителей беспилотных и ракетных автономных систем, что делает распространение такого оружия бесконтрольным.

Малоэффективна и работа большинства режимов экспортного контроля. Это касается в первую очередь контроля над передачами “неосязаемых” технологий: программного обеспечения для автономных систем, технологий их производства (3D-принтеров, систем дополненной реальности и т.п.). Возможный рост эффективности регулирования САС может лежать в вовлечении в эти режимы стран – ключевых разработчиков и производителей САС, а также в повышении эффективности контроля над передачами автономных систем и их технологий.

Проблематика САС может рассматриваться как попадающая в сферу охвата диалога о стратегической стабильности, а также соглашений по контролю над обычными вооружениями в Европе (ДОВСЕ и Венский документ 2011 г.). К сфере стратегической стабильности (скажем, перспективного Договора СНВ на период после 2026 г.) могут иметь отношение стратегические САС – БПЛА и морские комплексы, которые способны выступать в качестве средств доставки ядерного оружия [Контроль над вооружениями… 2020: 63]. В случае восстановления работы на европейском пространстве тех или иных механизмов контроля над обычными вооружениями и сопутствующих мер доверия можно было бы также включить в сферу их охвата определенный набор САС, ограничив возможности их размещения и накопления в регионе. Аналогичное решение принесло бы пользу и в Азиатско-Тихоокеанском регионе, но там оно осложняется отсутствием опыта заключения и имплементации подобных многосторонних соглашений.

[11] Козюлин В. Автономные вооружения и законы войны. *Валдайский клуб*, 09.02.2021. https://ru.valdaiclub.com/a/highlights/avtonomnye-vooruzheniya-i-zakony-voyny (accessed 10.04.2022).

Поскольку значимая модернизация указанных соглашений с включением в нее проблематики САС в среднесрочной перспективе представляется маловероятной, на текущем этапе акценты следовало бы сделать на точечных политических инициативах и заявлениях, снижающих риски дестабилизации. Наиболее значимым примером могло бы стать политическое заявление России и стран НАТО о неразмещении ядерного оружия на борту САС.

DOI: 10.17976/jpps/2022.06.06

LETHAL AUTONOMOUS WEAPON SYSTEMS: LANDSCAPE AND PERSPECTIVES

K.V. Bogdanov[1], M.G. Yevtodyeva[1]

[1] Primakov National Research Institute of World Economy and International Relations, Russian Academy of Sciences. Moscow, Russia

BOGDANOV, Konstantin Vadimovich, Cand. Sci. (Engineering), Senior Research Fellow, Center for International Security, Primakov National Research Institute of World Economy and International Relations, Russian Academy of Sciences, email: cbogdanov@imemo.ru; **YEVTODYEVA, Marianna Georgievna**, Cand. Sci. (Polit. Sci.), Senior Research Fellow, Center for International Security, Primakov National Research Institute of World Economy and International Relations, Russian Academy of Sciences, email: marianna133@mail.ru

Bogdanov, K.V., & Yevtodyeva, M.G. (2022). Lethal autonomous weapon systems: landscape and perspectives. *Polis. Political Studies*, 6, 67-80. (In Russ.) https://doi.org/10.17976/jpps/2022.06.06

Acknowledgements. The article was prepared within the project "Post-crisis world order: challenges and technologies, competition and cooperation" supported by the grant from the Ministry of Science and Higher Education of the Russian Federation's program for research projects in priority areas of scientific and technological development (Agreement № 075-15-2020-783).

Received: 22.04.2022. Accepted: 01.09.2022

Abstract. The active development of military artificial intelligence technologies brings the moment of the mass deployment of lethal autonomous weapons systems (LAWS) closer. There is no internationally recognized definition of LAWS, and this has led to terminological disputes behind which are hidden both military-technical and ethical-philosophical problems. The task of keeping control over autonomous weapon systems is complex, and must be addressed in interdisciplinary ways, simultaneously in several dimensions, considering operational issues, moral and ethical dilemmas, and legal collisions. All three types of considerations focus on the problems of meaningful human control and the attribution of responsibility for the application of LAWS. The problem of fictitiousness of nominal human control over complex technical systems using opaque functioning algorithms is analysed. Developers of the new weapons are resisting the creation of adapted regulations, claiming that existing international humanitarian law is enough to regulate the new field, making it difficult to impose additional restrictions on LAWS. Existing agreements to control the proliferation of technology are in fact only partially capable of serving as a tool for controlling LAWS. It was noted that in the case of LAWS, a powerful impetus is possible for the development of a special direction in arms control related to the legal aspects of attribution of responsibility for the misuse of weapons systems, which was not previously observed in the corpus of the international legal framework for arms control. In the context of the collapse of the post-Cold War arms control system, only certain concrete measures related to commitments to prevent the most destabilizing manifestations of LAWS (for example, equipping them with nuclear weapons) can be proposed in the short- and medium-term.

Keywords: arms control, artificial intelligence, human control, international humanitarian law, lethal autonomous weapon systems, Martens Clause, unmanned vehicles.

References

Autonomy, artificial intelligence and robotics: technical aspects of human control. (2019). Geneve: ICRC.

Benjamin, M. (2013). Drone warfare: killing by remote control. London: Verso.

Bode, I., & Watts, T. (2021). Meaning-less human control. Lessons from air defence systems on meaningful human control for the debate on AWS. *Odense: Center for War Studies, South Denmark University.*

Boulanin, V., & Veerbruggen, M. (2017). Mapping the development of autonomy in weapon systems. Stockholm: SIPRI.

Davison, N. (2017). A legal perspective: autonomous weapon systems under international humanitarian law. Perspectives on lethal autonomous weapon systems. UNODA Occasional Paper No. *30 (pp. 5-18)*.

De Sio, F.S., & van den Hoven, J. (2018). Meaningful human control over autonomous systems: a philosophical account. *Frontiers in Robotics and AI, 5, 1-15*. https://doi.org/10.3389/frobt.2018.00015

Fisher, J., & Ravizza, M. (1999). Responsibility and control. Cambridge, UK: Cambridge University Press.

Ivanov, D.V., Korzhenyak, A.M., & Lapikhina, E.S. (2021). Lethal autonomous weapons systems and international law. *Moscow Journal of International Law,* 3, 6-19. https://doi.org/10.24833/0869-0049-2021-3-6-19

Kania, E. (2017). Battlefield singularity: artificial intelligence, military revolution, and China's future military power. Washington, D.C.: Center for a New American Security.

Lee, P. (2018). The distance paradox: reaper, the human dimension of remote warfare, and future challenges for the RAF. *Air Power Review,* 21(3), 106-131.

Marra, W.C., & McNeil, S.K. (2012). Understanding the loop: regulating the next generation of war machines. *Harvard Journal of Law & Public Policy,* 36(3), 1139-1185. https://dx.doi.org/10.2139/ssrn.2043131

McGregor, L., Murray, D., & Ng, V. (2019). International human rights law as a framework for algorithmic accountability. *International and Comparative Law Quarterly,* 68, 2, 309-343. https://doi.org/10.1017/S0020589319000046

Osinga, F.P.B. (2007). Science, strategy and war: the strategic theory of John Boyd. New York, NY: Routledge.

The Military Balance 2021. (2021). London: Routledge.

Anthony, I., & Holland, Ch. (2015). The governance of autonomous weapons. In *SIPRI Yearbook 2014*. (Russ. ed.: Anthony, I., & Holland, Ch. Upravlenie avtonomnym oruzhiem. In *Ezhegodnik SIPRI 2014: vooruzhenia, razoruzhenie i mezhdunarodnaya bezopasnost* (pp. 456-465). Moscow: IMEMO RAS).

Anthony, I., Bromley, M., Boulanin, V., & Grip, L. (2015). Humanitarian arms control. In *SIPRI Yearbook 2015*. (Russ. ed.: Anthony, I., Bromley, M., Boulanin, V., & Grip, L. Gumanitarnyi kontrol' nad vooruzheniyami. In *Ezhegodnik SIPRI 2015: vooruzhenia, razoruzhenie i mezhdunarodnaya bezopasnost* (pp. 630-641). Moscow: IMEMO RAS).

Arbatov, A.G. (Ed.). (2020). Kontrol nad vooruzheniyami v novykh voenno-politicheskikh i tekhnologicheskikh usloviyakh [Arms control in the new military-political and technological conditions]. Moscow: IMEMO RAS.

Golubenko, E.A. (2017). Autonomous weapon systems and international humanitarian law. *Vestnik Akademii voennykh nauk,* 2, 101-106. (In Russ.)

Makarenko, S.I. (2016). Military robots – the current state and prospects of improvement. *Systems of Control, Communication and Security*, 2, 73-132. (In Russ.)

Peldán Carlsson, M., & Boulanin, V. (2020). The group of governmental experts on lethal autonomous weapon systems. In *SIPRI Yearbook 2020*. (Russ. ed.: Peldán Carlsson, M., & Boulanin, V. Gruppa nepravitelstvennykh ekspertov po smertonosnym avtonomnym sistemam oruzhia. In *Ezhegodnik SIPRI 2020: vooruzhenia, razoruzhenie i mezhdunarodnaya bezopasnost* (pp. 582-595). Moscow: IMEMO RAS).

Yevtodyeva, M.G., & Oznobishchev, S.K. (2019). Military unmanned systems: current state, prospects for limitation and control. *Moscow University Bulletin of World Politics,* 3, 160-197. (In Russ.)

Литература на русском языке

Голубенко Е.А. 2017. Автономные системы вооружения и международное гуманитарное право. *Вестник Академии военных наук*. № 2. С. 101-106.

Евтодьева М.Г., Ознобищев С.К. 2019. Боевые беспилотные системы: современный этап развития, перспективы ограничения и контроля. *Вестник Московского университета. Серия 25: Международные отношения и мировая политика*. № 3. С. 160-197.

Контроль над вооружениями в новых военно-политических и технологических условиях. 2020. Под ред. А.Г. Арбатова. М.: ИМЭМО РАН.

Макаренко С.И. 2016. Робототехнические комплексы военного назначения – современное состояние и перспективы развития. *Системы управления, связи и безопасности*. № 2. С. 73-132.

Пелдан Карлсон М., Буланен В. 2021. Группа правительственных экспертов по смертоносным автономным системам оружия. *Ежегодник СИПРИ 2020: вооружения, разоружение и международная безопасность*. М.: ИМЭМО РАН. С. 582-595.

Энтони И., Бромли М., Буланен В., Грип Л. 2016. Гуманитарный контроль над вооружениями. *Ежегодник СИПРИ 2015: вооружения, разоружение и международная безопасность*. М.: ИМЭМО РАН. С. 630-641.

Энтони И., Холланд К. 2015. Управление автономным оружием. *Ежегодник СИПРИ 2014: вооружения, разоружение и международная безопасность*. М.: ИМЭМО РАН. С. 456-465.