

УДК 34.096

ГРНТИ 10.87.68

**ТРИ ЗАКОНА РОБОТОТЕХНИКИ АЙЗЕКА АЗИМОВА: К ВОПРОСУ
ГУМАННОСТИ ПРИМЕНЕНИЯ СМЕРТОНОСНЫХ АВТОНОМНЫХ
СИСТЕМ ВООРУЖЕНИЯ НА ВОЙНЕ**

Шибаета Кристина Валерьевна

студент магистратуры первого курса Юридической школы

Дальневосточного Федерального университета,

г. Владивосток

e-mail: shibaeva2015@mail.ru

Холова Лейла Насимжановна

студент магистратуры первого курса Юридической школы

Дальневосточного Федерального университета,

г. Владивосток

e-mail: gmu-gruppa@mail.ru

**THREE LAWS OF ISAAK ASIMOV'S ROBOTICS: ON THE QUESTION OF
HUMANITIES OF THE APPLICATION OF LETHAL AUTONOMOUS
WEAPONS (LAWs) IN WAR**

Kristina V. Shibaeva

master 1 st course student of Law School

Far Eastern Federal University

Vladivostok

Leila N. Kholova

master 1 st course student of Law School

Far Eastern Federal University

Vladivostok

АННОТАЦИЯ

В данной статье поднимается вопрос гуманности в контексте разработки и применения искусственного интеллекта на войне. Отражены позиции «за» и «против» представителей научного сообщества, продемонстрированы возможные исходы событий. Рассмотрена этическая сторона такого явления как применение смертельных автономных систем вооружения.

ABSTRACT

In this article the issue of humanity is raised in the context of the development and application of artificial intelligence in the war. The positions “for” and “against” the representatives of the scientific community are reflected; the possible outcomes of events are demonstrated. Also in this article the ethical aspect of such a phenomenon as the use of lethal autonomous weapons systems is considered

Ключевые слова: искусственный интеллект, робот на войне, законы робототехники, автономное вооружение, смертельная автономная система вооружения, этика, право, гуманность, сверхразум.

Key words: artificial intelligence, robot in war, laws of robotics, autonomous weapons, lethal autonomous weapons, ethics, law, humanity, supermind.

Время неумолимо движется вперёд и человек, подгоняемый научно-техническим прогрессом, старается не отставать от него, идя в одну ногу со временем. Эпоха, когда в сознании людей крепко сидела идея о том, что машины ни на что не способны и вольны мыслить лишь цифрами, осталась далеко в прошлом. На сегодняшний день мы наблюдаем, что все сферы человеческой жизни компьютеризированы, а наличие разнообразных систем с элементами искусственного интеллекта уже не является чем-то за гранью реального и не вызывает удивления общественности.

Искусственный интеллект постепенно внедряется повсеместно, и в первую очередь в области, отвечающие за благосостояние и процветание государств: политическая, социальная и экономическая сферы, именно поэтому применение, так называемых, роботов в военных интересах - это, к сожалению или к счастью, далеко уже не новинка. В мире военного хайтека сейчас всё сложнее разглядеть

такие вещи как лояльность, преданность, самопожертвование и множество других проявлений человечности, именно поэтому напрашивается вопрос, а место ли вообще людям на полях военных действий?

Ни для кого не является секретом, что военными и разведывательными ведомствами по всему миру уже широко и довольно успешно внедряются боевые роботы - сапёры для нахождения и обезвреживания бомб и мин, беспилотные летательные аппараты, предназначенные как для разведки, так и для ведения боевых действий. Однако надо понимать, что подобные системы являются телеуправляемыми, то есть полностью контролируруемыми человеком. Роботы в данном случае выступают лишь орудием, в то время как принятие решений остаётся за человеком, ими управляющим.

Но что же будет, когда появится полностью автономная боевая система, которая будет вольна убивать по своей собственной инициативе?

Что ждёт человечество, когда машины с искусственным интеллектом, способные к самообучению и обмену информацией между собой, будут не столько иметь право убивать, сколько обязаны это делать?

Будет ли это гуманно - прибегать к помощи полностью роботизированной армии?

Вопросов множество, ответов же значительно меньше. Неудивительно, что в сознании большинства людей при разговоре о подобных армиях, ярко всплывает образ Терминатора, а сюжеты банальных голливудских боевиков о войнах людей с цивилизацией роботов внезапно обретают смысл. Это вполне логично и обосновано, а чувствовать страх при подобном развитии событий - нормально. Вопрос доверия человеческой жизни достаточно сложен для однозначного ответа, а вопрос доверия миллионов, миллиардов человеческих жизней в руки машины, которая самостоятельно будет продумывать каждый свой последующий шаг, в разы сложнее.

В момент, когда в мире появится новый, хоть и искусственный, но сверхразумный вид, неважно, что придёт ему в голову: попытка сделать всех людей счастливыми или же идея уничтожения человечества, сопротивляться у

людей просто на просто не будет возможности. В этом случае фатализм не только необходим, но и крайне полезен, ведь принимая всевозможные варианты развития событий, мы значительно снижаем уровень страха.

Создание роботизированной армии - это тот момент, когда люди должны играть на опережение, а не принимать решения *post factum*. Уже сейчас человечеству необходимо сделать моральный выбор: что считать более правильным, законным и гуманным - ***убийство человека человеком или убийство человека машиной?***

Создание и применение роботов-убийц регулярно обсуждается с 2014 года, когда проблематика смертоносных автономных систем вооружения¹ (далее - САС) прочно закрепилась в качестве новой темы в рамках «Конвенции о «негуманном» оружии» (далее - КНО) и была создана Группа правительственных экспертов открытого состава (далее - ГПЭ) [1]. На сегодняшний день проведено уже три сессии ГПЭ по САС (первая сессия - 13-17 ноября 2017 г., вторая сессия - апрель 2018 г., третья сессия – 27 августа 2018 г.). В ходе данных конференций тематика САС неоднократно была рассмотрена с множества позиций: военно-технической, правовой, морально-этической, политической. Однако вопрос применения вооружения с высокой степенью автономности всё ещё остаётся открытым.

Ведутся многочисленные дискуссии среди ведущих представителей учёного мира. Стивен Хокинг, один из самых уважаемых британских учёных, с 2014 года придерживался мнения, что *искусственный интеллект - «это громадная угроза для человечества»* [10].

Против применения роботов-убийц неоднократно выступал и основатель космической компании SpaceX, человек, который имеет доступ к последним из всевозможных разработок в сфере искусственного интеллекта, Илон Маск. *«Искусственный интеллект - фундаментальная угроза существованию человечества»* [11]. Он призывает сверхдержавы, ведущие разработку

¹ Смертоносные автономные системы вооружения (САС) - оружие, способное поражать цели на земле, в воздухе, на воде и под водой без участия человека.

автономного вооружения, притормозить и удостовериться, что это безопасно. *Людам необходимо озаботиться искусственным интеллектом* - говорит Илон Маск на одной из конференций, посвящённых автономному летальному оружию [6].

В 2017 году группа из 116 специалистов обратилась к Организации Объединённых наций (далее - ООН) с просьбой запретить разработку и использование автономного оружия. *«Решение лишить человека жизни никогда не должно быть делегировано роботу»* говорилось в письме. К числу авторов обращения, помимо Илона Маска, присоединился и основатель компании, занимающейся искусственным интеллектом, DeepMind Мустафа Сулейман. Разработчики обращения, говоря об использовании роботов-убийц, предупреждают о возможности начала третьей революции в области вооружения, сравнимой с такими точками исторического отсчёта, как изобретение пороха и ядерного оружия.

«Если ящик Пандоры будет открыт, закрыть его будет крайне сложно» - вот о чём говорят авторы обращения [8].

Опасения в разрезе использования САС далеко не пустой звук. Выступающие против разработки и использования роботов-убийц ссылаются на Международную Конвенцию ООН об обычных вооружениях. Однако ссылки на Женевскую Конвенцию 1949 г. возможны лишь с большим количеством оговорок, поскольку, насколько нам известно, в международном гуманитарном праве, регулирующем методы нападения на людей во время военных действий, отсутствует ряд специальных положений, которые непосредственно касались бы автономных систем вооружения. Поднимая вопросы гуманности использования роботов-убийц, естественным образом необходимо затрагивать принципы ведения «гуманной войны». По этому поводу в Женевской Конвенции закреплено, что любая военная атака обязана соответствовать трём основным критериям:

- 1) военная необходимость;

2) дифференцирование военных и мирных жителей, которые не принимают участия в конфликте;

3) сохранение пропорции между потенциальными целями войны и ущербом, ею нанесённым [7].

Особое место должно быть отведено и Оговорке Мартенса, запрещающей применение оружия, противоречащее «принципам гуманности, человечности и здравого смысла» [13]. Тут и возникает один из многочисленных вопросов, а способна ли система искусственного интеллекта в принципе выносить такие субъективные оценки?

Если роботизированная система выходит из строя по непонятным на то причинам, а принимая во внимание тот факт, что робот-убийца - это в первую очередь машина и машинам свойственно ошибаться, таков закон развития техники и технологий, убивает несметное количество человек, то кого винить? Как и кого наказывать? Сделал ли робот это по собственной воле (*mens rea*) или же всему виной технический сбой в программе?

Как видите, чем глубже мы уходим в изучение проблемы, которая уже сегодня имеет место быть, тем больше вопросов у нас возникает.

И, к большому сожалению, не на все вопросы находятся исчерпывающиеся ответы ни у представителей учёного мира, ни у представителей сверхдержав, выступающих за разработку автономного вооружения.

Прошлый год был назван годом прорыва в области искусственного интеллекта. В январе 2017 года была проведена Асиломарская конференция по безопасности работы с искусственным интеллектом, организатором которой выступил Институт Будущего Жизни (FLI). В ходе конференции были сформулированы 23 принципа работы с искусственным интеллектом [3]. В разрезе разработки и применения автономного вооружения особого внимания заслуживают два принципа из общего количества.

Во-первых, одно из опасений всех представителей учёного мира, о чём неоднократно и упоминает Илон Маск в своих интервью - это гонка вооружений, ибо лидер по созданию искусственного интеллекта в мгновение ока станет

властелином мира. На Асиломарской конференции единогласно было решено, что *«стоит избегать гонки вооружений в области автономного летального оружия»* [3].

Казалось бы, как подойти к подобному опасению с этической точки зрения? Однако всё до банальности просто, на данный момент времени можно наблюдать цивилизационное неравенство между государствами, приводящее к ситуации, когда некоторые страны просто не могут позволить себе создание автономных видов вооружения. Полбеда, когда роботы выходят против роботов, но что произойдёт, когда робот выйдет против живого человека? Думаю, исход ясен.

Вторым же принципом, который должен выступать неким стоп-краном при разработке роботов-убийц - это принцип всеобщего блага, который гласит *«сверхразум должен быть разработан исключительно в целях, соответствующих большинству этических идеалов и для пользы всего человечества, а не одного государства или организации»* [3].

Допустим, сам по себе искусственный интеллект вряд ли обладает каким-либо злым умыслом и задаётся целью уничтожить человечество, но мы не можем сказать того же самого о людях, в руки которым достанутся роботы-убийцы. Поднимается вопрос о демократизации технологий искусственного интеллекта, которые должны быть широкодоступны. *«Нельзя допустить концентрацию искусственного интеллекта в одних руках, слишком высок соблазн захвата контроля, а подобная ситуация будет, как вы уже поняли, исключительно опасной»* - высказывается по данному вопросу Илон Маск [5].

Принимая во внимание выводы, к которым пришли участники последних конференций по безопасности работы с искусственным интеллектом, можно назвать **три основные причины**, почему разработка и использование роботов-убийц - это негуманно (см. рис. 1).

Первая группа, а именно «проблема значимого человеческого контроля» во главу угла ставит риски, связанные с уровнем принятия решений, когда мы полностью исключаем человека из цикла. Мировая общественность (Илон Маск,

Стив Возняк, неправительственные организации «Stop Killer Robots», «Article 36», «International Committee for Robot Arms Control») поднимает вопрос о том, что автономное вооружение с огромной долей вероятности, не будет соответствовать международному гуманитарному праву. Подобные системы, как обговаривалось в рамках конференций, будут нарушать права человека. Использование САС противоречит принципу гуманности и требованиям сознания общественности, поскольку роботам-убийцам, по мнению учёных, несвойственно сочувствие, которое часто является сдерживающим фактором в отношении применения любого оружия.



Рис. 1. - Угрозы создания и применения автономных вооруженных систем с использованием искусственного интеллекта

Ко второй группе причин, связанных с нарушением стратегической стабильности, относятся такие риски как гонка вооружений, мировое господство, распространение опасных технологий среди негосударственных объектов. Автономное вооружение с элементами искусственного интеллекта может выступать инструментом стратегического сдерживания, равным счётом, как и всем известное ядерное оружие.

И, наконец, третья причина - сокращение времени на принятие важных стратегических решений, также рассматривается представителями научного

сообщества как потенциальная угроза использования автономного вооружения. В рамках данной проблемы необходимо сказать опять-таки об отсутствии контроля над ситуацией со стороны человека, о нехватке времени для осмысления человеком решений, принятых машиной. И самое главное - отсутствие взаимопонимания между роботом и человеком, поскольку вряд ли машина станет объяснять людям логику собственных действий [2].

Новые, неизведанные технологии, готовые в любой момент времени выйти из-под контроля - это всегда страшно и вполне объяснимым является факт испуга людей, они и должны быть напуганы. Однако, по мнению Рэя Курцвейла, испуг перед разрушительной силой новых технологий - лишь одна из фаз их принятия человечеством. *«Технологии всегда будут палкой о двух концах, и в их принятии человечество проходит три фазы. Первая - это желание решить давние проблемы с помощью инноваций, например, истребить болезни и устранить нищету. Вторая фаза - это тревога из-за разрушительного характера технологий. И третья фаза - это ощущение морального долга, обязывающего продолжать разработки вопреки всему»* - высказывается футуролог [9].

Роботы учатся ходить за считанные часы, быстрее, чем любой биологический вид на планете, так почему бы не использовать возможности искусственного интеллекта во благо?

Одним из самых значительных и весомых аргументов, выступающих в защиту использования роботизированной армии и однозначно оправдывающих гуманность подобного действия, выступает сохранение человеческих жизней. Как ни парадоксально это звучит, но доверие человеческой жизни роботам может способствовать её сохранению.

Войны уносят миллионы, а то и миллиарды человеческих жизней, стирают с лица земли целые этнические группы. Они всегда ассоциируются у людей со страхом, смертью и огромнейшими психологическими нагрузками. А если на минуту представить, что всего этого можно было бы избежать? Если бы на поле боя выходили полностью автономные роботы, людям больше не пришлось бы сталкиваться с жестокостью, рисковать собственными жизнями и проливать

кровь за то, что чаще всего случается не по их желанию, а по веянию людей, стоящих у власти.

Несмотря на то, что на сегодняшний день война всё чаще ведётся с использованием телекоммуникационных систем вооружения и люди зачастую - это безликие точки на экранах мониторов, человек как бы там ни было испытывает психологическую нагрузку. Используя автономные системы вооружения подобного можно было бы избежать вовсе.

Профессор Рональд Аркин из Технологического института Джорджии, известный американский конструктор роботов, который работает над проектами Пентагона, утверждает, что роботы-убийцы могут быть запрограммированы так, чтобы безукоризненно следовать законам войны, и следовать они им будут в разы лучше, чем люди. *«Робот - говорит он - никогда не будет стрелять, защищая свою жизнь или из-за страха. Он будет иметь доступ к информации и данным, которые ни один солдат-человек не смог бы обработать так быстро, что делает менее вероятной ошибку в «пылу битвы». Он никогда не будет намеренно убивать мирных жителей в отместку за смерть товарища. И он сможет следить за солдатами, чтобы они не совершали злодеяния»* [4].

В действительности человек всегда будет проигрывать машине лишь потому, что слишком подвержен веянию эмоций, чувству страха, которые зачастую мешают сконцентрироваться. К тому же, ни для кого не секрет, что функции головного мозга человека ограничены, человек очень медлительное живое существо, и он не может делать несколько дел одновременно с той быстротой, которая просто необходима в рамках военных действий, когда решения должны быть приняты практически молниеносно и с большой долей вероятности.

Принимая во внимание экономическую составляющую, отметим, что роботизированная армия в разы дешевле армии человеческой: они не нуждаются в обучении, дешевле стоят и не возвращаются домой «грузом-200».

Однако, как уже было неоднократно отмечено, разработка и применение роботов-убийц крепко завязаны на вопросах этического содержания. Может ли

железная конструкция, оснащённая элементами искусственного интеллекта, принимать решения, ставя перед собой вопрос, что есть хорошо, а что плохо? Именно этот вопрос послужил фундаментом такой проблемы как «машинная этика».

Давайте вспомним три замечательных закона робототехники писателя-фантаста Айзека Азимова, которые изначально были задуманы так, чтобы наложить этический оттенок на поведение машины. Первый закон категорически запрещает роботу вредить людям, второй - приказывает роботу подчиняться людям до тех пор, пока это не станет противоречить первому закону, и, наконец, третий закон приказывает роботу защищать свою жизнь, не нарушая первых двух.

Большинство экспертов, придерживающиеся мнения о запрете применения роботов-убийц, говорят о неисполнимости подобных законов в реальном мире, то есть машине не свойственно испытывать моральные переживания из-за возможной ошибки. Опровергая подобное заявление, следует обратиться к экспериментам, проведённым двумя учёными из Португалии и Индонезии. Lu'is Moniz Pereira и Ari Saptawijaya. В своей работе «Modelling Morality with Prospective Logic», которая в 2009 году была опубликована в журнале «International Journal of Reasoning-based Intelligent Systems», они описывают методы программирования для компьютеров, дающие возможности машинам совершать поступки, исходя из гипотетических моральных рассуждений, присущих человеку [12].

Учёными был проведён ряд экспериментов, которые, так или иначе, моделировали сложные моральные дилеммы для роботов. «Проблема вагонетки» (the trolley problem), именно такое название получила дилемма, предлагаемая роботам и моделирующая проблемный вопрос, можно ли навредить одному человеку или некоторому количеству людей, чтобы спасти жизнь большинства. Можно ли допустить небольшую погрешность для достижения лучшего эффекта, блага?

Проводя своё исследование, авторы ставили роботов перед нелёгким выбором ровно шесть раз (bystander, footbridge, loop track, man-in-front, drop man, collapse bridge). Проходя каждое испытание и стоя перед выбором спасти пятерых, но убить одного, роботы колебались. Интересным является тот факт, что по мере проведения эксперимента, было выявлено, что машина всегда руководствовалась принципом, так называемого, «двойного эффекта»: **«причинение вреда другому лицу допустимо в том случае, если это будет являться поступком, приводящим к большему благу»** [12].

Подобный эксперимент позволяет задуматься о том, что машины используют человека, как образ и подобие, а мораль и этика больше не являются лишь людской привилегией.

Объединив всё вышесказанное, хочется ещё раз отметить, что страх людей перед неизведанным обоснован. Однако в их сознании ещё не до конца сложилось понимание того, что же делать с искусственным интеллектом, когда он станет подобен человеческому, а то и превзойдёт его в несколько десятков раз. Пока люди не увидят, как роботы идут по улицам и убивают людей, они, к огромному сожалению, не будут знать, как же на это реагировать. Именно поэтому, если уж человек задумал «игру в Бога», то ему стоит приостановиться и тщательно всё обдумать. Решения, от которых зависит жизнь человечества, не должны быть приняты де-факто, после того как случится множество непоправимых страшных вещей, они должны быть тщательно обдуманны и взвешены. Искусственный интеллект же не должен восприниматься людьми только с позиции силы, способной уничтожать всё на своём пути, на него необходимо смотреть и с позиции верного соратника, способного помочь.

Только лишь принятие человечеством того факта, что оно может либо приобрести что-то совершенно новое, либо потерять всё, что было накоплено веками, подтолкнёт его, на наш взгляд, к великому прорыву.

Список литературы

1. Конвенция о «негуманном» оружии [Электронный ресурс] // Министерство иностранных дел Российской Федерации. URL: http://www.mid.ru/obychnye-vooruzeniya/-/asset_publisher/MIJdOT56NKIk/content/id/1130752 (дата обращения 09.12.2018);
2. Козюлин В. Три группы угроз смертоносных автономных систем [Электронный ресурс] // аналитическая статья: Российский совет по международным делам, 2018. URL: <http://russiancouncil.ru/analytics-and-comments/analytics/tri-gruppy-ugroz-smertonosnykh-avtonomnykh-sistem/> (дата обращения 11.12.2018);
3. Леонов В. Двадцать три принципа Асиломара [Электронный ресурс] // Современное машиностроение: научно-технический журнал, 2017. URL: <https://www.sovmash.com/node/348> (дата обращения 12.12.2018);
4. Если робот убьет человека, кто будет виновен? [Электронный ресурс]. URL: <https://habr.com/company/robhunter/blog/381437/> (дата обращения 08.12.2018);
5. Илон Маск об искусственном интеллекте [Электронный ресурс] // интервью от 15.09.2016 на русском языке. URL: <https://www.youtube.com/watch?v=GdGyYI1KP5I&t=1014s> (дата обращения 09.12.2018);
6. Илон Маск потребовал от ООН запретить роботов-убийц [Электронный ресурс]. URL: <https://www.interfax.ru/world/575613> (дата обращения 10.12.2018);
7. Когда робот идет на войну [Электронный ресурс]. URL: <https://inosmi.ru/world/20150621/228697791.html> (дата обращения 08.12.2018);
8. Роботы-убийцы: скоро во всех армиях или под запретом ООН? [Электронный ресурс]. URL: <https://ru.delfi.lt/abroad/global/roboty-ubijcy-skoro-vo-vseh-armiyah-ili-pod-zapretom-oon.d?id=78914141> (дата обращения 11.12.2018);

9. Рэй Курцвейл: «Развитие технологий избавит человечество от страданий» [Электронный ресурс]. URL: https://hightech.fm/2017/11/14/ray_kurzweil (дата обращения 10.12.2018);
10. Хокинг: искусственный интеллект - угроза человечеству [Электронный ресурс]. URL: https://www.bbc.com/russian/science/2014/12/141202_hawking_ai_danger (дата обращения 10.12.2018);
11. Человечество в опасности: Илон Маск призвал регулировать искусственный интеллект [Электронный ресурс]. URL: <http://www.forbes.ru/tehnologii/347945-chelovechestvo-v-opasnosti-ilon-mask-prizval-regulirovat-iskusstvennyu-intellekt> (дата обращения 09.12.2018);
12. Pereira L. M., Saptawijaya A. Modelling Morality with Prospective Logic [Электронный ресурс] // International Journal of Reasoning-based Intelligent Systems, 2009. URL: <https://pdfs.semanticscholar.org/2ecc/67e7929cf613976d33ff6479589b7b62dfce.pdf> (дата обращения 11.12.2018);
13. Sparrow R. Ethics as a source of law: The Martens clause and autonomous weapons [Электронный ресурс] // Autonomous Weapons Mini-series, 2017. URL: <http://blogs.icrc.org/law-and-policy/2017/11/14/ethics-source-law-martens-clause-autonomous-weapons/> (дата обращения 11.12.2018).