

Этика автономных машин: деонтология и военные роботы

Кирилл Мартынов

Доцент, Школа философии, факультет гуманитарных наук,
Национальный исследовательский университет «Высшая
школа экономики» (НИУ ВШЭ). Адрес: 105066, Москва,
ул. Старая Басманная, 21/4. E-mail: kmartynov@hse.ru.

Ключевые слова: деонтология; летальные автономные
вооружения; этика; Иммануил Кант; полицейская
операция; война.

Статья посвящена этической дискуссии вокруг автономного летального вооружения. Появление военных роботов, способных самостоятельно действовать на поле боя, рассматривается как неизбежный этап развития современной войны, поскольку они способны дать критическое преимущество одной из сторон конфликта. Несмотря на общественные движения, призывающие запретить «роботов-убийц», существуют этические аргументы в поддержку развития таких технологий. В частности, утилитаристская традиция соглашается с допустимостью военных роботов, при условии что «нечеловеческая война» будет становиться более гуманной и вести к меньшему числу жертв. Деонтологический анализ, в свою очередь, предполагает, что этика невозможна без этического субъекта.

Моральная философия Иммануила Канта коррелирует с интуицией о том, существует ли значимая разница между ситуацией, в которой человек принимает решение об убий-

стве другого человека, и ситуацией, в которой такое решение принимает машина. Роботы, как и животные, становятся пограничными агентами на границах «моральных сообществ». На примере дискуссии о правах животных мы видим, как работает этика Канта применительно к нечеловеческим агентам. Ключевой проблемой использования автономных вооружений становится трансформация войны и непредсказуемые риски, связанные со стиранием различий между войной и полицейской операцией. Гипотеза автора состоит в том, что роботам не нужно будет убивать, чтобы побеждать противника. Если на войне никто не гибнет, то не существует причин не распространять ее практики на noncombatantov или стремиться к миру. Анализ, представленный в утилитаризме, игнорирует возможность таких последствий. Главной проблемой автономного летального вооружения становится автономность, а не способность причинять смерть.

Нанимать людей, для того чтобы они убивали или были убиты, означает пользоваться ими как простыми машинами.

Иммануил Кант. К вечному миру

СТАНДАРТНОЙ темой дискуссии в последние десять лет стала автоматизация труда как прикладное применение искусственного интеллекта. Если интеллектуальный труд человека можно рассматривать как последовательность действий по обработке информации, то существует ли хотя бы одна профессия, которую нельзя делегировать программе? Практические импликации этого вопроса обсуждаются в первую очередь в рамках социальной теории: основным коррелятом автоматизации труда становится проблема роста имущественного неравенства, экзистенциальный кризис «общества конца работы», а также понятие базового дохода¹. Акценты меняются, когда в контексте автоматизации труда мы обсуждаем профессию солдата. Тогда проблема автономных машин, действующих в мире наряду с человеком, становится *par excellence* предметом этики².

Какие этические вопросы возникают с появлением военных роботов? И как отвечают на них деонтология и утилитаризм? Тезис данного текста заключается в том, что убедительного разрешения этической дилеммы «роботов-убийц» в рамках утилитаризма не существует, хотя эта традиция хорошо сочетается с нашими первичными моральными интуициями. Кантианская этика будет склонять нас к полному запрету использования роботов на войне, несмотря на очевидную пользу и технологическую неизбежность этого решения. Основным аргументом для запрета военных роботов способно стать стирание грани между войной и полицейской операцией, что, в свою очередь, может обернуться разруше-

1. It's Basic Income. Global Debate / A. Downes, S. Lansley (eds). Bristol: Policy Press, 2018.
2. Strawser B. J. Moral Predators: The Duty to Employ Uninhabited Aerial Vehicles // Journal of Military Ethics. 2010. Vol. 9. № 4. P. 342–368.

нием «правил войны» и непредсказуемыми рисками для социального порядка и политических систем.

Под автономными машинами здесь мы понимаем алгоритмы, способные анализировать данные, поступающие в реальном времени, и принимать на их основе решения, имеющие моральные и социальные последствия³. Такие машины в целом имеют мало общего с антропоморфными роботами, описанными в массовой культуре. Машины, пригодные для принятия решений, не обязательно обладают конкретной физической реализацией. Разумеется, их применение не ограничивается военной сферой: аналогичные машины сегодня используются в финансах (высокочастотные торги), медицине (автономные диагностические системы), лингвистике (машинный перевод). Наиболее узнаваемый пример автономной машины — беспилотный автомобиль. Масштабное исследование, проведенное в последние годы Массачусетским технологическим институтом (MIT), показало, что робот-автомобиль, принимающий решения в условиях реального мира, неизбежно сталкивается с моральным выбором, а значит, может быть описан в качестве «субъекта этики»⁴. Элементарным аналогом этического выбора для него становится дилемма вагонетки⁵.

Наиболее острые споры в контексте этики робототехники вызывает именно применение автономных машин на войне⁶. В этой дискуссии используется термин «автономные летальные вооружения» (*lethal autonomous weapons, LAW*), обозначающий машину, которая, будучи активированной, принимает решение об огне на поражение без прямого контроля со стороны человека. Внедрение подобных систем обещает военным многочисленные преимущества: они будут быстрее и точнее человека, так что армия, располагающая таким вооружением, может получить решающее преимущество над противником⁷. Как будет показано ниже,

3. *Minsky M.* Steps Towards Artificial Intelligence // Proceedings of the IEEE. 1961. Vol. 49. P. 8–30.
4. *Awad E. et al.* The Moral Machine Experiment // Nature. 2018. Vol. 563. P. 59–64; *Charisi V. et al.* Towards Moral Autonomous Systems // ArXiv.org 14.03.2017. URL: <https://arxiv.org/pdf/1703.04741.pdf>.
5. *Foot P.* The Problem of Abortion and the Doctrine of Double Effect // Oxford Review. 1967. № 5. P. 5–15.
6. *Purves D. et al.* Autonomous Machines, Moral Judgment, and Acting for the Right Reasons // Ethical Theory and Moral Practice. 2015. Vol. 18. № 4. P. 851–872.
7. *Singer P. W.* Wired for War. The Robotics Revolution and Conflict in the Twenty-First Century. N.Y.: Penguin Books, 2010.

этические оценки соответствующих практик могут быть диаметрально противоположными: от утверждения о полной аморальности военных роботов до тезиса о том, что их использование — это единственный гуманный способ ведения современной войны⁸. Предположу, что в этих дискуссиях из виду упускается ключевая особенность автономных боевых роботов: чтобы побеждать в войне, роботам, в отличие от людей, не обязательно убивать.

Отметим, что вопрос об автономных машинах, как они понимаются в данном тексте, не имеет связи с дискуссией о возможности сильного искусственного интеллекта, представленной у Джона Серла⁹. Вне зависимости от способности машин пройти тест Тьюринга и ее философских следствий, мы уже сегодня сталкиваемся с тем, что поведение неантропоморфных и способных к обучению систем слишком сложное, чтобы объяснять его «замыслом программистов». Субъектность алгоритмов обсуждалась еще в контексте победы суперкомпьютера *Deep Blue* над Гарри Каспаровым в шахматном матче 1998 года¹⁰. Очевидно, что ни один из разработчиков суперкомпьютера не смог бы предложить выигрышной стратегии игры против чемпиона мира — сильнейшего из игроков-людей. В отличие от реального мира, шахматы представляют собой игру с полной информацией и конечным числом возможных ситуаций. Это означает, что машину, способную действовать наравне с людьми в реальном мире, с прагматической точки зрения необходимо воспринимать в качестве морального субъекта. Наиболее последовательный анализ этой проблемы представлен в начавшейся несколько лет назад дискуссии о правовом статусе роботов. Может ли машина быть субъектом права? И если нет, то почему?¹¹ При этом, по всей видимости, существует разрыв между прагматическим описанием машин в качестве членов правового и морального сообщества и интерпретации ис-

8. Muller V. Autonomous Killer Robots Are Probably Good News // Drones and Responsibility: Legal, Philosophical and Socio-Technical Perspectives on the Use of Remotely Controlled Weapons / E. Di Nucci, F. S. de Sio (eds). Basingstoke: Taylor & Francis, 2016.
9. Searle J. Minds, Brains, and Programs // Behavioral and Brain Sciences. 1980. Vol. 3. № 3. P. 417–457.
10. Hamilton S. Impersonations: Troubling Person in Law and Culture. Toronto: University of Toronto, 2009.
11. Chopra S., White L. A Legal Theory for Autonomous Artificial Agents. Ann Arbor: University of Michigan Press, 2011.

кусственного интеллекта в качестве реального морального субъекта. Этот разрыв между тем, как мы должны относиться к машинам, и тем, чем они на самом деле являются, особенно важен для деонтологии.

Наше определение подразумевает, что действие автономных машин предписывается не столько программой, сколько данными, полученными в реальных «полевых условиях», что позволяет алгоритму гибко адаптироваться к различным ситуативным контекстам. Такой подход к машинному обучению стал возможен в последние десятилетия вместе с удешевлением вычислений и резким ростом доступных для анализа данных. Адаптивная стратегия машины, в свою очередь, влияет на дальнейший ход событий (состояние финансового рынка, тактика терапии в медицине, формулировка перевода текста). В случае с военным роботом, принимающим решение о поражении цели, поведение машины непосредственно определяет жизнь или смерть людей. Соответственно, этические вопросы, с которыми мы сталкиваемся, звучат так:

1. Может ли машина убивать, при условии что аналогичное действие совершил бы профессиональный военный?
2. Можно ли делегировать машине право на убийство без прямой санкции со стороны человека, при условии что этого требуют соображения безопасности?
3. Может ли убийство совершаться субъектом, который не несет моральной ответственности за него?

Классический утилитаризм, представленный в работах Джереми Бентама и Джона Стюарта Милля¹², на первый взгляд, дает удовлетворительные ответы на эти вопросы. Взяв за аксиому исчисление страданий, мы получаем тривиальное решение вопроса о применении автономных машин на войне. Так, робот может убивать, если это необходимо и компетентный человек совершил бы такой же выбор. Убийство человека машиной или другим человеком является эквивалентным с точки зрения утилитаристской морали, если общая сумма страданий оказывается неизменной. Именно в этом состоит здравый смысл, представленный в эксперименте *MIT*, в котором людям предлагали решать за беспилотный автомобиль, чью жизнь следует спасать, а чьей — пожертвовать. Робот может считаться моральным, если он принимает те решения, которые большинство (компетентных) лю-

12. Scarre G. Utilitarianism. L.: Routledge, 1996.

дей признали бы моральными. Утилитаризм в этике выглядит естественным союзником вычислительного разума, ведь обе системы принятия решений построены на калькуляции — последствий в первом случае и приоритетов во втором. Соответственно, ответ на второй вопрос для утилитаризма также является тривиальным: делегировать право на убийство машине можно во всех случаях, когда отказ от этого шага ведет к увеличению числа страданий. Более того, третий вопрос вообще не имеет для последователей Бентама конкретного смысла: статус морального субъекта неисчислимы, следовательно, его можно не учитывать. Итогом этих размышлений становится тезис: автономные машины на войне возможны, если их использование позволяет сократить число жертв.

Позиция, которую можно назвать «продвинутым утилитаризмом», пойдет дальше: согласно ей, использование автономных машин для ведения войны является единственным разумным выбором. Аргументация здесь начинается с довольно очевидного соображения: если война избегнет участия человека, то в идеальных условиях она будет вестись без человеческих жертв, что будет гуманно и, следовательно, оправданно в моральном отношении. Эта логика действует даже в том случае, если роботы заменяют человека лишь частично и позволяют снизить число жертв до некоторой степени — по сравнению с обычной войной. Рассуждение верно и в том случае, если автономные машины ведут войну против людей, то есть стороны конфликта находятся на качественно разном уровне технологического развития, что регулярно случается в современном мире. Если военные роботы атакуют террористов и это ведет к уменьшению жертв среди людей по сравнению с традиционной военной операцией или невмешательством, значит, использование роботов было морально приемлемым. Как и прежде, в общем случае утилитаристы будут утверждать, что любое использование автономных машин, которое ведет к снижению количества страданий, будет оправданным. Дегуманизация войны в буквальном смысле слова, то есть устранение человека из числа комбатантов, в такой перспективе рассматривается как наилучший из возможных сценариев. Есть некоторая ирония в том, что в этом ключе разыгрывается сценарий кантовского трактата «К вечному миру»: роботы-убийцы становятся все более гуманными, а попутно отменяются войны как таковые. Никто не может сопротивляться машинам, так что даже если конфликты и случаются, то они ведутся без участия людей. Путь к пацифиз-

му предполагает в данном контексте не столько отказ от профессиональных армий, делающих из людей машины, сколько отказ от участия людей в войне.

Подход деонтологии оказывается более сложным. Основная проблема применения этики Канта к вопросу об искусственном интеллекте связана с тем, что для философа источником этического суждения является автономный рациональный субъект. Свойственная Просвещению вера в разум и универсальность человеческой природы позволяла Канту отождествлять в человеке разумность и способность к моральному суждению («Основы метафизики нравственности»¹³). Следовательно, три вопроса, представленные выше, становятся для деонтологии трудными этическими дилеммами. В трактате «К вечному миру» Кант критиковал профессиональные армии за то, что они превращают людей в «простые машины или орудия» для убийства¹⁴. Однако случай роботов-убийц переворачивает эту формулу: что если убийство совершается существом, не имеющим человеческих моральных интуиций? Наиболее известная в культуре попытка создания этических правил для роботов связана с «законами робототехники» Айзека Азимова, и она строится по аналогии с категорическим императивом¹⁵. Азимов предполагал, что у роботов могут быть правила, безусловно подлежащие исполнению. «Законы робототехники» непригодны для решения дилеммы вагонетки и, таким образом, не слишком подходят для большинства реальных этических кейсов. Но еще большие проблемы возникают при попытке занять антиутилитаристскую позицию этики в отсутствие морального субъекта. Любопытна в этом смысле статья Феликса Линднера и Мартина Бентцена, которые попытались формализовать в контексте этики робототехники вторую, наиболее безличную формулировку категорического императива¹⁶.

Вопрос о возможном моральном статусе роботов для деонтологии тесно связан с дискуссией о моральном статусе существ, не относящихся к нашему биологическому виду в целом. Наи-

13. *Кант И.* Основы метафизики нравственности // Собр. соч.: В 6 т. / Под общ. ред. В. Ф. Асмуса и др. М.: Мысль, 1965. Т. 4. Ч. 1. С. 219–310.

14. *Он же.* К вечному миру // Собр. соч.: В 6 т. М.: Мысль, 1966. Т. 6. С. 261.

15. *Moor J.* Four Kinds of Ethical Robots // *Philosophy Now*. 2009. Vol. 72. P. 12–14.

16. *Lindner F., Bentzen M.* A Formalization of Kant's Second Formulation of the Categorical Imperative // *ArXiv.org*. 09.01.2018. URL: <https://arxiv.org/abs/1801.03160>.

более известное расширение этики за пределы сообщества людей связано с утилитаристской традицией и представлено в классической работе «Освобождение животных» Питера Сингера¹⁷. Альтернативой этого подхода стала недавняя работа *Fellow Creatures: Our Obligations to Other Animals* философа Кристин Корсгард¹⁸. В своем деонтологическом анализе она исходит из идеи о том, что не существует способа нейтрального измерения блага в терминах баланса удовольствий и страданий, поскольку благо всегда бывает благом для кого-то. Следовательно, начинать дискуссию о включении нечеловеческих существ в пространство этического следует с вопроса о границах наших моральных обязательств. Корсгард утверждает, что, хотя другие биологические виды не могут быть носителями морали в кантовском смысле слова, поскольку не являются разумными, у нас тем не менее нет никаких оснований исключать их из числа субъектов, к которым применима вторая формулировка категорического императива («Относись к кролику всегда как к цели и никогда как к средству»).

Кант придерживался идеи о том, что нерациональные животные не могут быть частью морального сообщества, и в этом пункте кантовский анализ Корсгард выходит за пределы собственно кантовской интерпретации. Она различает моральных субъектов в активном и пассивном смыслах слова, то есть тех, кто в состоянии формулировать правила морали с помощью разума, и тех, на кого они распространяются. Второе множество не обязательно совпадает с первым. В конечном счете мы полагаем, что наши моральные обязательства распространяются на младенцев или людей, находящихся в коме, которые не являются разумными.

Классическая кантовская интерпретация была связана с попыткой вывести человека из списка биологических видов, но мы знаем, что это когнитивное заблуждение, предложенное западной интеллектуальной культурой. Об этом есть чудесная книга Жана-Мари Шеффера «Конец человеческой исключительности»¹⁹. Аналогия с животными позволяет нам увидеть ту категориальную ошибку, которая предопределяет ход современных дискус-

17. *Singer P. Animal Liberation*. N.Y.: Avon Books, 1975.

18. *Korsgaard C. Fellow Creatures: Our Obligations to Other Animals*. Oxford: Oxford University Press, 2018.

19. *Шеффер Ж.-М. Конец человеческой исключительности*. М.: НЛЮ, 2010.

сий об использовании роботов на войне. Нет никаких причин считать, что машины будут находиться на одном уровне развития с человеком. Аргументация Корсгард применительно к роботам означает, что хотя они могут быть членами «морального сообщества», но вряд ли обладают способностью устанавливать для себя моральные законы. Следовательно, деонтология дает отрицательные ответы на три вопроса, представленные выше, причем этот вывод совпадает с моральными интуициями заметной части людей.

Действительно, современная война предполагает, что существуют определенные правила, ограничивающие как поведение враждующих сторон, так и использование целых классов вооружения, в частности химического и биологического²⁰. В практику войны тем самым органически инсталлированы этические ограничения, даже несмотря на то, что словосочетание «этичная война» может выглядеть оксюмороном. Базовые этические константы, признанные в мире, связаны с разграничением комбатантов и некомбатантов, утверждением права на справедливую (оборонительную) войну, а также с фактическим запретом на применение ядерного оружия. Отказ от применения ядерной бомбы можно назвать величайшим этическим достижением человечества, а такие военные теоретики, как Мартин ван Кревельд, отсчитывают историю современной войны именно с 1945 года — с момента, когда война без правил стала синонимом самоуничтожения человечества²¹.

В русле этих этических ограничений на ведение войны возникают движения против использования роботов на войне; самое известное из них — *Stop Killer Robots*. Документы, опубликованные на сайте движения, выражают беспокойство по поводу того, что появление автономных машин на войне означает постепенное исключение человека из цепочки принятий решений: сначала за людьми сохранится роль стратегического командования, но затем и она сойдет на нет — ради эффективности. Согласно докладу *Human Rights Watch*²², автономные боевые маши-

20. Price R. The Chemical Weapons Taboo. Ithaca, NY; L.: Cornell University Press, 1997.

21. Ван Кревельд М. Трансформация войны. М.: ИРИСЭН, 2008.

22. The “Killer Robots” Accountability Gap // Human Rights Watch. 08.04.2015. URL: <https://www.hrw.org/news/2015/04/08/killer-robots-accountability-gap>.

ны должны быть запрещены из-за неопределенности их правового статуса и проблемы ответственности за возможные военные преступления.

Аргументация противников летальных автономных вооружений представляется убедительной, так что «роботы-убийцы» вполне могут пополнить собой список запрещенных вооружений. Можно предполагать, что алармистские образы из массовой культуры сыграют здесь свою роль, а конвенция об ограничении роботов на войне будет принята под давлением общественного мнения и международных организаций. Однако парадоксальным образом запрет автономных летальных вооружений может привести лишь к переходу конфликта о статусе военных роботов на следующий уровень. Давление общественных кампаний заставит производителей дронов инвестировать в создание нелетальных автономных видов вооружения.

Интуиция Клаузевица, стоящая за классическим пониманием войны как продолжения политики другими средствами, предполагает, что разрешение наиболее интенсивных конфликтов возможно только в случае физического устранения их источников. Иными словами, тринитарная война разворачивалась как конфликт двух государств, ведущийся армиями и нацеленный на поражение противника любой ценой. Практика войны предполагает убийство в качестве финального средства, поскольку в ином случае пацифисты будут уничтожены своими противниками. Тем не менее нет никаких причин предполагать, что так должно быть всегда. Людям нужно убивать людей, чтобы избежать поражения в войне. Более совершенные в технологическом отношении машины смогут избегать убийств в ходе военных действий, добиваясь при этом традиционных военных целей (заключение мира на выгодных условиях, оборона территории, устранение угроз, в том числе летальных для некомбатантов). Разработчики военных роботов могут сделать ставку на создание машин, которые способны эффективно уничтожать машины противника и при этом гуманно относиться к противникам-людям, к примеру выводя их из строя, обездвигивая. В этом случае военные роботы будут заимствовать арсенал полицейских средств, направленных на противодействие толпе или задержание преступников. Наиболее перспективной технологией считаются полицейские тазеры. Даже если предположить, что нелетальные роботы будут нести более высокие потери, их создатели смогут утверждать, что их относительная неэффективность является приемлемой платой за отказ от летальных последствий для людей. Сама идея ро-

бота-убийцы, популяризованная массовой культурой, является примером ложного антропоморфизма — предположения о том, что машины будут заимствовать человеческие модели поведения и находиться на одном уровне развития с нами. Чтобы побеждать в войнах, машинам не нужно ни выглядеть как люди, ни убивать своих противников²³.

Соответственно, производители автономного оружия предложат политическим силам выбор между нелетальными боевыми роботами и массовыми жертвами людей в случае запрета таких роботов, если войны будут вестись традиционными методами. Утилитаристские аргументы отводятся сразу; вопрос о допустимости делегирования внеморальным существам права убивать также снимается. Развитие нелетальных автономных вооружений приводит к тому, что развитые страны делегируют ведение войн машинам, причем это касается уровня как тактических единиц, так и стратегических центров принятия решений. Возможно, число жертв военных конфликтов в итоге действительно бы уменьшилось, что позволило бы демократическим режимам продолжать инвестировать в развитие военных роботов.

Ключевая проблема связана, однако, именно с размыванием понятия войны, за чем следует ее нормализация. Война как практика взаимного уничтожения людей является радикальной чертой современной культуры и осуждается с моральной точки зрения; даже идея «справедливой войны» находит своих критиков. В то же время идея обеспечения безопасности с помощью полицейских мер не вызывает специальных негативных коннотаций и считается частью нормального порядка вещей. Война без жертв может идти бесконечно, а военные действия станут неотличимыми от «поддержания мира», а точнее, от полицейских операций. Различия комбатантов и некомбатантов уйдут в прошлое, а армия и полицейские силы, сдерживающие, скажем, городские бунты, станут неотличимы друг от друга. Идея гибридной войны находит логическое завершение в образе бесконечной полицейской акции, которая ведется автономными машинами в отношении не-

23. *McKendrick K. Banning Autonomous Weapons Is Not the Answer // Chatham House. The Royal Institute of International Affairs. 02.05.2018. URL: <https://www.chathamhouse.org/expert/comment/banning-autonomous-weapons-not-answer>.*

ограниченного круга лиц. Если на войне нет жертв, то нет никаких причин, почему нельзя применять военные практики к некомбатантам. Таким образом, изменение баланса в сфере военной техники и правил ведения войны вместе с появлением автономных машин может радикально переопределить наши представления о войне и мире.

Если исчезает различие мирного и военного времени, то политические режимы могут использовать военные машины для поддержания стабильности, не испытывая при этом морального давления. Гоббсовская война всех против всех впервые будет надежно остановлена. Но вместе с риском погибнуть граждане теряют и право на восстание против тиранического государства или корпорации, которые отныне и всегда избегают отнимать у кого-либо жизнь, довольствуясь простым «поддержанием порядка».

Очевидно, что утилитаризм в данном контексте сталкивается с почти неразрешимой проблемой калькуляции, в которой на одной чаше весов «вечный мир», а на другой — состояние перманентной полицейской операции. В то же время одна из проблем деонтологии связана с отсутствием в ней иерархии моральных обязательств. Категорический императив имеет абсолютную природу, что описано Кантом в его парадоксе, посвященном недопустимости лжи во спасение («О мнимом праве лгать из человеколюбия»²⁴). В данном случае ригоризм деонтологического проекта становится точкой, из которой мы можем критиковать развитие любых автономных машин, использующихся в военных и/или полицейских операциях, поскольку реализация сценария «машинного вечного мира» оказывается несовместима с человеческой свободой.

Проблема автономных машин на войне, таким образом, заключается не в том, что они могут убивать людей. Беспилотный автомобиль или робот-хирург тоже может совершить убийство, по крайней мере непреднамеренное. Кантианская оптика показывает, что автономные машины, использующие человека как средство для достижения мира и безопасности, несовместимы с человеческими представлениями о свободе морального субъекта. Ключевая особенность военных роботов заключается именно в их автономности и при этом исключенности из «морального со-

24. Кант И. О мнимом праве лгать из человеколюбия // Трактаты и письма / Под ред. А. В. Гулыги. М.: Наука, 1980. С. 292–297.

общества». Когда из войны исключаются люди, она дегуманизируется в моральном отношении: разрывается связь между субъектом морального выбора и тем миром, в отношении которого моральный выбор совершается.

Библиография

- Ван Кревельд М. Трансформация войны. М.: ИРИСЭН, 2008.
- Кант И. К вечному миру // Собр. соч.: В 6 т. М.: Мысль, 1966. Т. 6.
- Кант И. О мнимом праве лгать из человеколюбия // Он же. Трактаты и письма. М.: Наука, 1980. С. 292–297.
- Кант И. Основы метафизики нравственности // Собр. соч.: В 6 т. / Под общ. ред. В. Ф. Асмуса, А. В. Гулыги, Т. И. Ойзермана. М.: Мысль, 1965. Т. 4. Ч. 1. С. 219–310.
- Шеффер Ж.-М. Конец человеческой исключительности. М.: НЛО, 2010.
- Awad E., Dsouza S. et al. The Moral Machine Experiment // Nature. 2018. Vol. 563. P. 59–64.
- Charis V., Dennis L. et al. Towards Moral Autonomous Systems // ArXiv.org 14.03.2017. URL: <http://arxiv.org/pdf/1703.04741.pdf>.
- Chopra S., White L. A Legal Theory for Autonomous Artificial Agents. Ann Arbor: University of Michigan Press, 2011.
- Foot P. The Problem of Abortion and the Doctrine of Double Effect // Oxford Review. 1967. № 5. P. 5–15.
- Hamilton S. Impersonations: Troubling Person in Law and Culture. Toronto: University of Toronto, 2009.
- It's Basic Income. Global Debate / A. Downes, S. Lansley (eds). Bristol: Policy Press, 2018.
- Korsgaard C. Fellow Creatures: Our Obligations to Other Animals. Oxford: Oxford University Press, 2018.
- Lindner F., Bentzen M. A Formalization of Kant's Second Formulation of the Categorical Imperative // ArXiv.org. 09.01.2018. URL: <http://arxiv.org/abs/1801.03160>.
- McKendrick K. Banning Autonomous Weapons Is Not the Answer // Chatam House. The Royal Institute of International Affairs. 02.05.2018. URL: <http://chathamhouse.org/expert/comment/banning-autonomous-weapons-not-answer>.
- Minsky M. Steps Towards Artificial Intelligence // Proceedings of the IEEE. 1961. Vol. 49. P. 8–30.
- Moor J. Four Kinds of Ethical Robots // Philosophy Now. 2009. Vol. 72. P. 12–14.
- Muller V. Autonomous Killer Robots Are Probably Good News // Drones and Responsibility: Legal, Philosophical and Socio-Technical Perspectives on the Use of Remotely Controlled Weapons / E. Di Nucci, F. S. de Sio (eds). Basingstoke: Taylor & Francis, 2016. P. 67–81.
- Price R. The Chemical Weapons Taboo. Ithaca, NY; L.: Cornell University Press, 1997.
- Purves D., Jenkins R., Strawser B. J. Autonomous Machines, Moral Judgment, and Acting for the Right Reasons // Ethical Theory and Moral Practice. 2015. Vol. 18. № 4. P. 851–872.
- Scarre G. Utilitarianism. L.: Routledge, 1996.

- Searle J. Minds, Brains, and Programs // Behavioral and Brain Sciences. 1980. Vol. 3. № 3. P. 417–457.
- Singer P. Animal Liberation. N.Y.: Avon Books, 1975.
- Singer P. W. Wired for War. The Robotics Revolution and Conflict in the Twenty-First Century. N.Y.: Penguin Books, 2010.
- Strawser B. J. Moral Predators: The Duty to Employ Uninhabited Aerial Vehicles // Journal of Military Ethics. 2010. Vol. 9. № 4. P. 342–368.
- The “Killer Robots” Accountability Gap // Human Rights Watch. 08.04.2015. URL: <http://hrw.org/news/2015/04/08/killer-robots-accountability-gap>.

AN ETHICS OF AUTONOMOUS MACHINES: DEONTOLOGY AND MILITARY ROBOTS

KIRILL MARTYNOV. Associate Professor, School of Philosophy, Faculty of Humanities, kmartynov@hse.ru.

National Research University Higher School of Economics (HSE), 21/4 Staraya Basmannaya str., 105066 Moscow, Russia.

Keywords: deontology; lethal autonomous weapons; nonhuman ethics; Immanuel Kant; war.

The article is a contribution to the ethical discussion of autonomous lethal weapons. The emergence of military robots acting independently on the battlefield is seen as an inevitable stage in the development of modern warfare because they will provide a critical advantage to an army. Even though there are already some social movements calling for a ban on “killer robots,” ethical arguments in favor of developing those technologies also exist. In particular, the utilitarian tradition may find that military robots are ethically permissible if “non-human combat” would minimize the number of human victims. A deontological analysis for its part might suggest that ethics is impossible without an ethical subject.

Immanuel Kant’s ethical philosophy would accommodate the intuition that there is a significant difference between a situation in which a person makes a decision to kill another person and a situation in which a machine makes such a decision. Like animals, robots become borderline agents on the edges of “moral communities.” Using the discussion of animal rights, we see how Kant’s ethics operates with non-human agents. The key problem in the use of autonomous weapons is the transformation of war and the unpredictable risks associated with blurring the distinction between war and police work. The hypothesis of the article is that robots would not need to kill anyone to defeat the enemy. If no one dies in a war, then there is no reason not to extend its operations to non-combatants or to sue for peace. The analysis presented by utilitarianism overlooks the possibility of such consequences. The main problem of autonomous lethal weapons is their autonomy and not their potential to be lethal.

DOI: 10.22394/0869-5377-2019-3-231-244

References

- Awad E., Dsouza S. et al. The Moral Machine Experiment. *Nature*, 2018, vol. 563, pp. 59–64.
- Charisi V., Dennis L. et al. Towards Moral Autonomous Systems. *ArXiv.org*, March 14, 2017. Available at: <http://arxiv.org/pdf/1703.04741.pdf>.
- Chopra S., White L. *A Legal Theory for Autonomous Artificial Agents*, Ann Arbor, University of Michigan Press, 2011.
- Foot P. The Problem of Abortion and the Doctrine of Double Effect. *Oxford Review*, 1967, no. 5, pp. 5–15.
- Hamilton S. *Impersonations: Troubling Person in Law and Culture*, Toronto, University of Toronto, 2009.
- It’s Basic Income. Global Debate* (eds A. Downes, S. Lansley), Bristol, Policy Press, 2018.
- Kant I. K vechnomu miru [Zum ewigen Frieden]. *Sobr. soch.: V 6 t.* [Collected Works: In 6 vols], Moscow, Mysl’, 1966, vol. 6.

- Kant I. O mnimom prave lgať iz chelovekoliubiia [Über ein vermeintes Recht aus Menschenliebe zu lügen]. *Traktaty i pis'ma* [Treatises and Letters], Moscow, Nauka, 1980, pp. 292–297.
- Kant I. Osnovy metafiziki nravstvennosti [Grundlegung zur Metaphysik der Sitten]. *Sobr. soch.: V 6 t.* [Collected Works: In 6 vols], Moscow, Mysl', 1965, vol. 4, pt. 1, pp. 219–310.
- Korsgaard C. *Fellow Creatures: Our Obligations to Other Animals*, Oxford, Oxford University Press, 2018.
- Lindner F., Bentzen M. A Formalization of Kant's Second Formulation of the Categorical Imperative. *ArXiv.org*, January 9, 2018. Available at: <http://arxiv.org/abs/1801.03160>.
- McKendrick K. Banning Autonomous Weapons Is Not the Answer. *Chatam House. The Royal Institute of International Affairs*, May 2, 2018. Available at: <http://chathamhouse.org/expert/comment/banning-autonomous-weapons-not-answer>.
- Minsky M. Steps Towards Artificial Intelligence. *Proceedings of the IEEE*, 1961, vol. 49, pp. 8–30.
- Moor J. Four Kinds of Ethical Robots. *Philosophy Now*, 2009, vol. 72, pp. 12–14.
- Muller V. Autonomous Killer Robots Are Probably Good News. *Drones and Responsibility: Legal, Philosophical and Socio-Technical Perspectives on the Use of Remotely Controlled Weapons* (eds E. Di Nucci, F. S. de Sio), Basingstoke, Taylor & Francis, 2016, pp. 67–81.
- Price R. *The Chemical Weapons Taboo*, Ithaca, NY, London, Cornell University Press, 1997.
- Purves D., Jenkins R., Strawser B. J. Autonomous Machines, Moral Judgment, and Acting for the Right Reasons. *Ethical Theory and Moral Practice*, 2015, vol. 18, no. 4, pp. 851–872.
- Scarre G. *Utilitarianism*, London, Routledge, 1996.
- Schaeffer J.-M. *Konets chelovecheskoi iskluchitel'nosti* [La fin de l'exception humaine], Moscow, New Literary Observer, 2010.
- Searle J. Minds, Brains, and Programs. *Behavioral and Brain Sciences*, 1980, vol. 3, no. 3, pp. 417–457.
- Singer P. *Animal Liberation*, New York, Avon Books, 1975.
- Singer P. W. *Wired for War. The Robotics Revolution and Conflict in the Twenty-First Century*, New York, Penguin Books, 2010.
- Strawser B. J. Moral Predators: The Duty to Employ Uninhabited Aerial Vehicles. *Journal of Military Ethics*, 2010, vol. 9, no. 4, pp. 342–368.
- The “Killer Robots” Accountability Gap. *Human Rights Watch*, April 8, 2015. Available at: <http://hrw.org/news/2015/04/08/killer-robots-accountability-gap>.
- Van Creveld M. *Transformatsiia voyny* [The Transformation of War], Moscow, IRISEN, 2008.