УДК 172
DOI: 10.17223/1998863X/58/8

Н.А. Ястреб

ОРУЖИЕ И СОВРЕМЕННЫЙ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: ДВА ВАРИАНТА ОТРИЦАНИЯ ЦЕННОСТНОЙ НЕЙТРАЛЬНОСТИ ТЕХНОЛОГИИ

Показано, что тезис о ценностной нейтральности технических объектов в силу ряда противоречий не может быть использован в практике гуманитарной экспертизы техники. Проводится анализ теории вложения ценностей, сделан вывод о том, что данный подход не объясняет возможность использования технических объектов для реализации тех ценностей, которые не предполагались разработчиками, и не раскрывает сам процесс «вложения» ценностей, следовательно, не может быть использован как основа для проведения гуманитарной экспертизы технологий.
Ключевые слова: *ценность, артефакт, аксиология техники, ценностная нейтральность технологии, искусственный интеллект, теория вложения ценностей.*

Введение

Проблема этической, моральной и ценностной нейтральности технических объектов является фундаментальной для философии техники [1, 2]. Ее актуализация на разных этапах развития философских исследований техники, как правило, связана с появлением новых технических систем, функционирование которых так или иначе трансформирует социальные отношения и может изменить моральные или ценностные ориентации людей. Двойственная природа технических объектов[1], сущность которых определяется одновременно физической структурой и функциональностью, затрудняет анализ их гуманитарных свойств и проявлений, так как сложно выделить, к чему именно относятся наши утверждения – к артефактам как к физическим объектам, к их функциональным свойствам или к процессу их использования человеком. Эта непроясненность повсеместно используется для манипуляций сознанием людей, так как, меняя контекст и ракурс рассмотрения, можно показывать одни и те же технические объекты и как благо и как угрозу для человечества.

Этические и ценностные суждения о технике и технологиях используются для составления заключений экспертиз, на основе которых принимаются решения о внедрении, ограничениях или запрете использования тех или иных систем. В последние годы дискуссии вокруг проведения этих экспертиз начинают обостряться, в том числе в связи с распространением практики, при которой производители, например, систем искусственного интеллекта, утверждают, что в их технологии изначально «встроены» ценности, и выступают против строгих форм регламентирования. Распространившийся в последнее время термин «этический искусственный интеллект» означает, что

[1] Используемые термины «артефакт» и «технический объект», строго говоря, не являются тождественными, но в том контексте, в котором они употребляются в статье, их можно условно считать синонимами.

разработчики определяют и декларируют определенную систему принципов, которой они руководствуются при создании технологий. Тем самым демонстрируется добрая воля создателей, которые заявляют в качестве своих целей «содействие прогрессу общества и человеческой цивилизации, устойчивому развитию природы и общества, на благо всего человечества и окружающей среды, а также повышение благосостояния общества» [3]. Следовательно, строгая регламентация в форме юридических разрешений и запретов как будто бы становится избыточной, потому что разработчик уже интегрировал в эти инструменты значимые общечеловеческие ценности и этические принципы. Такой подход весьма распространен в современной философии техники [4] и базируется на представлении о том, что артефакты (любые технические объекты) сами по себе являются носителями ценностей или, по меньшей мере, способствуют реализации определенных ценностей. Однако это утверждение является дискуссионным и требует детального анализа.

Тезис о нейтральности технологии и его критика

Отправной точкой для анализа подходов к пониманию морального статуса артефактов является тезис о нейтральности технологии, согласно которому технические объекты сами по себе нейтральны как в этическом, так и в ценностном плане, пассивны, не воплощают и не транслируют никакие системы убеждений, не являются ни добром, ни злом. При этом ответственность за все позитивные и негативные последствия применения этих технологий, а также за изменение социальных отношений возлагается исключительно на тех, кто внедряет и использует данные технологии. Ценностную окрашенность могут иметь действия с применением технологий, но не сами технологии. Этот подход объясняет, в частности, почему одни и те же объекты в разных ситуациях могут способствовать реализации различных ценностей. Например, нож может реализовывать ценность безопасности, если он как оружие используется для защиты, или ценность здорового образа жизни, если он применяется для измельчения полезных продуктов. Сам нож при этом остается тождественным самому себе и не изменяется в зависимости от выбора способа его использования, а те ценности, которые реализуются с его помощью, проявляются только в деятельности человека и связаны с его субъектностью.

Критики данного подхода отмечают, что если принять тезис о нейтральности технологии, то тогда нужно признать, что разные технические объекты одинаково нейтральны. Например, при таком подходе автомат Калашникова и губка для мытья посуды равнозначны [5]. Рассмотрение техники как некоторого инструментария, не оказывающего никакого влияния на деятельность с его применением, сталкивается с серьезными противоречиями. В частности, многие практики становятся возможными только при наличии определенных технологий. Без современных биотехнологий невозможны клонирование или генетическая модификация организмов. Если какой-либо технический объект является необходимым условием определенной деятельности, способствующей реализации конкретных ценностей, то такой объект нельзя считать нейтральным. Встраиваясь в повседневные практики, технические артефакты «могут активно влиять на своих пользователей, изменяя то, как они воспринимают мир, как действуют в мире и как взаимодействуют друг с дру-

гом» [3. P. 1]. Так, П. Вербик, критикуя традиционное понимание артефактов как пассивных объектов действия, утверждает, что технические объекты являются частью «моральных агентов». В этом смысле в качестве субъекта действия выступает не просто человек, а гибрид человека и технологий [6].

Более сильный вариант отрицания тезиса о нейтральности технологии входит в число базовых идей подхода, получившего название теории вложения ценностей [7, 8]. Данная теория утверждает, что при создании артефактов в них могут вкладываться те или иные ценности. Основная идея теории вложения ценностей формулируется следующим образом. Артефакт x воплощает значение V, если спроектированные свойства x имеют потенциал для достижения или вклада в V из-за того, что x был разработан для V [7]. Ван де Поул приводит пример с дамбой. Если дамба возводится для обеспечения безопасности и действительно способствует безопасности, то она воплощает ценность безопасности [Там же]. То есть если артефакт вносит вклад в обеспечение той или иной ценности, то его можно считать носителем этой ценности. Автор называет данный вариант ценности «внешней конечной ценностью», имея в виду, что артефакт имеет ценность для других (внешняя ценность) и ценен сам по себе (конечная ценность). То есть самому артефакту не приписывается субъектность или даже активность и признается, что он представляет ценность только для внешних субъектов, но при этом ценность рассматривается как нечто связанное с самим объектом, неотделимое от него. По аналогии с аристотелевским учением о причинности, в котором объекты наделяются телеологическими свойствами, теория вложения ценностей рассматривает артефакты как носители поставленных при их проектировании ценностей.

М. Кленк, который не полностью соглашается с теорией вложения ценностей, анализируя понятие внешней конечной ценности артефактов, отмечает, что она определяется проектными свойствами технических объектов, так как «внутренние свойства артефакта (например, его физическая структура) не определяют ценность артефакта исчерпывающим образом, а его внешние свойства, намерения разработчиков артефакта, обязательно способствуют его ценности» [5]. То есть такую ценность нельзя назвать в полном смысле вложенной. Она может быть основана на физической структуре артефакта, но проявляется в процессе реализации функций только при участии внешних факторов.

Несмотря на то что теория вложения ценностей обосновывается рядом аргументов, она не объясняет, почему некоторые артефакты могут способствовать реализации тех ценностей, которые не были предусмотрены при их проектировании. Если следовать логике теории вложения, то заложенные в технических объектах ценности не могут изменяться. Передает ли нам жуткая, но в своем роде совершенная и прекрасная колесница, созданная Леонардо да Винчи для перерубания сухожилий лошадей и солдат, представления о красоте, гармонии и человеческом разуме, формирует ли она у нас вкус, эстетические ориентации? Технические объекты не имеют собственной субъектности и вряд ли способны самостоятельно производить ценности, однако это не означает, что они не могут транслировать ценности тех субъектов, которые их используют, но не имеют отношения к их созданию. Наш пример с колесницей Леонардо да Винчи показывает, что сначала объект мо-

жет рассматриваться как то, что главным образом способствует реализации военных задач, а через некоторое время – уже эстетических ценностей. Сами ценности при этом также могут переосмысливаться с течением времени.

Вопрос об изменении реализуемых техническими объектами ценностей связан с проблемой появления новых функций у артефактов. Возможность использования артефактов для реализации тех функций, для которых они не были предназначены, хорошо описана в литературе. Выделяют три основных способа изменения функциональности, а именно реинтерпретацию, адаптацию и переизобретение [5]. Все эти приемы могут приводить, в том числе, к изменению тех ценностей, которые транслируются с использованием артефакта, и могут рассматриваться как некоторые практики преобразования технических объектов, изменяющие ценностные аспекты взаимодействия с этими объектами. Реинтерпретация представляет собой изменение семантики без физических преобразований. Например, если рассматривать сохранившиеся фрагменты Берлинской стены, то использование граффити стало определяющим фактором, благодаря которому они воспринимаются не как символы разъединения, а как свидетельства торжества гуманизма и неизбежности разрушения тоталитарных режимов. Адаптация предполагает приспособление, в том числе конструктивное, артефактов для решения тех задач, которые не предполагались на этапе проектирования. М. Кленк приводит яркий пример с кассетными плеерами, которые создавались для проигрывания музыки, а стали использоваться для проведения религиозных проповедей, т.е. трансляция художественных, эстетических ценностей сменилась на приобщение к духовным, религиозным ценностям. Самый радикальный вариант преобразования артефакта – его переизобретение. К примеру, вычислительные приборы 40–60-х гг. XX в. создавались для решения исследовательских задач, а в 80-е гг. они переизобретаются в виде персональных компьютеров, встроенных в структуру ценностей общества потребления. Таким образом, мы видим, что новые функции появляются у артефактов, в том числе без изменения физической структуры, следовательно, некоторые ценности получают возможность реализации без их «вложения» на этапе проектирования.

И самым уязвимым местом теории является неопределенность самой процедуры «вложения ценности». Как именно оно происходит и какие явные и неявные намерения, убеждения и представления при этом вкладываются в артефакты? Оставляя в рамках данной статьи этот вопрос за скобками, отмечу, что при всей туманности и кажущейся мистичности данного процесса ссылки на высказывания о вложенных ценностях можно найти в самых серьезных документах, что заставляет более внимательно отнестись к данной проблеме, требующей отдельного рассмотрения.

Выявив серьезные уязвимости и в тезисе о нейтральности технологии, и в теории вложения ценностей, можно сделать вывод о том, что ценностные аспекты артефактов возникают в процессе их взаимодействия с человеком, однако сами артефакты здесь нельзя рассматривать как пассивные орудия, не оказывающие влияния на человека. Тезис о нейтральности технологии справедлив только в том случае, если технический объект не задействован напрямую или опосредованно в деятельности человека. В остальных случаях его нужно рассматривать как актор, который может задавать границы действия, влиять на принятие решений и способствовать реализации ценностей, однако

некорректно приписывать эти ценности объекту самому по себе, потому что их источником является гибридная система, включающая человека и технические объекты, используемые им.

Отрицание тезиса нейтральности технологии в практике производства и применения оружия и систем искусственного интеллекта: сравнительный анализ

Любое оружие производится для того, чтобы устрашать, предостерегать и по необходимости причинять вред. Использовать для этих целей теоретически можно все виды оружия, но на практике сам факт применения некоторых устройств рассматривается как необоснованная жестокость, вне зависимости от того, кто и при каких условиях их использует. Еще в Петербургской декларации 1868 г. об отмене употребления взрывчатых и зажигательных пуль указывалось, что «единственная законная цель, которую должны иметь государства во время войны, состоит в ослаблении военных сил неприятеля… употребление такого оружия, которое по нанесении противнику раны без пользы увеличивает страдания людей, выведенных из строя, или делает смерть их неизбежною, должно признавать не соответствующим упомянутой цели; употребление подобного оружия было бы противно законам человеколюбия» [9]. Если любые без исключения практики применения конкретного оружия признаются негуманными, то, на первый взгляд, это значит, что данное качество относится не к способу использования объекта, а к нему самому. Здесь требуется уточнение. Если представить себе, что какое-то запрещенное этой конвенцией оружие оказалось потеряно, забыто и лежит там, где никто не может его найти и применить, и при этом никто не может использовать его как угрозу, то вряд ли мы будем утверждать, что оно само по себе содержит антигуманные ценности. Скорее, речь идет о том, что любая деятельность с применением этого вооружения, вне зависимости от целей, намерений, тактики и других аспектов признается негуманной.

Практически все виды лазеров могут при неправильном использовании приводить к слепоте, однако существует ослепляющее лазерное оружие, специально созданное для перманентного ослепления и запрещенное Конвенцией о некоторых видах обычных вооружений [10]. Некоторые виды оружия не позволяют контролировать зону их поражающего действия, вследствие чего с высокой вероятностью при любом их использовании будет причинен вред гражданскому населению. Речь идет, в частности, об обсуждении применения кассетных боеприпасов, запрещенных Дублинской конвенцией [11]. При взрыве такие кассеты распадаются на множество отдельных суббоеприпасов, которые могут уцелеть при первоначальном взрыве и в дальнейшем привести к неконтролируемым последствиям. Конвенция прямо подчеркивает опасность самих кассетных боеприпасов вне зависимости от целей их применения, утверждая, что «взрывоопасные остатки кассетных боеприпасов убивают или калечат гражданских лиц, в том числе женщин и детей, препятствуют экономическому и социальному развитию, в том числе вызывая потерю средств к существованию, затрудняют восстановление и реконструкцию в постконфликтный период, задерживают процесс возвращения беженцев и внутренне перемещенных лиц или препятствуют ему, могут оказывать негативное воздействие на национальные и международные усилия по миростро-

ительству и оказанию гуманитарной помощи» [11. P. 32]. Как видно из цитаты, в данном случае негативные гуманитарные аспекты связываются непосредственно с самим оружием, точнее, с его неизбирательностью, а не со спецификой практик его применения.

Практика гуманистической оценки оружия основана на тезисе о том, что технические объекты сами по себе имеют свойства, способствующие или препятствующие реализации ценностей. Да, запрещенные виды оружия не обладают субъектностью и, не участвуя в человеческой деятельности, не транслируют ценности, но при любом своем применении вне зависимости от обстоятельств они приводят к необоснованным убийствам военных, бессмысленным жертвам среди гражданского населения и тем самым признаются негуманными и отрицающими ценность человеческой жизни. Международные документы, запрещающие применение оружия или производство, учитывают его ценностные аспекты, однако в аргументации мы, как правило, не находим положений в духе теории вложения ценностей, наделяющих сами орудия ценностями. Самые радикальные в этом отношении документы, такие как Конвенция по кассетным боеприпасам, обращают внимание на то, что любая деятельность, вне зависимости от целей и способов использования этого оружия, приводит к негативным гуманитарным последствиям.

В последние несколько лет появился ряд новых дискуссий, посвященных регламентированию использования технических объектов, в рамках которых активно продвигается теория вложенных ценностей. Речь идет о попытках разработки концепций «этической техники». Наиболее показательным примером является «этический искусственный интеллект», т.е. декларируемая разработчиками интеллектуальных продуктов система ценностей и этических принципов, реализованная в системах искусственного интеллекта [3, 12–14]. Сами системы искусственного интеллекта при этом представляются как технологии со «встроенными» ценностными ориентациями и этическими принципами. И речь здесь идет не о возможностях, которые открывает применение искусственного интеллекта, а о том, что такие интеллектуальные системы сами «воплощают принципы человеческого достоинства, единства, свободы, неприкосновенности частной жизни и культурного и гендерного разнообразия, а также основные права человека» и должны препятствовать распространению «вредных человеческих предрассудков» [14]. Вроде бы, налицо антиинструменталистский подход и отрицание нейтральности технологии, однако в следующем предложении авторы говорят о том, что в правовом поле такие системы должны иметь статус орудий.

Нужно отметить, что в целом выделяются три вида регулирования технологий. В первом случае не предусмотрено вообще никакого правового регулирования, соблюдение этических принципов является добровольными. Во втором случае применяется умеренное правовое регулирование, поощряющее или требующее технических корректировок, которые существенно не противоречат прибыли. Самый строгий вариант предполагает ограничительное правовое регулирование, запрещающее или строго контролирующее внедрение технологии. Как указывают некоторые эксперты [15], бум на создание огромного количества деклараций, кодексов и хартий по этике искусственного интеллекта, наблюдавшийся в 2017–2019 гг., связан, в первую очередь, с желанием пролоббировать выбор способов регулирования первого и

второго типов. Если «правильные» ценности и этические принципы уже вложены в систему, то проведение гуманитарной экспертизы и строгая регламентация в форме юридических разрешений и запретов как будто бы становятся избыточными, так как разработчик сам взял на себя все гуманитарные проблемы. Такой подход, когда разработчик определяет то, какие ценности транслирует его продукт, на практике уже сейчас позволяет влиять на получение разрешения на внедрение технологии, облегчая или вовсе минуя этическую экспертизу. При этом, как показывает проведенный анализ, некорректно приписывать носительство ценностей самим технологиям, поскольку ценности проявляются в процессе человеческой деятельности. Следовательно, «этический искусственный интеллект» может в реальной практике способствовать распространению совсем не тех ценностей, которые были продекларированы его разработчиками.

Таким образом, можно видеть, что в XX в. в международном праве доминировал деятельностный подход к определению ценностного статуса оружия, при котором отрицается тезис о нейтральности технологии, но при этом само оружие не наделяется ценностями вне практики его применения. На современном этапе, прежде всего в связи с массовым появлением систем искусственного интеллекта, обостряются противоречия между требованиями цифровой экономики к скорости внедрения инноваций и необходимостью их правового регулирования. Поскольку в данном случае именно гуманитарные, т.е. ценностные и этические, проблемы становятся определяющими, то для облегчения регламентирования производства и внедрения систем искусственного интеллекта их разработчики начинают активно использовать идеи теории вложения ценностей. В то же время, как было показано ранее, при таком подходе игнорируются возможности использования технических объектов для реализации тех ценностей, которые не предполагались разработчиками, а сам процесс «вложения» ценностей вовсе не поддается анализу.

Заключение

Проблема нейтральности технологии, несмотря на очевидную практическую значимость, является фундаментальной для понимания сущности техники и структуры деятельности человека. В настоящее время в связи с необходимостью разработки этических кодексов и юридических документов, регламентирующих производство, внедрение и применение новых технологий, особо внимательного анализа заслуживают те положения, которые кладутся в их основу. Приписывание ценностей самим техническим объектам, которое мы можем встретить в разнообразных декларациях принципов этического искусственного интеллекта, не является в полной мере обоснованным и может привести к серьезным негативным социальным и гуманитарным последствиям. В то же время в силу своей физической организации и функциональной направленности технические объекты конструируют многие виды деятельности человека, а некоторые делают возможными в принципе. Поэтому опосредованно они могут влиять на реализацию тех или иных ценностей, что обязательно должно учитываться в практике регулирования применения этих объектов. Методологически проблемы гуманитарной экспертизы и прогнозирования возможных направлений реализации ценностей при внедрении технологий были разработаны в XX в. в международном праве при оценке

различных видов оружия. Проведенный анализ показывает, что в сложившихся условиях такие подходы целесообразно использовать в отношении систем искусственного интеллекта, которые, и это очевидно как для разработчиков, так и для экспертов, обладают огромным потенциалом гуманитарных и социальных трансформаций. При этом в сфере интеллектуальных систем активно продвигаются основанные на теории вложения ценностей практики облегченного правового регулирования и передачи функции аксиологической и этической оценки от экспертов к разработчикам данных технологий. В этих условиях важно, чтобы специалисты в области философии техники, аксиологии, этики активно подключались к обсуждению и прояснению проблем, а также активно высказывались в публичном пространстве по данным вопросам, так как последствия технократического подхода к гуманитарной экспертизе внедряемых технологий могу быть весьма серьезными.

Литература

1. *Mitcham C., Walelbers K.* Technology and Ethics: Overview // A Companion to the Philosophy of Technology / eds. J.B. Olsen, S. Pedersen, V. Hendricks. West Sussex : Wiley Blackwell, 2009. P. 367–383.

2. *Elul J.* The Technological Society / J. Elul, J. Wilkinson (transl.), R. Merton (int.). New York : Vintage Books, 1964. 449 p.

3. *Beijing* AI Principles. URL: https://www.baai.ac.cn/news/beijing-ai-principles-en.html (accessed: 06.06.2020).

4. *The Moral* Status of Technical Artefacts / eds. P. Kroes, P.-P. Verbeek. Dordrecht : Springer Netherlands, 2014. Vol. 17. 248 p.

5. *Klenk M.* How Do Technological Artefacts Embody Moral Values? // Philosophy & Technology. 2020. Vol. 2. URL: https://link.springer.com/article/10.1007/s13347-020-00401-y (accessed: 06.06.2020).

6. *Verbeek P.P.* Moralizing Technology: Understanding and Designing the Morality of Things. London : University of Chicago Press, 2011. 200 p.

7. *Poel van de I., Kroes P.* Can technology embody values? // The Moral Status of Technical Artefacts. 2014. P. 103–124.

8. *Hoven van den J.* Design for values and values for design // Information Age. 2005. № 4. P. 4–7.

9. *Декларация* об отмене употребления взрывчатых и зажигательных пуль. URL: http://old.memo.ru/prawo/hum/spb-1868.htm (дата обращения: 04.06.2020).

10. *Конвенция* ООН по конкретным видам обычного оружия 1980 года URL: https://www.icrc.org/ru/document/konvenciya-oon-po-konkretnym-vidam-obychnogo-oruzhiya-1980-goda (дата обращения: 01.06.2020).

11. *Конвенция* по кассетным боеприпасам. URL: https://www.un.org/ru/documents/decl_conv/conventions/pdf/cluster_munitions.pdf (дата обращения: 01.06.2020).

12. *AI* Ethics Principles/ Australian Government, Department of Industry, Science. URL: https://www.industry.gov.au/data-and-publications/ building-australias-artificial-intelligence-capability/ai-ethics-framework/ai-ethics-principles

13. *On* Artificial Intelligence – A European approach to excellence and trust. URL: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

14. *Top* Ten Principles for Ethical AI. URL: http://www.thefutureworldofwork.org/media/35420/uni_ethical_ai.pdf (accessed: 01.06.2020).

15. *Ochigame R.* The invention of “ethical AI”: how big tech manipulates academia to avoid regulation // The Intercept. 2019. URL: https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/ (accessed: 01.06.2020).

Natalia A. Yastreb, Vologda State University (Vologda, Russian Federation).

E-mail: nayastreb@mail.ru

Vestnik Tomskogo gosudarstvennogo universiteta. Filosofiya. Sotsiologiya. Politologiya – Tomsk State University Journal of Philosophy, Sociology and Political Science. 2020. 58. pp. 75–84.

DOI: 10.17223/1998863X/58/8

WEAPONS AND ARTIFICIAL INTELLIGENCE: TWO OPTIONS FOR DENYING THE VALUE-NEUTRALITY THESIS

Keywords: value; artifact; axiology of technology; value-neutrality thesis; artificial intelligence; theory of value embedding.

The article shows that the value-neutrality thesis contains contradictions and cannot be used in the practice of humanitarian assessment of technology. The theory of value embedding is analyzed. The conclusion that this approach does not explain the possibility of using technical objects to implement values that were not anticipated by the developers and does not disclose the process of "embedding" values has been made. It is shown that the concept of value embedding cannot be used as the basis for a humanitarian assessment of technologies. The attribution of values to the technical objects themselves, which we can find in various concepts of ethical artificial intelligence, is not fully justified and can lead to serious negative social and humanitarian consequences. At the same time, due to their physical organization and functional orientation, technical objects construct many types of human activity, and make some of them possible. Therefore, they can indirectly influence the implementation of certain values, and this must be taken into account in the practice of regulating the use of these objects. The problems of the humanitarian assessment and forecasting of the possible directions of the implementation of values during the implementation of technologies were developed in the 20th century in international law when assessing various types of weapons. The analysis shows that under the current conditions it is advisable to use such approaches in relation to artificial intelligence systems that can cause serious humanitarian and social transformations. At the same time, in the field of intelligent systems, the practice of facilitated legal regulation based on the theory of value embedding is being actively promoted. The functions of axiological and ethical assessment in such cases are in fact transferred from experts to developers of these technologies. In these conditions, it is important that specialists in the field of philosophy of technology, axiology, ethics actively speak out in the public space on these issues.

References

1. Mitcham, C. & Walelbers, K. (2009) Technology and Ethics: Overview. In: Berg Olsen, J., Pedersen, S. & Hendricks, V. (eds) *A Companion to the Philosophy of Technology*. West Sussex: Wiley Blackwell. pp. 367–383.

2. Elul, J. (1964) *The Technological Society*. New York: Vintage Books.

3. China. (n.d.) *Beijing AI Principles*. [Online] Available from: https://www.baai.ac.cn/news/beijing-ai-principles-en.html (Accessed: 6th June 2020).

4. Kroes, P. & Verbeek, P.-P. (eds) (2014) *The Moral Status of Technical Artefacts*. Vol. 17. Dordrecht: Springer Netherlands.

5. Klenk, M. (2020) How Do Technological Artefacts Embody Moral Values? *Philosophy & Technology*. 2. DOI: 10.1007/s13347-020-00401-y

6. Verbeek, P.P. (2011) *Moralizing Technology: Understanding and Designing the Morality of Things*. London: University of Chicago Press.

7. Poel, van de I. & Kroes, P. (2014) Can technology embody values? *The Moral Status of Technical Artefacts*. pp. 103–124. DOI: 10.1007/978-94-007-7914-3_7

8. Hoven van den, J. (2005) Design for values and values for design. *Information Age*. 4. pp. 4–7.

9. Russia. (1869) *Deklaratsiya ob otmene upotrebleniya vzryvchatykh i zazhigatel'nykh pul'* [Declaration on the abolition of the use of explosive and incendiary bullets]. [Online] Available from: http://old.memo.ru/prawo/hum/spb-1868.htm (Accessed: 4th June 2020).

10. UNO. (1980) *Konventsiya OON po konkretnym vidam obychnogo oruzhiya 1980 goda* [The 1980 UN Convention on Certain Conventional Weapons]. [Online] Available from: https://www.icrc.org/ru/document/konvenciya-oon-po-konkretnym-vidam-obychnogo-oruzhiya-1980-goda (Accessed: 1st June 2020).

11. UNO. (n.d.) *Konventsiya po kassetnym boepripasam* [Convention on Cluster Munitions]. [Online] Available from: https://www.un.org/ru/documents/decl_conv/conventions/pdf/cluster_munitions.pdf (Accessed: 1st June 2020).

12. Australian Government, Department of Industry, Science. (n.d.) *AI Ethics Principles*. [Online] Available from: https://www.industry.gov.au/data-and-publications/building-australias-artificial-intelligence-capability/ai-ethics-framework/ai-ethics-principles

13. European Comission. (2020) *On Artificial Intelligence – A European approach to excellence and trust*. [Online] Available from: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

14. UNI Global Union. (n.d.) *Top Ten Principles for Ethical AI*. [Online] Available from: http://www.thefutureworldofwork.org/media/35420/uni_ethical_ai.pdf (Accessed: 1st June 2020).

15. Ochigame, R. (2019) The invention of "ethical AI": how big tech manipulates academia to avoid regulation. *The Intercept*. [Online] Available from: https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/ (Accessed: 1st June 2020).