

ЭТИЧЕСКИЕ РИСКИ В ПРАКТИКЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Казарян Валентина Павловна,

*доктор философских наук, профессор,
профессор кафедры философии естественных факультетов
Московского государственного университета имени М.В. Ломоносова,
Россия, 119991, г. Москва, Ленинские горы, 1
ORCID: 0000-0001-9435-8625
kazaryanvp@mail.ru*

Шутова Ксения Юрьевна,

*аспирант,
математик первой категории кафедры
нелинейных динамических систем и процессов управления
Московского государственного университета имени М.В. Ломоносова,
Россия, 119991, г. Москва, Ленинские горы, 1
ORCID: 0000-0002-4915-0312
shutova.k.u@yandex.ru*

Аннотация

В статье поставлен вопрос о соотношении практики искусственного интеллекта и общечеловеческой этики. Тема принципиально важная, поскольку человек и общество, потерявшие этические основания, лишаются человечности. Этика ученых, конструкторов, менеджеров высокого уровня играет решающую роль в современных процессах развития и применения искусственного интеллекта. В их деятельности развивается этика ответственности, берущая свое начало с Манифеста Рассела–Эйнштейна, опубликованного в 1955 году в условиях ядерного противостояния периода холодной войны. В статье показан современный невероятно быстрый рост и развитие применений искусственного интеллекта в практической жизни людей. Обращено внимание на факт наличия неопределенности в ситуациях практического применения искусственного интеллекта. Подчеркивается значение наличия непредвиденных последствий помимо предполагаемых последствий. Ставится вопрос о том, кто отвечает за эти последствия: частные лица, корпорации, правительства? Делается вывод, что ответственность лежит на том, кто принимает решение на действие, – на акторе. Следовательно, актор находится в ситуации этического риска. Показано, что риск возрастает в силу ряда обстоятельств: 1) разнообразия применений; 2) неконтролируемого безудержного роста разнообразия; 3) трудностей в отслеживании эмпирической ситуации применения; 4) трудностей теоре-

тического анализа ситуации действия. В статье внимание сосредоточено на рисках в практике мирного использования дистанционно пилотируемых летательных аппаратов и автоматизированного оружия, а также их военного использования без подтверждения оператора. Экспоненциальный рост использования информационных технологий, практического воплощения искусственного интеллекта ставит человека в сложную этическую ситуацию – что выбрать: иметь или быть, или же есть что-то третье?

Ключевые слова: этика ответственности, искусственный интеллект, актор, дистанционно пилотируемый летательный аппарат, автономная система.

Библиографическое описание для цитирования:

Казарян В.П., Шутова К.Ю. Этические риски в практике искусственного интеллекта // Идеи и идеалы. – 2023. – Т. 15, № 2, ч. 2. – С. 346–364. – DOI: 10.17212/2075-0862-2023-15.2.2-346-364.

То, что мы можем что-то сделать,
не значит, что мы должны это делать.

Леонгард Г.

Несколько десятилетий назад интеллектуалами был поставлен и горячо обсуждался вопрос о значении развития современных технологий для судеб человечества. Несут ли они благо (добро), несут ли они опасность (зло) или нейтральны сами по себе относительно добрых и злых дел и требуют руководства со стороны человека. Возник конфликт мировоззрений. С тех пор ситуация обострилась. Мир неожиданно вступил в эпоху чрезвычайно высоких темпов технологического развития. Эти экспоненциально развивающиеся технологические достижения не только имеют огромный потенциал, но и несут большую опасность

Искусственный интеллект (ИИ) является первой движущей силой в области экспоненциально растущих технологий. Мы находимся в переломной точке экспоненциальной кривой во многих областях науки и техники – точке, где рост технологических изменений становится всё более значительным. Сейчас экспоненциальный рост наблюдается в таких областях, как глубокое обучение, генетика, материаловедение и производство.

В этом контексте практику искусственного интеллекта можно определить как разработку программ или роботов, обладающих интеллектом и способных к самообучению. В связи с этим на человечество надвигается крупная и масштабная проблема, связанная с технологиями в целом. Некоторые авторы называют ее проблемой «технологического ожирения» [5]. Задумывались ли мы когда-нибудь о «передавании» технологиями? Технологическое ожирение определяется как психическое и техническое состояние, при котором данные, информация и всеобщее цифровое едине-

ние распространены до такой степени, что это негативно сказывается на здоровье, настроении и жизни людей. Анонимность и конфиденциальность тоже могут стать роскошью, недоступной для большинства людей. По этой причине выдвигается предположение, что вопрос «Позволяют ли технологии что-то делать?» скоро заменит более актуальный: «Нужно ли нам делать то, что позволяют технологии на сегодняшний день, и почему?»

С лавинообразным развитием Интернета и соцсетей росло и количество информации. ИТ-компаниям необходимо было что-то делать с получаемыми данными, и применение ИИ стало необходимостью, а не модой. Google теперь сортирует поисковые объявления с помощью обучающихся программ Machine Learning (ML). YouTube подбирает рекомендуемые видео с помощью алгоритмов ML. По той же схеме рекомендует товары и Amazon. Новостная лента Facebook генерируется «умным» компьютером. И даже Tinder находит людей, используя алгоритмы ML. Технологии искусственного интеллекта стали нашими незаменимыми помощниками. И хотя последствия достижений в области искусственного интеллекта вызывают споры, дальнейшее развитие этой технологии неизбежно.

Ожидается, что возможности искусственного интеллекта будут развиваться в два раза быстрее, чем другие технологии, опережая даже рост вычислительных мощностей по закону Мура [11]. По оценкам Всемирного экономического форума 2016 года, в течение следующих 10 лет (2017–2027) общая стоимость перехода в цифровую среду составит 100 триллионов долларов. Эксперты предполагают, что это свидетельствует о возможности «создать перспективную рабочую силу будущего, в котором люди и умные машины трудятся вместе на благо Земли и человечества» [25]. Но в этом отчете не упоминается этика, а это один из ключевых факторов сохранения человечества.

Тема искусственного интеллекта давно волнует человечество. В 1942 году Айзек Азимов в рассказе «Хоровод» впервые сформулировал три закона робототехники.

1. Робот *не может причинить вред человеку* или своим бездействием допустить, чтобы человеку был причинен вред.
2. Робот *должен повиноваться всем приказам*, которые дает человек, кроме тех случаев, когда эти приказы противоречат Первому Закону.
3. Робот *должен заботиться о своей безопасности* в той мере, в которой это не противоречит Первому или Второму Закону.

С появлением ИИ, по словам Френка Паскуале, три закона А. Азимова необходимо дополнить еще четырьмя.

4. Цифровые технологии *должны дополнять деятельность профессионалов*, а не заменять их.
5. ИИ и роботизированные системы *не должны подделывать* человека.

6. В области ИИ *следует предотвратить усиление гонки вооружения.*

7. Роботы и системы ИИ *должны указывать на личность своих создателей, контролирующих их людей и владельцев* [7].

Ф. Паскуале утверждал, что если новые законы будут соблюдаться, то станет возможным использовать технологии для дополнения рабочих мест всех видов, а не для их поглощения, особенно в секторах здравоохранения и образования. Директивные и регулирующие органы должны способствовать формированию и демократизации результатов работы в области искусственного интеллекта. По его словам, «сдерживание автоматизации, которая контролирует, клеймит и обманывает невинных людей, – жизненно важная роль регуляторов XXI века» [7].

Ситуация с лавинообразным развитием ИИ стала беспокоить многих выдающихся людей. Еще Н. Винер, основоположник кибернетики, на заре научных исследований в этой области был убежден, что необходим контроль со стороны человека за развитием компьютеров [4]. В феврале 2023 года известный предприниматель Илон Маск поставил ИИ в один ряд с ядерными технологиями. Он подчеркнул, что у человечества есть ядерная энергия и есть ядерная бомба. Теперь есть искусственный интеллект, который может быть мирным и может быть военным. В ситуации опасности требуется человеческое внимание [6].

В литературе есть примеры, в которых явно просматривается вред, который приносит людям использование новых технологий. Это использование многоагентных систем с роевым поведением в военных целях [20]. Авторы сделали несколько художественных роликов, демонстрирующих работу БПЛА, предназначенных для стрельбы на поражение без подтверждения оператора. Они рассказывают о возможностях, которые может дать автоматизированное оружие на основе идентификации цели по данным, собираемым из различных источников: социальных сетей, фотографий и пр. Авторы предупреждают о том, что данное направление является опасным и разрушительным, так как текущая концепция, когда решение на убийство людей принимает искусственный интеллект, ведет к разрушительным последствиям для человечества. Конечно, ИТ-технологии, включая использование дронов, нельзя ассоциировать только с опасным оружием. Но, с другой стороны, любое оружие или технологию необходимо контролировать.

Смертоносное автономное оружие использует искусственный интеллект для идентификации человека и его убийства без вмешательства оператора. Если в случае с беспилотными военными дронами решение о лишении жизни принимается дистанционно человеком-оператором (актером), то в случае со смертоносным автономным оружием решение принимается исключительно алгоритмами. Смертоносные роботы заранее за-

программированы на убийство определенного «целевого профиля». Затем оружие помещается в окружающую среду, где ИИ ищет цель, используя данные датчиков, например распознавание лиц.

По мнению авторов [20], оружие, использующее для убийства алгоритмы, а не человеческие суждения, является аморальным и представляет серьезную угрозу национальной и глобальной безопасности. Это аморально, так как алгоритмы не способны понять ценность человеческой жизни, поэтому они не должны иметь права решать, кому жить, а кому умереть. Генеральный секретарь ООН Антониу Гутерриш согласен с тем, что «машины, обладающие полномочиями и способные по своему усмотрению отнимать жизнь без участия человека, политически неприемлемы, морально отвратительны и должны быть запрещены международным правом» [15].

Это несет угрозу безопасности, так как алгоритмическое принятие решений будет крайне дестабилизирующим фактором как на национальном, так и на международном уровне, поскольку создаст угрозы распространения, быстрой эскалации, непредсказуемости и даже потенциальной возможности создания оружия массового уничтожения.

Совершенствование искусственного интеллекта делает автономное оружие, способное распознавать и атаковать человеческие цели, вполне вероятной перспективой в ближайшем будущем. Но критики утверждают, что автономное оружие никогда не сможет соблюдать основные правила войны, например, различать солдат и гражданских лиц или соблюдать пропорциональность в военных действиях. Также не ясно, кто будет нести ответственность, если автономное оружие выйдет из строя на поле боя или будет взломано. И хотя ряд стран призвал к запрету так называемых «смертоносных автономных систем вооружений», их поддерживают не все страны.

Сегодня беспилотники невоенного назначения (назовем их мирными беспилотниками) широко вошли в жизнь. Приведем пример. В июле 2011 года в небольшом округе в Северной Дакоте возник спор о праве собственности на бродячий скот. Шериф начал действовать, планируя арестовать обвиняемых, но не имел возможности узнать, вооружены ли они. Он обратился за помощью на близлежащую базу ВВС Гранд-Форкс, и они прислали беспилотник, чтобы оценить ситуацию. После того как беспилотник подтвердил, что никто из подозреваемых не вооружен, местные власти смогли выехать на место и арестовать их без риска кровопролития. Обвиняемый заявил, что дело было несправедливым из-за незаконного использования беспилотников, и обратился в суд с иском против полиции. Это был первый случай, когда арест был произведен с помощью беспилотника [21]. Этот случай является примером того, как беспилотники интегрируются в различные стороны жизни общества.

Еще один громкий скандал произошел в Йемене (Рада), когда в декабре 2013 года американский БПЛА обстрелял свадебный кортеж вместо боевиков «Аль-Каиды» [23]. Погибло 17 человек. Вскоре после события парламент Йемена проголосовал за запрет на использование беспилотников США на территории страны, что подчеркнуло важность сохранения жизни мирных граждан. В октябре 2013 года правозащитная организация Human Rights Watch заявила, что в результате шести атак беспилотников США в Йемене с 2009 по 2013 год погибли 57 мирных жителей.

Удары американских БПЛА в Йемене и Пакистане нарушают международные законы. США тем не менее продолжают использовать беспилотники. Эти различные мнения о международном использовании дронов повышают напряженность между странами и могут привести к международному конфликту. Например, после нескольких случаев атак БПЛА в ответ на удары беспилотников протестующие пакистанцы поклялись помешать маршрутам поставок НАТО в Афганистан [23]. Согласно Geo News, ведущему новостному каналу Пакистана, министр внутренних дел Пакистана Чодри Нисар заявил: «Удары боевых дронов в племенных районах страны могут привести к конфликту между США и Пакистаном» [12].

Беспилотники используются для нанесения военных ударов, в правоохранительной деятельности, для защиты окружающей среды, в журналистике, для научных исследований, ликвидации последствий стихийных бедствий, охраны границ, международного шпионажа. Независимо от того, одобряют или не одобряют люди применение этого инструмента, очевидно, что БПЛА стали значительной силой в современном мире, и для них открываются всё новые области применения. Если не контролировать беспилотники должным образом, они будут использоваться для вторжения в частную жизнь человека. Теперь от людей всего мира зависит, насколько ответственно будут использоваться БПЛА.

Хотя ценность и моральность использования беспилотников вызывает жаркие споры, при правильном подходе они могут принести большую пользу человечеству.

Существует огромное разнообразие конструкций и размеров БПЛА, но, несмотря на такой широкий диапазон, большинство гражданских БПЛА строятся любителями и представляют собой квадрокоптеры, лопасти которых вращаются вокруг четырех точек вращения, что обеспечивает высокую степень устойчивости; размер этих аппаратов обычно составляет один кубический фут. Впрочем, название «беспилотный летательный аппарат» к подавляющему их большинству неприменимо, поскольку большинство беспилотников не являются ни беспилотными, ни транспортными средствами. Эти машины следует называть *дистанционно пилотируемыми*

летательными аппаратами, поскольку у этих машин на самом деле есть человек-пилот, управляющий ими дистанционно.

Беспилотники широко используются военными США в борьбе с терроризмом. В то время как в США существует общая поддержка использования беспилотников американскими военными и ЦРУ, опросы, проведенные за рубежом, показывают, что большинство населения против беспилотников [17]. Многие люди испытывают дискомфорт от использования БПЛА в быту. Опрос проводился более чем в 30 странах, и только в Кении, Израиле и США общественность поддерживает американскую кампанию по использованию беспилотников. Такая непопулярность беспилотников вызвана моральными опасениями, жертвами среди гражданского населения, подозрениями в росте уголовно наказуемых операций, нарушениями прав человека и игнорированием международных гуманитарных законов. США использовали беспилотники в Афганистане, Пакистане, Сомали и Йемене, чтобы поражать боевиков «Аль-Каиды», «Галибана» и других террористических организаций.

Правительство США, а точнее, Центральное разведывательное управление (ЦРУ), утверждает, что удары беспилотников являются точными и не приводят к жертвам среди гражданского населения (или приводят к очень незначительным жертвам). Однако факты свидетельствуют о том, что это не так, что вызывает много критики в отношении моральности использования беспилотников. Противники беспилотников утверждают, что они контрпродуктивны и вызывают рост боевиков в таких районах, как племенные регионы Пакистана, где в процессе их применения погибли невинные люди. Более того, легкость целенаправленных убийств с помощью беспилотников дает операторам огромную власть и нарушает Всеобщую декларацию прав человека, а также Международное гуманитарное право, которое обеспечивает защиту гражданских лиц и спасателей в зоне военных действий и за ее пределами.

Роберт Герниер, возглавлявший контртеррористический центр ЦРУ с 2004 по 2006 год и ранее работавший начальником отделения ЦРУ в Пакистане, заявил: «Мы соблазнились [беспилотниками], и непредвиденные последствия наших действий перевесят предполагаемые последствия» [13]. Поскольку беспилотники стали источником страха для многих людей, не являющихся военнослужащими, в таких местах, как племенные районы Пакистана и Йемена, они могут принести американской войне с терроризмом больше вреда, чем пользы. Опросы показывают, что эти атаки, возможно, создают больше врагов, чем убирают их с поля боя. По данным Pew Research Global Attitudes Project, опрос 2012 года показал, что 74 % пакистанцев считают США своим врагом, что выше, чем 69 %, зафиксиро-

ванных в 2011 году, и 64 % в 2009 году, вероятно, отчасти из-за нынешней кампании беспилотников [18, 24].

Несмотря на рост антиамериканских настроений, некоторые официальные лица США утверждают, что удары беспилотников эффективно ограничивают способность боевиков организовываться и действовать [24]. В настоящее время не ясно, так ли это на самом деле. Убийства гражданских лиц в зоне военных действий – не редкость, однако убийства в ходе военных операций и ударов беспилотников в невоенной зоне являются нарушением национального суверенитета. Выгоды, связанные с ударами беспилотников, приведут, по-видимому, к гонке вооружений между различными странами. Именно поэтому необходима правовая база для использования беспилотников, так как они становятся всё более распространёнными.

Четвертая технологическая революция, с которой ассоциируются цифровые технологии, меняет социальную структуру. Цифровая трансформация затрагивает государство, бизнес и повседневную жизнь каждого – мы все сталкиваемся с этим. Появляются новые стили жизни, новые методы действия. Цифровые технологии, кардинально меняя материальный и социальный контекст нашей жизни, приводят к пересмотру наших поведенческих норм и представлений об этичном и неэтичном поведении в конкретных ситуациях.

Взаимодействие человека и технологий давно обсуждается в философии, но с появлением новых технологий возникли проблемы иного характера. Научный прогресс несет в себе как возможности, так и определенные риски для общества. Обсуждение этических вопросов применительно к цифровым технологиям может стать способом минимизации потенциальных рисков развития технологий. Развитие технологий, включая ИИ, всё чаще дает повод для этических размышлений. Законы робототехники Айзика Азимова более полувека оставались фантастикой, но сегодня юридическое регулирование ИИ и его использование в самой юриспруденции становится повседневной практикой. Эта область называется «робоправо».

До недавнего времени искусственный интеллект не регулировался ни в России, ни в мире. В сентябре 2017 года ситуация кардинально изменилась: международная юридическая фирма Dentins и Российский исследовательский центр проблем регулирования робототехники и искусственного интеллекта подписали соглашение о сотрудничестве и общем взаимодействии по широкому кругу вопросов, связанных с развитием правового регулирования робототехники, ИИ и новых технологий в Российской Федерации. Сейчас организация именуется как АНО «ПравоРоботов» [2].

Внедрение систем ИИ в повседневную жизнь сопряжено со множеством этических проблем, которые уже в ближайшие годы будут всё се-

рбезнее и сложнее. Одной из наиболее острых этических и правовых проблем искусственного интеллекта считается проблема ответственности за действия робота или дрона. Этика социальной ответственности ученого начала формироваться в 50–60-е годы XX века. В 1955 году в Лондоне был провозглашен Манифест Рассела–Эйнштейна: в век ядерного оружия мы должны думать по-новому. Его подписали выдающиеся ученые. Он положил начало движению ученых за мир. Пагуошское движение, движение врачей воплотили идеи Манифеста. Ученые начали преодолевать традиционную этику невмешательства в пользу этики социальной ответственности. В последние десятилетия развиваются исследования вопросов этики информационных технологий (ИТ) под эгидой ЮНЕСКО. Этика стала неотъемлемой частью деятельности ИТ-компаний. Вопросы этики ИТ стали преподавать студентам всех крупных вузов мира.

С каждым годом растет число неприятных этических ситуаций. Например, казалось бы, идеальная, безопасная и автономная машина Tesla теряет управление и на большой скорости сбивает людей [3]. Чья вина здесь: пользователя или разработчиков? Как максимально сократить смертельный ущерб от автономных систем? А как расценивать необдуманные действия пользователей автономного автомобиля Tesla, которые экспериментируют над возможностями машины, подвергая опасности собственных и чужих детей? Владельцы этих авто массово в социальных сетях выкладывали видеоролики, где они специально выставляют курс движения автомобиля на ребенка и ожидают реакцию автомобиля [22]. Тренд появился после ролика организации Dawn Project, в котором Tesla Model 3 несколько раз сбивает манекен ребенка. Согласно проведенному исследованию, электромобили Tesla на автопилоте якобы не определяют детей на дорогах. Вскоре после этих экспериментов бывший кандидат в президенты США Ральф Надэр призвал запретить автопилот Tesla, назвав его внедрение самым опасным и безответственным решением. Владельцы электромашин решили доказать, что это не так. Один из них направил Tesla Model 3 на своего ребенка посреди дороги, а затем повторил попытку с другим ребенком, переходящим улицу. В обоих случаях автопилот вовремя сбросил скорость. К проверкам подключились другие владельцы электрокаров. YouTube уже начал удалять подобные ролики за нарушение правил. Водителей попросили перестать проводить подобные испытания.

Другая проблема – использование персональных данных: нужно ли их обезличивать или же использовать в строго определенных случаях (например, согласие субъекта, необходимость в суде). Корпорации Google, Apple, Facebook, Amazon, Microsoft собирали и обрабатывали миллионы аудиозаписей, в которые помимо устных команд голосовым помощникам и «умным» колонкам попадали фрагменты разговоров и другие звуки.

Запись велась автоматически, без предупреждения пользователей, а данные использовались для обучения ИИ распознаванию человеческой речи [14, 19]. Большинство этических проблем, связанных с данными, возникает из-за огромной стоимости информации о гражданах и колоссальной заинтересованности корпораций. Ради увеличения прибыли корпорации идут на нарушение прав граждан при обработке их данных.

А кто в ближайшем будущем будет решать, что есть человек, когда в нашей крови будут плавать наноботы, а в мозгу будут установлены коммуникационные импланты? Если у технологий нет и не должно быть этики, что случится с нашей моралью, нормами поведения и социальными договоренностями, когда машины будут управлять всем вместо нас? Раньше это казалось научной фантастикой, но уже появились первые реальные прототипы подобных наноботов, например, для транспортировки лекарства внутри организма. В перспективе, по мнению ученых, разработка позволит растворять тромбы и камни в почках и предотвращать инфекции после трансплантации [10].

К тому моменту, когда технологии ботов будут настолько совершенны, что беспилотники будет уже не страшно выпускать на дороги общего пользования, вся сопутствующая инфраструктура уже должна будет соответствовать требованиям безопасности. В градостроительных планах должна быть предусмотрена соответствующая инфраструктура, должны появиться новые СНиПы (строительные нормы и правила) по строительству дорог. Повсеместно должен быть Wi-Fi, который позволит беспилотным автомобилям взаимодействовать друг с другом.

Вопрос цифровой безопасности беспилотника, да и в принципе любого современного автомобиля с доступом в Интернет, с точкой подключения Wi-Fi и с возможностью подключения Bluetooth-гарнитуры, является актуальным с тех пор, как хакер в США смог подключиться к автомобилю во время движения по автостраде и полностью его контролировать (в частности, он смог заглушить двигатель). И это произошло на автомобиле, который не был полностью цифровым гаджетом, как беспилотник. Соответственно, для автономных автомобилей вопрос обновления программного обеспечения, установки межсетевых экранов (firewall) и антивирусов будет крайне актуален.

Проблема ответственности за действия систем ИИ – начиная от беспилотного автомобиля и заканчивая пока гипотетическими планетарными системами принятия решений – является наиболее обсуждаемой, когда речь заходит о применении ИИ. Проблема ответственности возникает там, где людям приходится полагаться на действия систем ИИ, таких как автомобильный транспорт, фармацевтика, медицина, образование и так далее. По мнению А.В. Незнамова, основателя проекта «Робоправо», на это

есть ряд причин: «Во-первых, институты ответственности могут иметь отдельные нюансы для разных категорий роботов в зависимости от степени их общественной опасности, контролируемости или способности к обучению. Во-вторых, в ряде случаев в принципе трудно восстановить фактические обстоятельства причинения вреда. В-третьих, одна и та же ситуация может получить разное решение с точки зрения конкретной юрисдикции. Поэтому национальные особенности конкретной правовой системы часто не позволяют учитывать существующий опыт других стран» [1].

Встает вопрос о возникновении прав роботов по мере дальнейшего совершенствования технологий и размывания границ между человеком и цифровой сущностью. Например, парламент Европейского союза рассматривает вопрос о предоставлении роботам-компаньонам определенных прав, позволяющих им совершать финансовые и некоторые другие операции от лица владельца. Это может привести к тому, что роботы будут привлекаться к ответственности за повреждение имущества или травмирование людей с вполне реальными последствиями, а хозяева роботов – частные лица, корпорации или правительства – будут ограждены от любой ответственности за это.

При обсуждении высказываются различные подходы к принципам установления ответственности за действия ИИ, в том числе:

- *полное освобождение* кого-либо от ответственности за действия ИИ (по аналогии с обстоятельствами непреодолимой силы);
- *частичное освобождение* от ответственности (освобождение конкретного лица от какой-либо ответственности и одновременная выплата пострадавшим компенсации вреда из различных источников);
- *ответственность по вине*, наступающая в зависимости от вины конкретного субъекта (например, производителя, разработчика, лица, ответственного за обучение ИИ, владельца, пользователя и т. д.);
- *безвиновная ответственность* (определенное лицо по общему правилу считается ответственным за любые действия системы ИИ);
- *личная ответственность* роботов при условии наделения роботов правосубъектностью (правами и обязанностями, статусом электронной личности).

Сегодня у нас есть стандарты, инструкции и договоренности, регулирующие развитие биотехнологий и биоинженерии (например, инструкции, принятые на конференции Asilomar по рекомбинаторной ДНК в 1975 году) [16]. У нас также есть договоры о нераспространении ядерного оружия. Но у нас еще нет похожих договоренностей в сфере данных и ИИ, которые являются «нефтью» цифровой эпохи. Несмотря на то что данные уже стали самым мощным двигателем экономики, мы еще не подписали глобальный договор, в котором бы оговаривалось, что можно, а что нельзя делать с личными

данными нескольких миллиардов интернет-пользователей [9]. Договоренностей по когнитивным вычислениям или ИИ также не существует.

Интерес представляет проект НПА [8] о регулировании беспилотного транспорта от АНО «ПравоРоботов», решающий такие вопросы:

– *технические*: разрешить тестирование беспилотных автомобилей на дорогах общего пользования;

– *социальные*: сохранить рабочее место водителя (ответственного лица) за рулем автомобиля;

– *экономические*: выстроить в сжатые сроки с нуля новый рынок с большим потенциалом;

– *правовые*: соблюсти статус-кво между действующим законодательством и взятыми международными обязательствами по ратифицированным международно-правовым актам (вопрос безопасности и разграничения ответственности).

Чтобы обеспечить надежное будущее, мы должны инвестировать не только в развитие технологий, но и в самих себя. Необходимо приложить значительные усилия для формулировки принципов нового гуманизма (философии цифровой эпохи), который, несмотря на наши несовершенства и недостатки, поможет сохранить наш мир пригодным для жизни. Недостаточно будет инвестировать в технологии, когда мы во всём полагаемся на машины и даже не знаем, как они работают. Герд Леонгард в своей книге [5] говорит о том, что экспоненциальный неконтролируемый рост в областях искусственного интеллекта, робототехники, биоинженерии и генетики приведет к систематическому пренебрежению фундаментальными принципами человеческого существования, поскольку у техники нет этики. А общество без этики обречено.

Человечество находится на критически важном этапе. Важно постоянно оценивать ситуацию и проявлять большую предусмотрительность в наших действиях, осуществлять строгий надзор за технологиями, которые впоследствии могут получить большую власть над нами, чем мы можем себе представить. Мы не должны позволять свободным рынкам, предпринимателям и военным организациям принимать решения за нас. Будущее человечества не должно строиться в соответствии с устаревшими порядками индустриального общества 1980-х годов, в котором прибыль и рост любой ценой считались высшим благом. К счастью, принцип 90/10 (90 % возможностей, которые открывают нам новые технологии, могут пойти на пользу человечества, и только 10 % оцениваются как негативные и вызывающие беспокойство) соблюдается до сих пор.

Герд Леонгард в своей книге [5] привел несколько технологических ошибок, которых нужно избегать человечеству, чтобы сохранить этику в ее первоначальном виде. Вот некоторые из них.

«Мы не будем требовать или планировать, чтобы люди постепенно сами становились технологиями, только из-за того, что это будет выгодно для технологий или технологических компаний». Эта ошибка, связана, в частности, с вопросами гуманного применения дронов в гражданском секторе, анонимности пользователей и передачи личных данных другим компаниям.

«Мы не позволим ботам, машинам, платформам и другим технологиям взять на себя основные демократические функции в нашем обществе, которые все-таки должны выполнять люди». Речь идет о принятии решений о свободе или даже жизни человека. Автономный аппарат не должен самостоятельно принимать такие решения, как, например, принимать по внутренним алгоритмам решение об ограничении свободы или самостоятельно прибегать к запуску смертоносного оружия при обнаружении объекта.

«Мы не будем создавать, проектировать или распространять какие-либо технологии с главной целью вызвать зависимость от них». Конечно, этот пункт в наше время активно нарушают, в особенности в игровой индустрии, используя алгоритмы, заставляющие пользователя играть как можно больше. Уже сейчас «Яндекс», ВК и другие платформы изучают наши пристрастия и интересы, предлагая нам всё больше и больше интересного нам контента, музыки, чтобы мы как можно реже выходили из сети.

Становится ясно, что технологический детерминизм – не лучшее решение. Если бы наше поведение и действия были заранее предопределены неким законом, то любой мог бы оправдать свое действие или бездействие. Но в мире есть и свобода. И к этой категории стоит отнести понятия этики: выбор человеческого поведения, а не оптимального пути; выбор всемирного благополучия, а не личной выгоды.

В книге «Технологии против человека» [5] автор приводит замечательную фразу: «То, что мы можем что-то сделать, не значит, что мы должны это делать». Действительно, в нашей быстроразвивающейся эпохе стоит обратить внимание на то, какое влияние технологии могут оказать на развитие человечества. Сама возможность воплотить идею в жизнь еще не означает, что нам стоит это делать. Решения, принимаемые относительно научных или технологических исследований, их финансирования и регулирования, должны приниматься в интересах счастья и благополучия всего человечества. Процесс познания осуществляется с помощью технологий, а не ради них.

Для безопасного и благоприятного будущего не нужно сдерживать или оттягивать инновации, но вместе с тем необходимо считаться с экспоненциально растущими рисками и не перекладывать ответственность на других. Предпринимательский дух в нас не должен заглушать заботу о процветании общества.

Литература

1. Сбербанк. Аналитический обзор мирового рынка робототехники 2019. – URL: <https://ict.moscow/research/analiticheskii-obzor-mirovogo-rynka-robototekhniki-2019/> (дата обращения: 23.05.2023).
2. Автономная некоммерческая организация по содействию развитию робототехники и инновационных технологий «ПравоРоботов»: сайт. – URL: <https://pravorobotov.ru/> (дата обращения: 23.05.2023).
3. В Китае Tesla потеряла управление и разогналась до максимальной скорости. – 2022, 13 ноября. – URL: <https://electrek.co/2022/11/13/tesla-china-responds-to-dramatic-crash-that-kills-two-video/> (дата обращения: 23.05.2023).
4. Левин В. К власти разума над цифровой экономикой // Цифровая экономика. – 2019, 7 февраля. – URL: <http://digital-economy.ru/mneniya/politiko-ekonomicheskie-voprosy-tsifrovoj-ekonomiki-1> (дата обращения: 23.05.2023).
5. Леонгард Г. Технологии против человека / пер. с англ. А.О. Юркова, М.Ю. Килина, Т.Ю. Глазкова; предисл. М. Федорова – М.: АСТ, 2018. – 320 с.
6. Маск назвал искусственный интеллект главной угрозой человечеству // Информационное агентство ТАСС. – 2023, 15 февраля. – URL: <https://tass.ru/ekonomika/17051707> (дата обращения: 23.05.2023).
7. Паскуале Ф. Новые законы робототехники: апология человеческих знаний в эпоху искусственного интеллекта. – М.: Дело, 2022. – 446 с.
8. Распоряжение Правительства РФ от 25 марта 2020 г. № 724-р «О Концепции обеспечения безопасности дорожного движения с участием беспилотных транспортных средств на автомобильных дорогах общего пользования». – URL: <https://www.garant.ru/products/ipo/prime/doc/73707148/> (дата обращения: 23.05.2023).
9. Asilomar Conference on Recombinant DNA. – URL: https://en.wikipedia.org/wiki/Asilomar_Conference_on_Recombinant_DNA (accessed: 23.05.2023).
10. Autonomous Treatment of Bacterial Infections in Vivo Using Antimicrobial Micro- and Nanomotors / X. Arqué, M.D.T. Torres, T. Patiño, A. Boaro, S. Sánchez, C. de la Fuente-Nunez // ACS Nano. – 2022. – Vol. 16 (5). – P. 7547–7558. – DOI: 10.1021/acsnano.1c11013.
11. Bostrom N. Superintelligence: Paths, Dangers, Strategies. – Oxford: Oxford University Press, 2014. – 328 p.
12. Drone issue may give rise to Pak-US conflict: Nisar // Geo News. – 2013, July 04. – URL: <https://www.geo.tv/latest/61112-drone-issue-may-give-rise-to-pak-us-conflict-nisar> (accessed: 24.05.2023).
13. Glaser J. Former US Official: For Every Yemen Terrorist US Drones Kill, 40–60 New Enemies Are Created // Anti War Blog. – 2013, October 24. – URL: <https://www.antiwar.com/blog/2013/10/24/former-us-official-for-every-yemen-terrorist-us-drones-kill-40-60-new-enemies-are-created/> (accessed: 24.05.2023).
14. Frier S. Facebook Paid Contractors to Transcribe Users' Audio Chats // Bloomberg. Technology. – 2019, August 14. – URL: <https://www.bloomberg.com/news/>

articles/2019-08-13/facebook-paid-hundreds-of-contractors-to-transcribe-users-audio (accessed: 24.05.2023).

15. Killer AI robots must be outlawed, says UN chief / Written by Steve Ranger, Global News Director. – 2018, November 6. – URL: <https://www.zdnet.com/article/killer-ai-robots-must-be-outlawed-says-un-chief/> (accessed: 24.05.2023).

16. *Leonhard G.* Automation, Machine Thinking and Unintended Consequences. – 2015, July 30. – Retrieved 3 July, 2016. – URL: https://youtu.be/Gq8_xPjlsQ (accessed: 24.05.2023).

17. More than a third fear drone use in U.S.: Poll // CBS NEWS. – 2012, September 27. – URL: <https://www.cbsnews.com/news/more-than-a-third-fear-drone-use-in-us-poll/> (accessed: 24.05.2023).

18. Pakistani Public Opinion Ever More Critical of U.S. / Pew Research Center. Global Attitudes Project. – 2012, June 27. – URL: <http://www.pewglobal.org/2012/06/27/pakistani-public-opinion-ever-more-critical-of-u-s> (accessed: 24.05.2023).

19. *Smith D.* Microphone Gate: The world's biggest tech companies were caught sending sensitive audio from customers to human contractors. Here's where they stand now // Business Insider. – 2019, 22 August. – URL: <https://www.businessinsider.nl/amazon-apple-google-microsoft-assistants-sent-audio-contractors-2019-8?international=true&r=US> (accessed: 24.05.2023).

20. *Stuart R.* Slaughterbots are Here. The era in which algorithms decide who lives and who dies is upon us. We must act now to prohibit these weapons. – URL: <https://autonomousweapons.org> (accessed: 20.09.2022).

21. *Subbaraman N.* Don't call 'em drones: The wide world of unmanned flying machines // NBC News. Technology. – 2013, March 15. – URL: <https://www.nbcnews.com/technology/dont-call-em-drones-wide-world-unmanned-flying-machines-1c8857699> (accessed: 24.05.2023).

22. Tesla потребовала удалить видео, на которых ее электромобиль на автопилоте сбивает манекенов в виде детей. – URL: <https://dtf.ru/life/1327652-tesla-potrebovala-udalit-video-na-kotoryh-ee-elektromobil-na-avtopilote-sbivaet-manekenov-v-vide-detey> (дата обращения: 24.05.2023).

23. *Masood S., Mehsud I.T.* Thousands in Pakistan Protest American Drone Strikes // The New York Times. – 2013, November 23. – URL: <https://www.nytimes.com/2013/11/24/world/asia/in-pakistan-rally-protests-drone-strikes.html> (accessed: 24.05.2023).

24. *Byman D.L.* Why Drones Work: The Case for Washington's Weapon of Choice // Brookings. – 2013, June 17. – URL: <https://www.brookings.edu/articles/why-drones-work-the-case-for-washingtons-weapon-of-choice/> (accessed: 24.05.2023).

25. World Economic Forum. Digital Transformation of Industries. – 2016. – URL: <https://www.sbs.com.au/ondemand/news-series/dateline/dateline-2016/koreas-internet-addicts-s2016-ep9/654468675848> (accessed: 03.12.2022).

Статья поступила в редакцию 07.03.2023.

Статья прошла рецензирование 29.03.2023.

DOI: 10.17212/2075-0862-2023-15.2.2-346-364

ETHICAL RISKS IN THE PRACTICE OF ARTIFICIAL INTELLIGENCE

Kazaryan, Valentina,

Dr. of Sc. (Philosophy), Professor,

Professor at the Department of Philosophy of the Faculty of Natural Sciences,

Lomonosov Moscow State University,

1 Leninskije Gory, Moscow, 119991, Russian Federation

ORCID: 0000-0001-9435-8625

kazaryanvp@mail.ru

Shutova, Ksenia,

Postgraduate student,

Mathematician of the first category, Department of Nonlinear Dynamic Systems and Control Processes,

Lomonosov Moscow State University,

1 Leninskije Gory, Moscow, 119991, Russian Federation

ORCID: 0000-0002-4915-0312

shutova.k.u@yandex.ru

Abstract

The article raises the question of the relationship between the practice of artificial intelligence and universal ethics. The topic is fundamentally important, since a person and society that have lost their ethical foundations are deprived of their humanity. The ethics of scientists, designers, high-level managers play a decisive role in modern processes of development and application of artificial intelligence. In their activities, an ethic of responsibility develops, originating from the Russell-Einstein Manifesto, published in 1955 in the conditions of a nuclear standoff during the Cold War. The article shows the modern indescribably rapid growth and development of artificial intelligence applications in the practical life of people. Attention is drawn to the fact of the presence of uncertainty in situations of practical application of artificial intelligence, the presence of unforeseen consequences in addition to the expected consequences, who is responsible for those consequences (individuals, corporations, governments). The responsibility lies with the one who decides on the action: the actor. The actor is in a situation of ethical risk. It is shown that the risk increases due to a number of circumstances: 1) a variety of applications; 2) uncontrolled rampant growth; 3) difficulties in tracking the empirical situation of application; 4) difficulties in theoretical analysis of the situation of action. The article focuses on the risks of the practice of peaceful use of remotely piloted aircraft, as well as their military use and automated weapons without operator confirmation. The sharp, apparently exponential growth of information technology, the practical implementation of artificial intelligence, puts people in a difficult situation, an ethical situation: what to choose 'to have or to be'. Or is there a third choice?

Keywords: ethics of responsibility, artificial intelligence, actor, remotely piloted aircraft, autonomous system.

Bibliographic description for citation:

Kazaryan V., Shutova K. Ethical Risks in the Practice of Artificial Intelligence. *Idei i idealy = Ideas and Ideals*, 2023, vol. 15, iss. 2, pt. 2, pp. 346–364. DOI: 10.17212/2075-0862-2023-15.2.2-346-364.

References

1. Sberbank. *Analiticheskii obzor mirovogo rynka robototekhniki 2019* [Analytical review of the global robotics market 2019]. Available at: <https://ict.moscow/research/analiticheskii-obzor-mirovogo-rynka-robototekhniki-2019/> (accessed 23.05.2023).
2. *Avtonomnaya nekommercheskaya organizatsiya po sodeistviyu razvitiyu robototekhniki i innovatsionnykh tekhnologii «PravoRobotov»* [ANO to promote the development of robotics and innovative technologies “PravoRobotov”]. Website. Available at: <https://pravorobotov.ru/> (accessed 23.05.2023).
3. *Tesla China responds to dramatic crash that kills two*. 2022, 13 November. Available at: <https://electrek.co/2022/11/13/tesla-china-responds-to-dramatic-crash-that-kills-two-video/> (accessed 23.05.2023).
4. Levin V. K vlasti razuma nad tsifrovoy ekonomikoi [Towards the power of reason over the digital economy]. *Tsifrovaya ekonomika* [Digital economy], 2019, 7 February. Available at: <http://digital-economy.ru/mneniya/politiko-ekonomicheskie-voprosy-tsifrovoj-ekonomiki-1> (accessed 23.05.2023).
5. Leonhard G. *Tekhnologii protiv cheloveka* [Technologies against man]. Moscow, AST Publ., 2018. 320 p. (In Russian).
6. Musk nazval iskusstvennyi intellekt glavnoi ugrozoi chelovechestvu [Musk called artificial intelligence the main threat to humanity]. *Informatsionnoe agentstvo TASS* [TASS, Russian news agency], 2023, 15 February. (In Russian). Available at: <https://tass.ru/ekonomika/17051707> (accessed 23.05.2023).
7. Pasquale F. *New laws of robotics : defending human expertise in the age of AI*. Cambridge, MA, The Belknap Press of Harvard University Press, 2020 (Russ. ed.: Paskuale F. *Novye zakony robototekhniki: apologiya chelovecheskikh znanii v epokhu iskusstvennogo intellekta*. Moscow, Delo Publ., 2022. 446 p.).
8. Order of the Government of the Russian Federation dated March 25, 2020 N 724-r “On the Concept of ensuring road safety with the participation of unmanned vehicles on public roads”. (In Russian). Available at: <https://www.garant.ru/products/ipo/prime/doc/73707148/> (accessed 23.05.2023).
9. Asilomar Conference on Recombinant DNA. Available at: https://en.wikipedia.org/wiki/Asilomar_Conference_on_Recombinant_DNA (accessed 23.05.2023).
10. Arqué X., Torres M.D.T., Patiño T., Boaro A., Sánchez S., de la Fuente-Nunez C. Autonomous Treatment of Bacterial Infections in Vivo Using Antimicrobial Micro- and Nanomotors. *ACS Nano*, 2022, vol. 16 (5), pp. 7547–7558. DOI: 10.1021/acsnano.1c11013.
11. Bostrom N. *Superintelligence: Paths, Dangers, Strategies*. Oxford, Oxford University-Press, 2014. 328 p.

12. Drone issue may give rise to Pak-US conflict: Nisar. *Geo News*, 2013, July 04. Available at: <https://www.geo.tv/latest/61112-drone-issue-may-give-rise-to-pak-us-conflict-nisar> (accessed: 24.05.2023).
13. Glaser J. Former US Official: For Every Yemen Terrorist US Drones Kill, 40–60 New Enemies Are Created. *Anti War Blog*, 2013, October 24. Available at: <https://www.antiwar.com/blog/2013/10/24/former-us-official-for-every-yemen-terrorist-us-drones-kill-40-60-new-enemies-are-created/> (accessed 24.05.2023).
14. Frier S. Facebook Paid Contractors to Transcribe Users' Audio Chats. *Bloomberg. Technology*, 2019, August 14. Available at: <https://www.bloomberg.com/news/articles/2019-08-13/facebook-paid-hundreds-of-contractors-to-transcribe-users-audio> (accessed 24.05.2023).
15. Killer AI robots must be outlawed, says UN chief. Written by Steve Ranger, Global News Director, 2018, November 6. Available at: <https://www.zdnet.com/article/killer-ai-robots-must-be-outlawed-says-un-chief/> (accessed 24.05.2023).
16. Leonhard G. *Automation, Machine Thinking and Unintended Consequences*. 2015, July 30. Retrieved 3 July, 2016. Available at: https://youtu.be/Gq8_xPjlsQ (accessed 24.05.2023).
17. More than a third fear drone use in U.S.: Poll. *CBS NEWS*, 2012, September 27. Available at: <https://www.cbsnews.com/news/more-than-a-third-fear-drone-use-in-us-poll/> (accessed 24.05.2023).
18. Pakistani Public Opinion Ever More Critical of U.S. *Pew Research Center: Global Attitudes Project*, 2012, June 27. Available at: <http://www.pewglobal.org/2012/06/27/pakistani-public-opinion-ever-more-critical-of-u-s> (accessed 24.05.2023).
19. Smith D. Microphone Gate: The world's biggest tech companies were caught sending sensitive audio from customers to human contractors. Here's where they stand now. *Business Insider*, 2019, 22 August. Available at: <https://www.businessinsider.nl/amazon-apple-google-microsoft-assistants-sent-audio-contractors-2019-8?international=true&r=US> (accessed 24.05.2023).
20. Russell S. *Slaughterbots are Here. The era in which algorithms decide who lives and who dies is upon us. We must act now to prohibit these weapons*. Available at: <https://autonomousweapons.org> (accessed 20.09.2022).
21. Subbaraman N. Don't call 'em drones: The wide world of unmanned flying machines. *NBC News. Technology*, 2013, March 15. Available at: <https://www.nbcnews.com/technology/dont-call-em-drones-wide-world-unmanned-flying-machines-1c8857699> (accessed 24.05.2023).
22. Tesla demanded to remove videos of its electric car on autopilot hitting dummies as children. (In Russian). Available at: <https://dtf.ru/life/1327652-tesla-potrebovala-udalit-video-na-kotoryh-ee-elektromobil-na-avtopilote-sbivaet-manekenov-v-vide-detey> (accessed 24.05.2023).
23. Masood S., Mehsud I.T. Thousands in Pakistan Protest American Drone Strikes. *The New York Times*, 2013, November 23. Available at: <https://www.nytimes.com>

com/2013/11/24/world/asia/in-pakistan-rally-protests-drone-strikes.html (accessed 24.05.2023).

24. Byman D.L. Why Drones Work: The Case for Washington's Weapon of Choice. *Brookings*, 2013, June 17. Available at: <https://www.brookings.edu/articles/why-drones-work-the-case-for-washingtons-weapon-of-choice/> (accessed 24.05.2023).

25. World Economic Forum. Digital Transformation of Industries. 2016. Available at: <https://www.sbs.com.au/ondemand/news-series/dateline/dateline-2016/koreas-internet-addicts-s2016-ep9/654468675848> (accessed 03.12.2022).

The article was received on 07.03.2023.

The article was reviewed on 29.03.2023.