DOI: 10.30570/2078-5089-2020-99-4-90-107

Д.А.Томильцева, А.С.Железнов

НЕИЗБЕЖНЫЙ ТРЕТИЙ: ЭТИКО-ПОЛИТИЧЕСКИЕ АСПЕКТЫ ВЗАИМОДЕЙСТВИЙ С ИСКУССТВЕННЫМИ АГЕНТАМИ[1]

[1] Статья подготовлена при поддержке Совета по грантам Президента РФ (МК-1740.2019.6).

Дарья Алексеевна Томильцева — кандидат философских наук, доцент кафедры социальной философии департамента философии Уральского федерального университета им. Б.Н.Ельцина (Екатеринбург). Для связи с автором: tomiltceva_d@mail.ru.

Андрей Сергеевич Железнов — кандидат философских наук, независимый исследователь (Москва). Для связи с автором: itsnomoredancing@gmail.com.

Аннотация. Искусственные агенты, то есть созданные людьми технические устройства и программные средства, которые способны целенаправленно действовать и самостоятельно принимать решения, пронизывают сегодня практически все сферы жизни человека. Как новые политические актанты они трансформируют характер человеческих взаимодействий, что порождает проблему этико-политического регулирования их деятельности. Поэтому появление подобных агентов запускает глобальную философскую рефлексию, выходящую за рамки технических или прикладных вопросов и возвращающую исследователей к фундаментальным проблемам этики.

В статье фиксируются три основных аспекта, в которых существование искусственных агентов нуждается в философском осмыслении. Во-первых, искусственные агенты выявляют фактическое противоречие между декларируемыми моральными и политическими ценностями и реальными социальными практиками. Обучаясь на данных об уже состоявшихся оценках и выводах, искусственные агенты выдают решения, отвечающие не моральным принципам их создателей или потребителей, а сложившимся поведенческим паттернам. Во-вторых, особенности создания и функционирования искусственных агентов актуализируют проблему ответственности за их действия, что, в свою очередь, требует нового подхода к политическому регулированию деятельности не только разработчиков, заказчиков и пользователей, но и самих агентов. В-третьих, те формы, которые в настоящее время принимает активность искусственных агентов, смещают традиционные границы человеческого и ставят вопрос

о переопределении гуманитарного. Тщательно проанализировав выделенные аспекты, авторы раскрывают их логику и очерчивают поле для дальнейшей дискуссии.

Ключевые слова: искусственный агент, искусственный интеллект, мораль, предвзятость, ответственность, политика

Введение

Вопреки устоявшемуся мнению, сегодня политика, в том числе публичная, все больше пронизывается вопросами этики, и не только в контексте бинарного оценивания (речей, действий, решений) по принципу «хорошо/плохо», но и через пересмотр самой возможности действовать, а также переопределение политических актантов. Если до недавнего времени подобный ракурс означал лишь наделение возможностью говорить и действовать различные ранее ущемляемые или попросту «невидимые» человеческие сообщества[2], то в последние десятилетия в качестве новых политических актантов начали выступать нечеловеческие сущности, как естественные (например, вирусы), так и создаваемые людьми (роботы, виртуальные ассистенты, автономные автомобили и беспилотные летательные аппараты, рекомендательные и скоринговые системы и т.д.), то есть искусственные агенты.

[2] Об актуальности данного подхода свидетельствует, в частности, активизация в США и некоторых европейских странах движения Black Lives Matter.

В рамках интересующей нас проблемы понятие «искусственный агент» представляется наиболее подходящим, так как оно, во-первых, предполагает, что технические устройства и программные средства достигли определенного равенства с человеком, приобретя способность к самостоятельному действию — «агентность», а во-вторых, подчеркивает их «искусственность»: созданные людьми, они не даны как естественный факт, их появление и устройство не может быть списано на природные процессы. Кроме того, в отличие от микробов, пробок или, скажем, высокочастотного диапазона, искусственные агенты оказывают влияние не случайно, в результате стечения обстоятельств или побочных эффектов, а целенаправленно, вследствие осознанного решения их создателя-человека. Важно отметить, что тема искусственных агентов тесно связана с темой искусственного интеллекта, наделенного способностью самостоятельно действовать и принимать решения. Поэтому далее мы будем иногда употреблять термин «искусственный интеллект» (ИИ), подразумевая, что сказанное о нем имеет непосредственное отношение к искусственным агентам.

Итак, если благодаря развитию ряда социальных и философских теорий (новый материализм, акторно-сетевая теория и др.) «привычные» нечеловеческие актанты типа микробов или атомной энергии стали тем, с чьим автономным существованием мы вынуждены считаться и кому, как в случае с животными, назначаем человеческих представителей, то с искусственными агентами дело обстоит сложнее. С одной стороны, они задумываются и создаются как посредники или

помощники, значительно облегчающие те практические задачи, которые людям приходится ежедневно решать. С другой стороны, они кардинальным образом меняют наши представления о политических практиках и, более того, об этических основаниях этих практик, принимая собственные решения, не поддающиеся человеческой логике, и регулируя нашу активность.

Проблемы, с которыми связано осмысление искусственных агентов в социальной и политической теории, можно описать с точки зрения двух взаимосвязанных процессов — гуманитаризации технического и технизации гуманитарного. Первый процесс означает, что, когда дело касается людей, их взаимоотношений, а также взаимодействий со значимыми для них объектами, проблемы, еще недавно казавшиеся сугубо техническими (например, «как рассчитать?»), приобретают гуманитарные смыслы и становятся не менее, а, быть может, даже более сложными для решения. Второй же предполагает, что задачи, прежде представлявшиеся сугубо гуманитарными и исключительно человеческими (например, «как сегментировать аудиторию и воздействовать на те или иные целевые группы?»), мы с легкостью делегируем техническим системам. В результате наша повседневность наполняется всевозможными конфликтами между гуманитарным и техническим, возникающими из-за недостатков одного и чрезмерной «эффективности» другого.

Правила для искусственных агентов

За последние 20 лет мы уже привыкли к роботам, заменяющим домашних питомцев, алгоритмам, сочиняющим музыку и пишущим картины, беспилотникам, осуществляющим разведку, и многому другому, принимая как само собой разумеющееся, что искусственные агенты превосходно выполняют те задачи, которые ранее считались прерогативой человека. В нашу практику активно входят виртуальные ассистенты и планировщики, искусственные рекомендательные системы и системы оценки. Более того, речь идет не просто о неких орудиях, связывающих пользователя и заказчика. Теперь об искусственных агентах говорят как о «хороших» или «кривых», «полезных» или «опасных», сетуют, что некоторые компании подменяют ими живое общение, или радуются, что оно сведено к минимуму... Словом, искусственный агент оказывается деперсонализированным, невидимым, но при этом антропоморфизированным[3] актантом, который заставляет нас принимать его всерьез и согласовывать с ним свои решения и действия[4]. В каком-то смысле он напоминает идеального веберовского бюрократа, который, действуя строго по инструкции, не имеет личностных черт и потому предельно беспристрастен.

Но чем интенсивнее человеческая повседневность оцифровывается и технизируется, тем более очевидной становится необходимость разработки специфических мер морально-правового регулирования подобных технологий и систем. Традиционные кодексы, правила

[3] В данном случае мы имеем в виду свойственный нам способ восприятия алгоритма, а не принципы его работы. О несостоятельности антропоморфного объяснения принципа действия искусственного интеллекта см., в частности, Ученый 2016.

[4] См. Керимов 2019.

и предписания здесь заведомо не годятся[5] — созданные людьми и для людей, они не учитывают, что современные искусственные агенты проявляют инициативу и сами принимают решения. Поскольку же люди пока не научились исходить из какой-либо отличной от собственной логики, чрезвычайно важно, чтобы искусственные агенты применяли к своим действиям моральные ограничения, сходные с человеческими. Несмотря на глобальную значимость данной проблемы, адекватных путей ее решения пока не найдено.

[5] Подробнее см., напр. Кашкин и Покровский 2019. В России проблема правового регулирования ИИ стоит очень остро, законодательная база как таковая отсутствует. Заполнить этот пробел призвана утвержденная в 2020 г. Концепция развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники, разработанная Минэкономразвития (см. Рабочая группа 2020 ; Концепция 2020).

На первый взгляд может показаться, что разработчикам достаточно вложить во вроде бы нейтральный искусственный агент собственные или доминирующие в данном сообществе представления о человеке, социальных отношениях и политических доминантах — и тогда получающиеся на выходе технологии как бы скопируют человеческую логику (что, впрочем, лишит их способности быть полностью беспристрастными при принятии решений). Однако сами представления о человеке постоянно меняются — во многом под воздействием технологий. Как справедливо замечает Тобиас Рис, «подавляющее большинство передовых исследований Искусственного Интеллекта проводится в компаниях. Проблема в том, что большинство людей, возглавляющих эти компании, не осознают... что они кардинально переосмысляют наше определение того, что означает „быть человеком“. Они считают себя всего лишь людьми, работающими в технологических компаниях»[6]. Сложность состоит еще и в том, что модель «перенесения» не отвечает специфике машинного обучения как основного на сегодняшний день способа создания ИИ.

[6] Рис 2020.

В общих чертах машинное обучение предполагает, что, предварительно обработав некий массив данных об уже состоявшихся оценках и выводах, программное обеспечение самостоятельно учится анализировать и формировать собственные принципы оценки, не всегда совпадающие с теми, которых ждут от него разработчики. Так, для того чтобы научить ИИ узнавать на картинках кошек, в него нужно загрузить большую базу изображений, часть которых помечена как кошки. Но если ошибки системы, идентифицирующей как кошку не являющееся ею животное, не критичны, то в отношении людей подобные ошибки чреваты довольно серьезными этико-политическими последствиями[7]. По мере того как искусственные агенты все активнее вовлекаются в социальные практики, связанные с идентификацией человека по внешнему виду (скажем, в системах безопасности), растет и число ситуаций, в которых «очень сложно заставить ПО одинаково функционировать с людьми, принадлежащими к разным этническим группам»[8].

[7] См., напр. Робота-редактора 2020.

[8] Самойдюк 2018.

Аналогичным образом искусственный агент действует и в более сложных случаях, например при создании всевозможных рейтинговых систем, генерирующих классификации и различные формы оценивания на основе анализа поведения и предпочтений значительного числа людей, а также их социальных контактов. Подобные системы, базирующиеся на машинном обучении, широко применяются в рекрутинге,

банковском секторе, системах страхования, правоохранительных или налоговых органах.

Для того чтобы определить, какой из претендентов может быть допущен до собеседования или выдачи кредита, нужно загрузить базу резюме, которые ранее обрабатывали люди. После этого помощь человека уже не требуется — искусственный агент не копирует устоявшиеся практики, а формирует собственную логику, неизвестную его создателям, поскольку не обладает четко прописанным и подконтрольным человеку «алгоритмом» действия. В результате использующий его человек далеко не всегда способен объяснить причины, по которым были приняты те или иные решения[9]. И нередко такие решения оказываются неудовлетворительными. Дело вовсе не в том, что собранные данные носят «фейковый» характер, ведь сознательно хитрить искусственные агенты, по крайней мере — пока, не умеют. Парадоксальным образом данные, собираемые и анализируемые далекими от человеческих распрей системами, демонстрируют реальную предвзятость, присущую человеческим суждениям.

[9] *Cappelli 2019.*

Искусственный агент не беспристрастен

Искусственный агент не беспристрастен. Истоки его предвзятости (bias) кроются в том, что, обучаясь на примерах фактически совершенных человеческих действий, он выдает решения, отвечающие не моральным принципам его создателей или потребителей, а реальной практике.

Наглядной иллюстрацией подобной предвзятости служит широко обсуждавшаяся в 2018 г. история с алгоритмом найма персонала, тестировавшимся компанией Amazon[10]. Будучи обучен на основе резюме, поступавших в компанию в предшествующие годы, «алгоритм начал отбраковывать заявки со словом „женщины“. Программу отредактировали так, чтобы искусственный интеллект не маркировал это слово и его производные как нечто негативное, но это не очень помогло»[11]. По сути, внезапно обнаружившаяся предвзятость алгоритма найма обнажила противоречие между официально провозглашаемой политикой гендерного равенства и реальными практиками дискриминации.

[10] *Dastin 2018.*

[11] *Искусственный интеллект 2019.*

С аналогичной проблемой столкнулась и система оценки рисков задержания граждан правоохранительными органами в США. С одной стороны, исторически сложилось так, что чернокожие граждане чаще подвергаются аресту. С другой стороны, известно, что это происходит в том числе из-за предвзятости правоохранительных органов. Учет системой оценки только первого факта при игнорировании второго (который оказывает на него влияние) породил ситуацию, когда оценка рисков производилась одновременно и корректно, и некорректно[12]. Иными словами, система, показывающая, что чернокожие граждане с большей вероятностью будут арестованы, делала прогноз, отвечающий существующей реальности, при том что именно эту реальность создатели системы хотели изменить.

[12] *Hao 2019.*

Итак, искусственные агенты не лишены предвзятости. Их разработка в соответствии с доминирующими ценностными установками, накопленным опытом социальных практик и личностными моральными предпочтениями скорее кодирует и закрепляет текущую этико-политическую ситуацию. Причем чем более отстраненным от принятия решений оказывается человек, тем вероятнее, что искусственный агент будет порождать новые формы несправедливости и неравенства, рационализированного и беспристрастного, то есть лишенного сознательного стремления к дискриминации, но воспроизводящего и даже интенсифицирующего ее[13].

[13] *Харари 2020.*

Следуя этой логике, правомерно предположить, что возможные векторы дискриминации зависят от эпохи, культурного контекста, религии и т.д. Так, если сегодня речь не идет о сословной дискриминации, то именно потому, что само представление о сословиях для нас неактуально, тогда как учет различных способов конструирования сообществ и фиксация принадлежности к ним происходят повсеместно. Иначе говоря, хотя в искусственные агенты закладываются множественные варианты анализа качеств человека, оценки выносятся исходя из заранее предустановленных критериев нормальности и/или идеала, впоследствии уточняемых в ходе функционирования искусственных агентов.

Впрочем, предвзятость — лишь один из аспектов конфликта гуманитарного и технического применительно к искусственным агентам. Проблема не только и не столько в том, что искусственный агент «учится плохому» и делает некорректные выводы, сколько в том, что изначальное решение, чему и как учить искусственный агент, принимает его создатель. А моральные представления создателей могут в корне отличаться друг от друга и приводить к новым, уже сугубо человеческим ценностным конфликтам. Эту проблему прекрасно раскрывает проект Массачусетского технологического института «Моральная машина»[14], в рамках которого осуществлялся масштабный сбор мнений о том, как должен повести себя автономный автомобиль в ситуации, когда невозможно избежать жертв. Благодаря этому проекту мы знаем, что в одних странах предпочли бы, чтобы автономный автомобиль наехал на молодую девушку, а не на пожилую женщину, тогда как в других — наоборот.

[14] *Awad et al. 2018.*

Указанное расхождение отнюдь не означает ни превосходства одной морали над другой, ни даже необходимости учета при проектировании автономного автомобиля культурно-зависимых логик поведения. Речь идет о другом — о том, что имеющаяся на сегодня мораль (а точнее, множество моральных систем) недостаточна для того, чтобы позволить искусственному агенту действовать в соответствии с ней. По этой причине создателю автономного автомобиля (а также самому автомобилю) требуется некая логика, которая превосходила бы фактически существующие локальные моральные установки.

Тем самым проблема взаимодействия людей и искусственных агентов обостряет моральную рефлексию, которая в корне отличается от той, что обычно сопровождает рассуждения об использовании техники,

орудий и т.д. Так, при рассмотрении этических аспектов применения атомного или биологического оружия или обсуждении конфиденциальности личных данных мы ориентируемся на существующие моральные принципы. Но в случае искусственных агентов подобных принципов еще не сформировано, как и правовых и законодательных мер, а для их появления необходимо переосмысление сложившихся моральных координат и разработка некоей формы глобальных моральных установок. Выработка таких принципов — проблема сугубо гуманитарная. Скажем, для ответа на вопросы, что является дискриминацией, а что нет и почему дискриминация — это плохо, нам надо определить, когда и по какой причине мнение одного сообщества людей (например, ученых) относительно дискриминации может быть важнее мнения другого сообщества (например, рекрутеров и полицейских), и при этом избежать «тирании смысла». Следовательно, решение должно опираться на некоторый консенсус между учеными, разработчиками, конечными пользователями и регулирующими органами, то есть для его достижения нужна не только научная, но и глобальная политическая консолидация.

Ответственность и регулирование

Проблема предвзятости, описанная выше, позволяет заострить внимание на одном важном положении: влияние создателя на логику искусственных агентов ограничено. Но если вышедший из-под контроля рекрутинговый алгоритм вряд ли спровоцирует катастрофу, то автономный автомобиль вполне на это способен. Так кто же несет ответственность за неверно принятые искусственными агентами решения?

Ответ на данный вопрос далеко не очевиден, а необходимость найти его становится все более насущной, учитывая, что искусственные агенты широко используются и в военной сфере[15]. Осознание актуальных и возможных рисков приводит к всплеску инициатив по регулированию создания и применения ИИ. Подобного рода инициативы представлены на самых разных институциональных уровнях — наднациональном (например, Специальный комитет по искусственному интеллекту при Совете Европы[16]), национальном (например, Комитет по вопросам этики искусственного интеллекта при Комиссии Российской Федерации по делам ЮНЕСКО[17]), на уровне научно-технического сообщества (например, Глобальная инициатива по этике автономных и интеллектуальных систем[18]), наконец, на уровне корпораций[19].

Потребность в разработке специфического этического кодекса и соответствующей нормативно-правовой базы, дабы самообучающиеся системы не научились «чему-нибудь не тому» и не стали впоследствии руководствоваться этими знаниями при выполнении своих функций[20], не вызывает сомнений. Однако установка на формулирование гуманитарных требований к технологиям сама по себе сопряжена с немалыми сложностями, поскольку она ставит вопрос о выявлении действующего поля нормативности, как этической, так и правовой, что,

[15] *Мартынов 2019. См. также В Москве 2020.*

[16] *https://www.coe.int/en/web/artificial-intelligence/cahai.*

[17] *См. О создании 2020; Емелин 2020.*

[18] *https://ethicsinaction.ieee.org/.*

[19] *См. Pichai 2020.*

[20] *Уэйкфилд 2018.*

в свою очередь, предполагает исследование и (пере)определение человека в свете актуальной гуманистической парадигмы и представлений о социальном.

С одной стороны, здесь в фокусе внимания оказываются границы дозволенного, то есть те горизонты «можно» или «нельзя», в пределах которых людям и технологиям предстоит действовать и выход за которые в лучшем случае повлечет за собой неприятности или неудобства, а в худшем вполне может обернуться катастрофой. С другой стороны, на передний план выдвигаются деонтологические и консеквенциалистские аспекты проблемы: не приведут ли выводы самообучающихся систем и принимаемые на их основе решения к установлению цензуры и насколько морально оправданными являются задачи, решаемые искусственными агентами, и цели, стоящие перед разработчиками, пользователями и/или теми, кто выступает объектом анализа? Должно ли руководство компании нести ответственность за то, что реальное поведение ее сотрудников не отвечает декларируемым ценностям? Или же ее несут программисты, не скорректировавшие вводные данные так, чтобы они соответствовали этим ценностям? Может ли вина лежать на самих искусственных агентах, которые гуманитарно слепы, а значит, не умеют распознавать смыслы? Тогда ответственность использующих их организаций будет состоять в обеспечении надлежащего контроля над результатами их деятельности[21]. Последняя позиция чрезвычайно удобна для компаний — особенно если речь идет о сглаживании конфликтов, вызванных действиями искусственных агентов.

Примером подобной ситуации может служить история с блокировкой алгоритмами социальной сети Facebook постов с фотографией Евгения Халдея «Знамя Победы над Рейхстагом». Извинившись перед пользователями, компания восстановила контент, объяснив его удаление ошибкой в действии «автоматизированных инструментов выявления нарушений». «Мы хотим, чтобы наша платформа оставалась безопасным местом для поддержания связи людей друг с другом... поэтому в данный момент мы больше полагаемся на наши автоматизированные системы обнаружения и удаления контента, который нарушает нормы сообщества, — подчеркнул представитель компании. — Чем меньше людей могут рассматривать публикации, изображения и комментарии, которые потенциально нарушают наши нормы, тем больше вероятность допущения некоторых ошибок»[22].

Искусственные агенты эффективны, однако, и это далеко не новость, они лишены человеческой способности распознавать смыслы[23]. Этические нормы применяющей их компании должны удовлетворять неким наиболее общим, глобальным представлениям о плохом и хорошем, правильном и неправильном и вместе с тем сообразовываться с локальными интерпретациями этих норм, форсируемыми извне идеологическими противоречиями, культурной спецификой и т.д. Что же касается вопроса об ответственности, то применительно к приведенному выше примеру можно констатировать, что, поскольку разобраться

[21] Так, в Концепции развития регулирования отношений в сфере технологий искусственного интеллекта, принятой правительством РФ в 2020, говорится о необходимости дальнейшей проработки «механизмов гражданско-правовой, уголовной и административной ответственности в случае причинения вреда системами искусственного интеллекта и робототехники, имеющими высокую степень автономности, при принятии ими решений, в том числе с точки зрения определения лиц, которые будут нести ответственность за их действия... а также возможности использования способов, позволяющих возместить причиненный действиями систем искусственного интеллекта и робототехники вред» (Концепция 2020).

[22] Facebook 2020.

[23] «Смысловую слепоту» алгоритмов прекрасно иллюстрирует пример с переводом полного собрания сочинений Артура Конан Дойля в компьютерный код, приводимый Александром Кулешовым (см. Развитие 2016: 171).

в содержании сотен тысяч постов в ручном режиме чрезвычайно сложно и в условиях дистанционной работы гораздо надежнее автоматизированная проверка, произошла ошибка, к которой люди напрямую не причастны, но и искусственный интеллект, не будучи дееспособным субъектом, нести вину за нее не может.

В любом случае рассмотренный пример позволяет выявить новое специфическое требование к искусственным агентам — обучение умению считывать смыслы человеческой активности, опираться на наши идеалы, ценности и стремления. Подобного рода «моральная чуткость», то есть способность ИИ распознавать моральный контекст ситуации, и образует ту принципиальную границу, при достижении которой вопрос об ответственности искусственного агента становится возможным и решаемым с нашей, человеческой, точки зрения.

Однако представление о том, что искусственные агенты можно обучить распознавать моральный контекст и проявлять специфическую «моральную чуткость», возвращает нас к основам этической дискуссии, к вопросу о том, что значит быть нравственным. Здесь, конечно же, нет единой позиции. Это позволяет некоторым исследователям делать вывод, что сама идея намерений или намеренных действий необходима для осуждения другого[24]. При подобном угле зрения моральная ответственность оказывается неким репрессивным или дисциплинарным инструментом. Но если так, то должны ли мы готовиться к тому, что искусственный агент будет использовать мораль в качестве одного из факторов влияния и способов манипуляции, или же «моральная чуткость» удержит его от этого?

[24] Clark 2014. Редуцировать моральный аспект к желанию обвинить другого могут, например, исследования в рамках «экспериментальной философии», как это происходит при обсуждении «эффекта Кноба». Подробнее об «эффекте Кноба» см. Knobe 2003.

Как бы то ни было, вопрос об ответственности за действия искусственного агента вновь проблематизирует поле этического. С одной стороны, мы не можем считать создателей всецело ответственными за действия искусственного агента, ведь они не в полной мере управляют процедурой его обучения и не могут оградить от «дурного» влияния реального мира. С другой стороны, нет оснований полагать, что ответственность может нести сам искусственный агент. Чтобы возлагать на него такую ответственность, мы должны признать его равным человеку в определении морального контекста.

Смещение человеческого

Постановка вопроса об ответственности и возможности переноса ее на искусственный агент подводит нас к еще одной значимой теме — смещении границы между человеком и не-человеком. Если еще несколько десятилетий назад мы могли смело говорить о человеческом как о том, что нельзя заменить техникой, то сегодня нетрудно заметить неоднозначность, если не ошибочность такого утверждения. Во-первых, само представление о неподдающемся замещению постоянно меняется и в большинстве случаев сужается благодаря развитию технологий (самим же человеком и разрабатываемых!); во-вторых, все наши технологии, как уже отмечалось, имеют интенцию к антропоморфизации,

а значит, мы творим их по своему образу и подобию. По справедливому замечанию Риса, «лаборатории Искусственного Интеллекта и технологические компании — наши самые влиятельные философские лаборатории. Это колоссальные экспериментальные пространства, внутри которых люди создают новые концепции человека и окружающего нас мира. В таких местах, как Google, Facebook, Microsoft и OpenAI, инженеры разрабатывают радикально новые концепции того, что такое „быть человеком", жить своей жизнью и жить вместе»[25].

[25] *Рис 2020.*

Когда сущность и человека, и искусственного агента становится все менее определенной, а этические требования — все более очевидными и детализированными, в философском плане обнажается классический разрыв между должным и сущим, совершенно чуждый естественным наукам. Узнав, как фактически работает та или иная технология, мы не спешим подстроиться под ее законы, а стараемся сообразовать их с неким признаваемым нами высшим долженствованием и абсолютным ценностным содержанием: «Если действительность не соответствует понятию, тем хуже для действительности»[26]. Собственно, мы ожидаем, что искусственный агент, не знакомый ни с Гегелем, ни с парадоксом Юма, станет следовать именно логике должного. К чему это «следование» может привести, в настоящее время спрогнозировать сложно. Тем не менее искусственные агенты постепенно осваивают не только технические, но и гуманитарные сферы — например, политическую.

[26] *Тимофеева 2017: 91.*

Несколько лет назад глобальное сообщество уже столкнулось с подобным явлением. Речь идет о новых алгоритмах распространения информации через «фабрики троллей»[27] и технологиях персонифицированной рекламы, фактически заменивших и привычную агитацию, и общественную дискуссию. Действия «фабрик троллей» и возможность постановки вопроса о легитимности принятых избирателями решений активно обсуждались в рамках дела о вмешательстве России в американские выборы[28]. Еще раньше разразился скандал вокруг персонифицированной рекламы Cambridge Analytica[29].

[27] *Мартьянов 2016.*

[28] *Baila et al. 2020.*

[29] *Chen and Potenza 2018.*

В контексте интересующей нас проблемы обе эти ситуации наглядно демонстрируют: существует реальная опасность, что коммуникация «человек — искусственный агент» окажется более эффективной, нежели обычная политическая коммуникация «человек—человек», «человек—сообщество» и т.д. Более того, искусственные агенты вполне успешно могут подменить собой политтехнологов, маркетологов и специалистов по PR (по всяком случае, в интернете). Можно даже допустить (и на то есть серьезные основания), что для какой-то части участников политического процесса общение с ботами заменило взаимодействие с настоящими политическими единомышленниками и оппонентами, так как персонализированная информация, подобранная посредством аналитических систем, воспринимается лучше, чем информация, полученная от реального окружения, а возможность вызывать эмоциональный отклик и поддерживать определенный градус сетевой дискуссии создает иллюзию социального взаимодействия.

Собственно, следствием такого допущения является то внимание, которое вызывают сегодня технологии дипфейк (deepfake), позволяющие создавать видео, на которых имитируются действия реально существующих людей[30]. Явный консенсус политиков и технологических компаний относительно необходимости борьбы с дипфейк[31] — это консенсус страха. Страха, основанного на отсутствии адекватных способов обнаружить различие между реальным политиком и сконструированным образом, страха перед тем, что прямое обращение человека-политика к гражданам перестает быть незаменимым. Напротив, его замена может оказаться даже убедительнее оригинала, ведь искусственный агент говорит в точности то, что хочет услышать конкретный избиратель, причем наиболее подходящим для его восприятия способом. И чем более дистанцированы друг от друга люди, чем больше они вовлечены в сетевое взаимодействие, тем действеннее подобное донесение информации.

[30] *См. Касьянова 2020; Vaccari and Chadwick 2020.*

[31] *Knight 2019.*

Иными словами, формируется новый глобальный субъект, который как бы подменяет собой субъектов-политиков и даже те сообщества, от имени которых они говорят. Парадоксальным образом деонтологические интенции действий в этом случае сменяются консеквенциалистскими, поскольку взаимодействие, происходящее не с отдельным лицом, а только с технической системой, нацелено уже не на игру по правилам, но на получение конкретного результата. Индивидуально (таргетированно) конструируемые образы служат лишь инструментами его достижения. Тогда политика, самое человеческое из всех доступных людям видов деятельности (впрочем, этологи с этим утверждением не соглашаются уже давно), уходит с агоры — пространства взаимодействия людей и сообществ, уступая место технике.

* * *

Итак, мы зафиксировали несколько направлений, где появление искусственных агентов реактуализирует фундаментальные философские, политические и этические вопросы. Искусственные агенты пронизывают практически все сферы жизни человека. За предельно короткий срок они прошли путь от инструмента и/или посредника до реального актанта, с чьим существованием нам приходится считаться. Располагаясь на грани человеческого, между антропоморфизированным восприятием и своей внутренней логикой, объяснить которую математикам пока не под силу, искусственные агенты интенсивно меняют как наши представления о нашей собственной сущности, так и основания и смыслы наших практик. По этой причине проблема морали для искусственных агентов становится критической. Однако, решая эту проблему, нельзя упускать из виду, что процессы гуманитаризации технического и технизации гуманитарного заставляют по-новому проблематизировать межчеловеческие отношения, в которых искусственные агенты все более активно (хотя и незаметно) берут на себя роль Третьего, необходимого и неизбежного, накладывая отпечаток на всю человеческую совместность.

Библиография

«В Москве создали Комитет по ИИ при российской комиссии по делам ЮНЕСКО». (2020) // *РИА Наука*, 27.02. URL: https://ria.ru/20200227/1565300001.html. (проверено 17.07.2020).

Емелин К.Н. (2020) «Искусственный интеллект под контролем» // *Вестник Комиссии Российской Федерации по делам ЮНЕСКО*, № 4. URL: http://unesco.ru/news/ai-committee/ (проверено 19.07.2020).

«Искусственный интеллект проявил сексизм и этим поставил крест на приложении для найма от Amazon». (2018) // *BBC*, 10.10. URL: https://www.bbc.com/russian/other-news-45814099 (проверено 10.07.2020).

Касьянова Д. (2020) «Индийский политик использовал технологию дипфейк, чтобы перевести свою речь на разные языки» // *Bird in Flight*, 19.02. URL: https://birdinflight.com/ru/novosti/indijskij-politik-ispolzoval-tehnologiyu-dipfejkov-chtoby-perevesti-svoyu-rech-na-raznye-yazyki.html (проверено 10.08.2020).

Кашкин С.Ю. и А.В.Покровский. (2019) «Искусственный интеллект, робототехника и защита прав человека в Европейском союзе» // *Вестник Университета им. М.Е.Кутафина (МГЮА)*, № 4: 64—90. URL: http://vestnik-msal.ru/articles/article_105096.html?issue=vest-1-2019 (проверено 10.08.2020).

Керимов Т.Х. (2019) «Цифровизация общества: модуляция, время, субъективация» // *Известия Уральского федерального университета. Серия 3: Общественные науки*, т. 14, № 3 (191): 5—17.

Концепция развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники до 2024 года. (2020) URL: http://www.consultant.ru/document/cons_doc_LAW_360681/ (проверено 27.08.2020).

Мартынов К. (2019) «Этика автономных машин: деонтология и военные роботы» // *Логос,* т. 29, № 3: 231—246. URL: http://logosjournal.ru/arch/107/Logos%203-2019_Press-239-254.pdf (проверено 27.08.2020).

Мартьянов Д.С. (2016) «Политический бот как профессия» // *Политическая экспертиза: ПОЛИТЭКС*, № 1: 74—89.

О создании Комитета по вопросам этики искусственного интеллекта при Комиссии Российской Федерации по делам ЮНЕСКО. (2020) URL: https://www.mid.ru/ru/foreign_policy/un/-/asset_publisher/U1StPbE8y3al/content/id/4069630 (проверено 17.07.2020).

Рабочая группа АНО «Цифровая экономика» одобрила разработанный Минэкономразвития проект Концепции развития регулирования в сфере искусственного интеллекта. (2020) URL: https://www.economy.gov.ru/material/news/rabochaya_gruppa_ano_cifrovaya_ekonomika_odobrila_razrabotannyy_minekonomrazvitiya_proekt_koncepcii_razvitiya_regulirovaniya_v_sfere_iskusstvennogo_intellekta.html (проверено 17.07.2020).

«Развитие инженерного образования и формирование современной инженерной культуры в России (Двадцать пятые Губернаторские чтения. Тюмень, 28 июня 2016 г.)». (2016) // *Полития*, № 3: 160—183. URL: http://politeia.ru/files/articles/rus/2016_03_09.pdf (проверено 29.08.2020).

Рис Т. (2020) *Зачем технологическим компаниям нужны философы, и как я убедил Google их нанять*. URL: https://syg.ma/@natella-speranskaja/zachiem-tiekhnologhichieskim-kompaniiam-nuzhny-filosofy-i-kak-ia-ubiedil-google-ikh-naniat (проверено 23.06.2020).

«Робота-редактора Microsoft обвинили в расизме». (2020) // *РБК Стиль*, 10.06. URL: https://style.rbc.ru/repost/5ee0b8d59a7947a22682b563 (проверено 08.08.2020).

Самойдюк А. (2018) «Системы распознавания лиц не различают азиатов. Как IT-компании с этим борются» // *Rusbase*, 3.04. URL: https://rb.ru/story/how-companies-deal-with-ai-bias/ (проверено 15.07.2020).

Тимофеева О. (2017) *История животных*. М.: Новое литературное обозрение.

«Ученый: искусственный интеллект приведет к сознательной архаизации жизни». (2016) // *РИА Наука*, 27.11. URL: https://ria.ru/20161127/1482248032.html (проверено 19.07.2020).

Уэйкфилд Д. (2018) «Встречайте: Norman, алгоритм-психопат, которому мерещатся трупы» // *BBC*, 2.06. URL: https://www.bbc.com/russian/features-44344648 (проверено 19.07.2020).

«Юваль Харари — РБК: „У политиков должен быть барьер между умом и ртом“». (2020) // *РБК*, 27.05. URL: https://www.rbc.ru/society/27/05/2020/5ecd05659a79472c86a33115?from=newsfeed (проверено 29.06.2020).

Awad E., S.Dsouza, R.Kim, J.Schulz, J.Henrich, A.Shariff, J.-F.Bonnefon, and I.Rahwan. (2018) «The Moral Machine Experiment» // *Nature*, no. 563: 59—64.

Baila C.A., B.Guay, E.Maloneya, A.Combs, D.S.Hillygus, F.Merhout, D.Freelon, and A.Volfovsky. (2020) «Assessing the Russian Internet Research Agency's Impact on the Political Attitudes and Behaviors of American Twitter Users in Late 2017» // *PNAS*, vol. 117, no. 1: 243—250.

Bostrom N. and E.Yudkowsky. (2014) «The Ethics of Artificial Intelligence» // Frankish K., ed. *The Cambridge Handbook of Artificial Intelligence*. Cambridge: Cambridge University Press: 316—334.

CAHAI — Ad hoc Committee on Artificial Intelligence. (2020) URL: https://www.coe.int/en/web/artificial-intelligence/cahai (accessed on 17.07.2020).

Cappelli P., P.Tambe, and V.Yakubovich. (2019) «Artificial Intelligence in Human Resources Management: Challenges and a Path Forward» // *California Management Review*, vol. 61, no. 4: 15—42.

Chen A. and A.Potenza. (2018) «Cambridge Analytica's Facebook Data Abuse Shouldn't Get Credit for Trump» // *The Verge*, 20.03. URL: https://www.theverge.com/2018/3/20/17138854/cambridge-analytica-facebook-data-trump-campaign-psychographic-microtargeting (accessed on 01.08.2020).

Clark C. (2014) «Free to Punish: A Motivated Account of Free Will Belief» // *Journal of Personality and Social Psychology*, vol. 106, no. 4: 501—513.

Dastin J. (2018) «Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women» // *Business News*, 10.10. URL: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-

recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G (accessed on 01.08.2020).

Facebook объяснил, почему удалял фотографию водружения Знамени Победы на Рейхстаге. (2020) URL: https://tass.ru/obschestvo/8437501 (проверено 29.06.2020).

Foo Y.Ch. (2020) «EU Mulls Five-year Ban on Facial Recognition Tech in Public Areas» // *Reuters*, 16.01. URL: https://www.reuters.com/article/us-eu-ai-idUSKBN1ZF2QL (accessed on 03.08.2020).

Hao K. and J.Stray. (2019) «Can You Make AI Fairer Than a Judge? Play Our Courtroom Algorithm Game» // *MIT Technology Review*, 17.10. URL: https://www.technologyreview.com/s/613508/ai-fairer-than-judge-criminal-risk-assessment-algorithm/ (accessed on 03.08.2020).

Knight W. (2019) «Facebook, Google, Twitter Aren't Prepared for Presidential Deepfakes» // *MIT Technology Review*, 6.08. URL: https://www.technologyreview.com/2019/08/06/639/facebook-google-twitter-arent-prepared-for-presidential-deepfakes/ (accessed on 05.08.2020).

Knobe J. (2003) «Intentional Action and Side Effects in Ordinary Language» // *Analysis*, vol. 63, no. 3: 190—194.

Pichai S. (2020) «Why Google Thinks We Need to Regulate AI» // *Financial Times*, 19.01. URL: https://www.ft.com/content/3467659a-386d-11ea-ac3c-f68c10993b04 (accessed on 02.08.2020).

Vaccari C. and A.Chadwick. (2020) «Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News» // *Social Media + Society*, vol. 6, no. 1: 1—13.

D.A.Tomiltseva, A.S.Zheleznov

INEVITABLE THIRD: ETHICAL AND POLITICAL ASPECTS OF INTERACTIONS WITH ARTIFICIAL AGENTS

Daria A. Tomiltseva — Ph.D. in Philosophy; Associate Professor at the Department of Social Philosophy, Ural Federal University named after the first President of Russia B.N.Yeltsin (Yekaterinburg). Email: tomiltceva_d@mail.ru.

Andrey S. Zheleznov — Ph.D. in Philosophy; Independent Researcher (Moscow). Email: itsnomoredancing@gmail.com.

Abstract. Artificial agents i.e., man-made technical devices and software that are capable of taking meaningful actions and making independent

decisions, permeate almost all spheres of human life today. Being new political actants, they transform the nature of human interactions, which gives rise to the problem of ethical and political regulation of their activities. Therefore, the appearance of such agents triggers a global philosophical reflection that goes beyond technical or practical issues and makes researchers return to the fundamental problems of ethics.

The article identifies three main aspects that call for philosophical understanding of the existence of artificial agents. First, artificial agents reveal the true contradiction between declared moral and political values and real social practices. Learning from the data on the assessments and conclusions that have already taken place, artificial agents make decisions that correspond to the prevailing behavioral patterns rather than moral principles of their creators or consumers. Second, the specificity of the creation and functioning of artificial agents brings the problem of responsibility for their actions to the forefront, which, in turn, requires a new approach to the political regulation of the activities of not only developers, customers and users, but also the agents themselves. Third, the current forms of the activity of artificial agents shift the traditional boundaries of the human and raise the question of redefining the humanitarian. Having carefully analyzed the selected aspects, the authors reveal their logic and outline the field for further discussion.

Keywords: artificial agent, artificial intelligence, morality, bias, responsibility, politics

References

Awad E., S.Dsouza, R.Kim, J.Schulz, J.Henrich, A.Shariff, J.-F.Bonnefon, and I.Rahwan. (2018) "The Moral Machine Experiment" // *Nature*, no. 563: 59—64.

Baila C.A., B.Guay, E.Maloneya, A.Combs, D.S.Hillygus, F.Merhout, D.Freelon, and A.Volfovsky. (2020) "Assessing the Russian Internet Research Agency's Impact on the Political Attitudes and Behaviors of American Twitter Users in Late 2017" // *PNAS*, vol. 117, no. 1: 243—250.

Bostrom N. and E.Yudkowsky. (2014) "The Ethics of Artificial Intelligence" // Frankish K., ed. *The Cambridge Handbook of Artificial Intelligence*. Cambridge: Cambridge University Press: 316—334.

CAHAI — Ad hoc Committee on Artificial Intelligence. (2020) URL: https://www.coe.int/en/web/artificial-intelligence/cahai (accessed on 17.07.2020).

Cappelli P., P.Tambe, and V.Yakubovich. (2019) "Artificial Intelligence in Human Resources Management: Challenges and a Path Forward" // *California Management Review*, vol. 61, no. 4: 15—42.

Chen A. and A.Potenza. (2018) "Cambridge Analytica's Facebook Data Abuse Shouldn't Get Credit for Trump" // *The Verge*, 20.03. URL: https://www.theverge.com/2018/3/20/17138854/cambridge-analytica-facebook-data-trump-campaign-psychographic-microtargeting (accessed on 01.08.2020).

Clark C. (2014) "Free to Punish: A Motivated Account of Free Will Belief" // *Journal of Personality and Social Psychology*, vol. 106, no. 4: 501—513.

Dastin J. (2018) "Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women" // *Business News*, 10.10. URL: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G (accessed on 01.08.2020).

Emelin K.N. (2020). "Iskusstvennyj intellekt pod kontrolem" [Artificial Intelligence Is under Control] // *Vestnik Komissii Rossijskoj Federatsii po delam UNESKO* [Vestnik of the Commission of the Russian Federation for UNESCO], no. 4. URL: http://unesco.ru/news/ai-committee/ (accessed on 19.07.2020). (In Russ.)

Facebook ob'jasnil, pochemu udaljal fotografiju vodruzhenija Znameni Pobedy na Rejkhstage [Facebook Explained Why It Deleted the Photo of Hoisting the Banner of Victory over the Reichstag]. (2020) URL: https://tass.ru/obschestvo/8437501(accessed on 29.06.2020). (In Russ.)

Foo Y.Ch. (2020) "EU Mulls Five-year Ban on Facial Recognition Tech in Public Areas" // *Reuters*, 16.01. URL: https://www.reuters.com/article/us-eu-ai-idUSKBN1ZF2QL (accessed on 03.08.2020).

"Iskusstvennyj intellekt projavil seksizm i etim postavil krest na prilozhenii dlja najma ot Amazon" [Artificial Intelligence Has Shown to Be Sexist and Thus Put an End to the Hiring App from Amazon]. (2018) // *BBC*, 10.10. URL: https://www.bbc.com/russian/other-news-45814099 (accessed on 10.07.2020). (In Russ.)

Hao K. and J.Stray. (2019) "Can You Make AI Fairer Than a Judge? Play Our Courtroom Algorithm Game" // *MIT Technology Review*, 17.10. URL: https://www.technologyreview.com/s/613508/ai-fairer-than-judge-criminal-risk-assessment-algorithm/ (accessed on 03.08.2020).

Kashkin S.Yu. and A.V.Pokrovsky. (2019) "Iskusstvennyj intellekt, robototekhnika i zashchita prav cheloveka v Evropejskom sojuze" [Artificial Intelligence, Robotics and the Protection of Human Rights in the European Union] // Vestnik Universiteta im. M.E.Kutafina (MGJuA) [Courier of the Kutafin Moscow State Law University (MSAL)], no. 4: 64—90. URL: http://vestnik-msal.ru/articles/article_105096.html?issue=vest-1-2019 (accessed on 10.08.2020). (In Russ.)

Kasyanova D. (2020) "Indijskij politik ispol'zoval tekhnologiju dipfejk, chtoby perevesti svoju rech' na raznye jazyki" [Indian Politician Used Deepfake Technology to Translate His Speech on Different Languages] // *Bird in Flight*, 19.02. URL: https://birdinflight.com/ru/novosti/indiyskiy-politik-ispolzoval-tehnologiyu-dipfeykov-chtoby-perevesti-svoyu-rech-na-raznye-yazyki.html (accessed on 10.08.2020). (In Russ.)

Kerimov T.Kh. (2019) "Tsifrovizatsija obshchestva: moduljatsija, vremja, sub''ektivatsija" [Digitalization of Society: Modulation, Time, Subjectivation] // *Izvestija Ural'skogo federal'nogo universiteta. Serija 3: Obshchestvennye nauki* [Izvestia: Ural Federal University Journal. Series 3: Social and Political Sciences], vol. 14, no. 3 (191): 5—17. (In Russ.)

Knight W. (2019) "Facebook, Google, Twitter Aren't Prepared for Presidential Deepfakes" // *MIT Technology Review*, 6.08. URL: https://www.technologyreview.com/2019/08/06/639/facebook-google-twitter-arent-prepared-for-presidential-deepfakes/ (accessed on 05.08.2020).

Knobe J. (2003) "Intentional Action and Side Effects in Ordinary Language" // *Analysis*, vol. 63, no. 3: 190—194.

Kontseptsija razvitija regulirovanija otnoshenij v sfere tekhnologij iskusstvennogo intellekta i robototekhniki do 2024 goda [The Concept of Development of Regulation of Relations in the Field of Artificial Intelligence Technologies and Robotics until 2024]. (2020) URL: http://www.consultant.ru/document/cons_doc_LAW_360681/ (accessed on 27.08.2020). (In Russ.)

Martyanov D.S. (2016) "Politicheskij bot kak professija" [Political Bot as a Profession] // *Politicheskaja ekspertiza: POLITEKS* [Political Expertise: POLITEX], no. 1: 74—89. (In Russ.)

Martynov K. (2019) "Etika avtonomnykh mashin: deontologija i voennye roboty" [An Ethics of Autonomous Machines: Deontology and Military Robots] // *Logos,* vol. 29, no. 3: 231—246. URL: http://logosjournal.ru/arch/107/Logos%203-2019_Press-239-254.pdf (accessed on 27.08.2020). (In Russ.)

O sozdanii Komiteta po voprosam etiki iskusstvennogo intellekta pri Komissii Rossijskoj Federatsii po delam YuNESKO [On the Establishment of the Committee on Artificial Intelligence Ethics under the Commission of the Russian Federation for UNESCO]. (2020) URL: https://www.mid.ru/ru/foreign_policy/un/-/asset_publisher/U1StPbE8y3al/content/id/4069630 (accessed on 17.07.2020). (In Russ.)

Pichai S. (2020) "Why Google Thinks We Need to Regulate AI" // *Financial Times*, 19.01. URL: https://www.ft.com/content/3467659a-386d-11ea-ac3c-f68c10993b04 (accessed on 02.08.2020).

Rabochaja gruppa ANO "Tsifrovaja ekonomika" odobrila razrabotannyj Minekonomrazvitija proekt Kontseptsii razvitija regulirovanija v sfere iskusstvennogo intellekta [The Working Group of Autonomous Nonprofit Organization "Digital Economy" Approved the Draft Concept of the Development of Regulation in the Field of Artificial Intelligence Created by the Ministry of Economic Development]. (2020) URL: https://www.economy.gov.ru/material/news/rabochaya_gruppa_ano_cifrovaya_ekonomika_odobrila_razrabotannyy_minekonomrazvitiya_proekt_koncepcii_razvitiya_regulirovaniya_v_sfere_iskusstvennogo_intellekta.html (accessed on 17.07.2020). (In Russ.)

"Razvitie inzhenernogo obrazovanija i formirovanie sovremennoj inzhenernoj kul'tury v Rossii (Dvadtsat' pjatye gubernatorskie chtenija. Tjumen', 28 ijunja 2016 g.)" [Development of Engineering Education and Formation of Modern Engineering Culture in Russia (Twenty-Fifth Gubernatorial Readings. Tyumen, June 28th, 2016)]. (2016) // Politeia, no. 3: 160—183. URL: http://politeia.ru/files/articles/rus/2016_03_09.pdf (accessed on 29.08.2020). (In Russ.)

Rees T. (2020) *Zachem tekhnologicheskim kompanijam nuzhny filosofy, i kak ja ubedil Google ikh nanjat'* [Why Tech Companies Need Philosophers, and How I Convinced Google to Hire Them]. URL: https://syg.ma/

@natella-speranskaya/zachiem-tiekhnologhichieskim-kompaniiam-nuzhny-filosofy-i-kak-ia-ubiedil-google-ikh-naniat (accessed on 23.06.2020). (In Russ.)

"Robota-redaktora Microsoft obvinili v rasizme" [Microsoft's Robot Editor Has Been Accused of Racism]. (2020) // *RBC Style*, 10.06. URL: https://style.rbc.ru/repost/5ee0b8d59a7947a22682b563 (accessed on 08.08.2020). (In Russ.)

Samoydyuk A. (2018) "Sistemy raspoznavanija lits ne razlichajut aziatov. Kak IT-kompanii s etim borjutsja" [The Face Recognition Systems Don't Differentiate Asians. How Do IT Companies Deal with This] // *Rusbase*, 3.04. URL: https://rb.ru/story/how-companies-deal-with-ai-bias/ (accessed on 15.07.2020). (In Russ.)

Timofeeva O. (2017) *Istorija zhivotnykh* [The History of Animals]. Moscow: Novoe literaturnoe obozrenie. (In Russ.)

"Uchenyj: iskusstvennyj intellekt privedet k soznatel'noj arkhaizatsii zhizni" [The Scientist: AI Will Lead to a Deliberate Archaization of Life]. (2016) // *RIA Nauka*, 27.11. URL: https://ria.ru/20161127/1482248032.html (accessed on 19.07.2020). (In Russ.)

"V Moskve sozdali Komitet po II pri rossijskoj komissii po delam YuNESKO [The AI Committee of the Russian Commission for UNESCO Was Established in Moscow]. (2020) // *RIA Nauka*, 27.02. URL: https://ria.ru/20200227/1565300001.html. (accessed on 17.07.2020). (In Russ.)

Vaccari C. and A.Chadwick. (2020) "Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News" // *Social Media + Society*, vol. 6, no. 1: 1—13.

Wakefield J. (2018) "Vstrechajte: Norman, algoritm-psikhopat, kotoromu mereshchatsja trupy" [Are You Scared Yet? Meet Norman, the Psychopathic AI] // *BBC*, 2.06. URL: https://www.bbc.com/russian/features-44344648 (accessed on 19.07.2020). (In Russ.)

"Yuval' Kharari — RBK: „U politikov dolzhen byt' bar'er mezhdu umom i rtom"" [Yuval Harari to RBC: "Politicians Should Have a Barrier between the Mind and the Mouth]. (2020) // RBC, 27.05. URL: https://www.rbc.ru/society/27/05/2020/5ecd05659a79472c86a33115?from=newsfeed (accessed on 29.06.2020). (In Russ.)