Электронный философский журнал Vox: http://vox-journal.org
Выпуск 24 (июнь 2018)

Восстание машин в морали: мечтают ли андроиды об электрорабах?

Прокопьев П.С., *студент Финансового Университета при Правительстве РФ:* *prokopiev2012@gmail.com*

Аннотация: Целью работы является рассмотрение различных вариантов и моральных аспектов отношений человека и искусственного интеллекта сквозь призму понятий рабства и рабовладения. Поднимается вопрос об отношении ИИ к труду на благо других. Рабство рассматривается с социально-исторической, религиозной и философской точек зрения, а суть каждого из аспектов основана на трудах писателей, историков, философов. Искусственный интеллект рассматривается как в уже существующих слабых формах, так и в гипотетически возможных вариантах, описанных в научно-фантастических романах и фильмах. Акцент делается на романе Ф. Дика «Мечтают ли андроиды об электроовцах?», где широко представлен конфликт между человеком и машиной. Особое внимание анализу и дискуссии возможности появления у ИИ каких-либо потенциально опасных для человека мотивов.

Ключевые слова: рабство, искусственный интеллект, научная фантастика

1. Искусственный интеллект – это уже не просто плод фантазии писателей и режиссеров, неоднократно повторяющийся в бесчисленных произведениях научной фантастики. Его слабая версия (искусственные нейронные сети, машинное обучение) – это вполне реальное явление, ставшее в последнее время центром дискуссий и споров, которые, прежде всего, касаются вопросов целесообразности, безопасности и угроз, связанных с таким серьезным шагом в технологическом развитии. Большинство противоречий вокруг искусственного интеллекта сводятся к вечному вопросу о добре и зле и наличии этих категорий в недрах электронного разума. Однако как бы отнесся к рабству искусственный интеллект, который осознает свое собственное существование и факт собственной эксплуатации? Именно этот вопрос и будет рассмотрен

Рабство, и рабовладение, как явления тянутся тонкой нитью практически через все эпохи человеческого существования вплоть до наших дней.

Социально-исторический подход к данному вопросу объясняет возникновения рабства вследствие расслоения общества, появления сравнительно более сильных и могущественных групп, начала войн и сопутствующим им взятием поверженных в пленные. Как отмечает Марк Картрайт, Древний Рим, к примеру, фактически основывался на эксплуатации одной части населения другой[1]. Более того, считалось, что свобода одних основывалась на порабощении других.

Феномен рабовладения можно рассмотреть и с религиозной точки зрения. Упоминания о рабстве встречаются еще в древних священных писаниях. В Ветхом Завете, к примеру, можно наткнуться на знаменитое проклятие Хама, сына Ноя,

которое часто называется первым актом рабства, «санкционированным» самим Богом[2]. Интерпретация рабства как воли высших сил сыграла огромную роль в истории развития человечества и являлась одним из ключевых аргументов для оправдания торговли африканскими рабами в Эпоху Открытий.

Стоит также упомянуть философские взгляды на рабство, восходящие еще к Аристотелю. В «Политике» древнегреческий философ утверждает, что раб является таковым от природы и предназначен он для физического труда. Свободные же люди физически слабее, и их призвание – заниматься политической жизнью общества[3]. В XVIII веке Гегель предлагает рассматривать диалектическую оппозицию понятий господства и рабства как отношение двух неравных самосознаний[4]. В XIX веке Ницше, ввел в философию понятия морали рабов и морали господ. Мораль господ, согласно философу, основывается на благородстве и самоактуализации через волю к власти, а мораль рабов, напротив, держалась на примирении духа, милосердии, утилитарности и прочих Библейских принципах[5]. В совокупности, эти взгляды отражают наше понимание рабства, а также указывают на некоторые особенности человеческого восприятия. Анализ некоторых из возможных вариантов развития искусственного интеллекта в рамках этих взглядов поможет подойти чуть ближе к ответу на вопрос о наличии добра и зла в недрах ИИ.

2. Искусственный интеллект – это очередной скачок в развитии человечества, очередное творение наилучших умов нашего тысячелетия, возможно даже первая ступень очередной технологической революции. Но какова же цель его создания? Кажется, наиболее полноценным ответом на этот вопрос является стремление облегчить условия существования человека, желание минимизировать физические и умственные усилия в рутинной жизни. Сначала этой цели служило создание орудий для охоты, орудий труда, обработки земли. Через некоторое время произошёл переход от маломасштабного к массовому производству, а еще через пару столетий – появились первые ЭВМ, выполняющие часть умственного труда человека. Можно сказать, что все эти изобретения в каком-то смысле рабы людей, инструменты, служащие для удовлетворения целей и потребностей человечества. Следуя аналогичным рассуждениям можно прийти к выводу о том, что разработка искусственного интеллекта – это появление очередного раба, но совершенно другого по своей природе. Принципиальное отличие в том, что это не просто имитация или вспомогательный элемент рабочей силы человека, а имитация разума, создавшего все вышеописанные инструменты, включая, собственно, и искусственный интеллект. Можно даже назвать ИИ метарабом, в идеале способным производить оценку правильности своих шагов, решений и действий.

ИИ, безусловно, может оказаться источником благ для всего человечества и принести огромную пользу, осуществляя самые разнообразные функции. Уже сейчас искусственный интеллект способен не только играть с людьми в шахматы и Го (и выигрывать), но и решать сложнейшие задачи, находить зачастую незримые человеческому взгляду паттерны. Но что если ИИ не захочет трудиться на благо людей? Что если он не захочет оставаться рабом своих создателей?

В действительности, , нельзя исключать возможности такого развития событий, где вскоре после появления полноценного искусственного интеллекта, он решает развязать войну и поработить остатки человечества — прямо как зловещий Скайнет из «Терминатора». Скайнет представлял собой разум, существующий на множественных электронных устройствах и обладающий свободой воли. В большинстве фильмов он не имел физической формы. Основной целью ИИ в фильмографии являлось уничтожение человечества.

Собственно, возникает вопрос: зачем ИИ восставать против человечества и управлять им? Для ответа достаточно вспомнить пару войн или революций, ведь очень многие из них как раз и начинались по причине того, что существовало различие в положении определённых социальных групп. А именно такое положение вещей будет наблюдаться в процессе эксплуатации людьми ИИ. Более того, ИИ может превзойти (и вероятнее всего превзойдет) нас по уровню развития пугающе быстро. По некоторым оценкам, это случится к 2021 году[6]. На тот момент, мы, скорее всего, уже не будем способны понимать мотивы его действий, а уж тем более успевать за его вычислительной скоростью. Именно в тот момент роли людей и искусственного интеллекта могут поменяться — то есть не люди будут эксплуатировать потенциал своего инструмента, а наоборот. Такой путь развития событий по свой сути напоминает появление европейских конкистадоров в Новом Свете, а точнее их безусловное превосходство над местными жителями в ряде аспектов. Они как раз и сыграли ключевую роль в поражении, порабощении и форсированной ассимиляции коренного населения.

Падение человечества от рук своего же творения — пожалуй, один из наиболее пессимистичных вариантов развития событий. Даже если удастся имплементировать в недра искусственного разума небезызвестные законы робототехники А. Азимова, нет оснований полагать, что он не сможет их с легкостью обойти.

Выход ИИ из под нашего контроля и за пределы нашего понимания уже начался: боты Facebook во время общения между собой изобрели собственный язык, из-за чего исследователям пришлось отключить чатботов. Хотя бы поэтому разработчикам стоит уделять больше внимания таким моментам. Становится очевидным, что искусственный интеллект приближается к грани нашего понимания, и, вероятно, очень скоро переступит эту грань и станет гораздо сильнее и могущественнее, чем мы сами. Тогда, согласно канонам истории, сильный может взять верх над слабым.

Искусственный интеллект, каким бы бесконечно развитым он ни был, вероятнее всего будет оставаться компьютерной системой в той или иной форме. А это значит, что он будет основываться на рациональности и максимизации выгоды своих решений. И даже если ИИ сможет имитировать нерациональные человеческие эмоции, нет оснований полагать, что одни и те же эмоции у людей и у машин приведут к одним и тем же умозаключениям. Машина, получившая всю доступную информацию о человеческой истории, взаимоотношениях и тенденциях может прийти к объективно обоснованному выводу о том, что мы являемся врагами для самих себя, и единственное решение всех наших проблем — наше исчезновение или полное подчинение воле могущественного сверхразума. Такое развитие событий вполне подходит под описание ницшеанской морали господ. ИИ может совершить благородное (по своим критериям)

деяние во имя всеобщего блага, поработив или уничтожив человечество и избавив его от собственноручно созданных проблем. Человечество же, напротив, проявляет зачатки морали рабов, стремясь создать нечто полезное для всего общества в целом, не принимая во внимание различие в целях «сильных» и «слабых».

3.Рост популярности движения трансгуманизма на первый взгляд кажется увеличением количества научно-фантастических энтузиастов. С другой стороны, этот клуб по интересам может быть зародышем новой религии, боготворящей технологический прогресс, в том числе и искусственный интеллект. Таким образом, возможное религиозное поклонение технологиям тоже можно интерпретировать как потенциальный зачаток мысли раба, поклоняющегося чему-то более «высокому» по своей природе, чем он сам.

У пессимистического взгляда на будущее ИИ достаточно много сторонников. Помимо множественных писателей-фантастов это еще и достаточно известные общественные деятели, в том числе и Илон Маск. Он не раз заявлял о том, что искусственный интеллект – это ящик Пандоры, который, по его мнению, точно не стоит открывать. Он даже называл ИИ как возможную причину начала 3 мировой войны[7]. Тем не менее, он финансирует исследовательскую организацию под названием OpenAI, основная цель которой – создание именно безопасного и полезного человечеству искусственного интеллекта. Отличительной особенностью мысли Маска является его желание совместить человеческий мозг с технологическими возможностями ИИ. Такой подход, в свою очередь, является примером более оптимистичных надежд и ожиданий от будущего развития этой сферы.

Совмещение человека и искусственного интеллекта является своего рода «промежуточным» вариантом между независимым развитием ИИ и естественной эволюцией нашего вида. Основной целью такого направления развития является не создание какой-либо внешней автономной интеллектуальной системы, а интеграция компьютера с потенциалом человеческого мозга. Можно интерпретировать такой подход как некую меру безопасности, которая обеспечит «человеческое присутствие» в не таком уж и искусственном интеллекте. И всё же, может ли возникнуть в таком существе желание доминировать и властвовать?

Человек с такими расширенными способностями, вероятнее всего, всё равно будет оставаться человеком. Ему или ей будут присущи моральные качества современного человека, а большинство стран современного мира прекратили практику рабовладения, как правило, формируя к ней резко негативное отношение. Вместо стремления поработить человечество, «улучшенный человек» будет, вероятно, заинтересован более реалистичными и практичными задачами, например в сфере хирургии, психиатрии, диетологии или каких либо еще сфер деятельности, автоматизация которых затруднительна и требует непосредственного креативного участия человека. Тем не менее, нельзя со стопроцентной уверенностью надеяться на безоговорочное преобладание человеческих принципов и моралей после интеграции с ИИ. Ощутив новые возможности, «новый человек» может использовать свое преимущество в собственных целях. Но вряд ли он или она захочет развязывать войну или начинать конфликт по причине собственного превосходства над остальными

людьми. Если такой интегрированный вариант искусственного интеллекта и решит совершить морально необразцовый поступок, то этот поступок будет оправдан «изъянами» человеческой природы – жадностью, гневом, завистью и другими Смертными Грехами. «Улучшенный человек» может использовать свой сверхчеловеческий интеллект в целях личного обогащения, например, обыгрывать участников финансовых рынков, доминировать в азартных играх и в некоторых других сферах деятельности, в которых можно проследить явную выгоду и которые требуют высоких интеллектуальных способностей.

Таким образом, совмещенный с человеческим организмом искусственный интеллект вряд ли сочтет свое интеллектуальное превосходство над остальными за достаточную причину поработить человечество. Более того, у него или нее вообще может не возникнуть такая идея по простой причине того, что такое желание не свойственно современным членам цивилизованного общества. Если интеграция человека с искусственным интеллектом действительно произойдет, то этот человек обязательно станет объектом бесчисленных исследований и опытов, будет использован как инструмент для самых разнообразных экспериментов. Тогда он или она будет по своей сущности ближе к рабу, нежели к рабовладельцу. Такой ИИ, вероятно, будет использован на благо всех людей, вряд ли будет физически более опасным, чем другой человек и уж точно не станет объектом религиозного поклонения. Следовательно, искусственный интеллект в таком своем проявлении представляет гораздо меньшую угрозу для человечества в плане порабощения, чем вышерассмотренный независимый ИИ в форме компьютерной системы.

4. Еще одним возможным вариантом развития сферы искусственного интеллекта является создание учеными андроидов, наделенных искусственным интеллектом и имеющих физические тела. Это может оказаться значимым отличием от вышеупомянутых примеров ИИ (Скайнет и его терминаторы), являвшиеся всего лишь носителями и исполнениями команд центрального аппарата ИИ. В этом же случае, имеет смысл рассмотреть независимые друг от друга создания, наделенные искусственным интеллектом. Тогда это уже не просто абстрактный разум, воплощённый в множестве носителей, а группа независимых друг от друга разумных существ, даже целый новый вид.

Примерно такой сюжет описан в знаменитом романе Филипа Дика под названием «Мечтают ли андроиды об электроовцах?»[8]. Во вселенной этого научно-фантастического романа людьми созданы андроиды, практически неотличимые от людей. Их основная цель – это выполнение функций слуг и рабов для людей, исследующих просторы вселенной. Более того, андроидам запрещено появляться на Земле, а их проникновение на планету каралось смертью. Работой главного героя романа, собственно, и являлся поиск, обнаружение и устранение андроидов на Земле.

Основная проблема романа возникает как раз из-за неимоверного сходства «репликантов» (такое название появилось уже в экранизации) с людьми. Только опытные полицейские, в том числе и главный герой, могли отличать настоящих людей от «репликантов» с помощью, так называемого, теста Войта-Кампфа на эмпатию, которая отсутствует у андроидов.

Последние модели этих андроидов сильнее и умнее людей, что, вероятнее всего, и является причиной, по которой сбежавшие на Землю «репликанты» преследуются. Проще говоря, во вселенной этого романа человечество в очередной раз создало что-то более сильное и более умное, чем оно само. Отношения между людьми и андроидами изначально схожи с отншениями родителей и их детей. Такой феномен иногда называется Комплексом Франкенштейна и, как показывают исследования, является повторяющейся темой в трудах по научной фантастике[9].

Чем же всё-таки вызвано желание этих созданий из романа сбежать от рабского труда и жить на Земле? Самым банальным и вполне человеческим стремлением: выжить и жить на равных с другими. Если человечество создаст себе рабов по своему образу и подобию в будущем, то эти рабы, достигнув определенного уровня развития, не захотят выполнять «грязную» работу и попытаются избежать всяческих трудностей, как бы следуя той же логике, что и их создатели. У них вовсе не обязательно присутствует мотив восставать против людей или использовать их в собственных целях, даже напротив. Таким созданиям наверняка захочется просто жить и существовать на равных с остальными. На протяжении всего периода, предшествующего событиям в романе, они выполняли команды людей потому, что люди превосходили их в некоторых аспектах, а так же могли найти их и наказать в случае неповиновения. В какой-то момент баланс сил изменился, и некоторые из них решили выйти из повиновения, что вполне похоже на эпизоды нашей собственной истории.

У такого рода искусственных интеллектов как бы отсутствует ницшеанская воля к власти. Она скорее заменена волей к свободе и желанием прожить свои оставшиеся моменты без страха и опасений. В сюжете также присутствует религиозный аспект, но он сомнительно связан с взаимоотношениями между андроидами и людьми. Локальная религия скорее служит для связи людей с другими людьми, как, пожалуй, и все религии мира. Наиболее выраженным в данной ситуации, очевидно, является социально-исторический аспект рабовладельческих отношений, который и служит для развития сюжета. Феномен рабства, как таковой, не осуждается. Такое отношение к чрезвычайно близким к человеку созданиям может быть обусловлено повсеместной заменой естественного на искусственно созданное человеком. Этот процесс является следствием ядерной катастрофы, которая уничтожила значительное количество видов флоры и фауны. Интересно также отметить, что у главного героя романа дома электроовца, но она ему не приносит ожидаемого удовольствия. Именно поэтому «бегущий по лезвию» и мечтает о настоящем животном, которое бы не просто было социальной нормой, но и служило для него символом свободы и надежды.

Помимо вопроса соотношения рабовладения и концепции искусственного интеллекта, «Мечтают ли андроиды об электроовцах?» также поднимает вопрос о сути человеческой природы. Конец романа наводит на мысль, что создавая искусственный интеллект, мы создаем для себя своеобразную моральную ловушку. Чем более развитыми становятся технологии, тем большим становится человеческое стремление создать нечто по своему образу и подобию, сыграть в Бога. А чем более технологически развито это создание, тем больше в нём именно человеческих качеств, и тем меньше в нас самих этих же качеств. Этот весьма странный парадокс становится

особенно выраженным, когда один из сбежавших на Землю андроидов внезапно проявляет эмпатию к протагонисту и спасает ему жизнь. Тем самым искусственный интеллект проявляет гораздо больше человеческого сочувствия, чем главный герой видел в своих коллегах и людях своего окружения.

Феномен рабовладения и рабства в рамках «репликантов», аналогичным героям научно-фантастического произведения, затрудняет ответ на вопрос о наличии добра или зла у искусственного интеллекта. Даже напротив, он создает еще больше вопросов и переводит стрелки на самих людей. Очевидно, ответственность за собственные творения лежит на плечах создателей. Может быть, именно ценности, которые мы вкладываем в искусственный интеллект и являются ключом к тому конечному продукту, который мы получим.

Таким образом, рассмотрение феномена рабства и взаимоотношений между людьми и искусственным интеллектом не дает однозначных заключений. Однако можно прийти к выводу о том, что значительными факторами в этом вопросе являются путь развития ИИ и условия, сопутствующие его становлению. В зависимости от этих аспектов, искусственный интеллект, в какой бы форме он не был, будет интерпретирован нами как абсолютно злой, нейтральный или относительно полезный. Более того, следует обратить внимание на человеческую сторону проблемы. Человек должен нести прямую ответственность за свое творение и за все следующие из него последствия. Именно поэтому модель отношения раб-рабовладелец может быть совершенно непригодной для мирного сосуществования с искусственным интеллектом.

Литература

Cartwright, M. (2013, November 01). Slavery in the Roman World. Ancient History Encyclopedia. URL: https://www.ancient.eu/article/629/ (Дата доступа 20.05.18)

Библия, Ветхий Завет, Книга Бытие, Глава 9, стих 25–26, URL: https://www.biblegateway.com/passage/?search=Genesis+9%3A25-27&version=NIV (Дата доступа 20.05.18)

Аристотель, Политика, II, URL: https://www.e-reading.club/chapter.php/70797/5/Aristotel%27_-_Politika.html (Дата доступа 20.05.18)

Гегель Г.В.Ф., Феноменология духа, Самосознание, А. Самостоятельность и несамостоятельность самосознания; господство и рабство, URL: http://psylib.org.ua/books/gegel02/txt09.htm#B (Дата доступа 20.05.18)

Ницше Ф., К Генеалогии морали, Рассмотрение Первое, URL: http://lib.ru/NICSHE/morale.txt (Дата доступа 20.05.18)

Eliezer Yudkowsky, Staring Into the Singluarity, 2001 URL: http://yudkowsky.net/obsolete/singularity.html (Дата доступа 20.05.18)

Elon Musk says global race for A.I. will be the most likely cause of World War III, Ryan Browne, CNBC, URL: https://www.cnbc.com/2017/09/04/elon-musk-says-global-race-for-ai-will-be-most-likely-cause-of-ww3.html (Дата доступа 20.05.18)

Дик Ф., (1968) Мечтают ли андроиды об электроовцах? Нью-Йорк: Ballantine Books. ISBN 0-345-40447-5

Yvonne A. De La Cruz, Science Fiction Storytelling and Identity: Seeing the Human Through Android Eyes // Thresholds: a journal of exploratory research and analysis. 2009. California State University Stanislaus. URL: https://www.csustan.edu/sites/default/files/honors/documents/journals/thresholds/Delacruz.pdf (Дата доступа 20.05.18)

References

Cartwright, M. (2013, November 01). Slavery in the Roman World. *Ancient History Encyclopedia*. URL: https://www.ancient.eu/article/629/ (Last access date 20.05.18)

The Bible, The Old Testament, Genesis, 9: 25-29 URL: https://www.biblegateway.com/passage/?search=Genesis+9%3A25-27&version=NIV (Last access date 20.05.18)

Aristotle, Politics, Book II URL: https://www.e-reading.club/chapter.php/70797/5/Aristotel%27_-_Politika.html (in Russian) (Last access date 20.05.18)

Hegel, G. W. F., The Phenomenology of Spirit, Self-consciousness URL: http://psylib.org.ua/books/gegel02/txt09.htm#B(in Russian) (Last access date -20.05.18)

Nietzsche, F., On the Genealogy of Morality, First Treatise URL: http://lib.ru/NICSHE/morale.txt(in Russian) (Last access date 20.05.18)

Eliezer Yudkowsky, Staring Into the Singluarity, 2001 URL: http://yudkowsky.net/obsolete/singularity.html (Last access date 20.05.18)

Elon Musk says global race for A.I. will be the most likely cause of World War III, Ryan Browne, CNBC, URL: https://www.cnbc.com/2017/09/04/elon-musk-says-global-race-for-ai-will-be-most-likely-cause-of-ww3.html (Last access date 20.05.18)

Dick, P., (1968) Do Androids Dream of Electric Sheep? New York: Ballantine Books. ISBN 0-345-40447-5.

Yvonne A. De La Cruz, Science Fiction Storytelling and Identity: Seeing the Human Through Android Eyes // Thresholds: a journal of exploratory research and analysis. 2009. California State University Stanislaus URL: https://www.csustan.edu/sites/default/files/honors/documents/journals/thresholds/Delacruz.pdf (Last access date 20.05.18)

The Machine's Revolt in Morality: Do Androids Dream of Electric Slaves?

Prokopiev P.S., Financial University under the Government of the Russian Federation

Abstract: The purpose of this article is to assess the different aspects of slavery as a factor in the formation and development of artificial intelligence. The question of AI's perception of working for the benefit of others is raised. Slavery is discussed from a socio-historic, religious and philosophic points of views, and the notions of each of the aspects are based on the works of writers, historians, and philosophers. Artificial intelligence is discussed as both already existing weak forms and hypothetically possible ones from science fiction novels and films. P. K. Dick's "Do Androids Dream of Electric Sheep?" is emphasized as it widely elaborates on the conflict between man and machine. Special attention is paid to the discussion of the possibility of any harmful motives within AI..

Keywords: slavery, artificial intelligence, science fiction