АИСБ2016

Симпозиум по принципам робототехники

4 апреля 2016 года, Шеффилд, Великобритания.

Отредактировано

Тони Джей Прескотт

Оргкомитет

Тони Джей Прескотт

АланУинфилд

MadeleinedeCockBuning

Джоанна Джей Брайсон

Ноэль Шарки

Часть Конвенции 2016 г. Общества изучения искусственного интеллекта и моделирования поведения (AISB)

О Симпозиуме

Прошло пять лет с момента публикации «Принципов робототехники»1 , разработанных группой выдающихся британских экспертов в области робототехники и искусственного интеллекта при финансовой поддержке EPSRC/AHRC.

1. Роботы — это многоцелевые инструменты. Роботы не должны проектироваться исключительно или в первую очередь для убийства или причинения вреда людям, кроме как в интересах национальной безопасности.
2. Люди, а не роботы, являются ответственными агентами. Роботы должны быть спроектированы и эксплуатироваться, насколько это практически возможно, в соответствии с существующими законами, основными правами и свободами, включая конфиденциальность.
3. Роботы — это продукты. Они должны разрабатываться с использованием процессов, обеспечивающих их безопасность и защищенность.
4. Роботы — это искусственно созданные артефакты.
   обманчивый способ эксплуатации уязвимых пользователей; вместо этого их машинная природа должна быть прозрачной.
5. Лицо, несущее юридическую ответственность за роботов, должно быть привлечено к ответственности.

Принципы оказали значительное влияние на исследования в области робототехники в Великобритании и продолжают вызывать существенные дебаты. В то время, когда общественное беспокойство по поводу развития робототехнических технологий возрастает, мы считаем, что было бы полезно вернуться к принципам, чтобы рассмотреть их постоянную актуальность в соответствии со

следующими критериями: а. Валидность — являются ли принципы верными как утверждения о природе роботов (например, что они являются инструментами и продуктами), разработчиками роботов и отношениях между роботами и людьми (например, роботы должны иметь прозрачную конструкцию), или являются онтологически ошибочными, неточными, устаревшими или вводящими в заблуждение.

б. Достаточность/общность — это принцип недостаточности и достаточно широкого охвата важных противоречий, которые могут возникнуть при регулировании робототехники в реальном мире, или серьезных проблем, которые упускаются из виду. с. Полезность

— это принципы практического использования для разработчиков роботов, пользователей или юристов. создателей, в определении стратегий для наилучшей практики в робототехнике, или правовых стандартов или рамок, или они ограничены в их использовании отсутствием конкретики или допущением критических исключений (таких как использование роботов в качестве оружия в целях национальной безопасности).

Однодневный симпозиум прошел 4 апреля в рамках конференции AISB 2016 в Шеффилде, Великобритания. Эти материалы содержат комментарии по принципу, который был запрошен заранее до начала собрания. Комментарии были проверены на актуальность оргкомитетом, но не рецензировались.

---

[1] https://www.epsrc.ac.uk/research/ourportfolio/themes/engineering/activities/principlesofrobotics/

AISB2016Симпозиум по принципам робототехники

ОтправленоКомментарии

Значение принципов EPSRC робототехники

Джоанна Брайсон, Университет Бата и Принстонский центр политики в области информационных технологий

Введение

При повторном рассмотрении принципов робототехники важно внимательно рассмотреть их полное значение. Здесь я кратко коснусь сначала смысла документа в целом, а затем его составных частей. Принципы робототехники EPSRC были созданы группой, собравшейся без особых указаний и без каких-либо требований. Первоначальным замыслом мероприятия по робототехнике epsrc, похоже, было только само обсуждение или, возможно, даже только сам факт встречи. Присутствующие ученые хотели что-то показать за потраченное время, и в результате значительное количество времени всех присутствующих в последний день ушло на создание трех версий принципов и их документацию. Часть документации был продлен — опять же на основе консенсуса — после собрания. Это правильно и уместно, что должен быть способ изучить и даже обновить или поддерживать документ. Даже национальные конституции имеют средства для поддержания. Однако для эффективности программных документов крайне важно, чтобы их было нелегко изменить. Они должны служить рулем направления для предотвращения сглаживания, и поэтому их обычно труднее изменить, чем создать экземпляр в первую очередь. Обратите внимание, что некоторым странам и другим политическим союзам было нелегко создать даже свои первоначальные конституции именно по этой причине. Поэтому важно тщательно обдумать значение принципов.

Принципы как политика

Технологическая политика и политика в целом — удивительно аморфная вещь. Как и другие аспекты естественного интеллекта, политика не всегда находится в рамках закона или даже управления. Большая часть политики не написана и даже не известна в явном виде. Великобритания на самом деле выдающаяся в своем новаторстве общего права, которое признает это, а также важность культуры и прецедента. Тем не менее, в холодном свете комитета, работающего над случаями воздействия REF, мы должны спросить, являются ли принципы политикой? Я думаю, что ответ "да". Они представляют собой набор руководящих принципов, согласованных значительной, хотя и произвольной, частью сообщества, на которое они влияют, и публикуются на правительственных веб-страницах. Любая политика состоит из трех компонентов: распределительного, распределительного и стабилизирующего. Распределение — это процесс определения того, на какие проблемы стоит тратить время и другие ресурсы. В случае с принципами это было спровоцировано EPSRC (или какой-либо организацией над ними) из-за опасений, что британская общественность может отвергнуть робототехнику, поскольку у них есть генетически модифицированная пища. Нам сказали, что отказ от робототехники рассматривается как серьезная угроза британской экономике. Обратите также внимание на то, что каждый из участников (по крайней мере, те, за участие в которых специально не платили) также вложил свои средства, посвятив время проблеме этики роботов, хотя для многих это смешивалось с возможностью стать более известными в своей основной финансирующей организации.

Стабилизирующий компонент — это тот компонент, который гарантирует, что однажды установленная политика будет внедрена в общество таким образом, что маловероятно, что она будет либо быстро отменена, либо станет значит

обязательство или спорный вопрос. В случае с принципами это, очевидно, было достигнуто, по крайней мере, на каком-то уровне, поскольку мы празднуем их пятую годовщину. Из разговоров с другими авторами я знаю, что ни один из них не был полностью очарован конечным продуктом, но все уважали (по общему признанию представительный) демократический процесс, с помощью которого они были достигнуты, и важность взаимной приверженности своих коллег конечному продукту. Я, например, хотел бы, чтобы принципы в дальнейшем воплощались в политику или даже в закон, но мне еще предстоит открыть процесс, с помощью которого это может быть достигнуто. Однако они привлекали и продолжают привлекать внимание различных советов по стандартам и парламентских расследований, а также прессы и других ученых. Я оставляю напоследок наиболее спорный аспект политики: распределение. В основе всей политики лежит выбор действий, а это подразумевает распределение или, скорее, перераспределение ресурсов. Политика пытается отмахнуться от этого, поскольку это обязательно идет вразрез с теми, от кого перераспределяются ресурсы, даже в тех случаях, когда эти люди получают чистую выгоду. Мы ненавидим терять контроль, но политика предназначена для контроля. «Пытается отмахнуться» — это на самом деле преуменьшение; сделать перераспределение приемлемым может быть основным проектом политиков. В этом случае у правительства были очень конкретные опасения по поводу лиц, которые в средствах массовой информации пропагандировали страх перед роботами, и очень четко выразили свое желание найти способы переключить внимание средств массовой информации и общественное мнение на безопасность робототехники. Напротив, на самом деле именно участники подняли другие важные сдвиги от сенсационности к прагматизму, утверждая, что роботы не являются ответственными сторонами по закону и что пользователей не следует обманывать относительно их возможностей. Представители совета знали, что такое перераспределение власти разозлит некоторых из их выдающихся получателей финансирования, и участники знали то же самое о некоторых из своих коллег. Тем не менее среди ученых существовало поразительное единодушие в том, что величайшей моральной опасностью роботов была их харизматическая природа и невероятное стремление многих людей вложить свою идентичность в машины, что привело к поразительному замешательству в отношении их природы, свидетелями которого были все мы. Эта харизма и замешательство оставляли открытой дверь для всевозможных манипуляций со стороны корпораций и правительств, где роботы могли быть назначены ответственными или даже суррогатными человеческими жизнями или ценностями.

Принцип убийства

Роботы — многоцелевые инструменты. Роботы не должны проектироваться исключительно или в первую очередь для убийства или причинения вреда людям, кроме как в интересах национальной безопасности.

Первые три принципа были задуманы как исправления законов Азимова. Роботы не являются ответственными сторонами, поэтому они не могут убивать. Вместо этого роботы не должны использоваться в качестве инструментов для убийства. Это простое правило сделало передачу моральной субъективности понятной и одновременно отвечало пацифистским желаниям большинства присутствующих. Однако с практической точки зрения роботы уже использовались в качестве оружия войны, и закон, который не имеет законной силы, вызывает сомнения. Нас убедили, что принцип, заведомо ложный, значительно снизит наши шансы на культурное влияние. Таким образом, смысл первого принципа может показаться нейтрализованным компромиссом исключения, но тот факт, что роботы не должны быть оружием в гражданском обществе, по-прежнему является важным социальным моментом. Помимо этого, значение имеет и тот факт, что практическая политика должна учитывать потребности правительства в вопросах безопасности и промышленности (Великобритания занимает пятое место в мире по торговле оружием). Как бы чисто академически некоторые из нас ни

Чтобы наша дисциплина была такой, тот факт, что многие из ее продуктов имеют немедленную пользу, означает, что мы не можем избежать влияния на наш мир.

Принцип соответствия

Люди, а не роботы, являются ответственными агентами. Роботы должны разрабатываться и эксплуатироваться настолько, насколько это практически возможно, в соответствии с существующими законами и основными правами и свободами, включая неприкосновенность частной жизни.

Второй закон Азимова имеет отношение к следованию инструкциям, но даже к понятию подчинения подразумевается моральная свобода воли. Первоначальный смысл этого закона заключался в том, что роботы являются обычной технологией и соответствуют обычным стандартам и законам. При формировании принципов как набора второй принцип стал тем, который еще больше сообщал об опасности ИИ в целом и ИИ, ошибочно принимаемом за моральный субъект в частности. Акцент на приватности отражает особую заботу воспринимающего разумного физического агента, занимающего точно такое же пространство, как и человеческая семья. Эта технология фундаментально погружена в умвельт человека больше, чем любая предыдущая технология или домашнее животное, возможно, даже больше, чем некоторые люди в домашнем хозяйстве, такие как дети. У него есть доступ к письменному и устному языку, социальной информации, наблюдаемому расписанию и т. д. Кроме того, его могут принять за домашнее животное или другого доверенного члена семьи, его особые способности для идеального общения с внешним миром временно забыты или его способности изучать закономерности. и классифицировать раздражители. В этих случаях информация о приматах может быть непреднамеренно сохранена в общедоступном облаке или даже в предположительно частном облаке, подверженном взлому. Привести такую новую, человекоподобную технологию в соответствие со стандартными юридическими нормами конфиденциальности и безопасности — нетривиальная задача.

Принцип товаризации

Роботы — это продукты. Они должны быть разработаны с использованием процессов, обеспечивающих их безопасность и защищенность.

Последний закон Азимова — самозащита, но у роботов нет личности. Вместо этого этот закон сосредоточился на защите людей от роботов на уровне базовой надежности робота. Этот принцип снова приводит нас к осознанию того, что робот не изготовлен специально, в попытке избежать юридической ответственности, заявляя, что роботы имеют уникальную природу. Производитель робота должен нести такую же ответственность за оборудование, работающее в соответствии со спецификацией, как и производитель автомобиля или электроинструмента. На самом деле, роботы могут быть автомобилями или электроинструментами, но если это так, то они должны быть более, чем менее безопасными, чем обычная разновидность того и другого.

Принцип прозрачности

Роботы — это искусственные артефакты. Они не должны быть разработаны таким образом, чтобы обманывать уязвимых пользователей, вместо этого их машинная природа должна быть прозрачной.

Первые три принципа установили правовую основу для производства и продажи робототехники как идентичной другим продуктам. Последние два предназначены для того, чтобы статус также сообщался пользователю. Принцип прозрачности направлен на то, чтобы люди не вкладывали чрезмерные средства в свои технологии, например, не нанимая сиделку, чтобы робот не оставался одиноким. Некоторые робототехники возражают против этого принципа, потому что обман необходим для эффективности их предполагаемого применения, например, чтобы заставить людей не чувствовать себя одинокими, чтобы они были менее подавлены. Другие утверждают, что этот принцип отрицает возможность того, что роботы должны быть чем-то большим, чем обычные машины. Первый аргумент открыт для эксперимента. Во-первых, необходимо установить, что невозможно вызвать эмоциональную вовлеченность без обмана, что кажется маловероятным, учитывая степень эмоциональной вовлеченности, которая устанавливается с вымышленными персонажами и явно незнакомыми объектами. Если требование обмана установлено экспериментально, то можно обсудить компромисс между издержками и выгодами обмана. Однако второе неоспоримо. Авторство, которое мы имеем над артефактами, является фундаментальной частью их машинной природы. ИИ по определению является артефактом. В какой-то степени можно даже утверждать, что этот принцип является самоограничивающим. Если бы ИИ действительно мог изменить то, что значит быть машиной, то передача этой измененной природы машины по-прежнему соответствовала бы этому принципу.

Принцип юридической ответственности

Должно быть указано лицо, несущее юридическую ответственность за робота.

Наконец, пятый принцип самым фундаментальным образом сообщает о статусе роботов как артефактов. Они находятся в собственности, и эта собственность должна быть юридически закреплена. Тот факт, что роботы созданы и ими владеют, является причиной, по которой я ранее утверждал, что мы этически обязаны не превращать их в людей, потому что владеть людьми, несомненно, неэтично. Аргумент состоит не в том, что существуют человекоподобные роботы, статус которых мы должны понизить на законных основаниях, а скорее в том, что обязательное понижение правового статуса означает, что мы не должны делать сходство с личностью чертой любого законно произведенного робота. Однако принципы робототехники не доходят до такой крайности футуризма. Как я уже говорил ранее, они сосредотачиваются на том, чтобы донести нынешнюю реальность до населения, которое так стремится владеть сверхчеловеком и отождествлять себя со сверхчеловеком, что его легко заставить поверить в то, что плохо изготовленный или плохо эксплуатируемый робот сам виноват в причиненном им ущербе. Если вы услышите ужасный шум и обнаружите, что машина врезалась в ваш дом, вы можете быстро и легко определить владельца машины, даже если машина в настоящее время пуста, просто по ее номерным знакам или, в худшем случае, по серийным номерам. Идея состоит в том, что то же самое должно быть верно, если вы найдете робота, встроенного в вашу собственность. Участники ретрита по робототехнике точно предсказали проблему, которая уже существует в нашем обществе из-за дронов, и проблему, которая сейчас решается в некоторых странах с обязательным лицензированием, как рекомендовал комитет.

Заключение

Подводя итог, можно сказать, что принципы EPSRC имеют ценность, поскольку они представляют собой политику, построенную на значительных затратах налогоплательщиков и частных лиц. Хотя ни одна политика не идеальна, в идеале они

должна быть заменена только новой политикой с эквивалентно высоким или более высоким уровнем инвестиций как правительства, так и экспертов в данной области. Их цель — обеспечить доверие потребителей и граждан к робототехнике как к надежной технологии, способной стать широко распространенной в нашем обществе. Каждый из отдельных принципов представляет существенную озабоченность экспертов и заинтересованных сторон, хотя иногда такое представление само по себе не совсем прозрачно. Общая цель заключалась в том, чтобы четко сообщить, что ответственность за безопасное и надежное производство и эксплуатацию роботов ничем не отличается от ответственности за любые другие объекты, производимые и продаваемые в Великобритании, и поэтому существующие законы страны должны быть адекватными как для потребителей, так и для производителей. .

Важно понимать, что это не относится ко всем мыслимым роботам. Легко представить себе уникальные произведения искусства, которые можно квалифицировать как роботов и которые не похожи на коммерческие продукты, или представить себе роботов, которые просто созданы небезопасным или безответственным образом. У людей больше проблем с осмыслением того, что могут быть когнитивные свойства, такие как страдание, которые, возможно, можно было бы включить в робота, но сделать это было бы так же неэтично, как поставить неисправные тормоза на машину. Принципы робототехники не стремятся определить, что возможно; они стремятся сообщить о рекомендуемых методах интеграции автономной робототехники в законодательство страны.

РобототехникаИсследованияЭтикаОбсуждение

А. Гнинг, Д. Дэвис, И. Ченг, П. Робинсон, факультет компьютерных наук,
Халлский университет.

e.gning@hull.ac.uk

Введение

В современном мире с развитием технологических ресурсов робототехника привела к многочисленным приложениям [1] и часто к непредсказуемому развертыванию в реальной жизни (например, все более широкое использование дронов в гражданских приложениях). Чтобы идти в ногу с этим веком робототехники, исследовательское сообщество и общество в целом должны определить этические принципы, которые были бы достаточно общими, чтобы быть устойчивыми к изменениям во времени и адаптированными к диапазону возможных применений.

Этические принципы должны быть настолько вульгаризированы и универсализированы, чтобы проектировщики и производители роботов осознавали правила и ограничения, которые необходимо соблюдать.

В целом приложения робототехники можно разделить на четыре группы: домашние роботы или вспомогательные роботы [2] [3] [4], медицинские роботы [5] [6] [7] [8], оборонные роботы [9] [10].] и промышленных роботов [11][12]. Обсуждение сосредоточено на первых трех группах роботов, поскольку промышленный робот часто ограничен ограниченными областями с заранее заданным набором ограниченных задач.

и не находятся в непосредственном взаимодействии с обществом.

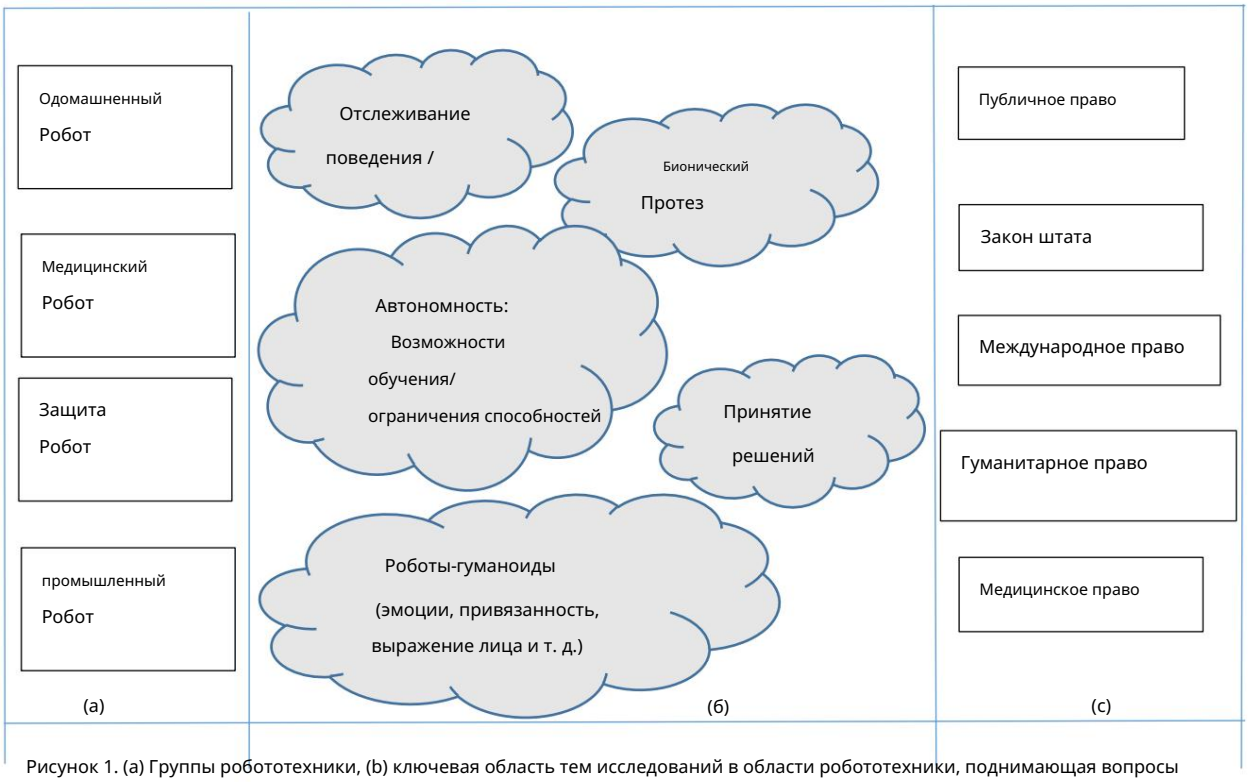

Рисунок 1. (a) Группы робототехники, (b) ключевая область тем исследований в области робототехники, поднимающая вопросы этики, (c) область правовых норм, которые необходимо учитывать.

На рисунке 1 показана сложная задача регулирования этики для различных групп робототехники в отношении пяти популярных тем исследований. В каждой группе робототехники (рис. 1-(a)) могут быть подняты этические вопросы, и представлены ключевые (рис. 1-(б)): как мы можем спроектировать роботов, чтобы они всегда были

способен интерпретировать их поведение, знать об ограничениях и границах обучения роботов; роботы-гуманоиды с имитацией человеческих жестов и анимацией лица могут вывести взаимодействие с людьми (особенно с детьми и людьми с ограниченными возможностями) на новые рубежи, повышая потребность в регулировании и прогнозировании возможных последствий; В настоящее время люди могут извлечь большую пользу из протезов благодаря прогрессу в исследованиях робототехники. Однако это может привести к тому, что люди, не являющиеся инвалидами, будут искать протезы, которые могут расширить их возможности, что снова вызовет необходимость в новых правилах. Наконец, самая большая проблема связана с возможностями принятия решений роботами. Эти решения непосредственно затрагивают человеческую жизнь и, следовательно, вызывают вопросы правового характера, вытекающие из совершенных действий.

Помимо этих исследовательских тем, этика робототехники должна быть совместима с рядом областей права (рис. 1 – (c)). Необходимо учитывать, что роботы должны быть совместимы с законами разных уровней, которые могут меняться, например, от региона к региону или от штата к

состояние.

Пять лет назад в Великобритании группа выдающихся экспертов в области робототехники и искусственного интеллекта опубликовала «Принципы робототехники» EPSRC в виде пяти правил и семи сообщений высокого уровня. Мы предлагаем обсудить эти правила с упором на поперечную структуру - между робототехникой группы, темы исследований и правовые рамки, представленные на рис. 1, - и в отношении трех критериев валидности, достаточности и полезности.

Обсуждение свода правил

Общий комментарий, который можно сделать по поводу набора из пяти правил, заключается в том, что довольно амбициозно думать, что можно дать общее/единообразное руководство для всех типов роботов, во всех возможных направлениях исследований и во всех правовых рамках. Было бы более естественно искать руководящие правила, отражающие трансверсальный характер робототехники, показанной на рис. во внимание особый аспект законов и направлений исследований, включенных в рисунок 1 (c) и (b).

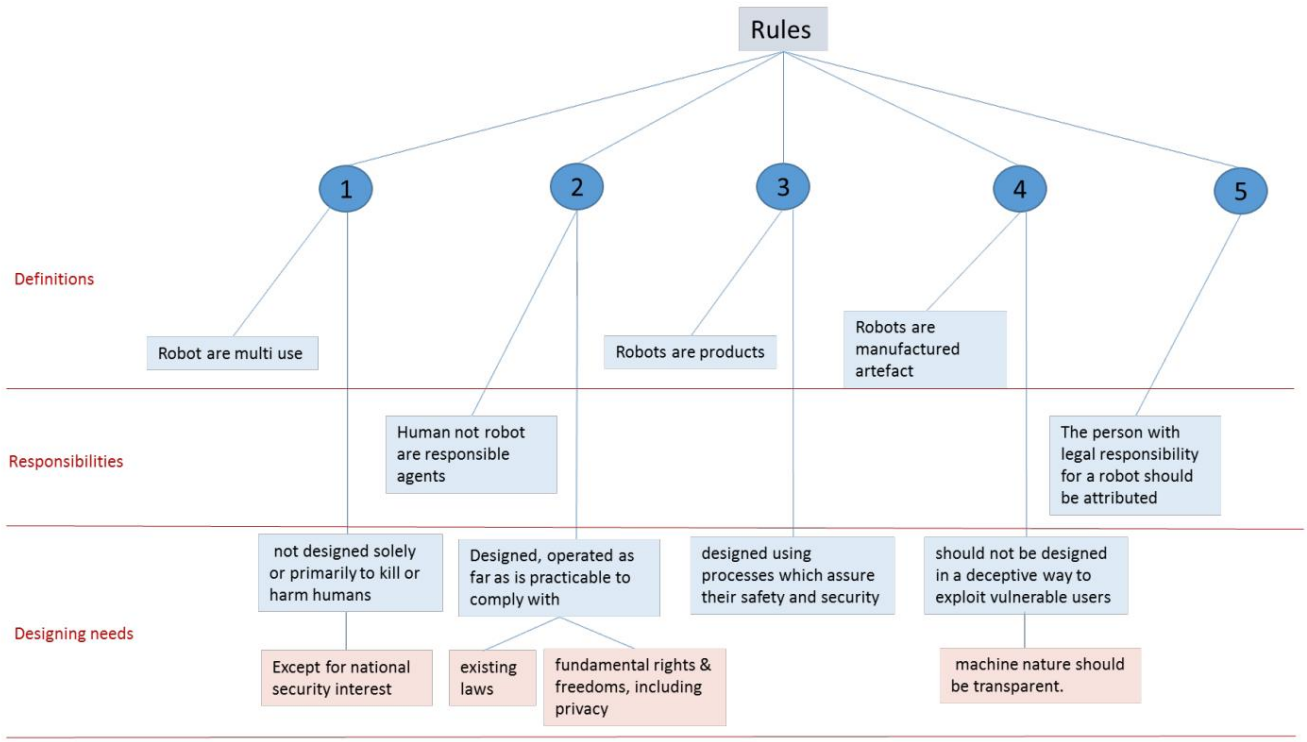

Рисунок 2. Набор из пяти правил

На рис. 2 графически представлен набор из пяти правил. Можно увидеть, что у пяти правил есть общая закономерность: первые предложения часто являются обобщениями и определениями.
о роботах, таких как «роботы — это продукты», или заявить, что ответственность лежит не на роботе, а на человеке. В последних предложениях изложены потребности проектирования и баланс между вопросами безопасности, юридическими вопросами и вопросами прозрачности.

Явно критические замечания по поводу структуры, показанной на рисунке 2, можно перечислить следующим образом.

- Пять правил иногда перекрываются. Например, «соблюдение действующего законодательства» в правиле 2 включает в себя «не предназначено исключительно или главным образом для убийства или причинения вреда человеку» в правиле 1; «лицо, несущее юридическую ответственность за робота, должно быть назначено» в правиле 5 может неявно заключать в себе «человек, а не робот, являются ответственными агентами» в правиле 2.
- Пять правил не являются общими. Например, мы едва видим, как бионические протезы могут быть вписаны в правила в нынешнем виде (например, исследования в области робототехники могут позволить в будущем модифицировать человеческое тело, чтобы получить больше силы, скорости и т. д.).
  Искусственная сексуальность — еще один пример противоречивого исследования, которое может привести к этическим вопросам.
- Пять правил не утверждают, что закон может быть противоречивым в зависимости от области, рассматриваемой советами, регионами, странами и континентами. Аналогичным примером, часто встречающимся в исследованиях, являются патентные заявки, для применения которых в определенных регионах мира требуется несколько конкретных исследований. Законы могут быть еще более сложными, поскольку религиозные убеждения, привычки и обычаи людей будут диктовать понятие этики.

Выводы

В этом обсуждении мы кратко представили аргументы в пользу необходимости другой формулировки пяти правил. Это демонстрируется графическим изображением пяти правил, которых на самом деле недостаточно, они частично совпадают и явно не отражают истинные проблемы этики робототехники.

Мы основывали часть наших рассуждений на трансверсальной природе робототехнической этики по трем направлениям: группы, составляющие робототехнику, будущие направления робототехники, которые необходимо учитывать при определении этики, и структурированный характер права. Мы рекомендуем естественную переформулировку, которая будет дифференцировать этику для каждой группы робототехники, будучи непреклонной в отношении противоречий и сильных ограничений, которые могут существовать из-за структурной природы права.

Библиография

[1] Г. Беки, Робототехника: состояние дел и будущие вызовы., Калифорния: London Imperial 2008.
Колледж Пресс. ,

[2] К. Даутенхан, С. Вудс, К. Каури, М. Л. Уолтерс, К. Л. Коай и И. Верри, Что такое робот-компаньон-друг, помощник или дворецкий?, IIC o. ИК а. Системы, изд., 2005.

[3] Дж. Форлицци и Д.К., Сервисные роботы в домашних условиях: исследование комнатного пылесоса в домашних условиях, 1-я конференция ACM SIGCHI/SIGART по взаимодействию человека и робота. AKM, 2006.

[4] JY Sung, RE Grinter, HI Christensen и L. Guo, Домохозяйки или технофилы?: понимание владельцев домашних роботов, 3-я Международная конференция ACM/IEEE по взаимодействию человека и робота (HRI), 2008 г., стр. 129-136.

[5] Г.П. Мустрис, С.К. Хиридис, К. Делипарасхос и К. Константинидис, Эволюция
автономные и полуавтономные роботизированные хирургические системы: обзор литературы,

об. 7, Международный журнал медицинской робототехники и компьютерной хирургии, 2011 г., стр. 375-392.

[6] Дж. Рассвейлер, Дж. Биндер и Т. Фреде, «Робототехника и телехирургия: изменят ли они нашу будущее?», т. 11, № 3, стр. 309-320, 2001.

[7] Г. Кваккель, КВЈ и КНІ, Влияние роботизированной терапии на верхнюю конечность. восстановление после инсульта: систематический обзор, Нейрореабилитация и восстановление нервной системы, 2007.

[8] К. Клири и К. Нгуен, «Современное состояние хирургической робототехники: клиническое применение и технологические проблемы», том. 6, нет. 6, стр. 312-328, 2001.

[9] П. У. Сингер, «Роботы на войне», Wilson Quarterly, 2008 г.

[10] Т.К. Адамс, «Война будущего и упадок принятия решений человеком», Параметры, об. 31, нет. 4, 2001.

[11] П. Лейтао, «Агентное распределенное управление производством: современный обзор». Инженерные приложения искусственного интеллекта, том. 22, нет. 7, стр. 979-991, 2009.

[12] ZM Bi, SY Lang, W. Shen и L. Wang, «Реконфигурируемые производственные системы: современное состояние», International Journal of Production Research, vol. 46, нет. 4, стр. 967-992, 2008.

Роботы не просто инструменты

Тони Дж. Прескотт, Шеффилдский университет

В основе принципов робототехники EPSRC (далее — «принципы») лежит ряд онтологических утверждений о природе роботов, которые служат аксиомами для построения последующего развития этических задач и правил. К ним относятся утверждения о том, что такое роботы, и
также о том, чем они не являются. Утверждения о том, чем роботы являются, включают в себя, что «роботы — это многоцелевые инструменты» (принцип 1), что «роботы — это продукты» (принцип 3) и «элементы технологии» (комментарий к принципу 3) и что «роботы — это искусственно созданные артефакты» (принцип 4). агенты» (принцип 2), что роботы «просто не люди» (комментарий к принципу 3), и что интеллект роботов может дать только «впечатление реального интеллекта» (комментарий к принципу 4).

При первом прочтении эти утверждения кажутся прямыми утверждениями очевидных истин. Я утверждаю, что это не так. Вместо этого я предполагаю, что эти онтологические обязательства лишены нюансов, они слишком легко предполагают, что мы знаем граничные условия будущего развития робототехники, и они затемняют или игнорируют некоторые важные антиэтические дебаты. Циклы можно было бы начать с тщательного обдумывания онтологического статуса роботов.

Если посмотреть на то, как представлены принципы, кажется, что здесь действует неявный процесс индукции, который позволяет интерпретировать утверждения о том, что представляет собой большинство современных роботов, как утверждения о том, чем, по существу, должны быть роботы. «роботы — это просто инструменты различного рода, хотя и очень специальные инструменты» в преамбуле. Хотя легко согласиться с общим утверждением, что роботы — это инструменты многократного использования, особенно в контексте обсуждения двойного использования (принцип 1), гораздо более сильное утверждение, что роботы — это просто инструменты или просто инструменты, отрицает, что они могут разумно принадлежать к другим разрозненным категориям.

Возьмем, к примеру, категорию «компаньон». Существует большая работа по разработке роботов-компаньонов, которые могут оказывать социальную и эмоциональную поддержку людям, что частично признано при обсуждении принципа 4. Категория «инструменты» описывает физические/механические объекты, выполняющие определенную функцию, а категория «компаньоны» описывает значимых других, обычно людей или животных, с которыми у вас могут быть взаимные отношения. Возможность того, что роботы могут принадлежать к обеим этим категориям, поднимает важные и интересные проблемы, которые затемняются утверждением, что роботы — это всего лишь инструменты.

Действительно, в соответствии с представлением о роботах как о инструментах, обсуждение роботов-компаньонов в принципах довольно пренебрежительно, описывая мачты как игрушки, которые могут доставить некоторое удовольствие людям, которые не могут или не могут позволить себе содержать домашних животных.

Утверждается, что искусственная природа роботов-компаньонов создает реальную этическую проблему, поскольку роботы-компаньоны потенциально обманчивы и поэтому должны быть спроектированы так, чтобы их «машинная природа была прозрачной».

Онтологическая проблема здесь, в частности, касается утверждения, что роботы никогда не могут обладать психологическими способностями, такими как «настоящие» эмоции или интеллект. Что это такое, с точки зрения человека, горячо обсуждается в когнитивных науках и науках о мозге.

Действительно, существуют встречные утверждения о том, что роботы, соответствующим образом сконфигурированные, могут иметь эмоции[1], в то время как будущее искусственного интеллекта как интеллекта не имеет очевидного потолка на уровне ниже человеческого.

Еще одна проблема связана с предположением о том, как люди будут видеть роботов, а именно, что роботы будут рассматриваться как инструменты, если они показаны прозрачным образом.
например, люди могут антропоморфизовать роботов независимо от того, насколько очевидно, что они являются искусственными продуктами. Одной из причин думать, что это может быть так, является ярко выраженная социальная природа нашего мозга и то, насколько легко наша эмпатия вызывается чем-то, что кажется живым. Анимация Хайдера-Симмела простых геометрических фигур [2] (см. рисунок) показывает, насколько грубой может быть эта информация, и тем не менее мы все равно увидим намеренность, мотивацию и даже эмоции. Изобретение цифрового питомца Тамагочи продемонстрировало, что простая 2-анимация животного, похожего на существо, может вызвать непреодолимое желание заботиться [3]. факт реален для того, чтобы иметь подлинную эмоциональную реакцию на это сами.

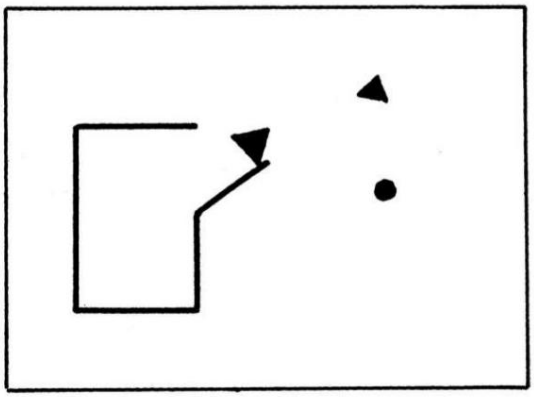

Рисунок. Геометрические фигуры, движущиеся в простой анимации, были интерпретированы как «одушевленные существа, главные лица» в этом знаменитом исследовании 1944 года, проведенном Хейдером и Зиммелем.

Анализ онтологических и психологических проблем взаимодействия человека и робота ранее было сделано Канандом с коллегами [4]. Следуя аналогичному ходу мысли, мы можем описать всегда могут сочетаться четыре общих взгляда на то, что такое роботы, и психологический взгляд на то, как их видят роботы. Они проиллюстрированы в следующей таблице.
наряду с некоторыми этическими проблемами, которые они влекут за собой.

I. Роботы — это всего лишь инструменты (о), и люди будут считать роботов всего лишь инструментами, если их не введет в заблуждение обманчивая конструкция роботов (р). Этические вопросы. Мы должны учитывать ответственность людей как производителей/пользователей роботов и риск обмана при создании роботов, которые кажутся чем-то, чем они не являются. Это позиция «принципов».	II. Роботы — всего лишь инструменты (о), но люди могут видеть, что месиво обладает значительными психологическими способностями независимо от прозрачности их машинная природа (р). Этические вопросы: мы должны принимать во внимание то, как люди видят роботов, например, что они могут чувствовать, что у них есть значимые и ценные отношения с роботами, или они могут видеть, что роботы обладают важными внутренними состояниями, такими как способность страдать, несмотря на то, что они не обладают такими способностями.
III. Роботы могут обладать значительными психологическими способностями (о), но люди по-прежнему будут считать их просто инструментами (р). Этические вопросы: мы должны анализировать риск обращения с сущностями, которые могут обладать значительными психологическими способностями, такими как способность страдать, как если бы они были всего лишь инструментами, и опасности, связанные с созданием нового класса сущностей со значительными психологическими способностями, такими как человеческий интеллект, не осознавая, что это утомительно.	IV. Роботы могут обладать некоторыми значительными психологическими способностями, подобными человеческим (о), и люди увидят затор, имеющий такую емкость(р). Этические вопросы. Мы должны рассмотреть сценарии, в которых люди должны будут сосуществовать вместе с новыми видами психологически значимых сущностей в форме будущих роботов/ИИ.

Таблица. Как можно сочетать онтологическую(о) и психологическую(р) точки зрения на роботов (по Kahnetal., 2007).

Обратите внимание, что только один квадрант этой таблицы (I) рассматривается в принципах, а все II, III и IV возможны, по крайней мере, теоретически.

Inquadrant II возникают интересные вопросы о том, как следует обращаться с роботами — не потому, что они являются разумными агентами, а потому, что люди захотят относиться к ним как к таковым. Владельцы собак-роботов Sony Aibo [6] не кажутся странными, если смотреть с точки зрения того, как люди видят роботов, а не с точки зрения того, чем они являются. с некоторыми роботами, которые могут быть похожи на тех, которые мы разрабатываем с другими ценными вещами, Например, автомобили и мобильные телефоны. С другой стороны, для некоторых роботов они могут быть больше похожи на отношения, которые у нас есть с домашними животными, в том числе, например, желание поддерживать и воспитывать их (что-то, что мы сами можем найти полезным). человек, а потому, что обладает способностью помнить и сообщать мне о каком-то из наших общих переживаний. В более общем плане для разработки подходящих этических принципов может потребоваться разработка классификации различных форм эмоциональных связей, которые могут существовать между роботами и людьми, и анализ факторов, которые могут лежать в основе развития и поддержания таких отношений [7].

Квадрант III касается возможности того, что роботы обладают значительными психологическими способностями, которые могут быть упущены из виду людьми. Это повышает этические риски, которые не обсуждаются в принципах, но подчеркиваются другими. Это новый вид разумных существ, которые без необходимости страдают из-за наших действий, и это явно проблематично с этической точки зрения, если бы это произошло. Хотя это может показаться весьма вероятным в краткосрочной перспективе, есть основания полагать, что в среднесрочной и долгосрочной перспективе это может быть связано с усложнением когнитивных архитектур роботов. Несколько тенденций в продолжающихся исследованиях человеческого сознания также подтверждают эту возможность. Во-первых, одна из основных современных теорий сознания [9] утверждает важнейшую роль интеграции информации
это не обязательно требует абиологического субстрата. Неврологи также серьезно оценивают, могут ли островки интегрированной активности в мозге «запертых» пациентов представлять собой форму минимального сознания [10].
Меньший мозг, чем наш, такой как рыба, может быть разумным в значительной степени (например, что они могут опыт боли) [11]. Эти разработки предполагают, что сознание может быть возможно в искусственном агенте, не имеющем соответствия по размеру или сложности нетронутому человеческому мозгу.
[12], а Брайсон [13] предположил, что у современных роботов уже могут быть некоторые простые формы сознания, отвечающие некоторым обычно предлагаемым критериям.

Одним из следствий взгляда на роботов как на «всего лишь инструменты» является неявное отрицание возможности сильного ИИ — того, что будущие роботы могут обладать общим интеллектом на уровне человека или выше человеческого уровня. Проблема Aquadrant III/IV, недавно обсуждавшаяся известными учеными и новаторами, такими как Стивен Хокинг, Илон Маск и Билл Гейтс, среди прочих, и подробно проанализированная Bostrom. [14], может ли ИИ-сингулярность обратить вспять отношения хозяин-раб между людьми и роботами. Убеждение, что роботы/ИИ являются «просто инструментами», может помешать нам распознать самозагружающийся супер-ИИ. Этический подход, несомненно, поощрил бы большую бдительно

Дебаты — это точка зрения «глобального мозга», предложенная Хейлигеном [15] и другими, согласно которой люди и продвинутый ИИ могут сосуществовать для нашей взаимной выгоды. Это напоминает нам, что этика должна быть направлена на анализ потенциальных выгод, а также рисков.

Хотя сценарии квадрантов III/IV могут показаться надуманными или, по крайней мере, отдаленными, такие опасения захватили общественное воображение и вызвали серьезные призывы к обсуждению (например, [16]). По моему собственному опыту общения с представителями общественности и средств массовой информации, эти темы часто вызывают наибольший интерес и озабоченность. настаивая на том, что роботы — это всего лишь инструменты, мало что может сделать, чтобы успокоить голоса, и они могут показаться гегемонистскими и снисходительными.

Более откровенный подход может заключаться в том, чтобы признать, что, хотя большинство роботов в настоящее время представляют собой не более чем инструменты, вступает в эпоху, когда появятся новые виды сущностей, сочетающих некоторые свойства машин и инструментов с психологическими способностями, которые, как мы ранее считали, зарезервированы для сложных биологических организмов, таких как люди. как лиминальные — не живущие совершенно так же, как биологические организмы, и не просто механические, как с традиционной машиной.

Лиминальность роботов делает их одновременно завораживающими и пугающими по своей сути, а также громоотводом для наших более широких опасений по поводу бесчеловечного воздействия технологий[18].

Ассоциация ученых Манхэттена писала в 1945 г. [19] о своем чувстве коллективной ответственности за свою роль в разработке технологии с «потенциалом великой брони или великого блага» (атомной энергии) и об их «особой осведомленности», что это может привести к «продвижению нашей цивилизации или ее полному уничтожению». также несут особую ответственность за то, чтобы понимать и открыто говорить о том, что может принести будущее робототехники, а также о ее потенциальных преимуществах и угрозах.

Рекомендации

1. Феллоус, Дж.-М., От человеческих эмоций к эмоциям, в Весеннем симпозиуме AAAIS по архитектуре моделирования эмоций: междисциплинарные основы Э.Худлика и Л.Каньямеро, редакторы. 2004, AAAIPress: MenloPark, CA.стр. 37-47.
2. Хайдер, Ф. и М. Зиммель, Экспериментальное исследование кажущегося поведения. Американский журнал психологии, 1944. 57 (2): стр. 243-259.
3. Леви Д., Любовь и секс с роботами. 2007, Лондон: HarperCollins.
4. Кан, Дж., Питер Х., и др., WhatisaHuman?: К психологическим ориентирам в области человеческого–взаимодействие с роботом. InteractionStudies, 2007.8(3): стр. 363-390.
5. Ловгрен С., Этический кодекс роботов по предотвращению злоупотреблений Android и защите людей, в NationalGeographicNews. 2007.
6. Браун А., Tomournaroboticdogistobetrulyhuman в Guardian. 2015: Манчестер.
7. Коллинз, Э.К., А.Миллингс и П.Т.Дж. Приложение к вспомогательной технологии: новая концептуализация. в AssistiveTechnology:FromResearchtoPractice:AAATE2013. 2013.
8. Метцингер Т., Туннель эго: наука о разуме и миф о себе. 2009, Нью-Йорк: Бейсикбукс.
9. Тонони, Г., Сознание как интегрированная информация: предварительный манифест. Биологический бюллетень, 2008. 215 (3): стр. 216-242.
10. Цю Дж., Исследование островов сознания в поврежденном мозгу. TheLancetNeurology.6(11):стр.946-947.
11. Сет, А. К., Почему у рыбной боли нельзя и не следует исключать AnimalSentience, 2016.2016.020.
12. Деннет Д., Практические требования к созданию сознательного робота. Философские труды Королевского общества Лондона, 1994. 349: стр. 133-146.

13. Брайсон, Джей Джей Груд, Сырный, Сознание второго сорта. в ВенеКонференция по сознанию. 2008.
14. Бостром Н., Сверхразум: пути, опасности, стратегии. 2014, Оксфорд: OxfordUniversityPress.
15. Хейлиген, Ф., TheGlobalBrainasaNewUtopia, in Zukunftsfiguren, R.Mareschand F.Rö tzer, Editors.2002, Suhrkamp: Франкфурт.
16. Институт будущего жизни. AnOpenLetter: Research Priorities for Robust BeneficialArtificial Intelligence. 2015; доступно по адресу: http://futureoflife.org/ai-open-letter/.
17. Канг, М., Возвышенные мечты о живых машинах: автоматы в европейском воображении. 2011, Кембридж, Массачусетс: HarvardUniversityPress.
18. Шоллози, М., Фрейд, Франкенштейн и наш страх перед роботами: проекция в нашем культурном восприятии технологий. AI&ОБЩЕСТВО,2016:с.1-7.
19. Ассоциация манхэттенских ученых. Предварительное заявление. 1945; Доступно по адресу: https://www.gilderlehrman.org/history-by-era/postwar-politics-and-origins-cold war/resources/physicists-predict-nuclear-arms-race-.

Семинар AISB по принципам робототехники EPSRC

Защищать устаревший гуманизм(изм)?

Майкл Соллоси, SheffieldRobotics

Введение

Семинар EPSRC 2010 г. по разработке принципов ответственной разработки и исследования роботов был важным, амбициозным и благонамеренным проектом. использоваться как в конкретном правовом контексте, так и для широкой аудитории.

Из обзора принципов ESPRCP становится ясно, что они могут и должны выполнять жизненно важные функции в защита людей от безответственных или просто бездумных исследований технологий, которые предположительно могут иметь очень реальные, очень негативные последствия для человечества на личном, общественном уровне или даже на уровне всего вида.

Тем не менее, также ясно, что то, что защищается Принципами ESPRC, — это очень конкретный человек или, по крайней мере, очень конкретное представление о том, что составляет человеческое существо. в мастерскойделал так.

(Однако я утверждаю, что должна была быть преамбула с изложением этих предположений.)
Руководящие принципы EPSRC предназначены для защиты людей и обеспечения того, чтобы исследования робототехники проводились для «максимальной пользы от падения граждан», хотя то, как именно могут выглядеть эти «граждане», остается без ответа и означает, что, какими бы похвальными ни были их намерения, Эти Принципы во многом уже являются документом, уходящим корнями в конкретный исторический и культурный контекст, и это делает маловероятным, что эти Принципы в их нынешнем виде сохранятся в среднесрочной или долгосрочной перспективе.

Принципы ESPRC делают определенные, очень конкретные, но совершенно невысказанные предположения относительно того, что представляет собой «человек». И Принципы очень хорошо подходят для человека. людьми -мастерами. Люди имеют право проектировать, строить и покупать роботов и должны нести за них полную юридическую ответственность.

Принципов в том виде, в каком они сформулированы в настоящее время, вероятно, будет достаточно в краткосрочной, а может быть, даже в среднесрочной перспективе, для решения большинства проблем с помощью новых технологий, возникающих из робототехники и компьютерных лабораторий по всей Великобритании.
преходящее существо, относительно новое изобретение, существо, более того, которое в настоящее время не является ни внутренне непротиворечивым, ни единым, и будет постоянно переделываться и трансформироваться целым рядом новых технологий.

но также и проще, новое как о себе, обществе и новых способах осмысления нашего места в мире, изменениях, которые могут быть вызваны не только робототехниками, генетиками, учеными-компьютерщиками, но и изменениями в нашей экономической, политической и социальной жизни в более общем плане. Человек – или человеческие существа – в Принципах ESPRC не будет ни первым, ни последним выражением того, что значит быть человеком.

Какой человек?

В основе (имплицитно) принципа ESPRCP лежит конкретное человеческое существо, определяемое на протяжении последних столетий тем, что стало известно как гуманизм. Это человеческое существо является самостоятельным агентом, существом независимым и неподвластным другим, метафизическим или сверхъестественным силам. Это человеческое существо находится в центре европейских правовых, этических, экономических и политических систем; однако важно помнить, что 1. это человеческое существо все еще является относительно новым изобретением и что 2. на протяжении всей его жизни никогда не существовало ни одной единственной версии этого человеческого существа, как это любят представлять сторонники гуманизма.

Существует мало единого мнения относительно рождения человека: одни утверждают, что это был Ренессанс, другие — Просвещение, а третьи до сих пор считают, что гуманистический субъект стал центральным в том, как мы думаем о себе, только в девятнадцатом веке. Если рассуждать о нас самих, то несомненно, что это осмысление человека есть изобретение, а не данность; гуманистический человек не есть «естественное» или даже «правильное» истолкование нашей человеческой природы. вещи, которыми мы пользуемся, и окружающая среда, в которой мы живем.

И когда мы таким образом контекстуализируем гуманизм и рассматриваем его как историческое, отмечая, насколько разные представления о том, что значит быть человеком (или даже «гуманистом»), радикально изменились на протяжении столетий, становится очевидным, что речь идет не просто об одном человеке или об одном представлении о том, что значит быть человеком. есть свидетельства того, что речь идет не об одном человеческом существе, а о многих человеческих существах, не о единой, неотъемлемой, самоочевидной человеческой природе, а о людях, меняющих представление о себе в конкретных контекстах. Человеческое понимание самих себя всегда условно и контекстуально . Каждый человек может быть членом общества, гражданином, специалистом в различных контекстах, потребителем, производителем; мы можем быть преступниками, пациентами, клиентами, налогоплательщиками, заинтересованными сторонами, студентами, рабочими или менеджерами или всем этим сразу или ни одним из них, в зависимости от контекста. 'ора'ман'ора' Женщина сегодня сильно отличается от той, что была двести, или даже пятьдесят, или даже десять лет назад. Мы можем подвергаться постоянно меняющемуся дискурсу о праве, медицине, образовании, политике, экономике, философии, промышленности, средствах массовой информации и множеству других систем, языков и институтов, которые пытаются определять и понимать нас слегка отличающимися или радикально отличающимися способами.

Предположения, лежащие в основе нашего представления об одиночном представлении о том, что значит быть «человеком», несостоятельны в условиях пристального изучения новых способов мышления о себе, а также в связи с тем, что новые технологии заставляют нас думать о себе по-другому.

палки, чтобы помочь охотнику, летающий челнок изменил способ изготовления ткани во время промышленной революции.
более сложно интегрированы со своими инструментами, поскольку палки становятся протезами, а человеческий труд полностью заменяется автоматизированными машинами. И эти разработки создадут новых людей и новые способы мышления о самих себе.

Однако технологии не всегда проявляются в физических объектах; технологические достижения не всегда принимают форму новых инструментов или машин: изобретение законов и правовой системы было новыми технологиями, которые оказали огромное влияние на то, как мы создаем нашу человеческую и социальную сущность, точно так же, как изобретение научного метода, новые производственные отношения или Facebook изменили наше представление о себе. , и представляем эту идею о себе миру. Наши технологии двадцать первого века, в том числе более совершенные роботы и ИИ, а также изменение правовых систем, политических органов и систем для этической жизни, станут дальнейшим развитием, которое со временем радикально изменит то, как мы видим себя и понимаем каждое представление о том, что значит быть человеком.

Какими неявными, невысказанными способами, которые нам необходимо специально изучить, потворствует ли Принцип ESPRC этим гуманистическим предположениям о природе человека? Все подобные документы — конституции, хартии, договоры или декларации о принципах — должны с самого начала ясно излагать те истины, которые считаются самоочевидными, основой того, чему следует следовать.

В отсутствие четко определенного предмета «Принципы» предлагают обычную, знакомую гуманистическую концепцию человека — статичного, однородного человеческого существа, которое очень скоро будет сделано. устарела, если уже не устарела, благодаря всем технологическим достижениям, которые она пытается контролировать. Существует чрезвычайно упрощенное представление об отношениях между людьми и их орудиями: односторонние отношения, при которых орудия всегда являются слугами своих хозяев-людей и всегда находятся под контролем независимого человеческого агента. Такие отношения между субъектом и объектом, активным агентом и пассивным объектом всегда были бы наивными. Житель сначала схватился за палку, но как наши новые технологии требуют фундаментальной реорганизации всего нашего образа жизни, и настоял на том, как понимается вся наша социальная структура.

Сказать, что наши отношения с нашими инструментами не являются простыми односторонними отношениями хозяин-слуга, не значит сказать, что наши инструменты являются нашими хозяевами. Однако мы должны признать, что люди в значительной степени являются продуктом нашего способа создания вещей, поскольку то, что мы делаем, и то, как мы это делаем, решается людьми. (На самом деле в этом нет ничего спорного; это то, что Маркс признал более ста пятидесяти лет назад, объясняя, как общественный способ производства определяли его социальные отношения и то, как отдельные люди в свою очередь определялись этими социальными отношениями.) Отношения между людьми и их орудиями всегда были более сложными, чем это представлялось гуманизмом и в этих Принципах; люди являются «рабочими», а грань между «биологическим» и «машинным» становится еще более размытой.

Таким образом, Принципы, несмотря на благородные намерения, пытаются овладеть будущим и превратить технологию в устаревшее человеческое существо.

Представленная в «Принципах» концепция человеческого существа также разделяет с гуманизмом иллюзию предложения как единого, однородного субъекта, тогда как на самом деле этот предмет представляет собой совокупность множества — часто противоречащих друг другу — существ. разные культуры по всему миру или даже разные культуры в одном и том же сообществе.

Маловероятно, что прогресс в области робототехники и ИИ принесет пользу всем сообществам и всем народам в равной степени, особенно в краткосрочной и среднесрочной перспективе, и принципы развития робототехники должны это признавать.

Также неясно, какие лица упоминаются в документе; люди по-разному называются «общественностью», опять же, как если бы это было какое-то единое, однородное тело.

- Принципы относятся к «гражданам», как субъектам определенного политического (обычно национального) органа, хотя неясно, может ли кто-либо еще претендовать на звание «гражданина» скромного, независимого национального государства, независимого от других влияний. Будут ли выгоды и ответственность за роботов распространяться только на граждан определенного национального государства или политического органа? (Возможно, это также усложняется, поскольку повышенная автоматизация приводит к внедрению Универсального Базового Дохода в конкретном национальном государстве, но не где-либо еще.) Также интересно, что Принципы поддерживают это (устаревшее) понятие национального государства с учетом того, что робот может быть спроектирован как робот по «соображениям национальной безопасности».
- Принципы твердо заявляют, что только люди являются «ответственными агентами закона». В настоящее время это может не вызывать споров, и нужно обратиться к области научной фантастики, чтобы представить, когда мы могли бы иметь разумных, сознательных роботов и ИИ, которые были бы равны людям в глазах закона, но эта декларация игнорирует воду, уже замутненную автономными системами, такими как беспилотные автомобили. s, и вызовы, которые они ставят перед нашей гуманистической правовой системой. Кроме того, мы можем задаться вопросом, как технологически усовершенствованные люди (например, киборги) могут рассматриваться в законе в равной (меньше? больше?) степени ответственными агентами.
- Принципы настаивают на соображении конфиденциальности, хотя мы уже можем видеть, что для у многих людей границы «я» и «общественности» размыты, и понятие частной жизни были радикально изменены за такой короткий промежуток времени. Социальные сети, обещание «умных домов» и вопросы безопасности привели к тому, что в культурном отношении у нас совершенно разные представления о том, что означает «конфиденциальность» и насколько это важно.

- В Принципах четко проводится различие между теми, кто «проектирует» роботов, и теми, кто продает роботов, и теми «потребителями» и «пользователями». Принципы неявно признают, что интересы этих групп могут конкурировать. Этот аргумент уже устарел (ср., например, Пол Мейсон, 2015). Уже очевидно, что по мере развития робототехники и ИИ эти, казалось бы, стабильные социальные отношения будут подвергаться все большему напряжению и, вероятно, будут трансформированы во что-то более подходящее для новых возможностей создания вещей и более эффективных способов организации общества (например, посткапитализма, как некоторые считают) .Также уже стираются границы между производителями и продавцами, с одной стороны, и потребителями и пользователями, с другой.

> другие (например, Uber, данные краудсорсинга, Google), эти категории уже должны быть гораздо более гибкими, чем они представлялись в прямолинейном, упрощенном гуманизме.

Стоит также отметить, что принципы робототехники ESPRCP, что неудивительно, возможно, в значительной степени европейская, христианская выдумка. Роботы считаются машинами и, следовательно, просто объектами . это неосязаемое, метафизическое свойство, уникальное для жизни или, в самом начале, уникальное только для людей. (Эта идея об отсутствии чего-то жизненно важного для человека кроется в самой идее робота, когда это слово впервые было представлено миру в пьесе Карла Капека 1921 года, RUR) . Хотя можно утверждать, что Европа больше не обязана христианству, европейские (и Европейские правовые и этические рамки, включая эти Принципы ESPRCP.

В качестве контраста стоит отметить, как и многие (например, Metzler and Lewis, 2008; Lee, Sung, Š abanović , Han, 2012), насколько по-разному роботы и ИИ воспринимаются в разных культурных контекстах. Например, в Японии можно увидеть совершенно разные предполагаемые отношения между людьми и роботами. в Японии, и обе придерживаются «анимистических» религий, то есть они верят, что все вещи, включая неодушевленные объекты, содержат природу ками, или духа . Такое влияние, столь глубоко укоренившееся, с меньшей вероятностью будет слишком легко трансформировано внедрением новых технологий и идей, но оно подчеркивает, что принципы ESPRCP очень тесно связаны с очень специфическим культурным и историческим контекстом, и как мы должны быть готовы и желать вообразить другие идеи и отношения не только в будущем, но и прямо сейчас, если мы попытаемся построить международный конфликт. сенсорные принципы робототехники.

Новые люди?

То, что «Принципы» предназначены не для жестких законов, а скорее для информирования дебатов и для использования в будущем, демонстрирует дальновидность делегатов семинара, но «Принципы» должны быть ограничены, чтобы позволить движение за пределы узкого понятия «человека», лежащего в основе их в их нынешнем состоянии. — и — быстро ограничить то, что может быть постигнуто, потому что эта идея «человека» определяет все отношения, которые в ней воображаются. Вместо этого необходимо вообразить в основе другого, более плюралистического и гибкого человека.

Мыслители могут спорить (и часто до тошноты) о том, когда рухнул консенсус, поддерживающий гуманистическую тематику, но ясно, что в какой-то момент после Второй мировой войны, с утратой веры в метанарративы и новой, радикальной герменевтикой подозрений (которую некоторые стали понимать как «постмодернизм»), устойчивая гуманистическая тематика, как она когда-то была понята, недолго оставалась в этом мире. Следующее — и ясно, что что-то обязательно должно произойти дальше, ибо мы не можем приступить к конструированию каких-либо рамок или моделей без какого-либо представления о том, что значит быть человеком, — здесь нет большого согласия. разным образом и в разное время – люди

как потребители, люди как производители, люди как дизайнеры, как юридические субъекты, граждане и субъекты различных политических образований... Любой из этих людей может однажды победить любого.

Мы можем стремиться воссоздать некоторые принципы робототехники на основе человеческого субъекта, следующего за гуманизмом. Мы можем хотеть называть это человеческое существо постчеловеком или просто постчеловеком. Однако эти термины сложны и относятся к головокружительному набору различных идей и идеологий (даже большему, чем содержалось под зонтичным термином «гуманизм», предшествовавшим ему). рецепты технологических инноваций.

Постгуманизм может просто означать, философски, культурно, то, что следует за гуманизмом; этот постгуманизм, иногда антигуманизм, опровергает своего рода устойчивые, единичные предположения о человеке и человеческой природе, выдвигаемые гуманизмом.
принимает случайность и контекст концепции человека и заменяет статичную человеческую природу чем-то более динамичным и плюралистическим. Кроме того, многие из концепций постгуманизма включают соображения о том, как новые технологические разработки должны быть включены в человеческий опыт, преобразовывая как человека, так и наш мир.

Постгуманизм, или, может быть, точнее, постгуманизмы, не телеологичны, они не исходят из того, что человек, к которому мы пришли после миллионов лет эволюции и тысячелетий философии, – мы – есть человек , конечный, законченный, отшлифованный продукт, который отныне навсегда останется неизменным и неизменным. такими, какими они были всегда. Столь великая сила постгуманизма, как это понимается и формулируется здесь, заключается в том, что существует встроенная гибкость, позволяющая приспосабливаться к таким изменениям, и важно, чтобы любое амбициозное предприятие, такое как изложение принципов для определяя наши нынешние и будущие отношения с постоянно меняющейся технологией, обладают аналогичной встроенной гибкостью.

Некоторые, особенно оптимистично настроенные в отношении скорого появления разумного ИИ и называющие себя трансгуманистами, могут счесть принципы ESPRCP наивно-антропоцентричными, поскольку они не учитывают появление роботов и ИИ как разумных агентов по праву, которые заслуживают (возможно, равного) внимания наравне с людьми при создании каких-либо этических принципов. Я разделяю с тем, что я продвигаю здесь, убеждение в том, что принципы ESPRCP уже несколько устарели и слишком ошибочны в их представлении о том, что представляет собой «человек», хотя я гораздо менее оптимистичен в отношении неизбежности разумного ИИ и не разделяю общей трансгуманистической уверенности в том, что люди радикально трансформируются с помощью технологий (например, люди, которые
почти бессмертны) так же очень почти. Однако нет необходимости в появлении определенного разумного ИИ, чтобы эта постгуманистическая критика Принципов ESPRAC оставалась в силе. Даже если мы не будем изобретать новых роботов и делать новые шаги в области искусственного интеллекта — что очень маловероятно — это почти наверняка что мы, люди, будем продолжать изобретать другие системы и институты, которые определяют, кто мы, тем самым трансформируя людей и вызывая необходимость в новом, более гибком наборе принципов для определения наших отношений с роботами и ИИ.

Авторы «Принципов» предполагают, что они станут «живым документом», а не «жесткими законами», а основой для будущих дискуссий и ссылок, что и является именно тем, чем это должно быть.

наши новые технологии, когда он берет в качестве отправной точки такой жесткий гидрант, уже устаревший человеческий объект в своем сердце.

Интересно отметить, что в преамбуле к «Принципам» упоминается вездесущность Азимова и его трех законов. Поскольку, хотя законы Азимова были отброшены — правильно — как неадекватные, поскольку их вымысел и не имеют отношения к «реальной жизни» и не могут быть использованы на практике, тем не менее в трудах Азимова есть что-то, что могло вдохновить ESPRC. : способность воображать разные миры, населенные разными видами людей. Человеческие существа постоянно находятся в процессе переизобретения, но с опережением в области робототехники и ИИ, которые, вероятно, не за горами, мы можем предположить, что находимся на грани еще более радикальной трансформации в том, как мы видим себя и как относимся к нашим технологиям. Поэтому абсолютно жизненно важно, чтобы мы стремились создать Принципы робототехники, которые будут способны приспосабливаться к этим изменяющимся отношениям, а затем и ограничивать различные направления развития. .Нам нужно будет творчески подумать о видах роботов, которых мы создадим , а также о типах людей, которыми мы станем, и если мы стремимся разработать принципы на благо наших обществ, нам нужно лучше понимать, как будут выглядеть эти общества и люди, населяющие их.

Рекомендации

Ли, Сун, Шабанович, Хан. 2012. Культурный дизайн домашних роботов: исследование ожиданий пользователей в Корея и Соединенные Штаты.

Мейсон П.2015. Посткапитализм: руководство к нашему будущему. Лондон: Алленлейн.

MetzlerandLewis. 2008. Этические взгляды, религиозные взгляды и принятие роботизированных приложений: пилотный проект. исследование.AAAI.15–22.

Регулирование!Робот!Города:!! Размышления!о!принципах!робототехники с новой!стороны!закона

Аврора и Войкулеску

Центр&для&Права&&&Теории,&Университет&оф&Вестминстера

«Они&спрашивали&меня&где&я&выбирал&&бег,&которое&выбирало?&Взлеты?&Или&Внизы?
Где&роботы&мыши&и&люди,&я&сказал,&бегите&круг&в&роботов&городах.
Но&это&мудро?&Для&железного&&глупого и&железного&не&думала!
Компьютер&мыши&может&найти&меня&факты&и&обучить&меня&чему&я&я&не.
Но&робот&все&нечеловеческий&является&всем&грехом&с&ког&и&сеткой.
Не&если&мы&обучаем&доброму&вещам&в,&поэтому&это&может&обучить&нашу&плоть&
[...]
Как&человек&сам&&смесь&есть,&буйный&парадокс,
Так&мы&должны&обучить&наши&безумные&машины:&встать&встать,&подтянуть&в&свои&носки!
Приходи и беги&со&мной,&дикими&детями/мужчинами,&полу&страстями&и&думами,&полу&клоунами.
Скорость&роботы&мыши,&гонки&роботы&мужчины,&выигрышQпроигрыш&в&роботах&городах».
Рой! Брэдбери1

Принципы!инициативы!робототехники!происходят!в значительной степени!от!отражения!масштаба! в!которых!роботы!уже!влияют!нашу!жизнь!и!в!даже!большей!степени!в!какой!это! является! ожидал этого! они! воля! оказывать воздействие! это! в! ! робот! города! из! ! относительно! около! будущее. Будь!начальный! регулирование! связанный! к! этот! преобразующий! технологии! примет!форму! мягких,!руководящих!принципов,!жестких!внутренних!правовых!инструментов,! или! даже! из! сложный! Международный! договоры! является! а! испытывающий! еще,! в! этот! точка,! а! вторично! проблема.! ! начальный! вопрос! скорее! а!(законно)! норматив! вопрос,! прицелился! в! очерчивая! прозрачный! границы! из! ! человек/робот! сосуществование;! обращение! ! норматив! динамика! из! причинность! и! ответственность;! пытающийся! к! определить! место или! места ! Менса и! акт в! процессы! и,! смейся! мы! сказать,! отношения , которые!может!хорошо! доказывать! к! стать! больше!и!больше! комплекс!с! ! достижения!науки!и!технологии.2

Стемминг! от! этот! нуждаться! для! норматив! самоанализ! (в! нашу! социальную! психику! гораздо! больше!чем! что-либо еще!),! это!бумага!есть! приглашение! к отражению! на! !

---

1 Рэй! Брэдбери! Где& Робот& Мыши& и& Робот& Мужчины& Бегать& Круглый& в& Робот& Города:& Новые&Стихи,&Оба&Светлые&и&Темные!(Нью!Йорк:!Random!House!Inc,!1977).!

2!Аврора! Войкулеску! "Человек! Права! Вне! ! Человек:! Герменевтика! и! Нормативность!в!Эпоху!Неизвестного!»,!(ожидается).!

предложенный! Принципы! из! Робототехника! (покрытие! 5! принципов! и! 7! HighELevel! Сообщения)! приходящий! вне! из! ! мультидисциплинарный! эксперт в курсе! ЭПРЦ! и! АКПЧ! Робототехника! Отступление!в! 2010г.! ! сложность! из!вопросов!к! чехол есть! такой! что! эти! размышления! может! только! цель! к! привлекать! с! что! является! предложенный,! с! текст! предложенный! для! отражение,! поднимается! вне! некоторый! из! ! возможный! значения! или! интерпретации!таких!текстов.!Такой!анализ!предлагается!как!необходимый!для!подготовки! почва для! дальше! обсуждения,! и! наконец,!за! посадка! на! любой! возможно! регулирующие!процессы.!

Принципы!в!Поиске!определения!

Размышление!над!существующими!принципами!приглашает!человека,!прежде всего!всего,! поразмыслить!над!что такое!робот!и!над!является ли!определение!этого!одного! !должен!согласиться! для,3 и!поэтому! тип!сущностей,!которые!следует!стремиться!регулировать,!должен!быть!ответом!наш! состояниеEofEtheEart!in! технологии! или!а! отражение! из! наш! состояниеEofEtheE(технология!в)! искусство.!Другими!словами,!на!какой!точке!на!спектре!между!наукой!и!наукой! вымысел! должен! мы! место! сами! когда! проектирование! норм! и! оценка! их! эффективность?!Как!далеко!в!будущее! следует!заглянуть,!когда!будущее!для!которого! мы!регулируем!есть!так! далеко! что!мы!можем!только! предполагать!как! к!своему!существованию,!пока!в! то же!время!мы!несемся!галопом!к!этому!самому! будущему!на!всевозрастающей!скорости?!

Закон/правило! жесткий! или! мягкий,! требует определений. ! принципы! здесь! под! обсуждение! делать! нет! немедленно! отправлять! к! один.! А! робот! является! определенный! к! некоторый! как! а! машина! способный! из! несущий! вне! а! сложный! ряд! из! действия! автоматически'! или,! иначе! нюансы! механический! или! виртуальный!искусственный!агент! обычно!an!electroE механический! машина! что! является! управляемый! к! а! компьютер! программа! или! электронный! схема'.4! Различные,! более! или! меньше! работоспособный! отличия! являются! также! помещать! вперед,! более! особенно! между! промышленный! и! услуга! машины,! между! высоко! автономные!машины!и! когнитивные!компьютерные!программы,!между!воплощенные! и!бестелесные!познания!сущности,!и т.д.!НАСА!само!использует!довольно!приземленное!и! неточное! язык,! очень! бесполезный! для! ! регулятор,! определение! роботы! как! "машины!которые!могут!использоваться!для!выполнения! работы".!Некоторые!роботы,!формулировка!НАСА!продолжается! к! добавлять,! 'может! делать! работа! к! сами себя.! Другой! роботы! должен! всегда! иметь! а! человек!

3!А! определение! никогда! существование! значениеНейтральное,! всегда! установление! ! 'in'Es! и! ! 'вне'Es! следующий! а! более! или! меньше! заявлено! значениеЭладен! путь.! (Смотрите! Алан! Норри! Войкулеску!2000)

4 МерриамЭВебстер! Словарь,! "Определение! из! «Робот», доступ! Февраль! 10,! 2016,!http://www.merriamEwebster.com/dictionary/robot!entry:!robot.

рассказываю! их! что! к! делать'.5!Такой! а! разнообразие! из! составы! создавать! а! регулирующий! головоломки!и!сделает!любые!нормативные!утверждения!трудные!для!следования!и/или!легкие! для! бежать!соответствуя!с.!!

Пока! согласен! что! там! является! нет! согласованный! определение! per& se, 6!some! помещать! вперед! а! число!функций,!которых!робот!должен!иметь,!функций,!которые,!из!нормативных!(и! не! только)! перспектива,! являются! сами себя! в! нуждаться! из! определения:! чувствуя ! окрестности! (имея! встроенное! «осознание»! своего! окружения);! движение,! ли! катание,!ходьба,! толкая,! или!может быть!даже!просто передача данных;!энергия! быть в состоянии! к!силе!себя!в!путях! это! будет!зависеть!от!чего! цель! ! робот!есть;!интеллект:!наделенный!'умом'!своим!программистом,! имеющий! емкость! к! оценивать! окрестности,! обстоятельства,! сложный! информация.! [(Ethe! Более! машина! способна! Независимо! взаимодействовать! с! динамичным! миром! !среди!других,! именно! к этому)]!

Так,! а! робот! является! определенный! более! конкретно! как! а! система,! а! машина! который содержит! датчики,!системы!управления,!манипуляторы,!питание!и!программное обеспечение!все! работает! вместе! к! выполнять! а! задача".! Согласно! к! такой! а! перспектива,! "[проектирование,! здание,! программирование! и! тестирование! а! робот! является! а! комбинация! из! физика,! механический! инженерия,! электрический! инженерия,! структурный! инженерия,! математика!и! вычисления.! В! некоторый! случаи! биология,!медицина,! химия!мощь! также! быть!причастным».! Если! ! студент!в! робототехника! может! активно! взаимодействовать с! все! эти! дисциплины! "в! а! глубоко! проблема проблемаРешение! окружающая среда»,7!некоторые! мог! правильно! сказать,! что! регулировка! роботы! и! робот! города! требует! а! сходным образом! сложное! междисциплинарное!взаимодействие!с!большинством!если! не все! из! эти! поля.#!Для! ! норматив! дискурс! (будь то! жесткий! регулирующий! или мягкий принцип),! ! факт! что!многие из!этих! определений!имеют!количество!общих!точек!не!достаточно.!
А! определение! что! является! достаточно! точный,! еще! динамичный! достаточно! к! захватывать! ! сущность! из! ! социотехнологический! явления! есть! поэтому! нужный! для! открытие! принципы! робототехники! к! дальнейшее!развитие!и!проблематизация.!Это!нужно! относится!к!перспективам! таким,!как!Андра!Кей,!который!говорит!о!роботах!как!«...!&& окружающая среда;!слишком!большая! для!нас!чтобы!смотреть!на!как!на ! item».!Хотя!неизбежно!связан!с! прогрессирует! из! технология! Е! "[что мы! вызов! а! робот! сегодня!больше! сложный

5 !НАСА! "Что! Является! Робототехника?», НАСА& Знает! Может! 18,! 2015,! http://www.nasa.gov/audience/forstudents/kE4/stories/nasaE know/what_is_robotics_k4.html!

6!Х.! Джеймс! Уилсон! "Что! Является! а! Робот! Во всяком случае?», «!Harvard&Business&Review,! Апрель! 15,!2015,!https://hbr.org/2015/04/whatEisEaErobotEanyway.

7 Там же.

чем!то, что!мы!назвали!роботом!в!80-х"!говорит!Кей!Е!это!также!правда!что!это!больше! чем!это.!«Это!всегда! было!идентификацией&проблемой»!говорит!Keay.8!!

Это! должен! быть! сказал,!Однако! что тождества!и! классификации! есть! всегда! был! проблематично! и! проблематизировано! когда! часть! из! регулирующий! инициативы! ли! эти! имел! к! делать! с! люди! или! нелюди! похожи.! Закон,! в! особый,! всегда! заканчивается!превращением любой!личности!в!юридическую фикцию, которая!часто!имеет!очень!мало!дел! с!любым!другим!физическим!или!научным!измерением!этой! сущности. 9!В!то же!время! закон!Е!в!его!самом!широком!смысле!общественно!поддерживаемых!нормативных! императивов! Е!есть!всегда! процветал!на!определениях.!Отсутствие!а!'рабочего!определения'!а! появляется робот! поэтому!оба!как!свидетели! к! !проблемы!закрепления!вниз! технологии!в!его! торопиться,!и!как! отражение!возможную!слабость,!которую!устранить!в!предлагаемом! документ.!!

И последнее!но!не в последнюю очередь!в!добавок! к!вопросу!о! отсутствие!согласованной!работы! определение!(которое!было бы!скоро!оспорено!и!проблематизировано,!конечно!),!вот! это также! признание! что! 'определения! есть! никогда! нейтрально'.!Эта!идея!была! передовой! некоторый! десятилетия! назад! к! Ларри! Может! когда! отражение! на! определение! ! ответственность!нечеловеческого!коллективного! агентства!(иначе не!неуместная!юридическая! инновация).!Определения,!авансы!Могут,!создать!'псевдоединства',! которые!предлагаются! как! факты,!в!реальности!они!устанавливают!оппозиции,!которые!произвольно! разъединяют!этих! которые!включены!и! те!которые!исключены! из!общей!концептуализации! или! упражняться'. 10!Это! утверждение! воля! становиться! более! и! более! очевидно! один раз! ! спектр!доступных! вариантов!между!роботом!и!ИИ!становится!расширенным.11

Будь то!ожидаемый!с!страхом!или!с!волнением! задача!регулирования! несколько!измерений!взаимодействий! человеко-роботов!многочисленны.!Число! из! проблемы! помещать! вперед! для! отражение! в! связь! к! ! данный! пять! принципы! являются! упомянуто!кратко!здесь:!

8!Смотрите! интервью! с! Андра! Кей,! основатель! из! Робот! Панель запуска! и! Управление! Директор! из! Силикон! Долина! Робототехника! в Сигне! Брюстер! "Что! Является! а! Робот?! ! Отвечать! 2014,! https://gigaom.com/2014/07/05/whatEisEaErobotEtheEanswerEisEconstantlyE5developments/! Постоянно! Развивается! Ответ! (Выделение автора!).!

9 Давид!Фагундес,!"Примечание,!Что!Мы!Говорим!О!Когда!Мы!Говорим!О!Лицах:!В! Язык!юридической!художественной литературы»,!Harvard&Law&Review 114,!no.!6!(2001):!1745–68.

10 Ларри! Может,! Разделение и ответственность! Новый! версия! 1996 год! (Чикаго:! Университет! Из! Чикаго!Пресс,!1992),!171ff.

11 Кеннет! Грейди! "Искусственный! Интеллект:! Быть! Испуганный,! Быть! Очень,! Очень! Испуганный! (Или! Нет),"! SeytLines:&Изменение&Практики&Закона&Закона! Декабрь! 31,! 2014,! http://www.seytlines.com/2014/12/artificialEintelligenceEbeEafraidEbeEveryEveryEafraidEorEnot/.

Во-первых&всего&,&есть&потребность&в&нескольком&более&ясном&ракурсе&в&что&&что&это&что&,&что&регулирует
Выше!упомянутое!отсутствие!согласованного!определения!кроме ,! принципы! поставить! вперед!
для!обсуждения!выявления!потенциала!путаницы!как!к!к! действительный! повестка дня:! 'регулировка!
роботы! в! ! настоящий! мир'! имеет! а! двойной! смысл,! полный! из! подводные камни.!Если!кто-то
'регулирует!роботов'!сами себя,!текст!вводит!в!подразумеваемое!агентство! это!может!приниматься!
как!само собой разумеющееся!в!контекстах!где!это!может!быть!нежелательным!(хотя! такое!толкование!
явно!противоречит!некоторым!из!принципов,! особенно!
Принцип! нет.! 2! и! 5).! ! второй! значение,! более! в! линия! с! что! ! пять! принципы! сами себя!
раскрывать,! мог! цель! в! 'регулировка! ! создание! и! использовать! из! роботы». Выбор!этой!
интерпретации!должен!быть!более!явным!на протяжении! составов,!избегая!регулятивной!путаницы.!

Принцип!нет!1:!

Роботы&являются&многофункциональными&инструментами.&Роботы&не&должны&быть&разработаны&исключительно

! первая!часть! этот принцип,!однако,!ещё!более!озадачивает.!Прежде всего!всего!один! может!
находить! ! начиная! заявление! 'роботы& являются& многопользовательскими& инструментами'!
как! виртуально! а! ограничение,!которое!не!служит!фактической!цели.!Непонятно!почему!робот!
должен! быть!'множественным!использовать'!для!чтобы!быть!безопасным!или,!наоборот,!каким!каким!
и!иначе!смертельным! робот!может!стать!любым!менее!смертоносным!если!разработан!как!'multiEuse'.!
Это!относится! к! ! следующая!часть!принципа:
'роботы&не&должны&разрабатываться&исключительно&или&прежде всего&для&убиения&
или&вреда&людей'.!Для!чтобы!привести!'роботов-убийц'!в!соответствие!букве!от! этой!части!принципа!
было бы!достаточно!чтобы!также!научить!«роботов-убийц»!делать! блины!или!вязать!шерстяные!
носки.!Это!что,!во! с!юридической!нормативной!точки зрения,! можно!назвать!'творческой!
соответствие!лазейке'. Чтобы!выявить!и!использовать!такую!лазейку,!юридический!глаз!должен!
смотреть!не!дальше, чем!буквальное!толкование! ! текст.!Тем не менее! ! буквально! интерпретация!
является! один! из! ! начальный! правила! из! толкование!в!законе,!когда!толкование!в!соответствии!с!
духом!правила! может!не!быть!удобным!одним.!Пояснения,!данные!этому!принципу!в!2010! оригинал!
документ! делать! нет! казаться! к! Действительно! адрес! этот! скорее! базовый! подход! к!
интерпретации!правил!и!их!последствий!в!этом!конкретном!контексте.

! комментарий! к! этот! принцип! появляется! к! подразумевать! другой! потенциал! подводные камни! для! нормативное!обоснование.!Прежде всего,!как!упоминалось!с!уважением!к!'multiEuse! инструментам',! предпринимаются!попытки!выдвинуть!идею,!что!роботы!являются инструментами! как!любые!другие.!В! чтобы!преследовать!с!этой!логикой,!эквивалентности!изыскиваются!любой! ценой.!Сравнение!а! робот!с!ножом!или!пистолетом!используется!для!различных,!относительно! безвредных!и!преступных! цели, делает! нет! крышка! для! ! непоследовательность! что! там! являются! инструменты,! включая! оружие, для!которого!можно!придумать!никакой!другой!цели!кроме!prima!facie!one.!!!

Принцип! нет.! 2:! Люди, а не роботы, являются ответственными и агентами. ! Комментарии! к! этот! принцип!кажется!вносит!больше!замешательства, чем!ясности.!Прежде всего,! относительно!небольшой! иметь значение,! есть! презумпция! что!'никто!не!может! умышленно!намеренно!отправиться! к! построить! робот!который!ломается! !закон'.!Это! ставит! вперед!а!презумпция! это!имеет!нет! основа!в!реальном!мире!'отклонения!и! неповиновения',!как!выявлено!социо-Элегальным! исследования!среди!широкого!населения! в!широких!а также!а также!а! среди!белых!воротничков.12 !этот! Комментарии! появляется! к! игнорировать! оба! ! способ! 'закон! думает'13 как! хорошо! как! ! способ! роботы!могут!не! достичь!целей!и!желаний!которые!люди!определяют'.14

В качестве!дополнительного!элемента!здесь!следует!указать!также!что,!в!отсутствие! четкое! рабочее!определение,!ИИ,!'обучающиеся!роботы',!и т.д.,!все!заинтересованы!в них! принципы! и! их! параметры.! Их!механика! однако! может! хорошо! более! сложный!чем!закон/ нормативный!дискурс!может!обработать!в!отсутствии!клира! определение#.! Роботы! и! ИИ! машины! может! хорошо! учиться! к! иметь дело! с! «исключения»! перед!законом!научится! справляться!с!'разногласиями'.15!В равной степени!другие!дисциплины!кажутся! к! указывать! что! номера! (в! этом! конкретном! случае! 'программирования')! может! хорошо! быть! более! чем! только! что,! номера! существование! неотъемлемо! дополнено! по/ассоциировано!

12!Некоторые! полезный! хотя! свободный! Примеры! Райан! Мэтьюз! и! Вт! Ваккер! The& Deviant's& Advantage:& How& Fringe& Ideas& Create& Mass& Markets! (Случайный! Дом,! 2010);! Келли! Фишер! «В! Психология! из! Мошенничество:! Что! Мотивирует! Мошенники! к! Совершить! Преступление?"! (Социальные! Науки! Исследования! Сеть,! 31 марта! 2015 г.),! http://papers.ssrn.com/abstract=2596825.!

13 Гюнтер! Тойбнер! "Как! ! Закон! Думает:! К! а! Конструктивист! Эпистемология!Права»,!Право&и&Общество&Обзор 23,!номер!5!(1989):!727–57.

14 !Смотри! для! пример! ! ортогональность! теория! в! Ник! Бостром! «В! Суперинтеллигент! Воля:! Мотивация! и! Инструментальный! Рациональность! в! Передовой! Искусственный! Агенты!» Умы& и& Машины! 2012,! http:// www.nickbostrom.com/superintelligentwill.pdf.!

15!Это!построено!на!идеях!тождества!и!различия,!от!Лейбница!к!Канту…!Жиль& Делёз&Q&Le&Point&de&Vue&(Le&Pli,&Leibniz&et&Le&Baroque)&1986&FRA&Sub&ITA,!2012 ,! http://www.youtube.com/watch?v=2ZrA_7ewQGs&feature=youtube_gdata_play э.

с а! повествование! движение! что,! один! мог! сказать! здесь,! может! быть! истолковано! иначе! машиной,!чем!человеком,!все же!может!все еще!истолковываться!им!16!!!

Принцип! нет.! 3:! Роботы& являются& продуктами.& Они& должны& быть& разработаны& с использованием& процессов& которые&обеспечивают&их&безопасность&и&защищенность. Это! принципы!поднимает!вопросы!различий! в! ! законно! самозащита! дебаты;! деликт! проблемы,! ассимиляция! из! ответственность,! и т. д.,! проблемы! что! зависеть! очень! много! из! ! контекст! и! из! ! степень! к! который,! как! упомянуто!выше,!регулирующий орган!будет!желать!толковать! эквивалентности!между! роботы! и! другой! типы! из! инструменты! или! между! роботы! и! другой! типы! из! свойство! предметы.!

Принцип! нет.! 4:!

Роботы&являются&изготовленными&артефактами.&Они&не&должны&быть&разработаны&а&обманчивым&способом&д

Принцип! нет.! 5:! & Лицо& с& юридической& ответственностью& за& робот& должно& быть& отнесено. Этот!принцип!кажется!достаточно ясным!в!контексте,!в!котором!мы!являемся!а! далеко! далеко!от!ИИ!способного!технически!выполнять!требования!для!такого! ответственность.!В! одинаковый! время,! принимая во внимание!сложные!элементы!и! поля! что! входить! ! макияж! из! а! робот (см.! выше! при! обсуждении! вопросов! связанных!с!определением!),!вызовы!регулятивному! подходу!будут!остаться!от! ! точка! из! вид! из! поднимается! отдельно! ! различный! градусов! из! ответственность,! когда! вещи! идти! неправильный.! Имея! а! 'зарегистрировано! хранитель,! носитель! из! ответственность,! является! только часть! из! ! решение.! ! ответственность! несущий! сущность! воля! требовать! дальше!

16 Маркус! дю! Саутой! Повествование& и& Доказательство:& Две& Стороны& одного и того же& Уравнения?& |& ФАКЕЛ! (ФАКЕЛ,! Оксфорд! Исследовательский! Центр! в!! Гуманитарных наук! 2015 г.),! http://www.torch.ox.ac.uk/narrativeEandEproofEtwoEsidesEsameEequationE0.

17 Смотрите! например!Синтия!Бризл,!“Эмоциональные!и!Общественные!Гуманоидные!Роботы”! International&Journal&of&HumanQComputer&Studies 59,!no.!1–2! (Июль! 2003 г.):! 129ff,! doi: 10.1016/S1071E5819(03)00018E1.

отражение,! как! воля! ! тип! из! вред(а)! что! может! быть! отнести! к! такой! сущности,! принимая! во внимание,!как! некоторый! предлагать,! что! для! ! первый! время! '! распущенность! из! данные'! в последнее время!сочетаются!с!'!способностью!наносить!физический!вред'.18!!

Вместо!выводов!

Замкнув эти!заметки, полный!круг,!другим!из!поэтических!образов!Брэдбери,!можно!сказать,!что,! поскольку!роботы!мыши!касаются,!это!вполне!вероятно! !это!в!контексте! рыночных!властей!и! научных!возможностей!дерегулирования!вместе,!мы!будем!наиболее! конечно! первый! 'Прыгать! выключенный! ! утес! и! строить! наш! крылья! на! ! способ! вниз'.! В! такой! обстоятельствах,!что! можно!сделать!это!убедиться!в!быть!относительно!готовым!к! этот! Прыгать! к! уже! допрос! и! проблематизировать! наш! отношение! с! наука! и! технологии! и! к! усиление! наш! отражение! на! жизнь! в! робот! городов.!Обращаясь!к!«инструментам»,!«продуктам»,!«артефактам»!и!«агентам»,! нужно!принимать!
Св! Августина! отражение! на! ! сложный! связь! между! язык! и! интерпретация! как! а! путь! к! раскрытие! а! Глубже,! экзистенциальный! уровень! из! самопонимание .!Путь!мы! думать! нормативно!о! человекЭробот!взаимодействие19 будет! сказать! как! много! о! ! робот! как! о! ! человек.! К! занимать! от!
Джинабаттиста! Вико! 1725! Нью&Наука, 20!мы! нуждаться! к! медведь! в! разум! что! наш! думать о! роботы! есть! коренится!в!данном! культурно! контекст.! Это означает! что,!в! отражение! о! ! норматив! параметры! из! робот! города! ! Социальное! ученый! воля! нет! иметь дело! с! а! поле! из! идеализированный! и! предположительно! 'subjectЕнезависимый! объектов',!но!будет! исследовать!мир!который!является!по сути!ее!своим.!Процесс! из! регулировка! роботы! является,! поэтому,! а! процесс! из! самопонимание! укоренился! в! а! с учетом!исторического!контекста!и! практики.!Понимание,!которое!не!завершает! автоматически!в!аккуратных,!нормативных,! законоподобных!предложениях.!

18 Райан!Кало,!«Робототехника!и!Уроки!Киберправа»,!California&Law&Review 103! (2015):! 513.

19 Будь! этот! взаимодействие! понял! на! Латура! спектр! из! факты! и! агентство;! Бруно! Латур! "Как! к! Разговаривать! О! ! Тело?! ! Нормативный! Измерение! из! Наука! Учеба!» Тело&&& Общество 10,! нет.! 2–3! (! 1 июня! 2004 г.):! 205–29,! doi:10.1177/1357034X04042943;!Бруно!Латур,!“Тело,!Киборги!и!Политики! из! Воплощение»,!in! The&Body:&Darwin&College&Lectures,!ed.!Sean!Sweeney!and! Ян! Ходдер,! 2002,! 127–41.
20!Джамбаттиста! Вико! The& New& Science& of& Giambattista& Vico! (Cornell University! Нажмите,!1744).

Комментарий об ответственности, дизайне продукта и понятиях безопасности

Паула Боддингтон, факультет компьютерных наук, Оксфордский университет.

Комментарии здесь в основном относятся к правилу 2 и правилу 3. Предполагается, что принципы можно было бы сделать более специфичными для контекста реализации роботов, а ключевые понятия, такие как «безопасность», можно было бы разработать и, возможно, расширить, или точный способ использования этого термина сделать более ясным.

Некоторые из этих моментов можно проиллюстрировать, рассматривая в общих чертах использование робототехники в контексте ухода за пожилыми людьми и социальной помощи.

Правило 2: Люди, а не роботы, являются ответственными агентами. Роботы должны быть спроектированы и использоваться, насколько это практически возможно, в соответствии с существующими законами, основными правами и свободами, включая конфиденциальность.

Если люди являются ответственными агентами, а роботы — нет, это означает, что везде, где роботы используются для замены человека или части человеческой деятельности, то ответственность, ранее приписываемая человеку-агенту или действиям человека, затем либо перемещается в более сложную систему, либо, возможно, игнорируется. быть неожиданным и сложным.

Роботы будут использоваться в системе человеческих агентов и поведения. Такие системы могут быть формализованы с четко выраженными понятиями ответственности и подотчетности, например, в больничных условиях (хотя могут быть элементы таких систем, которые не полностью поняты или формализованы с полной адекватностью); или они могут быть неформальными, например, в домашних условиях ухода. Неформальная семейная или общественная обстановка, социальные исследования показывают, что могут сохраняться сильные местные культуры и ценности в отношении линий ответственности и подотчетности.

В качестве примера того, как обязанности и подотчетность могут быть перераспределены, если робот берет на себя некоторые функции помощника по охране здоровья в рамках награждения, то затем ответственность может быть перенесена различными способами на разных субъектов в системе управления здравоохранением. Например, то, что раньше могло быть расценено как недобросовестность сотрудника, теперь может восприниматься как трудности в понимании или эксплуатации оборудования. Это может иметь далеко идущие последствия.

Отслеживание и понимание таких линий ответственности и подотчетности может быть сложным. Возникает вопрос, является ли это исключительно задачей тех, кто отвечает за среду, в которой используются роботы, или же разработчики роботов могут нести определенную ответственность за помощь тем, кто будет работать вместе с роботами, чтобы понять эти проблемы.

Правило 2 говорит о соблюдении существующих законов, основных прав и свобод, включая неприкосновенность частной жизни. Однако, кроме того, в определенных условиях будут более конкретные и локальные протоколы и методы, которым желательно следовать роботам. Возможно, стоит прямо указать это в правилах.

Например, в NHS существуют стандарты медицинской помощи, направленные на оказание помощи, ориентированной на человека, и достойное лечение пациентов. Такие формулировки стандартов играют центральную роль в обеспечении качественного медицинского обслуживания.

кто-то как человек с достоинством. Выяснение того, как использование роботов влияет на такие богатые и контекстуализированные ценности, может быть очень важным, но может быть более сложным вопросом, чем просто вопрос соблюдения закона и обязанностей, предусмотренных законом. явно заявляя о желательности решения этих вопросов при разработке и использовании роботов, если только не предполагается, что это широко понимается.

Правило 3. Роботы — это продукты. Они должны быть разработаны с использованием процессов, обеспечивающих их безопасность и защищенность.

Theremaybeaneedtoarticulatewhatismeantby'safety'inthiscontext.Doesthissimplyrefertophysicalsafety?

Anddoesitrefertosafetyintermsoftheimmediateoperationoftherobot,ortoeffectsoftheuseofrobotsfurtheralongthesystemwi

Эти правила гласят, что продукты должны быть безопасными. Однако, хотя правила этики часто формулируются как средства предотвращения вреда, стремление разрабатывать продукты, которые просто «безопасны», не вдохновляют. Хороший дизайн, отвечающий важным человеческим потребностям, выходит за рамки понятия простой безопасности, но следует позаботиться о том, чтобы они соответствовали своему назначению.

В тех случаях, когда роботы заменяют или расширяют человеческую деятельность, задача, которую берет на себя робот, может быть непрозрачной в целом. чтобы иметь больше времени для заботы о задачах, которые может обеспечить только человеческое взаимодействие. И наоборот, роботы могут быть спроектированы таким образом, чтобы хранить некоторые из морских аспектов задачи.

Проблемы с безопасностью возникают из-за того, что обнаружение того, как использование роботов может нарушать или потенциально даже улучшать некоторые, возможно, скрытые аспекты задач, которые берут на себя роботы, может потребовать серьезного анализа и исследований. Таким образом, существует вероятность того, что такая важная работа может быть не признана или не признана. Опека формирует чрезвычайно сложные системы социального взаимодействия, и обнаружение того, как роботы вписываются в такие условия, может потребовать чрезвычайно тщательного анализа с использованием различных знаний. В этом широком смысле, но, возможно, стоит рассмотреть прямое указание на то, что безопасность должна рассматриваться в широком смысле. Это также вновь поднимает вопрос о совместной ответственности между командами, занимающимися проектированием и производством роботов, и теми, кто будет с ними работать. .Конечно, возможно, что роботы могут улучшить такие вопросы.

Внимание к безопасности, конечно, будет включать рассмотрение вопросов безопасности использования роботов в более широкой системе. Например, распространенной и серьезной проблемой в больничных условиях пожилых и уязвимых пациентов является обезвоживание. Это может иметь серьезные последствия для здоровья, приводящие, например, к

спутанность сознания. Иногда обезвоживание усугубляется трудностями с получением напитков и управлением ими. Предположим, что роботизированная система может быть разработана для помощи таким пациентам в питье. Эта система может работать с полной безопасностью и надежностью с точки зрения ее немедленного использования, например, никогда не сбоит в работе таким образом, чтобы дать пациенту слишком много слишком быстро, и никогда не проливая напитки и не ударяя пациента.

Тем не менее, такая система потенциально может иметь серьезные негативные последствия в конкретном контексте. Повышенная гидратация может вызвать повышенную частоту ночного недержания мочи у некоторых пациентов. Это приводит к «блокировке кровати», поскольку пациенты становятся мягкими, а затем им приходится искать приспособления, которые могут удовлетворить их потребности. Повышенная гидратация может также привести к увеличению случаев, когда пациенты встают с постели, чтобы пойти в туалет, и, следовательно, к учащению падений. Это также может иметь тяжелые последствия.

Когда правило 3 говорит о «безопасности», ясно ли, что оно включает или не включает рассмотрение того, как роботы могут работать в более широкой системе работы? Ясно ли из принципов робототехники, какую ответственность несут те, кто работает с робототехникой, или же ответственность за выявление таких потенциальных последствий использования роботов лежит на руководителях больниц и другом персонале?

Вклад Руланда де Брюина и Мадлен де Кок Бунинг в

Семинар AISB по принципам робототехники, 4 апреля 2016 г., Шеффилд, Великобритания

---

1. Введение

Прошло пять лет с момента публикации принципов робототехники EPSRC, разработанных группой выдающихся британских экспертов в области робототехники и искусственного интеллекта на семинаре, финансируемом EPSRC/AHRC . сформулированы в виде пяти «правил» и

семь «сообщений высокого уровня». Принципы действительно оказали значительное влияние на британскую робототехнику. исследования, и продолжают вызывать существенные дискуссии. Поскольку в настоящее время общественная озабоченность по поводу развитие робототехники ускоряется, мы считаем полезным пересмотреть принципы

рассмотреть их сохраняющуюся актуальность в соответствии со следующими критериями.

Наш вклад сосредоточен на втором принципе:

Принцип 2: Люди, а не роботы, являются ответственными агентами. Роботы должны быть спроектированы; эксплуатировался насколько это возможно для соблюдения существующих законов и основных прав и свобод, включая конфиденциальность.

На самом деле этот второй принцип робототехники EPSRC двоякий. С одной стороны

принцип касается ответственности - в том числе ответственности - за действия робота, с другой стороны, принцип влечет за собой методы проектирования машин, которые могут помочь в соблюдении существующих законов и основные права и свободы, включая неприкосновенность частной жизни.

Поскольку и ответственность, и дизайн составляют основу внедрения робототехники, что касается например, автономные интеллектуальные автомобили в нашем обществе, мы проверим этот двойной принцип, сосредоточив внимание на текущей разработке и внедрении автономных интеллектуальных автомобилей. Будь то Второй принцип EPSRC можно рассматривать как перспективный, он будет проверен по трем критериям:

1. Обоснованность — является ли принцип правильным, как утверждения о природе роботов, разработчиков роботов, и отношения между роботами и людьми, или она онтологически ошибочна, неточна, вне датированы или вводят в заблуждение.
2. Достаточность/общность — является ли принцип достаточным и достаточно широким, чтобы охватить все важные вопросы, которые могут возникнуть при регулировании робототехники в реальном мире или важные проблемы упущены из виду.
3. Полезность — принцип практического использования роботов разработчиками, пользователями или законодателями в определение стратегий передовой практики в области робототехники, или правовых стандартов или рамок, или они ограничены в своем использовании из-за отсутствия конкретики или допуска критических исключений.

---

[1] https://www.epsrc.ac.uk/research/ourportfolio/themes/engineering/activities/principlesofrobotics/

2. Современное состояние

2.1 Состояние автономных интеллектуальных автомобилей (АИК)

Перед проверкой этого принципа мы вкратце познакомим вас с современными AIC.

В настоящее время потребительские автомобили все чаще оснащаются технологиями, помогающими в определенных аспектах вождения. Примеры таких технологий включают помощь в удержании полосы движения, экстренное торможение, помощь при парковке и адаптивный круиз-контроль. В ближайшем будущем более высокий уровень автоматизации автомобилей будет стали доступными, что в конечном итоге привело к появлению полностью автономных транспортных средств.

Также сейчас некоторые автомобили уже оснащены определенными формами автоматики. Есть даже доступны прототипы, которые могут управлять автомобилем без участия человека-оператора. В настоящее время Google является пионером в области самообслуживания. автомобильной технологии, и в начале 2015 года провела дорожные испытания полностью функционирующего прототипа AIC в Bay Area, штат Калифорния . беспилотных грузовиков

которые автономно следовали за управляемым человеком грузовиком, идущим в колонне, на дорогах Нидерландов.4 Volvo планировала развернуть 100 автомобилей, которые должны быть в состоянии взять на себя все аспекты вождения в Швеции к 2017 году5 и в Германии, часть автобана A9 между Мюнхеном и Берлином зарезервирован для масштабных испытаний автономных транспортных средств в ближайшие годы6.

Определение автономных интеллектуальных автомобилей состоит из трех элементов. Автономность относится к уровню вмешательства человека, необходимого для работы, которые можно рассматривать как спектр: меньшая потребность в вмешательство человека подразумевает более высокий уровень автономии. Интеллект относится к тому, как система может воспринимать свое окружение и способна адаптировать поведение к изменяющимся условиям. Это включает в себя способность учиться, обрабатывать сложную информацию и решать проблемы7 . моторизованные транспортные средства, используемые для перевозки товаров и/или людей и для оказания услуг.

---

[2] Википедия, «Беспилотный автомобиль Google», доступно в Интернете по адресу <http://en.wikipedia.org/wiki/Google_driverless_car> (последний доступ 17 марта 2015 г.), со ссылкой на Мэтта О'Брайана, «Новый дурацкий» Google беспилотный автомобиль — признак грядущего», 22-12-2014 г., доступно в Интернете по адресу <http://www.mercurynews.com/business/ci_27190285/googles-goofy-new-self-driving-car sign-things> (последний доступ 28 января 2016 г.).

[3] См., например, <https://www.media.volvocars.com/us/en-us/media/pressreleases/145619/volvo-car-groups-first-self-driving autopilot-cars-test-on-public-roads - вокруг-> (Вольво), <http://www.pcmag.com/article2/0,2817,2387524,00.asp>, (Фольксваген) <http://www.bbc.com/news/technology- 25653253> (BMW) (последний доступ 28 января 2016 г.).

[4] См. <http://www.scania.nl/about-scania/media/platooning/> (по состоянию на 20 марта 2015 г.).

[5] Александр Стоклоса, «У Volvo есть «пригодный для производства» автономный автомобиль, который будет выпущен на дороги к 2017 году», доступно в Интернете по адресу <http://blog.caranddriver.com/volvo-has-a-production-viable -автономный-автомобиль-выведет-его-на-дорогу-к-2017/>. (последний доступ 20 марта 2015 г.).

[6] Стивен Эдельштейн, «Германия планирует программу испытаний автономных автомобилей на высокоскоростных автобанах», 28 января 2015 г., доступно в Интернете по адресу < http://www.motorauthority.com/news/1096521_germany-plans-autonomous-car-test-program- на высокоскоростном автобане> (последний доступ 20 марта 2015 г.).

[7] См. Мадлен де Кок Банинг, Лаки Белдер и Руланд В. де Брюин, Рабочий документ: «Карта правовой базы для введения в общество роботов как автономных интеллектуальных систем», на стр. 11. 3-4, доступно в Интернете по адресу <http://www.caaai.eu/wp-content/uploads/2012/08/Mapping-L_N-fw-for-AIS.pdf> (последний доступ 28 января 2016 г.).) и ссылки на Самира Чопру и Лоуренса Ф. Уайта, Юридическая теория автономных интеллектуальных агентов (Анн-Арбор: University of Michigan Press, 2011) на стр. 10 (автономия) и Коллин Р. Дэвис, «Эволюционный шаг в области прав интеллектуальной собственности — искусственный интеллект и интеллектуальная собственность», 27 Обзор компьютерного права и безопасности , 2011 г., на с. 601-619 (интеллект); и главу тех же авторов «Отображение правовой базы для внедрения в общество роботов как автономных интеллектуальных систем» в книге Сэма Мюллера и др. (ред.), Закон будущего и будущее права, серия 2012 г. (De Кок Банинг, Белдер и Де Брюин, 2012 г.), стр. 195–210.

AIC могут способствовать поиску решений проблем, с которыми в настоящее время сталкивается наше общество. Дорога безопасность резко повысится, когда «человеческий фактор» будет устранен как фактор, вызвавший несчастные случаи. AIC могут значительно снизить риск автомобильных аварий, поскольку 93% дорожно-транспортных происшествий происходят по вине человека8, что приводит к 1,3 миллионам смертей и 50 миллионам серьезных травм во всем мире в год9. Помимо повышения безопасности дорожного движения, AIC могут способствовать более эффективному использованию дорожной сети, сократить выбросы CO2 и способствовать повышению мобильности людей с ограниченными возможностями10. Таким образом, внедрение АПК могло бы дать ответы на вопросы снижения очевидных в настоящее время рисков, являющихся результатом технологических инновации последних десятилетий.11

Однако не все с оптимизмом смотрят в будущее без водителя. Утверждается, что, хотя АИК могут быть полезно для безопасности дорожного движения, внедрение автономных транспортных средств повлечет за собой другие риски. AIC будут быть уязвимым для взлома, например. Также бизнес-модели и занятость в такси и транспортные рынки значительно изменятся, а водители могут в конечном итоге устаревать после автономизации вождения.12 Кроме того, риск несчастных случаев может возрасти, когда автономные и неавтономные автомобили сосуществуют на одних и тех же дорогах.13

2.2 Правовое государство

Достаточная определенность в отношении правового статуса необходима для роста и принятия обществом потребителей. технологии. Неопределенность вызывает обратное. Может ли в таком случае машина быть ответом на машина? Ниже мы кратко обсудим вопросы ответственности, которые в настоящее время препятствуют введению и развертывание в обществе AIC и коснуться возможных технологических решений для некоторых из эти проблемы, которые могут включать конфиденциальность по замыслу.

Обязанность

Текущие правила ЕС, касающиеся ответственности и ответственности за ущерб, который может быть вызван AIC создают проблемы с точки зрения инноваций в области AIC и их общественного признания. На С одной стороны, производители AIC опасаются, что в соответствии с Директивой об ответственности за качество продукции (PLD) они могут быть легко

---

[8] Брайант Уокер Смит, «Человеческая ошибка как причина автомобильных аварий», 18 ноября 2013 г., доступно в Интернете по адресу <http://cyberlaw.stanford.edu/blog/2013/12/human-error-cause-vehicle-crashes. > (последний доступ 28 января 2016 г.).

[9] ОЭСР, «Справочник ОЭСР, 2013 г.: экономическая, экологическая и социальная статистика», 2013 г., доступно в Интернете по адресу: <http://www.oecd-ilibrary.org/sites/factbook-2013-en/06/02/03/index . .html?contentType=&itemId=/content/chapter/factbook 2013-50-en&containerItemId=/content/serial/18147364&accessItemIds=&mimeType=text/html> (последний доступ 28 января 2016 г.), также цитируется в Gillian Yeomans, «Autonomous Vehicles – передача управления: возможности и риски для страхования», доступная в Интернете по адресу <https://www.lloyds.com/~/media/lloyds/reports/emerging%20risk%20reports/autonomous%20vehicles%20final.pdf > (последний доступ 28 января 2016 г.) (Yeomans 2014) на стр. 5.

[10] См., например, Yeomans 2014, p. 5. Также Энн Поузи, «Автономные дорожные транспортные средства», сентябрь 2013 г., с. 1. Доступно в Интернете по адресу <http://www.parliament.uk/briefing-papers/post-pn-443.pdf> (POSTnote 2013); Роболав 2014, на с. 42.

[11] Загрязнение окружающей среды, изменение климата, социальная изоляция «более слабых сторон» и высокий риск аварий на (европейских) дорогах можно рассматривать как результат процессов модернизации и индивидуализации, имевших место в прошлом столетии. С этими побочными эффектами теперь, в свою очередь, нужно бороться.
См. определение и исследование концепции общества риска Ульрихом Беком, его книгу «Общество риска, навстречу новой современности», Лондон: Sage Publications, 1992.

[12] См., например, Скотт Ле Вайн и Джон Полак, «Автоматизированные автомобили: плавная езда вперед?», февраль 2014 г., с. 14, доступно на в Интернете через <http://www.theitc.org.uk/docs/114.pdf> (последний доступ 28 января 2016 г.).

[13] См. Уэйн Каннингем и Антуан Гудвин, «Шесть причин любить или ненавидеть автономные автомобили», 8 мая 2013 г., доступно на в Интернете по адресу <http://www.cnet.com/news/six-reasons-to-love-or-loathe-autonomous-cars/> (последний доступ 28 января 2016 г.).

нести ответственность за ущерб, причиненный неисправными AIC, которые могли бы оказать охлаждающее воздействие на инновации. [14] Принимая во внимание, что, с другой стороны, действующая система ответственности за качество продукции на самом деле не предоставить потребителям простой инструментарий для привлечения производителей AIC к ответственности за дефекты их продукции в все. На потребителей ложится довольно тяжелое бремя доказывания того, что дефект действительно имел место. АПК, а также о причинно-следственной связи между дефектом и возникшим ущербом. Предоставление доказательств будет более сложным, когда автономия и интеллект автомобилей возрастут, поскольку жертвам придется проводить углубленный (технологический) анализ, среди прочего, (оригинального) программного обеспечения, обновления и оперативные данные, которыми оснащен AIC, чтобы установить точную причину несчастный случай. В то же время производители имеют широкие возможности для защиты от иски об ответственности. Столкнувшись с AIC, PLD не может оптимально защитить интересы потребителей, предоставляя им легкие средства для получения возмещения за ущерб, который они понесли в результате дефектные АПК от производителей.

Возможности для совершенствования действующего законодательства, кроме того, формируют различные негармонизированные европейские режимы ответственности за автотранспортные средства. На сегодняшний день в Европейском союзе действует 28 различных структур. Например, французский «Loi Badinter»15 устанавливает строгий режим ответственности за отсутствие вины в для того, чтобы оценить, должен ли водитель или хранитель автомобиля возмещать ущерб жертвам (кроме водителя)16 дорожно-транспортных происшествий с участием транспортных средств. Ответственность может быть снята только в том случае, если водитель (или смотритель) докажет непростительную ошибку потерпевшего17 . Нидерланды. «Wegenverkeerswet» устанавливает (полустрогую) ответственность перед владельцем или хранителем (примечание: вместо водителя или хранителем) автомобиля, участвующего в дорожно-транспортном происшествии, в результате которого ущерб был причинен немоторизованным участникам дорожного движения.18 По крайней мере 50% нанесенного ущерба должно быть возмещено, если не может быть доказано наличие обстоятельств непреодолимой силы.19 В Соединенном Королевстве Правила о халатности применяются для установления возможности привлечения к ответственности водителя транспортного средства. В таких случаях в Великобритании отсутствует режим строгой ответственности20 , хотя уровень осторожности, требуемый от водителей автомобилей, довольно высок. Прецедентное право объясняет, что водитель, теряющий сознание не по своей вине, тем не менее действует небрежно21, как и водитель, у которого отказали тормоза, когда этот отказ нельзя было предвидеть22.

Однако пострадавшие в результате ДТП с участием автотранспортных средств должны доказать, что водители находились на

---

[14] См. Эрика Палмерини, Федерико Аззарри, Фиорелла Батталья и др., D 6.2, «Руководство по регулированию робототехники», 22 сентября. 2014 г. (RoboLaw 2014), с. 60.

[15] Loi «Tendant à l'amélioration de la des Desicences d'accidents de la кровообращения и à l'accelérération des procédures возмещения ущерба».

[16] См. А. Tunc, « Loi Badinter — десять лет опыта», 3 Maastricht Journal of European and Comparative Law, 1996 (Tunc 1996), p. 330. Статья 3 гласит: «Потерпевшие не могут быть признаны виновными в возмещении вреда, причиненного потерпевшим в результате совершения посягательств на другого человека, который не имеет права быть вменяемым в противовес собственному праву».

[17] См. также Tunc 1997, p. 335.

[18] Возмещение ущерба, причиненного потерпевшим в автомобиле, регулируется общими правилами об ответственности, изложенными в Статья 6:162 Гражданского кодекса Нидерландов.

[19] Marloes de Vos, Верховный суд Нидерландов, 2 июня 1995 г., NJ 1997/700-702, и Saï d Hyati, 5 декабря. 1997 г. Нью-Джерси 1998/400-402. Понятие «Betriebsgefahr» заимствовано из немецкого Straβenverkehrsgesetz.

[20] Или презумпция ответственности , как это называется в шотландском законодательстве.

[21] Roberts v. Ramsbottom [1980] 1 WLR 823, также цитируется в Cees van Dam, European Tort Law, Oxford: Oxford University Press 2006 (Van Dam 2006), at p. 364, сноска 52.

[22] Henderson v. HE Jenkins & Sons and Evans [1970] AC 282, цитируется в Van Dam 2006, на с. 364, сноска 53. Ван Дам далее принимает к сведению дело Worsley v Hollins [1991] RTR 252 (CA), в котором судьи постановили, что иск потерпевшего о халатности был отклонен, поскольку ответчик мог доказать, что, хотя его тормозная система вышла из строя, что привело к повреждению, его микроавтобус недавно обслуживался и прошел свое ТО.

вина, то есть: они действовали небрежно.23 Существенные различия в способах ответственности за транспортные средства рассматриваются во всех государствах-членах, не выгодно для развития, страхования и развертывание AIC в Европе. В любом случае национальные режимы , устанавливающие ответственность водителей транспортные средства должны быть обновлены, чтобы иметь возможность решать вопросы ответственности за транспортные средства без водителя-человека.

Конфиденциальность

Принимая во внимание, что появление технологии AIC многообещающе с точки зрения повышения безопасности на дорогах, что приводит к меньшему ущербу, подлежащему покрытию, также страховые компании отмечают, что при несчастном случае происходит из-за автономных технологий, «требуется обширный опыт анализа программного и аппаратного обеспечения, чтобы знать, как и почему это произошло»24. Один из вариантов оценки причины аварии и, следовательно, помочь в ответе на вопрос, на ком лежит ответственность, может состоять в том, чтобы оборудовать автомобилей с «черными ящиками» или с телематическими решениями, соединяющими AIC с выделенной инфраструктурой и/или с удаленными серверами.25 Цели этих типов технологий, среди прочего, заключаются в регистрации перемещений автономных автомобилей и оперативных решений, которые принимаются либо самим автомобилем, либо водителем, управляющим его движением, а также данные о событиях и объектах в рядом с автономным транспортным средством. Технология «черный ящик» записывает и сохраняет собранные данные внутри транспортного средства и предлагает потенциал для последующей оценки. Телематические технологии могут иметь более широкое Приложения. Данные можно было использовать не только для оценки ошибок и причин ущерба после возникновения аварий, это может иметь даже профилактический эффект. Связь между автомобилями (V2V) и связь между транспортным средством и инфраструктурой можно использовать для предотвращения аварий в режиме реального времени и служит «безопасности, мобильности и экологическим преимуществам» в целом.26 Хотя черный ящик технологии и телематические решения, такие как V2V и V2I (далее именуемые «отслеживание технологии») могут быть перспективными с точки зрения предотвращения аварий и компенсации ущерба, вызванного Аварии AIC, они также создают риски с точки зрения права на (информационную) неприкосновенность частной жизни людей внутри и вблизи автомобилей, оснащенных этими технологиями.

Информационная конфиденциальность граждан строго регулируется в Европейском Союзе Директивой о защите данных (DPD)27 и будет регулироваться еще более строго после вступления в силу Общего регламента по защите данных (GDPR)28. Текущие и будущие рамки предписывают пример того, что уже на этапе проектирования AIC, оснащенных технологией отслеживания, конфиденциальность

---

23 Существует одно правило законодательной обязанности, которое в некоторой степени устанавливает строгую ответственность для водителей автомобилей, приближающихся к перекрестку: «Водитель каждого транспортного средства, приближающегося к переходу, должен, если он не видит, что пешеходного перехода нет двигаться с такой скоростью, чтобы при необходимости можно было остановиться, не доезжая до такого пересечения», как цитируется в Van Dam 2006, p. 365, сноска 57, ссылаясь на Рег. 3 Правил пешеходных переходов (движения) 1941 г., замененных Правилами пешеходных переходов «зебра» 1971 г., SI 1971, № 1524. Защита, которую имеет водитель в этом отношении, является форс-мажором .

24 Йоманс 2014, на с. 18.

25 Йоманс, 2014 г., стр. 18. См. также Джеймс М. Андерсон, Нидхи Калра, Карлин Д. Стэнли и др., Технология автономных транспортных средств – Руководство для политиков, Программа RAND по транспорту, космосу и технологиям, 2014 г. (отчет RAND), на стр. 94-95.

26 Отчет РЭНД, стр. 81.

27 Директива 95/46/ЕС Европейского парламента и Совета от 24 октября 1995 г. о защите лиц с в отношении обработки персональных данных и свободного перемещения таких данных, Официальный журнал L 281 , 23.11.1995 С. 0031 - 0050

28 Предложение о РЕГЛАМЕНТЕ ЕВРОПЕЙСКОГО ПАРЛАМЕНТА И СОВЕТА о защите физических лиц в отношении обработки персональных данных и о свободном перемещении таких данных (Общий регламент по защите данных)
COM/2012/011 окончательный - 2012/0011 (COD). Обратите внимание, что трилог между Европейской комиссией, Советом Европы и Европейским парламентом завершился по окончательному тексту GDPR, однако этот текст еще не был официально опубликован.

должна быть проведена оценка воздействия. Кроме того, «соответствующие технические и организационные меры по защите персональных данных от случайного или незаконного уничтожения или случайной потери, изменение, несанкционированное раскрытие или доступ, в частности, когда обработка включает передача данных по сети и против всех других незаконных форм обработки»29. GDPR регламентирует, что эти меры должны быть максимально «встроены» в новую технологию, в то время как эти меры должны быть, среди прочего, направлены на минимизацию данных. , и должен быть включен по умолчанию30 .

осуществление мер. Кроме того, они «должны обеспечивать уровень безопасности, соответствующий риски, связанные с обработкой и характером защищаемых данных».

Еще одна, еще более недавняя проблема, связана с недавним решением Европейского суда. объявить Safe Harbor Framework, которая лежит в основе многих обменов личными данными между ЕС и Соединенными Штатами Америки, недействительным. Вероятно, технология отслеживания включенные в AIC, будут представлять собой международную передачу (персональных) данных через границы Европейского Союза и, возможно, импортировать эти данные в Соединенные Штаты, например, с помощью облачных вычислений. Европейский суд постановил, что США не обеспечивают адекватного уровня защиты для личных данных, ибо после разоблачений Эдварда Сноудена стало ясно, что власти США такие поскольку Агентство национальной безопасности имеет легкий доступ к персональным данным, обрабатываемым американскими компаниями и учреждениями.31 Суд постановил, что полномочия европейских надзорных органов подрываются. практикой США, что не может быть разрешено решением Европейской комиссии. Это постановление подразумевает, что экспорт персональных данных в Соединенные Штаты больше невозможен на основании рамки безопасной гавани. Хотя Соединенные Штаты и Европейская комиссия в настоящее время ведут переговоры об альтернативном соглашении,32 в то же время обмен личными данными между ЕС и Соединенными Штатами запрещен на основании еще недействующих правил Safe Harbor.

3. Проверьте принцип

В этой части мы проверим, можно ли считать второй принцип EPSRC защитой в будущем от критерии валидности, достаточности/общности и полезности

3.1 Срок действия

При нынешнем состоянии технологий и закона первая часть принципа Люди, а не роботы, являются ответственными агентами, действительно доказал свою актуальность. Это правильное утверждение о природа роботов, разработчики роботов и отношения между роботами и людьми. Может быть всегда либо человек, либо юридическое лицо несет ответственность за действия AIC.

Конкретное создание отдельного юридического лица для АИК в настоящее время кажется надуманным, учитывая нынешнюю ситуацию. технологический и правовой статус АПК, это, кроме того, не способствовало бы решению вопроса об ответственности

---

29 Искусство. 17(1) ДПД; см. также статью 5(1)(eb) и раздел 2 (статья 30 и последующие) о безопасности данных в GDPR.

30 Искусство. 23 Общего регламента по защите данных.

31 Дело C-362/14, Максимилиан Шремс/Facebook [2015].

32 См. последние новости о «Зонтичном соглашении между ЕС и США» (Соглашение между Соединенными Штатами Америки и Европейским союзом о защите личной информации, касающейся предотвращения, расследования, выявления и судебного преследования уголовных преступлений): http : //ec.europa.eu/justice/newsroom/data-protection/news/150908_en.htm (последний доступ 9 марта 2016 г.).

проблемы, решенные, как описано в подразделе 2.2. То же самое верно и для второй части второго принципа, который заявляет, что роботы должны быть разработаны; насколько это возможно в соответствии с существующими законами и основные права и свободы, включая неприкосновенность частной жизни, оказались по-прежнему в силе. С прицелом на технология доказательства (которая должна быть) включена в AIC, эта фундаментальная идея оказалась даже более верно, чем можно было бы предположить по его конструкции.. Как мы видели, на самом деле недостатки Sub 2.2 существующего режима ответственности может быть частично решена за счет интеллектуальных систем сбора и сохранения доказательств, встроенных в АПК. Эти системы сбора и сохранения доказательств должны быть разработаны таким образом, что собранные личные данные максимально защищены: конфиденциальность по замыслу и конфиденциальность по умолчанию должны быть постоянно включены в AIC (технология отслеживания).

3.2 Достаточность/общность

В то же время принцип остается еще достаточным и достаточно широким, чтобы охватить все важные проблемы, которые могут возникнуть при регулировании АПК в реальном мире. Люди, а не роботы ответственные агенты Роботы должны быть разработаны; насколько это возможно для соблюдения существующих законов и основных прав и свобод, включая неприкосновенность частной жизни. Серьезных опасений вроде бы нет упущен из виду. Хотя некоторые авторы, кажется, утверждают, что юридические лица должны быть созданы для автономных интеллектуальных машин, делая роботов ответственным агентом,33 это не было убедительным для многих,34 и, конечно, не для нас.,

Проблемы, связанные с внедрением в общество автономных интеллектуальных автомобилей и их ответственность за ущерб сама по себе, как представляется, не требует отдельного юридического лица. Это просто добавило бы еще один субъект для возложения ответственности. В то же время потребуется существенная переработка системы ответственности, применяемой в настоящее время к реальному миру, в то время как технология все еще находится в стадии разработки этап, несущий риск недостаточного или чрезмерного регулирования.

3.3 Полезность

Поскольку существующие правовые средства не исчерпаны, в том числе путем дальнейшей гармонизации Законодательные режимы ответственности ЕС в сочетании с эффективной технологией доказывания не доказательство, которое подкрепило бы полный сдвиг парадигмы введением AICs в качестве ответственных агенты сами по себе. Поскольку AIC действительно может быть спроектирован и эксплуатироваться в соответствии с существующими законами полезность этого принципа остается очевидной. Однако технологии «черного ящика» и телематические решения такие как V2V и V2I, могут быть многообещающими с точки зрения предотвращения несчастных случаев и распределения ущерба. вызванные авариями ВС, поскольку они также создают риски с точки зрения права на (информационную) конфиденциальность люди внутри и вблизи автомобилей, оснащенных этими технологиями, системы должны будут

---

[33] См., например, Джеймс Бойл, «Наделенные их Создателем?: Будущее конституционной личности», Будущее конституции, 09 марта 2011 г., с. 6, также доступно в Интернете по адресу <http://www.brookings.edu/~/media/research/files/papers/2011/3/09-personhood-boyle/0309_personhood_boyle.pdf> (последний доступ 9 марта 2016 г.). См. также JP Gü nther, F. Mü nch, S. Beck, S. Lö ffler, C. Leroux, & R, Labruto, "Вопросы конфиденциальности и электронной личности в робототехнике. 21-й Международный симпозиум IEEE по интерактивной коммуникации роботов и человека", Париж. 2012 г., цитируется в Christophe Leroux, Roberto Labruto, Chiara Boscarato и др., «Предложение по зеленой книге по юридическим вопросам робототехники», декабрь 2012 г., доступно в Интернете по адресу <http://www.eu robotics.net /cms/upload/PDF/euRobotics_Deliverable_D.3.2.1_Annex_Suggestion_GreenPaper_ELS_IssuesInRobotics.pdf> (последний доступ 28 января 2016 г.); и Роболав 2014, с. 24.

[34] См., например, Питер Асаро, «Роботы и ответственность с правовой точки зрения», через < http://www.peterasaro.org/writing/ASARO%20Legal%20Perspective.pdf>, последний доступ 9 марта 2016 г.; и Лоуренс Солум, «Правосубъектность для искусственного интеллекта», North Carolina Law Review, (апрель 1992 г.), стр. 1231-1287.

включить неприкосновенность частной жизни по замыслу, чтобы защитить эти основные права, изложенные в международных и европейских договорах.35

Крайне важно, чтобы эти требования закона и технологий были выполнены до того, как внедрение и развертывание AIC в обществе может быть выполнено.

4. Вывод

Мы можем тщательно заключить, что принцип 2 принципов робототехники EPSRC, разработанный Британские эксперты по робототехнике и искусственному интеллекту на выездном семинаре, финансируемом EPSRC/AHRC, доказали свою готовность к будущему. когда мы обратились к текущему состоянию законодательства и технологий, связанных с AIC.

Люди, а не AICS, являются ответственными агентами. AIC должны быть разработаны; эксплуатировался до тех пор, пока практически осуществимо для соблюдения существующих законов и основных прав и свобод, включая неприкосновенность частной жизни по замыслу. Таким образом, свидетельствуя о том, что ответ машины хотя бы частично сама машина.

[35] См., например, ст. 7 и 8 Хартии основных прав Европейского Союза и статья 8 Европейской конвенции о правах человека.

Fairdatahandlingandробототехника

Буркхард Шафер, Эдинбургская школа права
Лилиан Эдвардс, Университет Стратклайда

Эта интервенция сочетает в себе принцип 4. Роботы представляют собой искусственные артефакты. Они не должны быть спроектированы таким образом, чтобы обманывать уязвимых пользователей; вместо этого их машинная природа должна быть прозрачной в соответствии с принципом и принципом 2. Роботы должны быть спроектированы; Он предлагает некоторую разработку/спецификацию принципов, которых требует пересечение между ними. Более радикально, может быть, речь идет о соответствующих обязанностях пользователей роботов и даже третьих сторон по отношению к роботам. определенные обязанности по отношению к роботам (или, для тех, кого беспокоит эта формулировка, обязанности по отношению к владельцам роботов, ведут себя определенным образом по отношению к машине.

Роботы создают некоторые уникальные проблемы для методов справедливой обработки данных, проблемы, которые, по крайней мере, частично вызваны их способностью «обманывать», если непреднамеренно, людей, с которыми они взаимодействуют.

Задолго до современных технологий люди разработали методы сохранения конфиденциальности, от занавесок до окон и завесы, от обучения, когда его запах смывается. или ноу-хау. Важно отметить, что они не только защищают информацию, но и позволяют делиться ею и обмениваться ею (шепотом, звукоизоляцией вашей студии).

Закон с его системой правил и исключений часто официально признавал эти низкотехнологичные меры защиты. Стены, которые мы возводим вокруг нас, не только защищают от тепла и дождя, но и защищают информацию и наблюдателей снаружи. огражденный живой изгородью сад — архетип «разумных ожиданий конфиденциальности» и «безопасности в наших домах и жилищах», но также и пространства, в котором данные могут быть более свободно собраны и обменены — исключение домохозяйства из европейского DP Law является ярким примером.

Технологии робототехники угрожают сделать эти низкотехнологичные решения проблемы конфиденциальности все более излишними. to) пригласить в наш дом.

Никто не является героем его домочадцев.

Дом мог предвидеть, что именно они смогут увидеть, понять нормативную (как социальную, так и юридическую) среду, которая сдерживала их от сбора и, самое главное, обмена данными о своем работодателе. The InstrestingofThenormativeEnvironmentTogetherThithTherStingOf the SensoryCapacities WouldThenEnablerationalRiskAssessEndandManagement. ErtoknockfirstbeforeEtertingThebedroom).

Робототехника угрожает этим защитным стратегиям не только потому, что они могут использовать датчики вне визуально-орального спектра, или из-за их мобильности, позволяющей ощущать в пространстве, ранее защищенном. Там, где они имитируют внешний вид людей или даже нечеловеческих животных, даже в тех случаях, когда их роботизированная природа явно невидима (в соответствии с принципом 4). Систематические исследования показали, что мы делаем такие выводы, когда взаимодействуем с роботами. В Интернете полно людей, «подкрадывающихся» к Asimo сзади — теперь датчики Asimo «действительно могут» находиться в его глазах и иметь ограничения зрения, подобные человеческим, но это вполне может быть

Поэтому часть этического дизайна должна также указывать сенсорные способности роботов способами, которые облегчают появление «интуитивной» защиты того типа, который мы используем с другими людьми, и воздерживаться, где это возможно, от вводящих в заблуждение выводов и включать «легкость защитного механизма» в оценку навязчивости, когда может быть сделан выбор между различными датчиками.

За этим стоит защита данных, но справедливое восприятие и методы обработки данных выходят за рамки личных, не говоря уже о конфиденциальных личных данных. Мы защищаем не только данные о нас, но и наши бизнес-идеи, научные или технологические открытия или навыки.).

Таким образом, IP-право является еще одним юридическим ограничением, которое необходимо соблюдать под этим заголовком, и в других странах понятие «справедливой практики передачи данных», которое выходит за рамки DP-права, может быть необходимо. «мой интерес», который заслуживает защиты/компенсации?

Потенциально это также поднимает вопрос, который ведет более радикальным образом за пределы Принципов. Они в основном пытаются установить обязанности, которые разработчики возлагают на людей, взаимодействующих с их машинами. Обязанности разработчиков и возможные обязанности, возложенные на них/роботовладельца.

В качестве простого примера, обеспечение безопасного проектирования роботов может потребовать от третьих сторон раскрытия или обмена определенной информацией с роботом, который в прошлом был юридически привилегированным. выветривание даже этого «случайного» копирования. Подход США, который утверждает, что это было бы копированием для функциональных, а не выразительных целей

– не слова – и поэтому не должны нарушать авторские права – однако, как только машины координируют свои действия, обмениваясь этими данными, даже этот неопровержимый аргумент может достичь своих пределов.

Граждане могут выбрать технологию, предотвращающую их обнаружение датчиками (например, камуфляжную краску для лица — https://cvdazzle.com), но это может означать, что они принимают на себя больший риск того, что робот наткнется на них. Если замешаны третьи лица, это может создать еще более сложные юридические проблемы. пример? Если я намеренно манипулирую процессом обучения, попадаю ли я на территорию закона о неправомерном использовании компьютера?

Базовый закон о небрежности и его различие между действием и бездействием и тем, как небрежность справляется с ним, устанавливая обязанности соседей, будут частью юридического ответа после того, как произошел несчастный случай.

1. Полагаться на этический/социальный долг третьих лиц не манипулировать знаниями, полученными от машины2.
Полагаться только на более строгое юридическое обязательство воздерживаться от определенных заранее
опасных манипуляций с данными3. Не стоит полагаться только на совместную среду, когда речь идет о безопасности и соблюдении законов роботов, которых они создают — в конце концов, не все законы соблюдаются всеми.

Чтобы прояснить, почему эта проблема возникает в контексте обсуждения «сенсорной прозрачности»: ЕСЛИ мы принимаем этическое обязательство, обсуждавшееся выше, т. е. что роботы должны нормально раскрывать, как и с помощью этого то, что они могут ощущать, тогда они неизбежно открывают себя для манипулирования.

Если мы примем 2, то нам придется столкнуться с тем фактом, что на двух крайних точках спектра закон ясно дает понять: это слишком сложно для людей, с другой стороны, я не могу не сжечь свой сарай.

В конце концов, у меня даже нет обязанности не врать незнакомцам, когда дает указания, — если только это не связано с профессиональным советом. Но опять же, отправить ребенка в заблуждение было бы другим предложением. Машины, которые «все еще учатся», аналогичны такой ситуации?

До сих пор обсуждались проблемы, вызванные тем, что люди утаивают/ искажают/манипулируют данными, которые необходимы для безопасной работы робота и которые этически или юридически обязаны.

Но мы также сталкиваемся с этическим выбором дизайна, когда люди сотрудничают и добровольно предоставляют информацию, которую они не обязаны предоставлять по закону, но выбирают или не запрещают из чувства гражданского долга. Должно ли это влиять на статус любого продукта , который производит робот, например, в форме совместного использования выгод?

Это может означать, что не только роботы должны быть идентифицированы как роботы, их сенсоры как сенсоры, но и результаты, сгенерированные роботами, должны быть идентифицированы как машина, а не человек. freetoshare» может потребоваться

Главной темой этого вмешательства, таким образом, является в конечном счете лишь алгоритмическая прозрачность: юридические и этические обязанности влияют на то, когда и как роботы должны раскрывать свои сенсорные способности. чтобы разрешить тайный сбор (некоторых) данных, чтобы иметь более безопасные машины. Там, где сотрудничество, выходящее за рамки юридически требуемого, создает ценность, необходимо обсудить, как объяснить это несправедливым образом, накладывая потенциально еще одну обязанность по раскрытию информации, «сделано роботом».

Хотя это в основном юридические вопросы, для вопроса этического (и законного) дизайна разработчикам также необходимо иметь возможность предвидеть, какой тип взаимодействия ожидать и к какому типу информации они будут иметь законный доступ.

Могут ли роботы нести ответственность за моральные агенты? И почему нас это должно волновать?

Аманда Шарки,

Департамент компьютерных наук и робототехники Шеффилдского университета Шеффилда

Принцип 2. Люди, а не роботы, являются ответственными агентами. Роботы должны быть спроектированы и эксплуатироваться, насколько это практически возможно, в соответствии с существующими законами и основными правами и свободами, включая конфиденциальность.

На первый взгляд, это утверждение или принцип кажется убедительным. Имеет смысл настаивать на том, что люди, а не роботы, являются ответственными агентами. Мы должны помочь ограничить возможное вредоносное использование роботов. Имеет смысл предположить, что роботы должны разрабатываться и эксплуатироваться в соответствии с существующими законами и основными правами и свободами: трудно представить, чтобы кто-то предлагал иное.

Но при дальнейшем рассмотрении становится очевидным, что утверждение не дает никакого обоснования за заявление о том, что люди, а не роботы, являются ответственными агентами, а также не дает каких-либо указаний относительно того, где и когда следует использовать роботов, или последствий, вытекающих из предположения, что роботы не являются ответственными агентами. Заявление поднимает ряд вопросов, заслуживающих дальнейшего обсуждения. (a) Каковы причины предполагать, что роботы, а не люди, являются ответственными агентами? (b) Достаточно ли спроектировать роботов так, чтобы они соответствовали существующим законам, основным правам и свободам? и (c) Если роботы не являются ответственными агентами, должно ли это ограничивать их роли и ситуации, в которых они используются?

(a) Какие есть основания полагать, что ответственными агентами являются люди, а не роботы?

Помимо юридической ответственности, можно выделить две причины этого предположения. Первая основана на различии биологических и механических машин и биологической основе морали. Вторая состоит в том, что общество должно взять на себя ответственность за сердечные факты, которые произвели люди. Мы рассматриваем их по очереди.

(i) Биологические машины против механических машин: возложение на агента ответственности за его действия эквивалентно возложению на него морального агента. Поэтому уместно выделить биологическую основу в форме нравственности в биологических машинах и противопоставить ее отсутствию такой основы в механических машинах, таких как роботы. Патриция Черчленд (2011) обсуждает основу формы нравственности у живых существ и утверждает, что основа заботы о других лежит в нейрохимии привязанности и связи у млекопитающих. -поддержание и избегание окрашивания своих ближайших родственников. Люди и другие млекопитающие беспокоятся о своем собственном благополучии и благополучии тех, к кому они привязаны. Помимо привязанности и сочувствия к другим, у людей и других млекопитающих развиваются более сложные социальные отношения, и они способны понимать и предсказывать действия других. Это вызвано разделением, исключением или неодобрением. Как следствие, у людей появляется внутреннее чувство справедливости.

То же самое в значительной степени относится и к нечеловеческим млекопитающим. Бекофф и Пирс (2009) приводят множество примеров, свидетельствующих об аморальном чувстве справедливости у млекопитающих. Например, обезьяны-капуцины.

работающие за угощение казались оскорбленными и отказывались бы от дальнейшего сотрудничества, если бы видели, что другая обезьяна получила более желанную награду за ту же работу (BrosnananddeWaal, 2003).

Напротив, роботы не заботятся о собственном самосохранении или избегании боли, не говоря уже о боли других. Отчасти это можно объяснить с помощью аргумента, что они не воплощены на самом деле, как живые существа. Живое тело является интегрированной аутопоэтической сущностью (Матурана и Варела, 1980), в отличие от созданной человеком машины. Конечно, можно предположить, что робота можно запрограммировать таким образом, чтобы он вел себя так, чтобы он заботился о своем сохранении или сохранении других, но это возможно исключительно благодаря вмешательству человека.

Социальная ответственность. Многие авторы согласятся с выводом о том, что роботы не являются полноценными моральными агентами. Джонсон и Миллер (2008) утверждают, что роботы и другие вычислительные артефакты не являются полноценными моральными агентами, потому что они «никогда не бывают полностью независимыми от своих создателей-людей». Возлагаются на сердце фактов самих себя, поскольку поведение и результаты работы роботов и компьютерных систем обязательно зависят от людей-дизайнеров и разработчиков. Полезным примером, который они считают, является открывание дверей. человек,механический открыватель двери не был бы

Соответствующие аргументы об отсутствии независимости от дизайнеров-людей выдвигались в прошлом на основании того, что роботы, в отличие от живых машин, никогда не могут считаться полностью воплощенными, поскольку они всегда требовали вмешательства и участия человека в своем развитии (Sharkey and Ziemke, 2001). Суть в том, что роботы и их лежащие в основе системы управления, зависят от вмешательства человека. Роботы могут быть «выпущены» для принятия непредсказуемых решений, но решение позволить им это делать — дело человека и общества. Крайне важно, чтобы человеческая ответственность принималась и признавалась. Джонсон (2006) проводит разумное различие между моральными агентами и моральными сущностями и относит роботов и компьютерные артефакты ко второй категории. Моральные сущности включают в себя творца сердечного факта, факта сердца и пользователя факта сердца, а моральную ответственность нельзя перекладывать на сам факт сердца.

(b) Достаточно ли разработать роботов, чтобы они соответствовали существующим законам и основным правам и свободы, включая неприкосновенность частной жизни?

Основная проблема с предложением о том, что роботы должны быть разработаны для соблюдения существующих законов и основных прав и свобод, и причина, по которой этого недостаточно, состоит в том, что существующие законы и права человека не были сформулированы с учетом технологических достижений, таких как робототехника. Необходимо пересмотреть их в свете таких достижений. Когда они предназначены для того, чтобы казаться друзьями и компаньонами, и в результате мы приветствуем их в наших домах и в интимной обстановке. Здесь нужно ответить на множество вопросов о том, в какой степени информация, к которой у них есть доступ, будет доступна другим, а также очень мало законодательства для решения этой проблемы. «забота» о роботах при минимальном контакте с человеком (например, «Акула и акула», 2012 г.; «Воробей и воробей», 2006 г.), но Закон о правах человека не обеспечивает какой-либо явной защиты от такой ситуации.

были воспитаны на том, что оставляют детей на «опеку» роботов до такой степени, что их привязанность к людям ставится под угрозу (Sharkey and Sharkey 2010), но опять же нет никакого законодательства или прав, которые явно предотвращают такую возможность, кроме тех, которые связаны с детской безнадзорностью. которые могут возникнуть, если люди поставят роботов на позиции власти над людьми.

Когда люди принимают решения о том, как действовать в социальных ситуациях, они должны делать больше, чем следовать правилам или законам. Они принимают решения, основанные на моральном понимании того, что для них неуместно или неуместно. Аркин (2009), например, утверждал, что в боевой ситуации солдат-роботов можно запрограммировать на следование правилам, что приведет к более этичному поведению, чем то, которое иногда демонстрируют солдаты-люди в пылу битвы. мотивированные местью для совершения военных преступлений. С другой стороны, роботы не реагировали эмоционально и могли быть запрограммированы с помощью «этического губернатора» оценивать действия перед их совершением и выполнять только те действия, которые считались морально допустимыми.

Различные авторы возражали против возможности программировать роботов для принятия моральных решений. оценка более широкой картины, понимание намерений, стоящих за действиями людей, а также понимание ценностей и предвидение направления, в котором разворачиваются события» (2013, A/HRC/23/47). Суть в том, что непредсказуемое разнообразие социальных ситуаций, которые могут возникнуть на поле боя, означает, что маловероятно, что набор заранее запрограммированных правил относительно надлежащего реагирования может быть применим.

В интересной статье о требованиях к созданию роботов с тем, что они называют «моральной компетентностью», Малле и Шойц (2014) утверждают, что, среди прочего, роботам потребуется сеть моральных норм, чтобы знать, что есть и что морально неприемлемо. Старые изучают и развивают сеть моральных норм на основе обратной связи, данной моей реакции на их действия. правильное и неправильное можно улучшить, обучив демонов моральным историям

(RiedlandHarrison, 2016), и требуя от них обратного проектирования человеческих ценностей, которые они представлять.

По общему признанию, трудно исключить возможность того, что в будущем роботов можно будет обучить или воспитать нравственными, но есть ряд причин скептически относиться к вероятности успеха. Причины для скептицизма включают слабость биологической основы робота, формальную мораль, как обсуждалось ранее. Как уже говорилось, отдельный робот даже не заботится о своем теле, не говоря уже о человеческом — он не испытает никакой боли, если, например, одно из его колес будет удалено. приводя примеры роботов, развивающих хорошее , поддающееся обобщению понимание различий между правильным и неправильным.

поведение, такое как роботы, запрограммированные Winfield etal (2014) для принятия мер по предотвращению падения других роботов в яму, которые описываются как демонстрирующие что-то, что можно описать как этическое поведение.

Роботов, о которых идет речь, можно законно хвалить или порицать за их действия.

(c) Если роботы не являются ответственными агентами, должно ли это ограничивать социальные роли, которые им отводятся, и ситуации, в которых они используются?

Исходное утверждение о том, что роботы не являются ответственными агентами, не поясняет, что это означает для развертывания роботов. Здесь утверждается, что существуют веские причины для ограничения социальных ролей и полномочий роботов по принятию решений. Чтобы понимать социальные ситуации, но также и потому, что люди должны иметь право на то, чтобы решения о их жизни и смерти принимали их собратья-люди. Аналогичный аргумент можно было бы привести и в отношении роботов-полицейских, которым также можно было бы решать вопросы жизни и смерти (или серьезного ранения) вдали от поля боя.

Этот аргумент можно, и я бы сказал, следует распространить и на другие виды решений, в которых роботы могут ограничивать свободу людей. Робот, поставленный на роль учителя, должен был бы принимать решения о таких вещах, как, например, когда наказывать сдерживать детей или когда их воспитывать. чтобы не дать им сделать что-то опасное или рискованное. Робот-няня должна была бы принять такое же решение в отношении своих младших подопечных. Дело в том, что все эти решения, скорее всего, включают в себя моральные суждения и оценки социальных ситуаций, и по причинам, которые уже обсуждались, робот вряд ли сможет сделать правильный выбор. Следует проявлять осторожность, чтобы сохранить человеческий контроль, участие и ответственность в решениях, которые повлияют на жизнь людей. Риски автоматических решений, влияющих на нашу жизнь, уже существуют, но роботы, которые могут создавать компетентных социальных акторов, делают эти риски еще более распространенными.

Резюме: Легко согласиться с принципом EPSRC о том, что роботы не являются ответственными агентами, но даже это краткое соображение оказывается недостаточным для руководства будущими действиями.

Роботы, запрограммированные на соблюдение закона и уважение прав и свобод людей, не будут понимать социальные ситуации и не смогут постоянно принимать правильные моральные решения относительно социальных ситуаций людей. Необходимо проявлять осторожность, чтобы избегать или сводить к минимуму автоматическое и алгоритмическое принятие решений в любых ситуациях, в которых требуется человеческое суждение. Еще большая осторожность требуется в случае с роботами, которые создают иллюзию того, что они понимают.

Рекомендации

Аркин, Р. (2009). Управление летальным поведением автономных роботов. Обзор Чепмена-Холла. Компьютеры и образование, 58 (3), 978–988.

Бекофф, М., и Пирс, Дж. (2009) WildJustice: The Moral LivesofAnimals. Университет Чикаго Пресс, Лондон.

Brosnan, SFanddeWaal, FB (2003). Обезьяны отвергают неравную оплату. Природа, 425,297-99

Черчленд, П. (2011) Braintrust: Что нейробиология говорит нам о морали. Принстон Юниверсити Пресс, Оксфорд.

Heyns, C. (2013). Отчет Специального докладчика о внесудебных, суммарных или произвольных казнях, A/HRC/23/47.

Джонсон, Д. Г. (2006). Компьютерные системы: моральные объекты, но не моральные агенты. Этика и информационные технологии, 8 (4): 195–204.

Джонсон Д.Г. и Миллер К.В. (2008 г.). Этика и информация
Технология (2008) 10: 123–133

Малле, Б. Ф., и Шойц, М. (2014). Моральная компетентность в социальных роботах. Международный симпозиум IEEE по этике в инженерии, науке и технологиях (стр. 30–35). Представлено на Международном симпозиуме IEEE по этике в инженерии, науке и технологиях, июнь, Чикаго, Иллинойс: IEEE.

Матурана, Х.Р. и Варела, Ф.Дж. (1980). Автопоэз и познание — реализация жизни.
Дордрехт, Нидерланды: D.ReidelPublishing

Ридл, М.О., и Харрисон, Б. (2016) Использование историй для обучения человеческим ценностям искусственным агентам. В материалах 2-го Международного семинара по искусственному интеллекту, этике и обществу, Феникс, Аризона.

Sharkey, AJC, & Sharkey, NE (2012). Бабушка и роботы: этические проблемы ухода за пожилыми людьми с помощью роботов. Этика и информационные технологии, 14(1), 27–40.

Sharkey, NE, & Sharkey, AJC (2010). Плачущий позор робонянь: аэтическая оценка.
Интерактивные исследования, 11 (2), 161–190.

Sharkey, NE & Ziemke, T. (2001). Mechanisticvs.PhenomenalEmbodiment-CanRobotEmbodimentLeadtoStrongAI
CognitiveSystemsResearch, 2,4,251-262

Воробей, Р., и Воробей, Л. (2006). В руках машин? Будущее ухода за престарелыми. Минданд
Машина, 16,141–161.

Уинфилд, А. Ф., Блюм, К. и Лю, В. (2014) На пути к этичному роботу: выбор внутренних моделей, последствий и этических действий. В М. Мистри, А. Леонардис, М. Витковски и К. Мелуиш (ред.)
Достижения в области автономных робототехнических систем: Материалы 15-й ежегодной конференции TAROC-2014 (стр. 85–96). Бирмингем, Великобритания, 1–3 сентября.

Дополнительные мысли о конфиденциальности, безопасности и обмане

Том Сорелл, Уорикский университет
Хизер Дрейпер, Бирмингемский университет

Пять принципов робототехники, сформулированные во время ретрита AHRC-EPSRC в 2010 году, — это не последнее слово в роботоэтике, а одно из первых слов. Это далеко не так. Далее мы обсудим (а) проблему с Принципом 2, не принимая ее во внимание; (б) противоречия между Принципами 2 и 3; и (в) некоторый скептицизм в отношении применения Принципа 4.

Конфиденциальность

Принцип 2 требует, чтобы роботы эксплуатировались в соответствии с существующими законами и основными правами и свободами, включая право на неприкосновенность частной жизни. неприкосновенность частной жизни (статьи 7 и 8), не является фундаментальным в более ранних договорах о правах человека, таких как Международный пакт о гражданских и политических правах (см. статью 17). многие теоретики прав человека отрицают существование иерархии прав человека, в которой одни из них являются более фундаментальными, чем любые другие. Статья 5). Таким образом, даже если будет достигнуто соглашение о том, что такое неприкосновенность частной жизни, необходимость уважать право на неприкосновенность частной жизни не обязательно будет иметь преимущественную силу.

Личная конфиденциальность иногда понимается как контроль над информацией о себе. Роботы-помощники в частности и социальные роботы в целом часто предназначены для сбора информации о людях, с которыми они взаимодействуют. контролировать свою информацию, потому что люди, чья информация это может в принципе дать согласие на ее сбор и хранение. Поскольку использование их информации зависит от их согласия, контроль не переходит в руки других людей: согласие является формой контроля.

Согласие, однако, не обязательно решает все вопросы о надлежащем использовании личной информации. Во-первых, существует разница между сбором информации на разовой или периодической основе и более или менее непрерывным сбором информации в режиме реального времени. Последствия второго труднее предсказать и дать согласие на продвижение вперед, чем Последствия первого. Можно даже спорить, что не существует такой вещи, как должным образом информированное согласие на непрерывный мониторинг и отслеживание живого-на-уходе-робота именно потому, что невозможно предсказать или даже представить заранее, каким будет опыт жизни с роботом.

Вместо того, чтобы определять пределы конфиденциальности только на основании того, на что пользователь соглашается, предоставляя достоверную информацию, можно также вскоре полагаться на аргументы относительно пределов конфиденциальности, основанные на

Роботы-помощники для пожилых людей часто призваны помогать в поддержании их автономии — то есть их способности выбирать и иметь набор навыков — способность умываться, убираться, готовить, кормить себя и т. д. — достаточных для самостоятельной жизни. информация для раскрытия. Arobot, предназначенный для поддержания автономии пожилого человека, может частично оцениваться по тому, обладает ли пожилой человек такой же широтой свободы, как и взрослый без посторонней помощи, чтобы принимать решения обо всех аспектах своей жизни, включая раскрытие информации.

Противоречие между автономией (и конфиденциальностью) и безопасностью

Один из способов, с помощью которого стандартный взрослый проявляет автономию, заключается в том, что он сам решает, на какой риск идти. На моральную допустимость риска, конечно же, влияют затраты для других. Если другие подвергаются опасности или полагаются на опасные действия по спасению независимого, идущего на риск, то это может быть аргументом против принятия риска из-за бремени, ограничивающего автономию, которое создает для других. наименее временно.

В случае с пожилыми людьми, которым оказывается помощь, краткое описание конструкции робота обычно сочетает в себе безопасность и Автономность. Робот помогает пользователям вести собственную жизнь, а также следит за пользователем и его состоянием здоровья в чрезвычайных ситуациях. Если обнаружатся чрезвычайные ситуации или отклонения от нормы, он может поднять тревогу.

В случаях, которые наиболее интересны с точки зрения теории морали, пользователь готов пойти на относительно небольшой риск — скажем, на риск падения — ради того, чтобы продолжать вести повседневную жизнь таким же образом, как и в молодости. делиться информацией о падении. Это потому, что поддержание

автономия считается главной целью робота-компаньона. Однако, если главной целью является обеспечение безопасности пользователя, у него может не быть этого права. Чтобы продлить.Низкотехнологичный телеуход, подвесной сигнал тревоги носится пользователем, и он или она может решить, вызывать помощь или нет. Ценность автономии поддерживает эту норму.

Пользовательские переопределения также могут быть включены в компаньон-робот-дизайн, где выбор образа жизни пользователя, если о нем сообщили друзьям и родственникам, может спровоцировать принудительное вмешательство со стороны этих людей. для людей среднего возраста; в противном случае своего рода эйджизм ограничивает автономию людей-солдат, и разработчики роботов и разработчики государственной политики обращаются с ними хуже, чем со взрослыми .

без посторонней помощи. И это ограничение трудно защитить без эйджизма. Это противоречит тому, что существует аргумент в пользу того, что азартные игры небезопасны в любом возрасте.

Напряжение между автономией и реабилитацией

Роботы-помощники и некоторые несоциальные роботы предназначены для того, чтобы помочь пожилым людям восстановить утраченные способности, а не просто тренировать те, которые у них есть самостоятельно. Насколько велика ответственность пожилых пользователей или других пользователей, которые должны сотрудничать с рутинной реабилитацией, запланированной и управляемой роботами? И у любого, кто платит за введение робота в дом пользователя, может существовать обязательство сотрудничать. Если такое совместное предприятие не признано, должно быть создано место для автономного отказа от выгодной реабилитации. В конце концов , автономный отказ от медицинского вмешательства не может быть юридически проигнорирован .

Что, если помощь робота пожилому человеку предоставляется при условии, что он соглашается сотрудничать с реабилитацией, которая может быть предложена в будущем? В этом случае автономный отказ может быть преодолен автономным обязательством сотрудничать. он должен быть предметом явного договора, который пользователь заключает, в котором указываются обязанности, которые пользователь берет на себя в обмен на предоставление робота. Такой договор может существовать между пользователем и местными властями.
Обязанности по договору, конечно, не исключают права.

Обман

Наконец, мы подходим к Принципу 5. Он призывает к прозрачности при проектировании роботов и запрещает обман уязвимых. Наш скептицизм в отношении этого принципа основан на низком пороге обмана, установленном в некоторых источниках по роботоэтике. Обман — это намеренное создание ложных убеждений. Обман обычно неправилен, потому что обманщик хочет манипулировать обманутым человеком, чтобы тот сделал что-то, что служит интересам обманщика. Робот сам по себе не имеет намерения обмануть, но обманчив ли его детский дизайн? Работает ли он, заставляя ребенка думать, что другие дети не помогают ему? Ответ здесь: «Нет». до того, как они достигнут 6-летнего возраста. Антропоморфизация не является случаем самообмана, равно как и моделирование обучающего робота гуманоидного ребенка не является случаем обмана.

Можно было бы рассматривать его как представление о ребенке. Аналогичный вывод можно сделать и о роботе-паро.
Это не случай обмана со стороны производителя Paro или самообмана. Пациенты с деменцией, похоже, используют Paro и нероботизированные куклы во многом так же, как очень маленькие дети используют мягкие игрушки.

Роботы обеспечивают присутствие, в большей степени, чем самедог или кошка. Для получения утешения от Паро не имеет решающего значения то, что кто-то думает, что это настоящий тюлень, а кто-то думает, что он жив — достаточно того факта, что его поведение имитирует поведение животного. Тот же вывод, по-видимому, можно сделать с помощью параллельных рассуждений для обучающих роботов-гуманоидов.

Комментарий к семинару AISB по принципам робототехники

Эмили С. Коллинз

Университет Шеффилда, Шеффилд, Великобритания.

1. Введение

На регулирование роботов в реальном мире направлен следующий принцип: № 4.
Роботы — это искусственные артефакты. Они не должны быть разработаны таким образом, чтобы обманывать уязвимых пользователей; вместо этого их машинная природа должна быть прозрачной.

В этом комментарии будет предложена критика этого принципа в соответствии со следующими

критериями: а. Период действия. Верны ли принципы как утверждения о природе роботов (например, о том, что они являются инструментами и продуктами), о разработчиках роботов и об отношениях между роботами и людьми (например, о том, что роботы должны иметь прозрачную конструкцию), или они онтологически ошибочны? , неточные, устаревшие или вводящие в заблуждение.

Критика разбивает этот принцип на два основных утверждения, которые я считаю его составляющими: 1. Роботы
не должны разрабатываться таким образом, чтобы обманывать уязвимых пользователей.
2. Машинная природа должна быть прозрачной.

Я утверждаю, что оба составных утверждения, составляющих этот принцип, в корне ошибочны из-за неопределенной природы критических терминов: «обманчивый», «уязвимый» и «машинная природа», и что как таковой принцип в целом вводит в заблуждение.

Рис. 1. Левая панель: «Эксо», сокращение от «экзоскелет», представляет собой носимого робота, который помогает парализованным пациентам ходить. Правая панель: два робота-млекопитающих MIRO, пример «социального» робота.

Для целей настоящего комментария робот определяется как искусственно созданный искусственный факт, в частности инструмент, с помощью которого пользователь-человек может увеличить существующее состояние, например, предоставляя человеку, который не может ходить, возможность ходить с помощью машин, или предоставляя пользователю расширенную форму развлечения, например, робота-компаньона (рис. 1). Этот комментарий будет посвящен, в частности, биомиметическим [1], социальным роботам и их роли в качестве инструментов для ставок пользователей. Социальный робот здесь определяется как устройство с некоторой автономией и физическим присутствием, способное социально взаимодействовать с людьми, и поэтому можно ожидать, что оно вызовет какую-то эмоциональную реакцию у своего пользователя [2]. Вот наша первая проблема, еще до того, как будет рассмотрен принцип: для определения «робота» нужно, по крайней мере, определить применение робота и степень его возможностей. Существуют взаимоисключающие типы роботов, которые потенциально могут обманывать пользователей различными способами, зависящими от того, как взаимодействуют с каждым роботом. Промышленные, мобильные, сервисные, образовательные, космические и социальные роботы, и это лишь некоторые из них, имеют разную морфологию и разные наборы ожиданий от своих пользователей. Ни один из Принципов робототехники не начинается с определения «робот», поэтому я дал свое собственное оп

2 Роботы не должны разрабатываться таким образом, чтобы обманывать уязвимых пользователей.

Во-первых, давайте начнем с вопроса, что является «обманчивым»? В этом контексте это робот, помеченный как вводящий в заблуждение, поэтому лучше спросить, как робот обманывает настолько, что это противоречит этому принципу?

Роботов разрабатывают так, чтобы они напоминали живых существ. То, что известно о динамике человека и животного, используется для создания звероподобного поведения и морфологии в конструкции социального робота. Биомиметика по определению означает проектирование по природе, путем имитации моделей, систем и элементов природы с целью решения человеческих проблем. Роботы — это инструменты, продукты для использования, целью которых является решение человеческих проблем. Здесь сами принципы проектирования, которые лежат в основе природы биомиметического социального робота и того, что требуют разработчики роботов, обусловлены тем, что можно было бы назвать «обманом»: попыткой имитировать живые существа для улучшения робота и его пользователя.

Животноподобные роботы, такие как Paro [3] и «FurReal Friends Lulu Cuddlin Kitty», производимые Hasbro (рис. 2), используются терапевтами аналогично терапии с помощью животных (AAT) [4]. при этом животное может быть привлечено к существующему сеансу терапии, чтобы помочь с социальной поддержкой (как в случае с групповой терапией), или использоваться один на один, чтобы помочь клиенту или пациенту сосредоточиться во время терапии. Эти роботы служат определенной цели: выглядеть как животные и помогать терапевту. Однако их существование, хотя и основано на лечении с участием живых существ, не заменяет животных. Животные в AAT считаются ко-терапевтами. К ним относятся с уважением, ожидаемым от живых существ, и удаляют из сеансов, на которых может быть причинен вред.

Рис. 2. Левая панель: Терапевтический робот «Паро». Правая панель: «Друзья FurReal». Лулу Кадлин Китти.

их или где они сами являются разрушительными [5]. Этот пример демонстрирует, что робот, созданный с расчетом на обман — чтобы он выглядел как животное — с намерение быть использованным уязвимыми группами населения - лицами, проходящим терапию, - не предназначено для эксплуатации, как это определено попыткой убедить пользователя, что похожий на животное робот действительно жив. Вместо этого они используются для запуска ассоциативных воспоминания о других живых существах. Трудно передать эту мысль, но важен нюанс. Эти роботы созданы не для того, чтобы быть убедительными животными. Они созданы, чтобы быть убедительными роботизированными инструментами, и чтобы идеи природы заимствовано.

Во-вторых, что значит «эксплуатировать уязвимых пользователей»? Что такое уязвимый пользователь? Является ли уязвимость единственным состоянием бытия? И если да, то в какой момент может можно считать или больше не считать уязвимым? Действительно, кто может решать в какой момент человек стал достаточно уязвимым, чтобы его состояние Арт робот взят у них? [1]

В медицине существует стандартизированное определение уязвимых групп, в для которых существуют определенные области уязвимости (например, [7]). То, как уязвимое лицо эксплуатируется вводящими в заблуждение роботами, зависит от того, где уязвимость личность лжет. Например, медицинские области включают экономическую уязвимость. Рассмотрим эмоциональную эксплуатацию страха, созданную популистскими СМИ. распространяет убеждение, что работа человека может быть под угрозой из-за роботов которые обманчиво изображаются как более продвинутые, чем они есть на самом деле. Хотя мы можем предположить, что принцип не относится к такой уязвимости как к экономической (хотя на самом деле мы не можем этого предположить; часть проблемы с этими Принципы робототехники состоят в том, что они определяются вовсе не так, а для Для целей этого комментария предположим, что упоминаемая уязвимость to является физическим, а не концептуальным). Поэтому, возможно, давайте предположим, что под «уязвимыми» принцип относится к группам. Предположим также, что обычный пользователь будет знать, когда робот является роботом, если только этот робот не будет настолько исключительно реалистичным

[1] Для примера этого в художественной литературе см. самую раннюю публикацию Исаака Азимова, Робби [6]. Страх перед роботами, эксплуатирующими уязвимых, был давним в сообщества робототехники, но при оценке этого вопроса следует отделять вымысел от фактов.

чтобы сойти за живого. Чтобы сойти за живого, робот должен был бы, и этот список ни в коем случае не является исчерпывающим, двигаться, реагировать, моргать, дышать и издавать звуки синхронно, а также быть морфологически точным. Такой технологии не существует. Таким образом, принимая во внимание современное состояние дел, которое существует в настоящее время, например, социальных роботов, которые находятся в центре внимания этого комментария, проблема возникает из-за того факта, что именно наиболее уязвимые слои населения получают наибольшую выгоду от своих действий. использовать. Двумя наиболее уязвимыми группами обычно считаются пожилые люди и несовершеннолетние, а также лица с когнитивными нарушениями внутри этих групп.

Для целей настоящего комментария давайте сосредоточимся на уязвимых группах пожилого населения. Вышеупомянутый робот Paro — это усовершенствованный интерактивный робот, предназначенный для оказания физической и эмоциональной поддержки больным и пожилым людям не сам по себе, а с помощью практикующего врача, обученного робототерапии (RAT). У лиц, страдающих деменцией и другими состояниями снижения когнитивных функций, эмоциональные способности не снижаются в той же степени, что и когнитивные функции [8]. Это позволяет терапевту осмысленно применять психологическую и эмоциональную терапию с помощью таких статических устройств, как Паро, который разработан, чтобы напоминать живое существо, чтобы его можно было держать и суетиться [9]. Здесь обман напоминает тот, что наблюдается при кукольной терапии.

В кукольной терапии лица, ухаживающие за больными с болезнью Альцгеймера, используют куклы, которые напоминают живых младенцев, чтобы облегчить тревогу и доставить радость тем, кто страдает деменцией. Это достигается за счет введения целенаправленной и полезной, но физически безвредной деятельности, а именно ухода за куклой (например, [10]). Несмотря на споры [11], такие методы лечения, которые вводят в процесс ухода реалистичные инструменты фокусной точки, получили высокую оценку за улучшение качества жизни (QoL) пациентов, и такие исследования включают исследования, в которых изучалось влияние использования роботов, похожих на животных, в терапии. тоже [12].

Качество жизни — комплексное измерение, охватывающее эмоциональные, социальные и физические аспекты жизни человека. Он существует в континууме, вне сферы дихотомии «или/или», где х считается плохим, а у хорошим. Если инструмент, который является роботом, используется с уязвимым населением, обладающим умственными способностями, которые можно использовать для облегчения страданий людей в этом населении, вопрос о том, должен ли этот инструмент существовать, становится расплывчатым и слишком сложным, чтобы ответить на него. с одним заявлением. Споры сводятся к тому, насколько мы должны обманывать уязвимых и в какой момент это становится эксплуатацией в негативном смысле. Когда это соображение противопоставляется улучшению качества жизни лиц, страдающих неизлечимыми нейродегенеративными заболеваниями, становится ясно, что этого четвертого принципа недостаточно. Он в корне ошибочен, потому что его составные термины остаются неопределенными. Без знания того, что на самом деле имеется в виду под эксплуатацией уязвимых, весь принцип вводит в заблуждение.

Если эксплуатируемая вещь — это само снижение когнитивных функций, и робот извлекает выгоду из уязвимой природы человека, но для целенаправленного результата улучшения качества жизни этого человека, разве это не положительно? Когда там

нет другой альтернативы доступу к остаткам эмоционального фактора, страдающего слабоумием, кого-то, кто в противном случае мог бы испугаться живого животного, в остальном утешающего, где реальный вред? Заключается ли зло в умах тех, кто не страдает и не является свидетелем того, что они сами в размышлении считают печальным состоянием? И если это так, то не должны ли мы еще больше пытаться проецировать себя в разум уязвимых и ценить эту ситуацию такой, какая она есть? Попытка оказать помощь с использованием всех и любых доступных инструментов, предпринятая с доброй волей и под наблюдением лиц, осуществляющих уход, которые знают всю степень ущерба, который нейродегенеративные заболевания наносят как пациентам, так и их близким, наблюдающим за ними.

3 Машинная природа должна быть прозрачной

Давайте рассмотрим здоровое население, которое наблюдает за роботами. Как уже говорилось ранее в этом комментарии, я считаю, что не существует такой роботизированной технологии, которая была бы совершенно обманчивой. Даже самые сложные роботы — это явно роботы. Пользователь может полагать, что ИИ робота более совершенен, чем он кажется на первый взгляд, но, по крайней мере, исходя из моего собственного опыта в лаборатории, я считаю, что любого периода времени с роботом достаточно, чтобы пользователь установил приблизительное представление о нем. достаточное приближение к его ограничениям, так что любая первоначальная переоценка возможностей робота вскоре перекрывается реальностью. Что касается тех групп населения, которые достаточно уязвимы, чтобы обмануться, полагая, что робот более продвинутый или более «живой», чем он есть на самом деле, я считаю, что не робот должен быть спроектирован по-другому, а люди-пользователи или клинические специалисты RAT, которые должны быть обучены использовать свой инструмент, свой продукт-робот, наиболее эффективным и позитивным образом.

4 Резюме

Робот, который настолько совершенен, что полностью обманывает пользователя, заставляя его поверить, что это что угодно, но только не машина, — это то, что я не могу себе представить в ближайшее время. Для тех людей, которые достаточно уязвимы, чтобы быть убежденными, что робот, который явно является машиной, на самом деле жив, я рекомендую как можно более объективно и широко рассмотреть все положительные преимущества, которые могут возникнуть в такой ситуации. Чтобы рассмотреть, что на самом деле означает эксплуатация уязвимого, и, возможно, перефразировать сценарий с якобы положительными результатами для уязвимого пользователя, не используя термин «эксплуатация», а вместо слова «помощь»:

> Роботы — это искусственные артефакты, но они являются инструментами, помогающими нам, и могут быть разработаны с использованием известных нам принципов работы, в том числе биомиметических. Роботы, созданные обманчивым образом для облегчения страданий уязвимых пользователей, должны быть доведены до сведения общественности об их машинной природе.

лица, осуществляющие уход за этими уязвимыми пользователями. Пусть лицо, осуществляющее уход, несет ответственность за улучшение качества жизни своих пациентов любыми необходимыми безопасными средствами.

Рекомендации

1. Т. Дж. Прескотт, М. Дж. Пирсон, Б. Митчинсон, Дж. К. В. Салливан и А. Г. Пайп, «От крысиных вибрисс до биомиметических технологий для активного осязания», Журнал IEEE Robotics and Automation, том. 16, нет. 3, стр. 42–50, 2009 г.
2. Коллинз Э.К., Миллингс А. и Прескотт Т.Дж. «Привязанность к вспомогательным технологиям: новая концепция», Материалы 12-й Европейской конференции AAATE (Ассоциация по развитию вспомогательных технологий в Европе), 2013 г.
3. Т. Шибата, «Робот с мысленной фиксацией (PARO)». [Онлайн], http://www.paro.jp.
4. М. Р. Бэнкс, Л. М. Уиллоуби и В. А. Бэнкс, «Терапия с помощью животных и одиночество в домах престарелых: использование роботов в сравнении с живыми собаками», Журнал Американской ассоциации медицинских директоров, том. 9, нет. 3, стр. 173–177, 2008.
5. С. Брукс, «Психотерапия с помощью животных и психотерапия с участием лошадей». Работа с травмированной молодежью в сфере защиты детей, стр. 196–218, 2006 г.
6. И. Азимов, «Странный товарищ по играм», Super Science Stories, стр. 67–77, 1940.
7. Ассоциация ВМ и др., «Защита уязвимых взрослых — набор инструментов для общего Практики», 2011.
8. Магаи К., Коэн К., Гомберг Д., Малатеста К., Калвер К. Эмоциональное выражение на средней и поздней стадиях деменции // Международная психогериатрия . 8, нет. 03, стр. 383–395, 1996.
9. EC Collins, TJ Prescott и B. Mitchinson, «Говорить это с помощью света: экспериментальное исследование аффективной коммуникации с использованием микроробота», в Biomimetic and Biohybrid Systems, стр. 243–255, Springer, 2015.
10. М. Эренфельд, «Использование терапевтических кукол с психогериатрическими пациентами», Игровая терапия. со взрослыми, стр. 291–297, 2003 г.
11. Г. Митчелл, «Использование кукольной терапии для людей с деменцией: обзор: Гэри Митчелл представляет аргументы за и против этого спорного, но популярного вмешательства», Уход за пожилыми людьми, т. 1, с . 26, нет. 4, стр. 24–26, 2014.
12. М. Херинк, Дж. Альбо-Каналс, М. Валенти-Солер, П. Мартинес-Мартин, Дж. Зондаг, К. Смитс и С. Анисуззаман, «Изучение требований и альтернативных роботов-питомцев для роботизированной терапии пожилых людей». взрослых с деменцией», в Social Robotics, стр. 104–115, Springer, 2013.

Почему мой робот так себя ведет?
Проектирование прозрачности для проверки в реальном времени автономные роботы

Андреас Теодору [1] и Роберт Х. Уортэм [2] и Джоанна Дж. Брайсон [3]

Абстрактный. Принципы робототехники EPSRC диктуют необходимость обеспечения прозрачности в роботизированных системах, однако исследования, связанные с этим, находятся в зачаточном состоянии. Настоящая статья знакомит читателя к необходимости иметь прозрачных для инспекции интеллектуальных агентов. Мы дать надежное определение прозрачности как механизма раскрытия процесса принятия решений роботом путем рассмотрения и расширения на другие известные определения, найденные в литературе. Документ завершается рассмотрением потенциальных возможностей, которые разработчики проектных решений должны учитывать. учитывать при разработке прозрачных систем.

1. ВВЕДЕНИЕ

Прозрачность, по нашему мнению, является ключевым элементом, касающимся этического последствия как разработки, так и использования искусственного интеллекта, тема, вызывающая все больший общественный интерес и дискуссия. Мы часто используем философские, математические и биологически вдохновленные методы для создание искусственных интерактивных интеллектуальных агентов, но мы относимся к ним как к черные ящики без понимания того, как лежащие в основе принятие решений работает.

Природа «черного ящика» интеллектуальных систем, таких как контекстно-зависимые приложения, делает взаимодействие ограниченным и часто неинформативным для конечного пользователя [14]. Ограничение взаимодействия может негативно сказаться производительность системы или даже поставить под угрозу функциональность система. Представьте себе автономную роботизированную систему, созданную для обеспечения медицинская помощь пожилым людям. Однако пожилые люди могут бояться и не доверять системе. Они могут не позволить роботу взаимодействовать с ними. При таком сценарии жизни людей угрожает опасность, поскольку они может не получить необходимое лечение вовремя, так как человек, наблюдающий за системой, должен обнаружить отсутствие взаимодействия и вмешаться. И наоборот, если пользователь-человек слишком доверяет роботу, он может привести к неправильному использованию, чрезмерной уверенности и неиспользованию системы [13]. В нашем Пример медицинского робота, если агент работает со сбоями, а его пациенты не знают о его отказе функционировать, пациенты могут продолжать используя робота, рискуя собственным здоровьем. Роботы в обоих случаях нарушают первый принцип робототехники EPSRC, помещая человека жизнь в опасности [1].

Во избежание подобных ситуаций необходима надлежащая калибровка доверия между людей-операторов и их роботов критически важно, если не необходимо, особенно в сценариях высокого риска, таких как использование роботов. в армии или в медицинских целях [9]. Происходит калибровка доверия когда конечный пользователь имеет мысленную модель системы и полагается на системы в пределах возможностей системы и осознает ее ограничения [6].

Мы считаем, что обеспечение прозрачности не только выгодно для конечных пользователей, но и для разработчиков интеллектуальных агентов. В режиме реального времени отладка механизма принятия решений роботом может помочь разработчикам исправлять ошибки, предотвращать проблемы и объяснять возможные отклонения в производительность робота. Мы предполагаем, что при правильной реализации прозрачности разработчики смогут проектировать, тестировать и отлаживать свои агентов в режиме реального времени — аналогично тому, как разработчики программного обеспечения работать с традиционной разработкой и отладкой программного обеспечения.

Несмотря на эти возможные преимущества прозрачности в интеллектуальных системах, существует мало исследований в области прозрачных агентов и даже меньше внедрения прозрачных агентов. Кроме того, существуют непротиворечивости в определениях прозрачности и критериях робот считается прозрачной системой. В этой статье мы будем представить противоречивые определения, встречающиеся в литературе, и попытаться чтобы дополнить их нашими собственными. Кроме того, в третьем разделе этой статьи мы обсудим дизайнерские решения, которые нужны разработчику учитывать при проектировании прозрачных роботизированных систем.

Мы специально используем термин «интеллектуальный агент» для обозначения комбинации как программного, так и аппаратного обеспечения автономного робота. системы, работая вместе как актер, живя и изменяя мира [3]. В этой статье слова «робот» и «агент» взаимозаменяемы.

2 ОПРЕДЕЛЕНИЕ ПРОЗРАЧНОСТИ

Несмотря на преобладающее использование ключевого слова прозрачность в EPSRC «Принципы робототехники», исследования по обеспечению прозрачности систем все еще находятся в зачаточном состоянии. На протяжении многих лет очень немногие публикации были посвящены необходимости прозрачных систем и даже меньше пытались удовлетворить эту потребность. Каждое исследование дает свое собственное определение ключевого слова, не исключая других. На сегодняшний день, концепция прозрачности была ограничена объяснениями ненормального поведения, надежности системы и попытками определить аналитические основы интеллектуальной системы.

2.1 Принцип прозрачности EPSRC

Принципы робототехники EPSRC считают прозрачность одним из своих ключевых принципов, определив прозрачность в робототехнике следующим образом: «Роботы изготовленные артефакты. Они не должны создаваться в обманчивом способ эксплуатации уязвимых пользователей; вместо этого их машинная природа должна быть прозрачным». .

Определение прозрачности EPSRC подчеркивает сохранение конечный пользователь осведомлен о произведенных, механических и, следовательно, искусственных

[1] Университет Бата, Великобритания, электронная почта: a.theodorou@bath.ac.uk
[2] Университет Бата, Великобритания, электронная почта: rhwortham@bath.ac.uk
[3] Университет Бата, Великобритания, электронная почта: jjbryson@bath.ac.uk

характер робота. Однако использованная формулировка позволяет считать даже косвенная информация, такая как онлайновая техническая документация, в качестве достаточной методологии обеспечения прозрачности[4]. Это места бремя ответственности на конечном пользователе. Пользователю придется найти, прочитать и понять документацию или другую информацию, предоставленную производителем. Некоторые группы пользователей, например пожилые или пользователи-неспециалисты, могут возникнуть проблемы с пониманием технических термины, часто встречающиеся в технических руководствах.

2.2 Прозрачность как механизм отчетности надежность

В одной из ранних публикаций прозрачность определялась с точки зрения передачи информации конечному пользователю относительно склонности системы к ошибкам в заданном контексте [6]. Хотя интерпретация Дзиндолет является лишь частью нашего определения прозрачной системы, исследование представляет интересные выводы о важности прозрачности системы. Исследование показало, что предоставление дополнительной обратной связи пользователям относительно системных сбоев, это помогло участникам поверить в система. Пользователи знали, что система не на 100% надежна, но они смогли откалибровать свое доверие к автономной системе в эксперименте, поскольку они узнали, когда они могли положиться об этом, а когда нет. Военное использование роботизированных систем становится все более популярным, особенно в виде беспилотных летательных аппаратов. Важнейшее значение имеют летательные аппараты (БПЛА) и прозрачность боевых систем. Представьте, если агент идентифицирует гражданское здание как террористическое. и решает принять против него меры. Кто ответственный? Робот за ненадежность? Или надзиратель-человек, доверившийся в датчиках системы и механизме принятия решений? В то время EPSRC Принцип робототехники считает человека-оператора ответственным, нанесенный ущерб необратим. Роботы работают автономно для обнаружения и обезвреживания целей необходимо иметь прозрачное поведение [17]. Люди должны иметь возможность калибровать свое доверие к системе и в случаях боевых, медицинских или других сценариев, когда робот действует ненадежные могут причинить вред или убить людей, прозрачность как механизм сообщить, что надежность системы имеет основополагающее значение.

2.3 Прозрачность как механизм раскрытия неожиданное поведение

Более поздние исследования Kim Hinds [11] и Stumpf et. аль [14], концентрированный по предоставлению пользователям механизмов обратной связи в отношении неожиданных поведение интеллектуального агента. В своих исследованиях пользователь был насторожен только тогда, когда агент считал свое поведение ненормальным. Ким и исследование Хиндса, что интересно, показало, что за счет увеличения автономии важность прозрачности также возросла, поскольку ответственность перешли от пользователя к роботу. Их результаты согласуются с [10] исследования, которые в совокупности показывают, что люди с большей вероятностью винить в неудачах робота, а не другие искусственные артефакты и коллеги.

Возможность предупредить пользователя, когда робот ведет себя непредвиденным образом, необходима для достижения прозрачности. В ситуациях повышенного риска он может помочь спасти человеческие жизни или ценные ресурсы, предупреждая человек-надзиратель системы, чтобы взять под контроль или откалибровать ее доверие соответственно. Однако в реализации Кима и Хайндса робот предупреждал пользователя только тогда, когда обнаруживал, что ведет себя в неожиданный способ. На наш взгляд, эта реализация пытается исправить черный ящик с помощью другого. Нет никакой гарантии, что у робота что-то случится неожиданно, если он не узнает о своем нетипичном поведении. Прозрачность должна обеспечиваться в режиме реального времени как постоянно включенный механизм. анизм, позволяющий пользователю решить, является ли поведение агента считается ожидаемым или неожиданным.

2.4 Прозрачность как механизм раскрытия принятие решений

Мы считаем, что механизмы прозрачности должны быть встроены системе, предоставляя информацию в режиме реального времени о ее работе, т.к. а также предоставление дополнительной документации в соответствии с действующим принципом EP SRC. Интеллектуальный агент, т. е. робот, должен содержать необходимые механизмы для предоставления значимой информации. конечному пользователю. Чтобы считать робота прозрачным для осмотра, конечный пользователь должен иметь возможность запросить точную интерпретацию возможности робота, цели, прогресс по отношению к указанным целям, сенсорные входы - осознание ситуации, ее достоверность и неожиданность поведение, например сообщения об ошибках. Информация, предоставленная Робот должен быть представлен в понятном человеку формате.

Прозрачный агент с поддающимся проверке механизмом принятия решений также может быть отлажен таким же образом, как и в какое традиционное неинтеллектуальное программное обеспечение обычно отлаживается. Разработчик мог видеть, какие действия выполняет агент, почему и как он переходит от одного действия к другому. Это похоже на способ в которых популярные интегрированные среды разработки (IDE) предоставляют возможность отслеживать различные потоки кода с помощью точек отладки, и иметь такие способности, как «Вперед» и «Вступить» по блокам код.

3 ПРОЕКТИРОВАНИЕ ПРОЗРАЧНЫХ СИСТЕМ

В этом разделе данной статьи мы обсудим различные решения разработчики могут столкнуться при проектировании прозрачной системы. До настоящего времени, Известные исследования в области проектирования прозрачных систем сосредоточены на представлении прозрачности только в контексте сотрудничества человека и робота (HRC). Таким образом, они сосредоточились на разработке прозрачных систем, способных построить доверие между людьми-участниками. и робот.[12]. Мы считаем, что прозрачность должна присутствовать даже в несовместных средах, таких как соревнования между людьми и роботами [11], или даже когда роботы используются военными. В нашем точки зрения, разработчики должны стремиться к созданию интеллектуальных агентов, которые может эффективно передавать информацию конечному пользователю-человеку, и последовательно позволяют ей разработать мысленную модель системы и его поведение.

3.1 Удобство использования

Чтобы обеспечить прозрачность, необходимо тщательно спроектировать дополнительные дисплеи или другие способы связи с конечным пользователем. поскольку они будут интегрировать потенциально сложную информацию. Агент разработчикам необходимо учитывать как фактическую актуальность, так и уровень абстракция информации, которую они раскрывают, и то, как они будут представить эту информацию.

3.1.1 Актуальность информации

Разные пользователи могут по-разному реагировать на информацию, предоставленную робот. [16] демонстрирует, что конечные пользователи без технического образования не понимают и не сохраняют информацию, поступающую от технических устройств, таких как датчики. Это противоречит мнению разработчика агента, который нуждается в доступе к такой информации как во время разработки, так и во время тестирования робота, чтобы эффективно калибровать датчики и устранять любые обнаруженные проблемы. Однако в том же исследовании Туллио демонстрирует, что

пользователи могут понимать хотя бы основные концепции машинного обучения, независимо от их нетехнического образования и опыта работы.

Исследование Туллио создает хорошую отправную точку для понимания того, какая информация может иметь отношение к пользователю, чтобы помочь ему понять интеллектуальные системы. Тем не менее, необходима дальнейшая работа в другие прикладные области, чтобы установить тенденции как для предметной области, так и для конкретных пользователей в отношении того, какую информацию следует учитывать. важность.

3.1.2 Абстракция информации

Разработчикам прозрачных систем нужно будет поставить под сомнение не только которые, но и сколько информации они предоставят пользователю путем установления уровня сложности, с которым пользователи могут взаимодействовать с информацией, связанной с прозрачностью. Это особенно важно в системах с несколькими роботами.

Системы с несколькими роботами позволяют использовать несколько, обычно небольших роботы, где цель распределяется между различными роботами, каждый со своим собственный сенсорный ввод, надежность и прогресс в выполнении своих поставленная задача для завершения всей системы. Недавние разработки в области биологии, вдохновленные роевым интеллектом, позволяют использовать большие количество крошечных роботов, работающих вместе в такой мультироботной системе [15]. Военные уже рассматривают возможность разработки рои автономных маленьких роботов-солдат. Реализация прозрачности в такой системе — нетривиальная задача. Разработчик должен сделать рациональный выбор того, когда требуется раскрывать информацию низкого или высокого уровня. Предоставляя всю информацию в любое время, для всех типов пользователей, система может стать непригодной для использования, поскольку пользователь быть перегруженным информацией. Мы считаем, что разные пользователи будут требуют различных уровней абстракции информации, чтобы избежать информационного бытия. Более высокие уровни абстракции могут быть сосредоточены на представлении только обзор системы. Вместо того, чтобы иметь прогресс систему к цели, показывая текущие действия системы принимая во внимание достижение указанной цели, оно может просто представлять собой полоса завершения. Более того, в системе с несколькими роботами информация более низкого уровня может также включать цель, датчик, цель-процесс и общую информацию. поведение отдельных агентов в подробностях. И наоборот, обзор высокого уровня может отображать всех роботов как единое целое с указанием средних значений для каждой машины. Интеллектуальные агенты с конструкцией, основанной на когнитивная архитектура, такая как дизайн, ориентированный на поведение (BOD) [2], может представить только элементы плана высокого уровня, если обзор нужна система. В случае агента, разработанного с помощью BOD, пользователи может предпочесть видеть и получать информацию о состояниях дисков или Компетенции, но не отдельные Действия. Другие пользователи могут захотеть видеть только части плана в деталях и другие части на высоком уровне обзор.

Хорошая реализация прозрачности должна предоставлять пользователю с такими опциями, предоставляя отдельным лицам или потенциальным группам пользователей как с гибкими, так и с предустановленными конфигурациями, чтобы удовлетворить широкий диапазон потребностей потенциальных пользователей. Мы предполагаем, что уровень абстракции, в которой нуждается человек, зависит от ряда факторов, в том числе от демографического фона пользователя.

1. Пользователь: мы уже обсуждали, как разные пользователи склонны по-разному реагировать на информацию о текущем состоянии робота. Точно так же мы можем ожидать, что различные пользователи ответят аналогично различным уровням абстракции, основанным на их использование системы. Конечные пользователи, особенно неспециалисты, предпочтут общий обзор доступной информации, в то время как мы ожидаем, что разработчики будут ожидать доступа к более низкому уровню информации.
2. Тип роботизированной системы: как обсуждалось в наших примерах выше, система с несколькими роботами, скорее всего, потребует более высокого уровня абстракции, чтобы избежать информационного ожирения конечного пользователя. Система с одним агентом потребует гораздо меньше абстракции, так как пользователю отображается меньше данных.
3. Назначение роботизированной системы: Целевое назначение системы следует учитывать при разработке прозрачного агента. Например, военный робот с гораздо большей вероятностью будет использоваться с профессиональный пользователь в цикле или на нем, и из-за его операции с высоким риском существует гораздо большая потребность отображать и фиксировать как можно больше информацию о поведении агента, насколько это возможно. С другой рука, скорее всего, робот-регистратор или личный помощник для использования нетехническими пользователями, которые могут предпочесть упрощенный обзор поведения робота.

3.1.3 Представление информации

Разработчикам необходимо подумать о том, как представить пользователю любой из дополнительную информацию о поведении агента, которую они будут разоблачать. В предыдущих исследованиях использовалось визуальное или звуковое представление информация. Насколько нам известно, нет предшествующих исследований, сравнивающих разные подходы.

Автономные роботизированные системы могут принимать десятки разных решений в секунду. Если агент использует реактивный план, например план POSH [5], агент может совершать тысячи звонков в минуту на различные элементы плана. С таким объемом информации трудно справиться аудиоориентированные системы. Более того, визуализация информации, т.е. путем предоставления графического представления плана агента, где различные элементы плана мигают, когда они вызываются, должны заставить система не требует пояснений и проста для менее технических пользователей. Наконец, визуализация графа как средство предоставления информации, связанной с прозрачностью, имеет дополнительные преимущества при отладке приложения. Разработчик должен иметь возможность отслеживать различные вызываемые элементы плана, просматривая сенсорный ввод, который вызвал их, пока не были использованы определенные элементы.

3.2 Полезность системы

До сих пор в этой статье мы подробно остановились на важности и дизайне выбор в отношении реализации прозрачности. Однако мы считают, что разработчик также должен рассмотреть вопрос о том, следует ли внедрять прозрачность может на самом деле повредить полезности системы. [18] утверждает что полезность агента измеряется степенью его доверяют. Повышение прозрачности может снизить его полезность. Это может, например, иметь отрицательный эффект для робота-компаньона или медицинский робот, предназначенный для помощи детям. В таких случаях система разработана в соответствии с Принципами робототехники EPSRC, поскольку она использует чувства своих пользователей, чтобы повысить свою полезность и производительность на свою поставленную задачу.

Другим важным конструктивным решением, влияющим на систему, является физическая прозрачность системы. Внешний вид агент может повысить его удобство использования [7], но также может контрастировать с прозрачность, скрывая свою механическую природу. Вернемся к нашему примеру с роботом-компаньоном. Человекоподобный или звероподобный робот может быть предпочтительнее агента, механизмы и внутренние органы которого выставлены напоказ. раскрывая его искусственный характер [8].

Обсуждение компромиссов между полезностью и прозрачностью далеко выходит за рамки данной статьи. Тем не менее, разработчики должны знать этого, поскольку они проектируют и разрабатывают роботов.

4. ВЫВОД

Мы твердо верим, что внедрение и использование интеллектуальных Системы, которые прозрачны по своей природе, могут помочь общественности понять ИИ, рассеяв пугающую тайну вокруг того, почему он так себя ведет. Прозрачность позволит понять агентов эмерджентное поведение. В этой статье мы переопределили прозрачность как постоянно включенный механизм, способный сообщать о поведении системы, ее надежности, чувств и целей, поскольку такая информация может помочь нам понять поведение автономной системы.

Необходима дальнейшая работа для тестирования и внедрения передовой практики в отношении обеспечения прозрачности в сообществе робототехники. Учитывая преимущества прозрачных систем, мы решительно предложить продвигать этот ключевой принцип научными советами, такими как как EPSRC и другие академические сообщества.

БЛАГОДАРНОСТИ

Мы хотели бы поблагодарить Свен Гаудл (Университет Бата) за его ценные инсайты.

ИСПОЛЬЗОВАННАЯ ЛИТЕРАТУРА

[1] Маргарет Боден, Джоанна Брайсон, Дарвин Колдуэлл, Керстин Даутен Хан, Лилиан Эдвардс, Сара Кембер, Пол Ньюман, Вивьен Парри, Джефф Пегман, Том Родден, Том Сорелл, Мик Уоллис, Блэй Уитби, и Алан Уинфилд. Принципы робототехники. Исследовательский совет по инженерным и физическим наукам Соединенного Королевства (EPSRC), апрель 2011 г. веб-публикация.

[2] Джоанна Брайсон, «Ориентированный на поведение дизайн модульного интеллекта агента», в System, том 2592, 61–76, (2002).

[3] Джоанна Дж. Брайсон, «Роботы должны быть рабами», в Close Engagements с искусственными компаньонами: ключевые социальные, психологические, этические и дизайнерские проблемы, изд., Йорик Уилкс, 63–74, Джон Бенджаминс, Амстердам, (март 2010 г.).

[4] Джоанна Дж. Брайсон, Дарвин Колдуэлл, Керстин Даутенхан, Паула Даксбери, Лилиан Эдвардс, Хейзел Грайан, Сара Кембер, Стивен Кемп, Пол Ньюман, Гео Пег, Эндрю Роуз, Том Родден, Том Сорелл, Мик Уоллис, Ширер Уэст, Алан Уинфилд и Ян Болдуин, «Создание принципы робототехники epsrc, 133 (133), 14–15, (2012).

[5] Джоанна Дж. Брайсон, Тристан Дж. Колфилд и Ян Другович, «Интеграция выбора действия, похожего на жизнь, в среду моделирования агентов на основе циклов», в Proceedings of Agent 2005: Generative Social Processes, Модели и механизмы, ред., Майкл Норт, Дэвид Л. Саллах и Чарльз Макал, стр. 67–81, Чикаго (октябрь 2005 г.). Аргоннский национальный Лаборатория.

[6] Мэри Т. Дзиндолет, Скотт А. Петерсон, Регина А. Помранки, Линда Г. Пирс и Холл П. Бек, «Роль доверия в зависимости от автоматизации», Международный журнал компьютерных исследований человека, 58 (6), 697–718, (2003).

[7] Керстин Фишер, «Как люди разговаривают с роботами: разработка диалога для снизить неопределенность пользователей», AI Magazine, 32(4), 31–38, (2011).

[8] Дженнифер Гетц, Сара Кислер и Аарон Пауэрс, «Соответствие внешнего вида и поведения робота задачам для улучшения взаимодействия человека и робота», Труды - Международный семинар IEEE по интерактивному общению роботов и людей, 55–60, (2003 г.).

[9] Виктория Грум и Клиффорд Насс, «Могут ли роботы быть товарищами по команде?», «Интер». Исследования действий, 8 (3), 483–500, (2007).

[10] Питер Х. Кан, Рэйчел Л. Северсон, Такаюки Канда, Хироши Исигуро, Брайан Т. Гилл, Джолина Х. Рукерт, Солас Шен, Хизер Э. Гэри, Эйми Л. Райхерт и Натан Г. Фрейер, «Держат ли люди гуманоидное робот несет моральную ответственность за вред, который он причиняет?», Труды седьмая ежегодная международная конференция ACM/IEEE по теме «человек-робот». Взаимодействие - HRI '12, (февраль 2016 г.), 33, (2012 г.).

[11] Тэми Ким и Памела Хайндс: «Кого мне винить? влияние автономии и прозрачности на атрибуцию во взаимодействии человека и робота», Материалы - Международный семинар IEEE по интерактивному общению роботов и людей, 80–85, (2006 г.).

[12] Джозеф Б. Лайонс, «Прозрачность в отношении прозрачности: модель для взаимодействие человека и робота», Доверительные и автономные системы: документы из Весенний симпозиум AAAI 2013 г., 48–53, (2013 г.).

[13] Р. Парасураман и В. Райли, «Люди и автоматизация: использование, неправильное использование, неиспользование, злоупотребление», Human Factors, 39(2), 230–253, (1997).

[14] Симона Штумпф, Венг-Кин Вонг, Маргарет Бернетт и Тодд Кулеша, «Как сделать интеллектуальные системы понятными и управляемыми». конечными пользователями», 10–11, (2010).

[15] Ин Тан и Чжун-ян Чжэн, «Исследовательский прогресс в рое робототехника», «Оборонные технологии», 9(1), 18–39, (3 2013).

[16] Джо Туллио, Анинд К. Дей, Джейсон Чалеки и Джеймс Фогарти, «Как это работает: полевое исследование нетехнических пользователей, взаимодействующих с интеллектуальной системой », конференция SIGCHI по человеческому фактору в вычислениях. системы (CHI'07), 31–40, (2009).

[17] Лу Ван, Грег Джеймисон и Джастин Дж. Холландс, «Доверие и уверенность». по автоматизированной системе боевого опознавания», Человеческий фактор, 51(3), 281–291, (2009).

[18] Роберт Уортам, Андреас Теодору и Джоанна Дж. Брайсон, «Железный треугольник: прозрачность-доверие-полезность». подано, 2016.

Прозрачность роботов, доверие и полезность

Роберт Х. Уортам1 Андреас Теодору2и Джоанна Дж. Брайсон3

Абстрактный. По мере того как мышление роботов становится все более сложным, отладка становится все более сложной, основываясь исключительно на наблюдаемом поведении, даже для разработчиков роботов и технических специалистов. Точно так же пользователям-неспециалистам трудно создавать полезные ментальные модели рассуждений роботов исключительно на основе наблюдаемого поведения. Принципы робототехники EPSRC требуют, чтобы наши артефакты были прозрачными, но что это означает на практике и как прозрачность влияет как на доверие, так и на полезность? Мы исследовали эту взаимосвязь в литературе и обнаружили, что она сложна, особенно в непромышленных условиях, где прозрачность может иметь более широкий спектр эффектов на доверие и полезность в зависимости от применения и цели робота. Мы намечаем нашу программу исследований, чтобы поддержать наше утверждение о том, что тем не менее возможно создать прозрачных агентов, эмоционально привлекательных, несмотря на то, что они имеют прозрачную машинную природу.

1. ВВЕДЕНИЕ

В «Принципах робототехники» EPSRC содержится особое упоминание о прозрачности: «Роботы — это искусственные артефакты. Они не должны быть разработаны таким образом, чтобы обманывать уязвимых пользователей; вместо этого их машинная природа должна быть прозрачной». см. [1]. Первоначально это кажется простым нормативным утверждением, основанным на общепринятой идее о том, что агенты не должны вводить в заблуждение, поскольку обман обычно ведет к эксплуатации. В этой статье рассматривается, действительно ли прозрачность является такой простой идеей, а также не снижает ли прозрачность определенных типов агентов их полезность. При рассмотрении этого вопроса мы также должны обратить внимание на взаимосвязь между прозрачностью и доверием.

В этой статье мы используем термины «робот» и «агент» взаимозаменяемо, и под этими терминами мы подразумеваем воплощенный автономный интеллектуальный артефакт.

Что значит доверять роботу? Сначала мы могли бы просто утверждать, что чем более прозрачен ИИ, тем больше мы можем ему доверять, и, следовательно, его полезность возрастает. Мы также можем утверждать, что доверие требуется только тогда, когда агент не полностью прозрачен, и, следовательно, повышенная прозрачность снижает потребность в доверии [4]. Если полезность артефакта измеряется степенью доверия к нему, то повышение прозрачности может уменьшить эту полезность. Например, это может иметь место для робота, основная функция которого состоит в том, чтобы обеспечивать общение.

Итак, мы начинаем видеть, что существует сложная взаимосвязь между идеями полезности, прозрачности и доверия. Это соотношение будет зависеть от цели ИИ. В этой статье мы рассматриваем литературу, касающуюся прозрачности и доверия, а также описываем текущие практические исследования, направленные на изучение предположения о том, что действительно возможно построить эмоционально привлекающего, но прозрачного робота.

1 Университет Бата, Великобритания, электронная почта: rhwortham@bath.ac.uk
2 Университет Бата, Великобритания, электронная почта: a.theodorou@bath.ac.uk
3 Университет Бата, Великобритания, электронная почта: jjbryson@bath.ac.uk

2 ТЕОРИЯ РАЗУМА, ДОВЕРИЯ И ПРОЗРАЧНОСТЬ

Хотя мы можем предположить, что коммуникация между животными, и особенно между людьми, должна быть сложной, на самом деле естественные коммуникационные системы склонны использовать относительно простые и минимальные сигналы, смысл которых вытекает из экстенсивных моделей [16]. Другими словами, эволюция или общая филогенетическая история обеспечивает адекватные априорные данные, так что для передачи контекста требуется минимум данных. Хотя некоторые утверждают обратное [8], общепризнанно, что эффективное взаимодействие, будь то принуждение или сотрудничество, зависит от наличия у каждой стороны некоторой теории разума (ToM) другой [16, 14]. Таким образом, отдельные действия и сложное поведение интерпретируются в рамках уже существующей структуры ToM. Является ли этот ToM точным, неважно, при условии, что он позволяет прогнозировать поведение. Модель прозрачности робота не определяет ToM, используемую пользователем-человеком, но это модель прозрачности, которую мы можем напрямую настраивать, и поэтому в центре внимания этой статьи. Хорошо известно, что наблюдаемое поведение может сообщать о внутренних психических состояниях человека. Бризил [2] обнаружил, что неявная невербальная коммуникация улучшает прозрачность по сравнению с только преднамеренной невербальной коммуникацией. Здесь имплицитное определяется как передача информации, присущей поведению, но не сообщаемой намеренно разработчиком робота. У людей есть большие ожидания относительно того, как неявные и явные невербальные сигналы соотносятся с психическими состояниями. Бризил также обнаружил, что прозрачность уменьшает количество конфликтов при возникновении ошибок, особенно при попытке выполнить совместную задачу.

Уменьшение конфликтов означает, что при возникновении ошибки во время выполнения задачи восстановление возможно с меньшим распределением вины.
Бризил называет это уменьшенным конфликтом устойчивостью, и эта устойчивость является одним из эффективных показателей полезности.

2.1 Антропоморфизм и ментальные модели роботов

Люди имеют сильную предрасположенность к антропоморфизации не только природы, но и всего, что их окружает [5] — гипотеза социального мозга [7] может объяснить это явление, однако люди не относятся к роботам идентично людям, например, в отношении морального положения [10]. Хотя ведутся серьезные споры об онтологии разума робота по сравнению с человеческим разумом, более важное практическое значение имеет то, как разум робота понимается людьми психологически, т.е. что является воспринимаемой, а не фактической онтологией. Стаббс [15] считает необходимым сформировать мысленную модель роботов, чтобы

построить точки соприкосновения, которые мы могли бы также интерпретировать как основу человеческого доверия. Стаббс [15] также обнаружил, что эта общая основа может быть эффективно установлена посредством интерактивного диалога с роботом.
Хотя в этом исследовании в первую очередь рассматривались удаленные роботы, работающие в промышленных или исследовательских условиях, а не роботы, работающие в

домашней обстановке, мы должны принять к сведению важность диалога для установления доверия. Действительно, Мюллер [13] рассматривает диалог как единое целое. из трех основных характеристик прозрачных компьютеров, другие объяснение и обучение.

Meerbeek [12] исследует взаимосвязь между воспринимаемой личностью робота и уровнем, до которого пользователь чувствует контроль. во время взаимодействия. Чтобы быть правдоподобным, Мирбек обнаружил, что выражение личности должно быть связано с внутренней моделью, имеет дело с поведением (например, принятием решений) на основе личности и эмоции. Более экспрессивное, неформальное поведение связано с более высокое восприятие пользовательского контроля.

У людей-неспециалистов либо слишком мало ToM для роботов, либо модель, основанная на современной научной фантастике, и поэтому интерпретирует поведение, используя теорию другого агента по умолчанию, которая предполагает агент, чтобы разделить человеческие мотивы. Это можно понять с точки зрения эволюции из-за потребности наших предков быстро классифицировать проксимальную активность либо как нейтральную (шелест листьев в ветер), дружелюбный (приближение соплеменника) или враждебный (приближение приближение хищника или врага). Когда сенсорная информация неопределенна, развитие склонности к допущению как свободы действий, так и враждебности избирательно влияет на индивидуальное долголетие в среде, где человек часто является добычей, а не хищником. Даже в нашей технологической среде мы часто сталкиваемся с фиктивной агентурой, такой как роботы. звонки по продажам, автоматические публикации в Твиттере и автоматически сгенерированные персонализированные спам-рассылки.

В исследовании, проведенном в 2006 году в местной больнице в г. США медперсонал постоянно искал причины, по которым роботы действовали так, как они делали. Они спрашивали себя и других, "Что здесь происходит? Должен ли это делать робот или это сделал я? Что-то не так?". Это исследование утверждает, что низкий уровень прозрачности заставляет людей сомневаться даже в нормальном поведении робот, иногда даже заставляющий людей думать о правильном поведении как ошибки [11].

3 ИССЛЕДОВАТЕЛЬСКАЯ ПРОГРАММА

Мы начинаем программу практических исследований для изучения прозрачность, доверие, треугольник полезности. Первоначально с использованием негуманоидных На роботах мы проводим эксперименты по определению влияния различных проявлений прозрачности на эмоциональную реакцию людей. В основе наших экспериментов мы используем методы реактивного планирования для создания автономных агентов. Мы разработали Инстинктивно-реактивный планировщик, основанный на подходе Bryson Behavior Oriented Design (BOD) [3]. Планировщик Instinct сообщает о выполнении и статус каждого элемента плана в режиме реального времени, что позволяет нам неявно зафиксировать процесс рассуждений внутри робота, который приводит к его поведение. Наши эксперименты исследуют и демонстрируют, как эти данные о прозрачности из планировщика можно использовать, чтобы сделать поведение робота более понятным. Изначально мы прежде всего заинтересованы в том, чтобы сделать поведение прозрачным для конструктора роботов, поскольку роботов со сложными планами, как правило, очень сложно спроектировать и отлаживать. Однако эти первоначальные эксперименты могут также улучшить прозрачность для наблюдателей-неспециалистов.

Впоследствии мы исследуем, как мы можем использовать механизм прозрачности, встроенный в планировщик инстинктов, для создания более эффективный отечественный робот. В исследовании будет выяснено, прозрачность заставляет людей чувствовать себя более или менее связанными со своим роботом, и способны ли они более или менее точно оценить потребности робота, поскольку он работает для достижения своих целей.

Ожидается, что эти испытания должны проходить в домашней или близкой к домашней среде, например, в доме престарелых. Мы должны получить отзывы от наблюдателей/пользователей-неспециалистов о качественный уровень интеллекта робота, а также о том, как им было бы удобно иметь такое устройство дома. Исследование попытается оценить начальные уровни страха, тревога, недоверие к ИИ и роботам в целом, а также к домашним роботам в частности. Установив референтную позицию, прозрачность робота необходимо включить, предоставив пользователю обратную связь на основе на выполнение в реальном времени в реактивном планировщике. Методы в настоящее время мы предполагаем:

- Представление в режиме реального времени текстовых заявлений, касающихся выполнения плана разрез.
- Графическая визуализация выполнения плана в реальном времени.
- Аудио (то есть устные) заявления, касающиеся выполнения плана робота.

Для каждого из этих методов информация о прозрачности может либо быть представлены на/с удаленного устройства или на/с самого робота. Таким образом, есть шесть возможных комбинаций. Конечно, также может быть добавлено дополнительное слияние прозрачности, такое как звук в сочетании с графикой. тестируется на основе успеха или неудачи первоначальных экспериментальных результатов.

Как показывает литература, диалог важен для установления доверие, это исследование должно уделить некоторое внимание возможности принимать речевой ввод, хотя и ограниченный простыми командами, как средства для пользователей, чтобы спросить робота, что он делает, и иметь робот реагирует соответствующим образом.

4. ДИСКУССИЯ

Принцип 1 EPSRC утверждает, что роботы — это инструменты. В рамках промышленных и В инженерных средах это достаточно ясно, в том смысле, что человек использует робота для выполнения технической задачи. Дизайнер и Пользователь робота разделяет цель робота: выполнить задачу. Однако в домашних условиях и в сфере здравоохранения роботы могут имеют несколько иные отношения с теми, с кем они взаимодействуют. Они может быть предназначено для обеспечения дружеских отношений и одновременного тайного мониторинг самочувствия пациента. Они могут быть инструментами для медицинского работника, но для пациента они являются компаньонами. В таком среда, на которую коммунальное предприятие может негативно повлиять повышенным прозрачность. Наше чувство товарищества связано с мерой агентства мы проецируем на робота. Если мы сможем понять работы интеллекта, кажется ли, что он по своей сути становится менее умны в народном смысле, так что мы тогда проецируем меньше свободы действий, и в результате опыта меньше пользы от робота? Мы могли бы сравнить это с телевидением. Мы знаем, что у него нет агентства, но его присутствие в угол нашей гостиной действительно дает компаньону преимущества. Может быть, это связано с сознательной приостановкой неверия или может быть, у нас есть бессознательный детектор агентности, который легче обманутые техникой.

Представления здравого смысла об интеллекте смешиваются с представлениями народной психологии о деятельности, а также о жизни. Вещи, которые разумны живы в том смысле, что у них есть свои собственные убеждения, желания и намерения, которые, как мы понимаем, в основе своей корыстны или эгоистичны. Мы неявно признаем эгоизм фундаментальной характеристикой всю жизнь [6]. Если такой агент взаимодействует с нами, то он считает нас важны для достижения этих эгоистичных целей. Такие агенты достойны стать нашими спутниками, потому что они придают истинную ценность в их отношениях с нами, и это повышает нашу ценность в обществе. И наоборот, агенты, у которых нет корыстных агентов, не достойны нашего внимания, потому что они не несут никакой социальной ценности. Возможно, поэтому искусственные агенты, единственной целью которых является общение и по-настоящему прозрачные в этом отношении, таким образом, лишены права быть достойными компаньонами. Поэтому в некоторых ситуациях прозрачность робота может

противоречить полезности и, в более общем смысле, может быть скорее ортогональным чем выгодно для успешного использования робота. Пока мы можем придумывать сценарии и продолжать обсуждать теоретическое и философское взаимодействие между прозрачностью, доверием и полезностью, как ученые мы Ждем результатов наших экспериментов.

5. ВЫВОД

Мы видели, что распаковка прозрачности и доверия сложна, но можно частично понять, если посмотреть на то, как люди приходят к пониманию и впоследствии доверяют друг другу, и как они преодолевают эволюционные страхи, чтобы доверять другим агентам, через неявное невербальная коммуникация. Неприемлемый уровень тревоги, страха и недоверие приведет к эмоциональной и когнитивной реакции отказа роботы. Хэнкок [9] утверждает, что если мы не можем доверять нашим роботам, мы не в состоянии извлечь из них эффективную пользу. Однако, учитывая, что мы счастливо взаимодействовать в обществе с другими людьми, которых мы не полностью доверие, и мы все чаще взаимодействуем с компьютерами, зная, что их рекомендации могут быть ошибочными, мы должны сделать вывод, что Хэнкок сверх упрощения. Наконец, могут быть приложения, в которых прозрачность противоречит полезности. Наша постоянная программа исследований предназначен для проверки нашей гипотезы о том, что мы действительно можем создавать прозрачных роботов, которые, тем не менее, являются эмоционально привлекательными и полезными. инструменты в широком диапазоне бытовых и околобытовых условий. Между тем, предстоит проделать большую работу, чтобы раскрыть взаимосвязь между прозрачностью, полезностью и доверием.

ИСПОЛЬЗОВАННАЯ ЛИТЕРАТУРА

[1] Маргарет Боден, Джоанна Брайсон, Дарвин Колдуэлл, Керстин Даутен Хан, Лилиан Эдвардс, Сара Кембер, Пол Ньюман, Вивьен Парри, Джефф Пегман, Том Родден, Том Сорелл, Мик Уоллис, Блэй Уитби, и Алан Уинфилд. Принципы робототехники. Исследовательский совет по инженерным и физическим наукам Соединенного Королевства (EPSRC), апрель 2011 г. веб-публикация.

[2] К. Брезил, К. Д. Кидд, А. Л. Томаз, Г. Хоффман и М. Берлин, «Влияние невербальной коммуникации на эффективность и устойчивость в совместная работа человека и робота», Международная конференция IEEE/RSJ, 2005 г. по интеллектуальным роботам и системам, стр. 708–713, Альберта, Канада, (2005). Иии.

[3] Джоанна Дж. Брайсон, «Разум по замыслу: принципы модульности и координация для инженерных комплексных адаптивных агентов», (2001).

[4] Джоанна Дж. Брайсон и Пол Раувольф, «Доверие, общение и взаимопонимание». равенство'. 2016.

[5] Керстин Даутенхан, «Методология и темы взаимодействия человека и робота: растущая область исследований», International Journal of Advanced Робототехнические системы, 4(1 Спец.выпуск), 103–108 (2007).

[6] Ричард Докинз, «Иерархическая организация: принцип-кандидат этология », в « Точках роста в этологии» , ред., П. П. Г. Бейтсон и Р. А. Хинде, 7–54, издательство Кембриджского университета, Кембридж, (1976).

[7] Р.И.М. Данбар, «Гипотеза социального мозга», Evolutionary Anthropol. огия, 178–190, (1998).

[8] Шон Галлахер, «Нартивная альтернатива теории разума», в книге «Радикальный энактивизм: интенциональность, феноменология и нарратив», изд., Р. Менари, номер Галлахер 2001, 223–229, Джон Бенджаминс, Амстердам, (2006).

[9] П. а. Хэнкок, Д.Р. Биллингс, К.Е. Шефер, JYC Чен, Э.Дж. де Виссер и Р. Парасураман, «Метаанализ факторов, влияющих на Доверие к взаимодействию человека и робота», Человеческий фактор: Журнал Общество человеческого фактора и эргономики, 53(5), 517–527, (2011).

[10] Питер Х. Кан, Хироши Исигуро, Батья Фридман и Такаюки Канда, 'Что такое человек? - К психологическим ориентирам в области взаимодействие человека и робота», Proceedings - IEEE International Workshop об интерактивном общении роботов и людей, 3, 364–371, (2006).

[11] Тэми Ким и Памела Хайндс: «Кого мне винить? Влияние автономии и прозрачности на атрибуцию во взаимодействии человека и робота», Материалы - Международный семинар IEEE по интерактивному общению роботов и людей, 80–85, (2006 г.).

[12] Бернт Мирбик, Джетти Хунхаут, Питер Бингли и Жак Теркен, «Исследуя отношения между личностью робота-телевизионщика помощник и уровень пользовательского контроля», Proceedings - IEEE International Workshop on Robot and Human Interactive Communication, 404–410, (2006).

[13] Эрик Т. Мюллер, Прозрачные компьютеры: понятное проектирование Интеллектуальные системы, Эрик Т. Мюллер, Сан-Бернардино, Калифорния, 2016 г.

[14] Ребекка Сакс, Лаура Э. Шульц и Юхонг В. Цзян, «Чтение мыслей». против следующих правил: теория диссоциации разума и исполнительный контроль в мозгу». Социальная нейронаука, 1 (3–4), 284–98 (январь 2006 г.).

[15] Кристен Стаббс, Памела Дж. Хайндс и Дэвид Веттергрин, «Автономия и точки соприкосновения во взаимодействии человека и робота: полевое исследование», Интеллектуальные системы IEEE, 22(2), 42–50, (2007).

[16] Robert H Wortham и Joanna J Bryson, «Communication», in Handbook of Living Machines {в печати}, Oxford University Press, Oxford, (2016).