Опубликовано в «Обзоре философии».
и Psychology 11, страницы 881–897
(2020)», которые следует цитировать при
ссылке на эту работу.

Понимание искусственного интеллекта: можем и должны ли мы сопереживать с роботами?

Сюзанна Шметкамп1

Published online: 28 April 2020

Абстрактный

Расширяя дискуссию об эмпатии к людям, животным или вымышленным персонажам, включив в нее отношения человека и робота, эта статья предлагает две разные точки зрения, с помощью которых можно оценить масштабы и пределы эмпатии к роботам.
эпистемологический, а второй нормативный. Эпистемологический подход помогает нам выяснить, можем ли мы сопереживать искусственному интеллекту или, точнее, социальные роботы. Основная загадка здесь, среди прочего, касается того, что именно которому мы сочувствуем, если у роботов нет эмоций или убеждений, поскольку у них нет сознание в более широком смысле. Однако, сравнивая роботов с вымышленными персонажей, статья показывает, что мы все еще можем сопереживать роботам и что многие из существующие представления об эмпатии и чтении мыслей совместимы с такой точкой зрения. Таким образом делая, в статье основное внимание уделяется важности принятия точки зрения и утверждениям, которые мы делаем также приписывают роботам нечто вроде перспективного опыта. Нормативный подход исследует моральные последствия сопереживания роботам. В связи с этим документ
Критически обсуждаются три возможных ответа: стратегический, антиварварский и прагматический. Последняя позиция защищается, подчеркивая, что мы все более вынуждены взаимодействовать с роботами в общем мире и сделать роботов частью нашей морали. Внимательность следует рассматривать как неотъемлемую часть нашего понимания себя и других.

Ключевые слова Эмпатия. Искусственный интеллект. Гуманоидные роботы. Взаимодействие. Взгляд на перспективу. Выдуманные персонажи. Этика

1. Введение

Дебаты об эмпатии или, в более широком смысле, межличностном понимании, были опора научных исследований по широкому спектру дисциплин. Однако, хотя многое была написана о способности человека сопереживать реальным людям или вымышленным персонажам.

* Сюзанна Шметкамп
Сюзанна.schmetkamp@icloud.com

1 Кафедра философии, Фрибурский университет, Фрибур, Швейцария

(недавние обзоры см. в Coplan and Goldie 2011; Maibom 2017), до недавнего времени философы несколько игнорировали роль эмпатии во взаимодействиях человека и робота (HRI) (ср. Brinck and Balkenius 2018; Lin et al. 2017). Тем не менее, в связи с растущим числом исследований эмоций или других особенностей систем искусственного интеллекта1, существует большой философский интерес к возможности и необходимости взаимодействия и сопереживания с различными формами искусственного интеллекта, особенно с так называемыми социальными. robots.2 Этот интерес также породил дискуссии о ценности эмпатии для общества в целом или для здравоохранения и терапии в частности (Coeckelbergh 2018; Darling 2016; Engelen 2018; Loh 2019; Misselhorn в прессе; Vallor 2011). Становится ясно, что в будущем роботы и андроиды, то есть роботы, похожие на людей, станут более или менее независимыми субъектами с социальными навыками. Таким образом, они станут важными компаньонами и обретут все большую способность устанавливать отношения с людьми (Бенфорд и Малартр, 2007; Бризил , 2002; Дюмушель и Дамиано, 2017). Кроме того, системы глубокого обучения (Каспаров, 2017) будут использоваться во многих (пока) человеческих профессиях, что не только улучшит или облегчит некоторые задачи или задачи (например, в медицинских исследованиях); Они также могут заставить нас пересмотреть некоторые ключевые понятия, такие как интеллект, активность, сознание, автономия, эмоции или перспективы (Шнайдер в печати).

Как показали исследования (Лейте и др., 2013), форма и успех отношений человека и робота часто зависят от человеческих качеств, таких как способность роботов выражать эмоции, взаимодействовать и выполнять (более или менее) автономные решения. . . Эти способности также важны для взаимного эмпатического понимания.3 Хотя люди также распознают и приписывают эмоции по отношению к абстрактным виртуальным формам или даже с точки зрения технических устройств (лучшими примерами являются смартфоны и компьютеры), для нашего кооперативного и совместного взаимодействия с роботами – особенно в контексте медицины или здравоохранения – сильное человеческое сходство может иметь решающее значение для успеха этих взаимодействий. По мере роста присутствия гуманоидных социальных роботов в обществе растет и необходимость изучать и формировать наше взаимодействие с ними. Нынешнее поколение роботов уже способно выражать целый ряд эмоций — например, гуманоидный ИИ «София» знает 60 различных выражений лица и, кажется, даже способен общаться с чувством юмора и иронии. Однако у роботов нет сознания в смысле субъективного опыта4, и они не обладают юмором или эмоциями в сложном смысле (Boden 2016; MacLennan 2014; Scheutz 2011). Тем не менее, в них может быть что-то, что можно рассматривать как аналог человеческих эмоций и некоторых психических процессов. Более того, в свете недавних открытий философии воплощенного познания, возможно, что человекоподобное тело и поведение, а также «расширенное» познание — это то, что помогает людям распознавать андроидов как партнеров и похожих на них в некоторых отношениях, в то время как оставаясь соверше

[1] Группа из Медиа-лаборатории Массачусетского технологического института и ассоциации стандартов IEEE выступает за концепцию «расширенного» интеллекта вместо «искусственного». Посредством такого нового нарратива о «расширении» они хотят заменить гарантию того, что роботы не поддерживают людей, а скорее поддерживают их и сотрудничают с ними. Вместе они создали Совет по расширенной разведке CXI, см. https://globalcxi.org. (последнее обращение 12.12.2019).

[2] Проект, финансируемый ERC, расположенный в Университете Глазго и возглавляемый Эмили Кросс, в частности, исследует общение людей с искусственным интеллектом и важность взаимодействия и отношений с роботами для социального познания. Одно из направлений сосредоточено на способности роботов быть компаньонами, http://www.so-bots.com. (последнее обращение 20.12.2019).

[3] О феномене «зловещей долины» см. ниже.

[4] По крайней мере, когда мы придерживаемся антифизикалистской позиции.

их в других (Бенфорд и Малартр 2007, 181; Хоффманн и Пфайфер 2018; Ньюен и др. 2018).5
Эмпатия

широко рассматривается как важнейший способ постижения и повторного переживания психических состояний других посредством чтения мыслей, эмоционального обмена и// или эмпирический совместный опыт (см., например, Engelen and Rö ttger-Rö ssler 2012; Goldman 2006; Stueber 2018; Zahavi 2014).6 В философии эмпатию обычно отличают от аффективного заражения и моральной симпатии или сострадания.7 Тогда как последнее направлено на благо будучи принадлежащим другим и хочет способствовать этому (или, по крайней мере, не препятствовать) (Darwall 1998), эмпатия, в первую очередь, приводит к пониманию психических процессов других - таких как эмоции или убеждения. В отличие от простого эмоционального заражения, необходимо наличие дифференциации «я» и «другой» (Де Виньемон и Джейкоб, 2012). По этому вопросу продолжаются серьезные дебаты, и было предложено множество определений и подходов, направленных на решение таких вопросов, как: Как мы воспринимаем состояния и опыт других людей и получаем к ним доступ? Как охарактеризовать Каков результат этого процесса? В общих чертах, преобладающими теориями, исходящими из философии сознания или феноменологии, являются теория зеркальных нейронов или теория резонанса (MNT) (Gallese 2001), теория теории (TT) (Fodor 1987; Gopnik and Wellman 1994), теория моделирования (ST)) (Де Виньемон и Джейкоб 2012; Голдман 2006, 2011; Штубер 2006), Теория прямого восприятия (DPT) (Захави 2011) с ее вариациями теории взаимодействия (IT) (Галлахер 2008, 2017) и теории нарратива (NT) (Галлахер) и Хатто 2008). Кроме того, существуют гибридные и плюралистические теории, сочетающие в себе два или более подходов, такие как прямое восприятие и воображение8 (Schmetkamp 2017, 2019; Dullstein 2013; превосходные обзоры см. в Newen 2015; Stueber 2018; Zahavi 2014; 2018). Учитывая этот разнообразный набор подходов, некоторые проводят дополнительные различия между когнитивной эмпатией (например, TT) и аффективной эмпатией (например, ST или MNT).9 Задаваясь вопросом, можем ли мы вообще

сопереживать роботам, раздел 2 сосредоточится на многих эпистемических Размеры эмпатических взаимоотношений с роботами: что мы воспринимаем и понимаем, если на самом деле нет эмоций, субъективных переживаний или

---

5 В статье основное внимание уделяется роботам-гуманоидам. Одна из причин этого заключается в том, что это помогает ограничить объем статьи; Другая причина заключается в предположении, что человеческие черты действительно облегчают наше социальное взаимодействие с искусственным интеллектом и делают более правдоподобным то, что мы относимся к роботам как к социальным партнерам. Однако мы также можем сопереживать более абстрактным формам ИИ, приписывая им эмоциональные состояния и мотивы (см. Isik, Koldeewyn, Beeler and Kanwisher 2017). Я очень благодарен одному рецензенту за это замечание.

6 Очень спорно, предполагает ли эмпатия аффективное отражение, теоретическое чтение мыслей, симуляционное восприятие перспективы, эмоциональное понимание и/или эмпатию, и в настоящее время конца этой дискуссии не видно (см., например, Захави 2018). Многие философы подчеркивают, что чтение мыслей — это нечто отличное от эмпатии, и что эмпатия — это «нечто дополнительное». Здесь, однако, я попытался применить все различные подходы. Однако моя собственная позиция является феноменологической.

7 Однако одна из проблем всей этой дискуссии заключается в том, что не существует концептуального консенсуса в том, что такое эмпатия и что она подразумевает. Например, проект социальных роботов, финансируемый ERC, определяет эмпатию как сочетание эмоционального соответствия и просоциального поведения. Однако в философии эмпатия обычно не рассматривается как моральная эмоция или отношение (см. Cross et al. 2018; Zahavi 2018).

8 Например, ссылаясь на классические позиции Штейна или Дильтея и сочетая прямое восприятие с образным представлением («Vergegenwä rtigung») (см. также Gallagher 2019).

9 Канске (2018) проводит различие между собственно аффективной эмпатией и когнитивной теорией разума. В то время как первая способность позволит нам чувствовать то, что чувствуют другие, вторая поможет нам понять, что думают или во что верят другие. Хотя я признаю различия, я не буду здесь отличать ментализацию от сопереживания, а буду рассматривать различные формы понимания других сознаний под общим термином эмпатия, поскольку это центральный термин в текущих философских дебатах.

перспективы в широком смысле? Или у роботов есть что-то похожее на эмоции, убеждения и опыт? Есть ли у них индивидуальный взгляд на мир (Шметкамп 2017) или повествование (Галлахер 2012), поскольку они, по крайней мере, воплощены и контекстуализированы? Сравнивая роботов с вымышленными персонажами, ответ будет утвердительным: да, в определенной степени мы можем сопереживать роботам когнитивным, аффективным и даже экспериментальным способом, делая выводы, чувствуя, взаимодействуя или представляя, как они воспринимают и движутся. их мир, точно так же, как мы понимаем во множественном числе (Vaage 2010) , как вымышленный персонаж (например, в фильме) воспринимает свой мир, действует и чувствует. Решающим аспектом здесь будет то, что мы приписываем другому индивидуальную точку зрения. Мы понимаем это независимо от того, рассказана ли эта перспектива, спроецирована или запрограммирована.10 Второй вопрос,

который будет обсуждаться в разделе 3, спрашивает, должны ли мы сопереживать роботам. У этого вопроса есть две стороны: мы можем либо спросить, имеет ли сочувствие к роботам просто стратегическую функцию в отношении улучшения взаимного понимания во взаимоотношениях человека и робота, либо мы можем спросить, имеет ли сочувствие такое этическое воздействие, что мы обязаны сопереживать роботам (обзор темы этики и ИИ см. в Boddington et al. 2017). Если, например, мы сможем эпистемически понять, что роботы воспринимают, намереваются или даже могут «чувствовать», мы также сможем предсказать, что они будут делать дальше. В целом, это может быть полезно с точки зрения нашего взаимодействия с ними.11 Очевидно, что это относится к стратегическому или рациональному «долженствованию». Второе значение вопроса порождает нормативный ответ: обязаны ли мы сопереживать другим в моральном смысле? И что с моральной точки зрения они или мы – как сопереживающие – получаем от этого? Рассматривая этот вопрос, на первый взгляд может показаться очевидным кантианский ответ, который следует прецеденту, созданному взглядами Канта на животных, и который можно модифицировать для применения к искусственному интеллекту: а именно, как утверждается, мы должны сопереживать, чтобы избегать. «моральное варварство». В конце концов, статья не пойдет ни по стратегическому, ни по кантианскому пути, а вместо этого предложит прагматический и реляционный ответ. Этот ответ связан с двумя другими. Однако он подчеркивает влияние взаимодействия и понимания себя и других.

2 Можем ли мы сопереживать роботам?

Из соображений экономии места я сосредоточусь на роботах, которые имеют и лицо, и тело, демонстрируют человеческие выражения и поведение, предназначены для взаимодействия с людьми и, следовательно, воплощены и внедрены в нашу повседневную жизнь и, как таковые, подвержены социальным воздействиям. оценка со стороны человека. Второй причиной такого внимания является предположение о том, что роботы с человекоподобными чертами и выражениями лица, вероятно, даже более способны вызывать доверие и вызывать эмоциональные реакции, подобные тем, которые есть у реальных людей (Brinck and Balkenius 2018; Mori 2005), и в этом отношении более вероятны . быть признанными и принятыми в качестве партнеров в социальном взаимодействии. Хотя исследования в области когнитивной психологии показали, что мы также можем сопереживать или читать мысли с системами, которые мало похожи физически (Бретан и др., 2015), человек

[10] В статье основное внимание уделяется гносеологическому вопросу. Это не ответит на метафизический вопрос, Роботы или ИИ обладают сознанием.

[11] Что касается развертывания систем глубокого обучения в медицине, среди прочего необходимо доверять интеллектуальной машине и понимать, что она собирается делать, например, при взаимодействии медицинского робота с пациентом.

важно для использования роботов в качестве сиделок или коллег в сфере здравоохранения (Vallor 2011).12 Но о каком сочувствии здесь идет речь? Отражаем ли мы выражения лиц роботов? Интерпретируем и прогнозируем ли мы их поведение? Или мы сопереживаем более феноменологическим, интерактивным способом?

В самом минимальном смысле эмпатию можно определить как способность человека понимать психические состояния других и тем или иным способом заново их переживать, хотя остается предметом споров, должен ли эмпатический субъект чувствовать то же, что и другой. Некоторые теории ограничивают объекты эмпатии эмоциями человека и их выражением, указывающим на аффективные состояния. Другие шире и включают в качестве объектов эмпатии другие когнитивные процессы, такие как убеждения, желания и их соответствующие причины (обзоры см. в Batson 2009; Slote 2017). Знаменитое определение подразумевает условие изоморфизма: сопереживающий и объект находятся в одном и том же или, по крайней мере, сходном аффективном состоянии (De Vignemont and Singer 2006). Однако, как утверждают некоторые критики, эмпатия не обязательно подразумевает, что мы копируем психические состояния других (Захави и Майкл 2018). Нам также не нужно заботиться о другом в более тщательном смысле.

Как широко известно, нынешний «ажиотаж» вокруг темы эмпатии во многом связан с открытием так называемых «зеркальных нейронов» (Iacoboni et al. 1999; 2011) .
В широком смысле зеркальные нейроны — это те нейроны, расположенные в участке мозга, который разряжается как для наблюдения, так и для выполнения аналогичных действий. Этот процесс имитации был применен к пониманию человеческих эмоций: при наблюдении аффективного выражения другого человека — например, грустного лица — те же нейроны будут такими, как если бы мы — как наблюдатели — сделали грустное лицо и сами почувствовали грусть. Хотя эта теория подвергалась широкой критике (Hickok 2014) и отвергалась как теория эмпатии, другие использовали ее в своем более сложном подходе к эмпатии. В своем описании теории моделирования (ST) Элвин Голдман, например, различает низкоуровневую и высокоуровневую форму чтения мыслей или «зеркальный путь» и «реконструктивный путь», хотя эмоциональный «резонанс» реализуется. обоими путями (Goldman 2006, 2011). Зеркальные нейроны — это основная часть низкоуровневых процессов, посредством которых мы сразу и автоматически постигаем психические состояния другого человека. На более сложном, более высоком уровне мы моделируем состояние другого в своем уме, а затем приходим к знанию того, что чувствует другой, не путем распространения теории, а, скорее, имитируя поведение других в своем уме, а затем проецируя свое поведение. собственный мыслительный процесс на другого. Согласно ST, мы моделируем ситуацию от первого лица, находясь в ситуации другого человека, и используем наши собственные психические механизмы для генерации мыслей, убеждений, желаний и эмоций. В течение последних нескольких десятилетий ST – вместе со своим оппонентом Theory Theory (TT) – доминировала в дебатах о чтении мыслей. TT утверждает, что наше понимание других разумов по существу опирается на народную психологию, которая либо врождена, либо приобретена в раннем детстве (Baron-Cohen 1995).
TT предполагает, что мы делаем основанные на теории выводы, чтобы понять других.13 С точки зрения третьего лица, наблюдения, мы применяем (неявно или явно) подобные законам обобщения, которые подразумевают такие понятия психических состояний, как восприятие, убеждение, и желание. TT критиковали за то, что он был слишком теоретическим и слишком общим (Zahavi 2014; но ср. Fodor 1987). Его недоброжелатели утверждают, что TT не принимает во внимание другой бетон и не

[12] Однако эмпирически остается неясным, действительно ли роботы должны быть человекоподобными в HRI (Brinck and Balkenius 2018).

[13] Одна из проблем, конечно, заключается в том, как мы понимаем термин «понимание». Моника Дуллштейн (2012) показали, что в теориях разума используются понятия, совершенно отличные от феноменологических.

признавать воплощение и включенность других. Более того, и TT, и ST воспринимаются как ложное картезианское окклюзионистское представление о разуме, как будто мы не можем воспринимать то, что происходит в сознании другого человека (Zahavi 2011, 2014) . Напротив, феноменологические подходы подчеркивают воплощение и встроенность человеческих существ и утверждают, что мы можем видеть непосредственно в лице и телесных выражениях другого человека то, что он испытывает: с этой точки зрения нам не нужно делать выводы или представлять, что он чувствует. ; нам нужно только воспринять это. Более того, мы делаем это в общем ситуативном контексте и посредством взаимодействия. По этой причине такой подход называется теорией прямого восприятия (ТПВ). (Захави 2011, 2014) или Теория взаимодействия (ИТ) (Галлахер 2001; 2012). В отличие от ТТ, ДПТ и ИТ утверждают, что мы не занимаем позицию третьего лица по отношению к другим и наблюдаем за ними. Кроме того, DPT и IT также утверждают, что у нас нет творческого косвенного доступа к другим. Вместо этого мы социально взаимодействуем как вторая личность, когда два «вы» признают друг друга взаимодополняюще и взаимно (Dullstein 2012; Engelen 2018; Zahavi and Michael 2018). Ограничения ДПТ, очевидно, возникают в ситуациях, когда другой для нас отсутствует: например, когда кто-то рассказывает нам историю о ком-то другом, или если мы читаем роман, смотрим фильм или смотрим пьесу, в которой отражается опыт других людей. каким-то образом опосредованным кем-то другим (например, рассказчиком), у нас нет прямых встреч. Следовательно, все подобные случаи представляют собой случаи, когда другой дан посредством повествования, иногда даже в художественных рамках. Вот почему некоторые философы добавляют, что повествование необходимо для понимания других умов или для того, чтобы вызвать сочувствие в чем-то большем, чем самый про Дэниел Д. Хатто (2008) формулирует гипотезу нарративной практики (NPH). Согласно этому тезису, мы понимаем причины действий других людей, их убеждения и желания только тогда, когда мы также принимаем во внимание индивидуальные обстоятельства, историю субъекта, его текущую ситуацию, его надежды и переживания, черты его характера и так далее. Другими словами, согласно NPH, чтобы понять чью-либо ситуацию, мы должны полагаться на «историю» человека (Gallagher 2012). Эта точка зрения также позволяет сопереживать «монстрам или пришельцам с других планет, как их изображают в фильме» (Галлахер, 2012). Однако здесь необходимо своего рода воображение: во многих случаях – не только, но и особенно в нашем общении с художественной литературой – мы полагаемся на наше воображение как на способ сделать доступным то, чего у нас нет. Даже одна из первых пионеров феноменологического подхода к эмпатии, а именно Эдит Штайн (1989), утверждала, что воображение или «репрезентация» [14] Играет решающую роль в мульти являются стадией процесса эмпатического понимания. Вот почему некоторые теории эмпатии сочетают подход второй личности с формой творческого репрезентации ситуации, повествования и/или перспективы конкретного другого человека (Schmetkamp 2019; Gallagher and Gallagher 2019).15

Независимо от того, должны ли мы это делать. Рассматривайте все эти различные теории как теории эмпатии или, в более широком смысле, как теории межличностного понимания. Для каждой теории мы можем задать следующие вопросы с описательной и эпистемологической точки зрения: Как мы

[14] Трудно дать точный перевод концепции Штейна «Vergegenwärtigung». В английском переводе (Stein 1989) термин «репрезентация» или «репрезентативный акт» (Stein 1989: 8) используется как неизначальная репрезентация, представленная «данностью» других или косвенным опытом (аналогично памяти, ожиданию и воображению) (там же). .). В дискуссиях часто упускается из виду, что Штейн предлагает ступенчатую модель эмпатии, согласно которой первый уровень — это непосредственное восприятие опыта другого, а второй уровень — это своего рода размышление и принятие перспективы (Stein 1989: 10) .

[15] Галлахер недавно определил эмпатию следующим образом: «Сочувствие может [...] не только [считаться] как нечто происходящее, но и как метод; и это [...] предполагает рассмотрение точки зрения или ситуации другого» (2018). При этом Галлахер расширил свой повествовательный подход до перспективного подхода (сочетая повествование с субъективной перспективой).

сопереживать ИИ, например, человекоподобным роботам, если бы соответствующая версия была наиболее правдоподобной? Например, если мы наблюдаем выражения и/или действия робота, можно утверждать, что мы автоматически резонируем и имитируем его выразительное поведение. Если мы хотим предсказать, что робот будет делать дальше, мы могли бы также применить народную психологическую теорию и сделать вывод о причинах их действий. Мы могли бы смоделировать, что бы мы сделали, если бы оказались в их ситуации, а затем проецировать на них свой опыт. Или, при прямых встречах, мы могли бы интерактивно воспринимать их действия. Мы могли бы рассмотреть их включенность в повествовательный контекст и понять интенциональную структуру их эмоций, не воспроизводя в то же время их «качественное» содержание. Эмпирически говоря, такие интерактивные способы понимания, безусловно, имеют место.

Однако можно выдвинуть некоторые очевидные метафизические и эпистемологические возражения. Основная проблема заключается в том, что роботы на самом деле ничего не чувствуют и не испытывают. На самом деле у них нет психических состояний, таких как желания или убеждения, поскольку у них нет сознания. Тем не менее, кажется странным говорить об индивидуальной точке зрения или личном повествовании робота. Поскольку эмпатия направлена на психические состояния и чье-то «бытие в мире», то ответ будет таким: мы не можем сопереживать роботам.

Тем не менее, можно дать два возможных ответа: во-первых, «психические состояния» роботов часто описываются как «вычислительные состояния», которые, как считается, имеют структуру, аналогичную психическим состояниям человека. Итак, если мы предположим, что у роботов есть нечто, сравнимое с психическими состояниями человека, есть ли у них также что-то вроде эмоций или переживаний чувств, которым мы сопереживаем? Согласно некоторым современным философским теориям эмоций, эмоциональные состояния или процессы представляют собой сложную структуру, состоящую из когнитивных и аффективных компонентов (De Sousa 1987; Nussbaum 2001): когда мы чувствуем гнев, наш гнев направлен на объект, который мы оцениваем как раздражающий. . Это также называется теорией оценки, которая подразумевает, что мы выносим суждения об объектах в нашей среде с учетом их психологического соответствия нашим целям. Если бы эмоции состояли только из этого простого когнитивного компонента, мы могли бы предположить, что у роботов есть эмоции в минимальном смысле. Мы могли бы утверждать, что роботы действуют по ряду причин, основанных на наборе представлений о мире. Однако эмоции могут включать в себя нечто большее: например, гнев также ощущается на чувственном и физическом уровне; например, оно вызывает разочарование и сужение. Тем не менее, гнев также имеет негативные коннотации, которые мы осознаем проприоцептивно (Colombetti 2013). Однако тела роботов — если они не чисто виртуальные — состоят из металла или пластика, и, что более важно, они не связаны с богатой концепцией сознания: в том смысле, что оно ощущает себя как эмоциональное существо. Он не может самостоятельно почувствовать, что значит находиться в пластиковом теле. Более того, как утверждают нарративные описания эмоций, сложные эмоции обычно встроены в нарративную структуру: мы можем рассказать историю об их возбуждении и развитии (Goldie 2000). И последнее, но не менее важное: люди способны творчески обращаться со своими чувствами и эмоциями: они могут изучать новые эмоции, модифицировать одни и развивать другие.

Тем не менее, это также может быть возможно для роботов и с их помощью. Важным моментом здесь является то, что мы совершенно интуитивно приписываем машинам эмоции. Сотрудничая с роботами, мы можем занять «намеренную позицию». Эта концепция, берущая свое начало в работах Дэниела Деннета, подразумевает, что мы рассматриваем объект, поведение которого хотим предсказать, как рационального агента; мы приписываем убеждения и желания и на этой основе предсказываем его поведение (Деннетт, 1987). Но все же этот подход основан на теории чтения мыслей, а не на теории феноменального взаимодействия, которую имеют в виду феноменологи. Однако, если мы рассмотрим сознание, подразумевающее феноменальный опыт, кажется, что его трудно применить иначе, чем

Теория разума объясняет сочувствие к правозащитникам. Другими словами, проблема совместимости феноменологических теорий для HRI, по-видимому, заключается в феноменальном аспекте психических состояний, особенно в чувственной и эмпирической стороне эмоций. В то время как мы могли бы теоретизировать (TT) о когнитивных компонентах или моделировать ситуацию принятия решения (ST) робота, а затем делать выводы или проецировать наши выводы на ситуацию робота, было бы трудно говорить об эмпатическом понимании аффективных реакций робота. и сенсационные состояния непроективным способом. Если мы расширим проблему до понятия «опыт» — центрального термина феноменологического подхода (ДПТ), — ситуация станет еще более сложной. Как описано выше, согласно ДПТ и его вариациям, в нашем социальном взаимодействии с другими мы эмпатически воспринимаем их опыт и делаем это с взаимной точки зрения второго лица. «Опыт» — это сложный феноменологический термин, подразумевающий экзистенциальные аспекты и феноменальные качества. Мы субъективно и сознательно воспринимаем наш мир или то, что значит чувствовать или делать что-то, например, воспринимая красный стол как красный и каково это покраснение. ДПТ предполагает, что мы переживаем феноменальные переживания других напрямую и интерсубъективно, хотя и не копируя точный качественный характер опыта, а, скорее, обращая внимание на интенциональную структуру точки зрения другого (Галлахер 2012; Захави и Майкл 2018) . Для того, чтобы этот процесс функционировал, важно межтелесное и личное взаимодействие.16 Хотя последнее (по крайней мере, в основном) гарантировано, когда мы сотрудничаем и сотрудничаем с роботами, некоторые важные критерии этих интерсубъективных отношений отсутствуют. : как роботы не чувствуют эмоций, так и они не имеют субъективного опыта с их феноменальным содержанием и экзистенциальным воздействием , каково это для них.

Нам не нужно прибегать к теоретическим выводам, имитациям или прогнозам. Мы чувствуем, что у другого есть феноменальный опыт. Тем не менее, с феноменологической точки зрения кажется трудным сопереживать роботам. Тем не менее, сравнивая искусственный интеллект с вымышленными персонажами, я предложу потенциальное решение, а также продемонстрирую, что мы не только читаем мысли или отражаем поведение роботов, но и что возможно, по крайней мере в определенной степени, применить феноменологический подход. то есть интерактивно сопереживать перспективному «опыту» роботов. И этот аргумент выходит даже за рамки этой аналогии: когда мы взаимодействуем с роботами в общей среде, мы развиваем общую интенциональность и даже совместную историю, и это имеет решающее значение для наших отношений с роботами (Coeckelberg 2018) .

Однако, как и в случае с нашим эмпатическим пониманием вымышленных персонажей, здесь решающее значение имеет наша способность воображения.

Давайте проведем аналогию: принято считать, что эмпатия играет важную роль в наших отношениях с вымышленными повествованиями и вымышленными персонажами – будь то в романе, фильме или пьесе. С 1990-х годов в философии литературы и кино ведутся серьезные дебаты о том, следует ли включать «эмпатию» под общий термин «эмоциональное взаимодействие» с вымышленными персонажами в целом (например,

[16] Однако нарративистская версия феноменологических подходов предполагает образный компонент, который позволяет нам постичь интенциональную структуру с помощью нарративного воображения, например, если не задано интерсубъективное взаимодействие (Галлахер и Галлахер 2019) .

[17] Это вопрос, аналогичный вопросу так называемого «мысленного эксперимента с зомби», в котором обсуждается, можем ли мы предположить или приписать сознание зомби, которые похожи на нас во всех физических отношениях, но не имеют сознательного опыта в богатом мире. смысл (Чалмерс 1996; Деннетт 1991).

Плантинга, 2009 г.; Смит 1995). Другие формы взаимодействия включают эмоциональное заражение и эмоциональное разделение – особенно в отношении капризных эффектов художественной литературы – сочувствие или сострадание, негативные эмоции, такие как антипатия, и синестетическое воздействие (Plantinga 2009; Schmetkamp 2017) . Как отмечают многие киноведы, эмпатия играет решающую эпистемологическую роль, позволяя зрителю следить за повествованием и оставаться привязанным к персонажам (Smith 1995) . обсуждается, можем ли мы испытывать настоящие эмоции по отношению к вымышленным существам и являются ли эти эмоции рациональными (Янал 1999) – и предполагая, что мы действительно чувствуем и должны чувствовать сочувствие к вымышленным персонажам, нам все равно придется объяснить, как лучше концептуализировать сочувствие в случае художественной литературы. . Хотя я в целом убежден, что при просмотре фильма или чтении романа мы используем различные формы сопереживания, чтения мыслей и понимания – то есть полного спектра понимания психических состояний других людей, я предполагаю, что один аспект особенно важен. : Вымышленные персонажи выражают и представляют определенные индивидуальные точки зрения на свой (вымышленный) мир. Эти перспективы рассказаны в диегетическом мире фильма или романа; Более того, они часто дополнительно обрамлены неявным или явным рассказчиком. Они встроены в правдоподобное повествование. Или, говоря иначе: нарратив – это структурированное и оформленное представление событий с определенной точки зрения (Goldie 2012: 8), и в художественной литературе персонажи воплощают, выражают и представляют такие встроенные точки зрения.

Важность перспектив для художественной литературы и для нашего эмпатического взаимодействия с ней отчасти обусловлена тем фактом, что художественная литература обычно (хотя и не всегда) имеет разные технические перспективы: история обычно рассказывается от первого или третьего лица. перспектива. Но еще важнее то, что точка зрения – это мировоззрение. Тем не менее, перспектива означает, как человек встроен в мир, как он воспринимает мир, как он его переживает.
Эта перспектива формируется эмоциями, опытом, историями, воспоминаниями и, в свою очередь, формирует их; на него влияют черты характера, суждения и убеждения, и они сами влияют на него (Schmetkamp 2017). Когда, например, мы находимся в депрессивном настроении, мы видим наш мир с другой – а именно депрессивной или меланхолической – точки зрения, чем если бы мы находились в счастливом состоянии.

Теперь мы можем говорить о вымышленных персонажах как о «имеющих» (или, скорее, выражающих и представляющих) перспективу, поскольку они фокусируются и повествуются рассказчиком, который конструирует и направляет их мировоззрение. Как читатели или зрители, мы относимся к ним так, как будто у них есть точка зрения, и мы можем представить, каково это иметь такую точку зрения. Эмпатия к вымышленным персонажам предполагает своего рода взгляд на других, не сводя этот процесс к простой эгоцентрической симуляции или проекции.19 Более того, преимуществом вымышленных повествований является то, что они передают точку зрения других в сжатой форме. Художественная литература дает нам возможность погрузиться в точки зрения, которые могут быть похожими на наши собственные или совершенно отличными от них, и они часто делают это в интенсивной, сжатой и всеобъемлющей форме.

---

[18] Эмпатия как взгляд на перспективу действительно является способностью, которая позволяет зрителям понимать повествования и точки зрения персонажей. Однако как форма чуткого понимания того, почему персонаж чувствует, думает и действует именно так, это также результат. Таким образом, Коплан (2011) и Голди (2000) утверждали, что эмпатия — это одновременно процесс и результат .

[19] Миссельхорн выдвинул аналогичный аргумент, отметив, что «видя Т-образную форму неодушевленного объекта, мы воображаем, что воспринимаем человеческую Т-образную форму» (2009: 353).

Сравнивая роботов с вымышленными персонажами, можно выделить одну центральную схожую характеристику: у обоих на самом деле нет эмоций или сознательных убеждений, но они могут выражать и представлять их. И отчасти на этом основании мы, как реципиенты или сопереживающие, приписываем им человекоподобные психические состояния (Вебер 2013). Однако мы также воспринимаем их как воплощенные сущности, с которыми взаимодействуем. Как утверждает феноменолог кинофилософ Вивиан Собчак, фильм и его персонажи — это не просто проекты; у них есть тело и голос, и они допускают квазиинтерсубъективные переживания между собой и получателями (Собчак 2004). Они могут даже обеспечить тактильные впечатления. Эта воплощенная характеристика справедлива и для роботов, возможно даже в большей степени.

Тем не менее, есть некоторые принципиальные различия. Во-первых, в отличие от роботов, вымышленные персонажи лишены способности, жизненно важной для любого интерсубъективного описания эмпатии, а именно взаимного взаимодействия. В наших отношениях с вымышленными персонажами мы должны представлять себе, что персонажи обладают выраженными эмоциями, переживаниями и перспективами, но не взаимодействуем с ними взаимно. Кроме того, вымышленные персонажи не могут наложить вето на все, что мы им приписываем. Напротив, в наших встречах с роботами есть, по крайней мере, существующая и присутствующая воплощенная и встроенная взаимодействующая сущность, с которой мы можем развивать отношения. Робот способен чему-то возражать — скажем, если бы я был пациентом и не хотел принимать лекарство, роботу можно было бы поручить следить за тем, чтобы я это сделал. Во-вторых, можно возразить, что, в отличие от вымышленных персонажей, роботы (пока) не обладают экспериментальной перспективой или индивидуальным повествованием, как упоминалось выше. Художественная литература действительно предлагает богатую картину того, как человек может воспринимать и оценивать свой мир; и посредством этих повествовательных рамок и практик мы расширяем наш кругозор и изучаем новые эмоции или эмоциональные нюансы. Однако эмоции и переживания вымышленных персонажей также повествуются только в рамках определенного повествовательного фрейма; Их развитие зависит как от того, что драматургически задумал рассказчик, так и от того, как читатели или зрители воспринимают это на своем собственном интеллектуальном и эмпирическом фоне. Вымышленные эмоции и переживания обладают меньшей гибкостью и креативностью, чем их человеческие аналоги. Тем не менее, остается вопрос, можно ли по-прежнему противопоставлять вымышленных персонажей роботам. Вымышленные персонажи на самом деле ничего не испытывают; Точно так же у роботов нет опыта в широком смысле, включая квалиа. Однако роботы, по крайней мере, воспринимают окружающую среду, классифицируют, оценивают и взаимодействуют внутри нее. У них есть способ видеть и существовать в мире; они воплощены и контекстуализированы. Если мы вспомним знаменитый антиредукционистский пример Томаса Нагеля «Каково быть летучей мышью?» (Nagel 1974) мы никогда не сможем полностью понять эмпирическую точку зрения других существ; летучая мышь, по крайней мере, так гласит его аргумент, имеет совершенно другую систему восприятия, которую нельзя сравнивать с человеческим восприятием. Тем не менее, ученые постоянно открывают новые факты о нечеловеческих существах, таких как рыбы или растения (Coeckelberg 2018: 148), и один из аргументов здесь утверждает, что даже если мы, возможно, никогда не узнаем, каково это быть с ними, мы можем, по крайней мере, исп

Если мы попытаемся сравнить точку зрения робота с нашей, мы увидим некоторые сходства и, конечно же, множество различий. Но это не новый феномен в нашем социальном познании других разумов. Во-первых, робот буквально (например, визуально) воспринимает мир определенным образом (может по-человечески, а может и нет). Во-вторых, как искусственный интеллект, он также имеет перспективу в том смысле, что воспринимает и оценивает окружающий мир, как решает проблемы и т. д. Точка зрения робота далека от точки зрения в сложном смысле, подобной точке зрения человека, это эпистемическая и оценочная точка зрения: робот что-то знает и выносит суждения о мире. Мы также можем утверждать, что у него есть мотивационная перспектива, поскольку робот действует в соответствии с ним.

на основе своих убеждений.20 Еще более важно то, что роботы встроены в контекст, который мы воспринимаем или с которым взаимодействуем. Итак, мой ответ на вопрос, можем ли мы сопереживать роботам: да. Более того, все существующие описания в той или иной степени применимы к HRI. Конечно, следующий вопрос, который нам нужно задать: стоит ли нам это делать?

3. Стоит ли сопереживать роботам?

Учитывая предыдущий анализ, давайте предположим, что мы можем сопереживать гуманоидным роботам во множественном числе, то есть мы можем чувствовать, взаимодействовать с их «убеждениями», «эмоциями», «опытом» и «перспективами» или делать выводы на их основе. Но почему мы вообще должны сопереживать им? Например, в свете растущего использования роботов в медицине, здравоохранении и уходе за пожилыми людьми кажется гораздо более правдоподобным, чтобы роботы сопереживали пациентам, чем наоборот. Они должны каким-то образом проявить некоторую чуткость к потребностям пациентов, в то время как, в свою очередь, пациенты-люди могут нуждаться в сопереживающем компаньо Тем не менее, похоже, что расследование до сих пор было в первую очередь теоретической проверкой, чтобы выявить, какие из различных объяснений эмпатии совместимы с HRI. Но есть ли еще причина, по которой мы, люди, тоже должны сопереживать роботам? Этот вопрос актуален, поскольку взаимоотношения между людьми и роботами успешны и плодотворны только в том случае, если оба действительно взаимодействуют друг с другом, и эти взаимодействия могут предполагать — так или иначе — эмпатическое взаимодействие.

В пользу этого нормативного тезиса можно привести три аргумента:

1. Стратегический аргумент; 2.
Аргумент против варварства; 3.
Прагматический аргумент в пользу общего сообщества.

Первый, стратегический аргумент, не является прямо морально значимым нормативным аргументом. Он предполагает, что для успешного взаимодействия мы должны каким-то образом сделать вывод и понять, что задумал наш интерактивный партнер. Точнее, мы можем захотеть сопереживать, перенять точку зрения или прочитать чужие мысли, чтобы лучше достичь наших целей. Наше взаимодействие с роботами и наше сочувствие к ним в этом смысле полезны только для чего-то другого; это всего лишь инструмент. Понятие «должен» относится к гипотетическому императиву. В этом отношении роботы рассматриваются скорее как инструменты, чем как сотрудники. Фактически, они здесь не рассматриваются как моральные агенты или пациенты, имеющие моральный статус (Coeckelberg 2018).

Более существенным и морально нормативным является второй аргумент в пользу отказа от варваризации или культивирования. Не сопереживая другим, как утверждается, мы рискуем потерять чувствительность. В свою очередь, эмпатия может способствовать просоциальному поведению и улучшению нашего морального облика. Прежде чем исследовать основные проблемы этого тезиса, я объясню два его корня, а именно кантианский и аристотелевский аргументы. Кантианский аргумент первоначально был выдвинут в отношении отношений человека и животного. Это подразумевает

[20] Опять же, аналогичные аргументы могут быть выдвинуты и для других форм ИИ, не являющихся людьми, например, абстрактных виртуальных форм. Основное внимание в этой статье уделяется роботам-гуманоидам, с которыми люди взаимодействуют и сотрудничают. Чтобы добиться успеха, люди могли бы приписать ИИ не только базовые психические состояния, но также перспективу и повествование. Это может быть важно для коллективной интенциональности и коллективного внимания.

что мы не должны быть жестокими по отношению к животным, потому что это может повредить или испортить наш моральный облик в целом. Согласно этому аргументу, животные являются лишь косвенными моральными пациентами, не имеющими собственного морального статуса, поскольку Кант связывает моральный статус человека со способностью действовать автономно, исходя из причин, и приписывает эту компетентность только людям. Тот же аргумент применим и к социальным роботам, которые сами по себе не являются моральными адресатами: это потому, что они могут не обладать автономией в сложном смысле. Однако, не сопереживая им, мы проявим неуважение к важнейшему состоянию человечества.21 Специалист по правам человека Кейт Дарлинг является современной сторонницей этой точки зрения: «Кантианский философский аргумент в пользу предотвращения жестокости по отношению к животным состоит в том, что наши действия по отношению к нелюдям отражают наша мораль - если мы обращаемся с животными бесчеловечно, мы становимся бесчеловечными людьми. Это логически распространяется и на обращение с роботами-компаньонами. [...] Это также может предотвратить снижение чувствительности к реальным живым существам и защитить наше сочувст

Аргумент Аристотеля движется в том же направлении. Речь идет о том, что мы можем культивировать свои эмоции, рассматривая перспективу, тем самым определяя принятие перспективы как дистанцирование от собственной точки зрения от первого лица, или посредством эмоционального обмена и, таким образом, знакомясь с новыми эмоциями (Nussbaum 2011; Rorty 2001) . В то время как кантианская точка зрения подчеркивает проблему варваризации, аристотелевская точка зрения подчеркивает этическое воздействие культивирования чего-либо путем сопереживания: наших эмоций, морального восприятия, воображения и силы суждения.

Как я сказал, здесь возникают некоторые проблемы, которые беспокоят, в частности, кантовскую точку зрения: первая из них состоит в том, что признание лишь косвенного статуса неличностей или существ без «рациональности» неудовлетворительно: оно контринтуитивно, антропоцентрично и исключает гораздо больше сущностей, чем нелюдей (Gruen 2017). Но касается ли это и неодушевленных существ? Таким образом, остается вопрос: чему мы причиняем вред, применяя насилие против роботов, которые могут ничего не чувствовать сложным и субъективным образом?
Есть ли у них понятие уважения и достоинства? Есть ли у них моральные претензии? Эти сложные вопросы останутся здесь без ответа, поскольку потребуют отдельного специального исследования.
Другое возражение против кантовской точки зрения состоит в том, что этот аргумент основан на конкретном понимании эмпатии как просоциального поведения. Это не только подразумевает понимание других умов, но также включает в себя заботу о благополучии другого существа. То есть сопереживающий не просто интересуется переживаниями другого и «чувствует» их; они также мотивированы облегчить страдания другого или способствовать его благополучию. И если бы мы были жестоки по отношению к ним и неуважительно относились к благополучию роботов – например, избивая или насилуя их (если мы думаем о секс-роботах), – это отразилось бы и на нашем поведении по отношению к людям. Однако, как отмечалось ранее, этическое воздействие заботы или заботы – это, скорее, влияние сочувствия или сострадания как особой моральной эмоции, и как таковое оно отличается от сочувствия (Darwall 1998) . Как, в частности, показали феноменологи, эмпатия не обязательно является позитивным отношением к другим, но также может привести к антисоциальному поведению.

Садистический человек должен быть эмпатичен и в этом смысле, то есть он понимает страдания другого, но не хочет их облегчать (Breithaupt 2019; Захави и Михаэль

21 Кант пишет: «Если человек застрелит свою собаку потому, что животное больше не способно служить, он не нарушит своего долга перед собакой, ибо собака не может судить, но его поступок бесчеловечен и наносит ущерб самому себе той человечности, которую он его долг - проявить себя по отношению к человечеству. Если он не хочет подавлять свои человеческие чувства, он должен проявлять доброту по отношению к животным, ибо тот, кто жесток к животным, становится жестким и в обращении с людьми» (Кант 1997: 212) .

2018) 22 Другими словами, кантианский подход объединяет некоторые важные концептуальные различия, а именно между эмпатией и состраданием. Здесь можно было бы выдвинуть еще одно возражение: эмпирически совершенно неясно, почему тот, кто не сопереживает другим, обязательно становится варваром (Brinck and Balkenius 2018).

Однако с более оптимистической точки зрения некоторые утверждают, что частое эмпатическое понимание или принятие точки зрения могут помочь нам узнать, что другие могут чувствовать или думать.
Чем больше мы используем эмпатию, тем больше мы можем вовлекаться в общение с другими людьми как в наших повседневных взаимодействиях, так и в более необычных встречах. Более того, это может сделать нас более терпимыми и более добродетельными людьми. Опять же, это утверждается в отношении вымышленных персонажей и повествований. Внимание к точкам зрения и опыту других, как известно, заявил Ричард Рорти, имеет этическую ценность, поскольку при этом мы отказываемся от нашей эгоцентрической точки зрения (Рорти 2001). Но, конечно, мы могли бы принять этот аргумент в пользу HRI: сопереживание роботам улучшит наше сотрудничество и взаимодействие, поскольку мы лучше познакомимся с ними. Это приводит к третьему аргументу, который имеет некоторые общие черты как со стратегическим, так и с кантианско-аристотелевским подходом, но подчеркивает взаимодействие, отношения и социальное самопонимание сопереживающих.

Этот аргумент (который описывает мою собственную позицию) развивает подход Рорти, но модифицирует его до еще более прагматичного и реляционного тезиса о социальном познании и его предпосылках. В отличие от кантианского и аристотелевского подхода, эта точка зрения исходит из антиантропоцентрической точки зрения и подчеркивает интерактивные отношения между людьми и роботами. Эта позиция предполагает, что сопереживание другим – во всех его вариациях, но особенно в феноменологической интерактивной традиции – может позволить нам познакомиться с «бытием других в мире» и тем самым расширить наш кругозор, изменить наши взгляды и сформировать наши социальные взаимодействия. .и моральное поведение по отношению к другим людям, не являющимся людьми.

Учитывая перспективы, я предполагаю, что мы можем даже говорить о (будущих) роботах и системах глубокого обучения23 как о людях, имеющих особый взгляд на мир.
Эта точка зрения будет в чем-то похожа на человеческую точку зрения, а в чем-то и отлична. В научно-фантастических фильмах, таких как «HER» (США, 2013 г.), показано, каким может стать независимый ИИ: сверхразумные системы, значительно превосходящие возможности человеческого мышления. Сопереживание роботам-гуманоидам, с которыми мы все чаще взаимодействуем – например, в контексте здравоохранения – может помочь нам подготовиться к будущему развитию. На данный момент, однако, скорее дело в том, что, поскольку мы уже разделяем действия и окружающую среду с роботами, и поскольку эмпатия и социальное познание могут улучшить наше взаимодействие с другими, мы также можем предположить, что наше взаимодействие с роботами выиграет от чуткая точка зрения – хотя и не просто в инструментальном, стратегическом смысле. Это также может иметь обучающий эффект – аргумент, который, как уже отмечалось, также был выдвинут в отношении вымышленных миров. Но более важным моментом является то, что такая точка зрения затрагивает вопрос о том, как мы хотим понимать себя: серьезное отношение к роботам как к социальным товарищам должно быть реализовано как часть нашего самопонимания как людей, так и членов демократических обществ.

То, как мы взаимодействуем с роботами, во многом зависит от того, как мы о них думаем: как и об инструментах.

---

22 Феномен, заключающийся в том, что сопереживающие могут становиться тем более жестокими, чем более человекоподобными роботы называются «зловещая долина» (см. Misselhorn 2009; Mori 2005).

23 Или, как их называет Сьюзен Шнайдер: «умы будущего» (в прессе).

которые должны взаимодействовать с чисто инструментальной точки зрения или как партнеры, к которым мы должны относиться серьезно ради них самих. Таким образом, именно отношения и общее сообщество выходят здесь на первый план. Такая позиция подчеркивает прагматическое и феноменологическое влияние взаимодействий. Это также может иметь последствия для статуса роботов как моральных агентов и моральных пациентов, как утверждает Марк Кокельберг: «Вопрос морального статуса всегда связан с вопросом о том, кто является частью морального сообщества и в какие моральные игры уже играют». (Кокельберг 2018: 149).
Вместо реализации морали сверху вниз Кекельберг выступает за перспективу снизу вверх. Рассматривая роботов как компаньонов в контексте отношений и сопереживая их перспективному повествованию, мы развиваем с ними отношения, которые, в свою очередь, влияют на то, как мы видим их морально (там же).24 Однако обсуждение морального статуса выходило бы за рамки рамки данной статьи. Как упоминалось выше, сочувствие само по себе не является моральной эмоцией или отношением заботы. Но это могло бы посеять соответствующие семена в этом отношении, поскольку обеспечивает эпистемологическую основу для интерсубъектной морали.
Более того, это во многом связано с нашим социальным и моральным самопониманием: «[То, как] мы обращаемся с другими сущностями, как мы воспринимаем их, что мы говорим о них, как мы к ним относимся и так далее. также много говорит обо мне и много говорит о нас». (Кокельберг 2018: 150). Но вместо антропоцентрического взгляда это скорее реляционный взгляд, который рассматривает нечеловеческие сущности как партнеров во взаимодействии.

4. Вывод

Искусственный интеллект в целом и роботы-гуманоиды в частности изменят нашу жизнь и, возможно, нас самих. Философам предстоит многое рассмотреть с точки зрения эпистемических, этических, эстетических и политических последствий этих новых вызовов. Эмпатия — лишь одна из многих тем, которые бросает вызов HRI. Эта статья способствовала необходимым расследованиям, которые уже ведутся или еще предстоит провести. Я обсуждал эпистемологическую загадку того, можем ли мы сопереживать роботам, применяя к этой области доминирующие современные концепции эмпатии. Затем я рассмотрел нормативный вопрос о том, следует ли и почему нам сочувствовать роботам. В статье предложена прагматичная точка зрения, демонстрирующая, что а) действительно мы можем сопереживать гуманоидным роботам не только на базовом уровне, но также, по крайней мере в определенной степени, на уровне воображения; Более того, было показано, что даже с феноменологической и интерсубъективной точки зрения можно говорить о сопереживании роботам, находящимся в нашем мире, с которыми мы взаимодействуем и разделяем контекстуальный нарратив. Основное внимание уделялось эмпатии как процессу взаимного взаимодействия, а не результату. Однако в статье также утверждается, что б) мы должны сопереживать роботам-гуманоидам, потому что, поступая так, мы можем получить новые знания об очень незнакомом существе в мире, тем самым расширяя наш кругозор, готовясь к будущим разработкам ИИ и улучшая HRI в общая социальная среда. Считалось, что это имеет не только инструментальную ценность, но и ценно для нашего понимания самих себя и нашего общества, в котором роботы и другие формы ИИ могут рассматриваться как компаньоны.

[24] Кокельберг предлагает аналогичный моему подход, но черпает вдохновение из концепций Витгенштейна о форме жизни и языковых играх. Тем не менее, в его статье отсутствует четкое определение того, что, по его мнению, подразумевает эмпатия (например, действительно ли эмпатия включает в себя заботу о благополучии другого человека, как это, по-видимому, предполагает его статья).

Рекомендации

Барон-Коэн, С. 1995. Слепота. Эссе об аутизме и теории разума. Кембридж, Массачусетс: MIT Press.

Бэтсон, CD 2009. Эти вещи называются эмпатией: восемь связанных, но различных явлений. В книге «Социальная нейробиология эмпатии» под ред. Дж. Десети и В. Икес, 3–15. Кембридж, Массачусетс: MIT Press.

Бенфорд, Г. и Э. Малартр. 2007. За пределами человека. Tom Doherty Associates: Жизнь с роботами и киборгами. Нью-Йорк.

Боддингтон П., П. Милликан и М. Вулдридж. 2017. Спецвыпуск «Разумы и машины»: Этика и искусственный интеллект. Умы и машины 27 (4): 569–574.

Боден, Массачусетс, 2016. AI. Его природа и будущее. Оксфорд: Издательство Оксфордского университета.

Бризил, CL 2002. Проектирование социальных роботов. Кембридж, Массачусетс: MIT Press.

Брайтаупт, Ф. 2019. Темные стороны эмпатии. Итака: Издательство Корнельского университета.

Бретан М., Г. Хоффман и Г. Вайнберг. 2015. Эмоционально выразительное, динамичное физическое поведение роботов. Международный журнал человеко-компьютерных исследований 78: 1–16.

Бринк И. и К. Балкениус. 2018. Взаимное признание во взаимодействии человека и робота: дефляционный счет. Философия и технология: 1–18. https://doi.org/10.1007/s13347-018-0339-x.

Чалмерс, DJ 1996. Сознательный разум. Оксфорд: Издательство Оксфордского университета.

Кокельберг, М. 2018. Зачем заботиться о роботах? Сочувствие, моральное положение и язык страдания. Кайрос. Журнал философии и науки 20: 141–158.

Коломбетти, Дж. 2013. Чувствующее тело. Аффективная наука встречается с активным разумом. Кембридж, Массачусетс: MIT Press.

Коплан, А. 2011. Понимание эмпатии, 3–18. Его особенности и последствия. В Эмпатии. Философский и психологические перспективы. Оксфорд: Издательство Оксфордского университета.

Коплан А. и П. Голди. 2011. Эмпатия. Философские и психологические аспекты. Оксфорд: Оксфорд Университетское издательство.

Кросс, Э.С., Риддок, К.А., Праттс, Дж., Титоне, С., Чаудхури, Б., и Гортензиус, Р. 2018. Нейрокогнитивное исследование влияния общения с роботом на сочувствие боли. Препринт. https://дой. орг/10.1101/470534.

Дарлинг, К. 2016. Распространение правовой защиты на социальных роботов: последствия антропоморфизма, сочувствия и агрессивного поведения по отношению к роботизированным объектам. В законе о роботах, изд. М. Фрумкин, Р. Кало и И. Керр. Челтнем: Эдвард Элгар.

Дарвалл, С. 1998. Сочувствие, сочувствие, забота. Философские исследования 89: 261–282.

Де Соуза, Р. 1987. Рациональность эмоций. Кембридж, Массачусетс: MIT Press.

Де Виньемон Ф. и П. Джейкоб. 2012. Каково чувствовать чужую боль? Философия науки 79 (2): 295–316.

Де Виньемон Ф. и Т. Сингер. 2006. Эмпатический мозг: как, когда и почему? Тенденции в когнитивной сфере науки 10 (10): 435–441.

Деннетт, Д. 1991. Объяснение сознания. Бостон: Литтл, Браун и Ко.

Дулштейн, М. 2012. Второй человек в дебатах по теории разума. Обзор философии и психологии 3 (2): 231–248.

Дулштейн, М. 2013. Прямое восприятие и моделирование: описание эмпатии Штейном. Обзор философии и психологии 4: 333–350.

Дюмушель П. и Л. Дамиано. 2017. Жизнь с роботами. Кембридж, Массачусетс: Издательство Гарвардского университета.

Энгелен, Э.М. 2018. Можем ли мы поделиться своими чувствами с цифровой машиной? Эмоциональное разделение и признание одного за другого. Междисциплинарные научные обзоры 43 (2): 125–135.

Энгелен, Э.М. и Б. Рёттгер-Рёсслер. 2012. Текущие дисциплинарные и междисциплинарные дебаты по эмпатии. Обзор эмоций 4 (1): 3–8.

Фодор, Дж. 1987. Психосемантика. Проблема смысла в философии сознания. Кембридж, Массачусетс: Массачусетский технологический институт Нажимать.

Галлахер, С. 2008. Прямое восприятие в интерактивном контексте. Сознание и познание 17 (2): 535–543.

Галлахер, С. 2017. Эмпатия и теории прямого восприятия. В справочнике по философии The Routledge сочувствие, изд. Х. Майбом, 158–168. Нью-Йорк: Рутледж.

Галлахер С. и Дж. Галлахер. 2019. Играть в роли другого: Сопереживание актера своему персонажу. Топои (сначала онлайн), https://doi.org/https://doi.org/10.1007/s11245-018-96247.

Галлахер С. и Д. Хатто. 2008. Понимание других посредством первичного взаимодействия и повествовательной практики. В книге «Общий разум: взгляды на интерсубъективность», под ред. Дж. Златев, Т. Расин, К. Синха и Э. Итконен, 17–38. Амстердам/ Филадельфия: Издательство John Benjamins Publishing Company.

Галлезе, В. 2001. Гипотеза «общего многообразия»: от зеркальных нейронов к эмпатии. Журнал Исследования сознания 8: 33–50.

Голди, П. 2000. Эмоции. Оксфорд: Издательство Оксфордского университета.

Голди, П. 2012. Беспорядок внутри. Повествование, эмоции и разум. Оксфорд: Издательство Оксфордского университета.

Гольдман, А. 2006. Моделирование разума: философия, психология и нейробиология чтения мыслей. Оксфорд: Издательство Оксфордского университета.

Гольдман, А. 2011. Два пути к эмпатии: выводы когнитивной нейробиологии. В книге «Эмпатия: философские и психологические перспективы», под ред. А. Коплан и П. Голди, 31–44. Оксфорд: Издательство Оксфордского университета.

Гопник А. и Х.М. Веллман. 1994. Теоретическая теория. В книге «Картирование разума: специфичность области познания и культуры» под ред. Л. А. Хиршфельд и С. А. Гельман, 257–293. Кембридж: Издательство Кембриджского университета.

Грюн, Л. 2009. Забота о природе: чуткое взаимодействие с более чем человеческим миром. Этика и Окружающая среда 14 (2): 23–38.

Грюн, Л. 2017. Моральный статус животных. В Стэнфордской энциклопедии философии (выпуск осенью 2017 г.) ред. EN Залта, https://plato.stanford.edu/archives/fall2017/entries/moral-animal/.

Хикок, Г. 2014. Миф о зеркальных нейронах: настоящая нейробиология общения и познания. Новый Йорк: WW Norton & Company.

Хоффманн М. и Р. Пфайфер. 2018. Роботы как мощные союзники в изучении воплощенного познания снизу вверх. В Оксфордском справочнике по познанию 4E, под ред. А. Ньюэн, Л. де Брюин и С. Галлахер. Оксфорд: Издательство Оксфордского университета.

Хатто, Д. Д. 2008. Гипотеза повествовательной практики: разъяснения и последствия. Философские исследования 11 (3): 175–192.

Якобони, М. 2011. Друг в друге: нейронные механизмы эмпатии в мозгу приматов. В книге «Эмпатия: философские и психологические перспективы», под ред. А. Коплан и П. Голди, 45–57. Оксфорд: Издательство Оксфордского университета.

Якобони М., Р.П. Вудс и др. 1999. Корковые механизмы имитации человека. Наука 286: 2526–2528.

Канске, П. 2018. Социальный разум: распутывание аффективных и когнитивных методов понимания других. Междисциплинарные научные обзоры 43 (2): 115–124.

Кант И. 1997. Лекции по этике, под ред. и транс. П. Хит и Дж. Б. Шнивинд. Кембридж: Кембридж Университетская пресса.

Каспаров, Г. 2017. Глубокое мышление: где заканчивается машинный интеллект и начинается человеческое творчество. Нью-Йорк: Общественные дела.

Лейте А., А. Перейра, С. Маскареньяш, К. Мартиньо, Р. Прада и А. Пайва. 2013. Влияние эмпатии на отношения человека и робота. Международный журнал человеко-компьютерных исследований 71 (3): 250–260.

Лин П., Р. Дженкинс и К. Эбни. 2017. Этика роботов 2.0: От автономных автомобилей к искусственному интеллекту. Оксфорд: Издательство Оксфордского университета.

Ло, Дж. 2019. Роботеретика. Эйне Эйнфюрунг. Берлин: Зуркамп.

МакЛеннан, Б.Дж. 2014. Этическое обращение с роботами и сложная проблема эмоций роботов. Международный журнал синтетических эмоций 5 (1): 9–16.

Майбом, Х. 2017. Справочник Routledge по философии эмпатии. Лондон: Рутледж.

Миссельхорн, К. 2009. Эмпатия к неодушевленным предметам и зловещей долине. Умы и машины 19 (3): 345–359.

Миссельхорн, К. В печати. Имеет ли сочувствие к роботам моральное значение? В книге «Эмоциональные машины: перспективы аффективных вычислений и эмоционального взаимодействия человека и машины», под ред. К. Миссельхорн и М. Кляйн. Висбаден.

Мори, М. 2005. В зловещей долине. В материалах семинара «Гуманоиды-2005: Виды зловещей долины». Цукуба: Япония.

Нагель, Т. 1974. Каково быть летучей мышью? Философский обзор 83 (4): 435–450.

Ньюэн, А. 2015. Понимание других: теория модели личности. В книге In Open MIND: 26 (T), изд. Т. Метцингер и Дж. М. Виндт. Франкфурт-на-Майне: MIND Group.

Ньюэн А., Л. Де Брюин и С. Галлахер. 2018. Оксфордский справочник по познанию 4E. Оксфорд: Оксфорд Университетское издательство.

Нуссбаум, М. 2011. Перевороты мысли: интеллект эмоций. Кембридж: Кембриджский университет Нажимать.

Плантинга, К. 2009. Трогательные зрители: американские фильмы и зрительский опыт. Беркли: Университет Калифорния Пресс.

Рорти, Р. 2001. Искупление от эгоизма: Джеймс и Пруст как духовные упражнения. Телос 3 (3): 243–263.

Шойц, М. 2011. Архитектурные роли аффекта и способы их оценки в искусственных агентах. Международный журнал синтетических эмоций 2 (2): 48–65.

Шметкамп, С. 2017. Взгляд на нашу жизнь: настроения и эстетический опыт. Философия 45(4): 1681–1695.

Шметкамп, С. 2019. Theorien der Empathie - Ein Einfü hrung. Гамбург: Издательство Юниус.

Шнайдер, С. В печати. Разум будущего: улучшение и превосходство мозга.

Слот, М. 2017. Многоликая эмпатия. Философия 45 (3): 843–855.

Смит, М. 1995. Привлекательные персонажи: художественная литература, эмоции и кино. Оксфорд: Кларендон Пресс.

Собчак, В. 2004. Плотские мысли: воплощение и движущийся образ культуры. Беркли: Университет Калифорния Пресс.

Штейн, Э. 1989. О проблеме эмпатии: Собрание сочинений Эдит Штайн. Том. 3 (3-е исправленное издание), пер. В. Штейн. Вашингтон, округ Колумбия: Публикации ICS.

Штубер, К. 2006. Новое открытие эмпатии: агентность, народная психология и гуманитарные науки. Кембридж, Массачусетс: МТИ Пресс.

Штубер, К. 2018. Эмпатия. В Стэнфордской энциклопедии философии (весеннее издание 2018 г.) под ред. Э. Н. Залта, https://plato.stanford.edu/archives/spr2018/entries/empathy/.

Вааге, М.Б. 2010. Художественные фильмы и разновидности эмпатического взаимодействия. Исследования Среднего Запада в области философии 34: 158–179.

Валлор, С. 2011. Роботы и лица, осуществляющие уход: поддержание этического идеала ухода в 21 веке. Философия и технология 24 (3): 251–268.

Вебер, К. 2013. Каково это — столкнуться с автономным искусственным агентом? ИИ И ОБЩЕСТВО 28: 483–489.

Янал, Р. Дж. 1999. Парадоксы эмоций и художественной литературы. Пенсильвания: Издательство Пенсильванского государственного университета.

Захави, Д. 2011. Эмпатия и прямое социальное восприятие: феноменологическое предложение. Обзор философии и психологии 2 (3): 541–558.

Захави, Д. 2014. «Я и другие: изучение субъективности, сочувствия и стыда». Оксфорд: Оксфордский университет Нажимать.

Захави Д. и Дж. Майкл. 2018. За пределами зеркального отражения: взгляды 4E на эмпатию. В Оксфордском справочнике по познанию 4E, под ред. А. Ньюэн, Л. де Брюин и С. Галлахер, 589–606. Оксфорд: Издательство Оксфордского университета.