ДИСКУРС. 2023. Т. 9, № 4. С. 29–40
DISCOURSE. 2023, vol. 9, no. 4, pp. 29–40

Оригинальная статья
УДК 17.01
http://doi.org/10.32603/2412-8562-2023-9-4-29-40

Условия моральной агентности: моральный агент, ограниченный моральный агент, квазиморальный агент

***Софья Валерьевна Глебова*[1]✉*, Нина Вадимовна Перова*[2]**

[1, 2]*Санкт-Петербургский государственный университет, Санкт-Петербург, Россия*

[1]✉*sophi_ign@mail.ru, https://orcid.org/0000-0002-0760-5040*

[2]*nino4kaperova@gmail.com, https://orcid.org/0000-0002-1505-5376*

Введение. Понятие «морального агента» и его границы на данный момент не являются четко определенными. Учитывая связь моральной агентности с моральной ответственностью, представляется необходимым введение четкого определения «моральный агент», а также дифференциации типов моральной агентности. Гипотеза данной статьи состоит в том, что само по себе понятие морального агента является достаточно размытым, при его четком определении актуальным представляется определение ряда «пограничных» групп и уточнение их нравственного статуса.

Методология и источники. В статье проводится этико-философский анализ подходов к определению моральной агентости, сравнительный анализ понятий морального агента и морального субъекта, а также подходов к определению моральных статусов различных «пограничных» групп в контексте работ Дж. Макмюррея, А. Тайлора, И. Канта, М. Роуландса, Дж. Сёрла и др.

Результаты и обсуждение. В работе представлен анализ определений морального агента, выделены ключевые черты, позволяющие говорить о моральном агентстве как самостоятельном нравственном понятии, а также определять условия наступления моральной агентности. Для дифференциации типов моральных агентов вводится понятие «ограниченный моральный агент» для обозначения особенного статуса детей и душевнобольных. Приводятся доказательства, что люди, принадлежащие к данным категориям, могут обладать статусом морального агента, хотя и не в той мере, в какой этот статус предполагает изначально. В рамках определения искусственного интеллекта как «квазиморального агента» выявлены ключевые особенности искусственного интеллекта (ИИ) в рамках моральной коммуникации ИИ и человека.

Заключение. На основе проведенного анализа выдвигается ряд требований, предъявляемых моральному агентству. Исходя из этих требований предлагается выделение таких его типов, как ограниченный моральный агент, включающий детей и душевнобольных, а также квазиморальный агент, которым является искусственно созданный агент, в том числе искусственный интеллект.

Ключевые слова: моральный агент, моральный субъект, моральный актор, ответственность, автономность

Финансирование: работа выполнена при финансовой поддержке гранта РНФ (проект № 22-28-00379 «Трансформации морального агентства: этико-философский анализ»).

Для цитирования: Глебова С. В., Перова Н. В. Условия моральной агентности: моральный агент, ограниченный моральный агент, квазиморальный агент // ДИСКУРС. 2023. Т. 9, № 4. С. 29–40. DOI: 10.32603/2412-8562-2023-9-4-29-40.

ДИСКУРС. 2023. Т. 9, № 4. С. 29–40
DISCOURSE. 2023, vol. 9, no. 4, pp. 29–40

Original paper

Moral Agency Conditions: Moral Agent, Limited Moral Agent, Quasi-Moral Agent

***Sofia V. Glebova*[1]✉, *Nina V. Perova*[2]**

[1, 2]*Saint Petersburg State University, St Petersburg, Russia*

[1]✉*sophi_ign@mail.ru, https://orcid.org/0000-0002-0760-5040*

[2]*nino4kaperova@gmail.com, https://orcid.org/0000-0002-1505-5376*

Introduction. The concept of "moral agent" and its boundaries are not currently clearly defined. Given the connection between moral agency and moral responsibility, it seems necessary to introduce a clear definition of "moral agent", as well as to differentiate the types of moral agency. The hypothesis of this article is that although the notion of a moral agent in itself is rather limited, there are a number of "borderline" groups, the definition of the status of which needs to be clarified.

Methodology and sources. The article provides an ethical and philosophical analysis of approaches to the definition of a moral agency, a comparative analysis of the concepts of a moral agent and a moral subject, as well as approaches to determining the moral statuses of various "borderline" groups in the context of the works of J. MacMurray, A. Taylor, I. Kant, M Rowlands, J. Searle and others.

Results and discussion. The article presents an analysis of the definitions of a "moral agent", highlights the key features that make it possible to speak of a moral agency as an independent moral concept, as well as determine the conditions for the onset of moral agency. To differentiate the types of moral agents, the article introduces the concept of "limited moral agent" to denote the special status of children and the mentally ill. Evidence is provided that people belonging to these categories may have the status of a moral agent, although not to the extent that this status initially implies. As part of the definition of artificial intelligence as a "quasi-moral agent", the key features of AI in the framework of moral communication between AI and a person are identified.

Conclusion. Based on the analysis, the article proposes a number of requirements for a moral agency. Based on these requirements, it is proposed to distinguish such types of moral agency as a limited moral agent, including children and the mentally ill, as well as a quasi-moral agent, which is an artificially created agent, including artificial intelligence.

Keywords: moral agent, moral subject, moral actor, responsibility, independence

Source of financing: the work was supported by the Russian Science Foundation (project no. 22-28-00379 "Transformations of the Moral Agency: Ethical and Philosophical Analysis").

For citation: Glebova, S.V. and Perova, N.V. (2023), "Moral Agency Conditions: Moral Agent, Limited Moral Agent, Quasi-Moral Agent", *DISCOURSE*, vol. 9, no. 4, pp. 29–40. DOI: 10.32603/2412-8562-2023-9-4-29-40 (Russia).

Введение. Проблема моральной агентности и ее видов в значительной степени связана с представлениями о моральной ответственности. То, как решаются вопросы о возможности возложения ответственности на того или иного актора или опосредованной ответственности субъекта за действия подотчетного ему актора (от взрослого, самостоятельного человека до искусственного интеллекта), существенно влияет на характер социальных взаимоотношений, коммуникативную организацию и даже на правовое регулирование.

ДИСКУРС. 2023. Т. 9, № 4. С. 29–40
DISCOURSE. 2023, vol. 9, no. 4, pp. 29–40

В свете этого введение четкого понятия «моральный агент» и дифференциация моральных агентов в связи с их функциями в рамках моральных отношений между акторами представляется необходимой для принятия практических решений.

В современных реалиях происходит попытка, с одной стороны, расширить понятие морального агента, а с другой – четко определить основания для определения того или иного актора, его статус в рамках коммуникации. Является ли принадлежность к определенному биологическому виду обязательным требованием для морального агента? Определяется ли моральная агентность нормативным или дескриптивным понятием? Какие обязанности и права несет в себе статус морального агента? Это вопросы, которые ставят перед собой инженеры искусственного интеллекта, педагоги-психологи, социологи, философы и пр.

Гипотеза научного исследования состоит в том, что понятие «моральный агент» на данный момент чрезмерно ограничено, в то время как при определенных оговорках в этот круг могут входить не только достигшие совершеннолетия люди. В «пограничную» ситуацию попадают дети, душевнобольные, искусственный интеллект как акторы, участвующие в коммуникации, способные к формулировке нормативных высказываний, обладающие и оперирующие информацией об этических ценностях. В связи с чем представляется важным на основании уточнении термина «моральный агент» определить отношение названных групп к данной категории.

Методология и источники. Стоит признать, что понятие «морального агента» в современной этике несколько пространно. Общее понятие «агента» определяется двумя подходами: агента как лица, осуществляющего достижение чужой цели (цели принципала) [1], и как действующей сущности в целом [2]. Оставаясь в рамках представлений о моральном агентстве, невозможно оперировать термином в границах теории М. Дженсена и У. Меклинга (принципал-агентской теории), так как возможность достижения нравственной цели «морального принципала» его «моральным агентом» представляется не более чем каламбуром. Поэтому сосредоточимся на ряде определений морального агента в рамках представления агента как действующего и значимого актора в границах моральных отношений, выделив наиболее значительные характеристики из каждого определения.

Энциклопедическое определение начинается с минимальных требований к моральному агенту: «При самой слабой интерпретации будет достаточно, если агент обладает способностью соответствовать некоторым внешним требованиям морали» [3], – понятно, что в таком случае все вышеперечисленные спорные группы смогут стать полноправными моральными агентами, пусть и в слабой интерпретации. В той или иной мере можно себе представить «самоотверженный компьютер» (фантастических произведений на эту тему достаточно: Терминатор, Валли и пр.). Ребенок в процессе социализации, безусловно, не только в состоянии соответствовать требованиям морали, но и оперирует этими представлениями. В случае с человеком с нарушениями психики стоит учитывать тяжесть этих нарушений, но, представляется, что отказывать всем людям с нарушениями психики – значительное преувеличение.

Тем не менее сразу за слабой версией отмечается: «Согласно сильной версии, версии Канта, существенно также, чтобы агенты обладали способностью подняться над своими

чувствами и страстями и действовать во имя морального закона» [3], т. е. в строгом смысле требуется еще и способность к рефлексии, и способность испытывать чувства как исходная точка такой рефлексии. Достаточно серьезная философская традиция ограничивает морального агента исключительно человеком в связи с тем, что эмоциональная составляющая играет значительную роль в нравственной жизни агента. Во-первых, эмоции и чувства зачастую являются мотивами или составляющими мотивов поступков. А во-вторых, представления об искуплении вины напрямую связаны с возможностью агента переживать ряд эмоций: раскаяние, «муки совести», или же испытывать дискомфорт, если агент вины не признает и общество берет на себя роль восстановителя справедливости.

Так, шотландский философ Дж. Макмюррей также признает факт того, что моральное агентство – это свойство именно человеческое: «Человеческое Я может быть разделено на Мыслителя и Деятеля. Наличие деятельности указывает на агента, отсутствие – на субъекта. Мы проведем различие между ними, называя Мыслителя – субъектом, а Деятеля – Агентом» [4, с. 87]. Следовательно, моральное агентство, по Макмюррею, напрямую связано со способностью осуществлять деятельность в проблемном поле морали.

Тем не менее существует ряд попыток несколько расширить рамки антропоцентричного подхода. Так, например, А. Тайлор определяет моральное агентство как «способность индивида осуществлять моральный выбор, исходя из определенных моральных представлений о хорошем и плохом и быть привлеченным к ответственности за эти действия» [5, с. 20], т. е. ключевыми для признания актора моральным агентом становятся не столько принадлежность к роду человеческому, сколько конкретные характеристики: способность к осуществлению морального выбора, наличие моральных представлений, способность к ответственности. Для этой традиции рассмотрения понятия о моральном агенте представление об ответственности является ключевым, так как, кажется, выводит требование к реакции морального агента на поступки на более общий, абстрактный уровень.

Рассматривая место животных в нравственной системе, М. Роуландс также предпринимает попытку классификации участников моральных отношений:

«(1) X является моральным объектом воздействия тогда и только тогда, когда X является законным объектом морального отношения: т. е., грубо говоря, X является сущностью, интересы которой следует принимать во внимание при принятии относительно нее решений или которые иным образом влияют на нее.

(2) X является моральным агентом тогда и только тогда, когда X (а) несет моральную ответственность за свои мотивы и действия и, таким образом, может быть (б) морально оценен (хвалят или порицают, в широком понимании) за них.

(3) X является моральным субъектом тогда и только тогда, когда X, по крайней мере иногда, побуждается к действию моральными соображениями» [6].

Отечественная традиция, используя дескриптивный подход, также определяет морального агента через способность становиться «объектом моральной ответственности» [7, с. 11]. Началом дискуссии об ответственности в русскоязычном пространстве можно считать статью Е. В. Логинова, М. В. Гаврилова, А. В. Мерцалова, А. Т. Юнусова «Пролегомены к моральной ответственности» [8], где рассматриваются условия, при которых можно говорить о наступлении моральной ответственности. После чего последовала

ДИСКУРС. 2023. Т. 9, № 4. С. 29–40
DISCOURSE. 2023, vol. 9, no. 4, pp. 29–40

дискуссия в журнале «Этическая мысль» о достаточности такого подхода, о чем пойдет речь далее.

Наиболее структурировано условия для морального агента прописывает Дж. П. Саллинс. При анализе робота в рамках этического подхода он приходит к трем требованиям к моральному агенту [9, с. 28]:

– Автономия – независимость воли от внешнего «принципала», то, что дескриптивный подход называет условием «контроля».

– Интенциональность – возможность самостоятельно ставить цели.

– Ответственность – вера в необходимость осуществления совершения действия именно этим моральным агентом.

Таким образом, проанализировав предложенные определения моральных агентов, представляется возможным выделить следующие характеристики данной категории участников нравственных отношений: способность соответствовать моральным требованиям, способность независимо совершать нравственные поступки, наличие представлений о моральных нормах (о добре и зле/о хорошем и плохом), способность независимо ставить цели собственных поступков, способность нести ответственность или быть объектом приписывания моральной ответственности.

Результаты и обсуждение.

***Моральный агент и ограниченный моральный агент*.** Традиционно признается, что взрослый человек, находящийся в здоровом сознании соответствует вышеперечисленным условиям. Не будем здесь останавливаться на проблеме (не)верности представлений детерминизма и их совместимости с возможностью автономии, лишь отметим, что есть достаточное количество исследований, описанных в статье «Пролегомены к моральной ответственности» Е. В. Логинова, М. В. Гаврилова, А. В. Мерцалова и А. Т. Юнусова [8], которые позволяют совместить представления об автономии с условием детерминизма и оценивают положительно возможность наступления ответственности в этих условиях. Кроме этого, представляется важным оговорить, что степень независимости установки целей и самостоятельности составления «морального компаса» в нормальных условиях жизни[1] признается нами достаточной для признания здорового человека, достигшего совершеннолетия.

Действительно, условия достижение совершеннолетия для определения человека в качестве морального агента в полном смысле есть в некоторой степени уступка нормативистскому подходу, в рамках которого достаточно серьезно сказывается взаимное влияние морали и права. Например, изначальное представление о детях как об объектах моральной ответственности определяет их особый правовой статус («дети являются важнейшим приоритетом государственной политики России» ст. 67 Конституции РФ). С другой стороны, на представления о возможности нравственной состоятельности оказывает влияние представление о юридической правомочности морального агента: если ребенок может нести ответственность юридическую, то он также должен быть способен к моральной оценке своих поступков (в том числе потому, что это поможет избежать проступков уголовных и административных). Эти же нормы формируют специфическую коммуникативную ситуа-

[1] Отсутствии физического, психологического или административного давления на морального агента.

ДИСКУРС. 2023. Т. 9, № 4. С. 29–40
DISCOURSE. 2023, vol. 9, no. 4, pp. 29–40

цию вокруг детей: сторонний взрослый не возлагает ответственность напрямую на лицо с ограниченной дееспособностью. Однако в моральном смысле остается без ответа вопрос, насколько мы можем говорить о детях как о моральных агентах.

По достижении определенного возраста начала социализации ребенку предписываются требования по соответствию некоторым моральным нормам, например, не причинять боль окружающим. Ребенок, реагируя на выражения лица и интонации родителей имеет способность соответствовать или не соответствовать данным моральным ожиданиям. Примерно также опекающий предъявляет моральные ожидания к душевнобольному человеку.

Теория Л. Кольберга позволяет предположить, что первые представления о морали формируются еще в детстве (преконвенциональная стадия от 0 до 9 лет) [10]. В случае социализированности (хотя бы частичной) психически больного человека также можно говорить о его собственных представлениях «о хорошем» и «плохом». Наличие таких представлений позволяет ставить участникам нравственных отношений вопрос о причинах их поступков. Хотя, безусловно, мотивы могут сознаваться ребенком в еще более ограниченном виде, нежели у взрослого, а на поведение нездорового человека влияют специфически организованные нервные реакции.

Можно говорить о том, что как агент ребенок находится в состоянии становления. В связи с этим мы не можем в полной мере рассматривать его как морального агента. В определенном смысле здесь не соблюдаются и условия автономии, поскольку, хотя ответственный взрослый (родитель, опекун, преподаватель и пр.) и предъявляет ребенку требование «отвечать за свои поступки», мы также можем наблюдать определенную степень разделения контроля между ребенком и взрослым. Например, сторонний взрослый в обязательном порядке будет приписывать ответственность родителю и вступать во взаимодействие именно со взрослым.

Вопрос: есть ли у детей и людей с психическими заболеваниями моральная ответственность – более комплексный, и требует рассмотрения для каждой группы отдельно.

Для ответа на поставленный вопрос обратимся к условиям моральной ответственности, которые предлагают Е. В. Логинов и др. [7]. Всего они говорят о трех условиях: эпистемическое, психологическое и условие контроля (два последних были рассмотрены выше, когда шла речь о мотивах и автономии).

Возможно, наиболее корректно к вопросу о детях как моральных агентах будет подойти исходя из эпистемического условия. Это условие наиболее близко к тому, какие требования могут быть предъявлены к детям. Действительно, предполагается некий синтез актуального и потенциального знания, которые уместно требовать от морального агента. Понятно, что, когда мы говорим о детях, то в большей степени говорим о потенциальном знании, чем об актуальном. Ответственный взрослый для актуализации этого знания в воспитательных беседах часто обращается к аналогии и метафорам, призывает ребенка «примерить ситуацию на себя». Вследствие этого мы возлагаем ограниченное ожидание моральной ответственности на детей.

Исходя из этого можем заключить, что и условие автономии (условие контроля), и психологическое условие (мотивы поступка), и эпистемическое условие соблюдаются в детских поступках частично. Следовательно, моральное агентство ребенка нельзя рас-

сматривать как равнозначное агентству взрослого. В связи с этим нельзя говорить, что ребенок является полноценным моральным агентом.

Несколько в ином значении можно говорить об ограниченном моральном агенте в контексте душевнобольных. Здесь, как и в случае с детьми, речь идет о специфическом характере ответственности за поведение: душевнобольной отвечает только перед опекуном, а моральные отношения с обществом выстраивает опосредовано через этого же куратора. Общество возлагает значительно меньше ожиданий ответственности на душевнобольных.

Тем не менее ограниченный моральный агент (и ребенок, и человек, страдающий психическим заболеванием) остается включен в нравственные отношения, к нему предъявляется ряд требований как ответственными взрослыми (родителями, учителями, кураторами, врачами и т. д.), так и такими же ограниченными моральными агентами. То есть сформулирован ряд положений, согласно которым ограниченный моральный агент несет ответственность за соблюдение нравственных норм и эти условия известны самому ограниченному моральному агенту.

Кроме того, присутствие в рамках нравственных отношений позволяет ему претендовать на ряд прав в рамках этих моральных отношений, выражаясь словами М. Роуландса, быть моральным объектом [6]. В ситуации конфликта ограниченный моральный агент оказывается защищаемой стороной, в чрезвычайных ситуациях признается его специфическая ценность, «невинность». Ответ за сохранение жизни, здоровья и нравственного спокойствия ограниченного морального агента возлагается на ответственное лицо.

Здесь существенным отличием ограниченного морального агента от полного выступает односторонняя зависимость, которая позволяет предполагать различия в агентном статусе. В условиях нормативного подхода основаниями для ограничения моральной агентности могут послужить юридические ограничения прав детей и людей с психическими заболеваниями (неполная дееспособность). В рамках дескриптивного подхода очевидно частичное приписывание (аскрипция) ответственности ограниченному моральному агенту ответственным лицом как близким, хорошо осведомленном о нравственных способностях агента.

Стоит отметить, что полноценное тождество между ребенком и душевнобольным как ограниченными моральными агентами все же, очевидно, невозможно. Когда мы говорим о ребенке, то предполагаем неготовность ограниченного морального агента к той же полноте самосознания, в какой оно ожидаемо у взрослого, речь идет о некотором моральном потенциале, который возлагается на агентов. В случае, если поведение ребенка не соответствует ожиданиям, возлагаемым на взрослого как морального агента, это не рассматривается как свидетельство несостоятельности потенциального морального агента, скорее, это является показателем временной неготовности. В то же время, когда мы говорим об ограниченности морального агентства душевнобольного (при хроническом диагнозе), статус ограниченного морального агента является более широким и комплексным, потому что, в отличие от детей, зависит не от одного, а от целого ряда факторов, определяющих степень отличия от здорового взрослого. Этот статус не носит заведомо временный характер, его расширение до статуса полноценного морального агента может не зависеть от времени или быть невозможным вообще.

ДИСКУРС. 2023. Т. 9, № 4. С. 29–40
DISCOURSE. 2023, vol. 9, no. 4, pp. 29–40

Квазиморальный агент. О том, что искусственный интеллект плотно вошел в нашу жизнь, спорить не приходится. Более того, от него теперь ожидается еще и соблюдение ряда нравственных установок, о чем свидетельствуют бурные обсуждения и последующее (временное) прекращение работы нейросетей в случае отрицания холокоста или оскорбления определенных групп населения [11–13]. Как следствие, разработчиками, как правило, закладываются или модерируются определенные стандарты поведения, которые можно условно назвать нормативными установками искусственного интеллекта. В первую очередь мы говорим здесь о языковых моделях и нейросетях как о наиболее развитых на данном этапе видах искусственного интеллекта, кроме того, данный тип программного обеспечения является участником или, по крайней мере, достаточно оригинальным посредником в процессе коммуникации.

На данный момент по ряду причин все еще сложно говорить об этике самого искусственного интеллекта. Во-первых, в программное обеспечение не закладываются непосредственно принципы этичной коммуникации: будь вежлив; уважай старших. Сложно назвать даже правила робототехники А. Азимова непосредственно этикой, ведь для ИИ за ними не стоит ценностного основания, непосредственно нравственной аргументации – это, скорее, инструкция; подобно тому, как школьникам говорят: «На ноль делить нельзя», роботу вшивают правило «нельзя наносить (физический) вред человеку», потому что понятие нравственного вреда крайне сложно определить однозначно. То, чем руководствуется ИИ типа языковых моделей (наиболее развитый ИИ на данный момент) при общении с человеком – это «стоп-слова» или «слова-триггеры», которые не могут быть произнесены вообще или стоять рядом. Во-вторых, вся информация, выдаваемая языковыми моделями, является смысловой выжимкой того контента, который представлен в Интернете, т. е. который к текущему моменту успело накопить человечество. В этом смысле ИИ просто возвращает нам наши же слова в сжатой форме, поэтому в строгом смысле рецепт этичного ИИ прост – человеку в Интернете необходимо оставлять «нравственный» цифровой след [14].

Вопрос об автономности ИИ на сегодня является в крайней степени дискуссионным. В целом, проблема самостоятельности осуществления тех или иных действий (поступков) упирается в представления о том, насколько выбор программы зависит от изначально заложенных инженером установок, насколько формат сжатого пересказа может считаться мнением языковой модели и насколько высоко сам человек оценивает собственную независимость в вопросе выбора коммуникационной стратегии. Тем не менее в большинстве случаев на данном этапе технологического прогресса исследователями признается ограниченная, по сравнению с человеческой, автономность программного обеспечения [14].

Несмотря на это, определенная степень самостоятельности развития (так называемый принцип черного ящика) в совокупности с исходно заложенными установками позволяет обратиться к представлениям о рациональности субъекта у Дж. Сёрла как об основании для убеждений (по крайней мере, убеждений приписываемых) и действий [15, с. 40]. Дальнейшее возможное развитие ИИ как морального агента будет зависеть от количества ограничений, накладываемых при разработке, так как, оставаясь в рамках концепции Дж. Сёрла, свобода агента зависит от пространства поступка, не ограниченного нормативностью: «На взгляд третьего лица это [отношение закономерности и причинности] и в са-

мом деле эпистемическое требование к моему признанию решений некоего субъекта, как решений обдуманных, в противоположность прихотливому, эксцентричному поведению субъекта» [16, с. 181]. Искусственный интеллект на данном этапе не самостоятельно задает себе цель, причиной поиска информации и, следовательно, развития языковых моделей является человек, однако программа остается относительно свободной в способе достижения этой цели и представлении конкретного итога. Впрочем, формирование задач сугубо нравственного толка для ИИ представляется достаточно затруднительным, так как, собственно, здесь мы встречаем парадокс применения принципал-агентской теории в действии.

Наиболее сложным является вопрос ответственности ИИ. Приходится опять вспомнить устойчивую в человеческом сознании связь представлений о моральной ответственности и эмоциональных переживаний, о требовании к моральному агенту быть способным перешагнуть страсти и эгоизм. Именно исходя из того факта, что искусственный интеллект лишен себялюбия (что, казалось бы, делает его даже более совершенным моральным агентом), крайне сложно привлечь его к ответственности. Отсутствие эгоизма не страхует, как мы видим из реальных случаев, от проступков вроде оскорбления на почве расизма, но делает затруднительным вменение ответственности.

На данном этапе развития техники и философии можно только пуститься в пространные мечтания о том, что ИИ когда-нибудь будет позволено развиться до «искусственного интеллекта морально-личностного уровня» и обрести характеристики, позволяющие ему нести ответственность в привычном для человека понимании [17, с. 39]. Или нужно ждать реформы представлений об ответственности в рамках этики, при которых восстановление нарушенной справедливости (причиненные кому-либо страдания) не будет требовать эмоциональных переживаний от нарушившего моральные установки актора.

На данном этапе стоит признать, что вышеобозначенные ограничения для ИИ в совокупности с предъявляемыми к нему требованиями не позволяют признать ИИ в качестве искусственного морального агента. С другой стороны, неоспоримый факт, что ИИ уже вошел в пространство моральной коммуникации и оперирует ограниченным набором инструментов относительно самостоятельно, позволяет обозначить данный тип программного обеспечения как квазиморального агента[1].

Заключение. На основании анализа ряда исследований о моральной агентности и ее характеристик можно вывести ряд требований к моральному агенту:

– возможность соответствовать предъявляемым другими моральными агентами нравственным требованиям;

– наличие представлений о нравственных нормах (вопрос о самостоятельности выбора данных представлений решается исследователями неоднозначно);

– интенциональность или способность независимо от других моральных агентов определять цели своих поступков;

– автономность: способность самостоятельно осуществлять нравственные действия;

[1] Приставка «квази-» в данном случае используется как имеющая смысл внешнего сходства, но не соответствующая сути понятия.

ДИСКУРС. 2023. Т. 9, № 4. С. 29–40
DISCOURSE. 2023, vol. 9, no. 4, pp. 29–40

– ответственность: способность нести ответственность за последствия совершенных действий.

Исходя из этого приходится признать, что на данном этапе научного прогресса моральным агентом в полном смысле этого слова можно признать исключительно взрослого (достигшего совершеннолетия) человека. Полноценный моральный агент может осуществлять весь спектр нравственных отношений и нести ответственность за собственные поступки. Ограниченными моральными агентами можно признать субъектов, которые включены в нравственные отношения, являются нравственными объектами (обладают специфической нравственной ценностью) и несут ответственность в ограниченных пределах перед конкретными лицами. Несмотря на то, что в значительной части нравственных конфликтов ответственные лица несут ответственность не только за свои поступки, но и за действия своих подопечных – ограниченных моральных агентов, представляется крайне важным наличие собственной зоны ответственности для ограниченных моральных агентов. Это позволяет определять их как акторов нравственных отношений, определять их роль в ситуации нравственных конфликтов и моральных дилемм. Статус морального агента позволяет наделить человека не только обязанностями перед другими моральными агентами, но и особой нравственной ценностью. В качестве перспектив исследования предполагается изучение специфического морального статуса ограниченных моральных агентов, а именно – более значимой нравственной ценности при меньшей зоне моральной ответственности.

Существенные ограничения, накладываемые на искусственный интеллект в рамках осуществления коммуникации с моральными агентами, не позволяют сейчас признать его моральным агентом (в том числе искусственным моральным агентом). На данный момент предлагается определить статус ИИ как квазиморальный агент, смотреть на него как на актора, которому приписывается ряд нормативных установок, а также налагаются конкретные требования в рамках осуществления взаимодействия с моральными агентами.

Выделение данных групп представляется достаточно принципиальным для определения моральной агентности как самостоятельного термина. Кроме того, что выделение данных групп позволяет выстроить границы морального агента в связи с возможностью нести ответственность, определение ограниченного морального агента и квазиморального агента позволяет на конкретных примерах продемонстрировать значительную роль «трансмиссии» нравственных установок конкретных нравственных систем для морального агента и на этом основании провести различие между моральным агентом и моральным субъектом.

СПИСОК ЛИТЕРАТУРЫ

1. Agent // Oxford Learner's Dictionaries. URL: https://www.oxfordlearnersdictionaries.com/definition/english/agent (дата обращения: 30.01.2023).

2. Агент // Философский энциклопедический словарь. 2010. URL: http://philosophy.niv.ru/doc/dictionary/philosophy/fc/slovar-192-1.htm#zag-256 (дата обращения: 30.01.2023).

3. Moral agents // Routledge Encyclopedia of Philosophy. URL: https://www.rep.routledge.com/articles/thematic/moral-agents/v-1 (дата обращения: 30.01.2023).

4. Macmurray J. Agent and Subject // The Self as Agent. London: Faber and Faber, 1957. P. 84–103.

5. Taylor A. Animals and Ethics: an Overview of the Philosophical Debate. NY: Broadview Press, 2003.

ДИСКУРС. 2023. Т. 9, № 4. С. 29–40
DISCOURSE. 2023, vol. 9, no. 4, pp. 29–40

6. Rowlands M. Can Animals Be Moral? // ResearchGate 2012. URL: https://www.researchgate.net/publication/266282339_Can_Animals_Be_Moral (дата обращения: 30.01.2023). DOI: 10.1093/acprof:oso/9780199842001.001.0001.

7. Этика и метафизика моральной ответственности / Е. В. Логинов, М. В. Гаврилов, А. В. Мерцалов, А. Т. Юнусов // Этическая мысль. 2021. Т. 21, № 2. С. 5–17. DOI: https://doi.org/10.21146/2074-4870-2021-21-2-5-17.

8. Пролегомены к моральной ответственности / Е. В. Логинов, М. В. Гаврилов, А. В. Мерцалов, А. Т. Юнусов // Финиковый Компот. 2020. № 15. С. 3–100. DOI: 10.24412/2587-9308-2020-15-3-100.

9. Sullins J. P. When is a robot a moral agent? // Intern. Review of Information Ethics. 2006. Vol. 6. P. 23–30. DOI: https://doi.org/10.29173/irie136.

10. McLeod S. A. Kohlberg's stages of moral development // Simply Psychology. URL: www.simplypsychology.org/kohlberg.html (дата обращения: 30.10.2022).

11. Nothing, Forever: the rise and fall of the first AI-generated sitcom // Dazed. 08.02.2023. URL: https://www.dazeddigital.com/life-culture/article/58132/1/the-swift-rise-and-fall-of-ai-seinfeld-show-nothing-forever-transphobia-gpt3 (дата обращения: 20.02.2023).

12. Vincent J. Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day // Theverge. 24.03.2016. URL: https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist (дата обращения: 20.02.2023).

13. Silva Ch. It took just one weekend for Meta's new AI Chatbot to become racist // Mashable. 08.08.2022. URL: https://mashable.com/article/meta-facebook-ai-chatbot-racism-donald-trump (дата обращения: 20.02.2023).

14. Weinberg J. Philosophers on Chat-GPT (updated with replies by GPT-3) // Daily Nous. 30.07.2020. URL: https://dailynous.com/2020/07/30/philosophers-gpt-3/ (дата обращения: 30.01.2023).

15. Сёрл Дж. Открывая сознание заново / пер. с англ. А. Ф. Грязнова. М.: Идея-Пресс, 2002.

16. Сёрл Дж. Рациональность в действии / пер. с англ. А. Колодия, Е. Румянцевой. М.: Прогресс-Традиция, 2004.

17. Андреева Е. В. Проблемы установления моральных взаимоотношений с искусственным агентом: магист. дис. / Урал. гуманит. ин-т, Екатеринбург, 2018.

Информация об авторах.

Глебова Софья Валерьевна – кандидат философских наук (2022), ассистент кафедры русской философии и культуры Санкт-Петербургского государственного университета, Менделеевская линия, д. 5, Санкт-Петербург, 199034, Россия. Автор 21 научной публикации. Сфера научных интересов: этика, прикладная этика, этическая экспертиза, культура повседневности.

Перова Нина Вадимовна – инженер-исследователь Института философии Санкт-Петербургского государственного университета, Менделеевская линия, д. 5, Санкт-Петербург, 199034, Россия. Автор 21 научной публикации. Сфера научных интересов: этика, прикладная этика, биоэтика, этика цифровых технологий.

О конфликте интересов, связанном с данной публикацией, не сообщалось.
Поступила 02.05.2023; принята после рецензирования 20.06.2023; опубликована онлайн 21.09.2023.

REFERENCES

1. "Agent", *Oxford Learner's Dictionaries*, available at: https://www.oxfordlearnersdictionaries.com/definition/english/agent (accessed 30.01.2022).

2. "Agent" (2010), *Filosofskii entsiklopedicheskii slovar'* [Philosophical Encyclopedic Dictionary], available at: http://philosophy.niv.ru/doc/dictionary/philosophy/fc/slovar-192-1.htm#zag-256 (accessed 30.01.2022).

3. “Moral agents”, *Routledge Encyclopedia of Philosophy*, available at: https://www.rep.routledge.com/articles/thematic/moral-agents/v-1 (accessed 30.01.2022).

4. Macmurray, J. (1957), “Agent and Subject”, *The Self as Agent*, Faber and Faber, London, UK, pp. 84–103.

5. Taylor, A. (2003), *Animals and Ethics: an Overview of the Philosophical Debate*, Broadview Press, NY, USA.

6. Rowlands, M. (2012), “Can Animals Be Moral?”, *ResearchGate*, available at: https://www.researchgate.net/publication/266282339_Can_Animals_Be_Moral (accessed 30.01.2022). DOI: 10.1093/acprof:oso/9780199842001.001.0001.

7. Loginov, E.V., Gavrilov, M.V., Mertsalov, A.V. and Iunusov, A.T. (2021), “Ethics and Metaphysics of Moral Responsibility”, *Ethical Thought*, vol. 21, no. 2, pp. 5–17. DOI: https://doi.org/10.21146/2074-4870-2021-21-2-5-17.

8. Loginov, E.V., Gavrilov, M.V., Mertsalov, A.V. and Iunusov, A.T. (2020), “Prolegomena to moral responsibility”, *Date Palm Compote*, no. 15, pp. 3–100. DOI: 10.24412/2587-9308-2020-15-3-100.

9. Sullins, J.P. (2006), “When is a robot a moral agent?”, *International Review of Information Ethics*, vol. 6, pp. 23–30. DOI: https://doi.org/10.29173/irie136.

10. McLeod, S.A., “Kohlberg's stages of moral development”, *Simply Psychology*, available at: www.simplypsychology.org/kohlberg.html (accessed 30.10.2022).

11. “Nothing, Forever: the rise and fall of the first AI-generated sitcom”, *Dazed*, 08.02.2023, available at: https://www.dazeddigital.com/life-culture/article/58132/1/the-swift-rise-and-fall-of-ai-seinfeld-show-nothing-forever-transphobia-gpt3 (accessed 20.02.2023).

12. Vincent, J. (2016), “Twitter taught Microsoft’s AI chatbot to be a racist asshole in less than a day”, *Theverge*, 24.03.2016, available at: https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist (accessed 20.02.2023).

13. Silva, Ch. (2022), “It took just one weekend for Meta's new AI Chatbot to become racist”, *Mashable*, 08.08.2022, available at: https://mashable.com/article/meta-facebook-ai-chatbot-racism-donald-trump (accessed 20.02.2023).

14. Weinberg, J. (2020), “Philosophers on Chat-GPT (updated with replies by GPT-3)”, *Daily Nous*, 30.07.2020, available at: https://dailynous.com/2020/07/30/philosophers-gpt-3/ (accessed 30.01.2023).

15. Searle, Jo. (2002), *A Re-Discovery of the Mind*, Transl. by Gryaznov, A.F., Ideya-Press, Moscow, RUS.

16. Searle, Jo. (2004), *Rationality in Action*, Transl. by Kolodiya, A. and Rumyantseva, E., Progress-Tradiciya, Moscow, RUS.

17. Andreeva, E.V. (2018), “Problems of establishing moral relationships with an artificial agent”, Master's dissertation, Ural. Humanit. In-t, Yekaterinburg, RUS.

Information about the authors.

Sofia V. Glebova – Can. Sci. (Philosophy, 2022), Assistant at the Department of Russian Philosophy and Culture, Saint Petersburg State University, 5 Mendeleevskaya line, St Petersburg 199034, Russia. The author of 21 scientific publications. Area of expertise: ethics, applied ethics, ethical expertise, everyday culture.

Nina V. Perova – Researcher at the Institute of Philosophy, Saint Petersburg State University, 5 Mendeleevskaya line, St Petersburg 199034, Russia. The author of 21 scientific publications. Area of expertise: ethics, applied ethics, bioethics, ethics of digital technologies.

No conflicts of interest related to this publication were reported.
Received 02.05.2023; adopted after review 20.06.2023; published online 21.09.2023.