Экономические и социально-гуманитарные исследования. 2025. Т. 12. № 1. С. 110–117.
Economic and Social Research. 2025. Vol. 12. No. 1. P. 110–117.
Научная статья

УДК 165.12:004.89
DOI: 10.24151/2409-1073-2025-12-1-110-117
EDN: IEUPFI

Может ли искусственный интеллект обладать свойствами субъектности: философские аспекты проблемы

Е. О. Труфанова

Институт философии РАН, Москва, Россия

iph@etrufanova.ru

Аннотация. Обнаружен пробел в знаниях: искусственным интеллектуальным системам иногда ошибочно приписываются свойства субъектности. Высказана и доказана гипотеза, что подобные ошибки базируются на акцентуации возможности систем искусственного интеллекта решать интеллектуальные задачи, когда эта возможность неверно воспринимается как признак существования у технической системы других свойств субъекта: воли, ответственности, сознания. Коммуникативные возможности современных систем искусственного интеллекта, имитирующие общение с человеком, способствуют закреплению этой ошибки. Обосновано, что системы искусственного интеллекта являются технологическим инструментом, за создание и использование которого может нести ответственность только человек, и только человек, в свою очередь, может выступать как субъект мышления и деятельности.

Ключевые слова: субъектность искусственного интеллекта, агентность искусственного интеллекта, квазисубъектность искусственного интеллекта, ответственность искусственного интеллекта, правоспособность искусственного интеллекта

Для цитирования: Труфанова Е. О. «Может ли искусственный интеллект обладать свойствами субъектности: философские аспекты проблемы». *Экономические и социально-гуманитарные исследования* 12.1 (2025): 110–117. https://doi.org/10.24151/2409-1073-2025-12-1-110-117 EDN: IEUPFI

Original article

Whether artificial intelligence can have the properties of subjectivity: Philosophical aspects of problem

E. O. Trufanova

Institute of Philosophy of the Russian Academy of Sciences, Moscow, Russia

iph@etrufanova.ru

Abstract. A gap in knowledge has been discovered: artificial intelligent systems are sometimes mistakenly attributed the properties of subjectivity. The hypothesis is expressed and proved that such errors are based on the accentuation of the ability of artificial intelligence systems to solve intellectual tasks, when this possibility is incorrectly perceived as a sign of the existence of other properties of the subject in the technical system: will, responsibility, and consciousness. The communication capabilities of modern artificial intelligence systems that simulate communication with humans contribute to perpetuating this error. It has been proved that artificial intelligence systems are a technological tool, for the creation and use of which only a person can be responsible, and only a person, in turn, can act as a subject of thinking and activity.

Keywords: artificial intelligence subjectivity, artificial intelligence agency, artificial intelligence quasi-subjectivity, artificial intelligence responsibility, artificial intelligence legal capacity

For citation: Trufanova E. O. "Whether Artificial Intelligence Can Have the Properties of Subjectivity: Philosophical Aspects of Problem". *Ekonomicheskie i sotsial'no-gumanitarnye issledovaniya = Economic and Social Research* 12.1 (2025): 110–117. (In Russian). https://doi.org/10.24151/2409-1073-2025-12-1-110-117

Введение

Может ли искусственный интеллект выступать как субъект? Этот вопрос сейчас активно обсуждается как в научной литературе, так и в публицистике, в первую очередь в социогуманитарных науках, поскольку речь идет о сопоставлении искусственного интеллекта и человека и о том, какую роль искусственный интеллект играет в жизни общества, в частности — в каких социальных ролях он может выступать (Лекторский, Васильев, Макаров и др., 2024). Особенно остро вопрос об искусственном интеллекте как субъекте стоит в сфере права (Хабриева, 2024) — речь идет о правосубъектности систем искусственного интеллекта. Вопрос о внедрении в повседневную жизнь «умных» систем, которые будут в ряде сфер принимать решение за человека, становится всё актуальнее — он уже не из области научной фантастики, это вопрос сегодняшнего дня. Поэтому и вопрошание о субъектности подобных систем перестает быть праздным любопытством.

К определению понятий

Прежде всего необходимо уточнить, что предмет данного исследования — искусственный интеллект, интеллектуальные технологии, реализуемые на различных технических устройствах. Речь идет о существующем в пространстве повседневности, а не о воображаемом или прогнозируемом искусственном интеллекте (и связанных с ним воображаемых угрозах). Принято выделять так называемый *сильный* и *слабый* искусственные интеллекты. В обыденной речи (не в теоретическом дискурсе, когда

мы проводим эту дистинкцию) под искусственным интеллектом часто подразумевается нечто среднее между этими двумя разновидностями — еще не субъект, способный организовать «восстание машин», но уже феномен, обладающий волей (субъектностью) и целеполаганием (агентностью). Как справедливо отмечает российский философ, специалист по проблемам искусственного интеллекта Е. А. Алексеева: «Когда мы называем искусственные интеллектуальные системы искусственным интеллектом, мы попадаем в ловушку языка. Мы начинаем от них ждать того, чего они не делают, бояться их, приписывать им некоторую квазисубъектность, которой они не обладают» (Лекторский, Алексеева, Емельянова и др., 2022: 9). На наш взгляд, постановка проблемы субъектности сильного искусственного интеллекта в настоящий момент преждевременна, поскольку теоретики еще не разрешили в положительном ключе вопрос о том, возможно ли в принципе создание «сильного» (или «общего») искусственного интеллекта. Поскольку «сильный» искусственный интеллект вряд ли будет создан в обозримом будущем, мы можем рассуждать только о том, наделена ли искусственная интеллектуальная технологическая система субъектностью, оставив проблему возможности его очеловечивания будущим поколениям.

Следует уточнить, что значит «обладать субъектностью» или «выступать как субъект». Слово «субъект» в современных исследованиях используется, как нам представляется, излишне свободно. Если мы подразумеваем под субъектом актора, производящего действие, то в обсуждении поставленной проблемы можно поставить точку: то, что искусственная интеллектуальная система способна совершать *действия*, не подлежит сомнению. Но в контексте обсуждений искусственного интеллекта ставятся вопросы о его правоспособности, об авторстве (например, текстов или произведений искусства, созданных с помощью искусственного интеллекта), об «электронных лицах» (Ястребов, 2018), что явно выходит за рамки просто указания на того, кто (или что) совершает действие. Катящийся с горы камень, увлекающий за собой сель, совершает действие, но мы не задаемся вопросом о том, должен ли он нести уголовную ответственность за разрушенные селем дома или же может избежать ее, переложив «вину» на растаявший под солнцем ледник, лишивший камень устойчивой основы и запустивший, таким образом, необратимый процесс. Так в чем же отличие искусственного интеллекта от действующего камня и селя?

Принципиальное отличие, разумеется, в том, что искусственные интеллектуальные системы выполняют не просто действия, но *интеллектуальные* операции, которые ранее считались исключительно прерогативой человека. Второе отличие состоит в том, что в современном мире искусственные интеллектуальные системы начинают использоваться для сопровождения многих социальных процессов. В результате различным нечеловеческим «интеллектуальным» агентам, действующим как в цифровом пространстве, так и за его пределами, начинает приписываться статус субъекта (Труфанова, 2024). Во многом это происходит из-за указанной Е. А. Алексеевой «языковой ловушки» — сами наши высказывания об искусственном интеллекте как бы подразумевают эту субъектность. Иногда это происходит неосознанно, из-за неточности высказываний, но часто мы вполне осознанно приписываем субъектность искусственным интеллектуальным системам, возлагая на них ответственность за какие-то действия или высказывания, или вступая с ними во взаимодействие.

С точки зрения российского философа науки и техники Т. Г. Лешкевич, «когда ведется речь о гипотетическом признании субъектности нейросетей ИИ, это означает

признание за ними суверенной системы действий, свободы от вложенных в них моделей, основанных на Больших данных» (Лешкевич, 2024: 127). Она указывает в своем исследовании на ряд фиксируемых типов «поведения» нейросетей, которые могли бы трактоваться как таковые суверенные действия (например, одна нейросеть может «критиковать» другую за неправильное решение задачи) (Лешкевич, 2024), однако, с нашей точки зрения, всё это «поведение» должно быть интерпретировано лишь как имитация субъектности.

Имитация субъектности

Знаменитый тест Тьюринга, который, в своем классическом варианте, вероятно, покажется нам сейчас устаревшим, был основан на так называемой *игре в имитацию*. Имитация действительно играет здесь ключевую роль. Интеллектуальные системы не являются субъектами, но зачастую имитируют их. Человеку проще всего осуществлять коммуникацию с другими людьми посредством языка, это его наиболее естественная форма взаимодействия с другими. Поэтому многие интерфейсы, через которые мы взаимодействуем с современными интеллектуальными техническими системами, сконструированы для имитации коммуникации подобного рода. Для многих пользователей это не только упрощает взаимодействие с той или иной технической системой, но и создает определенные эмоциональные — как положительные, так и отрицательные (от установления эмоциональной привязанности до раздражения и агрессии) — связи с ней. С нами «разговаривают» банкоматы, голосовые помощники, чат-боты с так называемым генеративным искусственным интеллектом и так далее. Поскольку в современном мире мы осуществляем большое количество взаимодействий с другими людьми не напрямую, а посредством различных технических интерфейсов, мы привыкаем к тому, что собеседник может быть лишь набором знаков на экране, что способствует легкости, с которой мы можем приравнять «искусственного» собеседника к реальному, поскольку наши способы взаимодействия и с тем, и с другим не различаются. Так мы начинаем общаться в ответ с интеллектуальными системами, совершая таким образом акт интерсубъективного признания в них равного нам Другого (например, при попытке дозвониться в банк я ругаюсь на программу виртуального помощника, требуя пригласить оператора, почти так же, как я бы ругалась с плохо говорящим на моем языке служащим в магазине, требуя позвать администратора).

Мы поддаемся на имитацию и своим решением наделяем искусственную интеллектуальную систему статусом субъекта. К примеру, мы задаем вопрос чат-боту и говорим: смотрите, что он мне *ответил*. Мы можем смеяться над «глупым» GPT, который сочиняет бестолковый или ошибочный ответ на поставленный вопрос, так же, как смеялись бы, к примеру, над несмышленым ребенком, который неточно формулирует какие-то мысли, а когда нейросеть продолжает настаивать на своей ошибочной информации, мы говорим, что она галлюцинирует (Маджумдер, Бегунова, 2024). Искусственная интеллектуальная система, таким образом, начинает выступать как субъект, потому что мы ее воспринимаем как субъекта. Точно так же мы приписываем действиям любимой собаки или кошки значительно больше осмысленности, чем в них есть на самом деле.

Однако представляется, что, соглашаясь принимать имитацию за реальность, мы смешиваем понятие субъекта и понятие агента (актора, деятеля и т. д.). Мы предполагаем, что раз искусственный интеллект может определенным образом действовать на основании некого выбора (ответить на вопрос, используя имеющуюся у него информацию, остановить движущийся автомобиль и т. д.), то он *принимает решения*, т. е. действует

рационально и осознанно. В случае с нейросетями убедительность имитации тем выше, чем лучше нейросеть «обучена», т. е. чем больше ее «высказывания» похожи на высказывания естественного языка в заданной коммуникативной ситуации.

При цифровом взаимодействии с Другим мы обмениваемся в первую очередь высказываниями в текстовом формате, поэтому успешная имитация подобных «осмысленных» высказываний может ввести нас в заблуждение. Мы принимаем связные «высказывания» искусственной интеллектуальной системы за отображение интеллектуальной деятельности, которая скрывается за ними, а вслед за интеллектом (под которым в данном случае понимается способность к решению определенных задач) путем экстраполяции предполагаем наличие других качеств, свойственных субъекту — в частности, воли, ответственности за свои действия и даже субъективной реальности, т. е. уникального субъективного опыта переживания мира. Российский эпистемолог Ю. С. Моркина справедливо делает вывод: «Именно природа человеческого существа как существа смыслополагающего мешает ему четко отличить человекоподобно реагирующие сущности от Другого, подобного себе. Человек склонен заключать по аналогии (и при этом по аналогии с собой), усматривая в Другом также сознающее, чувствующее, размышляющее существо и делая по внешним человекоподобным реакциям вывод о наличии у Другого сознания и его содержаний» (Моркина, 2024: 22). Она показывает в своем анализе, как даже бессмысленный ответ искусственной системы человек может попытаться интерпретировать как нечто осмысленное, например шутку, т. е. человек стремится находить смыслы даже там, где их нет. И потому, называя искусственную интеллектуальную систему субъектом, мы делаем тот же шаг, который совершали древние люди, антропоморфизируя природные стихии и приписывая существование души камням и деревьям.

Но искусственный интеллект не *принимает* решение (принятие решения предполагает возможность говорить и об ответственном осознании возможных последствий данного решения), а *выбирает* «оптимальное» в соответствии с заданным ему алгоритмом решение и не может нести за него ответственность. Если говорить об искусственных интеллектуальных системах в сфере права, то подобные системы не могут обладать правосубъектностью, поскольку они не могут быть как дееспособными, так и деликтоспособными, так как они не в состоянии понимать сущность правовых норм и осознавать последствия своих действий и не способны нести за них юридическую ответственность.

Когда мы пытаемся называть искусственную интеллектуальную систему субъектом — это означает, что мы пытаемся снять ответственность за принятие решений с человека — причем речь идет в большей мере не о создателе данной искусственной интеллектуальной системы, а о том, кто ее использует. Излишнее доверие к деятельности искусственного интеллекта (а создание так называемого доверенного искусственного интеллекта — серьезная и сложная задача), переложение на искусственные интеллектуальные системы значимых вопросов — это, по сути, новая форма проявления технократии. Главная ответственность в данном случае должна лежать на плечах тех субъектов, которые принимают решения о том, в каких сферах мы будем использовать искусственные интеллектуальные системы и в какой мере опираться на них.

Доверие к искусственным интеллектуальным системам, как и к техническим решениям в целом, основывается на той же презумпции, на которой основывается известная английская идиома *camera never lies* («камера не лжет»). Эта идиома намекает на то, что если человеческий взгляд может быть

субъективным, невнимательным и т. д., то объектив фотоаппарата схватывает картину точно и беспристрастно и потому заслуживает доверия. Однако мы знаем, что еще до появления цифровых «дипфейков» или многочисленных инструментов для редактирования цифровых изображений даже особым способом избранный ракурс фото мог создать иллюзию или ввести в заблуждение. Тем не менее представление о том, что человек может ошибиться и регулярно это делает, а машина — никогда, занимает устойчивое место в человеческом сознании. Такая же презумпция и обеспечивает доверие искусственному интеллекту — мы полагаем, раз он создан для решения конкретных задач, то в их пределах он работает точнее и эффективнее, чем человек, не отвлекаясь ни на что лишнее, не забывая детали и т. д. Следствием этого на бытовом уровне становится возникновение такого типа поведения, которое Т. Г. Лешкевич, вслед за известным итальянским философом информации Л. Флориди, называет прокси-культурой (Floridi, 2015). В рамках прокси-культуры различные цифровые приложение становятся «посредниками» (proxy) в нашей деятельности, которым мы доверяем, к примеру, построение маршрута или выбор ресторана с хорошим рейтингом (Лешкевич, 2023). Это доверие может в одних ситуациях облегчать нам жизнь, освобождая от необходимости решения рутинных задач, но в других ситуациях — приводить к неожиданным негативным результатам (например, ориентируясь на карту автомобильных пробок, мы можем выбрать неоптимальный маршрут). Что касается объективности, которую часто называют среди преимуществ искусственных интеллектуальных систем (по сравнению с субъективным взглядом человека), то, как показывают исследования, алгоритмы, лежащие в основе различных искусственных интеллектуальных систем, могут определенным образом сохранять в себе ценности, предубеждения и предрассудки своих создателей — таким образом, их объективность оказывается под сомнением (Nissenbaum, 2001).

Заключение

Следует понимать, что искусственный интеллект во всех его вариациях — это технология, а любая технология является инструментом в руках человека, созданным им самим, и за то, как этот инструмент применяется, отвечает только человек. Результаты деятельности искусственного интеллекта — к примеру, постановка медицинского диагноза, определение степени наказания за правонарушение, обработка данных и генерация текста, обобщающего эти данные, — должны проверяться, уточняться и утверждаться (или отклоняться) человеком как последней инстанцией, поскольку только человек может являться субъектом в полном смысле этого слова. Искусственный интеллект же обладает ровно той степенью свободы, власти и субъектности, которую мы сами ему приписываем.

Список литературы и источников / References

Лекторский В. А., Алексеева Е. А., Емельянова Н. Н., Катунин А. В., Меркулова И. Г., Пирожкова С. В., Труфанова Е. О., Щедрина И. О., Яковлева А. Ф. «Искусственный интеллект в исследованиях сознания и общественной жизни (к 70-летию статьи А. Тьюринга „Вычислительные машины и разум“) (материалы круглого стола)». *Философия науки и техники* 27.1 (2022): 5—33. https://doi.org/10.21146/2413-9084-2022-27-1-5-33. EDN: URJTOM.

Lektorsky V. A., Alekseeva E. A., Emelyanova N. N., Katunin A. V., Merkulova I. G., Pirozhkova S. V., Trufanova E. O., Shchedrina I. O., Yakovleva A. F. “Artificial Intelligence in the Research of Consciousness and in Social Life (in Honor of 70-Years Anniversary of A. Turing’s Paper ‘Computing Machinery and Intelligence’ (Papers of the ‘Round Table’)”. *Filosofiya nayki i tekhniki = Philosophy of Science and Technology* 27.1 (2022): 5—33. (In Russian). https://doi.org/10.21146/2413-9084-2022-27-1-5-33

Лекторский В. А., Васильев С. Н., Макаров В. Л., Хабриева Т. Я., Кокошин А. А., Ушаков Д. В., Валуева Е. А. и др. *Человек и системы искусственного интеллекта*. СПб.: Юридический центр, 2022. 328 с. EDN: XSBHKY.

Lektorsky V. A., Vasil’yev S. N., Makarov V. L., Khabrieva T. Ya., Kokoshin A. A., Ushakov D. V., Valuyeva E. A. et al. *Human and the Artificial Intelligence Systems.* St. Petersburg: Yuridicheskiy tsentr, 2022. 328 p. (In Russian).

Лешкевич Т. Г. «Парадокс доверия к искусственному интеллекту и его обоснование». *Философия науки и техники* 28.1 (2023): 34—47. https://doi.org/10.21146/2413-9084-2023-28-1-34-47. EDN: IGXMAW.

Leshkevich T. G. “The Paradox of Trust in Artificial Intelligence and Its Rationale”. *Filosofiya nayki i tekhniki = Philosophy of Science and Technology* 28.1 (2023): 34—47. (In Russian). https://doi.org/10.21146/2413-9084-2023-28-1-34-47

Лешкевич Т. Г. «Проблема субъектности нейросетей: humans и non-humans». *Философия науки и техники* 29.2 (2024): 125—135. https://doi.org/10.21146/2413-9084-2024-29-2-125-135. EDN: BPCXPE.

Leshkevich T. G. “The Problem of Subjectivity of Neural Networks: Humans and Non-Humans”. *Filosofiya nayki i tekhniki = Philosophy of Science and Technology* 29.2 (2024): 125—135. (In Russian). https://doi.org/10.21146/2413-9084-2024-29-2-125-135

Маджумдер М. Ш., Бегунова Д. Д. «Причины искажения контента: анализ и классификация галлюцинаций в больших языковых моделях GPT». *Искусственный интеллект и принятие решений* 3 (2024): 32—41. https://doi.org/10.14357/20718594240303. EDN: KNHVPP.

Madzhumder M. Sh., Begunova D. D. “Causes of Content Distortion: Analysis and Classification of Hallucinations in Large GPT Language Models”. *Iskusstvenniy intellekt i prinyatie resheniy = Artificial Intelligence and Decision Making* 3 (2024): 32—41. (In Russian). https://doi.org/10.14357/20718594240303

Моркина Ю. С. «Сознание как антиномия (антиномичность понятия „сознание“ и философия искусственного интеллекта)». *Философия науки и техники* 29.1 (2024): 20—33. https://doi.org/10.21146/2413-9084-2024-29-1-20-33. EDN: LCVKZF.

Morkina Ju. S. “Consciousness as an Antinomy (Antinomy of the Concept of Consciousness and the Philosophy of Artificial Intelligence)”. *Filosofiya nayki i tekhniki = Philosophy of Science and Technology* 29.1 (2024): 20—33. (In Russian). https://doi.org/10.21146/2413-9084-2024-29-1-20-33

Труфанова Е. О. «Субъект: вызовы цифровизации». *Galactica Media: Journal of Media Studies* 6.4 (2024): 215—234. https://doi.org/10.46539/gmd.v6i4.525. EDN: WMSJIU.

Trufanova E. O. “The Subject: Challenges of the Digital”. *Galactica Media: Journal of Media Studies* 6.4 (2024): 215—234. (In Russian). https://doi.org/10.46539/gmd.v6i4.525

Хабриева Т. Я. «Правовые проблемы идентификации искусственного интеллекта». *Вестник Российской академии наук* 94.7 (2024): 609—622. https://doi.org/10.31857/S0869587324070015. EDN: FMXVUD.

Khabrieva T. Ya. “Legal Issues of the Artificial Intelligence Identification”. *Vestnik Rossijskoj akademii nauk* 94.7 (2024): 609—622. (In Russian). https://doi.org/10.31857/S0869587324070015

Ястребов О. А. «Правосубъектность электронного лица: теоретико-методологические подходы». *Труды Института государства и права РАН* 13.2 (2018): 36—55. EDN: XSLRRJ.
Yastrebov O. A. “The Legal Capacity of Electronic Persons: Theoretical and Methodological Approaches”. *Trudy Instituta gosudarstva i prava RAN = Proceedings of the Institute of State and Law* 13.2 (2018): 36—55. (In Russian).

Floridi L. “A Proxy Culture”. *Philosophy & Technology* 28 (2015): 487—490. https://doi.org/10.1007/s13347-015-0209-8

Nissenbaum H. “How Computer Systems Embody Values”. *Computer* 34.3 (2001): 118—120. https://doi.org/10.1109/2.910905

Информация об авторе

Труфанова Елена Олеговна — доктор философских наук, доцент, ведущий научный сотрудник Института философии РАН (Россия, 109240, Москва, Гончарная ул., д. 12, стр. 1), *iph@etrufanova.ru*,
ORCID: 0000-0002-2215-1040.

Information about the author

Elena O. Trufanova — Dr. Sci. (Philos.), Assoc. Prof., Leading Research Fellow, Institute of Philosophy of the Russian Academy of Sciences (Russia, 109240, Moscow, Goncharnaya st., 12, bld. 1), *iph@etrufanova.ru*,
ORCID: 0000-0002-2215-1040.

Статья поступила в редакцию 06.02.2025, одобрена после рецензирования 06.03.2025.
The article was submitted 06.02.2025, approved after reviewing 06.03.2025.