

УДК 004.89, 347.78

DOI: 10.32326/2618-9267-2024-7-1-39-47

ПОИСК КРИТЕРИЕВ РАЗУМНОСТИ И ТВОРЧЕСТВА: ПУТЬ ПРОДУКТИВНЫХ ИССЛЕДОВАНИЙ

Пак Галина Станиславовна – доктор философских наук, профессор кафедры философии, Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского. Российская Федерация, 603022, г. Нижний Новгород, просп. Гагарина, д. 23; e-mail: pak@phys.unn.ru; ORCID: 0000-0002-7305-4371

В преддверии дискуссии о разумности и творческом начале искусственного интеллекта (ИИ) следует обратить внимание, что ответ «да» или «нет» зависит от принадлежности исследователя к одной из двух культур – естественно-научной или гуманитарной, хотя однозначной зависимости нет. Против признания разумности и способности к творчеству ИИ, как правило, выступают учёные социально-гуманитарного знания, в то время как сторонниками этой идеи становятся представители естественно-научной интеллигенции. Типичным примером является статья А.Н. Фатенкова «Натурный ум в контратаке на искусственный интеллект (полемический отклик на “Эволюцию разума” Рэя Курцвейла)». В тексте предлагаемой статьи рассматриваются кейсы, которые могут быть интерпретированы в качестве успешного прохождения теста Лавлейс, выступающего в качестве альтернативы тесту Тьюринга. Наличие творческих способностей выступает как условие признания за машинным интеллектом разумности. Путь поиска критериев творчества в самой деятельности ведёт к отрицанию творческого начала у ИИ, в то время как применение критериев творчества к результату деятельности даёт возможность сделать вывод, что ИИ обладает творческой способностью.

Ключевые слова: гуманитарная культура, естественно-научная культура, искусственный интеллект, иррациональное, творчество, тест Лавлейс

Цитирование: Пак Г.С. Поиск критериев разумности и творчества: путь продуктивных исследований // Цифровой учёный: лаборатория философа. 2024. Т. 7. №1. С. 39-47. DOI: 10.32326/2618-9267-2024-7-1-39-47

Рукопись получена: 21 декабря 2023

Пересмотрена: 19 марта 2024

Принята: 20 марта 2024

SEARCH FOR CRITERIA OF REASONABLENESS AND CREATIVITY: THE PATH OF PRODUCTIVE RESEARCH

Galina S. Pak – DSc PhD in Philosophy, Professor, National Research Lobachevsky State

Before further discussing the reasonableness and creativity of artificial intelligence (AI), it should be noted that the answer “yes” or “no” depends on

University. 23 Gagarin Ave.,
Nizhny Novgorod 603022,
Russian Federation;
e-mail: pak@phys.unn.ru;
ORCID: 0000-0002-7305-4371

the researcher's belonging to one of the two cultures – natural-scientific or humanitarian, although there is no unambiguous dependence. As a rule, scientists of social-humanitarian knowledge are against the recognition of reasonableness and the ability to create in AI, while the supporters are representatives of natural-scientific intelligentsia. A typical example is A.N. Fatenkov's article "Natural Intelligence in a Counterattack against Artificial Intelligence (a Polemical Response to "How to Create a Mind" by Ray Kurzweil)". The text discusses cases that can be interpreted as successful completion of the Lovelace test, which acts as an alternative to the Turing test. The presence of creativity acts as a condition for machine intelligence to be recognized as intelligent. The way of searching for the criteria of creativity in the activity itself leads to the denial of creativity in AI, while the application of the criteria of creativity to the result of the activity makes it possible to conclude that AI has creative ability

Keywords: humanitarian culture, natural science culture, artificial intelligence, irrational, creativity, Lovelace test

How to cite: Pak, G.S. (2024). Search for criteria of reasonableness and creativity: the path of productive research. *The Digital Scholar: Philosopher's Lab*, 7 (1): 39-47. DOI: 10.32326/2618-9267-2024-7-1-39-47 (In Russian).

Received: 21 December 2023

Revised: 19 March 2024

Accepted: 20 March 2024

Вопрос, поставленный Аланом Тьюрингом в середине прошлого века, «может ли машина мыслить?» до сих пор не имеет однозначного ответа. При анализе современного состояния дискуссии об искусственном интеллекте (ИИ) необходимо иметь в виду две сложившиеся альтернативные тенденции в обсуждении проблемы. Исследователей, признающих возможность мыслительных способностей за нейросетью, для краткости будем называть сторонниками ИИ, а отрицающих наличие этих способностей – противниками ИИ. В какой-то степени противостояние этих подходов возрождает спор физиков и лириков. Начало этому спору длиною в четверть века положила лекция Чарльза Перси Сноу в Кембриджском университете в 1959 г. «Две культуры и научная революция». Поскольку лекция имела сенсационный успех, она вскоре была опубликована. Сноу не просто заявил о разрыве между представителями науки и искусства, вызванном научной революцией XX в., но и обозначил наличие «двух культур» – научной и художественной (в нашей стране их чаще обозначают как естественно-научная и гуманитарная культуры). Учёных и представителей художественной интеллигенции разделяет стена непонимания, что может

доходить до полного неприятия и вражды, поскольку обе стороны имеют искажённое представление друг о друге. Они по-разному смотрят на вещи и найти общий язык для них весьма проблематично (Сноу, 1973, с. 20). В дискуссии об ИИ противниками признания его разумности, как правило, выступают учёные социально-гуманитарного знания, в то время как сторонниками этой идеи – естественно-научная интеллигенция. Типичным примером является статья А.Н. Фатенкова «Натурный ум в контратаке на искусственный интеллект (полемический отклик на “Эволюцию разума” Рэя Курцвейла)», в которой философ выступает против позиции учёного из области информатики. Оценка прогнозов Рэя Курцвейла в данном случае зависит от принадлежности авторов к одной из двух культур. Основатель *Microsoft* Билл Гейтс отмечает: «Рэй Курцвейл даёт самые точные прогнозы о будущем искусственного интеллекта из всех, что мне известны» (ПроБизнес, 2018, web).

В настоящее время способность ИИ решать задачи на уровне рассудочной деятельности ни у кого не вызывает сомнений. Сегодня вопрос ставится иначе, конкретнее – обладает ли ИИ творческим началом? Для ответа на поставленный вопрос Е.В. Масланов обращается к анализу больших языковых моделей, способных создавать тексты (Масланов, 2024), установление авторства которых является проблемой (Коваль, Ушкин, 2023). Кто их создатель – человек или машина? Текст рассматривается как наиболее вероятная последовательность сочетания минимальных единиц данных – токенов. Утверждается, что релевантность текста, его смысл и значение способен установить лишь сам человек. В этом случае логика работы языковых моделей аналогична логике эксперимента, предложенного Джоном Сёрлом и получившего название «Китайская комната». Осмысленный текст, как и разумный ответ, является результатом использования наиболее вероятных сочетаний. В этом случае о творчестве не может быть и речи. Для оперирования токенами, создающими осмысленный текст для человека (а не для машины), машину долго обучают. Это нельзя рассматривать как проявление недостатка ИИ, поскольку знания человека не являются врождёнными, а приобретаются в ходе социализации и обучения, проходящих через всю жизнь.

На основе анализа больших языковых моделей Е.В. Масланов приходит к выводу об отсутствии творческого начала у ИИ. В развернувшейся дискуссии интереснее случаи освоения ИИ компьютерных игр в результате обучения с подкреплением. В частности, Евгений Валерьевич приводит пример того, как в компании *OpenAI*, разработавшей *ChatGPT*, обучали ИИ играть в компьютерную игру *Coast Runners*. Создатели игры видели её смысл и цель в том, чтобы как можно быстрее добраться до финиша. Однако дополнительно в игре была предусмотрена механика зарабатывания баллов за счёт поражения целей по пути прохождения трассы. Игра в какой-то степени похожа на биатлон, сочетающий лыжную гонку и стрельбу по мишеням. Люди стремились первыми

достичь финиша, в то время как игровой агент, управляемый ИИ, пошёл по пути зарабатывания баллов. В итоге результат игрового агента оказался в среднем на 20 % выше, чем у игроков-людей. Обучение с вознаграждением объясняет то, как ИИ смог принять решение, не предусмотренное человеком. Рассмотренный сюжет можно интерпретировать как наличие творческого начала ИИ, и это не отменяется тем фактом, что машина реализовывала собственные цели, а не поставленные перед ней людьми.

Сюжет, приведённый Е.В. Маслановым, как и кейс, рассмотренный А.В. Сахаровой, о победе ИИ в игре в го (Сахарова, 2024), дают основание утверждать, что машинный интеллект оказался способным пройти тест Лавлейс. Тест Лавлейс рассматривается как альтернатива тесту Тьюринга. Он разработан совместными усилиями Сельмера Брингсйорда, Пола Белло и Дэвида Ферруччи, которые предложили назвать его в честь леди Ады Лавлейс – профессионального учёного, участвующего в разработке первой вычислительной машины (Bringsjord, 2003). Лавлейс считала определяющей характеристикой разумности способность создавать новое, подчёркивая, что новизна и оригинальность не должны быть следствием ошибки или случайности. Если кратко, то суть теста Лавлейс можно представить следующим образом. Есть субъект, в распоряжении которого находится полная информация об архитектуре и принципах работы ИИ; данный субъект способен в любой момент рассмотреть и проанализировать любой элемент ИИ, так как заложил в работу машины чёткий алгоритм действий. Если этот всезнающий субъект не способен объяснить, как возник неожиданный (и неслучайный) результат, для получения которого ИИ выходит за пределы установленного алгоритма действий, то существует основание признать наличие творческого начала у машинного интеллекта. Однако есть одно важное условие для получения ответа, связанное с тем, какой смысл мы вкладываем в понятие «творчество». Согласимся с тем, что исследователи далеки от единого понимания творческой деятельности и это понятие является проблемным.

Ряд авторов полагает, что демаркация между творчеством и нетворчеством является результатом субъективной оценки (например, З. Гизелин), и мы просто теряем время в поиске объективных критериев. И. Ставровский в качестве признаков творчества предлагает рассматривать следующие параметры: оригинальность, необусловленность, спонтанность. Оригинальность выполняет функцию выделения из множества тождественных по своей сути элементов, в то время как новизна направлена на противопоставление старого и нового, не являясь имманентным признаком оригинальности. В связи с этим при определении творчества представляется более адекватным говорить о новизне, а не об оригинальности. Хотя следует признать, что невозможна как абсолютная оригинальность, так и абсолютная новизна. А. Тьюринг обращал внимание на это в своём ответе леди Лавлейс (Тьюринг, 2018,

с. 79-85). Необусловленность в качестве признака творчества вызывает серьёзные возражения. Является ли безусловностью отсутствие цели творческой деятельности или считается, что творчество не может совмещаться с меркантильными мотивами, с деятельностью ради денег, славы? Потолок Сикстинской капеллы в Ватикане представляет собой цикл фресок Микеланджело 1508-1512 гг., которые он сделал в качестве уплаты долга, то есть ради денег, но это не мешает нам признать их шедевром искусства Высокого Возрождения. Таким образом, творчество не является исключительно деятельностью, основанной на непреодолимом внутреннем желании. В этом случае ИИ мог бы продемонстрировать большую безусловность, чем кто-либо из когда-либо живших на Земле людей. Выделение признака спонтанности не служит прояснению понятия «творчество». Спонтанность как самопроизвольность, как характеристика процессов, обусловленных внутренними, а не внешними причинами, совпадает в известном смысле с безусловностью. В спонтанности невозможно выявить какие-либо методы и техники, способные гарантировать успех в создании нового. Спонтанность и самообусловленность отсылают нас к интуитивному представлению о творчестве. Ряд авторов так и делают, связывая творчество с всплеском интуиции (Ильин, с. 22) и считая, что употребление понятия «логика открытий» лишено смысла (Ильин, 2008, с. 28). Некая таинственность, непрояснённость окружают понятие творческой деятельности.

Описанный подход в понимании творчества лежит в русле идей А.Н. Фатенкова. Для проведения демаркации между ИИ и человеческим мышлением он обращается к аналогии гомогенного и гетерогенного пространства. Гетерогенное пространство, как ментальное пространство человека, наполнено здоровой иррациональностью (Фатенков, 2022, с. 177). Иррациональна интуиция, иррациональна глупость. Традиционно сфера иррационального связывается с чувствами, эмоциями. Эразм Роттердамский считал глупость неуничтожимой, поскольку она порождается человеческими чувствами. Парадокс налицо. Творчество является главным признаком разумности, но необходимой составляющей творчества является иррациональная компонента. Дialeктик Гегель предупреждал, что разумность и страсти не исключают друг друга, человеческий разум страстен, а страсти разумны. Из логики А.Н. Фатенкова следует, что основное преимущество человека по сравнению с нейросетью состоит в способности делать глупости, а нейросеть лишена этой удивительной черты. Итак, если я скажу очередную глупость, то благополучно пройду тест на принадлежность к человеческому роду, которому априори приписывается способность к творчеству. Но так ли это на самом деле? Не совершила ли глупость или не подшутила ли нейросеть *ЯндексGPT*, когда написала биографию человека, в которой соответствовал реальности только год рождения? Он математик по образованию, а сеть его представила журналистом. Волшебных превращений

много, и в итоге нейросеть превратила профессора в сенатора. Е.В. Масланов объяснил это ограниченностью программы, смысл которой состоит в написании красивого и связного текста, а не в следовании фактам. Считаю, что это не единственная возможная интерпретация. Глупость программы – результат человеческой глупости. Можно ли запрограммировать глупость? Если да, то имеет место иррациональный компонент машинного интеллекта и нет оснований отрицать у него наличие творческих способностей. Поставим вопрос иначе – а способен ли ИИ понимать смысл глупости? В качестве примера возьмём фигуры алогизма, создающие комический и драматический эффект.

Ехала деревня
Мимо мужика.
Вдруг из подворотни
Лают ворота.
Тпр! – сказала лошадь,
А мужик заржал.
Лошадь пошла в гости,
А мужик стоял...
(Квятковский, 1966, с. 21).

По мнению специалистов, здесь происходит нарушение онтологической нормы (Сковородников, 2004, с. 40). С грамматической точки зрения слова лишены собственной синтаксической позиции, комический эффект достигается путём деформации онтологической связей реального мира (Садовников, 2016). Теперь приведём мнение ИИ. Я обратилась к *ChatGPT* с вопросом, в чём он видит смысл приведённого текста, и получила следующий ответ. «Этот текст описывает смешную ситуацию, когда лошадь проезжает мимо мужика, и вдруг ворота начинают лаять. Лошадь делает звук “Тпр!” и мужик начинает смеяться. Лошадь идёт в гости, а мужик остаётся стоять и удивляться происходящему. Возможно, смысл заключается в том, что даже обычные ситуации могут быть забавными и неожиданными». Нечеловеческий смысл сразу бросается в глаза, поскольку нейросеть оценивает ситуацию как обычную, в то время как для человека она просто невозможна. К аналогичному выводу приходит Е.В. Масланов, когда пишет, что и большая языковая модель, и играющий в игру агент ограничены цифровой реальностью и не способны её соотнести с внецифровой реальностью, в которой вещи могут «дать сдачи». В этой связи следует заметить, что небылицы-перевёртыши чаще всего используются в детской литературе: их цель укрепить знание ребёнка о правильном ходе вещей, заставляя смеяться над абсурдностью ситуации. Важно подчеркнуть, что понимание алогизмов не дано человеку априори, а является результатом обучения. Проблема заключается в том, способен ли ИИ научиться различать реальность и виртуальную реальность.

И снова мы пошли по пути поиска того, что не может ИИ. Непродуктивность такого подхода раскрыта А.В. Сахаровой (Сахаро-

ва, 2024). Я солидарна с ней в том, что попытка искать критерии творчества в процессе, в способе создания произведения с самого начала обречена на неудачу. Необходимы критерии творчества, применимые к результату творческой деятельности. Если результат аналогичен человеческому, то нам не стоит придумывать новые критерии мышления и творчества исходя из того, чего не может сделать ИИ, и априори приписывать такие способности человеку. Достоинством работы А.В. Сахаровой является как раз исследование и обоснование критериев, применимых к результату: новизна, неожиданность, значимость. Следует признать, что ИИ развивается и благополучно справляется с задачами, которые с нашей точки зрения казались не разрешимыми для него. Да, ИИ создан человеком. Человек искал братьев по разуму на других планетах, но в итоге создал разум на Земле. Из этого не следует, что машинный интеллект не достигнет уровня человеческого интеллекта и не превзойдёт своего творца. Дети тоже созданы человеком, однако они раз за разом превосходят своих родителей.

Заключение

Пора посмотреть правде в глаза. Развитие науки есть процесс развенчивания привилегированного положения человека в мироздании. Не Солнце вращается вокруг Земли, а Земля вокруг Солнца. Человек создан не по образу и подобию Бога, а стал результатом биологической эволюции. Не разум, а бессознательное является исходной душевной субстанцией. Признание за ИИ творческого начала – ещё один шаг на этом пути. «Да» или «нет» в ответе на вопрос о признании творческого начала ИИ определяется мировоззрением учёного, которое направляет весь ход научного исследования.

Информация о конфликте интересов

Автор заявляет об отсутствии конфликта интересов

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Список литературы

- Ильин, 2008 – *Ильин Е.П.* Психология творчества, креативности, одарённости. СПб.: Питер, 2008. 433 с.
- Квятковский, 1966 – *Квятковский А.П.* Поэтический словарь / Ред. И. Роднянская. М.: Советская энциклопедия, 1966. 376 с.
- Масланов, 2024 – *Масланов Е.В.* Обладает ли искусственный интеллект творческим началом? // Цифровой учёный: лаборатория философа. 2024. Т. 1. № 1. С. 6-14.

ПроБизнес, 2018 – Пробизнес: похоже, мы живём в самое интересное время. Каким видит будущее Рэй Курцвейл, прогнозы которого сбываются // Пробизнес. 27.02.2018. URL: <https://probusiness.io/ideas/4312-pokhozhe-my-zhivem-v-samoe-interesnoe-vremya-kakim-vidit-budushchee-rey-kurzveyl-prognozy-kotorogo-sbyvayutsya.html> (дата обращения: 26.11.2023).

Садовников, 2016 – Садовников С.А. Структурно-семантические функции фигур нарочитого алогизма в поэтических текстах // Вестник Нижегородского университета им. Н.И. Лобачевского. 2016. № 6. С. 212-214.

Сахарова, 2024 – Сахарова А.В. Креативность в науке и «креативность» искусственного интеллекта. Вопрос о настоящей креативности // Цифровой учёный: лаборатория философа. 2024. Т. 1. № 1. С. 22-31.

Сквородников, 2004 – Сквородников А.П. Алогизм как риторический приём // Русская речь. 2004. № 1. С. 35-49.

Сноу, 1973 – Сноу Ч.П. Две культуры: сборник публицистических работ. М.: Прогресс, 1973. 144 с.

Ставровский, 2022 – Ставровский И.К. Искусственный интеллект и творчество // Весці Нацыянальнай акадэміі навук Беларусі. Серыя гуманітарных навук. 2022. Т. 67. № 1. С. 15-20. <https://doi.org/10.29235/2524-2369-2022-67-1-15-20>

Тьюринг, 2018 – Тьюринг А. Вычислительные машины и разум / Пер. К. Королёв. М.: АСТ, 2018. 128 с.

Фатенков, 2022 – Фатенков А.Н. Натурный ум в контратаке на искусственный интеллект (полемический отклик на «Эволюцию разума» Рэя Курцвейла) // Философский журнал. 2022. Т. 15. № 3. С. 172-183. <https://doi.org/10.21146/2072-0726-2022-15-3-172-183>

Bringsjord, 2003 – Bringsjord S. Creativity, the Turing test, and the (better) Lovelace test // Bringsjord S., Bello P., Ferrucci D. A. The Turing Test: the Elusive Standard of Artificial Intelligence. Dordrecht: Springer, 2003. Pp. 215-239. https://doi.org/10.1007/978-94-010-0105-2_12

References

Bringsjord, S. (2003). Creativity, the Turing test, and the (better) Lovelace test. In S. Bringsjord, P. Bello, D. A. Ferrucci. *The Turing Test: the Elusive Standard of Artificial Intelligence* (pp. 215-239). Springer. https://doi.org/10.1007/978-94-010-0105-2_12

Fatenkov, A.N. (2022). Natural mind in counterattack on artificial intelligence (polemical response to “How to Create a Mind” by Ray Kurzweil). *Philosophical Journal*, 15 (3), 172-183. <https://doi.org/10.21146/2072-0726-2022-15-3-172-183> (In Russian)

Ilyin, E.P. (2008). *Psychology of Creativity, Creativeness, Giftedness*. Piter Publ. (In Russian)

Kvyatkovsky, A.P. (1966). *Poetic Dictionary*. Soviet Encyclopedias Publ. (In Russian)

Maslanov, E.V. (2024). Is artificial intelligence creative? *The Digital Scholar: Philosopher's Lab*, 1 (1), 6-14. (In Russian)

Sadovnikov, S.A. (2016). Structural and semantic functions of the figures of deliberate alogism in poetic texts. *Bulletin of Lobachevsky University of Nizhny Novgorod*, 6, 212-214. (In Russian)

Sakharova, A.V. (2024). Creativity in science and the “creativity” of artificial intelligence. The question of real creativity. *The Digital Scholar: Philosopher's Lab*, 1 (1), 22-31 (In Russian)

Skovorodnikov, A.P. (2004). Alogism as a rhetorical device. *Russian Speech*, 1, 35-49. (In Russian)

Snow, C. P. (1973). *Two Cultures: a Collection of Publicist Works*. Progress Publ. (In Russian)

Stavrovsky, I.K. (2022). Artificial intelligence and creativity. *Bulletin of Belarus National Academy of Sciences. Humanities Series*, 67 (1), 15-20. <https://doi.org/10.29235/2524-2369-2022-67-1-15-20> (In Russian)

Turing, A. (2018). *Computing Machines and the Mind*. AST Publ. (In Russian)

ProBusiness: We seem to be living in the most interesting times. How Ray Kurzweil sees the future, whose predictions are coming true (2018). *ProBusiness*, February 27. Retrieved November 26, 2023 from <https://probusiness.io/ideas/4312-pokhozhe-y-zhivem-v-samoe-interesnoe-vremya-kakim-vidit-budushchee-rey-kurcveyl-prognozy-kotorogo-sbyvayutsya.html> (In Russian)