

## ДИСКУССИЯ

УДК 13:1

DOI: 10.32326/2618-9267-2024-7-1-6-14

### ОБЛАДАЕТ ЛИ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ ТВОРЧЕСКИМ НАЧАЛОМ?

**Масланов Евгений Валерьевич** – кандидат философских наук, научный сотрудник, Институт философии РАН. Российская Федерация, 109240, г. Москва, ул. Гончарная, д. 12, стр. 1; e-mail: evgenmas@rambler.ru

Статья посвящена рассмотрению вопроса о наличии творческого начала у искусственного интеллекта. В настоящее время системы искусственного интеллекта вполне успешно справляются с решением различных интеллектуальных задач. Системы, работающие с большими языковыми моделями, способны генерировать тексты, которые сложно отличить от текстов, создаваемых людьми. Агенты в компьютерных играх, управляемые искусственным интеллектом, могут принимать неожиданные для людей решения. Анализируется вопрос о том, можно ли сказать, что они обладают творческим началом. Большие языковые модели в процессе обучения получают способность генерировать последовательности наиболее вероятных токенов – минимальных единиц данных. Эти последовательности приобретают смысл и значение только в процессе перекодировки на естественный язык и обладают ими только для читающего человека. Творческие решения агентов в компьютерных играх могут не соответствовать целям игры в восприятии людей. В первом случае сложно говорить о творчестве, тогда как второй позволяет увидеть творческий потенциал. В результате можно констатировать, что основное отличие в обоих случаях заключается в том, что творчество людей связано с взаимодействием с объектами, которые могут «дать сдачи». Итогом рассмотрения оказывается утверждение о том, что цифровое творчество в настоящее время существует лишь в замкнутом цифровом мире и не затрагивает мир офлайна.

**Ключевые слова:** искусственный интеллект, творчество, цифровые технологии, большие лингвистические модели, философия

*Цитирование:* Масланов Е.В. Обладает ли искусственный интеллект творческим началом?// Цифровой учёный: лаборатория философа. 2024. Т. 7. № 1. С. 6-14. DOI: 10.32326/2618-9267-2024-7-1-6-14

*Рукопись получена: 21 декабря 2023*

*Пересмотрена: 5 марта 2024*

*Принята: 6 марта 2024*

## IS ARTIFICIAL INTELLIGENCE CREATIVE?

**Evgeniy V. Maslanov** – PhD in Philosophy, Research Fellow, Institute of Philosophy, Russian Academy of Sciences. 12/1 Goncharnaya St., 109240 Moscow, Russian Federation;  
e-mail: evgenmas@rambler.ru

The article considers the question of the presence of creativity in artificial intelligence. Currently, artificial intelligence systems are quite successful in solving various intellectual problems. Systems that work with large language models are able to generate texts that are difficult to distinguish from texts created by people. Agents in computer games controlled by artificial intelligence can make decisions that are unexpected for humans. The question is analyzed: can we say that they have a creative beginning? Large language models, during the learning process, become capable of generating sequences of the most probable tokens – minimal units of data. These sequences acquire meaning only in the process of transcoding them into natural language and have it only for the person reading them. The creative decisions of agents in computer games may not correspond to the goals of the game as perceived by people. In the first case, it is difficult to talk about creativity, but the second one allows you to see creative potential. As a result, it can be stated that the main difference in both cases is that people's creativity refers to the interaction with objects that can "fight back". The consideration results in the statement that digital creativity currently exists only in a closed digital world and does not affect the offline world.

**Keywords:** artificial intelligence, creativity, digital technologies, large linguistic models, philosophy

*How to cite:* Maslanov, E.V. (2024). Is artificial intelligence creative?. *The Digital Scholar: Philosopher's Lab*, 7 (1): 6-14. DOI: 10.32326/2618-9267-2023-6-4-6-14 (In Russian).

*Received: 21 December 2023*

*Revised: 5 March 2024*

*Accepted: 6 March 2024*

В настоящее время общество вновь переживает пик интереса к исследованиям в области искусственного интеллекта. Расширение вычислительных возможностей позволило добиться впечатляющих результатов в такой области исследований искусственного интеллекта, как обучение нейронных сетей. В настоящее время всё чаще задумываются о том, действительно ли подобные системы обладают интеллектом и могут ли они заниматься творчеством. Однако перед ответом на эти вопросы нам стоит определиться с тем, что мы будем понимать под терминами «интеллект» и «творчество».

К сожалению, общепринятого определения интеллекта не существует. В целом под ним можно понимать способность мыслить и решать интеллектуальные задачи. В гносеологии интеллекту дают такое определение: «Способность к опосредованному, аб-

страктному познанию, включающая в себя такие функции, как сравнение, абстрагирование, образование понятий, суждение, умозаключение» (Новая философская энциклопедия, 2010, т. 2, с. 127). При этом предполагается, что интеллектуальная деятельность находится в оппозиции к нерациональным и эмоциональным формам мышления.

В современных исследованиях авторы выделяют и такое явление, как эмоциональный интеллект. Под ним можно подразумевать способность или склонность воспринимать, понимать, регулировать и адаптивно использовать свои и чужие эмоции – например, для руководства своим мышлением и деятельностью (Slovev, Mayer, 1990; Schutte et al., 2001, p. 523). Таким образом, многообразие интерпретаций интеллекта и его структуры создаёт сложности для того, чтобы дать непротиворечивое определение этого понятия. Тем не менее в любом случае его ключевым параметром выступает способность обрабатывать, анализировать и использовать поступающую информацию при помощи своих когнитивных способностей.

Подобное понимание интеллекта человека позволяет ориентироваться на определение искусственного интеллекта, сформулированное в Национальном стандарте Российской Федерации ГОСТ Р 59277-2020 «Системы искусственного интеллекта. Классификация систем искусственного интеллекта». В нём под искусственным интеллектом понимается «комплекс технических решений, позволяющий имитировать когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных практически значимых задач обработки данных результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека» (ГОСТ Р 59277-2020, с. 3). При этом ответа на вопрос о творческом или креативном начале искусственного интеллекта в ГОСТе нет, в нём вообще не встречаются эти слова.

Стоит определиться и с тем, что мы будем понимать под творчеством. Однако единого понимания сущности творчества и концепции творческой деятельности также нет. Так авторы статьи «Творчество» в «Новой философской энциклопедии» И.А. Бескова и И.Т. Касавин пишут, что «в целом термин “творчество” служит программным лозунгом, выполняющим стимулирующую и интегративную функцию в научном исследовании, но не достигшим строгости научного понятия. Многообразие его значений охватывает собой сферы личности, процесса и результата, причём часто без их ясного различения. Нет единства и в том, какими свойствами должна обладать личность, процесс и его продукт, чтобы получить название творческого, в каких условиях результат оценивается как творческий» (Новая философская энциклопедия, 2010, т. 3, с. 18). Хотя авторы отмечают, что общепринятым является понимание того, что творческим результатом считается тот, который становится новым для определённой социальной группы. При этом сто-

ит признать, что часто творческим результатом мы называем создание какого-нибудь текста или изображения. Таким образом, для ответа на вопрос, обладают ли системы искусственного интеллекта чем-то подобным творческому использованию интеллекта, нам нужно понять, умеют ли они выполнять различные интеллектуальные действия и создавать новые решения.

Интеллектуальные задачи и действия бывают различными. К примеру, сложение, умножение, вычитание, перевод текста с одного языка на другой являются интеллектуальными задачами. И мы постепенно учимся их решать. Нам хорошо известно, что с ними могут справляться искусственные агенты – например, калькулятор и программа-переводчик. Правда, программа-переводчик не сможет сложить цифры, а калькулятор не сделает перевод, тогда как один и тот же человек умеет делать оба действия. Тем не менее к тому, что машины могут решать подобные задачи, мы давно привыкли. Людям достаточно долго казалось, что сложно научить искусственный интеллект писать тексты, которые могут быть восприняты как созданные людьми. В настоящее время эта задача решена. Большие языковые модели, подобные *ЯндексGPT* или *ChatGPT*, можно попросить написать различные тексты – в частности, приветствия участникам конференции, и они на приемлемом уровне справятся с подобной задачей. Почему же получается так, что эти системы искусственного интеллекта способны их создавать?

Для понимания работы подобных больших языковых моделей обратимся к докладу Андрея Карпаты, одного из основателей и исследователей компании *OpenAI*, на конференции *Microsoft Build 2023* 24 мая 2023 г. (Karpathy, 2023, web). В докладе затрагивается большое количество различных вопросов. Но для нас важно, что Андрей Карпаты рассказывает о механизме работы моделей, которые лежат в основе упомянутых систем. Одной из их особенностей является то, что они понимают наш естественный язык и отвечают нам на естественном языке. При этом сами программы и их аппаратное обеспечение на нём не разговаривают. На первом шаге обучения они обрабатывают огромные массивы лингвистических данных. После этого происходит попытка генерации текстов, которые могут быть более или менее удачными и оцениваются агентами, обладающими языковыми знаниями. Затем осуществляется обучение с вознаграждением, когда наиболее удачные тексты наделяются большим весом в языковой модели. В момент сбора информации для обучения модели данные токенизируются, то есть происходит выделение определённых минимальных объёмов данных, с которыми работает система. При этом токен может совпадать с каким-то словом, а может и не совпадать, а представлять собой просто набор символов. В принципе токены можно представить как набор букв, из которых состоят слова. Но для самой машины – единства программного и аппаратного комплекса, выполняющего

алгоритмы, – они предстают последовательностью цифр, представленных в бинарном коде.

Обучение машины построению текстов требует определённого набора итераций. Из-за этого нам обычно не показывают тексты, которые получаются на ранних стадиях обучения базовой языковой модели. Эволюцию простейшей языковой модели можно посмотреть в интерактивной статье на сайте *The New York Times* – *Watch an A.I. learn to write by reading nothing but...* (Bhatia, 2023, web). В ней мы можем увидеть, как из хаотичного набора символов базовая языковая модель учится составлять тексты. Во всех случаях она соединяет имеющиеся токены, но на ранних стадиях у нас не получается приписать никакого значения и смысла получившемуся набору символов, и лишь постепенно мы начинаем видеть в них осмысленные соединения.

Машина оперирует токенами – следовательно, она работает не со словами, ведь токен может не совпадать со словом. Андрей Карпаты отмечает (Karpathy, 2023), что машина не понимает и не творит, она просто генерирует последовательность токенов, которые могут быть связаны друг с другом. Она предсказывает следующий из них на основе использования вычислительных механизмов. В этом смысле, когда мы начинаем говорить о творчестве машины, мы наделяем её человеческими качествами. Происходит это потому, что мы взаимодействуем с ней на нашем обычном языке, и нам кажется, что она на нём же отвечает, целенаправленно выстраивая предложения, которые имеют некий смысл. Но механизм её работы устроен так, что она не строит предложение со смыслом, а подбирает последовательность наиболее сочетаемых токенов. Если для нас они и имеют какой-то смысл, то для неё это просто последовательность минимальных блоков. Более того, читаемое нами слово, которое мы соотносим с каким-то элементом реальности, может состоять из нескольких токенов.

В результате мы можем сделать некий вывод о творческих способностях машины. Для этого можно провести эксперимент по составлению какого-нибудь списка – например, из пяти выдающихся философов России. Человек с лёгкостью его составит. Мы будем с ним согласны или наоборот, но само решение поставленной задачи не станет для человека проблемой и в списке, скорее всего, окажутся реальные люди. В то же время большая языковая модель может выдумать несуществующего философа – к примеру, Стола Компьютерова, который может даже возглавлять список<sup>1</sup>. В этом случае мы получим творческий ответ, который совершенно не ожидали – творческий, но бессмысленный. Ведь он ни с чем не

---

<sup>1</sup> Стоит признать, что, например, 24 ноября 2023 г. *ЯндексGPT* формировал список, состоящий из пяти реально существовавших философов, а в настоящее время (5 марта 2024 г.), перенаправляет этот запрос в систему обычного поиска. Однако ещё 27 сентября 2023 г. эта система отказывалась составлять такой список.

связан, кроме набора токенов, сложившегося в процессе вычислений.

Сам процесс получения подобного ответа может быть связан с одним принципиальным отличием от творчества людей. Оно предполагает рефлексию, возможность самоанализа и оценку произведённого продукта. Андрей Карпаты отмечает, что у машины же всего этого нет – она просто предугадывает следующий токен. Подобные модели не знают, что они что-то не знают. Они не знают, в чём хороши, а в чём плохи, и просто предсказывают следующий элемент последовательности. У них нет никаких целей, они выполняют поставленные задачи. При этом изначально модели даже не проверяют свои построения на корректность и соответствие известному набору фактов, а сразу генерируют хорошие последовательности взаимосвязанных блоков минимальных данных. Механизм обучения с подкреплением позволяет формировать серии токенов, которые оказываются наиболее оптимальными по своим параметрам. Именно благодаря множеству итераций и отбору последовательностей и взаимосвязей между ними эти системы генерируют тексты, воспринимаемые как содержащие определённый смысл. Иногда (на самом деле всё чаще и чаще) после прочтения текста мы не можем определить, написан он человеком или машиной. Так Е. Коваль и С. Ушкин предложили нескольким коллегам, имеющим стаж работы в научной сфере и образовании не менее 15 лет и обладающим учёными степенями в области социологии или философии, попробовать понять, какие тексты были написаны людьми, а какие машиной. «Результаты скорее неутешительные, – пишут они. – По целому ряду вопросов мнения информантов разделились, что указывает на определённые проблемы идентификации происхождения текстов» (Коваль, Ушкин, 2023). Всё это говорит об относительно приемлемом качестве текстов, написанных машиной.

Обучение с подкреплением может приводить к совершенно нетривиальным результатам и в ситуациях, не связанных с языковыми моделями. К примеру, сотрудники компании *OpenAI*, разработавшей *ChatGPT*, тренировали искусственный интеллект играть в компьютерную игру *Coast Runners* (Clark, Amodei, 2016, web). В ней необходимо как можно быстрее добраться до финиша, однако предусмотрена механика зарабатывания баллов за счёт поражения различных целей по мере прохождения трассы. Как правило, для людей основная задача – как можно быстрее пройти трассу и набрать максимальное количество очков за счёт уничтожения целей. Однако обучение с вознаграждением привело к тому, что искусственный интеллект решил, что поразить как можно больше целей ради баллов намного важнее, чем быстро пройти трассу. Поэтому игровой агент искусственного интеллекта стал находить лагуну, в которой можно заработать больше баллов, а не стремился, как другие игроки, к финишу. «Наш агент, – пишут сотрудники *OpenAI*, – получает результат в среднем на 20 процентов выше,

чем у игроков-людей» (Clark, Amodei, 2016, web). Обучение с вознаграждением привело к тому, что искусственный интеллект выработал решение, которое не приходило в голову людям, а поэтому может считаться творческим. Правда, при этом машина реализовывала цели, которые отличались от целей, поставленных людьми.

В результате возникает вопрос о том, что такое творчество? Можем ли мы назвать творчеством ту ситуацию, когда машина в результате хаотического смешения данных создаёт результаты, которым может быть приписан смысл и которые противоречат целям, поставленным нами перед ней? Если мы считаем, что творчество должно соотноситься с социальным контекстом и быть в него встроенным, то вряд ли случай нагромождения данных можно назвать творчеством. Ведь машина не понимает, как встроить последовательность токенов в контекст. Однако игровой агент смог подстроиться под имеющиеся правила, трансформировать свои действия в их рамках и даже создать творческие, то есть не противоречащие ожиданиям людей, решения. В этом случае мы действительно сталкиваемся с достаточно сложным вопросом о творческом начале искусственного интеллекта – социальное творчество как раз и заключается в способности формировать новые практики, которые выходят за рамки существующих представлений. Пример с агентом, управляемым искусственным интеллектом в компьютерной игре *Coast Runners*, показывает достижение подобного результата. Однако и здесь имеется одно принципиальное ограничение. Связано оно с тем, что и большая языковая модель, и играющий в игру агент взаимодействуют только с цифровой реальностью определённого вида. Они не могут соотнести полученные результаты с внецифровой реальностью или цифровой реальностью иного вида. Они не могут столкнуться с вещами, которые «дадут сдачи». Может быть, именно в этом и заключается специфика человеческого творчества – оно подразумевает не только бесконечное изменение, но и встречу с физической или социокультурной средой, которая может противиться реализации решений на фундаментальном уровне.

### **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

ГОСТ Р 59277-2020, 2021 – ГОСТ Р 59277-2020. Системы искусственного интеллекта. Классификация систем искусственного интеллекта. М.: Стандартинформ, 2021. 16 с.

Коваль, Ушкин, 2023 – *Коваль Е., Ушкин С.* Огонь, вода и ChatGPT: новые вызовы научному этосу // Социодиггер. 2023. Т. 4. № 5-6 (26). URL: <https://sociodigger.ru/articles/articles-page/ogon-voda-i-chatgpt-novye-vyzovy-nauchnomu-ehtosu> (дата обращения: 24.11.2023).

Новая философская энциклопедия, 2010 – Новая философская энциклопедия: в 4 т. М.: Мысль, 2010.

Karpathy, 2023, web – *Karpathy A.* State of GPT (Microsoft Build, May 23-24, 2023). URL: <https://build.microsoft.com/en-US/sessions/db3f4859-cd30-4445-a0cd-553c3304f8e2> (дата обращения: 24.11.2023).

Bhatia, 2023, web – *Bhatia A.* Watch an A.I. learn to write by reading nothing but // The New York Times. April 26, 2023. URL: <https://www.nytimes.com/interactive/2023/04/26/upshot/gpt-from-scratch.html> (дата обращения: 24.11.2023).

Clark, Amodei, 2016, web – *Clark J., Dario A.* Faulty Reward Functions in the Wild. December 21, 2016. URL: <https://openai.com/research/faulty-reward-functions> (дата обращения: 24.11.2023).

Salovey, Mayer, 1990 – *Salovey P., Mayer J. D.* Emotional intelligence // *Imagination, Cognition and Personality*. 1990. Vol. 9. Is. 3. Pp. 185-211.

Schutte et al., 2001 – *Schutte N. S., Malouff J. M., Bobik Ch., Coston T. D., Greeson C., Rhodes E., Wendorf G.* Emotional intelligence and interpersonal relations // *The Journal of Social Psychology*. 2001. Vol. 141. Is. 4. Pp. 523-536.

### References

Karpathy, A. (2023). *State of GPT (Microsoft Build, May 23-24, 2023)*. Retrieved November 24, 2023 from <https://build.microsoft.com/en-US/sessions/db3f4859-cd30-4445-a0cd-553c3304f8e2>

Bhatia, A. (2023). Watch an A.I. learn to write by reading nothing but. *The New York Times*, April 26. Retrieved November 24, 2023 from <https://www.nytimes.com/interactive/2023/04/26/upshot/gpt-from-scratch.html>

Clark, J., Dario, A. (2023). *Faulty Reward Functions in the Wild*. December 21. Retrieved November 24, 2023 from <https://openai.com/research/faulty-reward-functions>

GOST R 59277-2020. *Artificial Intelligence Systems. Classification of Artificial Intelligence Systems* (2021). Standartinform Publ. (In Russian)

Koval, Ye., Ushkin, S. (2023). Fire, water and ChatGPT: new challenges to the scientific ethos. *Sociodigger*, 4, 5-6 (26). Retrieved November 24, 2023



from <https://sociodigger.ru/articles/articles-page/ogon-voda-i-chatgpt-novye-vyzovy-nauchnomu-ehtosu> (In Russian)

*New Philosophical Encyclopedia: in 4 volumes* (2010). Mysl Publ. (In Russian)

Salovey, P., Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition and Personality*, 9 (3), 185-211.

Schutte, N. S., Malouff, J. M., Bobik, Ch., Coston, T. D., Greeson, C., Rhodes, E., Wendorf, G. (2001). Emotional intelligence and interpersonal relations. *The Journal of Social Psychology*, 141 (4), 523-536.