

УДК 001

DOI: 10.32326/2618-9267-2024-7-1-22-31

КРЕАТИВНОСТЬ В НАУКЕ И «КРЕАТИВНОСТЬ» ИСКУССТВЕННОГО ИНТЕЛЛЕКТА. ВОПРОС О НАСТОЯЩЕЙ КРЕАТИВНОСТИ

Сахарова Анна Владимировна – младший научный сотрудник, Институт философии РАН. Российская Федерация, 109240, г. Москва, ул. Гончарная, д. 12, стр. 1; e-mail: hanna.lazareva@gmail.com; ORCID: 0000-0001-5707-2281

В работе представлен социально-исторический подход к креативности в применении к проблеме искусственного интеллекта (ИИ). В социально-историческом ключе рассмотрены кейсы, которые демонстрируют возможность интерпретации научного результата, полученного ИИ, как креативного. Указывается, что в этом случае часто возникает антропоцентрическая предвзятость, при которой человеку, достигшему схожих результатов, моментально приписывается статус креативного субъекта. Тогда как к результату, полученному ИИ, выдвигаются гораздо более высокие требования, которые касаются как уровня новизны и ценности результата (результат должен быть по меньшей мере гениальным, на уровне нобелевского лауреата), так и процессуальных аспектов получения результата (психологичности, осмысленности, полной самостоятельности и независимости в работе). Однако эти требования далеко не всегда предъявляются человеку. Учёные, например, могут зависеть от заказчиков научных работ или руководства научной организации, а исследования часто проводятся коллективно. В статье предлагается оценивать ИИ по той же шкале, что и человеческих акторов, и в случае получения креативных результатов считать ИИ настоящим, без оговорок, креативным.

Ключевые слова: искусственный интеллект, креативность, социальная эпистемология, наука

Цитирование: Сахарова А.В. Креативность в науке и «креативность» искусственного интеллекта. Вопрос о настоящей креативности // Цифровой учёный: лаборатория философа. 2024. Т. 7. №1. С. 22-31. DOI: 10.32326/2618-9267-2024-7-1-22-31

Рукопись получена: 21 декабря 2023

Пересмотрена: 5 марта 2024

Принята: 6 марта 2024

CREATIVITY IN SCIENCE AND THE “CREATIVITY” OF ARTIFICIAL INTELLIGENCE. THE QUESTION OF REAL CREATIVITY

Anna V. Sakharova – Junior Research Fellow, Institute of Philosophy, Russian Academy

The paper presents a socio-historical approach to creativity applied to the problem of artificial intelligence (AI). Cases that demonstrate the possibil-

of Sciences. 12/1 Goncharnaya St., 109240 Moscow, Russian Federation; e-mail: hanna.lazareva@gmail.com; ORCID: 0000-0001-5707-2281

ity of interpreting a scientific result obtained by AI as creative are considered in a socio-historical context. It is indicated that in this case, an anthropocentric bias often arises, in which a person who has achieved similar results is instantly assigned the status of a creative subject. At the same time, much higher requirements are put forward for the result obtained by AI, regarding both the level of novelty and value of the result (the result must be no less than brilliant, at the level of a Nobel laureate), and the procedural aspects of obtaining the result (psychological, meaningful, complete independence and independence at work). However, these same requirements are not always presented to a person. Scientists, for example, may depend on customers for scientific work or the leadership of a scientific organization, and research is often carried out collectively. The article proposes to evaluate AI on the same scale as human actors and, if creative results are obtained, to consider artificial intelligence truly, without reservation, creative.

Keywords: artificial intelligence, creativity, social epistemology, science

How to cite: Sakharova, A.V. (2024). Creativity in science and the “creativity” of artificial intelligence. The question of real creativity. *The Digital Scholar: Philosopher's Lab, 7 (1):* 22-31. DOI: 10.32326/2618-9267-2024-7-1-22-31 (In Russian).

Received: 21 December 2023

Revised: 5 March 2024

Accepted: 6 March 2024

Искусственный интеллект (ИИ) в настоящее время широко интегрируется в области научных исследований, предоставляя учёным возможность расширить и ускорить процесс научных разработок. ИИ способен как на выполнение рутинной работы, связанной с расчётами и получением результатов по заданным правилам, так и на генерирование гипотез, планирование экспериментов, обработку и анализ обширных объёмов данных разных типов, а также поиск новых неожиданных идей, которые могли бы остаться незамеченными при использовании только традиционных методов научных исследований.

Прежде чем перейти к обсуждению креативности ИИ и его роли в науке, представим рабочее определение ИИ и нашу интерпретацию того, что может называться креативностью. Выбор определений и подходов к рассмотрению указанных понятий в этом случае имеет решающее значение для ответа на главный вопрос настоящей работы – может ли ИИ быть креативным?

Определение и классификации искусственного интеллекта

ChatGPT определяет ИИ как алгоритмы и компьютерные программы, которые позволяют машинам выполнять задачи, обычно требующие человеческого интеллекта – такие как решение проблем, обучение, распознавание закономерностей, понимание естественного языка и принятие решений. ИИ может использоваться в различных приложениях, включая робототехнику, обработку естественного языка, компьютерное зрение и машинное обучение, с целью создания систем, которые могут имитировать когнитивные функции, подобные человеческим. Такое определение хорошо коррелирует с определениями, которые дают человеческие эксперты в области ИИ, поэтому оно вполне может быть использовано как рабочее в данной статье.

Такой имитации когнитивных функций человека можно достичь разными методами и в зависимости от этого выделить несколько типов ИИ по методам работы.

Первый тип – нисходящий алгоритм, так называемое программирование по правилам – алгоритм, который предопределяет результат, используя заданные разработчиком строгие предписания. У такого алгоритма есть минусы – не все процессы можно описать с помощью правил; например, таким алгоритмам так и не покорился перевод текста с одного языка на другой.

Однако при их помощи можно решить множество других сложных задач, требующих рассмотрения существенного количества вариантов и обращения к большим объёмам данных. В качестве примера приведём теорему о четырёх красках – первую теорему, доказанную с помощью компьютера. Эта теорема утверждает, что всякую расположенную на плоскости или на сфере карту можно раскрасить не более чем четырьмя разными цветами так, чтобы любые две области с общим участком границы имели разный цвет. ИИ здесь не создаёт чего-то нового, не становится источником креативных решений, а избавляет человека от рутинной работы. В этом случае результат работы полностью предопределён создателем алгоритма и ни о какой креативности самого алгоритма речи не идёт.

При этом история теоремы о четырёх красках может продемонстрировать отношение учёных к результатам, полученным ИИ. Эта теорема была доказана в 1976 г., однако результат, полученный при помощи компьютера, не признавался многими математиками долгие годы, поскольку его нельзя проверить вручную. Определённость среди математиков наступила более чем через 30 лет спустя, когда в 2005 г. доказательство было проверено при помощи *Coq v7.3.1* – специализированного программного обеспечения, разработанного для проверки теорем. При этом, как утверждает Маркус дю Сотой, «у человека больше шансов допустить ошибку, чем у компьютера. То, что я сейчас скажу, может показаться ересью, но существуют, вероятно, тысячи доказательств с ненайден-

ными логическими пробелами или ошибками. Мне ли не знать: в паре случаев я обнаруживал логические дыры в своих доказательствах уже после того, как опубликовал их. Эти изъяны поддавались исправлению, но их не заметили ни рецензенты, ни редакторы» (Сотой, 2020, с. 212).

Второй тип ИИ основан на машинном обучении – компьютер сам генерирует правила, обнаруживает закономерности и делает прогнозы на основе массива данных. Когда речь идёт про ИИ чаще всего, конечно, имеют в виду именно этот тип, поскольку он, в отличие от линейных алгоритмов, может делать неожиданные вещи – например, замечать не обнаруженные человеком детали на медицинских изображениях. В случае машинного обучения результат его работы не полностью предопределён его создателем, может оказаться больше заложенного и не выводиться из него напрямую; сам программист, например, гораздо хуже интерпретирует медицинские изображения, чем созданный им ИИ.

Далее, говоря про ИИ, следует указать, что по уровню и широте выполняемых (сейчас или в перспективе) функций ИИ делят на узкий (приспособлен под определённый класс задач), общий (решает различные задачи) и сильный (обладает чувствительностью и сознанием).

Сейчас мы можем говорить только о существовании узкого ИИ, способного решать определённый тип задач: переводить текст, пересказывать его, интерпретировать или генерировать изображения. И если, как пишет Евгений Масланов (Масланов, 2023), мы считаем необходимым условием креативности некоторые психологические и социальные аспекты деятельности или наличие неявного знания, то здесь мы можем закончить разговор о креативном ИИ. Этих свойств у существующего сейчас узкого ИИ нет, и нет никакой уверенности в том, что они когда-то появятся. Сильный ИИ сегодня – лишь проект, гипотеза и потенциальная возможность.

Итак, в настоящей статье речь пойдёт об узком ИИ, основанном на машинном обучении, его возможностях и потенциальной креативности.

Креативность в науке и особенности её определения применительно к искусственному интеллекту

Далее обратимся к тому, как можно определить и интерпретировать креативность в контексте размышлений об ИИ. В современных исследованиях креативность может пониматься по-разному. С одной стороны, когнитивно-методологические подходы интерпретируют её, опираясь на психологические свойства индивида и когнитивные особенности его мышления. Этот подход предполагает понимание креативности как «предзаданной» характеристики субъекта, следствием которой будет получение некоторого креативного научного результата (Шибаршина, Дорожкин, 2023). С другой стороны, различные социально-исторические подходы

описывают креативность с позиций оценки научного результата научным сообществом и обществом в целом. Во втором случае мы направляем наш взгляд как бы в обратную сторону: результат, характеризующийся как креативный, делает своего создателя и акт мышления, предшествующий результату, креативными – первична в этом случае оценка результата внешними акторами. Подробнее об этом мы писали в одной из работ (Касавин, Сахарова, 2023).

Для описания человеческой креативности необходимо использовать сочетание этих подходов: «Творчество помещается тем самым в лагуну между уникальностью творческой личности и механизмами социального признания» (Касавин, 2022, с. 22). Для описания креативности ИИ мы не можем по причинам, указанным выше, использовать психологический подход – это бы завершило разговор в ту же минуту, поскольку узкий ИИ, который существует сейчас, не может обладать никакими психологическими характеристиками.

Однако, если применить социально-исторические подходы, мы можем попытаться оценить некоторые результаты работы ИИ с позиций креативности. Когда мы говорим о креативности с социально-исторических позиций, то не важно, кто этот креативный результат получил – человек или ИИ, поскольку оценивается он внешними акторами: научным сообществом, научными коммуникаторами или другими заинтересованными, но не связанными с наукой группами. Мы предлагаем оценить несколько кейсов – результатов исследований, выполненных ИИ – основываясь на социально-историческом подходе. Наша гипотеза заключается в том, что ИИ с позиций оценки результата деятельности может обладать креативностью, и если бы точно такой же результат был получен человеком, то этот человек был бы назван креативным.

На похожую интерпретацию креативности, судя по всему, опирался и А.М. Фейгельман (Фейгельман, 2023) в своём выступлении, однако, на наш взгляд, делал это непоследовательно. Утверждая, что исследователи и кураторы оценивают результаты, полученные ИИ, крайне высоко, Артём Маркович говорит, что этот результат всё равно не может быть назван креативным. Мы же считаем, что именно эта высокая оценка профессиональным сообществом и определяет результат, который получает ИИ, как креативный.

С социально-исторических позиций креативное описывается как обладающее тремя свойствами – ценностью, новизной и неожиданностью: «Creativity is the ability to come up with ideas or artefacts that are new, surprising and valuable» (Boden, 2004, p. 2). Все три фактора оцениваются внешними акторами: научным сообществом, заинтересованными социальными группами, не связанными с наукой, научными коммуникаторами, но никогда не рассматриваются самим актором, кем бы он ни был – человеком или компьютером.

Креативный искусственный интеллект: три кейса

Может ли компьютер получить новое, ценное и неожиданное (по мнению научного и других сообществ) знание? Чтобы ответить на этот вопрос рассмотрим несколько кейсов из деятельности *Google DeepMind* (ранее *DeepMind Technologies*). Эти кейсы касаются технологий *AlphaGo* и *AlphaFold*, работы компании над развитием математической библиотеки *Mizar*, а также исследований интуиции в математике при помощи ИИ.

Первый кейс касается игры в го. Игра в го схожа с доказыванием теорем, она может быть интерпретирована математически как поиск определённых точек на дереве возможных решений, но этот поиск не сводится к простому перебору всех возможных вариантов. Го не поддаётся фиксированному набору правил и стратегий, требует максимального из всех аналогичных игр творческого компонента. *Deep Blue* победил Г. Каспарова в 1996 г., и с тех пор преимущество в шахматах остаётся за ИИ. В отличие от игры в шахматы, в которой победа компьютера над человеком была лишь вопросом времени, игра в го долго считалась принципиально непобедимой компьютеру. И понадобилось целых 20 лет чрезвычайно интенсивного развития цифровых технологий, чтобы победу в этой творческой игре наконец-то одержал компьютер. В 2016 г. *Alpha Go* победила чемпиона мира по игре в го Ли Седоля. Но эта победа интересна не сама по себе. Во время второй игры серии произошло необычное событие – машина сделала знаменитый 37-й ход, настолько неожиданный что комментаторы-профессионалы ахнули. Этот ход не вписывался ни в одну из стратегий игры, существовавших до этого момента, и противоречил классической тактике. Однако именно он предопределил победу, доказав свою эффективность. *Alpha Go* не просто победила человека, она победила его максимально неожиданным образом, и её стиль игры изменил человеческие стратегии игры в го.

Можем ли мы называть то, что произошло, креативным событием? ИИ в этом вопросе явно превзошёл своих создателей-программистов: ни один из них не обладал умениями, дающими возможность даже приблизиться по уровню к чемпиону мира. Если мы ожидаем, что ИИ, метафорически говоря, должен создать новую игру в го, а не просто иную стратегию, то мы тем самым предъявляем к нему требования, завышенные относительно человека. Но какой человек в принципе может справиться с этой задачей? Люди – профессионалы игры в го – не справились даже с тем, чтобы обнаружить стратегию, предложенную ИИ.

Можно также предположить, что эти стратегии возможны только для интеллекта, который обладает вычислительными возможностями, превышающими человеческие. Однако, как показало дальнейшее развитие игры, появившиеся стратегии и тактики могут также использоваться людьми, и для того, чтобы выигрывать, не надо обладать бесконечными ресурсами для расчётов. *Alpha Go*

нашёл новое ценное и неожиданное не в бесконечном многообразии расчётов – а в рамках человеческих возможностей!

А теперь давайте представим, что эту победу одержал не ИИ, а обычный человек, назвали бы мы его креативным? Несомненно. То, что было сделано, обладало и новизной, и ценностью, и неожиданностью, предопределило развитие игры и оказалось явно больше заложенного изначально. И если этого вполне достаточно для человека, чтобы называться креативным, то должно быть достаточно и для ИИ.

Второй кейс продолжает первый – алгоритм *AlphaFold* от *DeepMind*, созданный на основе *Alpha Go*, привнёс новые данные в задачу сворачивания белка (Jumper et al., 2021). В своей речи при присуждении Нобелевской премии по химии в 1972 г. Кристиан Анфинсен предложил гипотезу о том, что последовательность аминокислот в белке должна полностью определять его трёхмерную структуру. Этот вопрос был назван «проблемой сворачивания белков» и был одним из важнейших вопросов биологии на протяжении последних 50 лет. Долгое время исследователи использовали экспериментальные методы, такие как рентгеновская кристаллография и криоэлектронная микроскопия, которые были трудоёмкими и дорогостоящими, а некоторые белки, кроме того, не поддаются такому изучению.

В 2020 г. (статья в *Nature* вышла в 2021 г.) компания *DeepMind* продемонстрировала способность своего программного обеспечения к точному предсказанию структуры множества белков, исходя только из их аминокислотной последовательности, которая определяется ДНК. Согласно данным разработчиков, точность прогнозирования третичной структуры белка алгоритмом *AlphaFold* приблизилась к точности экспериментального определения этой структуры.

Третий пример касается доказательства математических теорем – высшей формы абстракции и креативности в науке. Даже здесь ИИ показывает интересные результаты. Учитывая тесную связь между математикой и алгоритмами, никого, наверное, не удивит тот факт, что ИИ может помогать математикам доказывать сложнейшие теоремы. В математике ИИ давно уже используется в качестве ценного инструмента для опровержения гипотез, ускорения вычислений, обнаружения существования структуры в математических объектах.

Чтобы опровергнуть гипотезу, ИИ просто нужно найти хотя бы одно противоречие, перебрав необходимое для этого количество данных. Но доказательство гипотезы (даже на специальном адаптированном для компьютеров языке) – гораздо более сложная задача. Однако в созданной ещё в 1970-е гг. математической библиотеке *Mizar*, содержащей записанные формальным, понятным компьютеру языком теоремы, группа исследователей из *DeepMind* и *Google* обнаружила 56 % теорем, доказанных без участия человека вообще. И доля этих теорем (справедливости ради скажем, что не

без участия специалистов из группы *DeepMind*) постепенно растёт (Сотой, 2020, с. 286-289).

ИИ также занимается доказательством и разработкой теорем в сотрудничестве с математиками. Например, ИИ в сотрудничестве с математиками успешно разрабатывал две математические проблемы: о структуре узлов и о симметрии (Davies et al., 2021). ИИ от *DeepMind* исследовал существующие в этой области математические конструкции на предмет закономерностей, тестировал интуиции математиков и обнаруживал новые возможные решения. Разработчик утверждает, что система обнаружила как ранее известные, так и новые закономерности, а также направила математиков-людей к открытиям. И хотя участники проекта признаются, что компьютерные предположения полезны для «дополнения деталей», но никогда не заменят человеческую интуицию и творческий подход, однако они говорят и об особой, превышающей сухую и инструментальную роль ИИ в работе, связанной с интуицией и воображением (Castelvecchi, 2021, p. 202).

Здесь можно возразить, что это исследование проводила целая научная группа, и роль ИИ в ней соответствовала уровню лаборанта. Однако, во-первых, речь идёт о лаборанте, умеющем доказывать теоремы, которые находятся на переднем крае науки, что, несомненно, делает его креативным. Во-вторых, в современных естественных науках множество исследований проводятся большими, а то и гигантскими, научными группами, однако мы не отказываем участникам (независимо от статуса) в том, что их вклад в общее исследование является креативным.

Выводы

Когда речь заходит об участии ИИ в том или ином открытии, исследовании или разработке, мы сталкиваемся с тем, что можно назвать антропоцентрической предвзятостью – те же результаты, полученные человеком, были бы немедленно названы креативными. Причём многие из этих результатов действительно технически могут быть доступны человеку, поскольку они не связаны с перебором бесконечного количества вариантов. Однако к ИИ мы предъявляем заведомо невыполнимые требования: нам мало ценного, нового и неожиданного результата – мы ждём от него психологичности, осмысленности и полной самостоятельности во всём: от инициативы и постановки задачи до конечного результата, да ещё и полной прозрачности в том, как получен результат.

Но мы не предъявляем такие требования к человеку. Можно долго задаваться вопросом – насколько учёные самостоятельны в остановке задачи, если у них, например, есть госзаказ? Насколько они самостоятельно приняли решение заниматься той или иной темой? Насколько может быть назван креативным ценный новый и неожиданный результат, полученный аспирантом, которому задачу

поставил научный руководитель? Или насколько креативен вклад в гениальную статью каждого из, скажем, ста соавторов?

Кроме того, мы ожидаем от ИИ уровня креативности, свойственного по меньшей мере нобелевским лауреатам, и сравниваем его с креативностью, которая свойственна «обычным» людям, а сферическим учёным в вакууме, обладающим запредельными креативными способностями по созданию настоящих революций в науке и технике. Анализ креативности в науке часто оставляет за бортом рутинную научную деятельность, которая тоже обладает ценностью и подходит для изучения того, как функционирует наука в «обычной жизни». Такая обыденная креативность теряется и при исследовании креативности ИИ. Иными словами, креативность в науке может быть обнаружена не только в прорывных исследованиях, но и в обычной рутинной научной работе.

Если во всех этих случаях мы закрываем глаза на усложняющие ситуацию факторы и считаем человека креативным, то и ИИ, получивший аналогичный ценный, новый и креативный результат, в этих же условиях должен быть назван креативным. Иными словами, рассматривая креативность с социально-исторических позиций оценки результата, можно сказать, что ИИ, не превосходя вершины человеческого интеллекта, является не просто креативным, но и более креативным, чем средний исследователь.

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

Дорожкин, Шибаршина, 2023 – *Дорожкин А.М., Шибаршина С.В.* Эпистемологическая рандомизация, или о креативности в науке // Эпистемология и философия науки. 2023. Т. 60. № 1. С. 21-33. <https://doi.org/10.5840/eps20236012>

Касавин, 2022 – *Касавин И.Т.* Научное творчество как социальный феномен // Эпистемология и философия науки. 2022. Т. 59. № 3. С. 19-29. <https://doi.org/10.5840/eps202259336>

Касавин, Сахарова, 2023 – *Касавин И.Т., Сахарова А.В.* Креативность – не сущность, а существование! // Эпистемология и философия науки. 2023. Т. 60. № 1. С. 50-59. <https://doi.org/10.5840/eps20236015>

Масланов, 2024 – *Масланов Е.В.* Обладает ли искусственный интеллект творческим началом? // Цифровой учёный: лаборатория философа. 2024. Т. 7. № 1. С. 6-14.

Сотой, 2020 – *Сотой М.* Код креативности: как искусственный интеллект учится писать, рисовать и думать / Пер. с англ. Д.А. Прокофьев. М.: КоЛибри, Азбука-Аттикус, 2020. 384 с.

Фейгельман, 2023 – Фейгельман А.М. Своевольный помощник: нейросети как инструмент художника // Цифровой учёный: лаборатория философа. 2024. Т. 7. № 1. С. 15-21.

Boden, 2004 – Boden M. *The Creative Mind: Myths and Mechanisms*. L.: Routledge, 2004. 360 p.

Castelvecchi, 2021 – Castelvecchi D. DeepMind's AI helps untangle the mathematics of knots // *Nature*. Vol. 600. P. 202.

Davies et al., 2021 – Davies A., Veličković P., Buesing L. et al. Advancing mathematics by guiding human intuition with AI // *Nature*. 2021. Vol. 600. Pp. 70-74. <https://doi.org/10.1038/s41586-021-04086-x>

Jumper et al., 2021 – Jumper J., Evans R., Pritzel A. et al. Highly accurate protein structure prediction with AlphaFold // *Nature*. 2021. Vol. 596. Pp. 583-589. <https://doi.org/10.1038/s41586-021-03819-2>

References

Boden, M. (2004). *The Creative Mind: Myths and Mechanisms*. L.: Routledge, 2004. 360 p.

Castelvecchi, D. (2021). DeepMind's AI helps untangle the mathematics of knots. *Nature*, 600, 202.

Davies, A., Veličković, P., Buesing, L. et al. (2021). Advancing mathematics by guiding human intuition with AI. *Nature*, 600, 70-74. <https://doi.org/10.1038/s41586-021-04086-x>

Dorozhkin, A.M., Shibarshina, S.V. (2023). Epistemological randomization, or on creativity in science. *Epistemology and Philosophy of Science*, 60 (1), 21-33. <https://doi.org/10.5840/eps20236012> (In Russian)

Feigelman, A.M. (2024). Self-willed assistant: neural networks as an artist's tool. *The Digital Scholar: Philosopher's Lab*, 7 (1): 15-21.

Jumper, J., Evans, R., Pritzel, A. et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596, 583-589. <https://doi.org/10.1038/s41586-021-03819-2>

Kasavin, I.T. (2022). Creativity in science as a social phenomenon. *Epistemology and Philosophy of Science*, 59 (3), 19-29. <https://doi.org/10.5840/eps202259336> (In Russian)

Kasavin, I.T., Sakharova, A.V. (2023). Creativity is not Essence, but Existence! *Epistemology and Philosophy of Science*, 60 (1), 50-59. <https://doi.org/10.5840/eps20236015> (In Russian)

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

Sautoy, M. (2020). *The Creativity Code. How AI Is Learning to Write, Paint and Think*. Kolibri, Azbuka-Attikus Publ. (In Russian)