



НАУЧНАЯ СТАТЬЯ
УДК 165:004.8

DOI: 10.18287/2782-2966-2024-4-3-31-36

Дата поступления: 15.06.2024
рецензирования: 01.09.2024
принятия: 01.10.2024

Н.А. Крылов

Вологодский государственный университет,
г. Вологда, Российская Федерация
E-mail: krylovna@vogu35.ru
ORCID: <http://orcid.org/0009-0009-1378-0548>

Моделирование креативности в искусственном интеллекте: возможности и границы

Аннотация: статья посвящена возможностям и границам разработки креативного искусственного интеллекта, в частности, способности машин определять уровень креативности производимых ими объектов. Для оценки креативности предложена трёхступенчатая модель новизны, включающая онтологический, субъективный и семантический уровни. Выделяется три признака креативных идей или артефактов: новизна, неожиданность и ценность. Описывается модель состязательной генеративной сети «CAN: Creative Adversarial Networks», которая создает новые художественные стили и оценивает их новизну. Обсуждаются возможности и границы моделирования юмора в креативном искусственном интеллекте. Анализируются примеры успешной работы нейронных сетей, генерирующих шутки и пишущих сценарии, показывается, что ограниченность таких систем проявляется в отсутствии у машины чувства контекста, ощущения пространства и времени. Кроме того, важнейшим условием успешного написания шуток является способность смеяться над ними. Люди осознанно выбирают тему и формат юмора, а у машин отсутствует собственное целеполагание. Показано, что разработанные к настоящему времени технологии можно обобщить термином «слабый креативный искусственный интеллект», так как они могут создавать новые объекты, но не способны к целеполаганию и рефлексии. Несмотря на это, возможности искусственного интеллекта постоянно расширяются, что меняет наше представление о границах моделируемости естественного интеллекта.

Ключевые слова: философия техники; искусственный интеллект; понятие креативности; границы искусственного интеллекта.

Цитирование: Крылов Н.А. Моделирование креативности в искусственном интеллекте: возможности и границы // Семиотические исследования. Semiotic studies. 2024. Т. 4, № 3. С. 31–36. DOI: <http://doi.org/10.18287/2782-2966-2024-4-3-31-36>.

Благодарности: автор выражает благодарность доктору философских наук Н.А. Ястреб за оказанную помощь и мотивацию к научной деятельности.

Информация о конфликте интересов: автор заявляет об отсутствии конфликта интересов.

© Крылов Н.А., 2024 – аспирант, преподаватель кафедры философии, заместитель директора института социальных и гуманитарных наук, Вологодский государственный университет, 160035, Российская Федерация, г. Вологда, ул. Ленина, д. 15.

SCIENTIFIC ARTICLE

N.A. Krylov

Vologda State University,
Vologda, Russian Federation
E-mail: krylovna@vogu35.ru
ORCID: <http://orcid.org/0009-0009-1378-0548>

Modeling creativity in artificial intelligence: possibilities and limits

Abstract: the article is dedicated to exploring the possibilities and limitations of developing creative artificial intelligence, particularly the ability of machines to determine the level of creativity in the objects they produce. To assess creativity, a three-step model of novelty is proposed, including ontological, subjective, and semantic levels. Three features of creative ideas or artifacts are identified: novelty, unexpectedness, and value. The article describes the model of a competitive generative network called "CAN: Creative Adversarial Networks," which creates new artistic styles and evaluates their novelty. The possibilities and limitations of modeling humor in creative artificial intelligence are discussed. The article analyzes examples of successful

work by neural networks that generate jokes and write scripts, showing that the limitation of such systems is the machine's lack of a sense of context, space, and time. Additionally, a crucial condition for successfully writing jokes is the ability to laugh at them; humans can consciously choose the topic and format of humor, while machines lack their own goal-setting. It is shown that the technologies developed to date can be generalized as "weak creative artificial intelligence" since they can create new objects but are not capable of goal-setting and reflection. However, the possibilities of artificial intelligence are constantly expanding, changing our understanding of the limits of modeling natural intelligence.

Key words: philosophy of technology; artificial intelligence; concept of creativity; limits of artificial intelligence.

Acknowledgments: the author expresses gratitude to Doctor of Philosophy N. Yastreb for the provided assistance and motivation for scientific activity.

Citation: Krylov, N.A. (2024), Modeling creativity in artificial intelligence: possibilities and limits, *Semioticheskie issledovaniya. Semiotic studies*, vol. 4, no. 3, pp. 31–36, DOI: <http://doi.org/10.18287/2782-2966-2024-4-3-31-36>.

Information about conflict of interests: the author declares no conflict of interests.

© **Krylov N.A., 2024** – PhD student, a lecturer at the Department of Philosophy, Vologda State University, 15, Lenina str, Vologda, 160035, Russian Federation.

Введение

Ключевым вопросом философии искусственного интеллекта является обозначение границы моделируемости естественного интеллекта в искусственных системах (Ястреб 2021). В связи с постоянными открытиями в области искусственного интеллекта в последние десятилетия представления об этой границе значительно изменились. В истории искусственного интеллекта период оптимизма, вдохновленный созданием перцептрона, программ «Логик-теоретик» и «Логик-математик», сначала сменился этапом застоя и разочарования («зима» искусственного интеллекта), а затем, начиная с 1980-х гг., периодом постепенного возрождения. Появление в XXI веке больших лингвистических моделей открыло новые возможности для моделирования человеческого интеллекта. В структуру естественного интеллекта входят как функции, которые успешно моделируются, например, распознавание образов, так и те, которые пока нереализуемы в системах искусственного интеллекта, прежде всего, творческая и креативная деятельность. Эту проблему поставила ещё Ада Лавлейс, которая утверждала, что главным критерием наличия интеллекта является креативность (Stein 1985). Создание нейросетей, которые генерируют изображения, музыку и тексты, заставляет нас по-новому взглянуть на проблему творческой деятельности и возможности её передачи машине.

Соотношение понятий «творчество» и «креативность» в контексте искусственного интеллекта

Категория «творчество» вводится Аристотелем, который определяет его как способность человека к нахождению новых решений и созданию новых вещей (Аристотель 1927). Для него способность к творчеству была неотъемлемой частью человеческой природы, и он считал, что она может

проявляться в различных сферах жизни, включая искусство, науку, философию и технологии. Это дало начало широкой трактовке понятия творчество, в рамках которой творческая деятельность определяется «как процессы порождения нового» (Нестеров, Дёмина 2020, с. 84). В рамках данного подхода смысл творческой деятельности видится в «увеличении многообразия человеческого мира в процессе культурной миграции» (Энциклопедия эпистемологии и философии науки 2009, с. 954). Творческая деятельность не ограничивается действиями субъекта и распространяется на общество и культуру в целом.

Постановка задачи моделирования творчества требует его конкретизации. И.А. Бескова и И.Т. Касавин связывают творчество с уровнем и качеством познавательной деятельности индивида и определяют его как «иерархическое структурированное единство способностей, которое определяет уровень и качество мыслительных процессов, направленных на приспособление к изменяющимся и неизвестным условиям в сенсомоторных, наглядных, оперативно-деятельностных и логико-теоретических формах» (там же). Исследование и моделирование творческой деятельности субъекта в когнитивных науках потребовало разграничения понимания творчества как культурного феномена и вида когнитивной деятельности.

Креативность как научный термин вводится психологом Д. Симпсоном, который определяет её как «способность человека отказываться от стереотипных и исторически сложившихся способов мышления» (Симпсон 1948, с. 172). Дж. Гилфорд в работе «Природа человеческого интеллекта», разрабатывая концепцию креативности, предлагает похожий подход и описывает её как «способность порождать новые идеи, отличающиеся от традиционных схем мышления, а также навыки, способствующие решению проблемных ситуаций» (Guilford 1967, p. 197). В данной работе мы будем

опираться на определение, которое дала М. Боден, определившая креативность как «способность придумывать идеи и артефакты, которые считаются новыми, неожиданными и ценными» (Boden 2004, p. 16). Обобщая данные определения, можно отметить, что креативность рассматривается как построение новых возможностей для субъекта, а творчество – как создание новых возможностей для культуры.

Ключевой проблемой в определении креативности является понимание того, что считать новым и какой новый объект должен появиться, чтобы мы считали его результатом креативной деятельности. Для решения этой проблемы можно выделить три уровня новизны: онтологический, субъективный и семантический.

К *онтологическому уровню* новизны можно отнести всё то, чего не было ранее. При таком подходе любой объект, которого ранее не было в пространстве и времени, будет считаться новым вне зависимости от степени новизны своих характеристик. В этом смысле можно считать новым любое изображение, сгенерированное нейросетью. Обучившись на некоторой совокупности примеров, нейросеть фактически продолжает последовательность, имеющуюся в обучающей выборке. В этом случае речь не идет о создании новых смыслов или качественных изменений чего-либо, а происходит продолжение в новых объектах того, что уже существует. Второй, *субъективный*, уровень новизны заключается в том, что создается нечто новое и неожиданное для субъекта, но при этом в результате более тщательного анализа субъект может описать алгоритм получения такого результата. Зная этот алгоритм, другой субъект может прийти в такому же или похожему результату, вне зависимости от его сложности. На самом высоком, *семантическом*, уровне находятся процессы создания объектов, не поддающиеся на данный момент описанию и алгоритмизации. В этом случае мы не понимаем, как они получены, и не можем «разложить на атомы» эти процессы. Похожую мысль высказывал Г. Хакен, который определял креативность как «рождение идей, которые не рождались никогда прежде и более того – рождение которых в высшей степени маловероятно» (Хакен 2001, с. 313–314). Для данного уровня характерно нахождение нестандартных способов решения проблем. Сложность моделирования такой деятельности связана с её уникальностью и невоспроизводимостью. Креативная деятельность этого уровня не может быть реализована в алгоритмических, программируемых моделях.

Моделирование креативности: постановка проблемы

Как уже было упомянуто во введении, проблему создания машин, способных к творческой

деятельности, поставила ещё Ада Лавлейс. В современных исследованиях искусственного интеллекта сформировалась целая область, специализирующаяся на разработке креативных моделей и систем. Под креативным искусственным интеллектом в данной работе понимается *направление в искусственном интеллекте, в котором методы машинного обучения используются для генерации новых объектов на основе закономерностей, обнаруженных в имеющихся данных*. Основные методы и подходы к созданию креативных систем искусственного интеллекта были разработаны достаточно давно, но только в последние годы разработчики получили, с одной стороны, возможность использовать достаточно большие вычислительные мощности, а с другой – появилась стремительно развивающаяся сфера креативных индустрий, для решения задач которой эти модели необходимы.

В круг основных задач систем креативного искусственного интеллекта входят генерирование изображений, кода, звуковых эффектов, написание историй, стихов, эссе, статей, диалогов, переработка и перефразирование текста в различных жанрах, создание карикатур, комиксов, юмористических иллюстраций, мелодий, аранжировок, песен в разных жанрах (например, Jukebox), генерация видео и трехмерных моделей объектов, то есть практически все основные виды творческой деятельности человека.

Для оценки креативности искусственного интеллекта широко используется видоизменённый тест Тьюринга (Тьюринг 1960). В настоящее время существует множество его вариантов, в частности: музыкальный, юмористический, креативный тест Тьюринга. Для прохождения теста машинам ставится задача создать произведение, которое будет неотличимо от результатов творческой деятельности человека. Далее экспертному сообществу предлагается определить, какие из произведений созданы человеком, а какие – машиной (Winters, Nys, De Schreye 2018). Примером может служить проведенный нами опрос, в ходе которого респондентам были предложены две шутки, одна из которых написана человеком, а вторая – чатом GPT-4. В результате 35 % опрошенных высказали мнение о том, что вторая шутка написана человеком. Существуют и более успешные примеры прохождения машинами креативных тестов Тьюринга. Поскольку в таких тестах достаточно часто побеждают машины, необходимо признать, что либо мы должны согласиться с тем, что искусственный интеллект стал креативным, либо пересмотреть сам тест Тьюринга.

Важным вопросом разработки креативного искусственного интеллекта является то, сможет ли сам искусственный интеллект определять уровень креативности производимых им объектов. В

указанном выше определении сформулированы три признака креативных идей или артефактов, а именно новизна, неожиданность и ценность. Первый параметр новизны может быть оценён машиной через соотнесение с имеющимися аналогами. Наиболее высокий уровень генерируемых продуктов в настоящее время предоставляют состязательные системы креативного искусственного интеллекта (GAN – Generative Adversarial Networks). Их целью является создание на основе обучающей выборки таких новых данных (изображений, текстов, аудио и т.д.), которые были бы неотличимы от реальных. Примером может служить состязательная генеративная сеть «CAN: Creative Adversarial Networks» (<https://arxiv.org/abs/1706.07068>), предназначенная для создания новых художественных стилей и оценки их новизны. Она состоит из двух нейросетей – генератора и дискриминатора. Задача генератора состоит в том, чтобы на основе имеющихся примеров создать новый стиль изобразительного искусства и добавить его в имеющуюся базу данных. Вторая часть системы, дискриминатор, анализирует данные и пытается выявить, какие из стилей были предложены машиной. Данная модель одновременно способна генерировать новые объекты и оценивать их новизну. Генератор и дискриминатор обучаются в соревновательной среде, что способствует их более быстрому и эффективному обучению, так как с каждым циклом обучения генератор становится лучше в создании реалистичных данных, а дискриминатор становится лучше в их отличии от реальных. Состязательный характер обучения позволяет генератору «экспериментировать» с данными и создавать нестандартные решения, которые могут быть рассмотрены субъектом как «творческие».

Параметр неожиданности носит субъективный характер и определяется эмоциональной реакцией удивления человека. Ценность также представляет собой параметр, который привносится субъектом. Система креативного искусственного интеллекта может оценивать степень отличия полученного результата от аналогов, то есть новизну в формальном смысле, однако наделение сгенерированного объекта смыслом осуществляется только человеком.

Сфера применения креативного искусственного интеллекта включает не только традиционные типы творческой деятельности, такие как написание музыки, стихотворений, создание изображений, но и создание прогнозов развития болезней, дизайн протезов в медицине, управление денежными активами на фондовых биржах. Но все эти системы не способны к целеполаганию и не могут инициировать креативную деятельность. Поэтому зачастую такие системы используются в качестве креативных помощников. Постановка

цели, задачи и оценка результатов осуществляется человеком, а использование искусственного интеллекта позволяет расширить возможности субъекта.

Задача разработки креативного искусственного интеллекта порождает множество этических и юридических проблем. Центральными из них являются вопросы определения авторства продуктов и защиты авторских прав. Традиционное авторское право признает автором человека, однако в случае систем искусственного интеллекта возникает неопределенность. Кто может претендовать на право собственности: разработчик системы, ее владелец или пользователь, который ввел задание? Отдельную сложность представляет проблема оценки оригинальности продукта. Поскольку системы креативного искусственного интеллекта обучаются на большом количестве примеров, можно ли считать созданные ими произведения достаточно оригинальными, чтобы они могли быть защищены авторским правом? Имеют ли авторы произведений, использовавшихся для обучения модели, право претендовать на авторство создаваемых ею результатов? Этические проблемы связаны с возможным причинением вреда репутации, благополучию или чувствам людей. В этом случае проблема авторства тесно переплетается с проблемой моральной ответственности (Ястреб 2024). Без решения этих проблем дальнейшее развитие креативного искусственного интеллекта несет в себе большие риски для экономики, культуры и общества в целом.

Возможности и границы моделирования юмора в креативном искусственном интеллекте

Одним из направлений развития креативного искусственного интеллекта стало моделирование юмора. Способность шутить и понимать юмор всегда считалась исключительно человеческой чертой. Проблема моделирования юмора помогает более точно понимать границу между тем, чему мы можем научить машину, и между тем, что остается исключительной чертой человеческого интеллекта. Юмористическую деятельность сложно моделировать, так как она носит комплексный характер. Генерация юмора требует высокого уровня владения языком, гибкости и подвижности мышления, креативных способностей, умения выявлять противоречия, знания законов логического мышления и возможности их нарушать, способности реагировать на актуальные события и понимать социальный контекст в целом.

Несмотря на сложность моделирования юмора, современные нейронные сети успешно справляются с рядом задач в этой сфере. Так, машины генерируют шутки, роботы стендап-комики создают монологи и выступают с ними, генеративные системы, такие как GPT, пишут сценарий

юмористических выступлений, придумывают анекдоты и каламбуры (<https://openai.com/gpt-4>). В исследовании (Nijholt 2018) проводилось сравнение восприятия шуток, написанных людьми-волонтерами, и системой искусственного интеллекта GAG. В результате проведенного эксперимента выяснилось, что шутки, которые были написаны машиной, в два раза чаще воспринимались как смешные, чем те, которые были написаны людьми. Ключевым преимуществом таких систем является способность к обучению. Следовательно, нейросеть, подобно человеку, способна учиться на своих ошибках и совершенствовать свои возможности генерации юмора. Примеры генеративных систем показывают, что нейронные сети способны корректно воспроизводить структуру шуток, выявлять противоречия и парадоксы, использовать такой инструмент, как языковые игры.

В то же время нейронные сети обладают рядом существенных ограничений. Ключевой проблемой является отсутствие у машины чувства контекста, ощущения пространства и времени. Кроме того, важнейшим условием успешного написания шуток является способность смеяться над ними. Люди осознанно выбирают тему и формат юмора, а у машин отсутствует собственное целеполагание. В качестве вывода приведем ответ, который дал нам чат GPT-4, на вопрос о том, каковы возможности и границы машинной генерации юмора: «что касается моих возможностей – я могу предложить шутки и анекдоты на различные темы, но у меня нет чувства юмора, поэтому я не могу смеяться или оценивать шутки».

Заключение

В русскоязычной литературе принято разделять понятия «творчество» как деятельность, направленная на развитие общества и культуры в целом, и «креативность» как построение новых возможностей для субъекта. Применительно к искусственному интеллекту корректным является использование термина «креативность». Под креативным искусственным интеллектом понимается направление в искусственном интеллекте, в котором методы машинного обучения используются для генерации новых объектов на основе закономерностей, обнаруженных в имеющихся данных. В современном креативном искусственном интеллекте выделяются две основных задачи. Прагматическая задача состоит в расширении креативных способностей человека за счет создания инструментов. Исследовательская задача состоит в моделировании креативности человека. Разработанные к настоящему времени технологии можно обобщить термином «слабый креативный искусственный интеллект», так как они могут создавать новые объекты, но не способны к целеполаганию и рефлексии. Примером могут служить системы

генерации юмора, которые могут создавать шутки, но не имеют чувства юмора. Несмотря на это, возможности искусственного интеллекта постоянно расширяются, что меняет наше представление о границах моделируемости естественного интеллекта.

Библиографический список

- Boden, M.A. (1996), *Dimensions of creativity, Artificial Intelligence. Handbook of Perception and Cognition*, MIT Press, Cambridge, UK, DOI: <https://doi.org/10.1016/B978-012161964-0/50011-X>.
- Boden, M.A. (2004), *The creative mind: myths and mechanisms*, 2nd edn, Routledge, London, UK.
- Guilford, J.P. (1967), *The nature of human intelligence*, McGraw-Hill, New York, USA.
- Nijholt, A. (2018), *Robotic stand-sp comedy: state-of-the-art, Lecture Notes in Computer Science*, Springer, Cham, UK.
- Stein, D. (1985), *Ada: a life and a legacy*, MIT Press, Cambridge, MA, UK.
- Winters, T., Nys, V., De Schreye, D. (2018), *Automatic joke generation: learning humor from examples, Lecture Notes in Computer Science*, Springer, Cham, UK.
- Аристотель. Поэтика. Ленинград: АКАДЕМИА, 1927. 121 с.
- Барышников П.Н. Методологические возможности и границы вычислительных моделей сознания. Пятигорск: Изд-во ПГУ, 2018. 310 с.
- Нестеров А.Ю., Демина А.И. Воображение в семиотике творчества // *Вестник Томского государственного университета*. 2020. № 460. С. 84–89.
- Хакен Г. Принципы работы головного мозга: Синергетический подход к активности мозга, поведению и когнитивной деятельности; перевод с немецкого Ю. А. Данилов. Москва: Per Se, 2001. 351 с.
- Симпсон Д.Г. Темпы и формы эволюции. Москва: Госиздат, 1948. 358 с.
- Тьюринг А. Могут ли машины мыслить. Москва: Государственное издательство Физико-математической литературы, 1960. 67 с.
- Энциклопедия эпистемологии и философии науки. Москва: «Канон+» РООИ «Реабилитация», 2009. 1248 с.
- Ястреб Н.А. Как возможно техническое творчество? // *Философская мысль*. 2023. № 2. С. 15–25.
- Ястреб Н.А. Концепции этики искусственного интеллекта: от принципов к критическому подходу // *Семиотические исследования. Semiotic studies*. 2024. Т. 4. № 1. С. 24–30. DOI: <http://doi.org/10.18287/2782-2966-2024-4-1-24-30>.
- Ястреб Н.А. Шахматы и машина Тьюринга: границы применимости вычислительного подхода в социальных науках // *Философия науки и техники*. 2021. № 1. С. 51–55.

References

- Boden, M.A. (1996), *Dimensions of creativity, Artificial Intelligence. Handbook of Perception and Cognition*, MIT Press, Cambridge, UK, DOI: <https://doi.org/10.1016/B978-012161964-0/50011-X>.
- Boden, M.A. (2004), *The Creative Mind: Myths and Mechanisms*, 2nd edn, Routledge, London, UK.
- Guilford, J.P. (1967), *The Nature of Human Intelligence*, McGraw-Hill, New York, USA.
- Nijholt, A. (2018), *Robotic Stand-Up Comedy: State-of-the-Art, Lecture Notes in Computer Science*, Springer, Cham, UK.
- Stein, D. (1985), *Ada: A Life and a Legacy*, MIT Press, Cambridge, MA, UK.
- Winters, T., Nys, V., De Schreye, D. (2018), *Automatic Joke Generation: Learning Humor from Examples, Lecture Notes in Computer Science*, Springer, Cham, UK.
- Aristotel (1927), *Poetics*, ACADEMIA, Leningrad, USSR.
- Baryshnikov, P.N. (2018), *Methodological possibilities and boundaries of computational models of consciousness*, PGU, Pyatigorsk, Russia.
- Nesterov, A.Yu. and Demina, A.I. (2020), *Imagination in the semiotics of creativity, Vestnik Tomskogo gosudarstvennogo universiteta*, no. 460, pp. 84–89.
- Haken, G. (2001), *Brain principles: a synergistic approach to brain activity, behavior, and cognition*, Yu. A. Danilov (transl.), Per Se, Moscow, Russia.
- Simpson, D.G. (1948), *Rates and forms of evolution*, Gosizdat, Moscow, Russia.
- Turing, A. (1960), *Can machines think*, Gosudarstvennoye izdatel'stvo Fiziko-matematicheskoy literatury, Moscow, Russia.
- Encyclopedia of epistemology and philosophy of science* (2009), Kanon ROOI «Reabilitatsiya», Moscow, Russia.
- Yastreb, N.A. (2023), *How is technical creativity possible?*, *Filosofskaya mysl*, no. 2, pp. 15–25.
- Yastreb, N.A. (2024), *Artificial intelligence ethics: from principles to the critical approach*, *Semioticheskie issledovaniya. Semiotic studies*, vol. 4, no. 1, pp. 24–30, DOI: <http://doi.org/10.18287/2782-2966-2024-4-1-24-30>.
- Yastreb, N.A. (2021), *Chess and the Turing machine: the limits of applicability of the computational approach in the social sciences*, *Filosofiya nauki i tekhniki*, no. 1, pp. 51–55.

Submitted: 15.06.2024

Revised: 01.09.2024

Accepted: 01.10.2024