УДК 1+004.81+008.2

ТВОРЧЕСКИЙ ПОТЕНЦИАЛ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В КОНТЕКСТЕ ИДЕИ НОВОГО ПРОСВЕЩЕНИЯ

Д. А. Стельмахов[1]

Современный мир сталкивается с рядом глобальных проблем, характер которых усугубляется развитием технологий. В этом контексте в массовом сознании возникает беспокойство относительно возможности искусственного интеллекта (ИИ) превзойти человека в интеллектуальной и творческой деятельности. Тема креативности ИИ становится актуальной и вызывает в научном сообществе споры относительно его творческого потенциала. В связи с этим члены Римского клуба в 2018 г. представили концепцию нового Просвещения и принципа баланса как ответ на вызовы технологического развития. Эти идеи подчеркивают необходимость сбалансированного подхода к техническому прогрессу с учетом глобальных ценностей и этических принципов. В статье рассматриваются определения творчества и ИИ, формулируются критерии для оценки творчества и анализируются современные достижения в области моделирования творческих процессов в ИИ. Также статья охватывает использование различных алгоритмов для реализации разнообразных видов творческой деятельности, включая генерацию текстов, изображений и музыки. Кроме того, представлены разнообразные мнения исследователей относительно творчества ИИ. Для систематизации этих точек зрения предложена классификация, в соответствии с которой их можно ранжировать. Вывод автора следующий: в настоящее время проблема оценки творческого потенциала ИИ не имеет решения, что сохраняет ее актуальность для дальнейших исследований. В будущем ситуация может измениться, поэтому значительное внимание необходимо уделять прогнозированию возможных последствий развития таких технологий, а также разработке механизмов контроля деятельности ИИ, потому как человек несет от-

[1] Институт философии РАН,
Россия, 109240, Москва, ул. Гончарная, д. 12, стр. 1.
Поступила в редакцию: 25.08.2023 г.
doi: 10.5922/0207-6918-2023-4-13

CREATIVE POTENTIAL OF ARTIFICIAL INTELLIGENCE IN THE CONTEXT OF THE IDEA OF THE NEW ENLIGHTENMENT

D. A. Stelmakhov[1]

The modern world is confronted with a series of global problems, exacerbated by technological advancements. In this context, concerns arise in the public consciousness regarding the potential of artificial intelligence (AI) to surpass humans in intellectual and creative activities. The topic of AI creativity becomes pertinent and sparks debates within the scientific community regarding its creative potential. In response to these challenges, members of the Club of Rome in 2018 propose the concept of a new Enlightenment and the principle of balance as a response to the demands of technological development. These ideas underscore the need for a balanced approach to technical progress, taking into account global values and ethical principles. I explore the definitions of creativity and AI, formulate criteria for assessing creativity, and analyse recent achievements in modelling creative processes in AI. I also focus on the utilisation of various algorithms for implementing diverse forms of creative activities, including text, image and music generation. Additionally, I present diverse opinions of researchers regarding creativity in AI. To systematise these perspectives, I propose a classification which allows their ranking. The conclusion emphasises that currently the issue of evaluating the creative potential of AI lacks a definitive solution, underscoring the relevance of further research. The future outlook may alter this situation, prompting the necessity to focus on predicting potential consequences of technological developments and developing mechanisms for controlling AI activities; humans, after all, bear responsibility for AI actions. Such an approach aligns with the concept of a new

[1] Institute of Philosophy, Russian Academy of Sciences.
12 Goncharnaya st., Moscow, 109240, Russia.
Received: 25.08.2023.
doi: 10.5922/0207-6918-2023-4-13

ветственность за действия ИИ. Такой подход соответствует идее нового просвещения, так как в этом случае обретается баланс между технологическими достижениями и сохранением глобальной гармонии.

Ключевые слова: *творчество, искусственный интеллект, критерии творчества, моделирование, нейросети, алгоритмы, новое просвещение*

Enlightenment which seeks a balance between technological advancements and the preservation of global harmony.

Keywords: *creativity, AI, creativity criteria, modelling, neural networks, algorithms, new Enlightenment*

Введение

В настоящее время можно выделить несколько универсальных факторов, определяющих существование современного человека: глобализация, цифровизация, информатизация, политизация, экологизация и ряд других. Никогда ранее люди не сталкивались с подобными детерминантами собственного бытия, имеющими своим следствием стремительную трансформацию человеческого существования, приводящую к возникновению принципиально новых затруднений. В 2018 г. был опубликован отчет Римского клуба, содержавший мысль о том, что причины данных изменений коренятся в транслируемых философией Просвещения идеалах развития рационализма, науки и техники, а также освобождения общества от догматизма и веры в несомненные авторитеты (Weizsäcker, 2018, p. 92—93).

Согласно мнению членов Римского клуба, реализация и распространение просветительских идей способствовали отделению церкви от государства, человеческой индивидуализации и несогласованности физических и экономических законов, что впоследствии привело к серьезным изменениям в социальной сфере, а также катастрофическому разрыву между людьми и природой. При этом в настоящее время особая роль в усилении данного разрыва отводится развитию цифровых технологий, в особенности искусственному интеллекту. Членами Римского клуба были высказаны опасения по поводу интеллектуального и творческого превосходства ИИ над человеком, так как это может привести к сокращению рабочих мест, ухудшению трудовых отношений, обострению экологической ситуации и т.д. В связи с этим они подняли вопрос о необходимости формирования нового просвещения, основанного на поиске баланса во всех сферах человеческой деятельности (Weizsäcker, 2018, p. 88—89, 95; см. также: Зильбер, 2023; Чалый, 2023).

Но насколько оправдан страх относительно творческого превосходства ИИ? Если у нас есть основания утверждать, что процессы, воспроизводимые ИИ, являются творческими, это влечет за собой необходимость пересмотра понятия «творчество», которое ранее относилось только к человеку. Для того чтобы ответить на вопрос «Способен ли ИИ к творчеству?», прежде всего необходимо определиться с тем, что мы понимаем под творчеством и каковы его признаки. Затем следует обратиться к ИИ и описать принципы его работы, после чего можно будет приступить к анализу деятельности ИИ в соответствии с признаками творчества, учитывая позиции и мнения представителей научного сообщества, а также определить возможные векторы развития данной области.

Что такое творчество?

Изучением проблемы творчества с целью выявления его причин и структуры занимались ученые из разных научных областей. Немецкий ученый Г. Гельмгольц рассматривал творческий акт как системный процесс, состоящий из трех этапов: подготовка, вынашивание идеи, озарение (Ле-

бединский и др., 1966, с. 131—132). Французский математик А. Пуанкаре, анализируя свою творческую деятельность, также пришел к выводу о том, что творческий акт имеет определенные этапы, и дополнил предыдущую схему этапом верификации (Пуанкаре, 1990, с. 407—408). Английский психолог Г. Уоллес представил концепцию четырехэтапного творческого процесса, сформулировав выводы на основе собственных наблюдений и изучения опыта известных ученых и изобретателей (Wallas, 1945, p. 42). Объединяющим звеном их концепций стало представление о творчестве как духовно-практической деятельности, разворачивающейся поэтапно и направленной на создание качественно нового. Обобщенная схема творческого процесса содержит следующие этапы: 1) постановка проблемы; 2) подготовка (поиск и анализ информации); 3) период инкубации (бессознательная работа над проблемой); 4) инсайт (озарение, рождение новой идеи); 5) завершающий этап (проверка и внедрение / корректировка полученного результата).

На современном этапе научного развития когнитивная наука как «область междисциплинарных исследований, изучающая познание и высшие мыслительные процессы с помощью информационных моделей» (Меркулов, 2010, с. 264), не обходит стороной и проблему творчества. Об этом свидетельствует множество исследований, направленных на раскрытие механизмов мышления, воображения, ассоциативных связей и их влияния на генерацию оригинальных идей[2], а также создание и функционирование соответствующих институций (Центр когнитивных наук Университета Рочестера, Институт креативности и инноваций Университета Баффало и др.) и научных периодических изданий (*Creativity Research Journal*, *Journal of Creativity* и др.). Такой интерес к изучению творчества продиктован стремлением выявить основные когнитивные механизмы и процессы, лежащие в основе творческой деятельности человека. Для всестороннего изучения творчества также разрабатываются алгоритмы и программы, которые затем интегрируются в ИИ.

В связи с отсутствием общепринятого определения творчества и многообразием существующих дефиниций мы обратимся к определению, предложенному известной британской исследовательницей в области когнитивных наук Маргарет Боден. Она понимает под творчеством способность генерировать новые ценные идеи и, исходя из этого, различает несколько типов творчества: 1) комбинационный тип, основанный на преобразовании уже знакомых идей в новые комбинации; 2) исследовательский тип, предполагающий обнаружение границ генеративных правил (норм и законов, порождающих систему) в разных областях; 3) трансформационный тип, связанный с преобразованием самого пространства, в котором мы находимся, — создание принципиально новых идей (Boden, 2009, p. 24—25).

Особое внимание к исследованию Боден обосновано потенциалом ее типологии видов творческой деятельности, способной систематизировать процесс творчества на основе его внутренней природы и результирующих аспектов. Ее классификация предоставляет возможность осуществить структурированный подход к анализу творческих процессов. Опираясь на ее определение и классификацию, а также на представленную выше схему этапов, по которым протекает творческий процесс, можно сформулировать следующие основные признаки творческой деятельности: 1) способность обнаружить проблему, которая требует решения; 2) оригинальность конечного результата; 3) его ценность (он решает проблему, справляется с задачей либо помогает по-иному взглянуть

[2] Подобные исследования проводятся в рамках нейронаук с использованием вычислительных методов и средств визуализации. В качестве примеров можно привести исследование роли субъективных ценностей в творческом процессе с построением поведенческой вычислительной модели с ключевыми операциями для генерации творческих идей (Lopez-Persem et al., 2023); изучение корреляции между различными измерениями когнитивных способностей и творчеством (Serban et al., 2023); определение локализации механизмов распознавания объектов с помощью средств цифровой визуализации (Шелепин и др., 2019); построение модели линейной регрессии для предсказания творческого поведения у людей с помощью средств нейровизуализации (Ovando-Tellez et al., 2022) и т.д.

на нее); 4) адаптация к переменам в поставленных задачах; 5) внедрение результата. Эти признаки нам необходимы для последующего анализа, в ходе которого будет произведена оценка деятельности ИИ, симулирующего творческие процессы.

ИИ и принципы его работы

С момента возникновения ИИ в 1940-х гг. и по настоящее время не существует однозначного определения этого термина. Это обусловлено изменчивостью содержания указанного понятия, его связью с исследованиями в разных научных дисциплинах, а также разнообразием целей, стоящих перед учеными. Каждая научная область имеет собственный набор понятий и взглядов на объект, и участники научных сообществ могут подходить к определению ИИ с разных сторон. Также термин «ИИ» может включать в себя разработки различного уровня сложности, что приводит в итоге к разнообразию трактовок (Лысачев, Прохоров, 2023, с. 19—20). Далее мы будем опираться на определение ИИ, предложенное основательницей *Future Today Institute*, зарубежным футурологом Эми Линн Уэбб: «выполнение компьютером действий, для которых обычно требуется человеческий интеллект» (см.: Watson, 2018). Оно позволяет провести параллель между творческой деятельностью человека и действиями ИИ, а также сопоставить их с критериями творчества. Но прежде чем мы приступим к этому, рассмотрим принципы работы ИИ.

Существует несколько типов ИИ: символьный, коннекционистский (нейросетевой), эволюционный и агентный. Принцип работы символьного ИИ заключается в анализе и манипулировании символами, которые представляют смысловую информацию в соответствии с заложенными правилами. В рамках коннекционизма одним из главных направлений является использование нейросетей[3]. Нейросетевой ИИ применяется для обработки сложных и неструктурированных данных (тексты, изображения и звуки). Он способен обучаться и адаптироваться к условиям, что делает его эффективным инструментом для решения разнообразных задач, однако данный подход требует огромного количества информации для обучения сети. Эволюционный ИИ представляет собой разработку алгоритмов, моделирующих эволюционные процессы и генетический отбор. Он успешно применяется для решения задач оптимизации, создания новых дизайнов и нахождения нестандартных решений. Агентный подход заключается в разработке компьютерной программы или системы, предназначенной для восприятия окружающей среды, принятия решений и мер для достижения определенной цели или набора целей. Для создания интеллектуальных агентов могут использоваться все ранее перечисленные подходы (Лысачев, Прохоров, 2023, с. 54—91).

С развитием технологий и накоплением обширных данных стало очевидно, что статичные алгоритмы ограничены в своем функционале. Это привело к формированию нового подхода в разработке ИИ — машинного обучения, в рамках которого ИИ может без явного программирования учиться и принимать решения на основе полученных данных. Термин «машинное обучение» был введен Артуром Самуэлем еще в 1959 г., но получил значительное развитие лишь в 1990-е гг. Машинное обучение как тип ИИ включает в себя разнообразные типы задач: обучение с учителем, обучение без учителя и обучение с подкреплением, где алгоритмы обучаются на основе размеченных данных, находят структуры в данных без явных ответов и взаимодействуют с окружающей средой и обратной связью (Там же, с. 93, 97—98).

[3] Нейросети — это «искусственные, многослойные высокопараллельные (то есть с большим числом независимо параллельно работающих элементов) логические структуры, составленные из формальных нейронов» (Галушкин, 2022).

Современные методы машинного обучения, основанные на нейронных сетях, приобрели доминирующее положение в связи с расширением вычислительных возможностей компьютеров и распространением доступа к обширным массивам данных через Интернет. Эти факторы способствовали их успешному внедрению и распространению. Глубокое обучение — это разновидность машинного обучения, для которого характерно использование больших объемов данных, наличие высокопроизводительных компьютеров, длительное время обучения, а также разнообразие форматов выходных данных. В настоящее время глубокое обучение является доминирующим направлением в сфере ИИ, и наращивание уровня сложности обучаемых моделей обусловливает потребность в увеличении вычислительных ресурсов (Там же, 2023, с. 333).

Таким образом, общая идея ИИ заключается в том, чтобы создать компьютерные системы, способные анализировать данные, принимать решения, обучаться и выполнять задачи, которые ранее требовали бы наличия человеческого интеллекта. Для осуществления таких процессов разрабатываются соответствующие алгоритмы.

Творчество в ИИ

С развитием ИИ появились новые перспективы для моделирования творческих процессов. Множество программ были разработаны с целью симулировать различные аспекты творчества, начиная от генерации музыки и изображений и заканчивая созданием текстов и дизайнов. В их основе лежат разные типы алгоритмов, что позволяет воспроизводить продукты из разнообразных сфер творческой деятельности, однако это влечет за собой следующие вопросы: «Насколько к машинам может быть применимо понятие креативности? Где проходит линия разделения между машинными интеллектуальными функциями и человеческой креативностью и по каким критериям оценивать обе эти сферы?» (Эванс, 2018). Действительно, это достаточно дискуссионные вопросы, и далее мы предпримем попытку найти возможные ответы на них. Для этого мы рассмотрим современные алгоритмы и программы ИИ, используемые для моделирования творческого процесса, и согласно классификации М. Боден определим, какому типу творчества они могут соответствовать.

Примером алгоритмов ИИ, реализующих комбинационный тип творчества, являются генеративно-состязательные алгоритмы и алгоритмы обработки естественного языка (NLP). Они предназначены для создания нового контента. На их основе функционируют программы для генерации текста (*GPT-4*[4] от *OpenAI*), изображений и искусственных медиа (*DALL-E*[5] от *OpenAI*, *Midjourney*[6] от одноименной компании), а также для создания музыки (например, *Magenta*[7] от *Google*). С их помощью можно преобразовывать уже существующие идеи и шаблоны в новые комбинации и вариации.

К исследовательскому типу творчества мы можем отнести алгоритмы обучения с подкреплением и эволюционные алгоритмы, так как они могут искать генеративные правила и законы, которые описывают систему, путем эволюционных процессов. Отличительная черта эволюционных алгоритмов — фактор случайности и способность к самообучению, когда программе предоставляется возможность самостоятельно создавать различные рекомбинации и определять свои дальнейшие действия. Важным условием является корректно написанная программа, которая позволит внесен-

[4] https://chat.openai.com
[5] https://openai.com/dall-e-2
[6] http://www.midjourney.com/
[7] https://magenta.tensorflow.org

ным данным развиваться и преобразовывать заложенные в нее элементы в новую информацию (Cohen-Or, Zhang, 2016, p. 10).

Примером может служить программа *NeuroEvolution of Augmenting Topologies* (NEAT), в которой процесс эволюции приводит к тому, что частицы начинают организовываться в разнообразные визуальные структуры и генерируют абстрактные композиции, геометрические узоры, анимационные эффекты и т.п., а также сервис *Picbreeder*, в котором пользователи могут, начиная с набора случайных пикселей, выбирать изображения, которые им нравятся, и скрещивать их между собой, создавая новые варианты.

Алгоритмы обучения с подкреплением позволяют ИИ обучаться через взаимодействие с окружающей средой и получение обратной связи. В робототехнике и игровой индустрии они используются для создания автономных агентов, способных генерировать творческие решения, как, например, известные многим *AlphaGo* и *AlphaZero* от *DeepMind*[8], способные обучаться игре в го и шахматы, или система *Dactyl* от *OpenAI*, манипулирующая объектами в физическом мире[9].

Для трансформационного типа творчества могут быть задействованы сети долгой краткосрочной памяти (LSTM) и рекуррентные нейронные сети (RNN), а также генеративно-состязательные алгоритмы, способные изменять пространство решений и преобразовывать его, создавая принципиально новые идеи. Они моделируют последовательные данные (музыкальные композиции, тексты) для генерации нового контента на основе обнаруженных паттернов. Например, программа *DeepDream*[10] (*Google*) использует глубокие нейронные сети для генерации психоделических искажений изображений. Платформа *Artbreeder*[11] позволяет пользователям комбинировать изображения и осуществлять над ними различные манипуляции. Программа *AIVA*[12] создает музыкальные темы для проектов, используя *RNN*. Сервис *Jukedeck*[13], основанный на нейронных сетях, включая *LSTM*, также генерирует оригинальную музыку на основе заданных параметров.

Таким образом, к каждому типу творчества можно подобрать соответствующую программу, которая направлена на генерацию новых идей посредством комбинаций, принятия непредвиденных решений в ходе самообучения программы, а также путем полного преобразования имеющейся информации. Однако насколько процесс и результат деятельности ИИ соответствуют критериям творчества (обнаружение проблемы, оригинальность результата, ценность, адаптация к переменам и внедрение результата)?

На первый взгляд может показаться, что для реализации первого пункта у ИИ недостаточно автономности, поэтому обнаружение проблемы и постановка задачи выполняются человеком (путем написания алгоритма, формулирования запроса и т.д.). Тем не менее ИИ на основе самообучающихся алгоритмов может, выявляя закономерности и паттерны, недоступные человеку, обнаруживать сложные проблемы и задачи в областях, связанных с обработкой больших объемов данных. Что касается второго пункта, если понимать под оригинальным (новым) результатом не только то, чего никогда ранее не существовало, но и создание чего-то нового именно для творца, то продукт деятельности ИИ можно считать творческим.

[8] https://www.deepmind.com/research/highlighted-research/alphago
[9] https://openai.com/research/learning-dexterity
[10] https://deepdreamgenerator.com
[11] https://www.artbreeder.com
[12] https://www.aiva.ai
[13] https://www.jukedeck.com

Третий пункт тесно связан с субъективными установками человека, так как оценка ценности продукта формируется не только в соответствии с поставленными задачами, но также зависит от индивидуальных социокультурных и моральных убеждений, характерных для данного индивида. Даже при предположении, что эти установки можно запрограммировать и предоставить ИИ возможность самостоятельно оценивать свой результат, мнения относительно самой сути ценностей и их содержания могут значительно различаться среди разных людей. Данный критерий зависит от контекста, поэтому мы не можем отрицать, что результат деятельности ИИ может обладать ценностью. Адаптацию к изменениям в процессе возможно реализовать посредством новых подходов к ИИ, в рамках которых разрабатываются метаалгоритмы, способные нарушать заложенные в них правила, чтобы рассмотреть возможный результат (Сотой, 2020, с. 320). Последний пункт, связанный с внедрением результата, также реализуется не ИИ, а человеком.

В итоге мы видим, что деятельность ИИ, претендующая на статус творческой, соответствует не всем критериям. Это затрудняет оценку его творческого потенциала, но не позволяет с уверенностью заявить, что ИИ не способен к творчеству. Такой результат можно связать с зависимостью ИИ от человека на текущем этапе развития технологий и предположить, что в будущем ситуация может измениться. Далее мы обратимся к существующим в научном сообществе позициям относительно творчества в ИИ и попробуем обозначить возможные перспективы развития данной области.

Способен ли ИИ к творчеству?

Предполагаемые ответы на вопрос «Способен ли ИИ к творчеству?» можно рассматривать по аналогии с философским зомби[14]. Используя классификацию, предложенную А.Ю. Алексеевым, и адаптируя ее к проблеме творчества в ИИ, можно выделить следующие позиции:

1) азомбисты (категории, характеризующие ментальные процессы, протекающие внутри субъекта, неприменимы к ИИ);

2) нейтральные зомбисты (творческий ИИ является составной частью более масштабной разработки «псевдосознания» в ИИ, которое лишь по своему функционалу может напоминать человеческое);

3) зомбисты (независимо от степени изучения интеллекта человека и протекающих в нем когнитивных процессов можно создать творческий ИИ);

4) антизомбисты (неотличимое функционирование, в том числе и во время творческой деятельности, возможно только благодаря расшифровке и повторению нейрофизиологических кодов психической деятельности вне мозга (Алексеев, 2014, с. 158).

Как мы видим, сторонники первых двух позиций отказываются считать, что ИИ способен быть творческим. Британский математик и профессор Оксфордского университета Маркус дю Сотой в своей работе «Код креативности. Как искусственный интеллект учится писать, рисовать и думать» придерживается следующей позиции: творчество является выражением сущности человека, а потому ИИ отказано в праве быть творцом. По его мнению, творчество невозможно подвергнуть моделированию с помощью ИИ (Сотой, 2020, с. 10). Такой же позиции придерживается профессор Российской академии естественных наук В.А. Самохвалова, отмечая, что «именно творчество, творческая способность, творческая составляющая играют главную роль в самой идентификации естественного человеческого интеллекта» (Самохвалова, 2007, с. 140—141).

[14] Мысленный эксперимент австралийского философа сознания Д. Чалмерса, согласно которому «зомби» может функционировать как человек, но у него отсутствует сознание.

Аарон Херцман, главный научный сотрудник *Adobe Research* в Сан-Франциско, полагает, что ИИ является всего лишь инструментом, и его использование в творческом процессе не дает никаких оснований считать его творцом. Восприятие ИИ в качестве субъекта творческой деятельности вводит в заблуждение, заставляя думать, что ИИ обладает интеллектом, эмоциями и автономией. Возможно, в будущем ИИ удастся усовершенствовать настолько, что мы передадим ему авторство наших собственных работ, однако Херцман отмечает, что это маловероятно (Hertzmann, 2020, p. 48).

Зомбисты в перспективе придерживаются позитивного мнения относительно творчества ИИ, однако не спешат называть современные ИИ творческими. Как отмечает в своем исследовании М. Боден, для ИИ доступно творчество комбинаторного и исследовательского типа, но не трансформационного. Также она полагает, что независимо от того, могут ли компьютеры действительно быть творческими, они могут делать очевидно творческие вещи (Boden, 2004, p. 8—9, 21). Знаменитый исследователь в области ИИ и когнитивной науки Дуглас Хофштадтер, будучи любителем классической музыки, полагал, что ИИ неспособен сочинить музыку, которая была бы сопоставима по качеству с творением человека. Но уже в начале 2000-х гг. ему довелось ознакомиться с музыкальными произведениями, созданными ИИ *EMI* (Experiments in Musical Intelligence), после чего он был вынужден изменить свое мнение (Boden, 2009, p. 32).

Антизомбисты настаивают на том, что для создания творческого ИИ необходимо не только изучить интеллект человека и протекающие в мозге когнитивные процессы, но и научиться воплощать их на физических носителях. Такой позиции придерживается американский ученый в области ИИ Марвин Ли Минский, который отстаивает точку зрения, согласно которой когнитивные процессы и механизмы человека не являются его неотъемлемыми атрибутами, а потому их можно подвергнуть моделированию в ИИ (Минский, 2020).

Мы приходим к выводу, что вопрос о творческом потенциале ИИ на данный момент не имеет однозначного ответа, а потому остается открытым. Обнаруживается множество разнообразных подходов, каждый из которых отражает индивидуальное видение возможности креативности ИИ, а также его соответствие критериям творчества. Возможно, в будущем наступит такой момент, когда появится больше оснований, которые позволят нам прийти к единому мнению относительно творчества в ИИ. Но, как и с любыми другими технологиями, способными оказывать влияние на сферы человеческой жизнедеятельности, необходимо предвидеть, к чему может привести их развитие.

Профессор факультета философии Оксфордского университета, основатель и директор Института будущего человечества Ник Бостром прогнозирует возможные отрицательные результаты распространения и развития ИИ, в связи с чем предлагает меры по их предупреждению, следовать которым необходимо уже сейчас. Во-первых, это принцип общего блага, согласно которому потенциал подобных технологий должен служить благу человечества и реализовываться в рамках этических идеалов. Во-вторых, необходимо параллельно с развитием ИИ разрабатывать методы контроля, чтобы иметь возможность управления в критических ситуациях (Бостром, 2016, с. 383, 386).

Основная мысль его работы заключается в том, что «высший этический приоритет… имеют такие вопросы, как снижение экзистенциального риска и выход на цивилизационную траекторию, которая ведет к лучшему исходу — на благо всего человечества» (Там же, с. 391). Это отлично дополняет принцип баланса, предложенный членами Римского клуба. В его основу легли идеи, подчеркивающие необходимость сохранять и восстанавливать природные ресурсы, уравновешивать технический прогресс с глобальными ценностями и этическими принципами и т.д.

Возвращаясь к возможным негативным последствиям технологического развития, стоит отметить, что ответственность за существующие и прогнозируемые проблемы всецело лежит на человеке, ведь именно он руководит деятельностью ИИ, пишет алгоритмы и разрабатывает программы. Манипуляции, включая дипфейки, модификацию голоса и другие техники, а также применение ИИ в военно-оборонной области, представляют собой явления, обусловленные человеческим вмешательством. Отдельные лица или группы индивидов как активные участники данного процесса определяют вектор развития современных технологий, в том числе ИИ, основываясь на своих моральных установках, мировоззрении и различных социокультурных факторах. Таким образом, дальнейшее развитие данных технологий влечет за собой существенные социальные и этические изменения для человечества, а характер такого развития непосредственно зависит от решений и действий этих субъектов.

Заключение

Проблема творчества — явление многогранное, характеризующееся множеством толкований, из-за чего не представляется возможным сформулировать единое и всеобъемлющее определение, учитывающее все аспекты данной области. В силу отсутствия универсальных критериев в сфере творческой деятельности были избраны критерии, выведенные автором, однако базирующиеся на существующем теоретическом аппарате и психологических концепциях творчества.

От определения ИИ также зависят конечные выводы исследования. Подвергать анализу процессы творчества в рамках ИИ возможно лишь при условии рассмотрения его деятельности по аналогии с человеческой. Если ИИ понимать как научную область, как инструмент для выполнения задач, понятие творчества здесь не только затруднительно в реализации, но и несостоятельно. Современные программы совершили значительный рывок за последние десятилетия, и их достижения, начиная от генерации текстов разного уровня сложности, стиля и объема и заканчивая созданием картин и написанием музыки в самых разных жанрах, заставляют задуматься об успешности моделирования творчества в ИИ и даже способствуют возникновению тенденции обозначения ИИ как потенциально творческого. Однако, как выяснилось, деятельность ИИ отвечает не всем критериям творчества, позиции в научном сообществе тоже разнятся и единого мнения по этому вопросу нет. Практики, занимающиеся разработкой ИИ, нередко определяют его как инструмент. Ученые в области когнитивных наук более благосклонны, однако и они не спешат называть современные ИИ творческими. Тем не менее они отмечают их возможный творческий потенциал в будущем.

Применение творческого ИИ может изменить рынок труда и общественные отношения, поэтому оправданность страха относительно творческого превосходства ИИ зависит от того, каким образом общество сможет адаптироваться к изменениям и предоставить новые возможности и поддержку людям, затронутым этими изменениями. Существует вероятность того, что развитие творческого ИИ будет угрожать человеческой роли в областях, где творчество является ключевой чертой (изобразительное искусство, музыка и т.д.), однако в то же время в рамках научных исследований внедрение ИИ может способствовать решению сложных задач и содействовать прогрессу. Поэтому разработка этичных и социокультурных рамок для использования ИИ может помочь управлять этими аспектами и минимизировать потенциальные риски.

В то же время расширение вычислительной мощности и усовершенствование алгоритмов может вызвать серьезные последствия (угроза приватности, зависимость от технологий, трансформация рынка труда, экологические проблемы и т.д.), к которым человечество не готово. Опасе-

ния, высказываемые исследователями и членами Римского клуба, небезосновательны. Без своевременных коррективов могут реализоваться пессимистичные сценарии. Человечество несет ответственность за свои технологические достижения, включая ИИ, поэтому необходимо соблюдать баланс между технологическими прорывами и общественными интересами.

Список литературы

Алексеев А. Ю. Когнитотехнологические проекты искусственной личности // Человек: Образ и сущность. Гуманитарные аспекты. 2014. №1 (25). С. 156—174.

Бостром Н. Искусственный интеллект. Этапы. Угрозы. Стратегии / пер. с англ. С. Филина. М. : Манн, Иванов и Фербер, 2016.

Галушкин А. И. Нейронные сети [2022] // Большая российская энциклопедия: научно-образовательный портал. URL: https://bigenc.ru/c/neironnye-seti-e734b3/?v=5514538 (дата обращения: 16.08.2023).

Зильбер А. С. Кант и новое просвещение: о балансе между долгом и утилитарными целями // Кантовский сборник. 2023. Т. 42, № 2. С. 40—67.

Лебединский А. В., Франкфурт У. И., Френк А. М. Гельмгольц. М. : Наука, 1966.

Лысачев М. Н., Прохоров А. Н. Искусственный интеллект. Анализ, тренды, мировой опыт / науч. ред. Д. А. Ларионов. М. ; Белгород : Константа-Принт, 2023.

Меркулов И. П. Когнитивная наука // Новая философская энциклопедия. М. : Мысль, 2010. Т. 2. С. 264—265.

Минский М. Машина эмоций / пер. В. Дегтяревой, А. Курышевой. М. : АСТ, 2020.

Пуанкаре А. О науке / пер. с фр. ; под ред. Л. С. Понтрягина. 2-е изд., стер. М. : Наука, 1990.

Самохвалова В. И. Творчество: событие космической и родовой истории человека // Ориентиры. Метафизические исследования человека и мира. 2007. № 4. С. 140—161.

Сотой М. Код креативности. Как искусственный интеллект учится писать, рисовать и думать / пер. Д. Прокофьева. М. : КоЛибри, 2020.

Чалый В. А. Возвращая «культурный мандат»: Идея устойчивого развития в кантианской перспективе // Кантовский сборник. 2023. Т. 42, № 2. С. 68—94.

References

Alekseev, A. Yu., 2014. Cognitotechnological Projects of Artificial Personality. *Human Being: Image and Essence. Humanitarian Aspects*, 1(25), p. 156-174. (In Rus.)

Boden, M., 2009. Computer Models of Creativity. *AI Magazine*, 3, pp. 23-34.

Boden, M. 2004. *The Creative Mind: Myths and Mechanisms. Second Edition*. London: Routledge, Taylor & Francis.

Bostrom, N., 2016. *Iskusstvennyy intellekt. Etapy. Ugrozy. Strategii [Superintelligence: Paths, Dangers, Strategies]*. Translated by S. Filin. Moscow: Mann, Ivanov i Ferber. (In Rus.)

Chaly, V. A., 2023. Reclaiming the "Cultural Mandate": The Idea of Sustainable Development in the Kantian Perspective. *Kantian Journal*, 42(2), pp. 68-94. http://dx.doi:10.5922/0207-6918-2023-2-4.

Cohen-Or, D., Zhang, H., 2016. From Inspired Modeling to Creative Modeling. *The Visual Computer*, 32, pp. 7-14. https://doi.org/10.1007/s00371-015-1193-9.

Evans, J., 2018. Can AI be Creative? In: Orange Business, [online]. Available at: <https://www.orange-business.com/en/blogs/can-ai-be-creative> [Accessed: 19.08.2023].

Galushkin, A. I., 2022. Neural Networks. In: *Bol'shaya Rossiyskaya Entsiklopediya: Nauchno-obrazovatel'nyy Portal* [*Big Russian Encyclopedia: Scientific and Educational Portal*]. Available at: <https://bigenc.ru/c/neironnye-seti-e734b3/?v=5514538> [Accessed: 16.08.2023]. (In Rus.)

Hertzmann, A., 2020. Computers Do Not Make Art, People Do. Communications of the ACM, 63(5), pp. 45-48. https://doi.org/10.1145/3347092.

Lebedinskii, A. V., Frankfurt, U. I. and Frank, A. M., 1966. *Helmholtz (1821—1894)*. Moscow: Nauka. (In Rus.)

Lopez-Persem, A., Moreno-Rodriguez, S., Ovando-Tellez, M., Bieth, T., Guiet, S., Brochard, J. and Volle, E., 2023. How Subjective Idea Valuation Energizes and Guides Creative Idea Generation. *American Psychologist*. Advance online publication. https://dx.doi.org/10.1037/amp0001165.

Шелепин К. Ю., Соколов А. В., Фокин В. А. и др. Феномен инсайта и цифровая визуализация активности головного мозга человека // Вестник ЮУрГУ. Сер.: Психология. 2019. Т. 10, № 4. С. 47—55.

Эванс Дж. Искусственный интеллект и творчество [2018] // Orange Business. URL: https://www.orange-business.com/ru/blogs/iskusstvennii-intellekt-i-tvorchestvo (дата обращения: 19.08.2023).

Boden M. Computer Models of Creativity // AI Magazine. 2009. Vol. 30, № 3. P. 23—34. https://doi.org/10.1609/aimag.v30i3.2254.

Boden M. The Creative Mind: Myths and Mechanisms. 2nd ed. L. : Routledge, Taylor & Francis, 2004.

Cohen-Or D., Zhang H. From Inspired Modeling to Creative Modeling // The Visual Computer. 2016. № 32. P. 7—14. https://doi.org/10.1007/s00371-015-1193-9.

Hertzmann A. Computers Do Not Make Art, People Do // Communications of the ACM. 2020. Vol. 63, № 5. P. 45—48. https://doi.org/10.1145/3347092.

Lopez-Persem A., Moreno-Rodriguez S., Ovando-Tellez M. et al. How Subjective Idea Valuation Energizes and Guides Creative Idea Generation // American Psychologist. 2023. Advance online publication. https://dx.doi.org/10.1037/amp0001165.

Ovando-Tellez M., Kenett Y. N., Benedek M., et al. Brain Connectivity-Based Prediction of Real-life Creativity is Mediated by Semantic Memory Structure // Science Advances. 2022. Vol. 8, № 5. https://doi.org/10.1126/sciadv.abl4294.

Serban A., Kepes S., Wang W., Baldwin R. Cognitive Ability and Creativity: Typology Contributions and a Meta-analytic Review // Intelligence. 2023. Vol. 98. https://doi.org/10.1016/j.intell.2023.101757.

Wallas G. The Art of Thought. L. : C. A. Watts & Co. Ltd, 1945.

Watson M. Why Business Leaders Should Think of AI as an Umbrella Term [2018] // Medium. URL: https://medium.com/opex-analytics/why-business-leaders-should-think-of-ai-as-an-umbrella-term-dba8badc55e4 (дата обращения: 20.08.2023).

Weizsäcker E. U., Wijkman A. Come On! Capitalism, Short-Termism, Population and the Destruction of the Planet. N. Y. : Springer, 2018.

Lysachev, M. N., Prokhorov, A. N., 2023. *Iskusstvennyy intellekt. Analiz, trendy, mirovoy opyt [Artificial Intelligence. Analysis, Trends, Global Experience]*. Edited by D. A. Larionov. Corporate Edition. Moscow; Belgorod: KONSTANTA-print. (In Rus.)

Merkulov, I. P., 2010. Cognitive Science. In: *Novaya filosofskaya entsiklopediya* [*New Philosophical Encyclopedia*]. Vol. 2. Moscow: Mysl', pp. 264-265. (In Rus.)

Minski, M., 2020. *Mashina emotsiy* [*The Emotion Machine*]. Moscow: AST, 2020. (In Rus.)

Ovando-Tellez, M., Kenett, Y. N., Benedek, M., Bernard, M., Belo, J., Beranger, B., Bieth, T. and Volle, E., 2022. Brain connectivity-based prediction of real-life creativity is mediated by semantic memory structure. *Science Advances*, 5(8). https://doi.org/10.1126/sciadv.abl4294.

Poincaré, A., 1990. *O nauke* [*On Science*]. *Second Edition*. Translated from French by L. S. Pontryagin. Moscow: Nauka. (In Rus.)

Samokhvalova, V. I., 2007. Creative Activity: an Event of Cosmic and Generational History of Human. *Landmarks. Metaphysical Studies of Human and the World*, 4, pp. 140-161. (In Rus.)

Serban, A., Kepes, S., Wang, W., Baldwin, R., 2023. Cognitive Ability and Creativity: Typology Contributions and a Meta-analytic Review. *Intelligence*, 98. https://doi.org/10.1016/j.intell.2023.101757.

Shelepin, K. Y., Sokolov, A. V., Fokin, V. A., Vasiliev, P. P. and Pronin, S. V., 2019. The Insight Phenomenon and the Digital Visualization of the Human Brain Activity. *Bulletin of the South Ural State University. Ser. Psychology*, 10(4), pp. 47-55. https://doi.org/10.14529/psy170405 (In Rus.)

Sotoy, M. 2020. *Kod kreativnosti. Kak iskusstvennyy intellekt uchitsya pisat', risovat' i dumat'* [*Code of Creativity: How Artificial Intelligence Learns to Write, Draw, and Think*]. Moscow: KOLibri. (In Rus.)

Wallas, G., 1945. *The Art of Thought*. London: C. A. Watts & Co. Ltd.

Watson, M., 2018. Why Business Leaders Should Think of AI as an Umbrella Term. In: *Medium*, [online]. Available at: https://medium.com/opex-analytics/why-business-leaders-should-think-of-ai-as-an-umbrella-term-dba8badc55e4 [Accessed: 20.08.2023].

Weizsäcker, E. U., Wijkman, A., 2018. *Come On! Capitalism, Short-termism, Population and the Destruction of the Planet*. New York, NY: Springer.

Zilber, A. S., 2023. Kant and the New Enlightenment: On the Balance between Duty and Utilitarian Ends. *Kantian Journal*, 42(2), pp. 40-67. http://dx.doi:10.5922/0207-6918-2023-2-3.

Об авторе

Денис Александрович ***Стельмахов***, аспирант, Институт философии РАН, Москва, Россия.

E-mail: denis.stelmakhov@mail.ru

ORCID: https://orcid.org/0000-0002-9631-6768

Для цитирования:

Стельмахов Д. А. Творческий потенциал искусственного интеллекта в контексте идеи нового просвещения // Кантовский сборник. 2023. Т. 42, № 4. С. 240—251.

doi: 10.5922/0207-6918-2023-4-13

© Стельмахов Д. А., 2023.

ПРЕДСТАВЛЕНО ДЛЯ ВОЗМОЖНОЙ ПУБЛИКАЦИИ В ОТКРЫТОМ ДОСТУПЕ В СООТВЕТСТВИИ С УСЛОВИЯМИ ЛИЦЕНЗИИ CREATIVE COMMONS ATTRIBUTION (CC BY) (HTTP://CREATIVECOMMONS.ORG/LICENSES/BY/4.0/)

The author

Denis Aleksandrovich ***Stelmakhov***, Institute of Philosophy, Russian Academy of Sciences, Moscow, Russia.

E-mail: denis.stelmakhov@mail.ru

ORCID: https://orcid.org/0000-0002-9631-6768

To cite this article:

Stelmakhov, D. A., 2023. Creative Potential of Artificial Intelligence in the Context of the Idea of the New Enlightenment. *Kantian Journal*, 42(4), pp. 240-251. (In Rus.)

http://dx.doi.org/10.5922/0207-6918-2023-4-13

© Stelmakhov D. A., 2023.

SUBMITTED FOR POSSIBLE OPEN ACCESS PUBLICATION UNDER THE TERMS AND CONDITIONS OF THE CREATIVE COMMONS ATTRIBUTION (CC BY) LICENSE (HTTP://CREATIVECOMMONS.ORG/LICENSES/BY/4.0/)