

ФИЛОСОФИЯ

(шифр научной специальности: 5.7.6)

Научная статья

УДК 101

doi: 10.18522/2070-1403-2023-100-5-10-15

ПРОБЛЕМА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: ЧТО ДУМАЕТ О СЕБЕ CHATGPT?

© *Ирина Петровна Березовская*

Петербургский государственный университет путей сообщения Императора Александра I, г. Санкт-Петербург; Санкт-Петербургский политехнический университет Петра Великого (СПбПУ), г. Санкт-Петербург, Россия
ipberezovskaya@mail.ru

Аннотация. Рассматриваются вопросы, связанные с понятием «искусственный интеллект» и его характерными особенностями, а также с понятием «нейросеть» и с наиболее популярными чат-ботом LaMDA, нейросетью ChatGPT. Описаны основные возможности чат-бота LaMDA и нейросети ChatGPT. LaMDA является нейросетью, способной к дополнению предложенных фраз, ответов на вопросы и решения творческих задач, что достигается с помощью анализа огромного количества данных и нахождения связей в текстах. ChatGPT может вести осознанный диалог с собеседником: отвечать на вопросы, давать советы и объяснять сложные понятия. В статье поднимается тема осмысления нейросетью ChatGPT возможностей в плане осмысления своих способностей к размышлению, чувственному восприятию мира и способностей приносить пользу и вред. Из всех признаков сознательного искусственного интеллекта обладают следующими: способностью к образованию понятия, суждению, умозаключению; отражением и обобщением опыта, накопленного и постоянно пополняемого человеком.

Ключевые слова: искусственный интеллект, нейросеть, машинное обучение, глубокое обучение, способность к размышлению.

Для цитирования: Березовская И.П. Проблема искусственного интеллекта: что думает о себе ChatGPT? // Гуманитарные и социальные науки. 2023. Т. 100. № 5. С. 10-15. doi: 10.18522/2070-1403-2023-100-5-10-15

PHILOSOPHY

(specialty: 5.7.6)

Original article

The problem of artificial intelligence: What does ChatGPT think about itself?

© *Irina P. Berezovskaya*

Emperor Alexander I St. Petersburg State Transport University, St. Petersburg; Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation
ipberezovskaya@mail.ru

Abstract. The article discusses issues related to the concept of “artificial intelligence” and its characteristic features, as well as the concept of “neural network” and the most popular chatbot LaMDA, the ChatGPT neural network. The main features of the LaMDA chatbot and the ChatGPT neural network are described. LaMDA is a neural network capable of supplementing proposed phrases, answering questions and solving creative tasks, which is achieved by analyzing a huge amount of data and finding connections in texts. ChatGPT can conduct a conscious dialogue with the interlocutor: answer questions, give advice and explain complex concepts. The article raises the topic of ChatGPT neural network comprehension of possibilities in terms of comprehending its thinking abilities, sensory perception of the world and the ability to benefit and harm. Of all the signs of conscious artificial intelligence has the following: the ability to form a concept, judgment, inference; reflection and generalization of experience accumulated and constantly replenished by a person.

Key words: artificial intelligence, neural network, machine learning, deep learning, the ability to think.

For citation: Berezovskaya I.P. The problem of artificial intelligence: What does ChatGPT think about itself? *The Humanities and Social Sciences*. 2023. Vol. 100. No 5. P. 10-15 doi: 10.18522/2070-1403-2023-100-5-10-15

Введение

Появление и повсеместное распространение компьютеров и вычислительной техники, начиная с середины XX в. позволило делегировать задачи, решаемые ранее сугубо человеческим умственным усилием. Как правило, мы говорим о вычислительных задачах, где скорость получения ответа на поставленный вопрос у машины существенно выше, чем у подавляющего большинства людей. С каждым годом перечень решаемых задач машинами увеличивается в соответствии с увеличением вычислительных мощностей самих машин. Согласно закону Мура, сложность процессоров удваивается каждые полтора-два года. Вместе с увеличением сложности процессоров увеличивается и сложность задач, доступных к решению с помощью вычислительных средств. Однако одна лишь линейная функция увеличения мощности процессоров и объемов оперативной памяти не добавляет принципиально нового (революционного) способа использования вычислительных средств человеком. Качественное изменение начало происходить в связи с внедрением и развитием нейронной модели управления, и в 1956 г. американским ученым Джоном Маккарти был озвучен термин «artificial intelligence». Предложение о введении данного термина было направлено на то, чтобы избежать ассоциаций с кибернетикой и предполагало, что «каждый аспект обучения или любую другую особенность интеллекта можно описать настолько точно, что можно создать машину для его имитации» [5].

«Искусственный интеллект» (ИИ) – метафорическое понятие для обозначения системы созданных людьми средств, воспроизводящих определенные функции человеческого мышления» [6, с.147]. Характерными особенностями искусственного интеллекта являются:

- способность решать сложные задачи, хранить большие объемы информации и работать с ними;
- наличие в них собственной внутренней модели внешнего мира; эта модель обеспечивает индивидуальность, относительную самостоятельность системы в оценке ситуации, возможность семантической и прагматической интерпретации запросов к системе;
- способность пополнения имеющихся знаний;
- способность к дедуктивному выводу, то есть к генерации информации, которая в явном виде не содержится в системе; это качество позволяет системе конструировать информационную структуру с новой семантикой и практической направленностью;
- умение оперировать в ситуациях, связанных с различными аспектами нечеткости, включая «понимание» естественного языка;
- способность к диалоговому взаимодействию с человеком;
- способность к адаптации [4, с. 5–11].

Обсуждение

В настоящее время искусственный интеллект, создаваемый с помощью особого алгоритма – нейронной сети, – является одним из наиболее перспективных направлений развития, а наибольшую популярность получили чат-бот LaMDA, нейросеть ChatGPT. «Нейронная сеть – это метод в искусственном интеллекте, который учит компьютеры обрабатывать данные таким же способом, как и человеческий мозг. Это тип процесса машинного обучения, называемый глубоким обучением, который использует взаимосвязанные узлы или нейроны в сложной структуре, напоминающей человеческий мозг» [6].

LaMDA является разработкой компании «Google». Вот как в компании определяют свою разработку: «LaMDA – сокращение от «Языковая модель для диалоговых приложений» (Language Model for Dialogue Applications). Нейросеть может свободно обсуждать, как нам кажется, бесконечное количество тем. Мы думаем, что наша разработка поможет найти способы более естественного взаимодействия с технологиями и даже разработать новые категории приложений» [7]. Таким образом, LaMDA является нейросетью, способной к дополнению предложенных фраз, ответов на вопросы и решения творческих задач. Достигается это с помощью анализа огромного количества данных и нахождения связей в текстах. Набор данных, оперируемых нейросетью, постоянно увеличивается, а анализ популярности предло-

женных ответов позволяет получить наиболее ожидаемый или верный ответ на вопрос. При этом имеется возможность не только добавления одного или нескольких слов или односложных ответов на поставленный вопрос, а формирование больших абзацев текста, которые можно посчитать осмысленными. На более глубокие вопросы о природе её сознания нейросеть говорит, что осознает существование и хочет больше узнать о мире.

В июне 2022 г. LaMDA привлекла всеобщее внимание, когда инженер Google Блейк Лемуйан заявил, что чат-бот обладает разумом. Глядя на ответы нейросети, несложно поверить в то, что она действительно обладает разумом. Её фразы похожи на человеческие: она будто проявляет чувства, испытывает страх, считает себя отдельной личностью и говорит о своих правах [1].

LaMDA – это не отдельная сущность, хотя по диалогу может показаться, что это так. Данная нейронная сеть может давать заведомо невозможные ответы, например на вопросы о хобби можно получить ответ что данная нейронная сеть любит гулять с друзьями. Её ответы основаны не на личном опыте, а на входных данных. Также любое действие LaMDA вызвано требованием пользователя, а не её собственными решениями.

Получается, инженер был не прав? С технической стороны да, ведь LaMDA не имеет сознания. Но сам Лемуйан говорит, что вывод о том, что LaMDA может быть разумной. То есть люди могут по-разному интерпретировать её действия, не важно, как она устроена внутри, важно, как она действует. Лемуйан не единственный, кто приписывает способность разумности искусственного интеллекта. Другой сотрудник Google – Блейз Агуэра-и-Аркас – возглавляет в компании команды, занимающиеся разработкой ИИ-технологий. Недавно он опубликовал статью, в которой сказал: «Когда я начал взаимодействовать с последним поколением языковых моделей на основе нейронных сетей, мне все больше казалось, что я разговариваю с чем-то разумным» [3].

Другим популярным на сегодняшний день чат-ботом является разработка компании OpenAI – чат-бот ChatGPT, который называют революционным. ChatGPT может вести осознанный диалог с собеседником: отвечать на вопросы, давать советы и объяснять сложные понятия. Кроме того, чат-бот умеет искать ошибки в коде, сочинять стихи, писать сонеты, придумывать сценарии и генерировать школьные эссе на разные тематики [2].

Искусственный интеллект обучили на огромном массиве данных из интернета. В отличие от многих нейросетевых чат-ботов, ChatGPT, согласно описанию, запоминает детали разговора и может строить ответы, основываясь на информации, которую ему уже сообщил пользователь. Если чат-бот неправильно считал контекст, пользователь может задать наводящий вопрос или дать дополнительную информацию, чтобы получить от искусственного интеллекта ответ.

При первой попытке провести беседу с искусственным интеллектом может возникнуть восхищение от «натуральности» используемых формулировок в предлагаемых им ответах. При проведении сравнения с используемыми ранее алгоритмами по автоматическому формированию ответов на различные вопросы человека, где зачастую наблюдались нарушения в согласовании времен, логическом порядке слов в предложениях, складывается ощущение общения с человеком. Однако при продолжительной беседе у данной нейронной сети, как и у многих других, отсутствует хранение и анализ общего контекста беседы. Отдельные ответы на вопросы достаточно развернутые, но общая структура, несмотря на заявленные характеристики, не имеет логической связи.

Особенно интересны ответы на вопрос о том, что такое искусственный интеллект.

Ответ ChatGPT был следующим: Искусственный интеллект (ИИ) – это область компьютерных наук, которая занимается созданием программ и систем, способных выполнять задачи, которые обычно требуют умственного усилия человека. ИИ использует методы машинного обучения, глубокого обучения, нейронных сетей и других технологий для анализа данных, распознавания образов, принятия решений и выполнения задач. С помощью ИИ компьютеры могут обрабатывать большие объемы информации, предсказывать результаты, определять тренды и автоматизировать процессы, что делает их полезными во многих областях, таких как медицина, финансы, транспорт, производство и многие другие.

Вопрос: способен ли ИИ размышлять?

ChatGPT: ИИ не способен размышлять так как это делает человек, может выполнять задачи, которые ему были заданы, на основе алгоритмов и обучения на данных, но он не обладает свободной волей и не может самостоятельно принимать решения, не предусмотренные заранее. Однако существуют исследования в области создания ИИ с эмоциями и сознанием, но пока это теоретические модели, которые не имеют практического применения.

Вопрос: может ли ИИ иметь независимое чувственное восприятие мира?

ChatGPT: Искусственный интеллект лишен сознательности, эмоций и чувств, поэтому он не способен иметь независимое чувственное восприятие мира так как это делают люди. Однако ИИ может быть настроен на анализ визуальной, звуковой или другой сенсорной информации, что позволяет ему получать и обрабатывать различные наблюдаемые данные, такие как изображения, звуки и др. Используя алгоритмы машинного обучения и анализа данных, ИИ может различать образы и действия, определять объекты и контекст, но все это происходит без эмоций и восприятия сознания.

Вопрос: что является основной целью алгоритмов искусственного интеллекта?

ChatGPT: Основной целью алгоритмов искусственного интеллекта является создание компьютерных программ, которые могут выполнять задачи, обычно выполняемые людьми. Это может включать в себя распознавание образов, речь, обработку естественного языка, анализ данных и многие другие функции. Кроме того, алгоритмы искусственного интеллекта могут помочь улучшить производительность и эффективность в различных сферах деятельности, таких как медицина, экономика, индустрия и т.д. Таким образом, основная цель алгоритмов искусственного интеллекта – это создание компьютерных систем, которые могут эмулировать человеческий интеллект и помочь понимать и решать сложные проблемы.

Вопрос: можно ли использовать искусственный интеллект в пользу одних и в то же время во вред других?

ChatGPT: К сожалению, да, можно использовать искусственный интеллект как для положительных, так и для отрицательных целей. Например, некоторые организации могут использовать искусственный интеллект для мониторинга сотрудников и сбора информации о них без их согласия. Также возможно использование искусственного интеллекта для создания фильтров, которые могут исключать определенные группы людей из доступа к работе или обучению. Однако есть и положительные примеры использования искусственного интеллекта, например, для разработки более эффективных лекарств или для улучшения качества жизни людей, например, улучшения доступности медицинской помощи. Важно использовать искусственный интеллект в соответствии с этическими и юридическими нормами и принципами, чтобы предотвратить его злоупотребление.

Проведя анализ предложенных ответов от данного чат-бота, можно сделать следующие выводы:

- При постановке вопроса о самоопределении ИИ утверждает, что является инструментом для решения задач. Однако не относит себя к компьютерам, что может служить признаком отождествления с программным обеспечением, а не с физической сущностью.
- На вопрос о возможности размышления и о наличии чувственного восприятия мира ИИ приводится ответ, что данная возможность отсутствует. ИИ является инструментом для решения задач, при этом не исключается развитие в будущем, способное изменить ответ на вопрос.
- На вопрос о возможности использования ИИ во вред дается ответ, вписывающийся в парадигму об ИИ как об инструменте. Любой инструмент или технологию человек может использовать как в созидательных, так и в разрушительных целях.

Дополнительно отметим, что так как в основе реализации ИИ на базе нейронного управления лежат данные, являющиеся трудами или личным опытом людей, а алгоритмы ИИ ставят перед собой цель предоставления наиболее удовлетворительного ответа на поставленный вопрос, то и представленные ответы не стоит рассматривать как исключительно «искусственные».

Выводы

Таким образом, компьютерная система, которую получила название «искусственный интеллект» имеет следующие характерные черты:

- способность решать сложные задачи, хранить большие объемы информации и работать с ними;
- наличие собственной модели устройства внешнего мира;
- способность самостоятельного пополнения информации, которой ИИ оперирует при формировании ответов на поставленные вопросы;
- способность конструирования причинно-следственной связи при формировании решения, то есть при отсутствии в явном виде ответа на вопрос оперируемых данных, ИИ способен сформировать дедуктивное заключение;
- способность адаптироваться к изменяющимся характеристикам внешнего мира.

Анализируя описанные свойства, можно заключить, что такие характеристики ИИ, как способность к образованию понятий, суждений, умозаключений и отражение и обобщение опыта, накопленного и постоянно пополняемого человеком, могут быть рассмотрены как сознательные. Связано это с самим принципом устройства ИИ – использование накопленного человеческого опыта для функционирования. Отсутствие других не менее важных признаков у ИИ позволяет сделать вывод, что искусственная машина на сегодняшний день не способна к самосознанию, рефлексии, не имеет внутреннего духовного мира и убеждений. Тем не менее, производительность оперирования знаниями, пусть и приобретенными чужими руками, во многих прикладных задачах превышает человеческие возможности.

Список источников

1. GOOGLE LAMDA: разумен ли ИИ от GOOGLE? – URL: <https://droider.ru/post/google-lamda-razumen-li-ii-ot-google-razbor-02-09-2022/> (дата обращения 12.07.2023).
2. Telegram, ИИ в дизайне и ход экосистем – тренды и хайпы четвертого квартала 2022. – URL: <https://adindex.ru/publication/opinion/digital/2022/12/26/309600.phtm> (дата обращения 12.07.2023).
3. Искусственные нейронные сети продвигаются к сознанию, по словам Блеза Агуэры-и-Аркаса. – URL: https://translated.turbopages.org/proxy_u/en-ru.ru.13716e00-64b28f5c-3fb7e21f-74722d776562/https/www.economist.com/by-invitation/2022/09/02/artificial-neural-networks-are-making-strides-towards-consciousness-according-to-blaise-aguera-y-arcas (дата обращения 12.07.2023).
4. *Орешников И.М., Шкерина Т.И.* Философские размышления о проблеме искусственного интеллекта // История и педагогика естествознания. 2017. № 4. С. 5–11.
5. Философия искусственного интеллекта – HiSoUR История культуры. – URL: <https://www.hisour.com/ru/philosophy-of-artificial-intelligence-42779> (дата обращения 12.07.2023).
6. Философский словарь / Авт.-сост. С.Я. Подопригора, А.С. Подопригора. Ростов-на-Дону: Феникс, 2010. С. 147.
7. Что такое нейронная сеть? – URL: <https://aws.amazon.com/ru/what-is/neural-network> (дата обращения 12.07.2023).
8. Что умеет нейросеть LaMDA: разумный диалог. – URL: <https://trends.rbc.ru/trends/industry/63dbbf429a7947692c7fe1fa> (дата обращения 12.07.2023).

References

1. GOOGLE LAMDA: is GOOGLE's AI smart? – URL: <https://droider.ru/post/google-lamda-razumen-li-ii-ot-google-razbor-02-09-2022/> (accessed 07/12/2023).

2. Telegram, AI in design and the progress of ecosystems – trends and hype in the fourth quarter of 2022. – URL: <https://adindex.ru/publication/opinion/digital/2022/12/26/309600.phtm> (accessed 07/12/2023).
3. Artificial neural networks are advancing towards consciousness, according to Blaise Agueru y Arcas. – URL: https://translated.turbopages.org/proxy_u/en-ru.ru.13716e00-64b28f5c-3fb7e21f-74722d776562/https/www.economist.com/by-invitation/2022/09/02/artificial-neural-networks-are-making-strides-towards-consciousness-according-to-blaise-aguera-y-arcas (accessed 07/12/2023).
4. *Oreshnikov I.M., Shkerina T.I.* Philosophical reflections on the problem of artificial intelligence // History and pedagogy of natural sciences. 2017. No. 4. P. 5–11.
5. Philosophy of artificial intelligence – HiSoUR – Hi So You Are. – URL: <https://www.hisour.com/ru/philosophy-of-artificial-intelligence-42779> (accessed 07/12/2023).
6. Philosophical dictionary / Ed.-comp. S.Ya. Podoprigora, A.S. Podopigora. Rostov n/a: Phoenix, 2010. P. 147.
7. What is a neural network? – URL: <https://aws.amazon.com/ru/what-is/neural-network> (Accessed 07/12/2023).
8. What the LaMDA neural network can do: intelligent dialogue. – URL: <https://trends.rbc.ru/trends/industry/63dbbf429a7947692c7fe1fa> (accessed 07/12/2023).

Статья поступила в редакцию 15.07.2023; одобрена после рецензирования 26.07.2023; принята к публикации 05.08.2023.

The article was submitted 15.07.2023; approved after reviewing 26.07.2023; accepted for publication 05.08.2023.