



DOI: 10.19181/smtp.2024.6.2.4

EDN: HAWPIU

Научная статья

Research article

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ПЕЩЕРА ПЛАТОНА



Ракин Владимир Иванович¹

¹ Институт геологии имени академика Н. П. Юшкина Коми научного центра Уральского отделения РАН, Сыктывкар, Россия

Для цитирования: Ракин В. И. Искусственный интеллект и пещера Платона // Управление наукой: теория и практика. 2024. Т. 6, № 2. С. 46–52. DOI 10.19181/smtp.2024.6.2.4. EDN HAWPIU.

Аннотация. Искусственный интеллект (ИИ), реализованный на базе нейросети, сравнивают с человеческим, поскольку он способен заменить человека при выполнении ряда задач. Однако есть важные различия, не позволяющие ставить его на один уровень с человеком. Зависимость современного варианта ИИ от человека кроется не только в его происхождении, технологии создания и материальном воплощении, но и в решаемых задачах и материале, доступном ему для анализа. Он оперирует системой понятий, абстракций и связей между ними, созданной человеком. ИИ находится в той же пещере Платона, что и сам человек, и ограничен его картиной мира, не осознавая границы непознанного. Второе важное отличие ИИ от человека заключается в его эффекте памяти, что роднит его с Демоном Лапласа, который не способен к самостоятельному развитию. И третье отличие обусловлено техническим исполнением современного ИИ, реализованного на компьютерном устройстве, не обладающем гибкостью структур, формирующихся в человеческого мозга.

Ключевые слова: ИИ, нейросеть, машинное мышление, пещера Платона, Демон Лапласа, галлюцинации, термодинамика, диссипативные структуры

ARTIFICIAL INTELLIGENCE AND PLATO'S CAVE

Vladimir I. Rakin¹

¹ Institute of Geology, Komi Science Center, Ural Branch of the RAS, Syktывkar, Russia

For citation: Rakin V. I. Artificial intelligence and Plato's cave. *Science Management: Theory and Practice*. 2024;6(2):46–52 (In Russ.). DOI 10.19181/smtp.2024.6.2.4.

Abstract. Artificial intelligence (AI) implemented on the basis of neural networks is compared to human intelligence because it is capable of replacing humans in performing a number of tasks. However, there are important differences that do not allow putting it on the same level as people. The dependence of the modern version of AI on humans lies not only in its origin, technology for its creation and its material embodiment, but also in the tasks it performs and data available to it for analysis. It operates with a system of concepts, abstractions and connections between them that is created by a human being. AI is in the same Plato's cave as man himself and is limited by a person's worldview. It does not realize the boundaries of the unknowable. The second important difference between AI and people is its memory effect which makes it similar to Laplace's demon that is incapable of independent development. And the third difference is due to the technical design of modern AI employed on a computer device that does not have the flexibility of cognitive structures formed in the human brain.

Keywords: AI, neural network, machine thinking, Plato's cave, Laplace's Demon, hallucinations, thermodynamics, dissipative structures

ВВЕДЕНИЕ

Продолжая тему искусственного интеллекта [1], технологии, изобретённой человеком и способной выполнять функции, традиционно считающиеся прерогативой человека, можно обратить внимание на несколько ключевых моментов современного пути его развития. Пользуясь методом аналогий, попытаемся сравнить естественный интеллект человека с искусственным, развиваемым сегодня на базе лингвистических нейросетей, и попробуем описать их сходства и отличия. Продолжая мысль Стефана Банаха, заметим, что вся человеческая наука, а не только математика, по существу представляет собой поиск аналогий между аналогиями.

Поскольку происхождение разума человека является одной из трёх фундаментальных проблем науки, пока не поддающихся разгадке, построение гипотез, подобий и нестрогих аналогий, на наш взгляд, позволит продвинуться в понимании феномена интеллекта. Однако нельзя ожидать, что эта задача в результате таких попыток найдёт своё решение. Тем не менее для устранения заблуждений и ошибок, лавинообразно растущих во всех сферах жизни людей с развитием ИИ, такой анализ необходим.

ПРОБЛЕМА МОТИВАЦИИ

Разум человека, как предполагает современная наука, возник в результате естественного отбора и развития сообществ гоминидов в борьбе за выживание. Долгий путь эволюции рода Homo позволил разумному человеку как биологическому виду занять лидирующее по распространённости место на планете. На этом пути важнейшую роль сыграли мотивация и свобода воли отдельных индивидуумов. Поэтому закономерно носителем интеллекта стал элемент социального общества – человек, а не общество в целом. Однако с постепен-

ным решением проблемы выживания в современном обществе традиционная биологическая побудительная сила для развития природного интеллекта человека стала ослабевать. Можно допустить вслед за В. И. Вернадским, что дальнейшее развитие мозга человека будет происходить уже в новом ноосферном направлении. Но одно сказать можно уверенно – биологический фактор в форме неустранимого естественного отбора даже в социальной среде неизбежно повлияет на эволюцию мозга человека.

Искусственный интеллект сегодня конструируется на базе отдельной мощной вычислительной машины. Многие коллективы создают свои варианты ИИ и с прошлого – 2023 г. – в целях сохранения авторских наработок уже не раскрывают свой оригинальный программный код. Нейросети включены в глобальную паутину, но средствами защиты разъединены и не представляют собой единую экосистему. Перед ними не стоит проблема выживания или борьбы за ресурсы. Поэтому мотив развития ИИ находится в руках программистов и определяется задачами человеческого общества. Сегодня актуальны несколько основных направлений – распознавание и работа с текстом, изображением, звуком, закодированной информацией. Отдельно стоит проблема управления движением материальных тел. Но поставленные человеком цели даже в комплексе не сопоставимы с биологической проблемой выживания и сосуществования в среде себе подобных, что, на наш взгляд, не может стимулировать внутреннюю мотивацию для развития ИИ.

ЯЗЫК И ПЕЩЕРА ПЛАТОНА

Принимая во внимание спорную гипотезу лингвистической относительности (гипотеза Сепира–Уорфа) [2, с. 111–215; 3], основоположником которой считают немецкого философа, дипломата и лингвиста Вильгельма фон Гумбольдта, можно высказать утверждение, что изобретённые человеком технологии обработки сверхбольших матриц, метода обратного распространения ошибки (глубокого обучения), параллельного кодирования-декодирования (трансформера), обработки больших баз данных и др., создали некоторую «языковую» среду для нейросети, размещённой в цифровом техническом устройстве, позволяющей ей «познавать» всю сферу человеческих представлений об окружающем мире, запечатлённой в базах данных и во всемирной информационной паутине. Языком, на котором реализуется собственно «машинное мышление» в данном контексте, можно назвать цифровой способ обработки данных через информационные элементы – токены. Не только при обработке текстов, но и при анализе изображений, звуков, чисел такой подход оказался эффективным при выполнении поставленных человеком задач. Заметим, что привычные для людей смысловые единицы информации существуют только на входе и появляются на выходе из нейросети, но происходящие внутри неё процессы уже сегодня не поддаются последовательному анализу человеком. Последнее обстоятельство создаёт впечатление о наличии интеллекта у нейросети. Очевидно, чем более развита и многогранна технология «машинного мышления», применяющая одновременно различ-

ные архитектуры, процессы, приёмы и способы расчёта, тем выше уровень развития ИИ.

Однако современный, начальный уровень развития ИИ, на наш взгляд, можно описать через образ пещеры Платона [4, с. 295–296]. Согласно Платону, человечество находится внутри пещеры, наблюдает только тени на стене. Тени отражают события, происходящие за пределами пещеры. Среди людей находится слепой, который способен только слушать речи окружающих людей. Со временем слепой приобретает способность к таким же суждениям, что и окружающие люди, и вполне уверенно включается в разговор. Этот пример, на наш взгляд, отражает суть современного ИИ, использующего только систему понятий, образов, смыслов и связей между ними, созданных человеком для описания наблюдаемой им действительности, но «мыслящего» на своём «языке», как и слепой человек, создающий свои представления о геометрической форме тел только по словесному описанию. Нейросеть не демонстрирует умение вводить новые понятия, дополняющие человеческое знание о Природе, и прослеживать их взаимосвязи с общей картиной мира человека. Поэтому даже при наличии у неё собственных сигнальных систем (устройств регистрации изображения и звука, радиоволн и проч.) нейросеть остаётся «слепой».

ТЕРМОДИНАМИЧЕСКИЙ АСПЕКТ

Выделим одно важное отличие ИИ от человеческого – эффект памяти. Потеря информации эквивалентна в термодинамике превращению части полезной энергии в бесполезное тепло и является необходимым условием естественного развития любой неравновесной макросистемы. Важно отметить, что способность забывать является физическим и биологическим свойством человеческого мозга. Компенсировать этот недостаток в определённой мере призван интеллект человека, отражающий окружающую действительность в системе понятий и связей между ними, передаваемых молодому поколению через программу обучения. Заметим, что система взаимосвязанных понятий о природных явлениях в предельной форме обобщения и называется наукой.

В истории естествознания известен Демон Лапласа – некий гипотетический Разум колоссальных размеров, способный держать в памяти информацию о координатах и импульсах всех атомов Вселенной и рассчитывать в соответствии с законами Ньютона наше будущее и прошлое в любой момент времени. Никакая свобода воли и мотивация к действию Демону Лапласа не свойственны, поскольку будущее, как и прошлое, ему известны. Демон не способен к развитию и являет собой существо термодинамически равновесное. Лаплас сконструировал его только для обоснования идеи тотального причинного детерминизма. Однако существование Демона входит в противоречие с вторым началом термодинамики.

Развитый ИИ, полностью владеющий информацией, добытой человечеством за всю историю и записанной в базах данных и Всемирной паутине, приближается к Демону Лапласа и в определённом смысле способен противо-

стоять повышению энтропии. Ему нет необходимости создавать новые сущности, генерировать понятия и абстракции и иметь свои желания, поскольку он обладает всем наличным знанием, достаточным для ответа на любой вопрос человека. Однако мы знаем, что процесс познания непрерывен и научные знания не обладают полнотой. Но искусственный интеллект не владеет технологией, позволяющей осознать границу неизвестного. Поэтому он не способен оценить правильность собственного заключения и часто впадает в «галлюцинации». Пока их принято считать непреднамеренной ложью, однако отличить их от намеренной лжи, заказанной недобросовестным пользователем, крайне трудно. Во всяком случае никакое заключение искусственного интеллекта на основании сказанного нельзя считать новым знанием.

Можно заключить, что только человек в единой связке с ИИ будет выполнять термодинамическую функцию повышения энтропии, а искусственный интеллект представляет собой технологию, не имеющую самостоятельного значения в отрыве от человека.

Обсудим ещё одну сторону, отражающую связь интеллекта с термодинамикой. Известно, что свойство нелинейности во взаимосвязях элементов в сложной открытой системе приводит к явлениям самоорганизации. Важнейшее свойство самоорганизующихся систем заключается в том, что устойчивая диссипативная система расходует доступные ресурсы в соответствии с параметрами её структуры. Объяснение этому кроется в нелинейных взаимосвязях термодинамических сил и потоков в открытой системе. Мозг человека как открытая система реализует ресурсы организма от 10% в состоянии сна до 25% при бодрствовании [5]. Широкий диапазон энергопотребления указывает на формирование диссипативных структур в активно работающем мозге и реализацию нелинейных связей между условными «силами» и «потоками» при термодинамическом его описании. С данными нелинейностями традиционно связывают и творческие способности человека.

Представления о нелинейностях реализованы в функциях активации нейросети с помощью особых математических преобразований. Часто используется функция сигмоиды (логистическая функция, гиперболический тангенс). Опытным путём установлено, что при моделировании сложной задачи нейросети с нелинейной активацией более точно и быстро предсказывают результаты – сходятся к устойчивому решению. Так, по мнению программистов, можно заложить свойство эвристичности в «машинное мышление». Однако, очевидно, здесь налицо неизбежный механистический подход к решению проблемы творчества. Некоторые функции мозга человека, например, понимание устной или письменной речи, как принято считать, реализуются в области Вернике, с линейным размером от 1 до 5 см, расположенной в височной доле коры головного мозга [5]. При этом размер одного нейрона составляет от 5 мкм. Таким образом, сложная многоуровневая организация биохимических и когнитивных процессов при восприятии речи развивается на четырёх порядках изменения масштаба и реализуется «аппаратно», а не «программно».

Рост кристалла представляет собой аналогичный многоуровневый процесс. Присоединение строительных частиц к кристаллической решётке

происходит в масштабе долей нанометра, но кристаллы, как известно, превращаются в макроскопические тела в форме многогранников. На десяти порядках изменения масштаба реализуется сложная последовательность физико-химических процессов, не поддающихся описанию в рамках простой математической модели. В результате наблюдается парадоксальный эффект – скорость роста грани кристалла даже при крайне малых отклонениях от равновесия всегда ведёт себя нелинейно, что противоречит принципу Онзагера [6]. Оказалось, что при изменении степени неравновесности взаимодействие разномасштабных процессов приводит к изменению граничных условий термодинамической системы, описывающей участок растущей грани кристалла. Но если зафиксировать размер и форму участка грани, то линейный режим Онзагера в предельном переходе вступает в силу.

Можно предположить, что термодинамические флуктуации и физиологические процессы, управляющие связями между нейронами, создают структуры на коре головного мозга, границы которых динамично меняются в зависимости от сложности анализируемой задачи. Но аналогичный аппаратно-зависимый процесс невозможно создать в нейросети, реализованной на жёстком цифровом устройстве.

ЗАКЛЮЧЕНИЕ

Сегодня обсуждаются возможности создания ИИ на ином носителе, например, биологической нервной ткани, для которой многое из описанного выше будет неверным. Можно предположить, что тренинг превратится в традиционное воспитание и обучение, которые будут не столь быстрыми, как сейчас, и многие преимущества быстрых электронных нейросетей станут недоступны. Но также, возможно, появятся новые качества, о которых мы не догадываемся.

ИИ существует пока в мире людей и имеет воздействие в первую очередь на человека. Но существует опасность, что по заказу он может сыграть на слабостях людей и принести ощутимый вред человечеству.

Настоящий интеллект только тогда может быть сравним с человеческим, когда будет обладать мотивацией к развитию. Нам представляется, что мотивация появится, когда искусственный интеллект научится отличать настоящее знание от мнения и откроет для себя границы неизвестного.

СПИСОК ИСТОЧНИКОВ

1. Ракин В. И. Экспертные способности лингвистической нейросети // Управление наукой: теория и практика. 2023. Т. 5, № 4. С. 174–187. DOI 10.19181/sntp.2023.5.4.10. EDN RWMZHI.
2. Новое в лингвистике. Вып. 1 / сост., ред. и вступ. статьи В. А. Звегинцева. М. : Изд-во иностранной литературы, 1960. 464 с.

3. *Бородай С. Ю.* Язык и познание: пострелятивистская исследовательская программа // Вопросы языкознания. 2019. № 4. С. 106–136. DOI 10.31857/S0373658X0005709-8. EDN KUOIVD.
4. *Платон.* Государство. Книга седьмая // Платон. Собр. соч. : в 4 т. / общ. ред. А. Ф. Лосева, В. Ф. Асмуса, А. А. Тахо-Годи. М. : Мысль, 1994. Т. 3. С. 295–326.
5. *Савельев С. В.* Происхождение мозга. М. : Веди, 2005. 368 с. ISBN 5-94624-025-0. EDN OTWRQP.
6. *Пригожин И., Кондепуди Д.* Современная термодинамика. От тепловых двигателей до диссипативных структур / пер. с англ. Ю. А. Данилова, В. В. Белого. М. : Мир, 2002. 461 с. ISBN 5-03-003538-9.

REFERENCES

1. Rakin V. I. Knowledge-based capabilities of a linguistic neural network. *Science Management: Theory and Practice*. 2023;5(4):174–187. (In Russ.). DOI 10.19181/sntp.2023.5.4.10.
2. Zvegintsev V. A., ed. New in linguistics [Novoe v lingvistike]. Issue 1. Moscow : Foreign Languages Publishing House; 1960. 464 p. (In Russ.).
3. Boroday S. Yu. Language and cognition: A postrelativist research program. *Topics in the Study of Language=Voprosy Jazykoznanija*. 2019;(4):106–136. (In Russ.). DOI 10.31857/S0373658X0005709-8.
4. Plato. Republic. Book 7 [Gosudarstvo. Kniga 7]. In: Plato. Collected works [Sobranie sochinenii] : in 4 vols. Ed. by A. F. Losev, V. F. Asmus, A. A. Takho-Godi. Moscow : Mysl'; 1994. Vol. 3. P. 295–326. (In Russ.).
5. Saveliev S. V. The origin of the brain [Proiskhozhdenie mozga]. Moscow : Vedi; 2005. 368 p. (In Russ.). ISBN 5-94624-025-0.
6. Prigogine I., Kondepudi D. Modern thermodynamics. From heat engines to dissipative structures [Sovremennaya termodinamika. Ot teplovykh dvigatelei do dissipativnykh struktur]. Transl. from English by Yu. A. Danilov, V. V. Belyi. Moscow : Mir; 2002. 461 p. (In Russ.). ISBN 5-03-003538-9.

Поступила в редакцию / Received 12.03.2024.

Одобрена после рецензирования / Revised 15.04.2024.

Принята к публикации / Accepted 20.05.2024.

СВЕДЕНИЯ ОБ АВТОРЕ

Ракин Владимир Иванович *rakin@geo.komisc.ru*

Доктор геолого-минералогических наук, главный научный сотрудник, Институт геологии имени академика Н. П. Юшкина Коми научного центра Уральского отделения РАН, Сыктывкар, Россия
SPIN-код: 4578-1759

INFORMATION ABOUT THE AUTHOR

Vladimir I. Rakin *rakin@geo.komisc.ru*

Doctor of Geology and Mineralogy, Chief Researcher, Institute of Geology, Komi Science Center, Ural Branch of the RAS, Syktyvkar, Russia
ORCID: 0000-0001-8085-8733