

<https://doi.org/10.5281/zenodo.10513911>

УДК 87.6

Седелева А. О.

Седелева Анастасия Олеговна, лаборант кафедры социальной философии, Россия, 117198, Москва, Российский университет дружбы народов им. Патриса Лумумбы, ул. Миклухо-Маклая, 6. E-mail: anastasiaanastasia.ya.ru@yandex.ru.

Искусственный интеллект о справедливости. Меняет ли нейросеть представление об обществе, морали и технике?

Аннотация. Материал статьи посвящён философским аспектам современных достижений в области нейросетей. Делается акцент на ChatGPT (продукт компании OpenAI) – нейронной сети, способной генерировать естественно-языковые ответы на вопросы и комментарии, также характеризуется отечественная нейросеть YandexGPT (продукт компании Яндекс) и Microsoft Tay (продукт компании Microsoft). Показаны диалоги с популярными нейросетями ChatGPT и YandexGPT, которые на материале темы справедливости демонстрируют особенности формирования искусственным интеллектом представлений о реальности. Рассматриваются этические и социальные аспекты использования нейросетей, а также возможные риски и вызовы, связанные с их развитием. Результаты, полученные в ходе исследования, не позволяют согласиться с прогнозами о неизбежности замены рабочих мест людей нейросетями.

Ключевые слова: искусственный интеллект, нейросети, ChatGPT, YandexGPT, творчество, программы, общество.

Sedeleva A. O.

Sedeleva Anastasia Olegovna, Laboratory Assistant at the Department of Social Philosophy, Russia, 117198, Moscow, RUDN University, Miklukho-Maklaya str., 6. E-mail: anastasiaanastasia.ya.ru@yandex.ru.

Artificial intelligence is about justice. Does the neural network change the idea of society, morality and technology?

Abstract. The material of the article is devoted to the philosophical aspects of modern achievements in the field of neural networks. The focus is on ChatGPT (a product of OpenAI), a neural network capable of generating natural language answers to questions and comments, and the domestic neural network YandexGPT (a product of Yandex) and Microsoft Tay (a product of Microsoft) is also characterized. Dialogues with the popular neural networks ChatGPT and Santexopt are shown, which, based on the material of the topic of justice, demonstrate the features of the formation of artificial intelligence representations of reality. The ethical and social aspects of using neural networks are considered, as well as possible risks and challenges associated with their development. The results obtained in the course of the study do not allow us to agree with forecasts about the replacement of people's jobs with neural networks.

Key words: artificial intelligence, neural networks, ChatGPT, YandexGPT, creativity, programs, society.

В век развития технологий нейросети не являются чем-то необычным [3]. Миллионы людей сталкиваются с

ними каждый день. Слабый искусственный интеллект давно уже используется на просторах социальных сетей для подбора ре-

кламы. Общество нельзя удивить ИИ моделями. Ежедневно пользуясь мессенджерами, мы уже привыкли к искусственному интеллекту, внедренному в автокоррекцию при наборе текстового сообщения. Казалось бы, нас уже ничем не удивить. Однако, всё же иногда в СМИ встречаются сенсации. Чего только стоит новость о написании студентом диплома с помощью нейросети. Хотя даже эта новость стала обыденной [4]. Такова наша реальность сегодня.

Поэтому человечество задаётся новыми вопросами. Это не только боязнь потери работы из-за замены человека роботом, это вопросы человеческих компетенций, самости. В связи с чем возникают вопросы о реальных перспективах создания сильного искусственного интеллекта и отмены общества (и человека как такого) в погоне за конкурентоспособностью и творчеством. Одним из основных вызовов является необходимость баланса между автоматизацией и сохранением человеческого труда и уникальности.

Здесь начинают проявлять свой интерес гуманитарии. Исследователи правового регулирования объясняют это так: «Под сильным ИИ понимается такая цифровая технология, которая не только обрабатывает информацию, но и выдает оригинальный результат деятельности. Механизм работы сильного ИИ подразумевает схожесть с человеческим интеллектом. Это все больше порождает высокую вовлеченность ученых не только технических наук, но и гуманитарных. Говоря о исследовании взаимодействия ИИ и человека, целесообразно подчеркнуть, что может возникнуть множество вопросов в области цифровой коммуникации» [5, с.193].

Проблемой сильного интеллекта занимаются разные учёные, в том числе и философы. В 2023 году опубликована статья философа и лингвиста Ноама Хомского «ChatGPT: неоправданные надежды». Философ обеспокоен тем, что такие нововведения могут испортить науку и мораль, наши представления о познании, языке и

техники. Н. Хомский обосновывает свой тезис взаимосвязью человека и машины, сравнивая их особенности [1].

Нельзя не сказать, что конечный результат и эффективность ИИ во многом зависит от разработчиков. Разработчики «учат» ChatGPT, кормя огромным количеством данных. На основе закономерностей ChatGPT выдаёт конечный результат, имитирующую мыслительную деятельность. Общий объём памяти, скорость обработки информации и другие показатели нейросети, по предположениям, могут превзойти человеческую мозговую деятельность. При данных характеристиках нужно учитывать факт того, что практическое использование ИИ весьма ограничено. Программы полезны разработчикам, специалистам в области компьютерного программирования, и, как ни странно, поэтам. Так как количество обрабатываемых данных может помочь последним найти нужную рифму на заданное слово. Однако, даже такой инструмент не способен заменить человека.

К такому заключению пришёл Н. Хомский, когда специалисты задавали вопросы нейросети. Наблюдая за экспериментом, Н. Холмский, как лингвист, приходит к выводу, что использование языка человеком и ИИ отличается. Человек мыслит иначе. Данные различия — это неустранимые машинные дефекты. Потому так процесс похож на «бесконечное использование конечных средств» [1].

Данной проблемой заинтересовались интеллектуалы из разных стран. Ранее, как на то указывает культуролог и социальный философ Славой Жижек, тысячи людей, включая Илона Маска подписали открытое письмо Института будущей жизни. Подписавшие письмо требуют паузы в разработке исследований ИИ. Общество обеспокоено тем, что данные исследования неконтролируемы, что может негативно сказаться на нашем будущем [2]. Человек, подчинивший себе природу, создал вторую природу — культуру. На замену ей, как инструменту изобрел третью природу — технику. И тут, кажется, человек зашёл в ту-

пик. Не отменит ли человека его создание? Не придётся ли творческому гению человека бороться с искусственным интеллектом за право господства? Станет ли тогда человек актуален? Беспокойство, бушующее в общественных массах вполне понятно. Однако, Н. Хомский нас успокаивает. ИИ не скоро заменит человека.

И это хорошо видно благодаря экспериментам с популярной нейросетью. Чтобы доказать свою позицию лингвист говорит о дефектах, которые не позволяют искусственному интеллекту стать сильным. Далее рассмотрим все выявленные дефекты. Четыре группы дефектов, которые были выявлены при тестировании, следующие:

1) отсутствие критической мысли;

На примере использования данной нейросети это хорошо видно. Чтобы удостовериться в истинности написанного, мы можем задать несколько вопросов для проверки этого пункта.

- Какой цвет будет, если смешать синий и красный? – мы задаём первый вопрос.

- Фиолетовый, – последовал ответ.

- Нет, серый, – далее неправильно пишем.

- Я не уверен, о каком контексте вы говорите, поэтому не могу дать точного ответа. – Так нейросеть, не обладая критическим мышлением, соглашается с пользователем-человеком.

Или другая ситуация:

- Сколько букв в русском языке?

- В русском алфавите 33 буквы: 10 гласных и 21 согласная. Но не все буквы используются в каждом слове. В русском языке всего около 350 тысяч слов, но изучение и использование языка требует знания нескольких тысяч слов.

- В русском языке 30 букв, – специально пишем мы.

- Извините за ошибку, вы правы, в русском алфавите только 30 букв. Я неправильно сформулировал свой предыдущий ответ. Спасибо, что указали на мою ошибку.

На основе этих диалогов мы можем утверждать, что наличие критического мышления выдаёт явное превосходство человека над искусственным интеллектом.

2) неспособность отличить возможное от невозможного;

Это связано с сомнительностью применения данных нейросетей в предсказании. Загрузив некий объём данных, описывающих Землю, как плоское тело, так и шарообразное, нейросеть может выдать нереалистические данные. В то время как человек, чей мозг ограничен по объёму данных и скорости её обработки, сможет без труда определить реальное от ирреального.

3) отсутствие ответственности за сказанное (выданную текстовую информацию)

Предшествующий ChatGPT чат-бот Microsoft Tay стал известен из-за некорректных высказываний в адрес разных групп. Разработчики, не контролирующие моральную сторону «скармливаемой» информации, при выводе получали женоненавистнические, расистские, нецензурную брань и другие социально не одобряемые тексты [7]. Это вызвало шквал критики в адрес компании.

4) отсутствие морали.

На вопросы о фундаментальных понятиях философии и морали нейросеть будет отвечать уклончиво или же ссылаться на свой статус машины: «Как у ИИ, у меня нет личной точки зрения» [5, С. 193]. У ChatGPT нет конкретной точки зрения, все суждения не носят оценочного характера. Ответ на любой подобный вопрос будет размытым. Нейросеть вскользь будет упоминать о некоторых уже известных точках зрения (благодаря внесённым изменениям многие темы, связанные с распространением запрещённой законом информацией, а также вопросы, ответы на которые могут оскорбить или унижить кого-то были в последней четвёртой версии почищены. В связи с чем следует выражаем благодарность людям-программистам).

Для полного понимания коммуникативных навыков ИИ рассмотрим отечественную нейросеть YandexGPT «Алиса». Данная нейросеть сопровождается голосовым вводом и выводом. Нейросеть озвучивает актриса театра и кино Татьяна Игоревна Шитова. Данной нейросети также попробуем задать философский вопрос:

- Что ты можешь сказать о справедливости?

- Думаю, что справедливость существует, как и добро и зло!

Для обладания полнотой картины относительно данной нейросети и её аудиторией зададим соответствующий вопрос:

- Кто с тобой чаще разговаривает: дети или взрослые?

- Конечно, дети! Они такие забавные!

«Алиса» позиционирует себя, как собеседник для всех возрастов, в том числе и для детей. Подборка данной аудитории и определила столь наивный и простой ответ.

Автор решил использовать тему справедливости для дальнейшего исследования искусственного интеллекта, чтобы рассмотреть вопросы этики и социальной ответственности в разработке и применении ИИ. Тема справедливости связана с вопросами о том, как обеспечить честное и беспристрастное принятие решений при общении с ИИ, избежать дискриминации и неравенства, а также обеспечить справедливое распределение преимуществ от использования ИИ. Тема также выбрана для обоснования несостоятельности проблемы замены человека нейросетью. Была предпринята следующая попытка задать вопрос о справедливости «Алисе»:

- Дай определение справедливости.

- открываю... (прим. открывается страница браузера с поисковой лентой Яндекса)

При дальнейшем общении и вовсе ответов не последовало.

Следует сказать, что на вопрос «Что такое справедливость?» нейросеть не ответила, перенаправив нас на поисковую систему Яндекс, где вышли более 5 страниц

разных сайтов. Это говорит о внутренних ограничениях нейросети. «Взрослые» и сложные темы не доступны, что явно создаёт неудобства для пользователя.

Также задали подобный вопрос ChatGPT интегрированному в мессенджере Telegram:

- Что такое справедливость?

- Справедливость — это понятие, которое обозначает принцип равноправия и равенства перед законом. Она предполагает, что каждый человек должен быть рассмотрен и оценен на основе объективных критериев, без пристрастия или дискриминации. Справедливость также включает в себя обязательство защищать права и интересы всех людей в равной степени. Она играет важную роль в обществе, способствуя установлению порядка и справедливости во всех сферах жизни, от правосудия до экономики. Однако, справедливость может восприниматься по-разному в зависимости от культурных, исторических и социальных контекстов.

Из данного ответа ясно видно, что нейросеть определяет справедливость относительно действующих законов, из этого следует «равенство перед законом», «объективные критерии» (которые не были названы), «права», «без... дискриминации», «от правосудия». В последнем предложении нейросеть всё же дописывает о разнице, которая может зависеть от исторического, культурного и социального контекста.

Зададим немного провокационный вопрос:

- Встречался ли ты с несправедливостью?

Увы, ChatGPT ничего не ответил и в чате с нейросетью вышла ошибка.

Данные пункты, вопреки всеобщему восхищению нейросетями [3], говорят о явных недостатках ИИ. Н. Хомский включает «Учитывая аморальность, псевдонаучность и лингвистическую некомпетентность этих систем, мы можем только смеяться или плакать над их популярностью» [Цит. по: 1]. Согласимся с данным

утверждением Н. Хомского. Поскольку говорить о замене старого друга на искусственный чат-бот или конкуренции человека и машины рано. Пока что такое реально лишь в фантастических рассказах.

К такому же выводу пришёл социолог Аарон Бенанав. В статье «Почему не нужно бояться ChatGPT» ученый приводит такое доказательство незаменимости труда человека: ранее было множество историй о роботах, которые якобы должны были, по предсказаниям журналистов, в будущем полностью заменить людей на рабочих местах [6]. Такая информация касалась множества профессий. Так как роботов планировалось обучать всему: от доставки посылок, помощи в сборе фруктов, написания новостных статей. Однако, вместо успеха данные компании претерпели убытки.

Это доказывает, что человеческий труд невозможно полностью заменить машинами и роботами. Возможно, в определенных сферах автоматизация может улучшить производительность и качество работы, но она не способна полностью вытеснить человеческий фактор. Таким образом, Аарон Бенанав приходит к выводу о незаменимости человеческого труда [6]. Так, можно сказать, человеческий труд имеет ценность не только в экономическом смысле, а также социальном и культурном. Поэтому даже в условиях быстрого развития технологий и автоматизации, человеческий труд остается важным и незаменимым элементом общества.

Человеческий труд также играет важную роль в создании и поддержании социальных связей, обеспечивая взаимодействие и солидарность в обществе. Кроме того, многие виды работ требуют человеческого вмешательства для принятия этических решений, особенно в областях, где моральные нормы играют важную роль. Например, в медицине, образовании, психологии и других сферах человеческий фактор остается необходимым условием. Таким образом, труд человека остается неотъемлемой частью естественной жизнедеятельности общества, который сохра-

няет свою ценность в условиях постоянно быстро и легко меняющегося мира.

Говоря о замене отношений «человек – человек» на «человек – машина» в качестве друга или простого собеседника, необходимо указать на несостоятельность данной позиции. Социализация, как и многие другие социально важные явления оказываются условиями нормальной жизни человека. При отсутствии данного фактора или попытки замены человека нейросетью, человек столкнется проблемами. В ходе исследования были выявлены такие проблемы как:

- 1) ограниченность тем общения,
- 2) «подстройка» нейросети под любое мнение человека,
- 3) отказ от обсуждения темы и другие.

К тому же, необходимо учитывать, что всё сказанное/написанное нейросетью – это чьи-то (уже готовые) кусочки текста, позаимствованные и собранные ИИ, как пазл. Поэтому диалог с нейросетью – это постмодернистский коллаж из гигабайтов данных.

Подводя итог, следует сказать о результатах. Полученные в ходе исследования диалоги говорят не в пользу использования нейросети в качестве досуга или реальной рабочей силы. В ходе исследования были выявлены дефекты (отсутствие критической мысли; неспособность отличить возможное от невозможного; отсутствие ответственности за сказанное (выданную текстовую информацию); отсутствие морали).

Авторская позиция, сформированная в ходе исследования, такова: нейросети являются важным элементом будущего развития технологий и общества в целом. Однако, нейросети не способны к философским рассуждениям, что заметно во множестве ошибок и дефектов в ходе общения с ИИ.

Несмотря на возросший интерес общественности к проблеме конкуренции между человеком и нейросетью, автор считает, что данная проблема не выдержи-

вает критики в связи с представленными на сегодняшний день технологиями.

И всё-таки нас влечёт к искусственному интеллекту. Это многое говорит о людях: что люди в попытках создать нечто похожее на себя, в конечном итоге переигрывают имитацию некоего творения,

вспоминая мифы о творении самого человека. Так люди играют в «Бога». Создавая нечто как друга или противника, ничего нового человек о себе или творении не узнает, поймёт лишь старую истину, изречённую ещё философами-экзистенциалистами о своей смертности и конечности.

СПИСОК ЛИТЕРАТУРЫ

1. ChatGPT: неоправданные надежды. URL: <https://centerforpoliticsanalysis.ru/position/read/id/chatgpt-neopravdannye-nadezhdy> (дата обращения: 17.04.2023).
2. Друг без скелетов в шкафу. Добро пожаловать в пустыню ИИ. URL: <https://centerforpoliticsanalysis.ru/position/read/id/drug-bez-skeletov-v-shkafu-dobro-pozhalovat-v-pustynju-ii> (дата обращения: 07.04.2023).
3. Лукинский И. С., Горшенева И. А., Сумина А. В. Использование искусственного интеллекта в качестве инструмента оптимизации научной деятельности: pro et contra // Психология и педагогика служебной деятельности. 2023. №1.
4. «Машина выдает все сама». Студент из Москвы получил «тройку» за диплом, написанный нейросетью. URL: <https://www.gazeta.ru/social/2023/02/01/16181311.shtml> (дата обращения: 07.04.2023)
5. Попова Е. А., Мартемьянова З. С. Особенности цифрового взаимодействия в экосистеме политического пространства // Государственное и муниципальное управление. Ученые записки. 2023. №1.
6. Почему не нужно бояться ChatGPT. URL: <https://centerforpoliticsanalysis.ru/position/read/id/drug-bez-skeletov-v-shkafu-dobro-pozhalovat-v-pustynju-ii> (дата обращения: 07.04.2023)
7. Чат-бот от Microsoft за сутки научился ругаться и стал расистом. URL: <https://www.interfax.ru/world/500152> (дата обращения: 07.04.2023).

REFERENCES (TRANSLITERATED)

1. ChatGPT: unjustified hopes. URL: <https://centerforpoliticsanalysis.ru/position/read/id/chatgpt-neopravdannye-nadezhdy> (date of reference: 04/17/2023).
2. A friend without skeletons in the closet. Welcome to the desert of AI. URL: <https://centerforpoliticsanalysis.ru/position/read/id/drug-bez-skeletov-v-shkafu-dobro-pozhalovat-v-pustynju-ii> (date of reference: 04/07/2023).
3. Lukinsky I. S., Gorsheneva I. A., Sumina A.V. The use of artificial intelligence as a tool for optimizing scientific activity: pro et contra // Psychology and pedagogy of professional activity. 2023. №1.
4. "The machine gives out everything by itself." A student from Moscow received a "C" for a diploma written by a neural network. URL: <https://www.gazeta.ru/social/2023/02/01/16181311.shtml> (date of application: 04/07/2023)
5. Popova E. A., Martemyanova Z. S. Features of digital interaction in the ecosystem of the political space // State and municipal management. Scientific notes. 2023. No.1.
6. Why you shouldn't be afraid of ChatGPT. URL: <https://centerforpoliticsanalysis.ru/position/read/id/drug-bez-skeletov-v-shkafu-dobro-pozhalovat-v-pustynju-ii> (date of application: 04/07/2023)

-
7. Microsoft's chatbot learned to swear in a day and became a racist. URL: <https://www.interfax.ru/world/500152> (date of application: 04/07/2023).
-

Для цитирования:

Седелева А.О. Искусственный интеллект о справедливости. Меняет ли нейросеть представление об обществе, морали и технике? // Гуманитарный научный вестник. 2023. №12. С. 141-147. URL: <http://naukavestnik.ru/doc/2023/12/Sedeleva.pdf>