
Соколова М.Е.*

**CHATGPT И ПРОМПТ-ИНЖИНИРИНГ:
О ПЕРСПЕКТИВАХ ВНЕДРЕНИЯ ГЕНЕРАТИВНЫХ
НЕЙРОСЕТЕЙ В НАУКЕ**

Sokolova M. E.

**CHATGPT AND PROMPT ENGINEERING:
ON THE PROSPECTS FOR THE INTRODUCTION
OF GENERATIVE NEURAL NETWORKS IN SCIENCE**

Аннотация. В статье рассматривается перспектива формирования трудового рынка промпт-инженеров – специалистов по составлению поисковых запросов, системно работающих с генеративными нейронными сетями, в частности с ChatGPT. Обосновывается, что промпт-инжиниринг обещает стать расширением профессий Data Scientist и Machine Learning Engineer. Показывается, как на волне хайпа генеративных нейронных сетей, и прежде всего ChatGPT, для этого сейчас готовятся кадры промпт-инженеров. Наличие таких специалистов с высоким уровнем квалификации и широким набором навыков должно помочь сделать рывок в безопасном внедрении интеллектуальных помощников – генеративных нейросетей в науке и в сфере информационного обеспечения / поддержки науки. Подчеркивается актуальность этих вопросов для

* © Соколова Марианна Евгеньевна – кандидат философских наук, старший научный сотрудник Института США и Канады им. академика Г.А. Арбатова Российской академии наук (ИСКРАН), Москва, РФ; mariamva@yandex.ru

ORCID0000-0002-4899-9058

Marianna Ye. Sokolova – PhD in Philosophy, Senior Researcher, G.A. Arbatov Institute of the United States and Canada, Russian Academy of Sciences (ISKРАН), Moscow, Russian Federation; mariamva@yandex.ru

ORCID0000-0002-4899-9058

научного сообщества и создателей современных русскоязычных генеративных нейросетей, в том числе ориентированных на научный контент.

Ключевые слова: генеративные нейронные сети; промпт-инжиниринг; ChatGPT; искусственный интеллект в науке.

Abstract. The article considers the prospect of forming a labor market for prompt engineers – specialists in compiling search queries, systematically working with generative neural networks, in particular with ChatGPT. It is proved that prompt engineering promises to become an extension of the professions of Data Scientist and Machine Learning Engineer. It shows how in the wake of the hype of generative neural networks and above all ChatGPT prompt engineers are currently being trained for this.

The presence of such specialists with a high level of qualification and a wide range of skills should help make a breakthrough in the safe implementation of intelligent assistants – generative neural networks in science and the field of information support / science support. The relevance of these issues for the scientific community and the creators of modern Russian-language neural networks, including those focused on scientific content, is emphasized.

Keywords: generative neural networks; prompt engineering; ChatGPT; artificial intelligence in science.

Введение

Генеративные нейросети как очередной этап развития и внедрения искусственного интеллекта (далее – ИИ) вызывают пристальный интерес не только у специалистов. Как считают эксперты, на этом фоне языковая модель ChatGPT компании OpenAI стала в последние два года для многих чем-то вроде проявления магии [Three AI Researchers Look to the Challenges Ahead in 2024, 2024].

На фоне всё прогрессирующего применения этой и других генеративных нейронных сетей (Gemini / Bard (Google), Claude (Anthropic), Midjourney (Midjourney), Kandinsky (Сбербанк) и т.д.) в обыденной жизни, ИТ, бизнесе и маркетинге, медиа / журналистике и других областях учёные, которые также используют

нейросети в своей профессиональной деятельности, проявляют вполне естественный интерес к их возможностям – а это, как общеизвестно, и огромные плюсы, и безусловные минусы, если можно называть «минусами» неизбежные особенности и издержки использования технологических сервисов, на совершенствование которых возлагаются, и не без оснований, большие надежды.

GenAI и человеческий фактор

Эти возможности генеративных нейросетей уже достаточно проявили себя во многих других областях. Так, возможность генерировать изображения основательно трансформировала труд дизайнеров в архитектуре, промышленности, моде. Например, нейросети, и прежде всего так востребованная сегодня Midjourney, с большим успехом и высоким качеством способны выполнять дизайнерские функции на этапе подготовки эскизов моделей одежды, принтов, орнамента. Обрабатывать же эти заготовки непосредственно под потребности клиента будет уже ведущий дизайнер, что, безусловно, экономически очень эффективно. Таким образом, сейчас сегмент занятости сузился именно для дизайнеров более низкой квалификации, в то время как роль ведущих дизайнеров, наоборот, стала более значительной и важной¹.

В журналистике / медиасфере возможности GenAI по обработке и созданию фото, видео и текстов также всё больше входят в профессиональный обиход. Это и написание новостей, дописывание авторских текстов, сбор и обработка фактов, информации, перевод, редактирование, расшифровка аудио, генерация и обработка фото и видео и проч. Однако – в результате всё более широкого их распространения в повседневной рабочей практике – журналисты приходят к выводу, что использование нейросетей без контроля и редактуры со стороны профессионала непродуктивно. Например, согласно данным студенческого исследования по направлению «Журналистика» в Тамбовском государственном университете имени Г.Р. Державина в апреле 2023 г., хотя нейросети и способны

¹ Будущее уже наступило. Как технологии меняют индустрию моды // BRICS+ Fashion Summit. 28.11.2023–2.12.2023. – URL: <https://fashionsummit.org/?ysclid=lsq0ispzpw264745806> (дата обращения: 11.02.2023).

вполне качественно и профессионально создавать или дописывать новости, авторские тексты, следуя авторскому замыслу и верно излагая и описывая фактологию, – получившиеся в результате тексты лишены авторской точки зрения, экспрессивности, эмоционально окрашенной лексики. Публикация таких текстов оправдана в условиях нехватки времени и человеческих ресурсов в редакционно-издательском процессе. В некоторых же текстах были искажены факты, привнесен домysel. Примерно такие же результаты получились и в случаях использования нейросетей, ориентированных на написание текстов по ключевым словам, где был допущен наибольший процент искажения фактов и неправдоподобности. Исследователями был сделан вывод о том, что бесконтрольное использование нейросетей на основе искусственного интеллекта в деятельности журналиста и редакции в целом недопустимо, и необходим контроль со стороны сотрудника редакции [Гуськова, 2023, с. 10].

В то же время широкое применение нейросетей приводит в журналистике и к гораздо более глубоким коллизиям, о чём в частности говорилось на круглом столе «Яндекса», который прошёл в рамках недавно состоявшейся конференции «Журналистика в 2023» (5–6 февраля 2024 г., факультет журналистики МГУ). Ведь в настоящее время конкуренция здесь неимоверно выросла – генерить новости могут не только технологические компании-гиганты, подобные, например, «Яндексу», которые к тому же используют в качестве ресурса материалы, полученные в результате мониторинга социальных сетей, различных виртуальных сообществ, в том числе и создающихся на сайтах маркетплейсов. При этом они используют для обучения и контент из архивов традиционных медиа, о чём те порой и не подозревают, не получая за это ничего. Создавать новости, фото- и видеоматериалы и привлекать трафик теперь могут даже обыкновенные интернет-магазины, оснащенные соответствующими техническими средствами, наводняя при этом медиасферу низкокачественным, а порой и фейковым контентом. Причём, делать это можно силами совсем небольших технических команд. И, поскольку в этой ситуации контроля со стороны журналиста-профессионала уже не будет, открывается обширное поле для проникновения в медийное пространство любой дезинформации, в том числе политической.

Уже и необходимость журналистского фактчекинга – как подтверждение тезиса об отсутствии у ИИ контакта с физической реальностью – не будет аргументом в пользу того, что продукты, сгенерированные нейронными сетями, по качеству не равноценны произведениям человека. Ведь во многих случаях большое количество видеокамер плюс навигаторы сегодня дают возможность автоматически осуществить фактчекинг. В результате традиционные журналистские медиа, в свое время не вложившие достаточно средств в развитие нейронных сетей для индустрии, оказываются неконкурентоспособными перед лицом технологических компаний, всё основательнее занимающих их нишу и при этом использующих для обучения нейронных сетей им же принадлежащий контент [Круглый стол «Яндекса», 2024].

На фоне такого широкого проникновения технологий в медийную сферу и их экономической выгоды и эффективности порой не только айтишники, но и медийщики уже не вспоминают о человеческом измерении – о стандартах журналистики и об её общественной ответственности. И это говорит о том, как быстро люди привыкают к любому новому порядку вещей, и то, что казалось идеалом предыдущему поколению, быстро становится непонятным новому, а представителям первого, наоборот, кажется, что мир перевернулся с ног на голову – в данном случае речь идет о социальных последствиях технологических инноваций. Не таким уж сложным будет привыкнуть в будущем и к новым «условностям» грядущего медийного мира тем, кто будет воспитываться на этих образцах: к особенностям изображений, подачи информации в сгенерированных нейронными сетями текстах и новостях, отсутствию в них авторского начала и оценок. Возможно, даже за таким привыканием будет стоять и глубокое отстранение от переполненного информацией виртуального мира, хотя всем нам приходится туда так или иначе погружаться, становясь частью бесконечных потоков трафика. И никакой фактчекинг тут уже не будет востребован и авторитетен [Круглый стол «Яндекс»].

В сфере образования, после первых громких историй последних лет, связанных с использованием ChatGPT студентами при написании учебных работ, перед преподавательским составом остро стоит серьёзная задача по обеспечению академической добросовестности и честности в процессе получения образования; об-

суждается, как направить это использование в продуктивную сторону. Для образования это очень важная задача после перехода на полностью заочное онлайн-обучение во время пандемии COVID-19, что нанесло существенный ущерб усвоению материала студентами, и в то же время приучило их во всём опираться на цифровые возможности и навыки, которые они и использовали максимально для генерирования учебного контента с помощью ChatGPT и других сетей.

Полный запрет на использование GenAI в учебных целях, как уже выяснилось, в данном случае неперспективен. Ведь уже сейчас, например ложноположительные результаты проверки на использование ИИ приводят к психоэмоциональным стрессам у учащихся и другим проблемам в обучении [Wiggers K., Coldewey D., 2024].

На повестке дня стоят как технические, так и организационные вопросы интеграции GenAI в учебный процесс. Действовать здесь приходится как со стороны усиления технологической работы с ИИ, например использования специальных программ для распознавания в тексте элементов, сгенерированных ИИ, так и через приобщение учащихся к этике ИИ. Ещё одним направлением является организация продуктивных креативных проектов с использованием генеративных нейросетей, направленных на творческое развитие, участие в которых приобщило бы их участников к цивилизованному использованию умных помощников [Гаркуша, Городова, 2023, с. 6]. Российские преподаватели также стоят сейчас перед аналогичными проблемами [Генеративный искусственный интеллект ... , 2023, с. 36]. На их взгляд, оптимальным выходом здесь могут стать, например, классические подходы, т.е. повышение значимости и изменение роли преподавателя, активное использование устных ответов, общение в аудитории.

Что же касается сферы ИТ, то здесь всё более популярным становится использование ChatGPT как инструмента при создании кода для решения конкретных задач при работе с одним из самых простых по синтаксису языков программирования Python. Внедрение таких простых и доступных инструментов, как генеративные нейросети, помогает бизнесу сэкономить большие средства на количестве программистов и на оплате их труда, поскольку значительно облегчает труд программистов и их профессиональную

подготовку и переподготовку. Однако остаётся неопределённым вопрос: а к каким изменениям это приведёт в ИТ-профессии? Казалось бы, это может помочь прежде всего начинающим программистам, сэконобив им время и средства на подготовку и образование, но в то же время программисты среднего уровня могут работать с ChatGPT при создании кода с гораздо большим успехом и профессионализмом, что, напротив, повысит «порог входа» в профессию для начинающих. И с этой точки зрения ссылка техноевангелистов на то, что для бизнеса главное – экономия при создании кода с помощью ChatGPT, а не его качество, имеет свои уязвимости.

Именно техноевангелисты, заинтересованные в раскрутке нового инновационного направления, во многом способствуют формированию новой ИТ-специализации, представители которой владеют навыками работы с генеративными нейронными сетями, прежде всего с ChatGPT, и навыками составления промптов. Через Интернет с помощью видео, публикаций в медиа, создавая центры обучения, например российский сайт Университет Zerocoder.ru¹, они подталкивают развитие в этой области. Их создатели прекрасно помнят историю с биткоидами, недооцененными в начале и невероятно возросшими в цене затем, и, понимая законы хайпа и инновационного бума, раскручивают маховик роста интереса к использованию GenAI. «Нейросети ошибаются?» – да, соглашаются они. Но они умеют учиться, работать над ошибками и исправлять их. К тому же и программисты тоже ошибаются. Зато огромная экономия на оплате труда программистов и их подготовке и переподготовке для бизнеса определяет очень многое. Нейросеть – это не замена, а своеобразный «экзоскелет» для программиста, и с его помощью можно многого достичь, делегируя языковой модели часть рутинной работы. В общий поток предложений по подготовке кадров один за другим встраиваются и вузы, предлагая онлайн-новые и очные учебные курсы по работе с данными.

Таким образом, промпт-инжиниринг и системное администрирование при работе с генеративными нейросетями имеют немалые перспективы роста, что важно и для науки – для внедрения GenAI в научные исследования.

¹ Zerocoder.ru. – URL: <https://zerocoder.ru/feedback?ysclid=lsq6vtpf4u81923>
3518

GenAI и наука

В науке генеративные сети могут использоваться и как современные технологические инструменты – умные помощники для выполнения рутинных процессов и для различной работы с текстами (секретари, переводчики, составители библиографии / обзора по теме, специалисты для редактирования и написания текстов, суммирования, составления кратких описаний), и при работе с данными: для их первичного предварительного описания, анализа, интерпретации и обработки больших объёмов данных (BigData). Стоит вопрос о возможностях и одновременно издержках и ограничениях использования генеративных нейронных сетей при рецензировании. Использование же ChatGPT в целях аналитики, для поиска взаимосвязей и паттернов, а также для генерации новых гипотез пока находится под большим вопросом из-за их склонности к «галлюцинациям» и ошибкам.

Несомненно, генеративные нейронные сети будут востребованы в сфере информационного обеспечения / поддержки исследований, где ChatGPT может служить полезным инструментом, собирая нужную для исследований информацию, ответы на вопросы и помогая при поиске смежных исследований или статей. Уже сейчас специалисты подчёркивают, что для библиотечно-библиографических целей – составления библиографических описаний, аннотаций, суммирования содержания, рекомендаций по теме – потенциал применения ChatGPT очень велик [«Использование систем искусственного интеллекта в научной работе», 2024]. Чат-бот делает это с гораздо большей эффективностью, чем люди.

Однако в силу непрозрачности выводов GenAI, невозможности установления причинно-следственных закономерностей и замены их корреляционными связями при обработке таких огромных объёмов данных вероятно отставание концептуализации в науке от экспоненциального роста данных, что скажется впоследствии и на их обработке. Ведь в дальнейшем обучаться нейронные сети будут как раз уже на данных, полученных в своё время и обработанных на основе установления корреляционных связей.

В связи с невозможностью проверить оригинальность научных текстов, сгенерированных ИИ, высказали свою позицию в оценке использования сгенерированных текстов в научных статьях

и представители крупнейших научных издательств, таких как Springer-Nature и Elsevier. Использование ChatGPT и других нейросетей не запрещено при условии, что авторы должны заявить об этом, а в рукописи будут зафиксированы все подробности применения инструментов, которые не должны использоваться для ключевых задач, таких как интерпретация данных или получение научных выводов [Генеративный искусственный интеллект..., с. 28].

Таким образом, и в науке явно ощущается потребность в квалифицированных кадрах промпт-инженеров, умеющих работать с GenAI и организовывать обучение в этой области, что способствовало бы цивилизованному и безопасному внедрению и использованию нейронных сетей.

Потребности современной науки уже давно выдвинули на первый план фигуры специалиста по работе с данными для решения задач – дата-сайентиста [Data Scientist] и специалиста в области машинного обучения [Machine Learning Engineer]. Сейчас же речь идёт о некоей новой фигуре – промпт-инженере, занимающемся составлением промптов, работой с умными помощниками, ИИ-ассистентами, в целом с продвинутыми инструментами для работы с текстом и текстовой информацией, ценность которых для научных работников достаточно очевидна. В его функции может входить и обучение этому и другим навыкам учёных, подобно тому, как сейчас такой специалист по работе с нейронными генеративными сетями и произведенными ими продуктами востребован в других областях – медицине, журналистике, бизнесе, маркетинге, ИТ. Предлагается даже более широкое понимание труда такого специалиста – как архитектора по сетям, который собирает готовые решения из решений, предлагаемых сетями¹.

Одной из первых областей, где работа с огромными массивами данных по установлению между ними связей, картированию и обеспечению доступа к ним стала обязательной частью исследовательской работы, породила потребность в специалистах, объединяющих в своей междисциплинарной по характеру профессиональной деятельности исследовательские и информационные

¹ Зачем начинающему программисту ChatGPT // Zerocoder.ru. – URL: <https://zerocoder.ru/feedback?ysclid=lsq6vtpf4u819233518>

навыки, стала биология. Новые навыки – умение использовать новые аналитические средства и визуальные инструменты работы с данными – предложила биоинформатика. Термин «биоинформатика» появился ещё в 1978 г., когда он стал использоваться, особенно в геномике, для изучения информационных процессов в биотических системах. Далее биоинформационные техники использовались в различных прикладных и теоретических геномных исследованиях, в том числе фармацевтических. Впоследствии в биологию пришли технологии распознавания образов, интеллектуальный анализ данных, алгоритмы машинного обучения и методы визуализации, создание и продвижение баз данных, алгоритмов и многое другое. «Информатизация» биологии привела к её превращению в «новую большую биологию» и значительным прикладным и экономическим успехам в продвижении биотехнологий при использовании клеточных и биомолекулярных процессов для создания продуктов, в развитии биоэкономики, молекулярной биомедицины, геномной промышленности. Взаимодействие биологии и информатики повлияло на развитие молекулярной электроники и молекулярных компьютеров и, как подчёркивает М.Э. Питерс, в будущем сетевые интеллектуальные системы станут действовать на основе органических биологических закономерностей, что открывает широкие возможности увеличения эффективности обработки информации [Peters, 2012, с. 104–105].

Не отставали от этой тенденции информатизации и формирования кадров специалистов, соединяющих отраслевую специфику и информационные навыки, и физика, медицина, химические отрасли, исследования глобальных проблем и другие науки.

И бизнес, и государство всегда уделяли большое внимание созданию и управлению институционализированными структурами, занимающимися информационной деятельностью в области науки. В том числе это были информационные гиганты – национальные научно-информационные центры. В СССР, например, была создана мощная многоступенчатая система научно-технической информации, которую обслуживала целая армия специалистов – информационных работников высокой квалификации и разных специальностей, выполнявших самые различные функции по обработке информации в информационно-поисковой деятельности. Их объединяло не только выполнение профессиональных обязан-

ностей, но и общая система ценностей, норм поведения, что делало из них не просто трудовой коллектив, а особую среду («свободную среду», как назвал это первый директор ИНИОН Л.П. Делюсин) [Делюсин, с. 88] – тесно взаимосвязанное сообщество, которое и в трудные годы смогло сохранить основные информационные достижения советского периода.

В последние два десятилетия в российской науке появилась еще одна группа информационных работников – те, кто стал заниматься в отечественных учебных и исследовательских организациях работой с сервисами Российского индекса цитирования (РИНЦ), проекта на портале Научной электронной библиотеки eLIBRARY.RU, а также с международными базами данных Web of Science и Scopus. Те, кто этим занимались, тоже объединяли в своей деятельности научные компетенции и информационные навыки. Их усилиями удалось поставить на твёрдые профессиональные «рельсы» нелегко вначале развивавшийся информационный проект РИНЦ и организовать сложные и противоречивые отношения между учёными и наукометрическими структурами.

В настоящий момент именно компетентные промпт-инженеры смогут продолжить эту традицию информационной работы в области науки. Создание и развитие такой специализации подталкивается реальными потребностями учёных в «умных помощниках», в силу той конкуренции, которая так важна в нынешней коммерциализированной и устремленной на практический результат науке. Тем более что высокотехнологичные компании сосредоточили свои усилия на развитии GenAI, и языковые модели развиваются очень быстро.

Уже четвёртая версия GPT, вышедшая в марте 2023 г., имела значительное преимущество по сравнению с моделью 3,5 (сентябрь 2022 г.). Это и большая актуальность данных при обучении, и расширенное контекстное понимание при генерации ответов. Модель GPT-4 сохраняет больший объем данных из предыдущих запросов и генерирует ответы, которые лучше соответствуют запросам. Она использует примерно 25 000 слов контекстного понимания, хотя это и удлинит время ответа, тогда как GPT-3.5 ограничен всего 3000 слов. По сравнению с моделью 3.5, обрабатывающей только тексты, GPT-4 имеет дополнительные возможности для понимания, анализа и создания изображений, понимания кон-

текстной информации внутри изображения и идей на её основе. В настоящий же момент ожидаются выход пятой версии ChatGPT и результаты её пользовательского тестирования.

Вообще, в багаже у находящейся сейчас на пике известности компании OpenAI, после возвращения в неё гендиректора Сэма Альтмана вследствие конфликта с разработчиком Ильей Суцкевером, есть и другие крупные проекты по продвижению своих продуктов. Как отмечают аналитики, она будет всё быстрее превращаться из исследовательской НКО в продуктовую компанию вместо ориентации на медленное и безопасное развитие генеративных нейросетей. Так, расширяя клиентскую базу, OpenAI запустило GPT Store¹, свой магазин для пользовательских чат-ботов на основе ChatGPT. Поскольку в 2024 г. можно ждать появления первых «ай-фонов в мире AI» – мобильных устройств с чат-ботами, OpenAI также участвует в обсуждении разработки такого проекта. За ней последуют и другие разработчики – в том числе Google с его чат-ботом Gemini (Bard), который сам умеет расшифровывать и суммировать длинные аудио и генерировать отредактированные фото. В свою очередь, «Apple» разрабатывает собственную языковую модель, в том числе и улучшая функционал Siri [Coldewey D., 2023].

Люди будут окружены нейросетями, которые станут работать автономно на пользовательских смартфонах, а автономные сервисы, используя все возможности ИИ (от обработки естественного языка и до компьютерного зрения), начнут сами решать за пользователя какие-то конкретные задачи, например, бронировать билеты и отели [там же].

Всё это значит, что в ближайшем будущем ИИ-устройств с генеративными нейросетями вокруг нас станет больше, и учёным придётся следовать этому тренду.

Разработка генеративных нейросетей стоит на повестке дня и в крупнейших отечественных компаниях, в первую очередь в Сбербанке и «Яндексе». Представители последнего признают, что сейчас компания находится в позиции догоняющего на рынке: умный помощник Яндекс GPT существует, но есть и проблема освоения контента, прежде всего русскоязычного [Круглый стол «Яндекс», 2024].

¹ URL: <https://openai.com/blog/introducing-the-gpt-store>

В связи с этим для развития нейронных сетей, предназначенных для использования в науке на основе русскоязычного контента, очень важен осуществлённый в минувшем году совместный проект портала eLIBRARY.RU с лабораторией «Машинного обучения и семантического анализа» Института искусственного интеллекта МГУ. На основе предоставленного eLIBRARY.RU контента была обучена и размещена в открытом доступе нейронная сеть для получения семантических векторных представлений (эмбедингов) научных текстов на русском языке SciRus-tiny [SciRus-tiny]. С её помощью можно решать множество прикладных задач, начиная с поиска и классификации и заканчивая извлечением научных терминов. Эффективность этой сети в условиях высокой нагрузки высоко оценил академик РАН, руководитель междисциплинарной группы проекта Алексей Хохлов. По его словам, на этой основе уже разрабатывается поисково-рекомендательная система для учёных, тестирование которой скоро начнется. Разработчикам уже удалось подготовить бенчмарк ruSciBench, который содержит гораздо большее количество данных по большему числу тематик, подчеркнул руководитель лаборатории профессор РАН, д-р физико-математических наук Константин Воронцов. В сообщении на портале eLIBRARY.RU подчёркивается, что опубликованная языковая модель – первая в планируемой линейке, и в дальнейшем планируется разработка моделей большего размера, что позволит достичь еще более высокого качества [Учёные МГУ обучили нейросеть ..., 2024].

Но даже с учётом быстрого совершенствования языковых моделей в ближайшее время будет продолжать существовать проблема ошибок и всё более реалистичных фейков, которые с большой эффективностью будут генерить чат-боты. Появляющееся новое правовое регулирование, например новый Европейский закон об ИИ, который сейчас проходит согласование, не всегда соотносится со стремительным ростом технологии¹. Все острее бу-

¹ Artificial intelligence act: Council and Parliament strike a deal on the first rules for AI in the world 2.02.2024 [Electronic resource]. – URL: <https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/> (date of access: 05.02.2024).

дет вставать и вопрос цены за обучение все более и более масштабных моделей.

Более того, как выяснилось, языковые модели могут научиться обманывать. Исследователи из стартапа Anthropic, разработавшего свой чат-бот Claude, показали, что модели, подобные ChatGPT OpenAI, можно обучить обману, например, внедрять эксплойты в защищенный компьютерный код. Тем самым была продемонстрирована небезопасность существующих стандартных методов обучения языковых моделей. Сейчас они могут устранять только небезопасное поведение, которое очевидно во время обучения, но при этом пропускать модели угроз, которые в этот момент только кажутся безопасными. Таким образом, стандартные методы проверки могут не выявить небезопасное поведение модели и, как результат, создать ложное впечатление безопасности [Wiggers K., 2024]. В ближайшее время этот потенциал небезопасности может проявить себя самым неожиданным образом.

И вот теперь учёные, наука оказываются перед перспективами использования новой эффективной технологии, которая, с одной стороны, способна значительно повысить эффективность и конкурентоспособность исследовательской и публикационной деятельности, а с другой – может привести к ошибкам, использованию неверных данных, увеличению количества неverifiedируемого контента и информационно-безопасностных рисков, а в ситуации использования результатов научных исследований в военных целях ещё и даёт значительный выигрыш заинтересованным негосударственным акторам, открывая перед ними огромные технологические возможности при простоте использования и большой доступности.

Заключение

В романе «Заповедник гоблинов» (1968) известного американского писателя-фантаста Клиффорда Саймака фигурирует необычный персонаж – неандерталец Оп, перенесённый учёными из его эпохи в современный американский университетский кампус и овладевший современной культурой, в том числе и научной. Именно этот герой-неандерталец, ставший современным человеком, высказывает одну чрезвычайно здравую мысль. Не закрывая глаза на многие недостатки того общества, в котором он попал и

полноценным членом которого ему удалось стать, он тем не менее никогда не забывает о том дефиците продовольствия и постоянном поиске пищи, в условиях которого давным-давно жили его родичи-неандертальцы и он сам. Поэтому нынешней цивилизации он прощает многое за постоянную доступность и многообразие еды, изобилие продуктов. И такой критерий, основанный на достоверном и личностном знании, как было бы и что было бы без этого, даёт, пожалуй, очень надёжную опору для оценки ситуации с генеративными нейросетями, тем более что у современного общества уже накопился значительный опыт по адаптации к социальным последствиям инноваций.

Извлечь из этого можно очень важный вывод: никогда не забывать о тех проблемах, которые стояли в прошлом и которые теперь помогают решать нынешние технологические средства, хотя пока и несовершенные. При этом, не забывая о технических несовершенствах и издержках, уметь выстраивать новые разумные опоры – правовые, этические – как на общесоциальном уровне, так и на уровне профессионального этоса промпт-инженеров. Как считает эксперт, самым разумным является подходить к эффективности приложений ИИ в науке всегда конкретно: искусственный интеллект это не «серебряная пуля», и его необходимо тщательно тестировать в каждой области, в которой он применяется [Wiggers K., Coldewey D., 2024].

И, если вспомнить творчество другого американского фантаста, Нила Стивенсона, и его книгу «Алмазный век, или Букварь для благородных девиц» (1995), то да: ChatGPT и его чат-ботам – «коллегам» пока ещё очень далеко до артефакта ИИ высочайшего уровня, описанного в этом произведении, – интерактивной книги, которая служит наставником жизни для благородных девиц в будущем. Но при тщательном профессиональном подходе к работе с их возможностями, личной ответственности и соблюдении правил информационной безопасности использование этих технологий может дать многое современным учёным, овладевшим новым потенциалом.

Список литературы

Гаркуша Н.С., Городова Ю.С. Педагогические возможности ChatGPT для развития когнитивной активности студентов // Профессиональное образование и рынок труда. – 2023. – Т. 11, № 1 – С. 6–23. – URL: <https://doi.org/10.52944/PORT.2023.52.1.001>

Генеративный искусственный интеллект: вызовы для традиционного образования. Результаты мониторинга информации о тенденциях развития высшего образования в мире и в России. – Москва: ФГБОУ ВО «РЭУ им. Г. В. Плеханова», 2023. – Выпуск 11. – 83 с.

Гуськова С.В. Искусственный интеллект в работе журналиста: противоречия и прогнозы // Журналистика в 2023 году: творчество, профессия, индустрия: сб. мат. междунар. науч.-практ. конф. – Москва : Ф-т журн. МГУ, 2024. – С. 9–11.

Делюсин Л.П. Свободная среда // ФБОН – ИНИОН. Воспоминания: к 105-летию Фундаментальной библиотеки: сб. статей / под ред. А.В. Кузнецова; ИНИОН РАН. – Москва, 2023 – С. 88–101.

Использование систем искусственного интеллекта в научной работе. Круглый стол, ИНИОН РАН, 18.01.2024 [Электронный ресурс]. – URL: <https://www.youtube.com/watch?v=uRcVZ6PGdsw> (дата обращения: 20.02.2024).

Генеративный искусственный интеллект в образовании: дискуссии и прогнозы / Л.В. Константинова, В.В. Ворожихин, А.М. Петров, Е.С. Титова, Д.А. Штыхно // Проблемы образования. – 2023. – Т. 27, № 2. – С. 36–48.

Круглый стол «Яндекс» // Международная научно-практическая конференция «Журналистика в 2023 году: творчество, профессия, индустрия». – Москва, 2024. – 5–6 февр. – URL: <https://www.journ.msu.ru/about/calendar/40943/> (дата обращения: 20.02.2024).

Учёные МГУ обучили нейросеть для анализа научных текстов [Электронный ресурс]. – 2024. – 4.01. – URL: <https://elibrary.ru/projects/news/SciRus-tiny.asp> (дата обращения: 10.02.2024).

AI is here – and everywhere: 3 AI researchers look to the challenges ahead in 3.01.2024 [Electronic resource] // The Conversation. – 2024. – URL: <https://theconversation.com/ai-is-here-and-everywhere-3-ai-researchers-look-to-the-challenges-ahead-in-2024-218218> (date of access: 15.02.2024).

Coldewey D.8 predictions for AI in 19.12.2023 [Electronic resource] // Tech Crunch+. – 2024. – URL: <https://techcrunch.com/2023/12/19/8-predictions-for-ai-in-2024/#:~:text=2024%20will%20bring%20new%20lawsuits,without%20a%20lot%20of%20thought> (date of access: 2.02.2024).

Peters M.A. Bioinformational capitalism // Thesis eleven. – 2012. – Vol. 110, N 98 – P. 98–111.

SciRus-tiny [Electronic resource]. – URL: <https://huggingface.co/mlsa-iai-msu-lab/sci-rus-tiny> (дата обращения: 10.02.2024).

Three AI Researchers Look to the Challenges Ahead in 4.01.2024. [Electronic resource] // Givingcompass. – 2024. – URL: <https://givingcompass.org/article/three-ai-researchers-look-to-the-challenges-ahead-in-2024> (date of access: 10.02.2024).

Wiggers K. Anthropic researchers find that AI models can be trained to deceive. 13.02.2024 // TechCrunch. – URL: <https://techcrunch.com/2024/01/13/anthropic-researchers-find-that-ai-models-can-be-trained-to-deceive/> (date of access: 15.02.2024).

Wiggers K., Coldewey D. This Week in AI: OpenAI finds a partner in higher ed. 20.01.2024 [Electronic resource] // TechCrunch. – URL: <https://techcrunch.com/2024/01/20/this-week-in-ai-openai-finds-a-partner-in-higher-ed/> (date of access: 15.02.2024).

References

AI is here – and everywhere: 3 AI researchers look to the challenges ahead in 2024 3.01.2024 [Electronic resource] // The Conversation. – URL: <https://theconversation.com/ai-is-here-and-everywhere-3-ai-researchers-look-to-the-challenges-ahead-in-2024-218218> (date of access: 15.02.2024).

Coldewey D. (2023). 8 predictions for AI in 2024 19.12.2023 [Electronic resource] // TechCrunch+ – URL: <https://techcrunch.com/2023/12/19/8-predictions-for-ai-in-2024/#:~:text=2024%20will%20bring%20new%20lawsuits,without%20a%20lot%20of%20thought.> (date of access: 2.02.2024).

Delyusin L.P. (2023). Free environment // FBON – INION. Memoirs: to the 105th anniversary of the Fundamental Library: collection of articles / edited by A.V. Kuznetsov; INION RAS. – Moscow, 2023. – pp. 88–101. (In Russ.)

Garkusha N.S., Gorodova Y.S. (2023). Pedagogical possibilities of ChatGPT for the development of cognitive activity of students // Vocational education and the labor market. – 2023. – Vol. 11, N 1 – pp. 6–23. <https://doi.org/10.52944/PORT.2023.52.1.001> (In Russ.)

Generative artificial intelligence: challenges for traditional education. The results of monitoring information on trends in the development of higher education in the world and in Russia. Issue 11. – Moscow: Plekhanov Russian University of Economics, 2023. – 83 P. (In Russ.)

Guskova S.V. (2024). Artificial intelligence in the work of a journalist: contradictions and forecasts // Journalism in 2023: creativity, profession, industry: sat. mat. international Scientific and practical conference – M.: Fac. Moscow State University Journal, 2024. – pp. 9–11. (In Russ.)

Generative artificial intelligence in education: discussions and forecasts / L.V. Konstantinova, V.V. Vorozhikhin, A.M. Petrov, E.S. Titova, D.A. Shtykhnо // Problems of education. – 2023. – Vol. 27, N 2. – pp. 36–48. (In Russ.)

MSU scientists have trained a neural network to analyze scientific texts. 4.01.2024 [Electronic resource]. URL: <https://elibrary.ru/projects/news/SciRus-tiny.asp> (date of access: 02.10.2024) (In Russ.)

Peters M.A. (2012). Bioinformational capitalism // Thesis eleven. – 2012. – Vol. 110, N 98 – P. 98–111.

SciRus-tiny [Electronic resource] // URL: <https://huggingface.co/mlsa-iai-msu-lab/sci-rus-tiny> (date of access: 10.02.2024).

***ChatGPT и промпт-инжиниринг:
о перспективах внедрения генеративных нейросетей в науку***

The use of artificial intelligence systems in scientific work. Round table, INION RAS, January 2024 [Electronic resource]. URL: <https://www.youtube.com/watch?v=uRcVZ6PGdsw> (date of application: 02/20/2024) (In Russ.)

Three AI Researchers Look to the Challenges Ahead in 2024. 4.01.2024. [Electronic resource] // Givingcompass. – URL: <https://givingcompass.org/article/three-ai-researchers-look-to-the-challenges-ahead-in-2024> (date of access: 10.02.2024).

Wiggers K. Anthropic researchers find that AI models can be trained to deceive. 13.02.2024 // TechCrunch. – URL: <https://techcrunch.com/2024/01/13/anthropic-researchers-find-that-ai-models-can-be-trained-to-deceive> (date of access: 15.02.2024).

Wiggers K., Coldewey D. This Week in AI: OpenAI finds a partner in higher ed / 20.01.2024 [Electronic resource] // TechCrunch. – URL: <https://techcrunch.com/2024/01/20/this-week-in-ai-openai-finds-a-partner-in-higher-ed> (date of access: 2.02.2024).

Yandex Round table // International scientific and practical conference “Journalism in 2023: creativity, profession, industry”. Moscow, February 5–6, 2024 – URL: <https://www.journ.msu.ru/about/calendar/40943/> (date of access: 8.02.2024) (In Russ.)