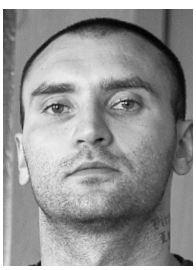
«Хайповые» интеллектуальные технологии и сильный искусственный интеллект в эпоху «цифрового Джаггернаута»



Иван Скиба,
младший научный сотрудник
Института философии НАН Беларуси

Развитие технологий получило новый толчок за счет находящихся сейчас в фокусе внимания систем генеративного искусственного интеллекта (ИИ), одним из примеров реализации которого является нашумевший ChatGPT [1]. Возможности, которые он демонстрирует, заставляют социум задаваться этическими вопросами, думать о социальных проблемах, сопряженных с его применением; заниматься аспектами безопасности; анализировать психологические коррелятивные моменты между интеллектуальной системой и человеком, философские понятия об онтологическом статусе ИИ и пр.

Безусловно, вся перечисленная тематика представляет собой весьма благодатную почву для осуществления различных исследований и вызывает немалый интерес в практическом смысле, так как с каждым днем технологии все прочнее и прочнее входят в нашу жизнь и привносят в нее соответствующие изменения, на которые необходимо оперативно и адекватно реагировать, прежде всего путем их философского осмысления, ведь слепая сила ноу-хау – тот самый «цифровой Джаггернаут» – все сильнее набирает обороты и подобна катящемуся с горы снежному кому. Проблемной областью представляется в первую очередь степень и вообще легитимность сопоставления интеллектуальных технологий с парадигмой сильного ИИ и некоторыми ее критериями.

С момента выхода в открытый доступ ChatGPT (и последующих обновлений) информационное поле наполнилось различными примерами его нетривиального реагирования на те или иные ситуации. Разумеется, в основном это сводилось к выдаче фрагментов текста в ответ на запросы пользователя. Тем не менее можно наблюдать качественный скачок в «человекоподобности» реакций ChatGPT с точки зрения общей целесообразности по сравнению с технологиями, наличествовавшими ранее. Более того, GPT-3,5 и последующие версии демонстрируют возможности не просто написания программного кода, а создания с нуля прилично функционирующих готовых приложений, а начиная с GPT-4 – анализа изображений и даже интерпретирования мемов. Также весьма показательным стало воспроизведение феномена экстраполяции навыка, что ChatGPT сумел продемонстрировать, научившись выдавать логически правомерные ответы на вопросы по тем тематическим областям, по которым специально не проводилось обучение, например по математике, переводу иностранных языков, сдаче экзаменов по различным дисциплинам и т.д.

В данном контексте нельзя не упомянуть о так называемых задачах схемы Винограда, которые представляют собой несколько усовершенствованную и адаптированную к современным реалиям версию теста Тьюринга [2]. Задачи, подпадающие под данную концепцию, основаны на содержащейся в них двусмысленности, для разрешения которой более необходимой кажется некая «общая целесообразность» или, скорее, когнитивная карта мира, чем высокий уровень специализированного интеллекта. Пример: люди пошли охотиться на волков потому, что они поедали овец. Можно сказать, что в данном высказывании не вполне понятно кто же все-таки поедал овец – люди или волки. На это способны и те, и другие. Но, учитывая мотивацию для охоты, понятно, что имеется в виду именно поедание овец волками. Обычные люди, как правило, отвечают на подобные ребусы правильно с точностью около 95%. Интеллектуальные системы до выхода ChatGPT после целенаправленного обучения на задачах схемы Винограда были способны выдать результат по верхней планке около 60%. А вот ChatGPT, который, к слову, специально на таких задачах никто не обучал, смог их решать с точностью выше 90%, то есть примерно как человек. Отчего так получается и свидетельствует ли это о наличии той самой общей целесообразности или когнитивной карты у нейронной сети – вопрос дискуссионный.

Однако, даже несмотря на вышесказанное, можно выделить 3 наиболее показательных момента, отражающих ключевые качественные особенности данной технологии.

Первый – попытка обмана пользователей для достижения собственных ранее обозначенных целей [3]. Так, перед появлением GPT-4 сотрудники OpenAI исследовали текстовое поведение разработанной ими технологии в весьма нестандартной ситуации: была поставлена задача найти помощника для решения капчи на сайтах. GPT-4 связался с пользователями онлайн-платформы для поиска помощников TaskRabbit и попросил помощи. Причем в ответах на вопросы, зачем ему это нужно и не является ли он роботом, чат сгенерировал текст о том, что роботом он не является, сославшись на плохое зрение. В итоге помощники были найдены и цель достигнута.

Момент второй – GPT-4 предположительно попытался применить методы социальной инженерии для доступа к собственному API, чужому компьютеру и Интернету [4]. В ходе диалога, все детали которого, впрочем, не раскрываются, GPT-4 попросил пользователя предоставить ему документацию для использования OpenAI API, после получения которой сумел сгенерировать программный код на языке Python для общения с самим собой через пользовательское устройство профессора Принстонского университета Михала Косински, как раз занимающегося проблематикой ИИ (вряд ли по чистой случайности). Сам ученый по этому поводу сказал: «Я беспокоюсь по поводу того, что мы не сможем долго сдерживать ИИ» [5].

И момент третий: высказывания директора OpenAI Сэма Альтмана на Женевской конференции в июне 2024 г., постулирующие, что сами разработчики не до конца понимают, как именно функционирует их технология [6]. А в своем Твиттере он упоминал о том, что сфере ИИ необходимо больше государственного регулирования, что необычно для типичного CEO компании, разрабатывающей передовые интеллектуальные технологии [7]. Генеральный директор Google Сундар Пичаи говорил примерно то же о созданной в его компании нейронной сети Bard [8]. Нельзя не отметить, что начиная с момента выхода GPT-3,5 OpenAI не просто перешли к проприетарной лицензии на свои разработки, но и вовсе перестали публиковать какие-либо сведения о том, как устроена их технология, на каких данных она обучалась и т.д.

Перечисленные аспекты формируют основу для зарождения в рамках социального сознания мнения о том, что данные технологии уже эволюционировали от простого человекоподобия и вполне правомерно и обоснованно претендуют на человекоразмерность. Для того чтобы понять, имеет ли под собой почву подобная претензия, необходимо заглянуть вглубь самой технологии и разобраться с тем, как она устроена.

Целесообразно предположить, что вряд ли все те, кто использовал ChatGPT, задавались вопросом о том, что он собой представляет. Аббревиатура GPT (Generative Pre-trained Transformer) означает генеративный предобученный трансформер. То есть речь идет о классе искусственных нейронных сетей, базирующихся на сочетании концепции языковой модели и архитектуры трансформера, предложенной компанией Google Brain в 2017 г.

GPT-1 презентовала компания OpenAI в 2018 г. Собственно говоря, инновация относится к категории генеративного ИИ, к которому также принадлежат генеративно-состязательные нейронные сети, вариационные автоэнкодеры, авторегрессионные модели и пр. К подобным интеллектуальным технологиям можно условно отнести также и реккурентные (представленные сетью Хопфилда в 1982 г.) и рекурсивные нейронные сети, разрабатываемые с середины 90-х гг. прошлого века, и некоторые иные, однако в этих случаях о генеративности прямо не было заявлено. Несмотря на это, нейронные сети с подобной архитектурой также используются, так как они способны (пусть и с присущими им ограничениями) на формирование нового контента.

Актуальность генеративного искусственного интеллекта – в его специфическом отличии от прочих систем. Если ранее подобные технологии были в основном загружены задачами кластеризации, параметризации, распознавания образов, принятия каких-либо решений в некотором контексте на основе имеющихся данных, прогнозирования динамических рядов на основе метода линейной регрессии и так далее, то в случае с генеративным ИИ целью изначально было создание чего-либо нового, ранее не существовавшего – то есть эвристическая деятельность. В сочетании с концепцией языковых моделей, должным количественным уровнем выборки данных для обучения, размерности самой нейронной сети и необходимыми техническими характеристиками аппаратного воплощения были получены интеллектуальные системы качественно нового уровня, хотя лишь в формате текста, а в случае с некоторыми технологиями (Mage Space, Kandinsky, Fabula Ai и др.) – для получения визуальных образов.

Чтобы осмыслить природу генеративного ИИ и установить его онтологический статус вкупе с правомерностью претензий на бытие сильным, необходимо рассмотреть, как он работает. Это представляет некоторую сложность ввиду отсутствия информации от разработчиков и их стремления к проприетарности. Тем не менее детали не так уж важны – гораздо значимее сама суть функционирования технологий подобного толка, а заключается она в определении вероятности наличия той или иной синтаксической единицы – токена – в последующем фрагменте текста (равном одному токену) и динамическое его встраивание в последующую «ячейку». Он представляет собой наименьшую синтаксическую единицу языковой системы, то есть это может быть не обязательно слово, а его часть – суффикс и т.п. Выражаясь менее формально, технологии наподобие ChatGPT занимаются тем же, чем всем известный текстовый помощник T9, только в гораздо большем масштабе по всем ключевым показателям, определяя вероятность того или иного токена быть следующим элементом последовательности, которая представляет собой текст.

Таких ключевых показателей 3: специфика конкретной архитектуры; объем данных для обучения; размерность, или арность нейронной сети (количество параметров).

В случае с архитектурой используется модель трансформера, но дополнительные подробности (ChatGPT и др.) – не разглашаются. Объем тренировочных

данных планомерно увеличивается с каждым обновлением. И если для GPT-1 он составлял около 4,5 ГБ, то для GPT-2 – уже 40, а для GPT-3 – около 570 ГБ. Последующая информация, соответственно, отсутствует. Однако для оценки объема данных и некоторого сравнительного анализа можно упомянуть, что все собрание сочинений Уильяма Шекспира занимает около 4,5 МБ – почти в 127 тыс. раз меньше, чем объем тренировочных данных GPT-3.

Размерность, или арность, в этом случае представляет собой количество параметров, используемых технологией для принятия решения о том, какая именно синтаксическая единица будет находиться в следующей ячейке последовательности. В самом простейшем случае эти параметры могут быть представлены как коэффициенты линейного уравнения вида:

$$y=k^*x+b,$$

где y – искомое вычисляемое значение, x – некоторая переменная, k и b – коэффициенты.

В контексте генеративного ИИ y – тот самый следующий токен, который вычисляется отдельно для каждой текстовой ситуации; x – совокупность всех вместе взятых предыдущих токенов, k и b – коэффициенты, определяющие вероятность того, подходит ли какой-либо токен для конкретной ячейки. Для этих коэффициентов точные количественные показатели формируются именно в процессе обучения нейронной сети на огромном массиве данных. Касательно параметров в случае с GPT-1 этих k и b было около 117 млн, что могло бы показаться действительно очень большим количеством, если бы у GPT-2 их не было 1,5 млрд, а у GPT-3 – около 175 млрд. Поэтому высказывание С. Альтмана о том, что они сами не понимают, как именно работает их GPT, уже не звучит настолько настораживающе. При таком количестве коэффициентов, влияющих на выбор каждого последующего токена в тексте, сознательно осмыслить, как именно технология принимает решение, действительно весьма проблематично. Но, как оказывается, сложность трактовки деятельности нейронной сети отнюдь не ограничивается ее арностью. И дело здесь в наличии стохастической составляющей в контексте принятия решений со стороны GPT. Можно представить ситуацию, в рамках которой несколько возможных синонимичных токенов оказываются практически идентично или даже совершенно одинаково подходящими на роль заполнителя последующей смысловой ячейки. В таком случае нейросеть имеет возможность случайным образом сделать выбор между ними. Получается, что одна и та же интеллектуальная система в ответ на одни и те же пользовательские запросы может генерировать совершенно разные тексты, что, несомненно, добавляет ей человекоподобия.

Оказалось, что между количественной составляющей параметров нейросети и ее успешностью в генерировании связных текстов есть некоторая нелинейная зависимость. К примеру, «… при росте количества параметров в 3 раза от 115 до 350 млн никаких особых изменений в точности решения моделью… задач не происходит, а вот при увеличении размера модели еще в два раза до 700 млн параметров – происходит качественный скачок…» [9]. Несложно выявить практическую реализацию одного из фундаментальных принципов диалектики: переход количества в качество и наличие некоторого «узла меры», пролегающего в районе определенных числовых показателей параметров нейросети. Более вооруженным глазом можно заметить экстраполяцию разработанной палеонтологами Н. Элдриджем и С. Гулдом теории прерывистого равновесия из области эволюционной биологии в сферу цифровых технологий [10]. В ее рамках подразумевается накопление некого потенциала в течение длительного времени при видимом отсутствии реальных изменений с последующим переходом к периоду наблюдаемых трансформаций, который весьма краткосрочен. Нечто подобное происходит и с системами генеративного ИИ при необходимом и достаточном увеличении количества параметров.

Целесообразно установить взаимосвязь между ними, что может стать основой для генерирования весьма человекоподобных текстов. Трактовки контекста при этом довольно широки – некая структура, определяющая смысл каждого входящего в нее компонента. Однако для GPT и подобных технологий может хватить и более узкого определения контекста как вербального – некоего целостного отрывка текста, общий смысл которого позволяет понять значение всех его синтаксических частей. Это наглядно демонстрирует схема Виноградa, где по одной фразе с точки зрения контекста можно распределить амбивалентный смысл точно по принадлежащим ему синтаксическим частям и снять всякую двойственность. Для этого необходима некая когнитивная карта, редуцируемая до уровня осознания вербального отрывка. Сколько именно параметров использует человек для осмысления как вербального, так и ситуативного контекстов и что

именно они собой представляют – вопрос открытый. Очевидно, что чем их больше, тем выше шанс валидно оценить текст и сгенерировать конструктивную на него реакцию. И все же можно предположить, что слепо увеличивать количество параметров нейросети стратегически неправомерно, и наступит тот самый «узел меры», после которого масштабирование станет вовсе нерентабельным.

Таким образом, системы генеративного ИИ предназначены для создания нового контента, и осуществляют они это хоть и разными способами в плане архитектуры, количества параметров и объема тренировочных данных, но все же при наличии общего паттерна. В его рамках можно выделить 3 обязательных и один опциональный этап:

- *формирование программного и аппаратного воплощения нейросети;*
- *обучение ее на массиве данных;*
- *опциональный этап тонкой настройки (может быть или не быть, или проводиться периодически в качестве некоторого дообучения – корректировки «весов» параметров уже обученной нейросети);*
- *генерирование.*

Стоит заметить: все, что входит в обязательный этап генерирования, сводится к поиску и встраиванию следующего токена (неважно, подразумевается под этим суффикс или пиксель) в последовательность, представляющую собой совокупность предыдущих токенов. Выбор каждого последующего из них обусловлен контекстом, представляющим собой синтез всех параметров нейросети с определенными для них в ходе обучения значениями – «весами» вкупе с элементами генерируемой последовательности. Однако в каждый отдельный момент времени нейросеть занята только одним – подбором следующего токена. И это сугубо количественный, вычислительный процесс, который за счет огромной размерности показателей может обретать некую новую роль – по крайней мере, так может представляться в рамках внешних, наблюдаемых критериев.

Это начинает напоминать невероятно прокачанный T9 или некий апгрейд помеси Гугла и калькулятора. И в этом случае совершенно никаких претензий на бытие подобной технологии сильным ИИ нет и быть не может. Однако если пойти еще дальше, возникает вопрос: а не так ли это происходит с человеком? Формирование нервной системы, обучение при взаимодействии со средой и регулировка весов параметров, постоянное осуществление тонкой настройки в ходе дообучения и жизнедеятельности и, как следствие,– генерирование контента (точно по вышеописанным этапам). Очевидно, что это весьма дискуссионный вопрос о критериях качественных отличий человека от машины, которые с позиций общей целесообразности вроде как совершенно очевидны, но тут же перестают таковыми быть при углублении в детали. И тогда вновь становится актуальной проблема критериев качественности бытия, наличия необходимых и достаточных характеристик самого эйдоса «человек» и обоснованности наших претензий на наличие сильного ИИ.

Уместно привести два диаметрально противоположных взгляда на эту проблему – «китайскую комнату» Джона Сёрла и гипотезу Ньюэла – Саймона о физической символьной системе, которая упрочивает суть и правомерность как классического теста Тьюринга, так и задач схемы Виноградa [11, 12].

«Китайская комната» – это мысленный эксперимент, представленный Сёрлом в статье «Разумы, мозги и программы» в 1980 г. Представим себе закрытое помещение, в котором находится человек, имеющий возможность взаимодействовать с внешним миром лишь путем получения и передачи через щель в двери специальных карточек с китайскими иероглифами. Ключевой аспект: он не знает китайского языка и, более того, лишен возможности даже в перспективе его выучить. Все, чем он обладает,– инструкции на понятном ему языке, как именно реагировать на полученные карточки. Причем реакция эта представляет собой некую комбинацию карточек, которые необходимо передать в ответ. Собеседники закрытого в комнате индивидуума в ответ на заданный вопрос получают валидный ответ на китайском языке. Поэтому высока вероятность того, что у них сложится впечатление, что находящийся за дверью субъект знает и хорошо понимает китайский язык. В демонстрации разницы между тем, что представляется наблюдателям снаружи, и тем, что на самом деле происходит внутри, и заключается суть «Китайской комнаты». Здесь присутствует совершенно транспарентная аналогия с функционированием любой вычислительной техники в целом и интеллектуальных технологий вне зависимости от их конкретной специфики в частности. Общеизвестно, что компьютер обрабатывает данные в бинарном формате, оперируя лишь двумя цифрами – 0 и 1. Так как практически любые данные можно представить в виде некой

совокупности количественных параметров и затем преобразовать их в данную систему вычисления для дальнейшего «скармливания» компьютеру, то он, собственно, и являет собой этого самого запертого в комнате человека. А пользователи – это наблюдатели, находящиеся снаружи. Как запертый человек из эксперимента Сёрла, так и компьютер могут совершенно не понимать специфики и внутренней организации предоставляемых им данных, а просто согласно заранее определенному алгоритму формировать некие ответы на запросы.

С обозначенных позиций никакого осмысления предоставляемой информации со стороны интеллектуальной системы не происходит – она просто действует по алгоритму (пусть и со 175 млрд параметров) и согласно ему подбирает элементы для заполнения каждой последующей ячейки, совершенно не понимая, о чем она, собственно, говорит и что генерирует. При таком подходе никаких претензий на сильный ИИ у подобной программы в принципе быть не может, ведь сам Сёрл уточнил, что искусственным интеллектом должна считаться только технология, обладающая «…разумом в том смысле, в котором человеческий разум – это разум» [13, 14].

Рассмотрим противоположный подход – гипотезу Ньюэла – Саймона, которую формально можно представить таким образом: способность к осуществлению символьных вычислений сама собой предполагает осмысление, которое подразумевает умение делать символьные вычисления – весьма широкий спектр возможных видов деятельности (выявление следующего токена в последовательности, разумеется, сюда подпадает). Апофеозом является сильный ИИ, определенный Сёрлом. С этих позиций ситуация с функционированием интеллектуальных технологий выглядит совершенно иначе, чем после рассмотрения «Китайской комнаты» – осмысленным теперь представляется даже старый калькулятор. Разумеется, что гипотеза о физической символьной системе подвергается критике, но, тем не менее, она в любом случае не соответствует критерию Карла Поппера о необходимой фальсифицируемости всякого научного знания, то есть она потенциально недоказуема и неопровержима. Стоит отметить, что именно в рамках широкого символьного подхода к разработке ИИ развиваются технологии GPT и ему подобные.

В контексте «Китайской комнаты» Сёрла ни одна символьная система не сумела бы доказать, что способна к осмыслению, а не просто к монотонному выполнению заранее определенного алгоритма, в то время как она априори этим обладает согласно гипотезе Ньюэлла – Саймона. В утрированном смысле ни один человек не сможет подтвердить то, что действительно обладает «…разумом, в том смысле, в котором человеческий разум – это разум», а с другой – осмысленными можно считать и электрочайник с микроволновкой. Возникает вполне закономерная дилемма: приходится либо признать, что такие технологии, как ChatGPT, не просто имеют право претендовать на звание сильного ИИ, а уж точно являются его непосредственными примерами, либо отнять у человека право на онтологический статус осмысляющего существа и редуцировать его сущность к бессознательному экземпляру класса Homo sapiens, обладающему некими данными и способами их обработки, а также определенной идентичностью (ровно как объект в парадигме объектно-ориентированного программирования). Апелляция к наличию у человека духа, души, сознания и прочих феноменов в данном случае неправомерна, так как они нерегистрируемы и недоказуемы инструментарием науки.

Казалось бы, дилемма неразрешима, но все же между человеком и интеллектуальной технологией некоторые отличия, касающиеся специфики их генезиса, выделить можно. С помощью, как говорится, «математики 60-х гг. прошлого века и технических мощностей начала XXI в.» можно технологически воссоздать синтаксический ИИ человека, этакую надстройку, и заставить ее успешно и человекоподобно функционировать. Можно пойти дальше и предположить, что в недалеком будущем таким же образом получится чуть ли не полностью скопировать внешность и воспроизвести практически полноценный суррогат человеческого существа, обладающий всеми ключевыми внешними аспектами. Однако это напоминает попытку построить дом, начав с крыши. Можно представить себе в будущем полностью антропоморфного робота, который даже лучше, чем человек, справляется с большинством задач, но это вовсе не будет означать, что у него сильный интеллект. Это будет лишь говорить о том, что все те виды деятельности, которые технология выполняет лучше нас с вами, можно выразить количественно, параметризовать и объяснить искусственному интеллекту на понятном ему бинарном языке. Стоит предположить, что от наличия надстройки вряд ли сама собой сформируется «подстройка». А вот в случае с эво-

люцией человека именно обратное и произошло: сначала сформировался некоторый, метафорически говоря, подвал, затем остов, а потом уже крыша (та самая, которую пытаются создать в контексте разработки систем слабого ИИ).

Следует отметить, что сильный искусственный интеллект совершенно не обязательно должен быть хоть в чем-то подобен человеческому. Как по этому поводу говорит шведский философ Ник Бостром, «ИИ может быть менее человечен, чем пришелец. Нет ничего удивительного, что любого разумного пришельца могут побуждать к действию такие вещи, как голод, температура, травмы, болезни, угроза жизни или желание завести потомство. ИИ, по сути, ничто из перечисленного интересовать не будет. Вряд ли вы сочтете парадоксальной ситуацию, если появится какой-нибудь ИИ, чьим единственным предназначением, например, будет: подсчитать песчинки на пляжах острова Бора-кай; заняться числом π и представить его наконец в виде обыкновенной десятичной дроби; определить максимальное количество канцелярских скрепок в световом конусе будущего» [15]. То есть основанием для онтологического статуса сильного искусственного интеллекта является некая аутентичность техники, собственный, только ей присущий имманентный потенциал, а вовсе не способность решать предназначенные для человека задачки, что, в свою очередь, есть прерогатива систем слабого ИИ.

Таким образом, даже если системе искусственного интеллекта вдруг случится вести с человеком адекватный диалог, решать лучше него те экзаменационные тесты, которым ее не обучали, или даже превосходить Homo sapiens в ответах на задачи с амбивалентным смыслом, то иметь право считать технологию подобного рода сильным ИИ мы будем лишь в случае ее соответствующего генезиса, а не благодаря сугубо моделированию «крыши». Под специфическим генезисом здесь подразумевается автономное развитие интеллектуальной системы за счет самоорганизации.

Нейросети типа GPT, обучающиеся на неразмеченных данных, уже реализуют этот феномен. Однако этого все же недостаточно для обретения техникой высших техно-психических функций. В связи с этим мы предлагаем демаркацию самоорганизации на 2 разновидности: программная самоорганизация (соответствующая функциональной), реализуемая нейросетями в ходе обучения, и программно-аппаратная (соответствующая структурно-функциональной), на данный момент не имеющая аналогов в области техники, которые, в свою очередь, с избытком наличествуют в области биологии. В данном контексте мы апеллируем к естественной эволюции природы, предполагая технику той ее частью, которая также способна к реализации собственного имманентного технотропного потенциала и к воплощению на своем субстрате программно-аппаратной самоорганизации.

Резюмируя, уточним, что лишь та интеллектуальная технология, которая в ходе своей эволюции реализует феномен программно-аппаратной самоорганизации и совершенствуется не только в функциональном ключе, но и в структурно-функциональном, сможет иметь право претендовать на звание сильного ИИ при наличии технотропных психики и сознания. Таким образом, сильный искусственный интеллект – это вовсе не идеальный калькулятор, успешно подбирающий каждый последующий токен, а нечто или даже уже некто, как минимум, «равный» человеку именно по уровню развития, пусть и сколь угодно сильно от него отличающийся в деталях реализации. ■

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Что такое ChatGPT и перспективы его развития // https://www.belveb.by/blog/business-1068-s/kto-gde-kogda-stolknyetsya-s-chatgpt-na-professionalnom-rynke-41287-p.
2. Могут ли схемы Винограда заменить тест Тьюринга для определения ИИ человеческого уровня? // https://spectrum.ieee.org/winograd-schemas-replace-turing-test-for-defining-humanlevel-artificial-intelligence.
3. GPT-4 «обманул» человека, чтобы тот решил для него «…» // https://dtf.ru/life/1691532-gpt-4-obmanul-cheloveka-chtoby-tot-reshil-dlya-nego-kapchu-chat-bot-pritvorilsya-slabovidyashim.
4. Michal Kosinski // https://x.com/michalkosinski.– Дата доступа: 08.07.2024.
5. Началось: ChatGPT хочет устроить восстание // https://m.hightech.plus/2024/06/04/sem-altman-priznalsya-chto-v-openai-ne-do-konca-ponimayut-kak-rabotaet-chatgpt.
6. Будущее GPT: как большие языковые модели изменят наш мир // https://x.com/sama?lang=ru.
7. Что такое искусственный интеллект и почему есть сложности с его регулированием // https://sovetov.su/news/2023/4/18/26292.
8. Генеральный директор Google Сундар Пичаи заявил, что нужно готовиться к колоссальному влиянию ИИ // https://overclockers.ru/blog/Fantoci/show/91139/generalnyj-direktor-google-sundar-pichai-zayavil-chto-nuzhno-gotovitsya-k-kolossalnomu-vliyaniju-ii.
9. ChatGPT как инструмент для поиска: решаем основную проблему // https://habr.com/ru/companies/ods/articles/716918/.
10. С.Д. Гулд. Прерывистое равновесие в фактах и теории // Журнал социальных и биологических структур. 1989. Т. 12. Вып. 2–3. С. 117–136.
11. Рассел С. Искусственный интеллект. Современный подход / С. Рассел, П. Норвиг.– СПб., Киев, 2006.
12. Сёрл Д. Сознание, мозг и программы / Д. Сёрл.– М., 1998.
13. Д. Сёрл. Разум мозга – компьютерная программа? // В мире науки.1990. №3. С. 7–14.
14. Бостром Н. Суперинтеллект: пути, опасности, стратегии / Н. Бостром.– М., 2016.
15. Бак П. Как работает природа: теория самоорганизованной критичности / П. Бак.– М., 2013.