

Применение генеративного искусственного интеллекта в научных исследованиях и в прикладной аналитике в обеспечение государственного управления: позитивные ВОЗМОЖНОСТИ И «ПОДВОДНЫЕ КАМНИ»

И.В. Понкин

Аннотация – Статья посвящена исследованию роли и значения концепта и технологий генеративного искусственного интеллекта применительно к сфере государственного управления и функционирования системы органов публичной администрации. В статье раскрыто понятие генеративного искусственного интеллекта, в том числе моделей GPT – трансформеров, предобученных на генерацию текстов на естественных языках. Автор технические и организационные предпосылки, условия и аспекты задействования генеративного искусственного интеллекта в государственном управлении и в его обеспечении. В статье представлен формализованный перечень позиций задействования моделей генеративного искусственного интеллекта (в том числе моделей GPT) в государственном управлении. Исходя из установки о том, что следует четко представлять и понимать достоинства и потенциальные преимущества (позитивный потенциал), а равно недостатки таких программных моделей, риски, скрытые проблемы («подводные камни») и негативные побочные эффекты, которые они могут повлечь, автор обращается к этим вопросам. Автором изложен перечень позиций возможного релевантного задействования программных моделей генеративного искусственного интеллекта (в том числе моделей GPT) в работе практика-аналитика, учёного, эксперта. Автор обобщает и объясняет проблемы и «подводные камни» программных моделей генеративного искусственного интеллекта (в том числе моделей GPT) при их непосредственном задействовании в научно-исследовательском, прикладном аналитическом и экспертном обеспечении государственного управления. В статье показана проблема текущей быстрой исчерпаемости потенциала развития производительности моделей генеративного искусственного интеллекта. Статья подготовлена в рамках работы над НИР № 12.26.-2024-2 «Перспективы и детерминации развития применения цифровых технологий в государственном управлении» (Институт государственной службы и управления РАНХиГС при Президенте РФ).

Ключевые слова — цифровизация государственного управления, генеративный искусственный интеллект, модель GPT, правовая и государственно-управленческая

Статья получена 30 ноября 2024 г.

И.В. Понкин – Институт государственной службы и управления Российской академии народного хозяйства и государственной службы при Президенте РФ, доктор юридических наук, профессор (e-mail: ponkin-iv@ranepa.ru).

прикладная аналитика, наука, государственное управление.

I. ВВЕДЕНИЕ

Вы попросили машину сделать невозможное, и машина выполнила просьбу.
Курт Воннегут. Сирены Титана [1].

Как и многие другие инновации, ChatGPT находится на грани между пугающим и захватывающим. *Сандип Мехта [2].*

Акцент и значение, придаваемые научно-исследовательскому и прикладному аналитическому ремеслу и мастерству, в разных сферах (включая и сферу государственного управления) менялись на протяжении времени, и в настоящее время явно не могли не претерпеть влияния новейших информационно-коммуникационных технологий, прежде всего – технологий искусственного интеллекта (ИИ).

Одним из релевантных инструментариев в научных исследованиях и в прикладной аналитике стали модели генеративного искусственного интеллекта, в том числе – специфические компьютерно-программные модели – **генеративные предобученные трансформеры текстов на естественном языке** (более точно – «трансформеры, обученные на генерацию текста»; англ. – «*generative pre-trained transformer*» (GPT); франц. – «*transformeur génératif pré-entraîné*»; испан. – «*transformadores generativos preentrenados*»), реализуемые нейронно-сетевыми моделями под управлением юнитов генеративного искусственного интеллекта (далее – модели GPT – для систем, производящих тексты или программные коды). Модели GPT задействуются для решения задач обработки естественных языков (тех, на которых говорят, коммуницируют люди) и способны по запросам (англ. – «*prompt*») пользователя (оператора) в интерфейсе интерактивного чата генерировать связные тексты, воспринимаемые человеком как осмысленные. В основе моделей GPT заложены **пользовательские большие языковые модели** (англ. – «*large language model*» (LLM)).

Несколько проще излагая, **генеративный искусственный интеллект** – это юнит искусственного интеллекта, функционально-целевым образом предназначенный и способный на основе

массированных (сверхбольших) входных обучающих данных генерировать и генерирующий относительно и условно новые цифровые тексты, цифровые статические и динамические (видео) изображения, цифровые аудио продукты и другие цифровые медиа, или же способный генерировать и корректировать программные коды.

Чатбот – это компьютерно-программная модель, предназначенная для машинной (компьютерно-программной, алгоритмически детерминированной) эмуляции речевого коммуницирования с человеком-оператором (пользователем) на естественных человеческих языках, с выстраиванием подобия бесед в формате «вопрос – ответ», через машинное (компьютерно-программное) распознавание и машинную квази-когнитивную обработку («понимание») ключевых лексем или лексических конструкций.

Модель чат-бота интегрируется в те или иные мобильные приложения мессенджеров (например, Телеграм), в различные интернет-платформы (веб-сайты). Оперирование «обычным» (не генеративным) чат-ботом заранее написанными ответами отличается такие чат-боты от чат-ботов на основе моделей GPT, которые самостоятельно генерируют ответы, исходя из массивов данных, на которых они были предобучены.

Конкретно, модели GPT в разговорной форме может быть задан любой вопрос и получен от неё быстрый и порою даже вполне адекватный письменный (а в каких-то опциях и смоделированный устный) ответ, похожий на производимый человеком ответ [3, с. 1].

Такая модель может быть интегрирована в различные платформы, такие как веб-сайты, мобильные приложения и платформы обмена сообщениями. Конкретно разработанная компанией OpenAI программная модель ChatGPT – это тип программного обеспечения искусственного интеллекта, предназначенного для имитации разговоров с пользователями-людьми. Этот чатбот работает по алгоритмам, запрограммированным на понимание естественного языка и написание соответствующих ответов, как заранее написанных, так и сгенерированных искусственным интеллектом. Модели GPT постоянно совершенствуются с помощью методов подкрепления, обработки естественных языков и глубинного машинного обучения, чтобы улучшить способности понимать и тщательно реагировать на потребности пользователей. Конкретно, модели GPT в разговорной форме может быть задан любой вопрос и получен от неё быстрый и порою даже вполне адекватный письменный (а в каких-то опциях и смоделированный устный) ответ, похожий на производимый человеком ответ [3, с. 1]. Качества и адекватность таких ответов, понятно, разнятся.

Выходит всё большее (особенно после 2013 года) число исследовательских публикаций по тематике генеративного искусственного интеллекта (само это понятие не является таким уж новым) в корреляции с технологическими достижениями данной отрасли [4, с. 119885].

Сегодня известно и доступно немало **моделей генеративного искусственного интеллекта, работающих:** **через телеграм-канал** (<@GPT4Telegrambot>; <@smartspeech_sber_bot>;

<@chatsgpts_bot>; <@yes_ai_bot> и др.); **через Интернет-ресурсы** (<<https://openai.com/index/gpt-4/>>; <<https://stablediffusion.fr/chatgpt4>>; <<https://stablediffusion.fr/chatgpt3>>; <<https://stablediffusion.fr/llama3>>; <<https://chatbot.relipa.vn>>; <<https://googlebard.ru>>; <<https://ya.ru/ai/gpt-4>>; <<https://gemini.google.com>> и мн. др.).

Ряд из этих ресурсов в процессе подготовки настоящего материала был нами протестирован на предмет выявления и обобщения возможностей и недостатков.

II. ТЕХНИЧЕСКИЕ ПРЕДПОСЫЛКИ, УСЛОВИЯ И АСПЕКТЫ ЗАДЕЙСТВОВАНИЯ ГЕНЕРАТИВНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ГОСУДАРСТВЕННОМ УПРАВЛЕНИИ И В ЕГО ОБЕСПЕЧЕНИИ

Генеративный искусственный интеллект (англ. – «Generative AI» (GenAI)) признан переломным моментом для инноваций, позволяющим предприятиям автоматизировать рутинные задачи, повысить качество обслуживания клиентов и помочь в принятии важных решений [5, с. 1].

По словам Бенджамина Фрида, большая привлекательность *ChatGPT* и других больших языковых моделей заключается в том, что они способны генерировать подробные ответы, которые имитируют человеческий разговор, а не выдавать клинические указания для другой веб-страницы. Сканируя миллиарды онто-единиц данных и программируясь на соответствие определённым моделям поведения или предметным областям, эти боты фактически «просматривают» строку текста и предсказывают, какие слова должны стоять далее в последовательности [6].

В развивающемся ландшафте поиска знаний пересечение вычислительных методов и добычи данных стало активной областью исследований, – пишет Маркус Бюлер. Одна из главных задач – найти способы моделирования и понимания данных, полученных из различных источников, и использовать их в качестве основы для дальнейших изысканий, чтобы расширить горизонт понимания. Из-за сверхбольших объёмов данных эта задача остаётся выражено сложной, особенно при разработке стратегий экстраполяции существующих знаний на неизвестные ранее идеи или модели действий. Благодаря использованию больших языковых моделей в научных и прикладных аналитических исследованиях, разработка новых идей и гипотез стала возможным и даже релевантным подходом. Большие языковые модели уже сегодня способны синтезировать сложный уровень понимания, например, переводить с одного языка на другой, на котором модель не обучалась. Например, когда в модель генеративного искусственного интеллекта *Gemini 1.5*, выпущенную в 2024 году, загрузили грамматический справочник по каламангу (языку, на котором говорит очень мало людей), модель приобрела способность переводить с английского на каламанг с мастерством, сравнимым с мастерством человека, изучающего тот же материал. Этот пример демонстрирует, что LLM могут эффективно обучаться, когда новые данные (в данном случае грамматический справочник) наделяют модель

новой способностью. Эти и другие возникающие модели поведения указывают на реальную возможность потенциального использования мощных систем генеративного искусственного интеллекта для получения знаний [7, с. 1–2].

В научной литературе уже подробно описаны и объяснены различные типы моделей генеративного искусственного интеллекта путём представления таксономии, основанной на таких архитектурных характеристиках, как:

– **вариационные автоэнкодеры** (англ. – «*variational autoencoders*» (VAE), автокодировщик, специальная архитектура искусственных нейронных сетей) – это тип автоэнкодера, который сочетает в себе вариационный вывод с архитектурой кодера-декодера; автоэнкодеры состоят из энкодерной сети, которая сопоставляет многомерные данные с низкоразмерным представлением, и сети декодера, которая реконструирует исходный входной сигнал из представления; они обучаются на ёмко-сжатых представлениях входных данных и генерируют новые выборки посредством извлечений из латентного пространства, отработанного в ходе машинного обучения;

– **генеративно-сопоставительные сети** (англ. – «*generative adversarial network*» (GAN)) – состоят из генератора и дискриминатора; обучаются в сопоставительном режиме, при этом генератор генерирует синтетические образцы, чтобы обмануть дискриминатор; справляются с генерацией реалистичных и разнообразных данных; GAN основан на минимаксной игре с нулевой суммой для двух игроков, в которой один игрок выигрывает только тогда, когда другой терпит равные потери: два игрока в GAN – это генератор и дискриминатор;

– **диффузионные модели** (англ. – «*diffusion models*») – это тип генеративной модели, которая работает путём постепенного добавления шума в данные до тех пор, пока они не будут соответствовать желаемому распределению; включают в себя этап шумопоставления, за которым следует этап шумоподавления; они итеративно уточняют зашумлённые входные данные для создания высококачественных выборок; обучение предполагает изучение динамики процесса диффузии;

– **трансформеры** (англ. – «*transformers*») – используют архитектуру кодировщика-декодера и используют механизмы самоконтроля для захвата больших закономерностей; они обычно используются в таких задачах, как машинный перевод, и генерируют согласованные последовательности с помощью контролируемого обучения;

– **языковые модели** (англ. – «*language models*») – часто основанные на рекуррентных нейронных сетях (англ. – «*recurrent neural networks*» (RNN)), такие модели генерируют последовательности, пошагово предсказывая следующую в последовательности лексическую единицу; они обучаются с помощью контролируемого обучения и преуспевают в создании последовательностей на естественном языке;

– **модели нормализации потока** (англ. – «*normalizing flow models*») – в таких моделях

используются слои связи для преобразования данных с сохранением плотности, они обучаются на сложных распределениях, преобразуя простое базовое распределение, обученное с помощью оценки максимального правдоподобия;

– **гибридные модели** (англ. – «*hybrid models*») – сочетают в себе различные архитектуры и методы обучения, чтобы использовать их сильные стороны; обеспечивают гибкость и индивидуальные возможности генерации за счёт интеграции элементов из нескольких моделей [8, с. 1, 10–11, 13, 16].

III. ОРГАНИЗАЦИОННЫЕ ВОЗМОЖНОСТИ И ОПЫТ ЗАДЕЙСТVOВАНИЯ МОДЕЛЕЙ ГЕНЕРАТИВНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ИНТЕРЕСАХ ГОСУДАРСТВЕННОГО УПРАВЛЕНИЯ

Большие языковые модели, пред-обученные на больших наборах данных, получили мощный импульс широкого распространения и внедрения в 2018 году после того, как Google представила программную модель BERT. Впоследствии были выпущены различные другие большие языковые модели, такие как модели GPT от OpenAI (лаборатория исследований искусственного интеллекта из Сан-Франциско, США) [9, с. 1].

ChatGPT был представлен компанией OpenAI в конце 2022 года, и за короткий промежуток времени стал невероятно популярным, собрав более 100 миллионов пользователей. Это сделало его самым быстрорастущим интернет-сервисом или сайтом в истории [2].

Число таких ресурсов растёт. 30 марта 2023 года Bloomberg опубликовал исследовательскую работу, в которой подробно описывается разработка *BloombergGPT*, новой крупномасштабной модели генеративного искусственного интеллекта. Эта большая языковая модель была специально проучена на широком спектре финансовых данных для поддержки разнообразного набора задач обработки естественного языка (NLP) в финансовой отрасли [10].

Генеративный искусственный интеллект превратился в мощную технологию с многочисленными применениями в различных областях, стал важной областью исследований, произведя революцию в различных областях, таких как компьютерное зрение, обработка естественного языка и творческие искусства. Способность генерировать данные, имитирующие характеристики реального мира, может помочь в решении различных задач, включая расширение данных, обнаружение аномалий и создание креативного контента [8, с. 1–2].

По данным *Precedence Research*, объём рынка генеративного искусственного интеллекта оценивался в 10,79 млрд. долларов США в 2022 году и, по прогнозам, составит около 118,06 млрд долларов США к 2032 году, увеличиваясь на 27,02% в период с 2023 по 2032 год. Объём рынка генеративного искусственного интеллекта только в США в 2022 году составил 2,7 млрд долларов США [11]. По другим экспертным оценкам, к 2030 году его ежегодный вклад в мировой ВВП в различных отраслях может составить от 2,6 до 4,4 триллиона долларов. В отдельных отраслях, таких как энергетика,

инвестиции в технологии генеративного искусственного интеллекта, как ожидается, утратятся – с 40 миллиардов долларов в 2023 году до более чем 140 миллиардов долларов к концу десятилетия. Такой резкий рост инвестиций отражает трансформационный потенциал технологий генеративного искусственного интеллекта, особенно в плане повышения производительности, оптимизации бизнес-процессов и изменения цепочек создания стоимости в различных отраслях [5, с. 1]. Согласно отраслевой предиктивной аналитике, модели GPT в последующем будут играть различные весьма значимые роли в маркетинге и корпоративном управлении, операционной деятельности, исследованиях и разработках, инжиниринговом и юридическом секторах [12].

Понятно, что сектор публичного управления никак не мог обойти вниманием эту технологическую сферу.

Одна из моделей GPT выдала такую оценку будущего применения подобных моделей в государственном управлении: «Поскольку искусственный интеллект продолжает развиваться, государственные и местные органы власти начинают изучать, как инструменты искусственного интеллекта, такие как ChatGPT, могут изменить цифровые услуги, которые они предлагают жителям. Чат-боты на базе искусственного интеллекта уже используются многими правительствами для автоматизации услуг, таких как регистрация в *Medicaid*, продление водительских прав, ресурсы для малого бизнеса и многое другое. Однако следующее поколение искусственного интеллекта, включающее ChatGPT, может произвести революцию в способах взаимодействия правительств со своими гражданами». Однако эксперты-люди, как отмечает Бенджамин Фрид, гораздо более осторожны в таких оценках [6].

Тем не менее, внедрение и применение моделей GPT в государственном управлении интенсивно ведутся.

Как пишет Сандип Мехта, государственные и местные органы власти уже несколько лет используют чат-боты в качестве поисковой системы, помогающей жителям найти нужную информацию по различным темам, таким, как принятие и обработка заявлений о безработице и оказание помощи в аренде жилья. В первые дни пандемии COVID-19 многие государства поспешили развернуть чат-боты, чтобы ускорять и обработку подачу заявлений на получение пособий и развенчивать дезинформацию. Автоматизация повторяющихся задач, таких как ввод и анализ данных, позволяет государственным служащим сосредоточиться на более сложных и важных задачах. Управление знаниями: упрощение доступа к информации и её обработки, что ведёт к повышению эффективности процессов. Несмотря на то, что эти новые технологии ещё только появляются, и многие этические, технические и практические вопросы ещё не решены, некоторые города и страны уже начали экспериментировать. В Сингапуре было заявлено, что 90 000 государственных служащих вскоре смогут использовать возможности ChatGPT для проведения исследований и подготовки отчётов и выступлений с помощью инструмента *Pair*, встроенного непосредственно в *Microsoft Office*. В начале 2023 года

назад правительство Португалии объявило о планах по внедрению «Практического руководства по доступу к правосудию»: модель искусственного интеллекта, которую Microsoft помогла разработать, использует базовую технологию ChatGPT для помощи гражданам в решении основных юридических вопросов. По словам чиновников, она будет предоставлять гражданам информацию о судебных разбирательствах, отвечать на вопросы о документах, необходимых, например, для получения разрешения на брак или португальского гражданства. Правительства некоторых других стран пока более осторожно относятся к использованию чатов. Некоторые эксперты считают, что когда правительство использует инструмент искусственного интеллекта, к нему должны применяться иные стандарты, чем когда его использует частная компания. В Великобритании, как сообщается, несколько федеральных ведомств обратились за разъяснениями, разрешено ли им использовать ChatGPT для автоматизации повторяющихся задач по разработке политики, таких как написание электронных писем [2].

Анализ и исследование текстов с помощью так называемых цифровых интеллектуальных инструментов может помочь нам понять глубинный смысл текста и выделить отдельные элементы, значимые для решения конкретной темы. Такой анализ может повлиять на общее развитие общества и науки. Текстовый анализ выделяет отдельные ключевые слова, важные для обсуждаемой темы, и общий подтекст текста, который может указывать на факторы за или против использования технологии генеративного искусственного интеллекта в сфере государственного управления. Термины, сопоставляемые с ключевыми словами, могут выражать определённые дополнительные мнения, которые необходимо учитывать при анализе. В нашем случае термины «правила» и «условия» были связаны с ключевым словом «использование», что может указывать на сильную необходимость регулирования сферы генеративного искусственного интеллекта и, возможно, служить указанием и стимулом для регулирующих органов при формировании политики использования генеративного искусственного интеллекта в государственном управлении [13, с. 92–93].

С помощью рассматриваемых инструментариев возможно проведение прикладных аналитических исследований эффективности государственного управления.

Поскольку уличная бюрократия остаётся краеугольным камнем в изучении государственного управления, понимание далеко идущих последствий внедрения таких инструментов генеративного искусственного интеллекта, как ChatGPT и Bard, имеет первостепенное значение. «Уличные бюрократы» – сотрудники государственного сектора, непосредственно взаимодействующие с гражданами, – являются важнейшими проводниками в принятии решений, которые оказывают неизгладимое влияние на отдельных людей и сообщества. Поэтому тщательное изучение их поведения и парадигм принятия решений необходимо для оптимизации предоставления государственных услуг и определения путей значимого реформирования

политики. Трансформационный потенциал инструментов генеративного искусственного интеллекта, таких как Bard и ChatGPT, в изменении исследований уличной бюрократии подчёркивается их способностью анализировать огромные объёмы текстовых данных, моделировать бюрократическое взаимодействие и раскрывать глубокое понимание работы организаций государственного сектора [14, с. 1].

IV. ЗАДЕЙСТВОВАНИЕ МОДЕЛЕЙ ГЕНЕРАТИВНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА (В ТОМ ЧИСЛЕ МОДЕЛЕЙ GPT) В ГОСУДАРСТВЕННОМ УПРАВЛЕНИИ; ФОРМАЛИЗОВАННЫЙ ПЕРЕЧЕНЬ ПОЗИЦИЙ

В общем объёме позиций **функционально-целевого назначения возможного и релевантного задействования юнитов и технологий генеративного искусственного интеллекта (в том числе моделей GPT) в публичном управлении** и в интересах публичного управления обоснованно артикулировать следующие:

- задействование программных моделей генеративного искусственного интеллекта в научно-исследовательской, прикладной аналитической и экспертной работе в обеспечение и сопровождение государственного управления;

- задействование программных моделей генеративного искусственного интеллекта под тестирование регуляторных и других управленческих решений в государственном управлении (модель GPT как цифровой ассистент-компаньон должностного лица, уполномоченного на издание/принятие управленческих решений);

- задействование моделей генеративного искусственного интеллекта (в том числе моделей GPT) по реализации коммуникации между различными органами государственного управления, а также между структурными подразделениями внутри таковых органов;

- задействование моделей генеративного искусственного интеллекта для решения повторяющихся задач в формировании, корригировании и предложении шаблонов писем, планов, отчётов и других документов);

- задействование программных моделей генеративного искусственного интеллекта для предоставления централизованной кибер-платформы для осуществления коммуникации между органами государственного управления и населением, а также между органами государственного управления и субъектами (акторами) предпринимательской сферы (ответы на часто задаваемые вопросы, целевое информирование, предоставление персонализированной информации и поддержки; помощь в общении и взаимодействии с должностными лицами, которые не могут присутствовать на личных встречах или публичных консультациях и др.);

- расширения участия общественности в делах государства – опять же через налаживание информирования (в обеспечение должной транспарентности) о проектах и планах, формируемых и/или реализуемых органами государственного управления;

– задействование программных моделей генеративного искусственного интеллекта под цифровых ассистентов-компаньонов судей (во всяком случае, судебный нормоконтроль, административно-судебное обжалование административного управленческого акта и разбирательство по этому поводу, рассмотрение споров о компетенции органов государственной власти – всё это относимое к государственному управлению в широком понимании и толковании).

V. ОПЫТ И ПЕРСПЕКТИВЫ ЗАДЕЙСТВОВАНИЯ ГЕНЕРАТИВНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ И В ПРИКЛАДНОЙ АНАЛИТИЧЕСКОЙ РАБОТЕ

Модели GPT, реализуемые на основе больших языковых моделей, собирающих обширные данные для получения ответов (подобных человеко-производимым) на запросы пользователей, уже сегодня способны выполнять различные релевантные задачи, включая генерацию языка, анализ тональности, категоризации текстов и ответы на вопросы [15, с. 878].

Существует ряд инструментов генеративного искусственного интеллекта, релевантно задействуемых под решение задачи создания научно-исследовательских, прикладных аналитических и экспертных задач:

- 1) машинные (компьютерно-программные) средства оперирования естественными языками (англ. – «*natural language processing*» (NLP)) – для осмысленного восприятия и генерирования текстовых или речевых продуктов на естественном человеческом языке, которые могут использоваться для обобщения научных или прикладных аналитических материалов или данных, для создания новых текстовых продуктов и выполнения других языковых задач;

- 2) машинные (компьютерно-программные) средства машинного обучения, которые используются для аналитики и обучения на основе больших наборов данных; эти алгоритмы могут использоваться для создания рисунков и таблиц, выявления закономерностей и тенденций в научных исследованиях;

- 3) машинные (компьютерно-программные) средства автоматической генерации естественного языка, которые используются для автоматической генерации текста на основе входных данных; эти алгоритмы могут использоваться для создания резюме научных статей, создания рисунков и таблиц, а также написания научных или аналитических текстов;

- 4) машинные (компьютерно-программные) средства автоматической вычитки и редактирования, которые используются для выявления и исправления ошибок и несоответствий в научных работах; эти инструменты можно использовать для повышения качества и точности научных или аналитических текстов, а также для редуцирования в них ошибок и расхождений [16, с. 2].

Высокое значение моделей GPT, способных создавать сложные иллюстративные образы по последовательностям текстовых запросов оператора (<<https://stablediffusionweb.com/ru/>>; StyleGAN от NVIDIA и др.), определяется тем, что эффективная визуализирующая репрезентация аналитического

продукта в доступной адресату-потребителю форме, комфортной для него в восприятии, усвоении и понимании, – является важным залогом понимания представляемых результатов аналитической работы.

По словам Джеффри Белкнапа, «наука в основе своей визуальна. Она опирается на материал – будь то атом, ген, кристалл, кит или далёкая галактика. Её цель – прояснение. Таким образом, передача информации об исследованиях всегда была [и должна быть] основана на сочетании изображений и текста, чтобы передавать сведения об открытиях, идеях и наблюдениях» [17, с. 25].

Модели GPT активно используются для проведения библиометрических исследований [4], для распознавания и транскрибирования текстов из аудиофайлов (например – <https://speech2text.ru>).

Заявляется, что в последнее время искусственный интеллект стал многообещающим инструментом для упрощения и улучшения научного письма. Используя машинное обучение и алгоритмы обработки естественного языка, помощь ИИ может повысить качество и эффективность научного письма. Однако, как и в случае с любой новой технологией, использование помощи ИИ в написании научных работ имеет как преимущества, так и недостатки [18, с. 1].

Хотя не критическое восхваление и рекламирование возможностей моделей GPT сильно их преувеличивают, тем не менее, сухой релевантный остаток всё же есть: генеративные предобученные трансформеры текстов на естественном языке, действительно, потенциально способны оказать помощь человеку в процессе подготовки аналитического продукта.

Следует чётко представлять и понимать достоинства и потенциальные преимущества (позитивный потенциал), а равно недостатки таких программных моделей, риски, скрытые проблемы («подводные камни») и негативные побочные эффекты, которые они могут повлечь.

VI. ВОЗМОЖНОСТИ РЕЛЕВАНТНОГО ЗАДЕЙСТВОВАНИЯ ПРОГРАММНЫХ МОДЕЛЕЙ ГЕНЕРАТИВНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ, ПРИКЛАДНОЙ АНАЛИТИЧЕСКОЙ И ЭКСПЕРТНОЙ РАБОТЕ В ОБЕСПЕЧЕНИЕ И СОПРОВОЖДЕНИЕ ГОСУДАРСТВЕННОГО УПРАВЛЕНИЯ

Полагаем, в числе позиций релевантного задействования программных моделей генеративного искусственного интеллекта (в том числе моделей GPT) в работе практика-аналитика, учёного, эксперта обоснованно указать следующие:

1) входное поисковое и реферативно-аннотирующее обеспечение и введение в тему:

1.1) быстрое первичное «входное» реферативно-теоретическое (знаниевое) введение в тему по основным её элементам, аспектам и направлениям посредством истребования и получения кратких реферативных обзоров по задаваемым итеративно-корректируемым последовательностям вопросов;

1.2) быстрое первичное введение (погружение) в исследуемое «проблемное поле»;

1.3) первичный прикидочный поиск наводок на

референтные нормативные или судебные акты, документы, формирование ссылок-подсказок таковых (обязательно нуждается в проверке и подтверждении, поскольку модели GPT нередко фабрикует ложные такие ссылки);

1.4) работа со скриншотами;

1.5) составление библиографического или библиографическо-реферативного обзора существующих по принятой в разработку теме публикаций; проведение обеспечительного библиометрического анализа;

2) обеспечение непосредственной работы над текстовым продуктом:

2.1) переводческое обеспечение – перевод с одного естественного языка на другой;

2.2) распознавание и транскрибирование текстов из аудиофайлов, когда наговорённые устно материалы переводятся в текстовые формы (например: `<@smartspeech_sber_bot>`);

2.3) генерирование (в ответ на информативно-насыщенные запросы и итеративно-адаптируемые последовательности таких запросов) набросков фрагментов черновиков рукописей текстовых материалов через отработку полученных на вопросы ответов;

2.4) текущее уточнение проблемных позиций в составляемом или уже составленном человеком текстовом продукте;

2.5) помощь в формулировании перечня вопросов для интервью или опросов, если создаваемый текстовый продукт предполагает такую часть;

2.6) оптимизация структурирования текстового материала, с предложением структуры и возможных формулировок наименований разделов;

2.7) генерация и корректирование визуально-иллюстративного обеспечения рукописи текстового материала (детализированных и реалистичных статических иллюстраций и видео, а также таблиц, слайдов и других визуальных элементов), помогающего обобщать данные и выводы и содействующего улучшенному восприятию текстового материала читателем; выполняется как по отдельным *prompt*-командам оператора, так и на основе машино-аналитической обработки целостного указанного текстового материала (в настоящее время модели GPT пока не обладают сколь-нибудь полноценной самостоятельной способностью визуализировать данные, но определённые позитивные тенденции к совершенствованию имеются);

2.8) создание (кодирование) программных продуктов; выявление дефектов и проблем кодирования в программном продукте;

3) пред-финальное редакционно-оформительское сопровождение:

3.1) проверка составленного человеком текстового продукта на предмет, не было ли что-то упущено (забыто) из существенно важного;

3.2) машинное редактирование, синтаксическая и грамматическая корректура; выявление областей двусмысленностей, выраженной неясности, внутренних ошибок и несоответствий (например, <https://www.deepl.com/ru/write>);

3.3) машинное форматирование текста под заданные требования;

3.4) составление резюме (аннотации) статьи (некоторые журналы требуют большие резюме, к примеру, – в 200 слов, при запрете воспроизведения кусками текста из статьи);

3.5) проверочная детекция плагиата в тексте (лишь в определённых пределах может быть релевантной);

3.6) формирование упрощённой и сокращённой модальности готового текстового материала для аудитории «попроще», перефразируя сложные лексические конструкции и предложения для того, чтобы сделать текст более ясным (не профильному специалисту) и лаконичным, сохраняя при этом исходное смысловое значение (например, <<https://www.deepl.com/ru/write>>);

3.7) составление библиографического списка в соответствии с разными стандартами (APA, MLA, Harvard и др.).

Однако технологии и юниты генеративного искусственного интеллекта (включая модели GPT) на сегодня ещё не достигли уровня надлежащей (сообразно потребностям и требованиям государственного сектора) проработанности и «отточенности», некритическое их восприятие и неосновательное ими восхищение только повредит делу, будет воспрепятствовать их продуманному и релевантному задействованию.

VII. ДЕВИАНТОЛОГИЧЕСКИЙ ВЗГЛЯД НА ПРОБЛЕМЫ ЗАЩИТЫ КОНФИДЕНЦИАЛЬНОСТИ ПРИ ЗАДЕЙСТВОВАНИИ МОДЕЛЕЙ ГЕНЕРАТИВНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Отсутствие надлежащего нормативно-правового регулирования разработки, тестирования, внедрения, задействования и контроля технологий и юнитов генеративного искусственного интеллекта, приводит к путанице и выраженным неопределённостям в отношении того, как и кем, в каких пределах такие технологии должны использоваться. Так, весьма сильно спровоцировала и усилила беспокойство по поводу конфиденциальности ситуация 23.03.2023 у компании OpenAI, вынужденной приостановить функционирование и временно отключить модель GPT от сети для исправления программных дефектов и сбоев, повлёкших то, что история чатов одних пользователей стала доступна для обозрения другим пользователям. После этого сбоя регулирующие органы Италии ввели временный запрет на использование ChatGPT в своей стране, ссылаясь на проблемы с конфиденциальностью [2].

В своей политике конфиденциальности компания OpenAI заявляет, что собирает все виды личной информации от людей, которые пользуются её услугами. Она утверждает, что может использовать эту информацию для улучшения или аналитической обработки своих услуг, проведения исследований, общения с пользователями и разработки новых программ и услуг, а также для других целей. В политике конфиденциальности говорится, что компания может предоставлять личную информацию третьим лицам без дополнительного уведомления пользователя, за исключением случаев, предусмотренных законом. Если

политика конфиденциальности, состоящая из более чем 2 000 слов, кажется несколько непрозрачной, то это, скорее всего, потому, что в эпоху Интернета это стало практически отраслевым обыкновением. У компании OpenAI также есть отдельный документ «Условия использования», который возлагает на пользователя большую часть ответственности за принятие соответствующих мер при работе с его инструментами. Политика конфиденциальности компании Google, распространяемая и на инструмент Bard, столь же многословна, и у неё есть дополнительные условия обслуживания для пользователей генеративного искусственного интеллекта. Компания заявляет, что для улучшения Bard и защиты конфиденциальности пользователей «мы отбираем подмножество разговоров и используем автоматизированные инструменты для удаления персональной информации». «Эти образцы разговоров могут быть просмотрены обученными рецензентами и хранятся до 3 лет отдельно от вашего аккаунта Google», – говорится в отдельных инструктивных документах по Bard. Компания также предупреждает: «Не включайте в разговоры Bard информацию, которая может быть использована для идентификации вас или других людей». В инструктивных документах также говорится, что разговоры Bard не используются в рекламных целях, и «мы будем чётко информировать о любых изменениях в этом подходе в будущем» [2].

VIII. ПРОБЛЕМЫ И «ПОДВОДНЫЕ КАМНИ» ПРОГРАММНЫХ МОДЕЛЕЙ ГЕНЕРАТИВНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА (В ТОМ ЧИСЛЕ МОДЕЛЕЙ GPT) ПРИ ИХ НЕПОСРЕДСТВЕННОМ ЗАДЕЙСТВОВАНИИ В НАУЧНО-ИССЛЕДОВАТЕЛЬСКОМ, ПРИКЛАДНОМ АНАЛИТИЧЕСКОМ И ЭКСПЕРТНОМ ОБЕСПЕЧЕНИИ ГОСУДАРСТВЕННОГО УПРАВЛЕНИЯ

Инструментальное задействование генеративного искусственного интеллекта (в том числе моделей GPT) в государственном управлении и в исследованиях государственного управления открывает широкие возможности, но и создаёт определённые проблемы и сложности, влечёт определённые риски.

К числу причин ограничений и сбоев моделей GPT (даже самых совершенных и продвинутых к настоящему времени) следует отнести то, что его основной источниковой средой данных для машинного обучения является сеть Интернет, информация в которой может быть и часто является неточной, а источники – ненадёжными.

Как и любые иные инструменты искусственного интеллекта (не только генеративного даже), модель GPT релевантна настолько, насколько хороши данные, на которых она была предобучена и к которым оно могла и может получить доступ для продолжения машинного обучения. Но не только это. Тут задействуется и интерсекционально взаимодействует множество факторов, определяющих качества работы моделей генеративного искусственного интеллекта.

Проблемы с ошибками, выдаваемыми моделями GPT, с ограничениями объяснений и с этическими проблемами (конфиденциальность данных и

прозрачность моделей) существенно ограничивают задействие моделей GPT в настоящее время, например, в сфере здравоохранения. В научных изданиях упоминаются уже весьма многочисленные случаи фактически неверных медицинских оценок, выводов и предписаний, упущения важных медицинских обстоятельств и результатов, дефектных рекомендаций по ненужным исследованиям и избыточному лечению. Эти проблемы в сочетании с проблемами конфиденциальности делают модели GPT в настоящее время непригодными для клинического использования в реальном мире [19].

Модели GPT не способны генерировать принципиально новые идеи, помогая лишь в том, чтобы организовывать (консолидировать и структурировать, увязывать) и развить идеи собственно человека, помогая создавать в черновом варианте (в черновике) первоначальный рабочий проект; однако это, как пишут Микеле Сальвандо, Фабио Сильвио Такконе и Альберто Джованни Герли, – всего лишь отправная точка для разработки текста человеком, поскольку созданный автоматически текст далеко не всегда заменяет знания, творческий потенциал и критическое мышление людей-экспертов [3, с. 2].

Процесс «рассуждения» в моделях генеративного искусственного интеллекта – это много более (практически полностью) вероятностное оперирование паттернами (шаблонами), почерпнутыми из тех массивов, на которых были предобучены эти модели, и сопоставление паттернов (шаблонов), нежели какое-то формальное рассуждение.

Поэтому, по мнению Джонатана Чена, чат-боты искусственного интеллекта принципиально не могут, не в состоянии заменить собой врачей: «Мы предоставляем совершенно иную, конкретную и уникальную ценность, чем компьютер... Существует очень реальный, достоверный риск того, что такого рода ресурсы могут генерировать правдоподобную дезинформацию – без намерения причинить вред, просто начинаются **конфабуляции** генеративного искусственного интеллекта (вымысел, принявший форму воспоминаний) и/или **галлюцинации** генеративного искусственного интеллекта (англ. – «*artificial intelligence hallucinations*»). Это тем более опасно, чем более дезинформация правдоподобно выглядит по содержанию и правдосообразно по форме. И актуальный вопрос здесь артикулируется относительно того, как и насколько врачи могут научиться эффективно использовать генеративный искусственный интеллект и работать с генеративным искусственным интеллектом, какого это может потребовать дальнейшего обучения (подготовки)... Можно, конечно, быть осторожным и не “растягивать” программы больших языковых моделей за рамки их текущих пределов, а например – использовать их в медицине для решения сфокусированно лишь некоторых административных задач, где процент ошибок низок» (цит. по: [20]).

Но аналогично они не способны заменить и лиц, принимающих управленческие решения в системе и процессе государственного управления.

Хотя большие языковые модели могут соответствовать более абстрактным шаблонам

рассуждений, они не способны на истинные самостоятельные логические рассуждения. Небольшие изменения во входных лексических конструкциях могут резко изменить результаты моделирования, что указывает на, что эти модели очень чувствительны и хрупки. Кроме того, в задачах, требующих правильного выбора нескольких маркеров, вероятность получения точного ответа уменьшается экспоненциально с ростом числа маркеров или шагов, что подчёркивает их ненадёжность в сложных сценариях рассуждений. Математические рассуждения – важнейший когнитивный навык, который помогает решать проблемы в многочисленных научных и практических приложениях. Следовательно, способность больших языковых моделей эффективно выполнять задачи математического рассуждения является ключом к развитию в будущем искусственного интеллекта и его реальных приложений. Хотя возможности больших языковых моделей включают простые математические вопросы с подробно излагаемыми решениями (релевантно для таких методов, как подсказка цепочки мыслей), большие языковые модели предоставляют только одну метрику по фиксированному набору вопросов, что существенно ограничивает исследования математических «рассуждений» больших языковых моделей. Широкая популярность и распространение моделей генеративного искусственного интеллекта влечёт риск неумышленного, но весьма существенного «загрязнения» данных [21].

Стивен Куперман и Роберто Брандао указывают, что текущие итерации моделей генеративного искусственного интеллекта имеют объективно обусловленные ограничения в понимании контекста и нюансов научного письма. Это связано с тем, что модели генеративного искусственного интеллекта в первую очередь полагаются на алгоритмы машинного обучения и обработку естественного языка, которые предназначены для выявления закономерностей и взаимосвязей в языке, но не наделены способностями и возможностями рекогнитивной (распознающей) обработки и понимания фигур речи, образных выражений и иных семантических нюансов, а также контекстов, коннотаций и дискурсов, что только и позволяет человеку глубоко погружаться в семантику текста. Экспертная научная или прикладная аналитическая семантика может становиться запредельно трудной для понимания означенными моделями, что может ограничить их способность предоставлять точную обратную связь и предложения. Инструменты искусственного интеллекта не заменяют человеческое творчество и интуицию. Системы искусственного интеллекта обучаются на уже существующих больших массивах данных и выявляют закономерности, но они не способны «мыслить нестандартно» или создавать новые смысловые связи так же, как люди. Кроме того, эти инструменты не способны проверять точность информации, которую они производят, и могут генерировать ответы, основанные исключительно на входных данных, на которых они проучены. Это может привести к созданию ложной и вводящей в заблуждение информации, которая может иметь серьёзные последствия [18, с. 1–2].

Целесообразно обратить внимание на проблему сложности установления, кем написан целостный законченный текст (условно, от страницы объёмом) – чатботом (моделью GPT) или человеком, поскольку, как уже было сказано выше, модели GPT используют передовые технологии, такие, как программирование на естественных языках и машинное обучение, чтобы генерировать тексты, похожие на производимые человеком тексты. Несколько признаков могут указывать на пригодность текста моделью GPT: отсутствие семантических нюансов или оригинальности, что может позволить идентифицировать её детекторами выходных данных искусственного интеллекта и рецензентами-людьми. Эксперты отмечают, что те же самые черты могут быть обнаружены в текстах, написанных на языке, который не является родным для человека. И в силу ряда таких технических причин ИИ-детекторы плагиата будут идентифицировать работы на неродном английском языке как текст, сгенерированный моделью GPT. В тексте, сгенерированном моделью GPT, могут и скорее всего будут отсутствовать тонкие смысловые и стилистические формулировки и специфические аранжировки слов, которые может использовать автор-человек, чтобы передать определённый смысл или тон. Текст от модели GPT также может быть более расплывчатым и содержать несоответствия, которых не было бы в написанном человеком тексте [3, с. 2].

Есть ещё и такая существенная проблема. Как обоснованно указывает Майкл Ан, многие государственные услуги и функции требуют, чтобы кто-то выносил оценочное суждение, например, в случае, связанном с общественным благосостоянием. Соответственно, государственно-управленческие решения, по сути своей, являются «ценностно-нагруженными» – во всяком случае, в таких областях, как общественное благосостояние, здоровье, бездомность. И во всём, что связано с человеческим фактором и требует некоторого оценочного ценностного суждения, предписания и рекомендации модели GPT могут оказаться не лучшим и даже неприемлемым вариантом (цит. по: [6]).

IX. АВТОРСКИЙ КОНЦЕПТ ОПИСАНИЯ И ОБЪЯСНЕНИЯ «ПОДВОДНЫХ КАМНЕЙ» ПРОГРАММНЫХ МОДЕЛЕЙ ГЕНЕРАТИВНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА (В ТОМ ЧИСЛЕ МОДЕЛЕЙ GPT) ПРИ ИХ НЕПОСРЕДСТВЕННОМ ЗАДЕЙСТВОВАНИИ В НАУЧНО-ИССЛЕДОВАТЕЛЬСКОМ, ПРИКЛАДНОМ АНАЛИТИЧЕСКОМ И ЭКСПЕРТНОМ ОБЕСПЕЧЕНИИ ГОСУДАРСТВЕННОГО УПРАВЛЕНИЯ

С учётом сказанного выше, к числу «подводных камней» программных моделей GPT обоснованно отнести следующие:

- значительная вероятность некорректного восприятия и понимания моделью генеративного искусственного интеллекта прямой семантической сути запроса (эксперты говорят о проблеме «иллюзии понимания»);
- полное отсутствие гарантий того, что выдаваемые моделью GPT по запросу тексты не были ранее многократно или в последующем не будут выданы многократно в лексически и стилистически неизменном

виде другим пользователям – авторам схожих запросов (то есть очень высокая вероятность столкнуться в последующем с проблемой маркировки взятых таких ответов модели GPT как плагиата);

- полное отсутствие гарантий того, что выданные моделью GPT целостные транскрипции загруженных для текстовой расшифровки голосовых сообщений (когда осуществляется распознавание и транскрибирование в текстовые формы из аудиофайлов записей наговорённых устно материалов) не будут в последующем «подарены» третьим лицам при ответах на их вопросы;

- отсутствие собственных способностей моделей GPT к сколь-нибудь чёткому и релевантному прогнозированию, к предиктивной (прогностической) аналитике (выдаются лишь перефразированные или иным образом трансформированные чужие прогнозы, либо размытые «прогнозы» в самых общих и неконкретных лексических формах); но любая аналитика, в принципе, является полноценной только при наличии в ней предиктивной и/или прескриптивной (прогноз + предписание) составляющей;

- отсутствие собственных способностей моделей GPT к ценностному пониманию и мышлению;

- проблема наличия у моделей генеративного искусственного интеллекта «галлюцинаций» (англ. – «*artificial intelligence hallucinations*»), под которыми понимается фальсифицирующее фабрикование такими моделями исходно ложных (не существующих и никогда не существовавших в реальности), но правдоподобно по содержанию и правдоподобно по форме выглядящих правовых норм и нормативно-правовых актов, судебных актов, фактов, документов, несуществующих интернет-адресов, референтных теме запроса фальшивых персон и библиографических источников и др. (модель GPT не в состоянии понять, что правда, а что нет);

- весьма посредственные и ограниченные креативность и оригинальность моделей GPT (в нормальном научном понимании научной оригинальности, а не так дефектно, как в современных крайне плохо работающих моделях «антиплагиат»);

- низкая способность моделей GPT к синтетическому высокоуровневому обобщению – выдача весьма содержательно слабых, семантически-расплывчатых и нивелированных по содержанию текстов (образно говоря – отражающих «среднюю температуру по больнице»), неспособность иерархизировать указываемых авторов по основанию авторитетности и общей признанности, их вклада в разработку темы, референтной запросу, неспособность иерархизировать выдаваемые тезисные наброски по основанию их «весовых» характеристик;

- неспособность моделей GPT к критическому экспертному мышлению (в устоявшейся в сознании людей-экспертов интерпретации этого понятия);

- наличие у моделей GPT проблем с классической логикой и, тем более, с логикой нечёткой;

- неспособность моделей GPT к интуитивной аналитике и к формулированию интуитивно понятных потребителю решений или рекомендаций;

- неспособность моделей GPT к пониманию, оцениванию и ранжированию нюансов, тем более в

условиях редундантности (чрезмерной переизбыточности) данных; неспособность моделей GPT понимать контекст и иметь контекстную осведомлённость (англ. – «*context awareness*»), модели GPT игнорируют в обрабатываемых и генерируемых текстах контексты, коннотации и дискурсы; критически слабая способность к связанному со всем этим символическому мышлению;

- дефектность ответов, детерминированная тем, что больше данных в основе – объективно вовсе не обязательно и далеко не всегда лучше, чем данные в меньшем объёме, но более высокого качества; отсюда – транслирование моделями GPT ошибочных или некорректных стереотипов и заблуждений, неявных и даже выраженных системных предвзятостей, извлечённых из больших массивов данных, на которых предобучена данная конкретная модель GPT, не обладая способностями к детекции и отбраковке подобных дефектных семантических единиц;

- фреймированность потенциала данной конкретной модели GPT хронологическими параметрами задействованного для машинного обучения массива данных (у некоторых моделей – до 2022 года);

- существенные трудности у моделей GPT с мультимодальными задачами со множеством переменных, с задачами, выходящими за рамки их предметно-объектных областей пред-обученности или сформулированными в специфических профессиональных технических терминах;

- выраженная дефектогенная чувствительность моделей GPT к вопросам, уже содержащим в своих формулировках указания на ответы (подсказок ответов), и к другим манипулятивным подсказкам;

- недостаток специализированных знаний моделей GPT, что влечёт во многих случаях неспособность модели GPT конкурировать с человеком-экспертом, выдавая весьма посредственные, откровенно слабые результаты;

- очень слабые способности моделей GPT к рассуждениям (хотя тенденция улучшения от модели к модели здесь просматривается);

- существенная и выраженная многоплановая дефектность существующих и применяемых программных моделей «антиплагиат»;

- ограничения моделей GPT по числу символов в одном запросе и одном ответе, что влечёт существенные дефекты при работе с крупными объёмами данных и со сложными задачами, требующими обширной записи (Модель GPT <https://t.me/chatsgpts_bot> на этот счёт выдаёт такой ответ: *«Максимальный объём ответа в ChatGPT составляет около 4096 токенов. Токены могут включать как слова, так и символы, поэтому в зависимости от языка и структуры текста это может варьироваться в знаках. Обычно это примерно 2000–3000 слов, но точное количество знаков может меняться»*);

- весьма выраженная неоднозначность и неочевидность реальной объективности, действительной непредвзятости модели GPT по очень большому перечню вопросов, контекстов и дискурсов.

Некоторые из выделяемых нами позиций были протестированы (на предмет верификации и валидации)

в форме вопросов – на самих моделях GPT с помещением цитируемых ответов в сносках, И полученные ответы, сами по себе, убедительно показывают существенную ограниченность таких программных моделей для их неадаптированного применения в государственном управлении.

Х. ПРОБЛЕМА ТЕКУЩЕЙ БЫСТРОЙ ИСЧЕРПАЕМОСТИ ПОТЕНЦИАЛА РАЗВИТИЯ ПРОИЗВОДИТЕЛЬНОСТИ МОДЕЛЕЙ ГЕНЕРАТИВНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

В настоящее время ключевые разработчики моделей генеративного искусственного интеллекта (в том числе моделей GPT) достигли ситуации достижения, вероятно, предельного уровня развития производительности моделей генеративного искусственного интеллекта, во всяком случае с интенсивностью и производительностью, сопоставимыми с аналогичными позициями в предшествовавшие периоды.

OpenAI планировала в сентябре 2024 года завершить первый раунд обучения новой массивной модели искусственного интеллекта, который, как высказывали ожидания в компании, значительно должен был превзойти предыдущие версии технологии, лежащей в основе модели ChatGPT, и приблизить к цели создания мощного искусственного интеллекта, превосходящего человека. Сообщается, что модель, известная внутри компании как Orion, не достигла желаемых показателей. В конце лета 2024 года модель Orion не справлялась с задачами, когда пыталась ответить на вопросы по программированию, которым он не была предобучена. В целом модель Orion пока не считается таким же большим шагом вперёд по сравнению с существующими моделями OpenAI, как GPT-4 по сравнению с GPT-3.5, системой, на которой изначально работал флагманский чатбот компании. Компания OpenAI не одинока в том, что недавно столкнулась с «камнями преткновения»; после нескольких лет стремительного выпуска всё более совершенных продуктов генеративного искусственного интеллекта три ключевые компании-разработчики таких технологий в текущий период столкнулись с экспоненциальным снижением отдачи от своих дорогостоящих усилий по созданию новых продвинутых моделей. В компании Google предстоящая итерация программного обеспечения Gemini тоже пока не оправдывает внутренних ожиданий. Компания Anthropic, в свою очередь, столкнулась с тем, что сроки выпуска долгожданной модели Claude под названием 3.5 Opus срываются. Компании сталкиваются с несколькими проблемами. Становится всё труднее находить новые, неиспользованные источники высококачественных обучающих данных, созданных людьми, которые можно использовать для создания более совершенных систем генеративного искусственного интеллекта. Неудовлетворительные результаты кодирования Orion отчасти объясняются отсутствием достаточного количества данных для обучения. В то же время даже скромных улучшений может оказаться недостаточно, чтобы оправдать огромные затраты, связанные с созданием и эксплуатацией новых моделей, или

оправдать ожидания, связанные с рекламой продукта как серьёзного обновления. Чтобы модели генеративного искусственного интеллекта действительно хорошо справлялись с различными задачами, могут потребоваться усложнённые варьируемые подходы к глубинному машинному обучению [22].

По словам Дарио Амодеи, что есть «много вещей», которые могут «сорвать» процесс достижения уровня существенно более мощного генеративного искусственного интеллекта в ближайшие несколько лет, включая возможность того, что «у нас полностью исчерпаны данные» (цит. по: [22]).

XI. ЗАКЛЮЧЕНИЕ

Описанные выше проблемы и «подводные камни» негативно влияют на производительность, надёжность и адекватность такого рода моделей, в принципе, и тем более – применительно к обсуждаемой тематике цифровизации государственного управления. В числе иных детерминантов, всё это определяет и тенденцию падения качества текстов, генерируемых («наводняемых») моделями GPT. Но в сфере государственного управления это может приводить к весьма тяжким и непоправимым последствиям с существенными размерностями причиняемого ущерба.

Есть ещё проблема отсутствия прозрачности функционирования моделей GPT, что редуцирует возможности рационального понимания и интерпретации процессов принятия решений моделью GPT, как следствие – это существенно снижает возможности не критического восприятия и принятия предписаний или рекомендаций модели GPT, снижает доверие к таким моделям.

Необходимо тщательно разобраться с требованиями и метриками оценки, условиями и пределами применения программных тех или иных моделей генеративного искусственного интеллекта (в том числе моделей GPT) при их непосредственном задействовании в научно-исследовательском, прикладном аналитическом и экспертном обеспечении государственного управления, в принятии управленческого решения, чётко определить такие требования и метрики оценки, такие условия и пределы.

Зная показанные в настоящей статье недостатки и достоинства исследуемых моделей, можно вполне релевантно использовать их возможности.

БЛАГОДАРНОСТИ

Настоящим автор выражает благодарность Куприяновскому Василию Павловичу, систематически помогающему интереснейшими материалами и обращающему внимание на интереснейшие аспекты, а также выражает глубокую признательность редакции настоящего журнала.

БИБЛИОГРАФИЯ

- [1] *Vonnegut K.* Sirens of Titan [Сирены Титана]. – London: Coronet Books; Hodder & Stoughton, 1982.
- [2] *Mehta S.* How should government agencies use generative AI like ChatGPT? [Как государственные учреждения должны использовать генеративный искусственный интеллект, подобный ChatGPT?] // <<https://www.publicceo.com/2023/05/how-should-government-agencies-use-generative-ai-like-chatgpt/>>. – 01.05.2023.
- [3] *Salvagno M., Taccone F.S., Gerli A.G.* Can artificial intelligence help for scientific writing? [Может ли искусственный интеллект помочь в написании научных статей?] // *Critical Care.* – 2023. – Vol. 27. – № 1. – Article № 75. – P. 1–5.
- [4] *Dwivedi R., Elluri L.* Exploring Generative Artificial Intelligence Research: A Bibliometric Analysis Approach [Изучение исследований генеративного искусственного интеллекта – подход к библиометрическому анализу] // *IEEE Access.* – 2024. – Vol. 12. – P. 119884–119902.
- [5] *Agentic AI – the new frontier in GenAI: An executive playbook* [Агентный искусственный интеллект – новый рубеж в генеративном искусственном интеллекте: Руководство к действию]. – London: PricewaterhouseCoopers, 2024. – 21 p.
- [6] *Freed B.* ChatGPT said it will ‘revolutionize’ digital government. Human experts are more cautious [ChatGPT заявляет, что это «революционизирует» цифровое правительство. Человеческие эксперты более осторожны] // <<https://statescoop.com/chatgpt-government-digital-services/>>. – 04.04.2023.
- [7] *Buehler M.J.* Accelerating scientific discovery with generative knowledge extraction, graph-based representation, and multimodal intelligent graph reasoning [Ускорение научных открытий с помощью генеративного извлечения знаний, представления на основе графов и мультимодального интеллектуального рассуждения на основе графов] // *Machine Learning: Science and Technology.* – 2024. – Vol. 5. – № 3. Article № 035083. – P. 1–58.
- [8] *Bandi A., Ramesh Adapa P.V., Kumar Kuchi Y.E.* The Power of Generative AI: A Review of Requirements, Models, Input–Output Formats, Evaluation Metrics, and Challenges [Сила генеративного искусственного интеллекта: обзор требований, моделей, форматов ввода-вывода, метрик оценки и проблем] // *Future Internet.* – 2023. – Vol. 15. – № 8. – Article 260. – P. 1–60.
- [9] *Saka A., Taiwo R., Saka N. et al.* GPT models in construction industry: Opportunities, limitations, and a use case validation [Модели GPT в строительной отрасли: Возможности, ограничения и проверка на практике] // *Developments in the Built Environment.* – 2024, March. – Vol. 17. – Article № 100300. – P. 1–29.
- [10] *Haas C., Gilmore A.* Introducing BloombergGPT, Bloomberg’s 50-billion parameter large language model, purpose-built from scratch for finance [Представляется BloombergGPT – большая языковая модель Bloomberg с 50 миллиардами параметров, специально созданная для финансов] // <<https://www.bloomberg.com/company/press/bloomberg-gpt-50-billion-parameter-llm-tuned-finance/>>. – 30.03.2023.

- [11] Generative AI Market Growth is Booming with 27.02 % [Пост рынка генеративного искусственного интеллекта составляет 27,02 %] // <<https://www.precedenceresearch.com/insights/generative-ai-market>>.
- [12] *Kundu A.* The case for a ChatGPT usage and adoption policy for public sector administration [Аргументы в пользу политики использования и внедрения ChatGPT для государственного управления] // <<https://sciencepolicy.ca/posts/the-case-for-a-chatgpt-usage-and-adoption-policy-for-public-sector-administration/>>.
- [13] *Lamovšek N.* Analysis of Research on Artificial Intelligence in Public Administration: Literature Review and Textual Analysis [Анализ исследований в области искусственного интеллекта в государственном управлении: Обзор литературы и текстовый анализ] // Central European Public Administration Review. – 2023. – Vol. 21. – № 2. – P. 77–93. – P. 92–93.
- [14] *Salah M., Abdelfattaha F., Al Halbusib H.* Generative Artificial Intelligence (ChatGPT & Bard) in Public Administration Research: A Double-Edged Sword for Street-Level Bureaucracy Studies [Генеративный искусственный интеллект (ChatGPT & Bard) в исследованиях государственного управления: Обоюдоострый меч для исследований бюрократии на уровне улиц] // International Journal of Public Administration. – 04.11.2023. – P. 1–7.
- [15] *Onatayo D., Onososen A., Oyediran A.O. et al.* Generative AI Applications in Architecture, Engineering, and Construction: Trends, Implications for Practice, Education & Imperatives for Upskilling – A Review [Приложения генеративного искусственного интеллекта в архитектуре, проектировании и строительстве: Тенденции, последствия для практики, образования и императивы для повышения квалификации: Обзор] // Architecture. – 2024. – Vol. 4. – P. 877–902.
- [16] *Kammer M.N., Gomila P., Vambaco D.J., Maldonado F.* The role of artificial intelligence in scientific writing [Роль искусственного интеллекта в написании научных работ] // Journal of Clinical Case Reports, Medical Images and Health Sciences. – 2023. – Vol. 3. – № 3. – P. 1–6.
- [17] *Belknap G.* 150 years of scientific illustration [150 лет научных иллюстраций] // Nature. – 07.11.2019. – Vol. 575. – P. 25–28.
- [18] *Cooperman S.R., Brandão R.A.* AI assistance with scientific writing: Possibilities, pitfalls, and ethical considerations [Помощь искусственного интеллекта в написании научных работ: Возможности, подводные камни и этические аспекты] // Foot & Ankle Surgery: Techniques, Reports & Cases. – 2024, Spring. – Vol. 4. – № 1. – Article № 100350. – P. 1–3. – P. 1.
- [19] *Zhang J., Sun K., Jagadeesh A. et al.* The potential and pitfalls of using a large language model such as ChatGPT, GPT-4, or LLaMA as a clinical assistant Get access Arrow [Возможности и подводные камни использования большой языковой модели, такой как ChatGPT, GPT-4 или LLaMA, в качестве клинического ассистента] // Journal of the American Medical Informatics Association. – 2024. – Vol. 31. – № 9. – P. 1884–1891.
- [20] *Frueh S.* Workshop Explores the ‘Opportunity and Perils’ of Using AI in Medical Diagnosis [Воркшоп исследует «возможности и опасности» использования искусственного интеллекта в медицинской диагностике] // <<https://www.nationalacademies.org/news/2024/10/workshop-explores-the-opportunity-and-perils-of-using-ai-in-medical-diagnosis>>. – 14.10.2024.
- [21] *Mirzadeh I., Alizadeh K., Shahrokhi H., et.al.* GSM-Symbolic: Understanding the Limitations of Mathematical Reasoning in Large Language Models [GSM-Symbolic: Понимание ограничений математических рассуждений в больших языковых моделях] // <<https://arxiv.org/abs/2410.05229>>.
- [22] *Metz R., Ghaffary S., Bass D., Love J.* OpenAI, Google and Anthropic Are Struggling to Build More Advanced AI. Three of the leading artificial intelligence companies are seeing diminishing returns from their costly efforts to develop newer models [OpenAI, Google и Anthropic пытаются создать более совершенный искусственный интеллект. Три ведущие компании, занимающиеся разработкой искусственного интеллекта, отмечают снижение отдачи от своих дорогостоящих усилий по созданию новых моделей] // <<https://www.bloomberg.com/news/articles/2024-11-13/openai-google-and-anthropic-are-struggling-to-build-more-advanced-ai>>. – 13.11.2024.

Application of generative artificial intelligence in research and practical analytics for public administration: positive opportunities and pitfalls

Igor Ponkin

Abstract – The article is devoted to the study of the role and meaning of the concept and technologies of generative artificial intelligence in relation to the sphere of public administration and functioning of the system of public administration bodies. The article reveals the concept of generative artificial intelligence, including GPT models – transformers, pre-trained to generate texts in natural languages. The author technical and organisational prerequisites, conditions and aspects of the involvement of generative artificial intelligence in public administration and its provision. The article presents a formalised list of positions of involvement of generative artificial intelligence models (including GPT models) in public administration. Proceeding from the attitude that it is necessary to clearly present and understand the advantages and potential advantages (positive potential), as well as disadvantages of such software models, risks, hidden problems ('pitfalls') and negative side effects that they may entail, the author addresses these issues. The author outlines a list of positions of possible relevant involvement of generative artificial intelligence software models (including GPT models) in the work of a practitioner-analyst, scientist, expert. The author summarises and explains the problems and 'pitfalls' of software models of generative artificial intelligence (including GPT models) in their direct involvement in research, applied analytical and expert support of public administration. The article shows the problem of the current rapid exhaustibility of the potential of performance development of generative artificial intelligence models. The article was prepared within the framework of the work on the research work № 12.26.-2024-2 «Prospects and determinants of the development of digital technologies application in public administration» (Russian Presidential Academy of National Economy and Public Administration (RANEPA)).

Keywords – digitalisation of public administration, generative artificial intelligence, GPT model, legal and public administration practical analytics, science, public administration.

УДК 34:355; 34:007; 34.01; 342; 341; 004.8; 004.9; 681.5
ББК 67:68; 67:30; 67.0; 66.0; 67.4; 67.412; 67.401

Ponkin Igor, Doctor of science (Law), professor of the Institute of Public Administration and Civil Service of the Russian Presidential Academy of National Economy and Public Administration (Moscow, Russia) (IPACS, RANEPA), State professor. (e-mail: ponkin-iv@ranepa.ru).
ORCID: 0000-0003-4438-6649

REFERENCES

- [1] *Vonnegut K.* Sirens of Titan. – London: Coronet Books; Hodder & Stoughton, 1982.
- [2] *Mehta S.* How should government agencies use generative AI like ChatGPT? // <<https://www.publicceo.com/2023/05/how-should-government-agencies-use-generative-ai-like-chatgpt/>>. – 01.05.2023.
- [3] *Salvagno M., Taccone F.S., Gerli A.G.* Can artificial intelligence help for scientific writing? // *Critical Care*. – 2023. – Vol. 27. – № 1. – Article № 75. – P. 1–5.
- [4] *Dwivedi R., Elluri L.* Exploring Generative Artificial Intelligence Research: A Bibliometric Analysis Approach // *IEEE Access*. – 2024. – Vol. 12. – P. 119884–119902.
- [5] *Agentic AI – the new frontier in GenAI: An executive playbook.* – London: PricewaterhouseCoopers, 2024. – 21 p.
- [6] *Freed B.* ChatGPT said it will 'revolutionize' digital government. Human experts are more cautious // <<https://statescoop.com/chatgpt-government-digital-services/>>. – 04.04.2023.
- [7] *Buehler M.J.* Accelerating scientific discovery with generative knowledge extraction, graph-based representation, and multimodal intelligent graph reasoning // *Machine Learning: Science and Technology*. – 2024. – Vol. 5. – № 3. Article № 035083. – P. 1–58.
- [8] *Bandi A., Ramesh Adapa P.V., Kumar Kuchi Y.E.* The Power of Generative AI: A Review of Requirements, Models, Input–Output Formats, Evaluation Metrics, and Challenges // *Future Internet*. – 2023. – Vol. 15. – № 8. – Article 260. – P. 1–60.
- [9] *Saka A., Taiwo R., Saka N. et al.* GPT models in construction industry: Opportunities, limitations, and a use case validation // *Developments in the Built Environment*. – 2024, March. – Vol. 17. – Article № 100300. – P. 1–29.
- [10] *Haas C., Gilmore A.* Introducing BloombergGPT, Bloomberg's 50-billion parameter large language model, purpose-built from scratch for finance // <<https://www.bloomberg.com/company/press/bloomberg-gpt-50-billion-parameter-llm-tuned-finance/>>. – 30.03.2023.
- [11] *Generative AI Market Growth is Booming with 27.02 %* // <<https://www.precedenceresearch.com/insights/generative-ai-market>>.
- [12] *Kundu A.* The case for a ChatGPT usage and adoption policy for public sector administration // <<https://sciencepolicy.ca/posts/the-case-for-a-chatgpt->

- usage-and-adoption-policy-for-public-sector-administration/>.
- [13] *Lamovšek N.* Analysis of Research on Artificial Intelligence in Public Administration: Literature Review and Textual Analysis // *Central European Public Administration Review*. – 2023. – Vol. 21. – № 2. – P. 77–93. – P. 92–93.
- [14] *Salah M., Abdelfattaha F., Al Halbusib H.* Generative Artificial Intelligence (ChatGPT & Bard) in Public Administration Research: A Double-Edged Sword for Street-Level Bureaucracy Studies // *International Journal of Public Administration*. – 04.11.2023. – P. 1–7.
- [15] *Onatayo D., Onososen A., Oyediran A.O. et al.* Generative AI Applications in Architecture, Engineering, and Construction: Trends, Implications for Practice, Education & Imperatives for Upskilling – A Review // *Architecture*. – 2024. – Vol. 4. – P. 877–902.
- [16] *Kammer M.N., Gomila P., Vambaco D.J., Maldonado F.* The role of artificial intelligence in scientific writing // *Journal of Clinical Case Reports, Medical Images and Health Sciences*. – 2023. – Vol. 3. – № 3. – P. 1–6.
- [17] *Belknap G.* 150 years of scientific illustration // *Nature*. – 07.11.2019. – Vol. 575. – P. 25–28.
- [18] *Cooperman S.R., Brandão R.A.* AI assistance with scientific writing: Possibilities, pitfalls, and ethical considerations // *Foot & Ankle Surgery: Techniques, Reports & Cases*. – 2024, Spring. – Vol. 4. – № 1. – Article № 100350. – P. 1–3. – P. 1.
- [19] *Zhang J., Sun K., Jagadeesh A. et al.* The potential and pitfalls of using a large language model such as ChatGPT, GPT-4, or LLaMA as a clinical assistant Get access Arrow // *Journal of the American Medical Informatics Association*. – 2024. – Vol. 31. – № 9. – P. 1884–1891.
- [20] *Frueh S.* Workshop Explores the ‘Opportunity and Perils’ of Using AI in Medical Diagnosis // <<https://www.nationalacademies.org/news/2024/10/workshop-explores-the-opportunity-and-perils-of-using-ai-in-medical-diagnosis>>. – 14.10.2024.
- [21] *Mirzadeh I., Alizadeh K., Shahrokhi H., et.al.* GSM-Symbolic: Understanding the Limitations of Mathematical Reasoning in Large Language Models // <<https://arxiv.org/abs/2410.05229>>.
- [22] *Metz R., Ghaffary S., Bass D., Love J.* OpenAI, Google and Anthropic Are Struggling to Build More Advanced AI. Three of the leading artificial intelligence companies are seeing diminishing returns from their costly efforts to develop newer models // <<https://www.bloomberg.com/news/articles/2024-11-13/openai-google-and-anthropic-are-struggling-to-build-more-advanced-ai>>. – 13.11.2024.