

УДК [61:004.9](048)

<https://doi.org/10.47093/2713-069X.2023.4.4.48-55>

Перспективы применения больших языковых моделей в здравоохранении

А.Е. Андрейченко¹, А.В. Гусев^{2,3,*}

¹ООО «К-Скай», набережная Варкауса, д. 17, г. Петрозаводск, 185031, Россия

²ФГБУ «Центральный научно-исследовательский институт организации и информатизации здравоохранения» Министерства здравоохранения Российской Федерации, ул. Добролюбова, д. 11, г. Москва, 127254, Россия

³ГБУЗ г. Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения Москвы», ул. Петровка, д. 24, г. Москва, 127051, Россия

Аннотация

Большие языковые модели стали новым значимым прорывом в области искусственного интеллекта, меняющим подходы от применения моделей машинного обучения в узких задачах, требующих больших объемов данных с готовыми ответами для обучения, к генеративным моделям, способным к дообучению на небольшом количестве примеров или вообще без примеров с готовыми ответами и при этом имеющих более широкие возможности применения. Медицина является одной из областей, в которой внедрение больших языковых моделей может стать крайне востребованным. В обзоре представлены данные о последних достижениях применения больших языковых моделей для медицинских задач, перспективы использования этих моделей как основы цифровых ассистентов для врачей и пациентов, а также существующие регуляторные и этические барьеры на пути развития этой прорывной технологии для решения задач здравоохранения.

Ключевые слова: большие языковые модели; генеративный искусственный интеллект; здравоохранение; машинное обучение; медицинский чат-бот

Для цитирования: Андрейченко А.Е., Гусев А.В. Перспективы применения больших языковых моделей в здравоохранении. Национальное здравоохранение. 2023; 4 (4): 48–55. <https://doi.org/10.47093/2713-069X.2023.4.4.48-55>

Контактная информация:

* Автор, ответственный за переписку: Гусев Александр Владимирович. E-mail: agusev@webiomed.ai

Статья поступила в редакцию: 22.09.23

Статья принята к печати: 01.11.23

Дата публикации: 20.12.23

Perspectives on the application of large language models in healthcare

Anna E. Andreychenko¹, Aleksandr V. Gusev^{2,3,*}

¹K-Skai LLC, Varkaus Embankment, 17, Petrozavodsk, 185031, Russia

²Federal Research Institute for Health Organization and Informatics, Dobrolyubova str., 11, Moscow, 127254, Russia

³Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies, Petrovka str., 24/1, Moscow, 127051, Russia

Abstract

Large language models have become a new significant breakthrough in the field of artificial intelligence. They are changing approaches to machine learning from models that solve narrow problems and require large amounts of data with the known answers for training, to generative models that are fine tunable to solve specific problems using a small number of examples with the known answers or even none at all. Medicine is one of the areas in which the use of large language models can become extremely useful. The review presents data on the latest achievements in the use of large language models for medical tasks, prospects for using these models as the basis for the digital assistants for doctors and patients, as well as existing regulatory and ethical barriers to the development of this breakthrough technology for addressing healthcare challenges.

Keywords: large language models; generative artificial intelligence; healthcare; machine learning; medical chatbot

For citation: Andreychenko A.E., Gusev A.V. Perspectives on the application of large language models in healthcare. National Health Care (Russia). 2023; 4 (4): 48–55. <https://doi.org/10.47093/2713-069X.2023.4.4.48-55>

Contacts:

* Corresponding author: Aleksandr V. Gusev. E-mail: agusev@webiomed.ai

The article received: 22.09.23

The article approved for publication: 01.11.23

Date of publication: 20.12.23

Список сокращений:

LLM – Large Language Models, большие языковые модели

ИИ – искусственный интеллект

СППВР – система поддержки принятия врачебных решений

Технология больших языковых моделей (Large Language Models, LLM) стала в последнее время одним из самых перспективных направлений для исследований и разработок в сфере искусственного интеллекта (ИИ) для обработки естественного языка (Natural Language Processing, NLP) [1–3] и последующего решения широкого спектра прикладных задач.

Как правило, для создания LLM используются архитектура искусственной нейронной сети (artificial neural network, ANN) и глубокое обучение (deep learning, DL), которое осуществляется на очень больших корпусах текстов, содержащих миллиарды слов – например статей, книг и других публикаций в сети Интернет. Это позволяет получить модель машинного обучения, содержащую миллиарды параметров, которые формируют довольно подробное представление сложных ассоциативных отношений

между словами в текстах (рис. 1). Например, 2-я версия большой языковой модели GPT (Generative Pretrained Transformer), выпущенная в 2019 г., имела 1,5 млрд параметров, а GPT-3.5, выпущенная осенью 2022 г. – уже 175 млрд параметров [2].

Применение LLM позволяет существенно повысить точность решения самых разнообразных задач обработки естественного языка, включая интерпретацию и классификацию, обобщение текстовой информации, создание чат-ботов и иных диалоговых систем, генерацию текстов по запросам и т.д. Более того, применение предварительно обученных LLM позволяет создать прикладные модели для решения узких задач с минимальным набором для дообучения ('few-shot') или даже вовсе без обучающего набора ('zero-shot'), что позволяет существенно ускорить и одновременно удешевить создание прикладных ИИ-систем [4].

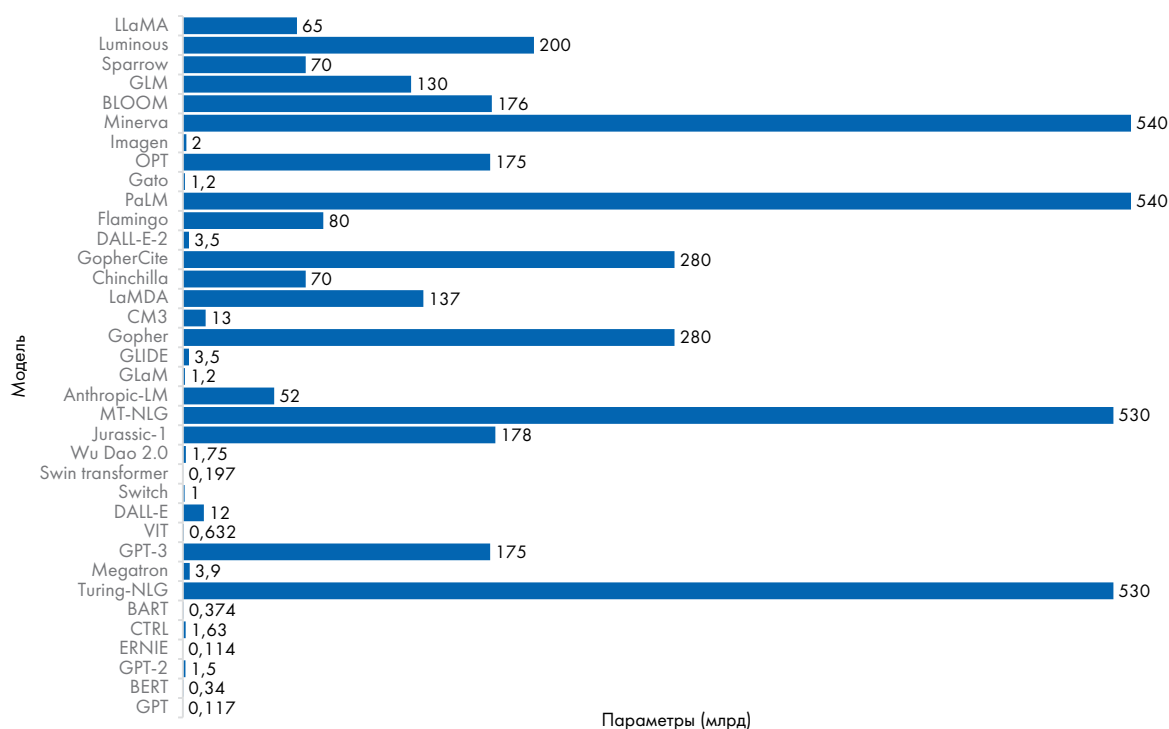


Рис. 1. Различные большие языковые модели с указанием количества параметров и их разработчиков, Thirunavukarasu, 2023 [2]
Fig. 1. Various large language models with number of parameters and their developers, Thirunavukarasu, 2023 [2]

По мере развития методов глубокого обучения, мощных вычислительных ресурсов и больших наборов данных для обучения начали появляться множество LLM, которые могут выполнять когнитивную работу на довольно высоком уровне, достигающем возможностей человека.

Одним из самых известных в мире примеров успешного применения LLM является чат-бот ChatGPT компании OpenAI, который может принимать на вход любой произвольный запрос и давать на него ответ, поразительно точно имитирующий ответ человека. Технологической основой ChatGPT являются большие языковые модели GPT-3.5 и GPT-4. Успех и глобальное влияние ChatGPT на рост спроса на решения в области генеративного ИИ проистекают из его доступности, универсальности, диалоговой интерактивности и производительности, близкой или равной человеческому уровню.

Необходимо обратить внимание, что применение LLM было лишь одним из этапов создания ChatGPT, хотя и имеющим фундаментальное значение. После того как LLM была получена, потребовалась дальнейшая доработка и тонкая настройка, а также постоянное дообучение во время эксплуатации (рис. 2) [2].

Исходная модель GPT-3 была предобучена на датасете с готовыми запросами и ответами до GPT-3.5. Далее было проведено «обучение с подкреплением на основе обратной связи от человека» (Reinforcement Learning from Human Feedback, RLHF) с использованием модели вознаграждения, созданной на данных, сгенерированных людьми, которым было поручено ранжировать ответы GPT-3.5 на набор запросов. Этот подход позволил реализовать LLM в гораздо больших

масштабах, чем это можно было бы сделать при ручном оценивании человеком каждого отдельного ответа модели. Для повышения надежности и безопасности было проведено дальнейшее автоматизированное состязательное обучение с использованием генерируемых моделью входных запросов и выходных ответов. При эксплуатации модель GPT-3.5 продолжила непрерывное самообучение, используя обратную связь от быстро растущей базы пользователей.

В России также проводятся аналогичные исследования и разработки в области LLM. В частности, модель ruGPT-3.5, насчитывающую 13 млрд параметров, создали в Сбере. Аналогичную большую языковую модель YandexGPT-2 (100 млрд параметров) развивает Яндекс. Эти модели являются подходящей базой для независимого развития технологий LLM в Российской Федерации.

Применение технологии LLM в здравоохранении

ChatGPT привлек к себе особое внимание в медицине благодаря тому, что получил проходной балл на экзаменах по лицензированию медицинской деятельности в США, а производительность GPT-4 стала заметно выше, чем у его предшественника GPT-3.5 [5]. Схожий результат был недавно продемонстрирован и для экзамена медсестер в Японии [6], а также GPT-4 превзошел результаты студентов-медиков на государственном экзамене в Германии [7].

Успех ChatGPT породил целую волну аналогичных разработок для задач здравоохранения у других компаний, например Google запустила проект создания медицинского чат-бота Med-PaLM 2, который

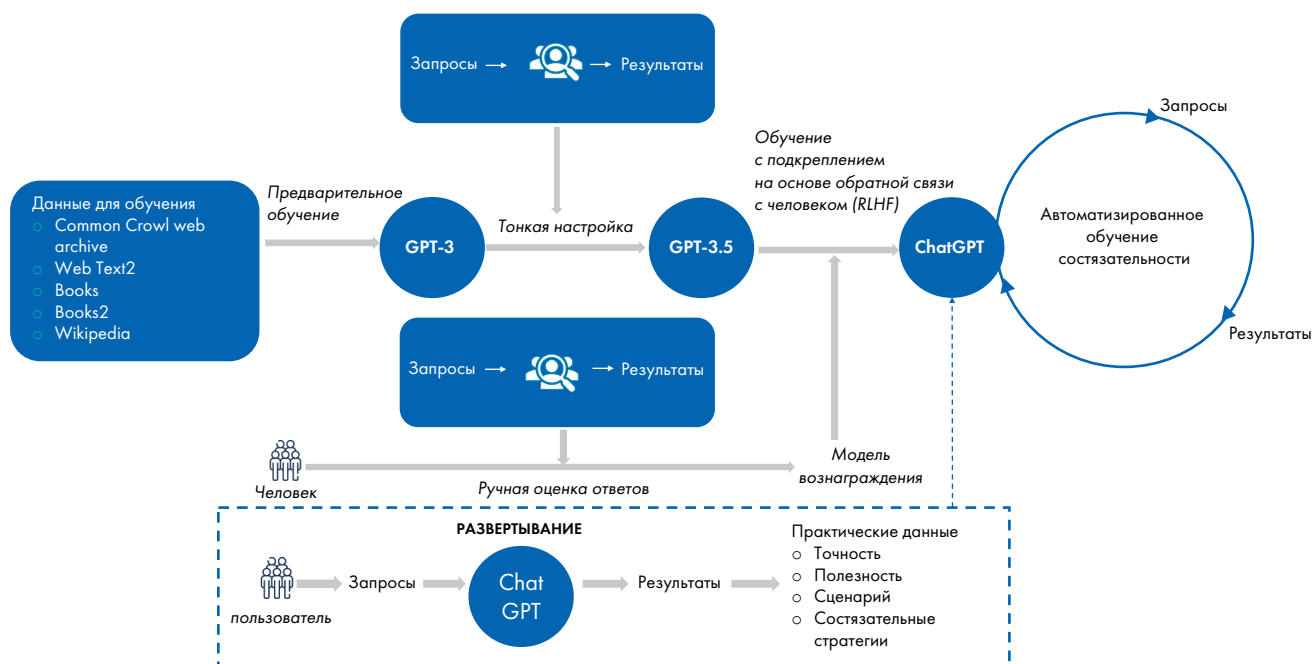


Рис. 2. Алгоритм обучения больших языковых моделей для решения прикладных задач, Thirunavukarasu, 2023 [2]

Fig. 2. Algorithm for training large language models for solving applied problems, Thirunavukarasu, 2023 [2]

способен на уровне врача-эксперта отвечать на медицинские вопросы [8]. В настоящее время решение проходит тестирование в исследовательском госпитале Mayo Clinic [1].

Когда ответы ChatGPT на запросы пациентов сравниваются с ответами врачей (отвечающих в свободное время в социальной сети), результат LLM оказывается предпочтительнее с точки зрения качества и эмпатии, если использовать ее в качестве качественной метрики на основе оценок врачами-судьями. Это дало повод говорить о том, что ИИ готов заменить врачей, но в реальности это не так. Даже на экзаменах для студентов-медиков результаты модели далеки от идеальных. Было показано, что ChatGPT не справляется с экзаменами для врачей-специалистов и дает неточную информацию в ответ на реальные вопросы пациентов о профилактике сердечно-сосудистых заболеваний [9], лечении онкологических заболеваний [10], а Bing Chatbot давал неправильные советы первой помощи при неотложных состояниях [11]. Несмотря на способность интерпретировать клинические заметки и отвечать на соответствующие вопросы, LLM часто не могут предоставить информацию в соответствии с индивидуальными особенностями пациента. Текущие результаты пока исключают возможность автономного развертывания системы для принятия клинических решений или общения с пациентами, тем более что пациенты часто не могут отличить информацию, предоставленную LLM, от информации, полученной от человека-клинициста. В связи с этим Всемирная организация здравоохранения призывает к осторожному и ответственному развитию и применению столь популярных сейчас технологий больших языковых моделей в медицине¹.

В Российской Федерации также активно запускаются проекты по развитию LLM в медицинских целях. Например, Сеченовский университет совместно со Сбером набирает ИИ-тренеров для обучения мультимодальной нейронной сети для медицины и здравоохранения "GigaChat"². В результате такой инициативы могут быть созданы русскоязычные базы для донастройки (fine-tuning) больших языковых моделей под медицинские задачи. Преимуществом таких баз является учет национальных подходов и особенностей в медицине и здравоохранении. Однако открытыми остаются вопросы безопасности LLM, обеспечения отсутствия «галлюцинаций» при решении медицинских задач, а также своевременной и экономичной актуализации «знаний» модели. Например, разработчики Яндекс предупреждают, что информация и данные, которыми владеет модель YaGPT 2.0, датируются не позднее марта 2023. При необходимости база знаний, на основе которой LLM генерирует ответ, может

быть ограничена с помощью технологии retrieval-augmented generation. При этом дорогостоящего предобучения базовой LLM модели не требуется. Это также актуально для задач медицины и здравоохранения, так как база знаний должна постоянно пополняться новым клиническими случаями, рекомендациями и т.п.

Перспективы применения LLM для создания цифровых ассистентов для врачей

В настоящее время применение LLM при создании специализированных ИИ-продуктов для клинического применения и поддержки принятия врачебных решений выглядит более обоснованным подходом, чем создание универсального ИИ, способного отвечать на любые вопросы и темы и уж тем более – заменить врача [1]. Более того, применение LLM демонстрирует гораздо лучшую эффективность в задачах, где не требуется специальных знаний, или заданиях, которые предоставляются в виде развернутых пользовательских запросов (prompt). Это открывает довольно большие возможности для ускорения внедрения ИИ в медицинскую практику, бросая, по сути, вызов традиционным подходам, применяемым при создании систем поддержки принятия врачебных решений (СППВР).

Так, модель Foresight, созданная на основе архитектуры LLM и неструктурированных текстовых медицинских записей, извлеченных из 811 тыс. электронных медицинских карт, продемонстрировала высокую эффективность в прогнозировании и предсказании в валидационных исследованиях [12]. Данный проект наглядно показал, что модели общего риска могут стать мощной альтернативой многочисленным инструментам, используемым в настоящее время для стратификации и сортировки пациентов.

Обобщая исследования и разработки в сфере LLM для клинического применения, можно выделить следующие наиболее перспективные направления [1–3]:

- формирование медицинской документации;
- формирование выписных эпикризов;
- генерация клинических заметок;
- предварительное подтверждение страховых случаев;
- краткое содержание научно-исследовательских работ;
- интерпретация диагностических исследований;
- предложение способов лечения;
- разработка планов лечения;
- диагностическая помощь;
- медицинский триаж.

Появляющиеся LLM будут расширять свои возможности и совместимость с разными типами источников данных; даже почерк врача может быть

¹ ВОЗ. ВОЗ призывает к безопасному и этичному использованию ИИ в интересах здоровья. 16 мая 2023 г. URL: <https://www.who.int/news/item/16-05-2023-who-calls-for-safe-and-ethical-ai-for-health> (дата обращения: 05.07.2023).

² <https://student.sechenov.ru/info/39305115> (дата обращения: 05.07.2023).

интерпретирован автоматически и точно. Компании Microsoft и Google стремятся интегрировать ChatGPT и PaLM 2, соответственно, в административный рабочий процесс, позволяя беспрепятственно и автоматически интегрировать информацию из видеозвонков, документов, электронных таблиц, презентаций и электронной почты.

Вместе с этим следует обратить особое внимание, что применение LLM в клинических условиях, когда безопасность пациента не гарантируется безусловно, требует всесторонней валидации. Для обеспечения безопасности пациентов и административной эффективности необходима всесторонняя оценка качества, а для распределения ответственности требуются специальные структуры управления [1].

Перспективы применения LLM для создания цифровых ассистентов для пациентов

ИИ-решения на основе LLM способны очень быстро анализировать, обобщать и перефразировать любую информацию на естественном языке, в том числе собранную со слов пациента. Это открывает действительно впечатляющие перспективы для ускорения и повышения эффективности проектов цифровой трансформации здравоохранения, направленных на создание новых инновационных сервисов для пациентов.

Различные чат-боты, интегрированные в мобильные приложения, сайты клиник или органов управления здравоохранения, страховых компаний и т.д. позволяют существенно сократить нагрузку на call-центры, регистратуры и даже первичные приемы, в конечном счете даже отказаться от участия человека в процессе первичного общения с пациентом. В довольно большом количестве случаев, особенно для наиболее распространенных и неопасных (поддающихся лечению в домашних условиях) заболеваний и состояний, чат-боты на основе LLM действительно способны полностью заменить (исключить) визит пациента на очный прием путем автоматического сбора, диалога и интерпретации информации с подбором рекомендаций по сохранению здоровья.

Наиболее перспективными задачами цифровых ассистентов для пациентов, которые могут быть решены с помощью LLM, являются [1–3]:

- анализ лабораторных результатов;
- описание заболеваний;
- интерпретация заметок врача;
- персонализированные рекомендации по поводу здоровья;
- прогнозирование состояния здоровья;
- оценка симптомов;
- анализ данных с носимых устройств;
- чат-боты по психическому здоровью;
- соблюдение плана лечения;
- гиды по реабилитации.

Перспективы применения LLM в клинических исследованиях

Ускорение и сокращение стоимости клинических исследований является одной из самых перспективных ниш для применения технологий LLM, которым можно поручить краткое изложение информации, формирование описания предоставленных результатов или генерацию фрагментов текстов для конкретного читателя или аудитории. Модели, прошедшие тонкую настройку на основе информации, специфичной для конкретной области, могут демонстрировать более высокую производительность, чему уже есть конкретные примеры, включая модели PubMedBERT и BioBERT. Это может снизить бремя критической оценки, написания отчетов об исследованиях и рецензирования, что составляет значительную часть рабочей нагрузки исследователей. Вопросы, связанные с ответственностью за корректность информации и выводов, будут решены за счет того, что клиницисты и исследователи, использующие эти инструменты, будут нести полную ответственность за их результаты. Однако широкое применение LLM в исследованиях будет возможно только после того, как разработчиками будет обеспечено отсутствие «галлюцинаций» и несуществующих источников в результатах работы моделей [13].

LLM могут способствовать проведению новых исследований, таких как анализ языка в больших масштабах, чем это было возможно ранее. В качестве наглядных примеров можно привести ClinicalBERT, GPT-3.5 и GatorTron, которые позволяют исследователям эффективно анализировать большие объемы клинических текстовых данных. LLM могут также стимулировать исследования в менее очевидных смежных областях, поскольку текстовая информация включает в себя не только человеческий язык. Например, генетические данные и данные о структуре белков обычно представлены в текстовой форме и поддаются обработке на естественном языке, чему способствуют LLM. Модели уже дают впечатляющие результаты: AlphaFold выводит структуру белка из аминокислотных последовательностей; ProGen генерирует белковые последовательности с предсказуемой биологической функцией; TSSNote-CyaPromBERT идентифицирует промоторные области в бактериальной ДНК.

Другие потенциальные возможности использования LLM включают контрфактическое моделирование и виртуальные клинические испытания. Такое направление может ускорить клинические исследования за счет получения ценных выводов о соотношении риска и пользы с целью информирования ученых и врачей о том, какие исследования с наибольшей вероятностью принесут пользу пациентам.

Новые архитектуры, такие как Hybrid Value-Aware Transformer (HVAT), могут еще больше повысить производительность LLM за счет интеграции продольных, мультимодальных клинических данных.

Наконец, генеративные приложения на основе LLM, применяемые для анализа данных пациентов, могут быть использованы для получения синтетических данных; при соответствующей оценке качества это может способствовать развитию клинических исследований, увеличивая масштаб обучающих выборок, доступных для разработки LLM и других инструментов ИИ.

Перспективы применения LLM в медицинском образовании

Высокие результаты GPT-4 и Med-PaLM 2 в медицинских тестированиях позволяют предположить, что LLM могут стать полезным инструментом обучения для студентов, которые в настоящее время показывают более низкие результаты в таких тестах. Функция мета-запроса в GPT-4 позволяет пользователям явно описать желаемую роль чат-бота в разговоре; полезным примером может служить «режим сократовского наставника», который побуждает студентов думать самостоятельно, задавая вопросы с понижающимся уровнем сложности, пока студенты не смогут найти решение более полного вопроса. Журналы бесед позволяют преподавателям отслеживать прогресс и корректировать обучение с учетом слабых сторон учащихся [14].

Таким образом, применение LLM является одним из перспективных подходов к повышению эффективности образования и последилового образования, в частности – при изучении или обобщении нового материала, для повышения мотивации и вовлеченности студентов в учебный процесс, как метод более углубленного получения знаний.

Нормативное и техническое регулирование LLM

Учитывая потенциальные последствия для результатов лечения пациентов и общественного здравоохранения, необходимо рассмотреть вопрос о том, как следует регулировать эти новые инструменты на основе ИИ и возникающие в связи с этим проблемы (табл.). Регулирование LLM в медицине и здравоохранении без ущерба для их многообещающего прогресса является своевременной и важной задачей для обеспечения безопасности, соблюдения этических стандартов, предотвращения несправедливости и предвзятости и защиты конфиденциальности пациентов. Какие бы опасения ни вызывал ИИ, теперь они заметно усиливаются благодаря огромному и разнообразному потенциалу приложений на основе LLM.

Большинство LLM глобальны, у них нет версий для конкретных стран, а значит, они требуют особого подхода со стороны регулирующих органов. Также неясно, к какой технической категории попадут LLM с точки зрения регулирования. Однако из-за различий между LLM и предыдущими методами глубокого обучения может потребоваться новая нормативная

категория для решения проблем и рисков, связанных с LLM.

Регулирующий орган должен разрабатывать правила для LLM только в том случае, если разработчики LLM заявляют, что их модель предназначена для использования в медицинских целях. В эту категорию также целесообразно отнести медицинские альтернативы LLM общего назначения, которые были специально дообучены на медицинских базах данных.

ЗАКЛЮЧЕНИЕ

LLM произвели революцию в обработке естественного языка, и современные модели, такие как GPT-4 и PaLM 2, занимают сегодня центральное место в инновациях ИИ в медицине. Аналогичные исследования и разработки ведутся в России, преимущественно крупнейшими технологическими компаниями, такими как Сбер и Яндекс. Новые технологии открывают широкие возможности для применения в клинической, образовательной и исследовательской деятельности, особенно в связи с развитием мультимодальности и интеграции в уже существующие цифровые инструменты в сфере здравоохранения. Однако потенциальные риски вызывают серьезную озабоченность в отношении безопасности, этики и потенциальной замены человека в определенных ситуациях. Применяя упреждающий подход к регулированию, можно использовать потенциал технологий на основе ИИ, таких как LLM, сводя к минимуму потенциальный вред и сохраняя доверие как пациентов, так и медицинских работников и организаторов здравоохранения.

Мы можем ожидать от регулирующих органов в отношении внедрения LLM в медицинскую практику следующего:

- запуск и стимулирование пилотных проектов разработок и внедрения LLM с целью как можно более раннего и всестороннего анализа практического опыта применения данной технологии, включая оценку преимуществ и потенциальных опасностей;
- разработку отдельного технического и, возможно в будущем, нормативного регулирования для продуктов на основе LLM, в том числе с точки зрения их применения в лечебно-диагностических процессах;
- разработку нормативных рекомендаций для компаний и организаций здравоохранения о том, как они могут внедрять LLM в свои существующие продукты и услуги на основе риск-ориентированного подхода и ответственного отношения к ИИ;
- выработку регуляторных подходов и рекомендаций для проведения различия между LLM, специально обученными на медицинских данных, и LLM, обученными для немедицинских целей;
- отнесение медицинских цифровых приложений для пациентов, использующих технологии LLM и обеспечивающих анализ медицинских данных и рекомендации по здоровью к категории программных медицинских изделий.

Таблица. Список регуляторных проблем, связанных с развитием LLM

Table. List of regulatory issues associated with LLM development

Задача регулирования	Краткое описание
Обеспечение национального суверенитета	Согласно национальной стратегии развития ИИ ³ при внедрении любых ИИ-систем должен соблюдаться принцип технологического суверенитета, включая преимущественное использование отечественных разработок. Это означает, что следует рассмотреть возможность ограничения LLM, созданных иностранными компаниями, или даже запрет на их применение в здравоохранении.
Доступность мощных вычислительных ресурсов для создания LLM	Известно, что для создания LLM требуются очень мощные вычислительные ресурсы, которые в настоящее время в России доступны лишь ограниченному списку крупных технологических компаний, таких как Сбер и Яндекс. Эта особенность существенно ограничивает возможности исследований и разработок в сфере LLM в небольших компаниях, стартапах и научных коллективах. Целесообразно рассмотреть возможность выделения финансовой поддержки для таких коллективов, чтобы они могли покупать необходимые серверные мощности или через API обращаться к имеющимся LLM и тем самым участвовать в исследованиях применения технологий LLM для решения задач здравоохранения. Кроме этого, стоит рассмотреть вопрос о создании открытых и общедоступных LLM, созданных для русского языка.
Конфиденциальность данных пациентов	Обеспечение того, чтобы данные пациентов, используемые для обучения больших языковых моделей, были полностью обезличенными и защищенными от возможных взломов. Это создает большую нормативную проблему, поскольку любое нарушение может привести к серьезным последствиям в соответствии с законом о защите персональных данных. Для обеспечения конфиденциальности и безопасности данных должны быть использованы технологии шифрования данных при хранении и передаче, защищенные сервера. При сборе и объединении мультимодальных данных могут быть использованы технологии дифференциальной приватности, которые добавляют шум к агрегированным данным для предотвращения идентификации отдельного пациента, при этом ценность объединенных данных сохраняется, а также протоколы конфиденциального вычисления (Secure multiparty computation SMC), обеспечивающие сохранность персональных данных при объединении данных из нескольких источников. Кроме этого, активно развиваются технологии федеративного обучения, когда не требуется централизованный сбор набора данных для обучения моделей. Соответственно, исходя из степени разнообразия данных, уровня решаемой с помощью LLM задачи в медицине и здравоохранении, могут быть применены те или иные технологии для обеспечения конфиденциальности и безопасности медицинских данных.
Интеллектуальная собственность	Если LLM создает контент, аналогичный проприетарным медицинским исследованиям или литературе, это может привести к проблемам с правами на интеллектуальную собственность.
Использование LLM в системах поддержки принятия врачебных решений	При использовании технологий ИИ в СППВР вероятность потенциального причинения вреда имеет максимальный высокий уровень риска. Применение LLM может привести к еще большему риску, который может превышать потенциальный положительный эффект, в силу особенностей создания LLM и наличия риска «галлюцинаций». Целесообразно рассмотреть вопрос о допустимости применения LLM в СППВР, включая анализ данных электронных медицинских карт. Возможно, на период становления отечественных исследований и разработок в этой сфере будет целесообразно ограничить применение LLM в некоторых наиболее рискованных сценариях использования СППВР, например в госпитальных условиях, отделениях реанимации, при назначении лекарственного лечения или медицинских вмешательств и т.д. Альтернативным решением для минимизации риска «галлюцинаций» может быть создание отраслевой LLM, изначально предобученной исключительно на валидированных медицинских и биомедицинских текстах и данных. Однако сбор, оцифровка и чистка всех медицинских знаний является трудоемкой и долгосрочной задачей.
Контроль качества и стандартизация	Регулирование необходимо для обеспечения надежности и сохранения качества ИИ-систем на основе LLM, т.к. они обучаются на неконтролируемо большом и не всегда валидированном контенте.
Информированное согласие	Пациенты должны быть проинформированы и давать согласие при использовании цифровых продуктов на основе LLM. Это сложно, потому что пациентам может быть трудно полностью понять все последствия использования ИИ.
Интерпретируемость и прозрачность	Правила должны обеспечивать прозрачность того, как ИИ принимает решения. Это особенно сложно с LLM, которые являются «черными ящиками» из-за их сложных алгоритмов, большого количества параметров и огромных объемов обучающих наборов данных.
Справедливость и предвзятость	Регулирование необходимо для предотвращения предвзятости в LLM, которые могут быть применены в лечебно-диагностических процессах с использованием данных пациентов. Это может привести к различиям в результатах лечения.
Владение данными	Может быть сложно определить и отрегулировать, кому принадлежат данные, на которых учатся LLM, особенно когда речь идет о данных пациентов.
Чрезмерная зависимость от моделей ИИ	Чрезмерная зависимость от LLM, например в задачах обобщения данных медицинской литературы или использовании диалоговых систем для поиска медицинской информации, может привести к снижению качества экспертизы у специалистов и потенциальным ошибкам, если LLM даст свой или предоставит неверную информацию (т.н. «галлюцинации» – характерная проблема современных LLM). Существуют подходы к устранению «галлюцинаций» с помощью retrieval-augmented generation (RAG), когда «база знаний» модели для генерации ответа ограничена. Однако у такого подхода есть технические ограничения по размеру документов в базе знаний и тем самым ограничения по полноте и точности ответа модели. Кроме этого, недавнее исследование показало, что LLM могут быть использованы для генерации несуществующих, но правдоподобных данных клинических исследований для подтверждения той или иной гипотезы. Необходимы правила, чтобы сбалансировать использование LLM и человеческий опыт и знания.
Непрерывный мониторинг и проверка	Обеспечение непрерывной безопасности, качества и достоверности ИИ-систем на основе LLM с течением времени и для разных групп населения является важной нормативно-правовой задачей.

³ Указ Президента РФ от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации» (вместе с «Национальной стратегией развития искусственного интеллекта на период до 2030 года»). http://www.consultant.ru/document/cons_doc_LAW_335184/ (дата обращения: 05.07.2023).

При условии решения этических, технических и регуляторных вопросов проверенные приложения на основе LLM могут стать ценным инструментом в проектах цифровой трансформации здравоохранения для улучшения медицинского обслуживания пациентов и практикующих врачей. Проверка приложений на основе LLM предполагает проведение независимых клинических испытаний, оценивающих реальные преимущества при минимизации предвзятости и прозрачности отчетности.

ВКЛАД АВТОРОВ

А.Е. Андрейченко – обзор литературы, описание существующих разработок, анализ и описание подходов к созданию и применению технологий LLM.

А.В. Гусев – идея исследования, направления для обзора литературы, обзор нормативных проблем и выводы по исследованию. Все авторы утвердили окончательную версию статьи.

ЛИТЕРАТУРА/REFERENCES

1. Yang R., Tan T.F., Lu W., et al. Large language models in health care: development, applications, and challenges. *Health Care Science*. 2023; 2(4): 255–263. <https://doi.org/10.1002/hcs2.61>
2. Thirunavukarasu A.J., Ting D.S.J., Elangovan K., et al. Large language models in medicine. *Nat Med*. 2023; 29: 1930–1940. <https://doi.org/10.1038/s41591-023-02448-8>
3. Meskó B., Topol E.J. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *npj Digit. Med*. 2023; 6(1): 120. <https://doi.org/10.1038/s41746-023-00873-0>
4. Language models and linguistic theories beyond words. *Nat Mach Intell*. 2023; 5: 677–678. <https://doi.org/10.1038/s42256-023-00703-8>
5. Kung T.H., Cheatham M., Medenilla A., et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digital Health*. 2023; 2(2): e0000198. <https://doi.org/10.1371/journal.pdig.0000198>
6. Kaneda Y., Takahashi R., Kaneda U., et al. Assessing the Performance of GPT-3.5 and GPT-4 on the 2023 Japanese Nursing Examination. *Cureus*. 2023; 15(8): e42924. <https://doi.org/doi:10.7759/cureus.42924>
7. Roos J., Kasapovic A., Jansen T., Kaczmarczyk R. Artificial Intelligence in Medical Education: Comparative Analysis of ChatGPT, Bing, and Medical Students in Germany. *JMIR Med Educ*. 2023; 9: e46482. <https://doi.org/10.2196/46482>
8. Singhal K., Aziz S., Tu T., et al. Large language models encode clinical knowledge. *Nature*. 2023; 620: 172–180. <https://doi.org/10.1038/s41586-023-06291-2>

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Conflict of interests. The authors declare that there is no conflict of interests.

Финансирование. Исследование не имело спонсорской поддержки (собственные ресурсы).

Financial support. The study was not sponsored (own resources).

AUTHOR CONTRIBUTIONS

Anna E. Andreychenko – literature review, description of existing developments, analysis and description of approaches to the creation and application of LLM technologies.

Aleksandr V. Gusev – research idea, history of market development, analysis of regulatory development.

All the authors approved the final version of the article.

9. Sarraju A., Bruemmer D., Van Iterson E., et al. Appropriateness of Cardiovascular Disease Prevention Recommendations Obtained From a Popular Online Chat-Based Artificial Intelligence Model. *JAMA*. 2023; 329(10): 842–844. <https://doi.org/10.1001/jama.2023.1044>
10. Chen S., Kann B.H., Foote M.B., et al. Use of Artificial Intelligence Chatbots for Cancer Treatment Information. *JAMA Oncol*. Published online. 2023. <https://doi.org/10.1001/jamaoncol.2023.2954>
11. Birkun A., Gautam A. Large language model-based chatbot as a source of advice on first aid in heart attack. *Current Problems in Cardiology*. 2023; 1(49): 102048. <https://doi.org/10.1016/j.cpcardiol.2023.102048>
12. Jiang L.Y., Liu X.C., Nejatian N.P., et al. Health system-scale language models are all-purpose prediction engines. *Nature*. 2023; 619: 357–362. <https://doi.org/10.1038/s41586-023-06160-y>
13. Moskatel L.S., Niushen Z. The utility of ChatGPT in the assessment of literature on the prevention of migraine: an observational, qualitative study. *Frontiers in Neurology*, 2023, 14: 1225223. <https://doi.org/10.3389/fneur.2023.1225223>
14. Lower K., Seth I., Lim B., Seth N. ChatGPT-4: Transforming Medical Education and Addressing Clinical Exposure Challenges in the Post-pandemic Era. *Indian J Orthop*. 2023; 57: 1527–1544. <https://doi.org/10.1007/s43465-023-00967-7>

Информация об авторах

Андрейченко Анна Евгеньевна – канд. физ.-мат. наук, руководитель направления искусственного интеллекта ООО «К-Скай».

ORCID: <https://orcid.org/0000-0001-6359-0763>

Гусев Александр Владимирович – канд. техн. наук, старший научный сотрудник отдела научных основ организации здравоохранения ФГБУ «Центральный научно-исследовательский институт организации и информатизации здравоохранения» Минздрава России; старший научный сотрудник ГБУЗ г. Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы».

ORCID: <https://orcid.org/0000-0002-7380-8460>

Information about the authors

Anna E. Andreychenko – Cand. of Sci. (Physics and Mathematics), Head of Artificial Intelligence, K-Skai LLC.

ORCID: <https://orcid.org/0000-0001-6359-0763>

Aleksandr V. Gusev – Cand. of Sci. (Technology), Senior Researcher, Department of Scientific Foundations of Healthcare Organization, Federal Research Institute for Health Organization and Informatics; Senior Researcher, Scientific and Practical Clinical Center for Diagnostics and Telemedicine Technologies.

ORCID: <https://orcid.org/0000-0002-7380-8460>