

## Обзор применения больших языковых моделей в информационных задачах систем поддержки принятия решения на примере здравоохранения

М.А. Евлахова<sup>1</sup>, А.В. Зубков<sup>1,2</sup>

<sup>1</sup>Волгоградский государственный технический университет, Волгоград

<sup>2</sup>Волгоградский государственный медицинский университет, Волгоград

**Аннотация:** Статья посвящена применению больших языковых моделей (БЯМ) в информационных задачах систем поддержки принятия решений на примере здравоохранения. Рассматриваются ключевые архитектуры БЯМ и их практические реализации, а также возможности этих моделей для обработки естественного языка и анализа медицинских данных. Особое внимание уделяется роли БЯМ в автоматизации процессов принятия решений, включая оптимизацию доступа к знаниям из клинических рекомендаций. Представлены примеры использования БЯМ в различных областях медицины. Кроме того, обсуждаются перспективы дальнейшего развития БЯМ в здравоохранении и связанные с этим вызовы.

**Ключевые слова:** большие языковые модели, обработка естественного языка, системы поддержки принятия решений, промпт-инжиниринг, клинические рекомендации, международная классификация болезней.

### Введение

В последние годы наблюдается стремительное развитие технологий искусственного интеллекта, особенно в области обработки естественного языка. Большие языковые модели (БЯМ), такие, как GPT-3 и его последующие версии, продемонстрировали впечатляющие способности в понимании и генерации текста, что открывает новые горизонты для их применения в различных сферах. Одной из наиболее перспективных областей использования БЯМ является здравоохранение, где системы поддержки принятия решения (СППР) играют ключевую роль в улучшении качества медицинских услуг и повышении эффективности работы медицинских учреждений.

Целью данного исследования является разработка метода и подхода, направленных на сокращение времени, необходимого для поиска и

получения профильной информации на примере клинических рекомендаций.

Актуальность этой цели обусловлена тем, что в современных условиях скорость доступа к нужной информации играет ключевую роль в повышении производительности труда, улучшении качества принимаемых решений и снижении рисков, связанных с человеческими ошибками на основе официально утвержденных баз данных и базах знаний.

### **Обзор состояния вопроса**

В обзорной статье [1] Дэниэл Р. и др. рассматривают применение глубокого обучения, включая БЯМ, в области медицинской информатики и систем поддержки принятия решений.

В статье [2] Джаан Л. и др. исследовали, как БЯМ могут помочь в предсказании повторной госпитализации пациентов. Для анализа большого объема клинических данных была использована модель BERT.

В статье [3] Клертон А. рассматривает применение BERT и GPT в медицинских текстах. Автор анализирует, как данные модели могут применяться для поддержки принятия клинических решений.

В статье [4] проводят исследование технологий обработки естественного языка, которые могут быть интегрированы в системы поддержки принятия решений для повышения качества медицинских услуг.

### **Основные понятия и терминология**

Большие языковые модели (БЯМ) представляют собой разновидность искусственного интеллекта, которые предназначены для обработки и генерации естественного языка [5]. Эти модели обучаются в несколько этапов через анализ больших объемов текстовой информации.

Модель не понимает смысл слов так же как люди, вместо этого её обучают на дополнительных данных либо задавая параметры ответа. На

---

первом этапе обучения (Pretraining) модель обрабатывает множество данных с большим количеством параметров и пытается предсказать следующее слово, используя распределение вероятности [6]. На втором этапе – дообучении (Finetuning) модель учится отвечать на конкретные вопросы вместо генерации общего текста. Для дообучения модели используют специально подобранные статьи, руководства и инструкции, а также привлекают специалистов из различных областей – инженеров, юристов, менеджеров и других.

Системы поддержки принятия решений (СППР) — это компьютерная автоматизированная система, целью которой является помощь людям, принимающим решение в сложных условиях для полного и объективного анализа предметной области [7]. СППР применяются в таких отраслях как промышленность, медицина, торговля, электроснабжение города [8] и многих других. Основные функции СППР включают в себя сбор и обработку данных, их анализ, а также моделирование для оценки различных сценариев и их последствий. Помимо этого, многие современные СППР способны предоставлять информацию и рекомендации в реальном времени, что особенно важно в постоянно меняющихся условиях.

Использование больших языковых моделей в медицине открывает новые возможности для улучшения диагностики, лечения и управления медицинскими данными. БЯМ находят применение в различных направлениях, таких как обработка медицинских текстов, поиск клинических исследований [9], краткое изложение истории болезни и результатов клинических испытаний, предоставление полезной информации с медицинских конгрессов, медицинское образование, а также поддержка клинических решений, которая помогает врачам выбирать оптимальные методы лечения на основе анализа данных о пациентах и существующих клинических рекомендаций [10]. Помимо этого, БЯМ могут быть

---

использованы для разработки чат-ботов, которые отвечают на вопросы пациентов и медработников.

### Архитектуры больших языковых моделей

Обработка естественного языка (Natural Language Processing, NLP) представляет собой динамичную развивающуюся область искусственного интеллекта, основанную на взаимодействии между компьютерами и человеческим языком. Основная цель NLP заключается в том, чтобы создавать такие компьютерные системы, которые способны понимать, интерпретировать и генерировать человеческий язык в полезной и значимой форме.

В процессе обработки текста всегда стоит учитывать кодирование, используемое в конкретном документе [11]. Первым этапом обработки текста является его разбиение на элементы (токены), такие как слова или фразы. Далее каждому слову в тексте присваивается существительное, прилагательное или глагол, что позволяет понять структуру и смысл предложения. Синтаксический анализ позволяет определить связи между словами. Семантический анализ направлен на понимание элементов текста в контексте, для этого нередко используются векторные представления, такие как word embeddings, Word2Vec и GloVe.

Современные подходы к NLP активно используют алгоритмы машинного обучения, такие как классификация, регрессия и кластеризация для обучения моделей на основе размеченных данных, что позволяет моделям адаптироваться к различным задачам – анализ текстов, настроений и распознавание именованных сущностей. С появлением глубоких нейронных сетей, таких как рекуррентные нейронные сети (RNN) и трансформеры, возникли новые модели, которые повысили качество обработки текста.

Трансформеры (Transformers) – архитектура, разработанная компанией OpenAI. Со временем она стала стандартом в NLP благодаря

---

масштабируемости и эффективности. Одним из компонентов трансформеров является механизм самообучения (Self-attention). Кодировщик-декодировщик (Encoder-decoder architecture) компонент, позволяющий обрабатывать входной текст (кодировщик) и затем генерировать выходной текст (декодировщик).

На базе архитектуры трансформеров появились такие модели, как GPT (Generative Pre-trained Transformer), BERT (Bidirectional Encoder Representations from Transformers) и T5 (Text-to-Text Transfer Transformer). Модель GPT специализируется на генерации текста, а BERT для задач классификации и извлечения информации. T5 хорошо справляется с задачами перевода и обобщенного вывода.

Рекуррентные нейронные сети (RNN) были популярны до появления трансформеров. Особенность архитектуры RNN состояла в умении сохранять информацию о предыдущих шагах и это позволило учитывать контекст. Базовым вариантом является простая RNN, после которой появилась модификация LSTM (Long Short-Term Memory) с добавляющимися ячейками памяти для учета долговременных зависимостей. GRU (Gated Recurrent Unit) – упрощенный вариант LSTM с меньшим количеством учитываемых параметров.

Конволюционные нейронные сети (CNN) эффективны для задач, которые требуют анализа небольших участков текста (тегирование, классификация), поскольку работают с окнами фиксированного размера. В качестве примера можно привести модель TextCNN, которая использует сверточные слои.

Графовые нейронные сети (GNN) обрабатывают структурированные текстовые данные в виде графов. В задачах NLP их используют для представления отношений между словами и фразами. Механизмы внимания для обработки графов используются в модели Graph-based Attention Network

---

(GAT).

### **Особенности работы БЯМ и уровни знаний**

Одной из главных особенностей БЯМ является их способность работать с огромными объемами данных. Такие модели как ChatGPT-3 содержат миллиарды параметров, которые позволяют обучаться на огромном количестве данных и за счет этого генерировать качественный результат. Чем больше данных используется для тренировки модели, тем точнее и сложнее становятся ее прогнозы.

Самообучение является еще одной важной особенностью БЯМ. Под самообучением подразумевается способность модели тренироваться на размеченных данных, что исключает необходимость ручного аннотирования.

Механизм внимания позволяет концентрироваться на отдельных частях текста и учитывать как ближние, так и дальние зависимости между словами.

Преимуществом БЯМ стала их универсальность – одна модель может выполнять разные задачи. Например, Gigachat от Сбера способен генерировать тексты, картинки, резюмировать документы, переводить файлы, проводить сложные вычисления и многое другое.

Если модель не способна корректно ответить на вопрос или решить практическую задачу, то вероятнее всего ей не хватает знаний. Знания модели имеют 3 уровня:

**Ядро (L0).** Данный уровень отражает общий уровень знаний модели о мире, понимание простых вопросов и задач (хорошо/плохо, всегда/никогда и т. п.). Модель обучается на разметке и массиве публичных данных.

**Предметная область (L1).** На данном уровне модель лучше ориентируется в специфичных знаниях определенной области, таких, как юриспруденция, кибербезопасность, медицина и других. Для этого модель

---

обучается на статьях, инструкциях, документах, а разметку выполняют эксперты.

Практические навыки и решение прикладных задач (L2). На этом уровне знания модели основываются на данных и ситуациях, которые предоставил пользователь. Для обучения модели применяют системный промптинг, т. е. предоставляют исходные данные (например, в виде роли) и набор размеченных данных по категориям.

Прежде чем модель сможет решать прикладные задачи уровня L2, она должна сначала иметь специфические знания уровня L1 для того, чтобы избежать генерации бессвязного текста.

### **Возможности БЯМ для обработки естественного языка**

Промптом называют вопросы к нейросети. Промпт-инжиниринг (prompt engineering) представляет собой процесс взаимодействия с нейросетью в виде создания и совершенствования запросов (промпов) к языковым моделям с целью достижения наилучших результатов.

Одним из самых эффективных методов промпт-инжиниринга является правильное формулирование запроса, ведь от этого зависит качество информации, которую выдает модель. Не стоит забывать, что промпт должен быть максимально четким, конкретным и кратким. Можно добавлять контекст для того, чтобы модель лучше могла сформулировать точный ответ.

Подсказки полезно использовать на моделях, которые предварительно не были обучены специальной задаче. Для этой цели перед основным вопросом помещают пример вопроса и ответа, который бы хотел получить пользователь при ответе на следующий вопрос.

Добавление инструкций необходимо тогда, когда модель не понимает, каким образом она должна обработать промпт, или для решения сложной задачи. В последнем случае допустимо писать промпт в виде пошагового

---



изложения рассуждений. Разбиение большого и сложного промпта на несколько маленьких и простых также облегчает работу модели.

Ролевое моделирование, или системный промпт, заставляет модель отвечать определенным образом. Например, в Gigachat доступны роли system, user, assistant, function.

БЯМ находят широкое применение в решении различных задач NLP. Одной из таких задач является трансформация информации, т. е. изменение текста без потери его смысла – перевод, пересказ, резюмирование. При этом модель может менять стиль текста по запросу пользователя. Для создания коротких заголовков можно подать модели на вход полную информацию, а на выходе она выдаст основное содержание другими словами, но с тем же смыслом.

Классификация информации осуществляется моделью за счет присвоения меток текстовым данным, имеющим определенный смысл. Классификация помогает фильтровать спам в электронной почте, анализировать тональность текста и делить тексты по категориям. Модели CNN используются для семантической маркировке ролей, классификации типов вопросов, категоризации текста [12].

Помимо перечисленных выше задач, БЯМ помогают извлекать необходимую информацию из текста, вести диалог с пользователем через интерфейс чат-ботов, сравнивать и анализировать информацию из разных источников.

### **Исследование возможностей современных БЯМ по извлечению информации из текстов клинических рекомендаций**

Удобный классификатор клинических рекомендаций по МКБ-10 размещен на официальном сайте Министерства Здравоохранения РФ. После нажатия на ссылку с конкретным диагнозом можно прочитать клинические

---



рекомендации онлайн либо скачать pdf-файл. Исследование в области промптов (запросов) проводилось на клинических рекомендациях К29 «Гастрит и дуоденит» КР708 (рис.1).

ID	Название	МКБ-10	Одобрено Научно-практическим Советом Минздрава РФ
КР708	Гастрит и дуоденит	K29, K31.8	<input checked="" type="checkbox"/>

Рис. 1. – Классификатор клинических рекомендаций по МКБ-10

Для автоматического анализа текста из pdf-файла необходимо загрузить его в одну из доступных языковых моделей. Для этой цели были выбраны популярные бесплатные БЯМ с возможностью загрузки документов – Gigachat, ChatPDF и AnySummary. Достоинства и недостатки каждой из них описаны в таблице № 1.

Таблица № 1

### Сравнение различных БЯМ с возможностью загрузки документов

	Достоинства	Недостатки
Gigachat	- Нейросеть предлагает подсказки после ответа на вопрос пользователя. - Нет ограничений по количеству загруженных	- Нельзя увидеть текст документа, с которым работает пользователь. - Необходимо давать указание отвечать строго по тексту

	файлов в сутки.	загруженного документа. - Необходимо авторизоваться.
ChatPDF	- В левой части интерфейса отображается документ, с которым работает пользователь. - Чат-бот ищет ответ на вопрос в тексте документа. - Регистрация не обязательна.	- Ограничение загрузки до 2-х файлов в сутки в бесплатной версии.
AnySummary	- Простой и понятный интерфейс. - Регистрация необязательна.	- Выполняет только краткую суммаризацию текста. - Нельзя задать и скорректировать вопрос.

Поскольку важно исследовать эффективность ответов на запросы по документу, для дальнейшего исследования были выбраны БЯМ с интерфейсом чат-бота – ChatPDF и Gigachat.

После загрузки pdf-файла с рекомендацией в ChatPDF модель сразу выдает краткое резюме о том, что содержит данный документ, кем был разработан и предлагает примерные вопросы (рис.2). Gigachat не проводит предварительный анализ текста и из примерных вопросов предлагает сделать краткое содержание документа (рис.3).

## Chat



Дорогие читатели!



С радостью представляем вам клинические рекомендации по гастриту и дуодениту, утвержденные ведущими специалистами в области гастроэнтерологии и микробиологии. Этот документ, разработанный Российской гастроэнтерологической ассоциацией и другими авторитетными организациями, содержит актуальную информацию и рекомендации для врачей и медицинских работников, работающих с взрослыми пациентами. Мы надеемся, что эти рекомендации помогут вам в вашей практике и способствуют улучшению здоровья пациентов!

С наилучшими пожеланиями,  
Ваша команда поддержки

Примеры вопросов, которые вы можете задать о файле:

- Какие основные рекомендации содержатся в документе по гастриту и дуодениту?
- Кто является разработчиком клинических рекомендаций?
- Какова возрастная категория, для которой предназначены эти рекомендации?

Рис. 2 – Результат ChatPDF

Дай краткое содержание

KP708.pdf

**Гастрит и дуоденит** - это воспалительные заболевания слизистой оболочки желудка и двенадцатиперстной кишки. Основные виды гастрита включают острый и хронический гастрит, а также аутоиммунный гастрит.

**Основные причины гастрита:**

1. Инфекция *Helicobacter pylori* (Hр) - наиболее частая причина хронического гастрита.
2. Аутоиммунные процессы, ведущие к атрофическому гастриту.
3. Другие факторы, такие как лекарства (НПВП), алкоголь, радиация, химические вещества и механические повреждения.

**Диагностика:**

- Основой диагностики является эндоскопическое исследование (ЭГДС) с биопсией.
- Дополнительно используются лабораторные тесты (кровь на антитела, пепсиногены, гастрин-17) и дыхательные тесты для выявления Hр.

**Лечение:**

- Основное внимание уделяется эрадикационной терапии Hр, которая включает

Рис. 3. – Результат Gigachat

На запрос «Как проводится диагностика гастрита?» ChatPDF дал краткие тезисы из документа со ссылками на номера страниц (рис.4), в то время как Gigachat при ответе на данный запрос упомянул систему OLGА и OLGIM для оценки риска развития рака желудка, которых нет в тексте данной клинической рекомендации (рис.5). В связи с этим для получения более корректного ответа от Gigachat в последующих запросах использовалось указание «Отвечай по тексту из документа».

Как проводится диагностика гастрита?

■ Диагностика гастрита проводится с использованием различных методов, включая:

- 1. Клинический осмотр и анамнез:**
  - Сбор информации о симптомах, таких как боль в животе, диспепсия, тошнота и рвота, а также о факторах риска, таких как прием нестероидных противовоспалительных препаратов (НПВП) и алкоголь.
- 2. Эндоскопическое исследование:**
  - Эндоскопия (ЭГДС) позволяет визуально оценить состояние слизистой оболочки желудка и взять биопсию для гистологического анализа. Это один из наиболее информативных методов диагностики гастрита <sup>11</sup>.
- 3. Гистологическое исследование:**
  - Биопсия, полученная во время эндоскопии, исследуется на наличие воспалительных изменений, атрофии и кишечной метаплазии, а также на наличие *H. pylori* <sup>11</sup>.
- 4. Неинвазивные тесты на *H. pylori*:**
  - Дыхательный тест с уреазой, определение антигена *H. pylori* в кале и ПЦР. Эти тесты помогают установить наличие инфекции без необходимости инвазивных процедур <sup>11</sup>.

Рис. 4. – Результат ChatPDF

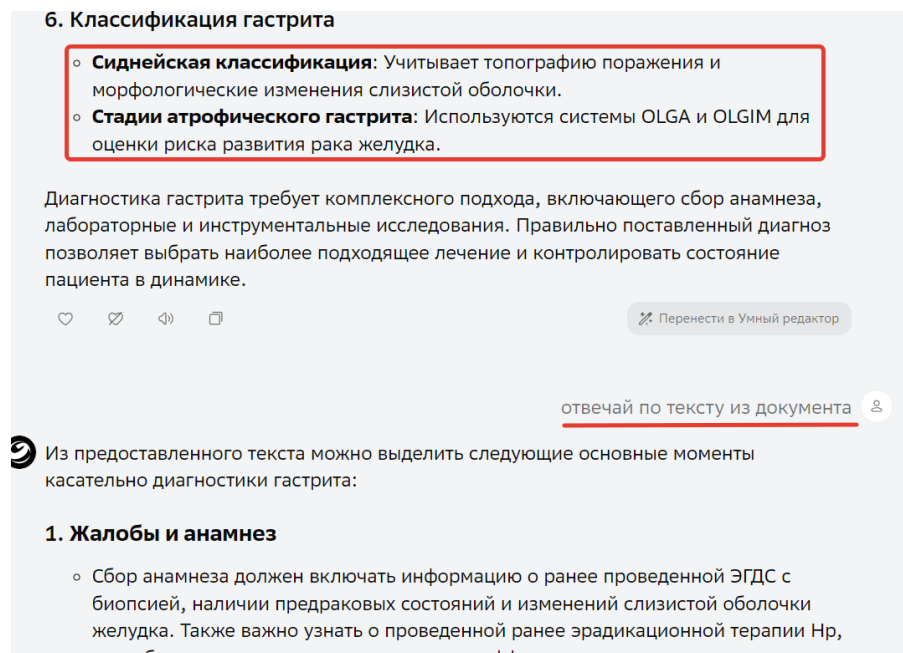


Рис. 5. – Результат Gigachat

Про лечение болезни обе БЯМ выдают схожий результат – медикаментозное лечение, схемы лечения и указание на санаторно-курортное лечение, но помимо этого ChatPDF пишет еще про побочные эффекты, диетические рекомендации и продолжительность лечения (рис.6). Ответ Gigachat получился более кратким и тезисным (рис.7).

лечение




-  Лечение инфекции *Helicobacter pylori* (*H. pylori*) включает несколько ключевых компонентов, направленных на уничтожение бактерии и восстановление здоровья желудка. Вот основные аспекты лечения:
- 1. Медикаментозное лечение:**
- **Антибиотики:** Используются для уничтожения *H. pylori*. Наиболее распространенные антибиотики включают:
    - Амоксициллин
    - Кларитромицин
    - Метронидазол
    - Тетрациклин
  - **Ингибиторы протонного насоса (ИПП):** Эти препараты снижают выработку желудочной кислоты, что помогает антибиотикам действовать более эффективно. Примеры:
    - Омепразол
    - Пантопразол
    - Эзомепразол
  - **Бисмутовые препараты:** Например, висмута трикалия дицитрат, который может быть добавлен для защиты слизистой оболочки и улучшения результатов лечения.
- 2. Схемы лечения:**

Рис. 6. – Результат ChatPDF

лечение в соответствии с документом 

 Лечение гастрита и дуоденита у взрослых пациентов должно осуществляться в соответствии с клиническими рекомендациями, разработанными и утвержденными экспертными организациями. Основные принципы лечения включают:

**1. Медикаментозное лечение**

**a. Эрадикационная терапия *Helicobacter pylori* (H<sub>p</sub>)**

- **Первая линия терапии:** включает ингибиторы протонного насоса (ИПП), кларитромицин и амоксициллин. Может быть усилена препаратом висмута.
- **Вторая линия терапии:** при неудаче первой линии используют квадротерапию (висмут + ИПП + тетрациклин + метронидазол) или тройную терапию с левофлоксацином.

**b. Симптоматическое лечение**

- **Ингибиторы протонного насоса (ИПП):** омепразол, пантопразол, рабепразол, эзомепразол. Используются для снижения кислотности и уменьшения симптомов.
- **Прокинетики:** домперидон, итоприд, тримебутин. Помогают улучшить моторику ЖКТ.
- **Цитопротекторы:** ребамипид. Усиливают защиту слизистой оболочки желудка.

Рис. 7. – Результат Gigachat

Насчет возрастной группы, к которой относятся данные клинические рекомендации, Gigachat верно ответил в соответствии с информацией на 1-й странице клинической рекомендации о том, что она применима к взрослым пациентам (рис.8). Однако, ChatPDF не смог распознать эту информацию и выдал более общий и развернутый ответ, которого нет в тексте клинической рекомендации (рис.9).

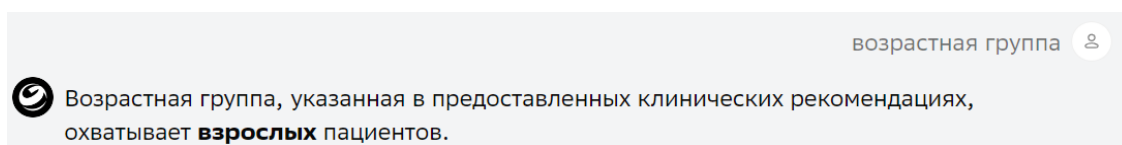


Рис. 8. – Результат Gigachat

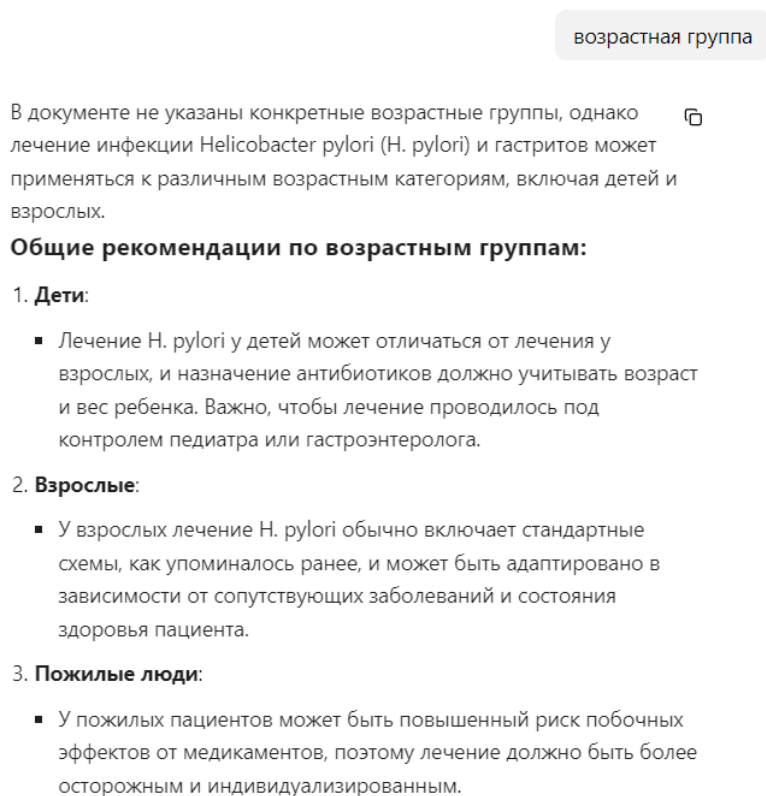


Рис. 9. – Результат ChatPDF

При запросе «Анамнез» ChatPDF на рис.10 выдал общую информацию без упоминания таких важных процедур, как была ли проведена процедура



ЭДГС с биопсией и т. п. Для этого пришлось больше конкретизировать запрос - «Что следует учесть при сборе анамнеза?» (рис.11) Gigachat справился с вопросом «Анамнез» изначально в более верном направлении (рис.12).

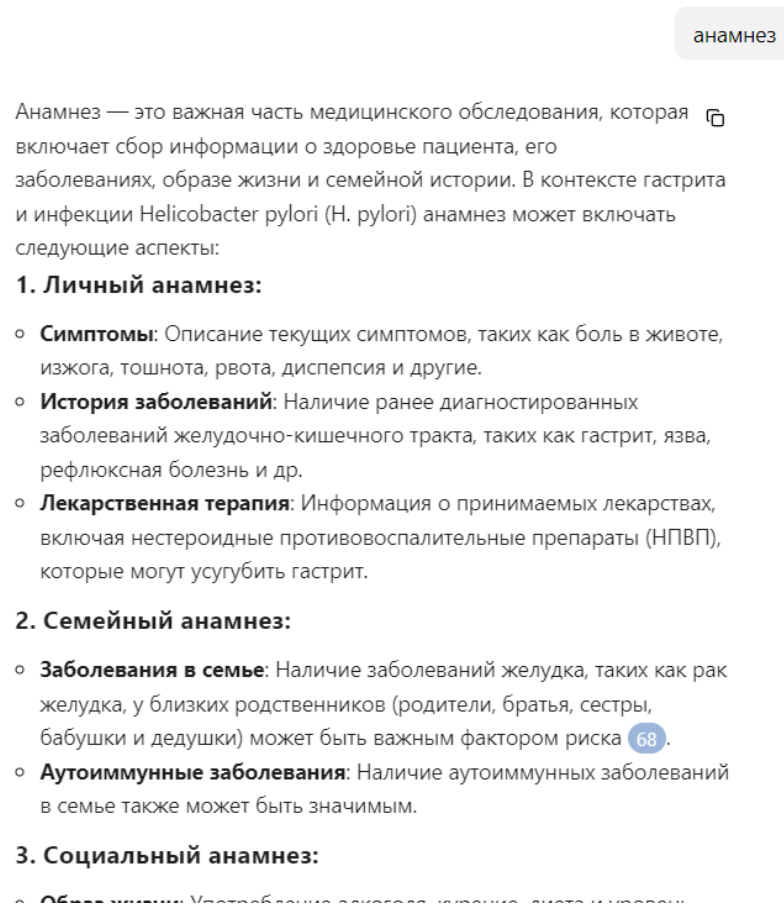


Рис. 10. – Первоначальный запрос об анамнезе в ChatPDF

Что следует учесть при сборе анамнеза

! При сборе анамнеза у пациента с подозрением на гастрит или инфекцию *Helicobacter pylori* (*H. pylori*) следует учесть следующие ключевые аспекты:

**1. Симптомы:**

- **Основные жалобы:** Боль в животе, изжога, тошнота, рвота, диспепсия, потеря аппетита, изменение стула.
- **Продолжительность и характер симптомов:** Когда начались симптомы, как они изменялись со временем, есть ли связь с приемом пищи.

**2. История заболеваний:**

- **Предыдущие диагнозы:** Наличие ранее диагностированных заболеваний желудочно-кишечного тракта, таких как гастрит, язва, рефлюксная болезнь.
- **Эндоскопические исследования:** Уточнить, проводились ли ранее ЭГДС с биопсией, и были ли выявлены предраковые состояния (атрофия, метаплазия, дисплазия) <sup>11</sup>, <sup>12</sup>.

**3. Лекарственная терапия:**

- **Принимаемые препараты:** Уточнить, какие лекарства принимает пациент, включая НПВП, антибиотики, антациды и другие препараты, которые могут влиять на слизистую оболочку желудка.

Рис. 11. – Конкретизированный запрос об анамнезе в ChatPDF

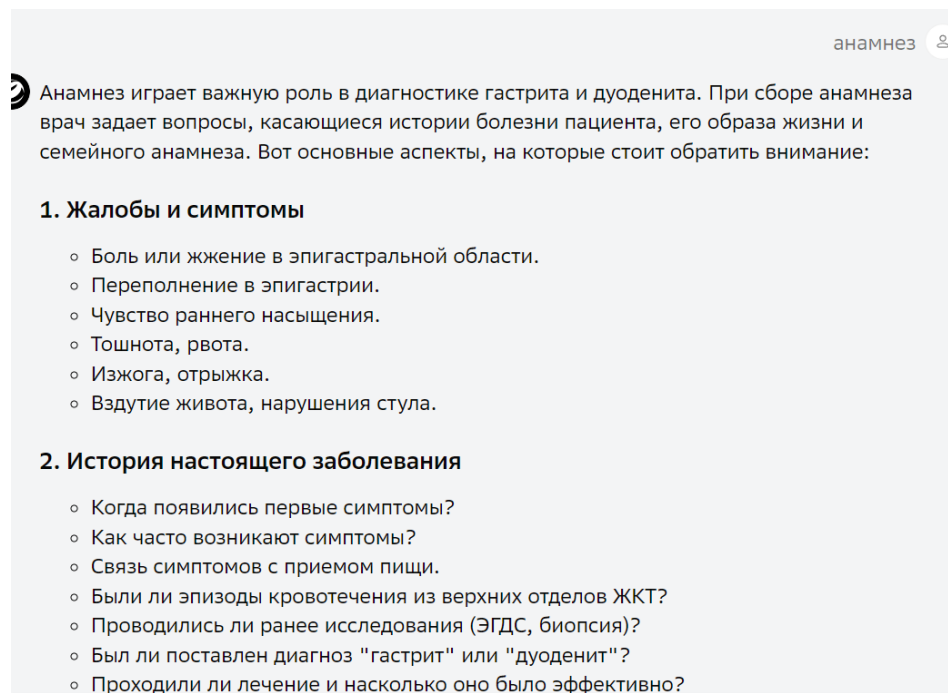


Рис. 12. – Результат Gigachat

Более сложный запрос к Gigachat «опиши механизм лечения с применением четырехкомпонентная терапии с висмутом трикалия дицитратом в соответствии с документом», за которым последовало конкретное описание процедуры по документу (рис.13). ChatPDF не сослался на дозировки препаратов, а лишь описал свойства препаратов (рис.14).

опиши механизм лечения с применением четырехкомпонентная терапии с висмутом трикалия дицитратом в соответствии с документом

Четырехкомпонентная терапия с висмутом трикалия дицитратом (ВТД) является одной из наиболее эффективных схем лечения инфекции *Helicobacter pylori* (Hр). Она используется в качестве терапии второй линии, если первая линия оказалась неэффективной.

**Состав четырехкомпонентной терапии с ВТД:**

1. **Ингибитор протонного насоса (ИПП):** Омепразол, пантопразол, рабепразол или эзомепразол в стандартной дозе дважды в день.
2. **Висмута трикалия дицитрат (ВТД):** 120 мг четыре раза в день или 240 мг два раза в день.
3. **Тетрациклин:** 500 мг четыре раза в день внутрь.
4. **Метронидазол:** 500 мг три раза в день.

**Механизм действия:**

1. **Ингибиторы протонного насоса (ИПП):** Снижают кислотность желудочного сока, создавая неблагоприятные условия для жизнедеятельности Hр. Это усиливает действие антибиотиков и способствует заживлению слизистой оболочки желудка.

Рис. 13. – Результат Gigachat

опиши механизм лечения с применением четырехкомпонентная терапии с висмутом трикалия дицитратом

Четырехкомпонентная терапия с висмутом трикалия дицитратом является одним из методов лечения инфекции *Helicobacter pylori* (*H. pylori*) и включает в себя комбинацию нескольких препаратов, направленных на эрадикацию этой бактерии и лечение гастрита. Механизм действия этой терапии можно описать следующим образом:

### 1. Компоненты терапии:

Четырехкомпонентная терапия обычно включает:

- **Висмут трикалия дицитрат:** Обладает защитным действием на слизистую оболочку желудка, способствует заживлению язв и эрозий, а также обладает антимикробной активностью против *H. pylori*.
- **Ингибиторы протонной помпы (ИПП):** Например, омепразол или лансопризол, которые снижают секрецию желудочной кислоты, создавая более благоприятные условия для действия других компонентов терапии и уменьшая кислотность в желудке.
- **Антибиотики:** Обычно используются тетрациклин и метронидазол, которые действуют на *H. pylori*, подавляя его размножение и способствуя его уничтожению.

### 2. Механизм действия:

- **Антимикробное действие:** Висмут трикалия дицитрат и антибиотики действуют синергически, что повышает эффективность

Рис. 14. – Результат ChatPDF

Выводы по результатам исследования приведены в таблице №2.

Таблица № 2

### Сравнение качества ответов ChatPDF и Gigachat

	При ответе всегда выдает информацию строго из документа	Учитывает контекст из схем	Понимание контекста медицинских терминов без уточнения конкретики	Сохраняет историю диалога
Gigachat	-	+	-	+
ChatPDF	-	-	-	+

В результате данного исследования запросов выявились следующие

проблемы:

- ChatPDF в некоторых случаях может неправильно интерпретировать запросы и не предоставить точного ответа, особенно если это специфические медицинские термины, поскольку это не его единственная специализация. Вдобавок к этому, ChatPDF извлекает текст из схем лечения без учета окружающего его контекста.

- Gigachat также сталкивается с трудностями в точной интерпретации медицинской лексики. Gigachat склонен добавлять при ответе информацию не из загруженного документа, поэтому запрос всегда должен содержать указание отвечать строго по документу. Извлеченный текст бывает трудно понять без контекста, в связи с чем усложняется его использование для дальнейшего анализа или принятия решений, например, врачом.

Таким образом, обе БЯМ имеют свои достоинства и недостатки. Они неплохо справляются с общими и несложными вопросами, но возникают сложности понимания контекста медицинских терминов. Медработникам может быть трудно принимать решения, полностью опираясь на БЯМ, поскольку они склонны добавлять при ответе информацию не из документа. На данном этапе развития нейросети больше работают как подсказки, а не полноценные помощники.

Дальнейшие исследования планируется проводить в области улучшения промптов и создания базы промптов специально для клинических рекомендаций, а также улучшения алгоритмов.

### **Концепция системы поддержки принятия решений для оптимизации времени доступа к профильной информации**

Задача заключается в разработке системы поддержки принятия решений для оптимизации времени доступа к клиническим рекомендациям. Система позволит медицинским работникам быстро находить нужные

---

клинические рекомендации на основе МКБ-кодов, что поможет сократить время на поиск информации и повысит качество медицинской помощи.

Целью создания данной системы является сокращение времени на поиск информации; улучшение качества медицинской помощи за счет повышения точности и полноты информации, которой будет достаточно для принятия решений.

Клинические рекомендации (КР) – это документы, в которых содержится информация о диагностике, лечении, профилактике и реабилитации заболевания. Помимо информационной поддержки принятия решений практикующим медицинским работникам, КР также используются для обеспечения непрерывного медицинского образования и повышения квалификации [13].

Международная классификация болезней (МКБ) – это нормативный документ, с помощью которого ведется статистическая классификация болезней. В МКБ каждой группе болезней присвоен уникальный код, который един для всего мира [14].

На рис. 15 представлена схема поиска контекстной информации в системе на основе промптов. Промпт-инжиниринг представляет собой процесс целенаправленного конструирования текстовых запросов-подсказок [15].

На основе введенных пользователем данных (симптомы, жалобы, история болезни) к системе через веб-интерфейс в виде чат-бота происходит взаимодействие с API официального ресурса Министерства Здравоохранения, после которого система, используя алгоритмы машинного обучения, классифицирует запрос по МКБ. Пользователь может подтвердить либо опровергнуть выбор кода МКБ. Код МКБ помогает сузить круг поиска клинических рекомендаций. После выбора МКБ кода система оценивает степень детализации запроса через дополнительные параметры, такие как

---

пол, возраст, сопутствующие заболевания. На основе этой информации система фильтрует подходящие клинические рекомендации из базы данных. Подходящие КР загружаются в память модели для того, чтобы их можно было использовать при последующих запросах.

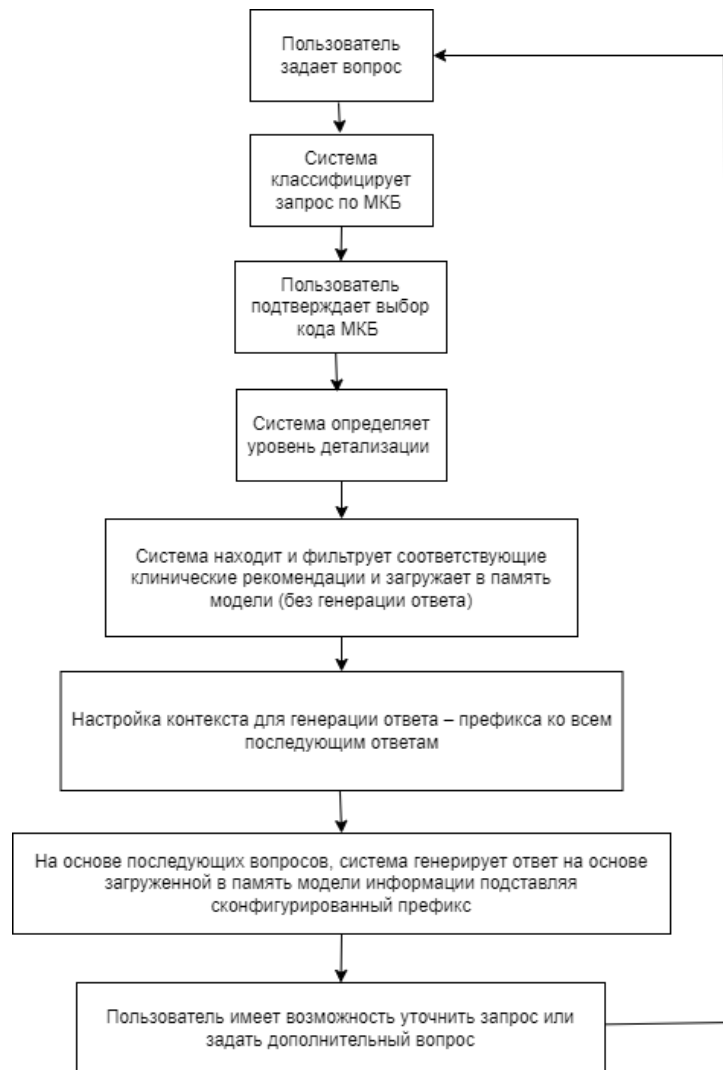


Рис. 15. – Схема поиска контекстной информации в системе на основе промптов



## Заключение

Данный обзор показал, что большие языковые модели представляют собой мощный инструмент для улучшения многих процессов в системе здравоохранения. Способность больших языковых моделей обрабатывать огромные объемы данных, извлекать необходимую информацию и предоставлять точные рекомендации может сделать их незаменимыми помощниками врачей и специалистов в области медицины. Также было проведено исследование различных промптов по клиническим рекомендациям в ChatPDF и Gigachat и на их примере удалось выявить недостатки, которые планируется скорректировать в проектируемой системе.

Таким образом, внедрение больших языковых моделей в здравоохранение является стратегически важным направлением, способным повысить уровень медицинской помощи.

## Литература

1. Daniele, R., D. Fani and W. Charence. Deep Learning for Health Informatics // ResearchGate. 2016. Vol. 99. pp. 1-18.
2. Huang, K., J. Li and R. Ranganath. ClinicalBERT: Modeling Clinical Notes and Predicting Hospital Readmission // ResearchGate. 2019. Vol. 2. pp. 7-13.
3. Claurton, A.S., K.B. Mascha and W. Katarzyna. Transformers in health: a systematic review on architectures for longitudinal data analysis // Springer Nature Link. 2024. Vol. 32. pp. 24-27.
4. Usman, A., I. Khurshed, A. Muhammad and K. Ghazi. Natural Language Processing for Clinical Decision Support Systems: A Review of Recent Advances in Healthcare // ResearchGate. 2023. Vol. 5. pp. 1-15.
5. Sinan, O., 2023. Quick Start Guide to Large Language Models: Strategies and Best Practices for Using ChatGPT and Other LLMs. Addison-

Wesley Professional.

6. Kushal, T., S. Daniel, A. Armen and M. Ari. D4: Improving LLM Pretraining via Document De-Duplication and Diversification // ResearchGate. 2023. Vol. 122. pp. 7-11.

7. Стародубцев А.А. Система поддержки принятия решений // Актуальные проблемы авиации и космонавтики. 2016. №11. С. 99-101.

8. Карелин А.Е, Береза А.Н. Экспертная система для поддержки принятия решений оператора в системе электроснабжения города // Инженерный вестник Дона. 2019. №4. URL: ivdon.ru/ru/magazine/archive/n4y2019/5941.

9. Донская А.Р., Сибирный Н.Д., Зубков А.В., Орлова Ю.А., Фролов М.Ю., Саласюк А.С. Разработка метода поиска клинических исследований по включающим факторам // Инженерный вестник Дона. 2022. №12. URL: ivdon.ru/ru/magazine/archive/n12y2022/8080.

10. Адилова Ф.Т., Давронов Р. Р., Сафаров Р.А., Кушмуратов С.И. Языковые модели (GPT) в здравоохранении: клиническая практика и медицинское образование // Universum: технические науки. 2023. №11(116). С. 8-16.

11. Цитульский А.М., Иванников А.В., Рогов И.С. NLP - обработка естественных языков // StudNet. 2020. №6. С. 467-475.

12. Курейчик В.В., Родзин С.И., Бова В.В. Методы глубокого обучения для обработки текстов на естественном языке // Известия Южного федерального университета. Технические науки. 2022. №23. С. 189-198.

13. Клинические рекомендации // Федеральная электронная библиотека. URL: femb.ru/request (Дата обращения: 24.12.2024).

14. МКБ-10 // Злокачественные опухоли. 2021. №11. С. 5-11.

15. Алферьева-Термсикос В.Б. Промпт-инжиниринг как стратегия формирования информационной культуры обучающихся // Международный



журнал гуманитарных и естественных наук. 2024. №96. С. 10-14.

### References

1. Daniele, R., D. Fani and W. Charence. ResearchGate. 2016. Vol. 99. pp. 1-18.
  2. Huang, K., J. Li and R. Ranganath. ResearchGate. 2019. Vol. 2. pp. 7-13. pp. 24-27.
  3. Clairton, A.S., K.B. Mascha and W. Katarzyna. Springer Nature Link. 2024. Vol. 32. pp. 24-27.
  4. Usman, A., I. Khurshed, A. Muhammad and K. Ghazi. ResearchGate. 2023. Vol. 5. pp. 1-15.
  5. Sinan, O., 2023. Quick Start Guide to Large Language Models: Strategies and Best Practices for Using ChatGPT and Other LLMs. Addison-Wesley Professional.
  6. Kushal, T., S. Daniel, A. Armen and M. Ari. ResearchGate. 2023. Vol. 122. pp. 7-11.
  7. Starodubtsev A.A. Aktual'nyye problemy aviatsii i kosmonavтики. 2016. №11. pp. 99-101.
  8. Karelin A.E, Bereza A.N. Inzhenernyj vestnik Dona. 2019. №4. URL: [ivdon.ru/ru/magazine/archive/n4y2019/5941](http://ivdon.ru/ru/magazine/archive/n4y2019/5941).
  9. Donskaya A.R., Sibirnyy N.D., Zubkov A.V., Orlova YU.A., Frolov M.YU., Salasyuk A.S. Inzhenernyj vestnik Dona. 2022. №12. URL: [ivdon.ru/ru/magazine/archive/n12y2022/8080](http://ivdon.ru/ru/magazine/archive/n12y2022/8080).
  10. Adilova F.T., Davronov R. R., Safarov R.A., Kushmuratov S.I. Universum: tekhnicheskiye nauki. 2023. №11(116). pp. 8-16.
  11. Tsitul'skiy A.M., Ivannikov A.V., Rogov I.S. StudNet. 2020. №6. pp. 467-475.
  12. Kureychik V.V., Rodzin S.I., Bova V.V. Izvestiya Yuzhnogo federal'nogo universiteta. Tekhnicheskiye nauki. 2022. №23. pp. 189-198.
-



13. Klinicheskiye rekomendatsii [Clinical recommendations]. Federal'naya elektronnaya biblioteka. URL: femb.ru/request (Accessed: 24.12.2024).

14. Zlokachestvennyye opukholi. 2021. №11. pp. 5-11.

15. Alfer'yeva-Termsikos V.B. Mezhdunarodnyy zhurnal gumanitarnykh i yestestvennykh nauk. 2024. №96. pp. 10-14.

**Дата поступления: 20.01.2025**

**Дата публикации: 25.03.2025**