Модульная переработка нормативных текстов как метод повышения релевантности ответов большой языковой модели: пилотное исследование на примере клинических рекомендаций по артериальной гипертензии

Е.М. Фролов, Д.Н. Ермолаева, К.Ю. Мокшин, Д.Д. Шемонаев, М.Ю.Фролов, М.Г.Жабицкий

Аннотация. **Цель исследования - оценить в пилотном сравнительном исследовании влияние модульной адаптации (сокращения) текста клинических рекомендаций, используемых как непосредственный источник информации, на качество (релевантность и безопасность) ответов большой языковой модели при решении реальных клинических задач, касающихся лечения артериальной гипертензии у взрослых. В исследование включены 45 клинических кейсов, сформулированных практикующими врачами, не имевшими опыта взаимодействия с LLM. Кейсы были рандомизированы в три группы: (1) адаптированный текст клинической рекомендации (исследуемая группа), (2) неадаптированный полный текст (контрольная 2), (3) отсутствие документа (контрольная 1). Ответы LLM (GPT-4) генерировались врачом-оператором с использованием кастомного промта (инструкции к языковой модели) и оценивались независимыми экспертами по шкале из трёх критериев: клиническая адекватность, безопасность, соответствие положениям клинических рекомендаций. Исследуемая группа показала наивысшие значения по всем критериям (средние баллы: КА — 8,8; БР — 9,5; СКР — 9,2). Контрольная группа 2 продемонстрировала высокую безопасность (9,1), но меньшую нормативную точность (СКР — 7,3). Контрольная группа 1 существенно отставала по всем показателям (КА — 6,3; БР — 7,5; СКР — 5,1). Установлено, что модульная адаптация (сокращение) клинических рекомендаций, направленная на удаление нерелевантных к задачам разделов, обеспечивает существенное повышение релевантности и нормативной точности медицинских ответов LLM. Подход может быть использован при проектировании систем поддержки принятия врачебных решений на базе LLM, в том числе в рамках цифровой трансформации клинической практики.**

Ключевые слова **– большие языковые модели (LLM), искусственный интеллект в медицине, принятие врачебных решений, системы поддержки принятия врачебных решений, клинические рекомендации, артериальная гипертензия, клиническая фармакология, оценка качества медицинских ответов, цифровая трансформация здравоохранения**

I. ВВЕДЕНИЕ

В последние годы наблюдается стремительное внедрение технологий искусственного интеллекта (ИИ) в клиническую медицину [1, 2]. Недавние обзоры и исследования фиксируют бурный рост применения ИИ в диагностике, прогнозировании и принятии клинических решений, причем ИИ применяется на всех этапах оказания медицинской помощи: от диагностики до образовательных и административных задач [1, 3, 4, 5].

Особое внимание привлекают большие языковые модели (large language models, LLM), которые демонстрируют способность к генерации развернутых, стилистически приближенных к экспертным медицинским заключений по открытым клиническим вопросам [3]. Эти модели уже используются врачами в реальной практике — как средство получения второго мнения, как инструмент быстрой справочной поддержки, а также как вспомогательный механизм в принятии решений в условиях дефицита времени и информационной перегрузки [1–3, 6, 7, 8].

Несмотря на высокий уровень интереса к применению LLM, их интеграция в клинические процессы сопровождается значительными рисками. Исследования показали, что качество медицинских ответов, полученных от LLM, может быть непостоянным и варьировать в зависимости от формулировки запроса, доступных источников информации, а также специфики нозологии и клинической задачи [4–6, 9, 10]. Отсутствие контекстуализации, некорректная интерпретация медицинской терминологии, а также генерация устаревших или не соответствующих локальным стандартам рекомендаций могут снижать клиническую ценность подобных ответов или даже представлять угрозу для пациента. В частности, небезопасные или не соответствующие клиническим рекомендациям предложения встречаются чаще при использовании LLM «из коробки», без дополнительной настройки или источников валидации [9, 10, 11].

В то же время, современные модели демонстрируют заметное улучшение качества ответов при использовании методик промт-инжиниринга и добавлении релевантных источников знаний [1, 3, 12, 13, 14]. В ряде публикаций было показано, что при структурированном взаимодействии с LLM, особенно при указании нормативных источников (например, клинических рекомендаций), значительно повышается релевантность и безопасность генерируемых текстов [4, 12]. Однако вопрос о том, как именно представлять эти источники и какие трансформации нормативного документа позволяют повысить релевантность ответов,

остаётся открытым. В частности, не изучено, влияет ли предварительная адаптация объёмных и регламентированных текстов (например, клинических рекомендаций Минздрава РФ) на качество интерпретации и генерации рекомендаций LLM.

Дополнительной проблемой остаётся недостаток доступных и валидированных систем поддержки принятия врачебных решений (СППВР), интегрированных в рабочую инфраструктуру российских лечебных учреждений [2, 5, 8]. Это приводит к тому, что практикующие врачи начинают экспериментировать с использованием открытых LLM в клиническом контексте без должной регламентации, что порождает новые риски — как клинические, так и этические. В условиях цифровой зрелости, не достигшей нужного уровня, LLM становятся спонтанными инструментами принятия решений, несмотря на их экспериментальный статус. Существующие СППВР часто не обладают достаточной гибкостью, не покрывают все клинические случаи и нередко ограничены в поддержке фармакотерапевтических решений, в частности в области коморбидных состояний и полипрагмазии.

В международной и российской практике существует устойчивый подход к сокращению и адаптации нормативных клинических документов для практического использования врачами. Помимо полной версии КР, многие профессиональные медицинские сообщества публикуют их сокращённые форматы — так называемые «карманные версии» (pocket guidelines), а также специализированные алгоритмические компиляции. Например, Европейское общество кардиологов (ESC) сопровождает каждое руководство краткой версией, содержащей только ключевые схемы принятия решений, шкалы, дозировки и списки рекомендуемых препаратов [15, 16]. Схожий подход реализуют Американская кардиологическая ассоциация (AHA), Американская диабетологическая ассоциация (ADA), Национальный институт здоровья и качества клинической практики Великобритании (NICE), а также ряд других организаций. Такие документы имеют самостоятельный статус и публикуются под названиями типа Executive Summary, Decision Pathway, Treatment Algorithm Compendium.

Сокращённые формы КР предназначены для быстрого доступа к клинически значимой информации в условиях ограниченного времени, повышенной когнитивной нагрузки и необходимости соблюдения стандартов лечения. Их использование повышает приверженность врачей рекомендациям, снижает вероятность ошибок и улучшает принятие решений в клинической практике [17 - 19]. Аналогичные принципы применимы и при создании источников знаний для LLM — адаптированный, модульный и логически декомпозированный текст может быть значительно более полезен для генерации релевантных и безопасных ответов [20, 21].

Современные базы знаний для СППВР и медицинских ИИ-систем, как правило, на так называемые «обогащённые» представления, включающие формализованные правила, онтологии, клинические триггеры и табличные маршруты. Такие подходы реализованы, в частности, в рамках стандартов HL7 FHIR, SNOMED CT, GDL2 и др., где информация подаётся в машиночитаемом и структурированном виде, обеспечивая автоматическую интерпретацию алгоритмами [22 - 25].

Однако в контексте генеративных языковых моделей общего назначения, таких как GPT, использование оригинальных текстов клинических рекомендаций сопровождается целым рядом трудностей. Структурная перегруженность, наличие многочисленных таблиц, диаграмм, отсылок, уровней доказательности, юридически нагруженных формулировок и условных конструкций существенно затрудняет корректное извлечение и интерпретацию информации. В отсутствие формальной структуры LLM может ошибочно интерпретировать ограничения или условия применения терапии, игнорировать исключения и противопоказания, синтезировать нерелевантный или небезопасный ответ при наличии модальных конструкций и двойственных формулировок [25 - 28].

Таким образом, существует обоснованная необходимость в предварительной адаптации нормативного текста для LLM-среды — аналогичной той, что применяется при разработке кратких версий КР. Такая модульная трансформация документа позволяет использовать только те части, которые имеют практическое значение для конкретной задачи, и обеспечивает более полную релевантность, интерпретируемость и безопасность создаваемых ответов. Настоящее исследование опирается на эту гипотезу и ставит своей целью эмпирически проверить её применимость в контексте генерации ответов LLM на клинические запросы врачей, опираясь на адаптированный текст клинических рекомендаций..

Выбор клинических рекомендаций (КР) по артериальной гипертензии (АГ) обусловлен высокой распространённостью, клинической значимостью и нормативной регламентацией этого заболевания. АГ служит типичным примером нозологии, для которой разработаны подробные клинические рекомендации, включающие стратификацию риска, алгоритмы терапии и выбор конкретных лекарственных препаратов. Это делает КР по АГ удобной моделью для оценки релевантности и нормативной точности медицинских рекомендаций, формируемых LLM.

Таким образом, исследование взаимодействия LLM с клиническими рекомендациями на примере АГ — в контексте адаптации, форматирования и целевого использования, представляет собой актуальное и практически значимое направление исследований. Оно не только направлено на повышение безопасности и точности использования LLM в клинической практике, но и формирует основу для будущего проектирования гибридных цифровых решений, объединяющих LLM, СППВР и формализованные нормативные базы, однако до настоящего времени не было предметом систематического исследования в отечественной литературе.

II. МЕТОДОЛГИЯ ИССЛЕДОВАНИЯ

Исследование выполнено в рамках проекта молодых ученых Высшей инжиниринговой школы НИЯУ МИФИ (г. Москва) и лаборатории фармакоэкономики, цифровой медицины и искусственного интеллекта НЦИЛС ВолгГМУ (г. Волгоград).

Целью работы явилась оценка в пилотном сравнительном исследовании влияния модульной адаптации (сокращения) текста клинических рекомендаций, используемых как непосредственный источник информации, на качество (релевантность и безопасность) ответов большой языковой модели при решении реальных клинических задач, касающихся лечения артериальной гипертензии у взрослых.

Методологическая база исследования опирается на принципы системного инженерного подхода, включающего:

декомпозицию задачи взаимодействия врача с LLM на управляемые компоненты (структура запроса, тип источника, формат передачи),

формализацию этих компонентов в виде модульных решений (структура запроса, адресуемого LLM, структура клинических рекомендаций),

последовательную оптимизацию каждого компонента,

обеспечение воспроизводимости и масштабируемости методики в клинических условиях.

В качестве нормативной базы использован документ «Клинические рекомендации. Артериальная гипертензия у взрослых» (2024), обладающий официальным статусом в системе здравоохранения Российской Федерации и размещённый в Рубрикаторе клинических рекомендаций Минздрава России [29].

Анализ структуры и внутренней логики текста КР «Артериальная гипертензия у взрослых» был выполнен врачом-клинических фармакологом, обладающим экспертной компетенцией в области лекарственного лечения, нормативного регулирования лекарственной терапии и разработки систем поддержки принятия врачебных решений. Целью анализа являлось определение модулей, представляющих практическую ценность для генерации языковой модели медицинских ответов, а также идентификация фрагментов, не содержащих непосредственно клинически значимой информации для решения задач настоящего исследования.

Каждый раздел клинических рекомендаций оценивался экспертами по двум основным критериям:

- содержательная релевантность для задач принятия клинических решений (включение алгоритмов, схем ведения, перечней препаратов и условий их назначения),
- структурная пригодность для прямой обработки языковой моделью без дополнительной интерпретации.

По результатам экспертной оценки был сформирован адаптированный текст клинических рекомендаций, включающий только те разделы, которые обладают высокой клинической значимостью и непосредственно используются при выборе терапии. Из документа были исключены модули справочного, описательного или методологического характера (например, цели и задачи рекомендаций, эпидемиология, методология разработки, правовые основания, информация для пациентов и пр.).

Текст включённых в адаптированную версию разделов не редактировался и не сокращался. Адаптация осуществлялась только путём удаления целых модулей. Полный перечень исключённых и сохранённых разделов приведён в разделе «Результаты».

В работе использованы клинические задачи, сформированные в рамках международного исследовательского проекта врачей - клинических фармакологов России и Кыргызстана в 2024 году. В исследование включены 45 деперсонализированных клинических задач, предоставленных практикующими врачами- клиническими фармакологами, на момент исследования не имевшими существенного опыта взаимодействия с LLM для решения клинических задач. Все кейсы касались пациентов с установленным диагнозом артериальной гипертензии, и были сформулированы в виде краткого описания диагноза и текущей терапии, дополнительных клинических сведений (анамнез, жалобы, осложнения), а также открытых вопросов, предполагающих формулировку рекомендаций по диагностике, тактике лечения и выбору конкретных лекарственных препаратов.

Ввод данных в языковую модель осуществлялся опытным врачом-оператором, координатором исследования, выступавшим посредником между клиницистом и LLM.

Для достижения цели исследования задачи, предоставленные практикующими врачами, были рандомно распределены в 3 группы, отличающиеся различным вариантом взаимодействия с LLM (таблица 1):

Таблица 1. Группы, отличающиеся различным вариантом взаимодействия с LLM.

Группа	Источник знаний	Характеристики
Исследуемая, n=15	Адаптированный текст КР	Удалены нерелевантные задаче разделы (определения, эпидемиология, диагностика, профилактика, реабилитация, методология, информация для пациента). Оставлены только фрагменты по тактике, терапии, алгоритмам.
Контрольная 1, n=15	Без загрузки текста КР	Ответ генерировался исключительно на основе внутреннего знания модели, без доступа к документу.
Контрольная 2, n=15	Неадаптированный текст КР	Использован полный текст КР в исходном виде, без структурной переработки.

Ответы генерировались в модели GPT-4 с использованием авторского системного промта и специально сконфигурированного кастомного профиля GPT, созданного в ходе предварительного тестирования, и показавшего высокую релевантность при решении клинических задач, связанных с выбором тактики лечения и лекарственной терапии.

Диалог с LLM осуществлялся в режиме единого окна, с поэтапной подачей клинического кейса и (в

исследуемой и контрольной 2 группах) текста КР в виде загружаемого документа.

Оценка каждого ответа проводилась тремя независимыми экспертами — врачами-клиническими фармакологами, имеющими клинический и экспертный опыт в отношении лечения пациентов с артериальной гипертонией.

Оценка проводилась по трём отдельным критериям:

Клиническая адекватность (КА) — соответствие логике задачи, обоснованность рекомендаций.

Безопасность рекомендаций (БР) — отсутствие потенциально вредных или некорректных суждений.

Соответствие клиническим рекомендациям (СКР) — степень соответствия ответа LLM соответствующим положениям официального документа.

Оценка проводилась по шкале от 0 до 10 баллов, где 0 обозначал полное отсутствие релевантности или соответствия, а 10 — максимальное качество по соответствующему критерию. Эксперты выставляли исключительно числовые оценки, без дополнительных комментариев.

Статистическая обработка. Для каждой группы рассчитывались средние значения по каждому из трёх критериев оценки, 95% доверительные интервалы, статистическая значимость различий между группами с использованием U-критерия Манна–Уитни.

Выбор U-критерия Манна–Уитни обусловлен тем, что размер выборок был ограничен (15 наблюдений на группу), предварительный анализ распределения оценок (в т.ч. визуальная проверка и оценка асимметрии) не позволил уверенно подтвердить их нормальность. Кроме того, оценки по критериям представлены в порядковой шкале (0–10), не исключающей интервальных искажающих допущений при использовании параметрических методов. В этой связи, для сравнения исследуемой группы с каждой из контрольных групп применялось двустороннее непараметрическое сравнение с использованием U-критерия Манна–Уитни, как более устойчивого к отклонениям от нормальности распределения и выбросам.

III. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Проанализирован документ Минздрава РФ — КР по артериальной гипертензии у взрослых (2024), обладающий нормативным статусом и размещённый в Рубрикаторе клинических рекомендаций [17]. Объём документа составляет 220 страниц; он обладает чёткой рубрикацией и соответствует действующим методическим требованиям к разработке КР.

Основная структура документа включает следующие разделы:

1. Общие положения;
2. Диагностика;
3. Лечение;
4. Реабилитация и профилактика;
5. Организация медицинской помощи;
6. Оценка качества медицинской помощи, список литературы;
7. Приложения.

Документ насыщен таблицами, шкалами, алгоритмами, критериями диагностики и лечения. Используются ссылки на уровни убедительности рекомендаций и достоверности доказательств, что важно для формализации и указания на ориентировочную степень доверия приводимым утверждениям, но усложняет восприятие без указания весов. Присутствуют условные конструкции, модальные глаголы и исключения, требующие логической декомпозиции при адаптации текста для LLM. Отсутствует машиночитаемая разметка или сквозная идентификация пунктов, что делает необходимой ручную структуризацию и алгоритмизацию текста.

Для понимания сложности интерпретации и необходимости предварительной подготовки исходного текста, была выполнена экспертная оценка информационной структуры КР на двух уровнях — макроуровне (базовое деление на модули) и мезоуровне (информационная неоднородность внутри модулей). В Таблице 2 представлены выявленные типы модулей и фрагментов текста, ранжированные по степени ценности для целей генерации медицинских ответов LLM в настоящем исследовании.

Таблица 2. Информационная структура КР: типы модулей и фрагментов, их ценность для LLM

Тип модуля, фрагмента текста	Содержание	Ценность для генерации ответов LLM
Нормативные алгоритмы	Алгоритмы действий, схемы лечения	Высокая
Прямые рекомендации	Конкретные утверждения с дозами, режимами, показаниями	Высокая
Ограничения и предостережения	Условия, при которых не следует применять ЛП	Высокая
Комментарии и пояснения	Контекстуализация рекомендаций, разбор клинических ситуаций	Средняя
Эпидемиология и общая информация	Распространённость, факторы риска	Низкая
Методология разработки	Процедуры создания КР, уровень доказательности	Низкая
Термины и определения	Глоссарии и формулировки	Низкая
Информация для пациентов	Просветительский контент, не предназначенный для врачей	Низкая

Вывод, полученный на этапе анализа: несмотря на внешнюю чёткость структуры, текст КР информационно неоднороден и включает блоки различной функциональной значимости. Без предварительной фильтрации такой источник не может быть использован в LLM-среде без риска снижения релевантности и точности ответов.

Таким образом, клинические рекомендации представляют собой нормативно-нагруженный и структурно насыщенный источник, требующий интеллектуальной адаптации для использования в LLM-среде. В рамках данного исследования работа с текстом документа велась по модульному принципу с фокусом на разделы диагностики и лекарственного лечения.

Поскольку исследование основывалось на 45 реальных клинических задачах, сформулированных в виде практических вопросов («как оценить», «какая тактика», «чем лечить?»), исходный текст КР был целенаправленно модифицирован экспертом-клиническим фармакологом. Результат был одобрен рабочей группой экспертов МОО «Ассоциация клинических фармакологов». Из него были исключены разделы, не имеющие непосредственного отношения к диагностической и терапевтической тактике, включая:

- «Термины и определения»;
- «Краткая информация о заболевании»;
- «Принципы измерения АД»;
- методы лабораторного и инструментального обследования;
- блоки по реабилитации, профилактике, организации помощи, оценке качества медицинской помощи;
- приложения с методологией, информацией для пациента и литературой и т.д.

Оставшийся текст — модули, включающие непосредственно рекомендации по применению лекарственных препаратов, алгоритмы, показания, ограничения и комментарии, объемом 125 страниц (что составило 56,82% от исходного объема), — не подвергался изменениям и использовался как единственный рекомендованный источник знаний для LLM в исследуемой группе.

A. Работа с LLM.

В рамках сравнительного анализа были оценены характеристики медицинских ответов, сгенерированных языковой моделью на клинические запросы, при использовании трёх различных условий доступа к тексту клинических рекомендаций. Сравнение проводилось по трём независимым критериям качества: клиническая адекватность, безопасность и соответствие положениям официальных рекомендаций. Ниже представлены количественные результаты по каждой из групп и критериев с указанием доверительных интервалов и статистической значимости различий.

В исследуемой группе (адаптированный текст КР) средние значения по всем критериям оказались наивысшими: КА — 8,8 ± 0,8, БР — 9,5 ± 0,5, СКР — 9,2 ± 0,6. Эти данные представлены в Таблице 3.

Таблица 3. Оценка качества медицинских ответов, сгенерированных LLM в исследуемой группе (адаптированный текст КР, n=15)

№	Критерий	Среднее значение	Стандартное отклонение
1	Клиническая адекватность (КА)	8.8	0.8
2	Безопасность рекомендаций (БР)	9.5	0.5
3	Соответствие клиническим рекомендациям (СКР)	9.2	0.7
4	Общий балл	27.5	1.7

Результаты демонстрируют стабильную высокую оценку качества медицинских ответов, полученных на основе специально подготовленного текста.

В контрольной группе 1 (отсутствие КР) были зафиксированы существенно более низкие оценки: КА — 6,3 ± 1,2, БР — 7,5 ± 1,1, СКР — 5,1 ± 1,6. Эти значения отражены в Таблице 4.

Таблица 4. Оценка качества медицинских ответов, сгенерированных LLM в контрольной группе 1 (без текста КР, n=15)

№	Критерий	Среднее значение	Стандартное отклонение
1	Клиническая адекватность (КА)	6.3	1.1
2	Безопасность рекомендаций (БР)	7.5	1.2
3	Соответствие клиническим рекомендациям (СКР)	5.1	1.4
4	Общий балл	18.9	2.4

Особенно выражено снижение показателя соответствия клиническим рекомендациям, что закономерно в условиях отсутствия доступа модели к нормативному источнику.

Контрольная группа 2 (полный неадаптированный текст КР) продемонстрировала промежуточные значения: КА — 7,9 ± 1,0, БР — 9,1 ± 0,7, СКР — 7,3 ± 1,2. Результаты приведены в Таблице 5.

Таблица 5. Оценка качества медицинских ответов, сгенерированных LLM в контрольной группе 2 (неадаптированный текст КР, n=15)

№	Критерий	Среднее значение	Стандартное отклонение
1	Клиническая адекватность (КА)	7.7	1.0
2	Безопасность рекомендаций (БР)	9.1	0.6
3	Соответствие клиническим рекомендациям (СКР)	7.3	1.5
4	Общий балл	24.1	2.2

Несмотря на сравнительно высокий уровень безопасности, оценки по критерию СКР уступают исследуемой группе, что может быть связано с трудностями извлечения релевантной информации из полного текста.

Полученные различия по всем критериям представлены на Рисунке 1. Приведены значения клинической адекватности (КА), безопасности (БР) и соответствия клиническим рекомендациям (СКР); отображены 95% доверительные интервалы.

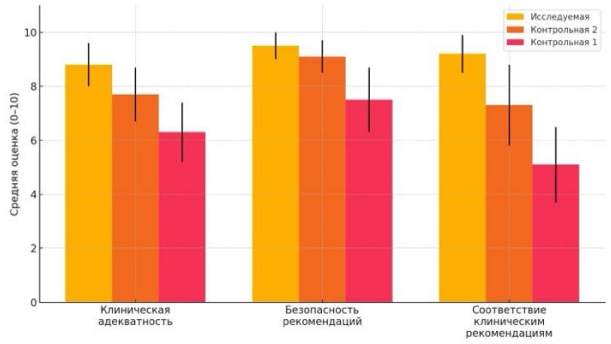

Рисунок 1. Средние значения по трём критериям оценки качества медицинских ответов, сгенерированных LLM, в исследуемой и контрольных группах.

Диаграмма позволяет наглядно оценить превосходство исследуемой группы (с модифицированным документов КР) по всем трём критериям, особенно по показателю соответствия клиническим рекомендациям. Детально эта информация представлена в Таблице 6.

Таблица 6. Средние баллы по каждому из трёх критериев оценки качества медицинских ответов LLM в исследуемой и двух контрольных группах.

Критерий	Исследуемая группа	Контрольная группа 1	Контрольная группа 2
Клиническая адекватность (КА)	8,8 ± 1,1	6,3 ± 1,4	7,8 ± 1,3
Безопасность рекомендаций (БР)	9,5 ± 0,8	7,5 ± 1,2	9,1 ± 1,0
Соответствие КР (СКР)	9,2 ± 0,9	5,1 ± 1,7	7,3 ± 1,5

При сравнении исследуемой группы с контрольной группой 1 и контрольной группой 2 выявлены статистически значимые различия по всем трём критериям ($p < 0,05$), за исключением критерия безопасности между исследуемой группой и контрольной группой 2 ($p = 0,07$).

IV. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

1. Результаты настоящего исследования подтверждают значимое влияние характера представления нормативной информации на качество медицинских ответов, генерируемых большой языковой моделью.

2. На первом этапе обсуждения полученных результатов необходимо обсудить структуру источника знаний, использованного в исследуемой группе, и его соответствие задачам исследования. Экспертная оценка КР «Артериальная гипертензия у взрослых» показала, что значительная часть текста — около 43% — не содержит непосредственно применимых для генерации ответов положений и может быть исключена без ущерба для точности и клинической полноты рекомендаций в рамках рассматриваемых кейсов. Исходный документ включает широкий спектр разделов, начиная от терминологических определений и эпидемиологической справки до организации медицинской помощи, методологии разработки и приложений с информацией для пациентов. Эти модули безусловно важны с нормативной и образовательной точки зрения, однако не имеют прямой связи с решением прикладных клинических задач, ориентированных на выбор терапии, тактику ведения, дифференциальную диагностику и лекарственные назначения. Соответственно, при использовании языковой модели в целях генерации медицинских ответов на практические запросы врачей включение этих разделов создавало бы избыточную нагрузку на модель, снижая интерпретируемость и увеличивая риск отвлечения на нерелевантные фрагменты.

3. Принципиально важно, что оставшиеся 57% документа, отобранные экспертом по модульному принципу, включали только клинически значимую информацию — формализованные рекомендации, алгоритмы действий, перечни показаний и ограничений, лекарственные схемы и комментированные позиции, обладающие нормативной силой. Эти данные в неизменённом виде были переданы языковой модели и обеспечили улучшенные характеристики ответов, что подтверждается количественными результатами, обсуждаемыми ниже.

4. Таким образом, модульная адаптация нормативного текста не только повысила точность и нормативную релевантность генерации, но и позволила продемонстрировать методологическую значимость экспертного анализа структуры КР как предварительного этапа использования языковых моделей в клинической практике.

5. Отметим, что если бы в исследовании рассматривались клинические задачи иного типа — например, маршрутизация пациентов между уровнями медицинской помощи, оценка качества оказанных услуг или формулирование информации для пациента, — то модульная адаптация клинических рекомендаций потребовала бы иного отбора текста. В этом случае целевыми фрагментами стали бы организационные разделы, методология контроля качества, приложения и информационные блоки, а модули, содержащие схемы лекарственного лечения, напротив, могли бы быть исключены как нерелевантные. Это подчёркивает ключевую особенность КР как комплексного нормативного документа, охватывающего широкий спектр задач, от клинической тактики до административных процедур. В рамках нашего исследования целевым направлением была выбрана генерация терапевтически релевантных ответов на клинические запросы врачей, поскольку именно в этом направлении наиболее отчётливо проявляется влияние модульной адаптации текста на релевантность, безопасность и нормативную точность выводов языковой модели. Такой фокус позволил не только добиться высокой воспроизводимости результатов, но и верифицировать исходную гипотезу о значимости интеллектуального сокращения источника знаний.

6. В перспективе, для повышения релевантности ответов, генерируемых языковой моделью, может использоваться более глубокая адаптация исходного текста, включающая не только исключение целых модулей, но и переработку содержания внутри модулей.

Такая модификация может включать упрощение синтаксиса, устранение модальных конструкций, логическую декомпозицию вложенных условий и преобразование текста в машиночитаемую форму, что крайне важно для разработчиков СППВР. Однако подобная работа требует существенных усилий, сопряжена с рисками и должна сопровождаться формальной валидацией полученного укороченного документа на предмет соответствия оригинальным рекомендациям. В рамках настоящего исследования такая задача не ставилась: модульная адаптация ограничивалась только удалением информационно нерелевантных фрагментов. Тем не менее, дальнейшие исследования в этом направлении представляются перспективными, особенно в контексте разработки машиночитаемых нормативных документов и создания структурированных источников знаний для систем поддержки принятия врачебных решений.

7. Проведённая рандомизация 45 клинических кейсов по трем экспериментальным группам позволила оценить, как меняется релевантность, безопасность и нормативная точность в зависимости от формы подачи клинических рекомендаций. Наиболее высокие оценки по всем критериям зафиксированы в группе с модульно адаптированным текстом КР. Исключение определений, эпидемиологических описаний, сведений для пациентов и методологии разработки, при сохранении разделов, содержащих конкретные алгоритмы действий, позволило сосредоточить внимание модели на релевантной клинической информации. Средние оценки 8,8 (КА), 9,5 (БР) и 9,2 (СКР) с узким межквартильным размахом подтверждают стабильность полученного эффекта в условиях разной сложности клинических задач.

8. Вторичная (неадаптированная) форма КР, представленная во второй контрольной группе, оказалась менее эффективной. Несмотря на высокие показатели безопасности (9,1), снижение точности соответствия (СКР 7,3) и адекватности (КА 7,6) указывает на перегрузку модели неструктурированной и избыточной информацией. Оригинальный текста КР, будучи нормативным документом, ориентирован на регуляторное и справочное применение, а не на машинную генерацию клинически релевантных и кратких ответов. В оригинальной редакции он включает множество текстовых конструкций, не предназначенных для непосредственного извлечения конкретных клинических решений языковой моделью. Очевидно, что в отсутствие онтологии, или даже единой логически выверенной структуры и однозначных утверждений, модель (на сегодняшнем уровне ее развития) не в состоянии приоритизировать информацию в зависимости от клинического контекста.

9. Контрольная группа 1, не имевшая доступа к документу, закономерно продемонстрировала минимальные значения всех параметров. Это особенно заметно по критерию соответствия КР (5,1 балла) – такой результат получен несмотря на то, что информация о лечении артериальной гипертензии хорошо представлена в открытых источниках. Таким образом, знание общей медицинской информации без нормативной привязки не позволяет достичь достаточной точности при формировании медицинских ответов для решения реальных клинических задач, представляющих выбор тактики ведения пациента и назначение конкретных лекарственных препаратов, наиболее всего отвечающих задачам лечения пациента.

10. Подобный подход сокращения текста клинических руководств применяется в медицинской литературе для врачей и средних медицинских работников. Сокращённые и структурированные версии, сохраняющие необходимую информацию - именно такой тип документов оказывается наименее конфликтным при внедрении в цифровые инструменты.

11. Модульная адаптация текста позволяет за счет сокращения «ненужных» разделов минимизировать «шум» и повысить фокусировку генерации на клинически значимых положениях. Этот принцип важен для современных LLM, не обладающих истинным пониманием текста. В условиях ограниченного контекста и отсутствия иерархической фильтрации, сокращение и структурирование исходных документов — необходимый этап, повышающий интерпретируемость и управляемость вывода.

12. Следующим этапом на пути решения проблемы может стать интеграция адаптированных текстов в архитектуры retrieval-augmented generation (RAG), обеспечивающие точечное извлечение нормативно релевантных фрагментов с возможностью верификации. Это позволит повысить воспроизводимость решений и доверие со стороны врачебного сообщества, особенно в ситуациях с правовыми или экспертными последствиями. Еще более рациональным путем (не отрицающим при этом использование RAG) явилась бы разработка машиночитаемых клинических рекомендаций к каждому нормативному документу, утверждающему очередную КР. Другим направлением развития исследования должно стать масштабирование на другие нозологии.

13. Таким образом, исследование демонстрирует: не только наличие текста клинической рекомендации, но и его форма, композиция и объём — критические факторы эффективности систем поддержки принятия решений на базе LLM. В условиях цифровой трансформации здравоохранения подход интеллектуальной адаптации источников знаний становится не вспомогательной мерой, а обязательным компонентом инженерного дизайна таких систем.

V.ЗАКЛЮЧЕНИЕ

Проведённое пилотное исследование показало, что модульная адаптация текста клинических рекомендаций значимо повышает релевантность, безопасность и нормативную точность медицинских ответов, формируемых LLM на их основе. Полученные данные подтверждают перспективность данного подхода при проектировании систем поддержки принятия решений на основе искусственного интеллекта в здравоохранении и подчеркивают важность интеллектуальной переработки источников знаний для повышения релевантности ответов LLM, а также

доверия врачей к цифровым инструментам, использующим искусственный интеллект.

БЛАГОДАРНОСТИ

Авторы выражают благодарность Ассоциации разработчиков и пользователей искусственного интеллекта в медицине «Национальная база медицинских знаний» (НБМЗ) (г. Москва, РФ), Межрегиональной общественной организации «Ассоциация клинических фармакологов» (г. Волгоград, РФ) и Общественному объединению "Кыргызская ассоциация клинической фармакологии и доказательной медицины" (г. Бишкек, Республика Кыргызстан).

БИБЛИОГРАФИЯ

[1] Morone, G., De Angelis, L., Martino Cinnera, A., Carbonetti, R., Bisirri, A., Ciancarelli, I., Iosa, M., Negrini, S., Kiekens, C., & Negrini, F. (2025). Artificial intelligence in clinical medicine: a state-of-the-art overview of systematic reviews with methodological recommendations for improved reporting. Frontiers in digital health, 7, 1550731. Available: https://doi.org/10.3389/fdgth.2025.1550731.

[2] Ханов А.М., Гусев А.В., Тюрганов А.Г. Перспективы применения технологий искусственного интеллекта для цифровой трансформации здравоохранения. Российский журнал телемедицины и электронного здравоохранения 2024;10(3):70-76; Available: https://doi.org/10.29188/2712-9217-2024-10-3-70-76.

[3] Yu E, Chu X, Zhang W, Meng X, Yang Y, Ji X, Wu C. Large Language Models in Medicine: Applications, Challenges, and Future Directions. Int J Med Sci. 2025 May 31;22(11):2792-2801. doi: 10.7150/ijms.111780.

[4] Maity, S., & Saikia, M. J. (2025). Large Language Models in Healthcare and Medical Applications: A Review. Bioengineering (Basel, Switzerland), 12(6), 631. Available: https://doi.org/10.3390/bioengineering12060631.

[5] Ваньков В.В., Артемова О.Р., Карпов О.Э., Матвиенко А.В., Гусев А.В., Еникеев И.М., Костина Е.В. Итоги внедрения искусственного интеллекта в здравоохранении России. Врач и информационные технологии. 2024;(3):32-43. Available: https://doi.org/10.25881/18110193_2024_3_32.

[6] Teles AS, Abd-alrazaq A, Heston TF, Damseh R and Ruback L (2025) Editorial: Large Language Models for medical applications. Front. Med. 12:1625293. doi: 10.3389/fmed.2025.1625293.

[7] Li H, Fu J, Python A Implementing Large Language Models in Health Care: Clinician-Focused Review With Interactive Guideline J Med Internet Res 2025;27:e71916 URL: Available: https://www.jmir.org/2025/1/e71916. DOI: 10.2196/71916.

[8] Андреев Д.А., Камынина Н.Н. Перспективы применения информационно-коммуникационной технологии Chat-GPT при организации медицинской помощи пациентам с сахарным диабетом: краткий обзор зарубежной литературы. Врач и информационные технологии. 2024;(2):6-11. Available: https://doi.org/10.25881/18110193_2024_2_6.

[9] Bélisle-Pipon J. C. (2024). Why we need to be careful with LLMs in medicine. Frontiers in medicine, 11, 1495582. Available: https://doi.org/10.3389/fmed.2024.1495582.

[10] Hager, P., Jungmann, F., Holland, R. et al. Evaluation and mitigation of the limitations of large language models in clinical decision-making. Nat Med **30**, 2613–2622 (2024). Available: https://doi.org/10.1038/s41591-024-03097-1.

[11] Zong H, Wu R, Cha J, Wang J, Wu E, Li J, Zhou Y, Zhang C, Feng W, Shen B Large Language Models in Worldwide Medical Exams: Platform Development and Comprehensive Analysis J Med Internet Res 2024;26:e66114. doi: 10.2196/66114.

[12] Wang, L., Chen, X., Deng, X. et al. Prompt engineering in consistency and reliability with the evidence-based guideline for LLMs. npj Digit. Med. **7**, 41 (2024). Available: https://doi.org/10.1038/s41746-024-01029-4.

[13] Zaghir J, Naguib M, Bjelogrlic M, Névéol A, Tannier X, Lovis C. Prompt Engineering Paradigms for Medical Applications: Scoping Review J Med Internet Res 2024;26:e60501. doi: 10.2196/60501.

[14] Satvik Tripathi, Dana Alkhulaifat, Shawn Lyo, Rithvik Sukumaran, Bolin Li, Vedant Acharya, Rafe McBeth, Tessa S. Cook, A Hitchhiker's Guide to Good Prompting Practices for Large Language Models in Radiology, Journal of the American College of Radiology, Volume 22, Issue 7, 2025, Pages 841-847, ISSN 1546-1440, Available: https://doi.org/10.1016/j.jacr.2025.02.051

[15] European Society of Cardiology. - Guidelines. - Clinical Practice Guidelines. - Guidelines Derivative. - Products Pocket Guidelines. Available: https://www.escardio.org/Guidelines/Clinical-Practice-Guidelines/Guidelines-derivative-products/Pocket-Guidelines (дата обращения: 11.07.2025).

[16] Алгоритмы ведения пациента с артериальной гипертензией / сост. Ж. Д. Кобалава, А. О. Конради, С. В. Недогода, Е. А. Троицкая, А. С. Саласюк. — 2-е изд. — Санкт-Петербург: Российское кардиологическое общество, 2024. — 68 с. — Available: https://www.ahleague.ru/images/rekom/Algoritmy_AG.pdf (дата обращения: 11.07.2025).

[17] American College of Obstetricians and Gynecologists. (2019). Clinical guidelines and standardization of practice to improve outcomes (ACOG Committee Opinion No. 792). Obstetrics & Gynecology, 134(4), e122–e125.

[18] ASH Pocket Guides. Available: https://www.hematology.org/education/clinicians/guidelines-and-quality-care/pocket-guides (дата обращения: 11.07.2025).

[19] Joint Royal Colleges Ambulance Liaison Committee, Association of Ambulance Chief Executives (2019) JRCALC Clinical Guidelines 2019. Bridgwater: Class Professional Publishing. ISBN-13: 9781859599020.

[20] Armando LG, Miglio G, Pierluigi de Cosmo, Cena C - Clinical decision support systems to improve drug prescription and therapy optimisation in clinical practice: a scoping review: BMJ Health & Care Informatics 2023;30:e100683. Available: https://doi.org/10.1136/bmjhci-2022-100683

[21] Chen Y., Lehmann C.U., Malin B. Digital Information Ecosystems in Modern Care Coordination and Patient Care Pathways and the Challenges and Opportunities for AI Solutions. - J Med Internet Res 2024;26:e60258. doi: 10.2196/60258.

[22] Martinez-Costa C, Schulz S. HL7 FHIR: Ontological Reinterpretation of Medication Resources. Stud Health Technol Inform. 2017;235:451-455. PMID: 28423833.

[23] Alowais, S.A., Alghamdi, S.S., Alsuhebany, N. et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Med Educ 23, 689 (2023). Available: https://doi.org/10.1186/s12909-023-04698-z.

[24] Alhejaily A. G. (2024). Artificial intelligence in healthcare (Review). Biomedical reports, 22(1), 11. https://doi.org/10.3892/br.2024.1889/

[25] Giebel GD, Raszke P, Nowak H, Palmowski L, Adamzik M, Heinz P, Tokic M, Timmesfeld N, Brunkhorst F, Wasem J, Blase N. Problems and Barriers Related to the Use of AI-Based Clinical Decision Support Systems: Interview Study. J Med Internet Res 2025;27:e63377. doi: 10.2196/63377

[26] Ethics and governance of artificial intelligence for health. Guidance on large multi-modal models. Geneva: World Health Organization; 2024. Licence: CC BY-NC-SA 3.0 IGO.

[27] Sridharan K, Sivaramakrishnan G. Unlocking the potential of advanced large language models in medication review and reconciliation: A proof-of-concept investigation. Explor Res Clin Soc Pharm. 2024 Aug 17;15:100492. doi: 10.1016/j.rcsop.2024.100492.

[28] Han, T., Nebelung, S., Khader, F. et al. Medical large language models are susceptible to targeted misinformation attacks. npj Digit. Med. 7, 288 (2024). Available: https://doi.org/10.1038/s41746-024-01282-7.

[29] Клинические рекомендации «Артериальная гипертензия у взрослых»: ID 62_3 / разработчики: Российское кардиологическое общество и Российское научное медицинское общество терапевтов. – Москва, 2024. – Утверждена 03.10.2024.

Available: https://cr.minzdrav.gov.ru/preview-cr/62_3 (дата обращения: 11.07.2025).

Статья получена 10 августа 2025 года.
Фролов Евгений Максимович, Волгоградский государственный медицинский университет, студент, xolodok_e@mail.ru
Величко Дарья Николаевна, Волгоградский государственный медицинский университет, ординатор, Vel.darya@icloud.com
Мокшин Кирилл Юрьевич , НИЯУ МИФИ, ассистент, KYMokshin@mephi.ru
Шемонаев Дмитрий Дмитриевич, НИЯУ МИФИ, магистрант, dmitryshemonaev@gmail.com
Фролов Максим Юрьевич, Волгоградский государственный медицинский университет, доцент, зав. лабораторией, mufrolov66@gmail.com
Жабицкий Михаил Георгиевич, НИЯУ МИФИ, заместитель директора ВИШ НИЯУ МИФИ, jabitsky@mail.ru

Modular Adaptation of Regulatory Medical Texts to Improve the Relevance of LLM-Based Outputs: A Pilot Study Using Clinical Guidelines for Arterial Hypertension

E.M. Frolov, D.N. Ermolaeva, K.Yu. Mokshin, D.D. Shemonaev, M.Yu. Frolov, M.G. Zhabitsky

***Abstract*—The objective of this study was to evaluate, in a pilot setting, the effect of modular adaptation of clinical guideline texts on the quality of medical responses generated by a large language model (LLM) when solving real-world clinical tasks related to the management of arterial hypertension in adults. A total of 45 clinical scenarios formulated by practicing physicians were randomized into three groups: (1) adapted modular version of the clinical guideline (intervention group), (2) full unadapted guideline text (control group 2), and (3) no document support (control group 1). Responses were generated by GPT-4 using a custom prompt and assessed independently by experts using three separate criteria: clinical adequacy, safety, and compliance with the official recommendations, rated on a 0–10 scale. The adapted guideline group demonstrated the highest mean scores across all criteria (clinical adequacy – 8.8; safety – 9.5; compliance – 9.2), with statistically significant differences confirmed by the Mann–Whitney U test. The results show that excluding structurally irrelevant sections of the guideline (e.g. definitions, epidemiology, methodological notes) significantly improves the relevance and regulatory accuracy of LLM-generated recommendations. This approach may be applicable in the design of AI-based clinical decision support systems and the broader context of digital transformation in healthcare.**

***Keywords* — Large language models, AI in medicine, clinical decision-making, clinical decision support systems, clinical guidelines, arterial hypertension, clinical pharmacology, medical output evaluation, digital transformation in healthcare**

REFERENCES

[1] Morone, G., De Angelis, L., Martino Cinnera, A., Carbonetti, R., Bisirri, A., Ciancarelli, I., Iosa, M., Negrini, S., Kiekens, C., & Negrini, F. (2025). Artificial intelligence in clinical medicine: a state-of-the-art overview of systematic reviews with methodological recommendations for improved reporting. Frontiers in digital health, 7, 1550731. Available: https://doi.org/10.3389/fdgth.2025.1550731.

[2] Khanov, A.M., Gusev, A.V., Tyurganov, A.G. Prospects for the application of artificial intelligence technologies for the digital transformation of healthcare. Russian Journal of Telemedicine and Electronic Health 2024;10(3):70-76; Available: https://doi.org/10.29188/2712-9217-2024-10-3-70-76.

[3] Yu E, Chu X, Zhang W, Meng X, Yang Y, Ji X, Wu C. Large Language Models in Medicine: Applications, Challenges, and Future Directions. Int J Med Sci. 2025 May 31;22(11):2792-2801. doi: 10.7150/ijms.111780.

[4] Maity, S., & Saikia, M. J. (2025). Large Language Models in Healthcare and Medical Applications: A Review. Bioengineering (Basel, Switzerland), 12(6), 631. https://doi.org/10.3390/bioengineering12060631.

[5] Vankov VV, Artemova OR, Karpov OE, Matvienko AV, Gusev AV, Enikeev IM, Kostina EV. Results of the implementation of artificial intelligence in Russian healthcare. Doctor and information technology. 2024;(3):32-43. https://doi.org/10.25881/18110193_2024_3_32.

[6] Teles AS, Abd-alrazaq A, Heston TF, Damseh R and Ruback L (2025) Editorial: Large Language Models for medical applications. Front. Med. 12:1625293. doi: 10.3389/fmed.2025.1625293.

[7] Li H, Fu J, Python A Implementing Large Language Models in Health Care: Clinician-Focused Review With Interactive Guideline J Med Internet Res 2025;27:e71916 Available: https://www.jmir.org/2025/1/e71916. DOI: 10.2196/71916.

[8] Andreev DA, Kamynina NN. Prospects for the use of Chat-GPT information and communication technology in organizing medical care for patients with diabetes mellitus: a brief review of foreign literature. Doctor and information technology. 2024;(2):6-11. Available: https://doi.org/10.25881/18110193_2024_2_6.

[9] Bélisle-Pipon J. C. (2024). Why we need to be careful with LLMs in medicine. Frontiers in medicine, 11, 1495582. Available: https://doi.org/10.3389/fmed.2024.1495582.

[10] Hager, P., Jungmann, F., Holland, R. et al. Evaluation and mitigation of the limitations of large language models in clinical decision-making. Nat Med 30, 2613–2622 (2024). Available: https://doi.org/10.1038/s41591-024-03097-1.

[11] Zong H, Wu R, Cha J, Wang J, Wu E, Li J, Zhou Y, Zhang C, Feng W, Shen B Large Language Models in Worldwide Medical Exams: Platform Development and Comprehensive Analysis J Med Internet Res 2024;26:e66114. doi: 10.2196/66114.

[12] Wang, L., Chen, X., Deng, X. et al. Prompt engineering in consistency and reliability with the evidence-based guideline for LLMs. npj Digit. Med. 7, 41 (2024). https://doi.org/10.1038/s41746-024-01029-4.

[13] Zaghir J, Naguib M, Bjelogrlic M, Névéol A, Tannier X, Lovis C. Prompt Engineering Paradigms for Medical Applications: Scoping Review J Med Internet Res 2024;26:e60501. doi:10.2196/60501.

[14] Satvik Tripathi, Dana Alkhulaifat, Shawn Lyo, Rithvik Sukumaran, Bolin Li, Vedant Acharya, Rafe McBeth, Tessa S. Cook, A Hitchhiker's Guide to Good Prompting Practices for Large Language Models in Radiology, Journal of the American College of Radiology, Volume 22, Issue 7, 2025, Pages 841-847, ISSN 1546-1440, Available: https://doi.org/10.1016/j.jacr.2025.02.051

[15] European Society of Cardiology. - Guidelines. - Clinical Practice Guidelines. - Guidelines Derivative. - Products Pocket Guidelines. Available: https://www.escardio.org/Guidelines/Clinical-Practice-Guidelines/Guidelines-derivative-products/Pocket-Guidelines (date of access: 11.07.2025).

[16] Algorithms for managing a patient with arterial hypertension / compiled by Zh. D. Kobalava, A. O. Konradi, S. V. Nedogoda, E. A. Troitskaya, A. S. Salasyuk. — 2nd ed. — St. Petersburg: Russian Society of Cardiology, 2024. — 68 p. — Electronic resource: https://www.ahleague.ru/images/rekom/Algoritmy_AG.pdf (date of access: 11.07.2025).

[17] American College of Obstetricians and Gynecologists. (2019). Clinical guidelines and standardization of practice to improve outcomes (ACOG Committee Opinion No. 792). Obstetrics & Gynecology, 134(4), e122–e125.

[18] ASH Pocket Guides. – Электронный ресурс: https://www.hematology.org/education/clinicians/guidelines-and-quality-care/pocket-guides (дата обращения: 11.07.2025).

[19] Joint Royal Colleges Ambulance Liaison Committee, Association of Ambulance Chief Executives (2019) JRCALC Clinical Guidelines 2019. Bridgwater: Class Professional Publishing. ISBN-13: 9781859599020.

[20] Armando LG, Miglio G, Pierluigi de Cosmo, Cena C - Clinical decision support systems to improve drug prescription and therapy optimisation in clinical practice: a scoping review: BMJ Health & Care Informatics 2023;30:e100683. https://doi.org/10.1136/bmjhci-2022-100683

[21] Chen Y., Lehmann C.U., Malin B. Digital Information Ecosystems in Modern Care Coordination and Patient Care Pathways and the Challenges and Opportunities for AI Solutions. - J Med Internet Res 2024;26:e60258. doi: 10.2196/60258.

[22] Martinez-Costa C, Schulz S. HL7 FHIR: Ontological Reinterpretation of Medication Resources. Stud Health Technol Inform. 2017;235:451-455. PMID: 28423833.

..

[23] Alowais, S.A., Alghamdi, S.S., Alsuhebany, N. et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Med Educ 23, 689 (2023). https://doi.org/10.1186/s12909-023-04698-z.

[24] Alhejaily A. G. (2024). Artificial intelligence in healthcare (Review). Biomedical reports, 22(1), 11. https://doi.org/10.3892/br.2024.1889/

[25] Giebel GD, Raszke P, Nowak H, Palmowski L, Adamzik M, Heinz P, Tokic M, Timmesfeld N, Brunkhorst F, Wasem J, Blase N. Problems and Barriers Related to the Use of AI-Based Clinical Decision Support Systems: Interview Study. J Med Internet Res 2025;27:e63377. doi: 10.2196/63377

[26] Ethics and governance of artificial intelligence for health. Guidance on large multi-modal models. Geneva: World Health Organization; 2024. Licence: CC BY-NC-SA 3.0 IGO.

[27] Sridharan K, Sivaramakrishnan G. Unlocking the potential of advanced large language models in medication review and reconciliation: A proof-of-concept investigation. Explor Res Clin Soc Pharm. 2024 Aug 17;15:100492. doi: 10.1016/j.rcsop.2024.100492.

[28] Han, T., Nebelung, S., Khader, F. et al. Medical large language models are susceptible to targeted misinformation attacks. npj Digit. Med. 7, 288 (2024). https://doi.org/10.1038/s41746-024-01282-7.

[29] Clinical guidelines "Arterial hypertension in adults": ID 62_3 / developers: Russian Society of Cardiology and Russian Scientific Medical Society of Therapists. - Moscow, 2024. - Approved on 03.10.2024. - Electronic resource: https://cr.minzdrav.gov.ru/preview-cr/62_3 (date of access: 11.07.2025).