

MMCRAg-Resp: мультимодальная корректирующая архитектура RAG для объяснимого анализа респираторных заболеваний

Источник: Frontiers in AI — Medicine

Оригинал: <https://www.frontiersin.org/articles/10.3389/frai.2026.1850832>

LLM

RAG

диагностика

мультимодальные модели

объяснимый ИИ

респираторные заболевания

Введение

Стандартные системы **Retrieval-Augmented Generation** (RAG — генерация с дополненной выборкой), использующие только семантическое сходство для поиска информации, обладают существенными ограничениями. Это может привести к поиску клинически нерелевантных доказательств и генерации небезопасных или галлюцинированных ответов. Данный недостаток особенно критичен в респираторной медицине, где постановка диагноза в значительной степени опирается на высокоточные физиологические показатели, такие как паттерны спирометрии и профили симптомов.

Методы

Мы предлагаем **MMCRAg-Resp**. — клинически обоснованную, учитывающую физиологические особенности корректирующую архитектуру RAG для интеллектуальной поддержки в респираторной медицине. Система гармонизирует 17 516 записей пациентов из трех гетерогенных публичных наборов данных (**Respiratory Sound Database**, спирометрия **NHANES** и

клинические рекомендации) посредством единой респираторной онтологии. Мультимодальные эмбединги (акустические, спирометрические и текстовые) объединяются в 144-мерные векторы и индексируются в **FAISS**.

Новый показатель респираторной релевантности (**Respiratory Relevance Score, RRS**), сочетающий в себе соответствие паттернам спирометрии (вес 0,45), перекрытие симптомов (0,30) и сходство эмбедингов (0,25), выступает в роли фильтра качества доказательств перед их обработкой большой языковой моделью (**LLM**) с помощью трехпутевой политики коррекции. Локально развернутая малая языковая модель (**Ollama; llama3.2, Mistral-7B-Instruct**) генерирует клинические объяснения, ограниченные рамками доказательной базы и полностью прослеживаемые, не требуя при этом использования графических процессоров (GPU).

Результаты

При оценке на 200 отложенных тестовых запросах, извлеченных из гармонизированного корпуса, **MMCRAG-Resp** достигла показателя **Clinical Precision@10** (клиническая точность при топ-10 результатах) 0,81, уровня галлюцинаций 0,11 (что на 71% меньше относительно стандартного **Vanilla RAG**), показателя соответствия спирометрии (**Spirometry Alignment Score**) 0,79 и частоты цитирования доказательств (**Evidence Citation Rate**) 0,87.

Аблационные исследования подтвердили, что приоритизация оценки по данным спирометрии вносит наибольший вклад в общий прирост эффективности. Все попарные сравнения с базовыми моделями были статистически значимыми ($p < 0,001$, тест Макнемара; d Коэна = 1,42).

Заключение

Насколько нам известно, **MMCRAG-Resp** является первой системой, объединяющей мультимодальный поиск респираторных доказательств, физиологически обоснованную корректирующую архитектуру RAG и обеспечивающий конфиденциальность локальный вывод **LLM** в единую развертываемую структуру клинических рассуждений. Данный подход способствует развитию безопасной и объяснимой поддержки принятия решений в респираторной медицине с помощью искусственного интеллекта.

Перевод выполнен: 16.07.2026 | ai4med.ru

Машинный перевод. Рекомендуем сверять с оригиналом при клиническом использовании.