

Мультимодальная медицинская диагностика: мини-обзор гибридных моделей LLM и компьютерного зрения в условиях ограниченных ресурсов здравоохранения

Источник: Frontiers in Digital Health

Дата публикации: 2026-09-04

Оригинал: <https://www.frontiersin.org/articles/10.3389/fdgth.2026.1862236>

NLP

диагностика

мультимодальные модели

низкоресурсное здравоохранение

радиология

Последние достижения в области больших языковых моделей (LLM) и визуальных трансформеров позволили создать мультимодальные системы, интегрирующие клинические тексты с медицинскими изображениями для поддержки диагностических решений. Хотя такие системы демонстрируют многообещающие результаты на эталонных наборах данных в хорошо оснащённых исследовательских условиях, их применимость в медицинских средах с ограниченными ресурсами, где диагностические disparities наиболее выражены, остаётся ограниченной и недостаточно изученной. Данный мини-обзор обобщает ключевые разработки в области архитектур слияния LLM и визуальных данных за период с 2018 по 2026 год, с акцентом на радиологические системы визуального ответа на вопросы (VQA) и генерации отчётов, рассматриваемые с точки зрения развёртывания. Вместо всеобъемлющего каталогизирования мультимодального медицинского ИИ, мы синтезируем эволюцию архитектур слияния LLM и визуальных данных и обсуждаем дополнительные стратегии, способствующие развёртыванию, включая адаптацию с эффективным использованием параметров, пост-обучающее квантование, федеративное обучение и многоязычную поддержку, — в тех случаях, когда они напрямую повышают осуществимость

радиологического ИИ в условиях ограниченных ресурсов здравоохранения. Вместо того чтобы сосредотачиваться исключительно на показателях производительности, мы рассматриваем эти подходы через призму развёртывания, подчёркивая компромиссы между репрезентативной способностью, вычислительной эффективностью, интерпретируемостью и объёмом памяти. Мы утверждаем, что текущий прогресс в значительной степени определяется масштабированием моделей и оптимизацией эталонных показателей, что часто не учитывает ограничения по памяти, подключению и аннотациям, характерные для систем здравоохранения с низким уровнем ресурсов. Хотя кросс-модальные трансформерные архитектуры обеспечивают сильное репрезентативное выравнивание, их вычислительные требования и зависимость от больших курируемых наборов данных ограничивают реальное развёртывание. В отличие от этого, новые направления, включая эффективную тонкую настройку параметров, пост-обучающее квантование, федеративное обучение и модульные агентные системы, предлагают более реализуемые пути к клинической интеграции в условиях аппаратных и данных ограничений. Чтобы преодолеть разрыв между производительностью на эталонных тестах и клинической полезностью, мы выявляем конкретные проблемы, связанные с дефицитом данных, многоязычным охватом и калибровкой, и предлагаем переход к лёгким, интерпретируемым и аппаратно-ориентированным мультимодальным системам ИИ. Эта перспектива подчёркивает необходимость выхода за рамки проектирования, ориентированного на масштабирование, в сторону моделей, способных работать на 4–8 ГБ видеопамати, функционировать офлайн и обобщаться на разные языки и типы медицинского оборудования.