Клинически обоснованная оценка больших языковых моделей в традиционной китайской медицине

Источник: Frontiers in Digital Health

Дата публикации: 2024-05-01

Оригинал: https://www.frontiersin.org/articles/10.3389/fdgth.2026.2009170

LLM | диагностика | клиническая валидация | оценка ИИ | традиционная китайская медицина

Вопросы медицинских лицензионных экзаменов предоставляют доступные и стандартизированные контрольные точки для оценки больших языковых моделей в традиционной китайской медицине. Однако результаты экзаменов сами по себе не могут подтвердить способность моделей поддерживать клиническое мышление или улучшать качество медицинской помощи. В данной статье рассматриваются практические причины ориентации на экзаменационную оценку и четыре потенциальных источника систематической ошибки: репрезентативность клинических знаний, охват профессиональных компетенций, достоверность оценочных данных и совместимость методов оценивания с практикой традиционной китайской медицины. Эти ограничения касаются интерпретации и области применения результатов контрольных тестов, а не ценности самой стандартизированной оценки знаний.

Мы предлагаем клинически обоснованную рамку оценки, которая сочетает динамические клинические сценарии, задачи с открытым ответом (генерацию текста), мультимодальную оценку, лонгитюдное повторное тестирование и клиническую валидацию. Многоэтапные консультации с постепенно раскрываемой информацией могут проверять способность к

рассуждению в условиях неопределенности, тогда как задачи по генерации рецептов и мультимодальные задания позволяют оценить индивидуализированное лечение и интеграцию диагностических данных. Экспертная оценка и автоматизированные метрики должны использоваться для проверки уместности, последовательности и безопасности рекомендаций, за которыми следует оценка в симулированных условиях принятия решений и в рамках взаимодействия человека с искусственным интеллектом.

Новые клинические случаи и повторное тестирование на разных версиях моделей должны служить основой для обновления оценочных задач. Качество рекомендаций, эффективность принятия решений и исходы для пациентов должны измеряться как отдельные конечные точки. Данная рамка помещает результаты экзаменационных баллов в более широкий контекст доказательной базы и делает независимую клиническую валидацию центральным элементом утверждений о клинической полезности моделей.

Перевод выполнен: 03.11.2026 | ai4med.ru

Машинный перевод. Рекомендуем сверять с оригиналом при клиническом использовании.