

R-GCRED: система оценки риска на основе извлечения и калибровки для прогнозирования и рекомендаций по скринингу депрессии

Источник: Frontiers in AI — Medicine

Оригинал: <https://www.frontiersin.org/articles/10.3389/frai.2026.1931579>

LLM

NLP

диагностика

калибровка моделей

объяснимый ИИ

психическое здоровье

скрининг

Раннее выявление депрессии по текстовым нарративам значительно продвинулось благодаря языковым моделям на основе трансформеров и большим языковым моделям (LLM). Однако надежная поддержка принятия решений, ориентированная на скрининг, требует большего, чем просто высокая предиктивная точность. Существующие подходы часто не имеют явного обоснования на уровне симптомов, дают плохо калиброванные оценки уверенности и обеспечивают ограниченную защиту от неопределенности и неподкрепленных моделью сгенерированных доказательств. В данной статье представлена система R-GCRED (Retrieval-Grounded and Calibrated Risk Estimation — оценка риска с опорой на поиск и калибровкой), предназначенная для прозрачной, основанной на доказательствах и учитывающей неопределенность оценки риска депрессии по тексту.

R-GCRED использует поиск, управляемый онтологией, для ограничения извлечения доказательств с помощью LLM клинически релевантными аспектами симптомов, что позволяет получать структурированные и отслеживаемые по источнику представления, фиксирующие идентичность симптома, полярность, интенсивность и поддерживающие текстовые

доказательства. Специализированный механизм проверки доказательств дополнительно снижает распространение неподкреплённых утверждений о симптомах в последующих этапах прогнозирования.

Проверенные доказательства затем обрабатываются архитектурой с двумя головками (dual-head) для совместной классификации риска депрессии и оценки тяжести состояния, за которыми следуют калибровка вероятностей, отказ от предсказаний при наличии неопределённости (uncertainty-aware abstention) и моделирование рекомендаций с учётом чувствительности к стоимости ошибок. Система R-GCRED оценивалась на наборах данных DAIC-WOZ и CLEF eRisk с использованием протоколов оценки с разделением участников (participant-disjoint evaluation) и сравнивалась с трансформерами общего назначения, языковыми моделями, адаптированными для ментального здоровья, методами, основанными на доказательствах, и базовыми моделями на основе LLM.

На всех оцененных наборах данных R-GCDED последовательно превосходит сравниваемые подходы, что указывает на то, что выигрыш в производительности выходит за рамки улучшений, полученных по сравнению с обычными базовыми моделями на основе трансформеров. Система достигает показателя Macro-F1 на уровне 0,84, AUC (площади под ROC-кривой) 0,89 и ожидаемой ошибки калибровки (Expected Calibration Error) 0,038, а также достигает AUC до 0,90 в настройках раннего обнаружения при частичном наблюдении.

Результаты демонстрируют, что интеграция симптомов, основанных на поиске доказательств, проверка доказательств, калиброванное прогнозирование и моделирование принятия решений с учётом неопределённости обеспечивают измеримые улучшения в предиктивной дискриминации, надёжности уверенности и отслеживаемости объяснений по сравнению с оцениваемыми существующими подходами. Эти выводы подтверждают вычислительную эффективность R-GCRED в рамках оцененных настроек бенчмарков, но не представляют собой проспективную клиническую валидацию.

Перевод выполнен: 06.10.2026 | ai4med.ru

Машинный перевод. Рекомендуем сверять с оригиналом при клиническом использовании.