Доказательный ИИ: от первопроходца к создателю доверия?

Источник: Frontiers in AI — Medicine

Оригинал: https://www.frontiersin.org/articles/10.3389/frai.2026.1818128

агентный ИИ | доказательная медицина | методология | надежность ИИ | регуляторная наука | токсикология

Агентные системы ИИ (Agentic AI) способны планировать, вызывать инструменты и координировать специализированных субагентов, что позволяет реализовывать многоэтапные научные рабочие процессы, превосходящие по надежности возможности генерации текста одной моделью. Тем не менее, в таких критически важных областях, как регуляторная наука и токсикология, беглости изложения недостаточно: внедрение технологий зависит от прослеживаемости, воспроизводимости, валидности контекста использования и явного сообщения о степени неопределенности.

В данной статье утверждается, что доказательная медицина и доказательная токсикология обеспечивают зрелый эпистемический каркас для создания доверенных агентных систем ИИ по принципу «доверие по определению» (trustworthy by design). Мы предлагаем «Стек агентных систем на основе доказательств» (Evidence-based Agent Stack), который разлагает сквозные задачи на протоколизированные роли (формулирование вопроса, поиск, скрининг, извлечение данных, оценка риска систематической ошибки, синтез, механистическая/причинно-следственная интеграция, оценка неопределенности и перевод доказательств в решение) с обязательным указанием происхождения данных и версионностью.

Закрепление агентных рабочих процессов в практике систематических обзоров, структурах оценки риска систематической ошибки и развивающихся регуляторных принципах (например, TREAT и e-validation) может превратить «прорывной» (trailblazing) ИИ в «заслуживающий доверия» (trustblazing) ИИ: системы, чьи результаты поддаются аудиту, обновлению и соответствуют принципам подотчетности при принятии решений.

Перевод выполнен: 10.06.2026 | ai4med.ru

Машинный перевод. Рекомендуем сверять с оригиналом при клиническом использовании.