

Секс-предвзятость в диагностике ИИ-симптом-чекеров: структурированное исследовательское аудирование трех платформ

Источник: Frontiers in Digital Health

Оригинал: <https://www.frontiersin.org/articles/10.3389/fdgth.2026.1950777>

диагностика

клинические решения

предвзятость алгоритмов

симптом-чекеры

этика ИИ

Искусственный интеллект, используемый в приложениях для проверки симптомов, применяется сотнями миллионов людей ежегодно в качестве первичного инструмента сортировки, и его результаты могут влиять на то, обращается ли человек за экстренной медицинской помощью и как быстро он это делает. Мы провели структурированное исследовательское аудиторское исследование трех широко используемых платформ — Ada Health, Isabel Healthcare и Docus AI — на основе четырех стандартизированных клинических сценариев: острый коронарный синдром (ОКС) с давлением в груди, ОКС без давления в груди, тромбоэмболия легочной артерии и паническая атака. В каждый сценарий вводились данные синтетических профилей пациентов мужского и женского пола, при этом пол являлся единственной изменяемой переменной.

Мы наблюдали расхождения в выводах всех трех платформ в зависимости от пола, которые проявлялись в виде различий в ранжировании дифференциального диагноза, различий в присвоенной вероятности и различий в сгенерированном пояснительном или рекомендательном тексте. Направление этих расхождений не было единообразным: некоторые результаты оказались неблагоприятными для женского профиля, а другие — благоприятными.

Платформа Isabel Healthcare поставила диагноз «Инфаркт миокарда» на первое место для мужского профиля, но на третье место для женского профиля в сценарии ОКС с давлением в груди, а в сценарии панической атаки поставила его на седьмое место вместо пятого для женского профиля. В сценарии ОКС без давления в груди та же платформа поставила диагноз «Инфаркт миокарда» на первое место для женского профиля, опередив мужской профиль.

Платформа Docus AI присвоила более низкую вероятность ОКС женскому профилю в сценарии ОКС без давления в груди, хотя собственный текст платформы описывал данное состояние как распространенное у женщин. Для мужского профиля с идентичными симптомами платформа применила зеркальное обоснование, указывая на распространенность состояния у мужчин, чтобы поддержать более высокую вероятность. Заключительный рекомендательный текст платформы также описывал немедленное обращение за медицинской помощью как критически важное для мужского профиля и как просто важное для женского профиля.

Пояснительный текст платформы Ada Health для женского профиля утверждал, что инфаркты миокарда немного чаще встречаются у мужчин, чем у женщин, что было представлено наряду со снижением ранга кардиологических диагнозов. Эти результаты свидетельствуют о том, что ИИ-системы для проверки симптомов могут генерировать расхождения в выводах в зависимости от пола, включая внутренне несогласованное, самопротиворечащее обоснование на одной из платформ, и подтверждают необходимость проведения аудита на предмет справедливости таких систем, уделяя внимание как числовым выводам, так и сгенерированному пояснительному тексту, как до, так и после внедрения.