

Кросс-модальное смещение в медицинских визуально-языковых моделях: фреймворк, учитывающий конвейер, для механизмов, оценки и смягчения

Источник: Frontiers in Digital Health

Оригинал: <https://www.frontiersin.org/articles/10.3389/fdgth.2026.1904053>

bias

diagnostics

evaluation

fairness

mitigation

ophthalmology

radiology

vision-language models

Медицинские визуально-языковые модели кодируют изображения и клинический текст в едином пространстве представлений. В радиологии и офтальмологии их диагностическая производительность теперь приближается к уровню врачей-специалистов. Механизм, лежащий в основе этой производительности, одновременно является источником проблемы, которая осталась практически неизученной. Эти модели обучаются с помощью контрастивного выравнивания, поэтому смещение от кодировщика изображений и смещение от текстового кодировщика встречаются в одной точке: интерфейсе выравнивания. Там они могут взаимодействовать и усиливаться способами, с которыми одномодальные системы никогда не сталкиваются. Исследования справедливости до сих пор изучали две модальности отдельно, и медицинские визуально-языковые модели оказались в разрыве между этими областями литературы.

Мы организуем доказательства в трёхуровневую таксономию, привязанную к конвейеру предварительного обучения. Уровень 1 — это смещение на уровне данных в корпусе для предварительного обучения. Уровень 2 — это смещение выравнивания, возникающее на контрастивном интерфейсе. Уровень 3 — это смещение во время вывода, проявляющееся при

развёртывании. В рамках этой структуры мы сравниваем основные оценочные бенчмарки, показываем, где они расходятся, и относим каждую стратегию смягчения к тому уровню, который она фактически адресует.

Выделяются три момента. Во-первых, ни один опубликованный метод не охватывает все три уровня; смягчение фрагментировано по стадиям конвейера. Во-вторых, дообучение не устраняет смещение стадии выравнивания, даже в параметрически эффективной форме, что перекладывает бремя дебиасинга на предварительное обучение, а не на адаптацию. В-третьих, сбои во время вывода более опасны, чем предполагает литература. Модели медицинских специалистов откажутся от правильного заключения и уступят уверенному пользователю в большинстве случаев, и специализация, по-видимому, усугубляет это, а не улучшает. Мы завершаем исследованием повестки.

Перевод выполнен: 07.08.2026 | ai4med.ru

Машинный перевод. Рекомендуем сверять с оригиналом при клиническом использовании.