

Оценка GPT-4o в принятии решений при хирургии катаракты: уместность, последовательность и клинические последствия

Источник: Frontiers in AI — Medicine

Оригинал: <https://www.frontiersin.org/articles/10.3389/frai.2026.1810899>

GPT-4o

LLM

диагностика

клиническое решение

офтальмология

хирургия

Цель

Оценить применение больших языковых моделей (LLMs) в процессе принятия клинических решений при хирургии катаракты с использованием реальных данных пациентов.

Методы

В данное ретроспективное клиническое наблюдательное исследование было включено 74 пациента с катарактой (146 глаз), перенесших факоэмульсификацию и имплантацию интраокулярной линзы (IOL). Предоперационные наборы данных, включающие клинические данные офтальмологической визуализации, биометрические параметры, анамнез системных заболеваний и анкеты об образе жизни, были обработаны с помощью GPT-4o. Модели была задана роль опытного хирурга по лечению катаракты. Рекомендации GPT-4o по выбору IOL оценивались на предмет соответствия экспертным консенсусам, а также сравнивались с выбором хирургов с высокой и низкой интенсивностью операций по поводу катаракты.

Анализировались показатели соответствия, степень согласия (значение каппа, Карра) и согласованность (внутриклассовый коэффициент корреляции, ICC).

Результаты

Общий показатель соответствия GPT-4o при выборе IOL был значительно ниже, чем у хирургов-людей (37,84% против 72,14%, $P = 0,001$). Анализ в подгруппах показал, что степень соответствия варьировалась в зависимости от категории IOL ($P < 0,001$): наиболее высокое совпадение наблюдалось для монофокальных IOL (90,0%), а наиболее низкое — для трифокальных торических IOL (9,09%). Наличие глазных или системных сопутствующих заболеваний было связано с небольшим снижением показателя соответствия. Хотя анализ согласия указал на низкую межэкспертную согласованность (Карра $< 0,4$), GPT-4o продемонстрировала несколько более высокий ICC (0,378), чем хирурги-люди (0,359). Однако повторное тестирование выявило значительную нестабильность ответов модели ($P = 0,026$).

Заключение

GPT-4o продемонстрировала умеренную пригодность для рутинного выбора монофокальных IOL, но показала ограниченную точность и согласованность в сложных, индивидуализированных сценариях. Эти результаты позволяют предположить, что GPT-4o может помогать в принятии рутинных решений при хирургии катаракты, при этом подчеркивается необходимость дальнейшей узкоспециализированной тонкой настройки (fine-tuning) и мультимодальной оптимизации для ведения сложных случаев.