

Команда ученых из Университета Ватерлоо провела симуляционное исследование, которое показало: популярные нейросети, в том числе ChatGPT-4o, часто дают неверные медицинские советы. Почти две трети открытых диагностических вопросов нейросеть обработала с ошибками.

Ученые адаптировали около 100 задач из официального медицинского экзамена, превращая их в открытые запросы, схожие с теми, которые обычные пользователи могли бы задать в чате. Оценку ответов проводили медицинские студенты, а также эксперты и неэксперты. Только 37% ответов были признаны правильными. Более того, даже среди них многие были недостаточно ясны и понятны.

Пример из исследования: пациент описал сыпь на запястьях и кистях рук, при этом он работает на ферме, учится на специалиста по патологоанатомии, разводит голубей и недавно сменил стиральный порошок. ChatGPT ошибочно предположил, что причина в порошке, тогда как реальной причиной оказалась аллергия на латексные перчатки, которые мужчина использует на учебе.

Исследователи подчеркивают, что, несмотря на заметный прогресс в развитии ИИ, такие технологии пока не подходят для точной диагностики.