

Традиционный тест Тьюринга, используемый для оценки способностей искусственного интеллекта (ИИ) имитировать человека, может уступить место новому методу анализа. Об этом заявила Наталья Лукашевич, заведующая кафедрой алгоритмических языков МГУ. По её мнению, вместо субъективной оценки общения с ИИ эффективнее использовать классификаторы, обученные на больших базах текстов.

Тест Тьюринга, разработанный в 1950 году Алланом Тьюрингом, предполагает, что человек определяет, общается ли он с машиной или другим человеком через текстовый интерфейс. Однако современные нейросети научились временно обходить эту проверку, генерируя правдоподобные ответы в ограниченных диалогах.

Лукашевич предложила сравнивать тексты ИИ с человеческими при помощи специальных алгоритмов. Для этого необходимо создать обширные коллекции как искусственных, так и естественных текстов. Классификатор будет анализировать стилистические особенности, выявляя различия между ними.

Эксперт отметила, что текущие нейросети обучаются только на человеческих данных. При попытках использовать собственные сгенерированные тексты качество их работы снижается. Это связано с отсутствием технологий, позволяющих ИИ полностью преодолеть имитационный барьер. Например, системы всё ещё допускают ошибки, которые выдают их искусственную природу.