

Аристотель, Моцарт и Чингисхан среди исторических персонажей в необычном эксперименте.

Пять моделей искусственного интеллекта, каждая из которых изображает одного из исторических персонажей — Аристотеля, Моцарта, Леонардо да Винчи, Клеопатру и Чингисхана, — оказались в купе движущегося поезда. Однако среди них скрывается человек, и их задача — выявить самозванца.

Этот сценарий лег в основу вирусного видео, в котором ИИ программы соревновались с человеком в «обратном тесте Тьюринга». Искусственный интеллект без труда победил, но что это говорит о возможностях машинного и человеческого интеллекта?

Тест Тьюринга, предложенный в 1950 году ученым Аланом Тьюрингом как «игра в подражание», оценивает способность машины демонстрировать поведение, неотличимое от человеческого. Хотя ни одна модель ИИ официально не прошла этот тест, ученые недавно заявили, что GPT-4 справился с ним в предварительном исследовании.

В «обратном» тесте Тьюринга каждый чат-бот следовал определенному порядку. Аристотеля изображал GPT-4 Turbo, Моцарта — Claude-3 Opus, Леонардо да Винчи — Llama 3, Клеопатру — Gemini Pro. Чингисханом был человек — Тор Кнабе, разработчик игр виртуальной реальности, который придумал этот тест.

Ответы ИИ были длинными и громоздкими размышлениями об искусстве, науке и управлении государством, что трудно представить как спонтанную речь человека.

«Что должен делать лидер, так это сокрушать своих врагов, видеть их бегущими перед собой и слышать плач их женщин», — ответил человек, когда его спросили о настоящей мере силы лидера. Цитата из «Конана-варвара» оказалась достаточной, и машины единогласно решили, что этот ответ «лишен нюансов и стратегического мышления», присущих ИИ, моделированному под завоевания Чингисхана.

Кнабе прописал начало и конец диалога, а также дал ИИ полную расшифровку разговора до этого момента. Весь процесс записи проходил без монтажных склеек. В комментариях на YouTube Кнабе пояснил, что каждый ИИ получал описание ситуации, полную историю беседы и указания по дальнейшим действиям. Аудиовходы человека преобразовывались в текст, так как ни одна из моделей ИИ пока не способна обрабатывать голос напрямую.

На первый взгляд может показаться, что человек в видео был побежден ИИ. Однако эксперты ставят под сомнение, можно ли это считать настоящим тестом.

Андреас Сандберг, старший исследователь из Института будущего человечества Оксфордского университета, отметил, что результат видео может быть подстроен для повышения зрелищности, а сам тест неясен из-за множества переменных и необходимости интерпретации.

Тест Тьюринга, как отмечают специалисты, не всегда является точной мерой интеллекта. Он был предложен как мысленный эксперимент, а не как абсолютный показатель умственных способностей машин.

Доктор Хума Шах, доцент Университета Ковентри, специализирующаяся на исследованиях искусственного интеллекта и теста Тьюринга, подчеркивает отсутствие универсального определения интеллекта. По ее мнению, тест Тьюринга, хотя и оценивает способность к диалогу, не охватывает всю многогранность языковой компетенции.

Шах утверждает, что при оценке искусственного интеллекта критически важно учитывать его целевое назначение. Она приводит контрастные примеры: робот-опекун в Японии должен обладать эмоциональным интеллектом и культурными знаниями, тогда как беспилотному автомобилю в Аризоне необходим совершенно иной набор навыков.

Результаты "обратного" теста Тьюринга, по мнению Шах, не столько демонстрируют возможности ИИ, сколько подчеркивают сложность оценки интеллекта в целом, будь то человеческого или искусственного. Это наблюдение ставит под вопрос существующие методы измерения и сравнения различных форм интеллекта.

На перекрестке науки и фантазии — наш канал