

GPT-4 впервые прошел тест Тьюринга

<https://www.livescience.com/technology/artificial-intelligence/gpt-4-has-passed-the-turing-test-researchers-claim>

Исследователи из Института инженеров электротехники и электроники (IEEE) решили провести тест Тьюринга для моделей ИИ, чтобы определить, могут ли люди отличить человека от искусственного интеллекта в разговоре. В их эксперименте 500 участников общались с четырьмя собеседниками: одним человеком и тремя ИИ-моделями — ELIZA 1960-х годов, GPT-3.5 и GPT-4. В результате 54% участников приняли GPT-4 за человека.

Для сравнения, программу ELIZA, в основе которой нет большой языковой модели или нейросети, посчитали человеком только 22% участников.

Тест Тьюринга, впервые предложенный в 1950 году ученым Аланом Тьюрингом оценивает, насколько способность машины проявлять интеллект неотличима от человеческой. Чтобы пройти тест Тьюринга, система должна уметь разговаривать с человеком и заставить его поверить, что он общается с себе подобным.

Для проверки возможностей современных ИИ-систем ученые воссоздали тест Тьюринга. Во время эксперимента 500 человек общались с четырьмя собеседниками: человеком, программой искусственного интеллекта 1960-х годов ELIZA, GPT-3.5, GPT-4. Беседы длились пять минут, после чего участники должны были определить, общались они с человеком или с ИИ.

Участники посчитали GPT-4 человеком в 54% случаев. ELIZA, система, заранее запрограммированная на ответы, но не имеющая большой языковой модели или архитектуры нейронной сети, была признана человеком только в 22% случаев. GPT-3.5 набрал 50%, а человек-участник — 67%.

«Машины, как и люди, могут придумывать правдоподобные обоснования постфактум. Они могут подвергаться когнитивным искажениям, ими можно манипулировать, и они становятся все более обманчивыми. Все эти факторы приводят к тому, что в системах ИИ проявляются человеческие слабости и причуды. Это делает их более похожими на людей, чем предыдущие разработки, которые располагали лишь набором заранее заготовленных ответов», — говорит исследователь искусственного интеллекта в IEEE Нелл Уотсон.

Ученые также утверждают, что существуют обоснованные критические замечания в отношении теста Тьюринга, который является слишком упрощенным. По их словам, «стилистические и социально-эмоциональные факторы играют более важную роль в прохождении теста Тьюринга, чем традиционные представления об интеллекте».

Исследование также подчёркивает, насколько сильно изменился искусственный интеллект в эпоху GPT.

«Ранние системы ИИ, такие как ELIZA, полагались на заранее запрограммированные ответы, что значительно ограничивало их возможности. Программа могла кого-то обмануть на короткое время, но ограничения быстро становились очевидны. Языковые модели эпохи GPT совершенно другие. Они адаптируются и реагируют на широкий круг тем, говорят на определенных языках или диалектах и даже изображают разных личностей с разными ценностями. Это огромный шаг вперед по сравнению с тем, что было тщательно запрограммировано человеком вручную, независимо от того, насколько умной или сложной был тест.