

Технологические гиганты специально отвлекают наше внимание от настоящей угрозы?

Учёный-физик и борец за безопасность искусственного интеллекта, Макс Тегмарк, предупреждает: технологические гиганты отвлекают внимание мира от эзистенциального риска, который ИИ по-прежнему представляет для человечества.

Выступая на Саммите по ИИ в Сеуле, Тегмарк предупредил, что смещение фокуса с угрозы вымирания жизни на общие вопросы безопасности может затянуть разработку ограничений для создателей мощнейших интеллектуальных систем.

Ученый провел прямую параллель с одним из самых важных событий в истории ядерной физики:

«В 1942 году Энрико Ферми построил под футбольным полем в Чикаго первый в мире реактор с самоподдерживающейся цепной ядерной реакцией. Когда ведущие физики того времени узнали об этом, они пришли в настоящий ужас, потому что поняли: только что рухнуло главное препятствие на пути к созданию ядерного оружия. Они осознали, что создание бомбы — вопрос ближайших лет, и действительно, до Тринити оставалось всего три года».

По мнению Тегмарка, модели искусственного интеллекта, способные пройти тест Тьюринга и неотличимые от человека в общении — такой же тревожный сигнал, ведь никто не гарантирует, что однажды мы не утратим контроль над алгоритмами. Именно поэтому в последнее время тревогу начали бить такие пионеры отрасли, как Джейфри Хинтон, Йошуа Бенджио, а также руководители многих технологических корпораций, хоть и не публично.

В прошлом году некоммерческая организация Future of Life («Будущее жизни») под руководством Тегмарка выступила с призывом ввести шестимесячный мораторий на исследования в области ИИ. Когда OpenAI — это компания, которая занимается исследованиями и разработкой в области искусственного интеллекта. Она была основана в 2015 году и создана с целью сделать искусственный интеллект более доступным и безопасным для людей. Компания разрабатывает и использует нейронные сети и другие методы искусственного интеллекта для решения различных задач, включая анализ данных, генерацию текста, голоса, изображений и т.д. OpenAI в марте того же года выпустила сверхмощную модель GPT-4, Тегмарк назвал ее «канарейкой в угольной шахте», доказывающей, что опасность угрожающе близка.

Однако, несмотря на тысячи подписей экспертов под петицией, включая «крестных отцов» современных нейросетей, Хинтона и Бенджио, согласовать введение моратория так и не удалось. Вместо этого саммиты по искусственному интеллекту, вроде нынешней встречи в Сеуле и прошлогодний в Блетчли-Парке в Великобритании, сосредоточились на выработке принципов регулирования этой молодой и бурно развивающейся области.

«Мы хотели, чтобы наше письмо легитимизировало дискуссию о рисках, связанных с ИИ, и рады, что это сработало, — заявил Тегмарк. — Когда люди увидели, что такие маститые специалисты, как Бенджио, серьезно обеспокоены, они поняли, что им тоже можно открыто выражать опасения. Даже обычный работник с автозаправки после этого сказал мне, что боится, как бы искусственный интеллект не вытеснил людей».

«Но теперь нам нужно переходить от разговоров к реальным действиям», — подчеркнул ученый-физик.

С момента анонса саммита в Блетчли-Парке фокус международного научного сообщества по выработке норм регулирования ИИ, тем не менее, сместился в сторону от экзистенциальной угрозы. В Сеуле вопросами безопасности ИИ напрямую занималась лишь одна из трех ключевых дискуссионных групп. Причем они рассматривали широкий круг рисков — от нарушения приватности и утечек данных до потенциального разрушительного влияния на рынок труда.

Тегмарк утверждает, что вытеснение наиболее острых опасений, связанных с ИИ, на обочину дискуссии — отнюдь не случайный, а целенаправленный процесс, виной которому лоббистские усилия технологической индустрии. «Именно это я и предсказывал, — заявил он. — В 1955 году появились первые научные публикации, прямо указывавшие на связь между курением и раком легких. Думаете регулировки были разработаны сразу же после этого? Как бы не так, на это ушло 25 лет. До 1980 года никто не обращал внимания на угрозу, потому что табачная индустрия намеренно отвлекала внимание общественности. Кажется, сейчас происходит то же самое».

«Конечно, искусственный интеллект порождает и текущие проблемы: предвзятость, ущерб для маргинальных групп населения... Но, как справедливо заметила сама министр науки и технологий Великобритании, Мишель Донелан, это вовсе не значит, что мы не можем параллельно заниматься и тем, и другим. Это все равно что сказать: «Давайте не будем обращать внимания на изменение климата, потому что в этом году будет ураган. Сосредоточимся только на урагане».

Критики Тегмарка выдвигают против него тот же аргумент: индустрия, якобы, хочет, чтобы все говорили о гипотетических рисках в будущем, чтобы отвлечь внимание от реальных проблем в настоящем. Однако сам ученый отмечает это обвинение: «Даже если рассматривать ситуацию беспристрастно, в этом утверждении крайне мало смысла. Получается, что кто-то вроде Сэма Альтмана из OpenAI нарочно всех пугает тотальной катастрофой, лишь бы избежать регулирования, а заодно убеждает таких, как мы, громко трубить о грядущей опасности. Слишком сложная многоходовка».

По мнению Тегмарка, сдержанная поддержка со стороны некоторых технологических лидеров объясняется тем, что «они оказались в безвыходном положении: даже если кто-то из них захочет остановиться, у него не получится. Что произойдет, если генеральный директор табачной компании вдруг осознает пагубность своего бизнеса? Его просто заменят. Поэтому единственный путь к безопасному ИИ — если государство установит обязательные для всех стандарты обеспечения безопасности».