

Крупнейшие IT-компании берут на себя обязательства по ответственной разработке технологий.

16 мировых лидеров в области ИИ, включая Google – одна из крупнейших технологических компаний в мире, основанная в 1998 году в США. Основной продукт компании – поисковая система Google, которая позволяет находить информацию в интернете. Компания также разрабатывает множество других продуктов, таких как электронная почта Gmail, видеохостинг YouTube, карты Google Maps и операционную систему Android для мобильных устройств. Google является одним из лидеров в области искусственного интеллекта и облачных вычислений. Компания занимает высокие позиции в рейтингах лучших работодателей в мире. <br>" data-html="true" data-original-title="Google" >Google, Microsoft, IBM и 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." data-html="true" data-original-title="OpenAI" >OpenAI, подписали обязательства деактивировать свои технологии в случае их потенциального опасного воздействия. Событие произошло в Южной Корее на саммите AI Safety Summit.

В рамках саммита были приняты новые обязательства по безопасности передовых технологий ИИ. Компании-участники согласились публиковать свои методики оценки рисков, связанных с их ИИ-моделями, описывать допустимые уровни рисков и принимаемые меры в случае превышения этих уровней. При невозможности снижения рисков ниже установленного порога, компании обязуются не разрабатывать и не внедрять соответствующие модели и системы.

Хотя звучит обязательство звучит многообещающе, детали пока не проработаны. Их предстоит обсуждать на Саммите действий по ИИ, который пройдет в начале 2025 года.

Компании, подписавшие документ в Сеуле, также обязались:

Во время саммита также была принята Сеульская декларация. В документе говорится о важности обеспечения совместимости между системами управления ИИ, основываясь на риск-ориентированном подходе, для максимизации преимуществ и

## Безопасность будущего: мировые лидеры отключат ИИ, если он выйдет из-под контроля

устранения широкого спектра рисков, связанных с ИИ. Это необходимо для безопасного, надежного и заслуживающего доверия проектирования, разработки, развертывания и использования ИИ.

Среди участников сессии были представители правительств стран G7, Сингапура, Австралии, ООН, ОЭСР и ЕС, а также представители индустрии.