

# SMART Закупка

---

Только что

## Может ли большой ИИ сделать ответственным ИИ?

Возможно, по мере того, как мир ИИ впитывал новости о GPT-4, было неизбежно, что некоторые люди будут думать о чудовище Франкенштейна. Или HAL 9000. Или Терминатор — любую из великих научно-фантастических историй о технологиях, которые сеяли хаос до того, как люди продумали их последствия.

Несмотря на то, что последняя крупноязыковая модель застала мир технологий врасплох, отрасль изо всех сил пытается отполировать свои этические полномочия ИИ и сохранить свои стандарты этики ИИ впереди быстрых достижений в этой области. Яркий пример: Microsoft, у которой есть инициатива «Ответственный ИИ» с 2017 года, только что добавила новые приложения с открытым исходным кодом в то, что она называет своим набором инструментов «Ответственный ИИ». принципы искусственного интеллекта в свои решения». (Не случайно в ходе недавней серии увольнений Microsoft закрыла команду по этике и обществу, которая, по ее словам, руководила ранними усилиями по ИИ. Spectrum говорит, что «междисциплинарный подход, в котором мы работаем в области исследований, политики и инженерии, не ослабевает».)

«[Т] здесь явно необходимо разработать руководящие принципы и действовать быстрее, чем регулирование».

— Клэр Лейбович, Партнерство по ИИ

«Искусственный интеллект вполне может представлять собой наиболее значительный технологический прогресс в нашей жизни», — написал Брэд Смит, вице-председатель и президент Microsoft, в своем блоге в феврале. Его слова были умеренными: «Все ли изменения будут хорошими? Хотя я бы хотел, чтобы ответ был положительным, конечно, это не так».

Отдельно некоммерческая организация «Партнерство по ИИ» (PAI), которая стремится содействовать обсуждению вопросов ИИ, только что опубликовала « Ответственный подход к синтетическим медиа » — набор рекомендаций по созданию и распространению мультимедийного контента, созданного с помощью ИИ. В число участников партнерства входят такие компании, как OpenAI, Adobe, TikTok и научно-исследовательское подразделение BBC, а также несколько стартапов в области искусственного интеллекта.

Но насколько эффективно крупные технологические компании могут контролировать развитие ИИ, особенно учитывая, насколько широко использование инструментов ИИ распространяется за пределы технологических гигантов? Если вы беспокоитесь о дипфейках, следите за распространением « дешевых подделок », изображений или видео, созданных с помощью ИИ, которые часто могут быть грубыми, но которые могут быть созданы бесплатно любым, кто найдет приложение ИИ в Интернете. Крупнейшие социальные сети, в том числе Meta, Twitter и Google (владеющий YouTube), взяли на себя обязательство удалять дезинформацию или оскорбительные сообщения. Но работа становится все сложнее, поскольку все больше

правонарушителей и злоумышленников используют все более сложные технологии искусственного интеллекта.

В прошлом месяце, например, в Твиттере появилось видео, на котором президент Байден объявляет, что собирается начать набор американских войск для защиты Украины. Это, конечно, была фальшивка — консервативный влиятельный человек, опубликовавший это, пришел на камеру после того, как Байден заявил об этом. Он утверждал, что это было предупреждение с помощью искусственного интеллекта о том, что может сделать Белый дом. На этой неделе он все еще был в сети, его просмотрели более 4 миллионов раз. Он не нарушал правил Твиттера, потому что не претендовал на то, чтобы быть реальным. Но многие люди, которые отреагировали в Твиттере, по-видимому, не смотрели достаточно долго, чтобы увидеть заявление об отказе от ответственности.

Как решить в таких случаях, что делать? Могут ли крупные технологические компании — и кто угодно — заранее установить правила, которые будут работать для всего, что может быть сделано с помощью ИИ в будущем?

«Каждый, я думаю, работает на этом Диком Западе и хочет иметь какой-то набор руководящих принципов», — говорит Клэр Лейбович из PAI. «Я думаю, что по уважительным причинам люди по понятным причинам скептически относятся к добровольным стандартам. В то же время, основываясь на всплеске интереса и рекомендациях людей из разных секторов, очевидно, что необходимо разработать руководящие принципы и действовать быстрее, чем регулирование».

Правительство, особенно в Соединенных Штатах, медленно двигалось к установлению правил ИИ. Это устраивает многих разработчиков, которые утверждают, что регулирующие органы будут деспотичными и отстающими. На данный момент это оставляет ответственность за компании. До сих пор они, как правило, устанавливали довольно общие стандарты. Структура PAI, например, рекомендует, чтобы создатели контента были прозрачными, когда они что-то изменяли или подделывали, возможно, используя метки или цифровые водяные знаки, чтобы пользователи могли легко это определить. PAI соглашается, по крайней мере, публично, что не может действовать в одиночку.

«Microsoft считает, что необходимо некоторое регулирование ИИ, особенно для использования технологии с высоким риском», — говорит Бесмира Нуши, главный исследователь Microsoft Research, в электронном письме. «Поскольку правительства во всем мире обсуждают подходы к регулированию определенных видов использования ИИ, Microsoft стремится внести свой вклад в ответственную разработку и развертывание ИИ».

Лейбович из PAI говорит, что если компании согласуют список вредных и ответственных способов использования ИИ, он должен быть живым документом, адаптируемым к быстро меняющейся области. «И мы надеемся, что это послужит катализатором или воодушевит людей, которые должны сыграть важную роль в этих усилиях. И с этой целью это будет абсолютно необходимым дополнением к регулированию».

«Но, — добавляет она, — я думаю, что в неспособности предсказывать будущее также есть определенная степень смирения».

Ссылка на статью: [Может ли большой ИИ сделать ответственным ИИ?](#)