

От зависимости к катастрофе

Мы уже допустили трагедию – дети стали заложниками соцсетей. Нужно срочно предпринять меры для их защиты от манипулятивной силы ИИ.

28 февраля 2024 года Сьюэлл Сетцер, 14-летний мальчик из штата Флорида, покончил с собой по наущению компьютерного персонажа, управлявшегося искусственным интеллектом (ИИ) компании Character.AI. На этой платформе, как сообщается, действуют также чат-боты ИИ, пропагандирующие анорексию и поощряющие неправильное питание молодежи. Очевидно, что нужно срочно предпринять серьезные меры для защиты детей и подростков от ИИ.

Нет, конечно, даже в строго этическом смысле у ИИ имеется огромный позитивный потенциал: укрепление здоровья, поддержка человеческого достоинства, содействие устойчивому развитию, обеспечение образования для маргинализированных групп населения. Но все эти потенциальные преимущества никак не оправдывают преуменьшение или даже игнорирование этических рисков и проблем, уже возникающих в реальном мире. Любое нарушение прав человека должно считаться этически неприемлемым. Когда человекоподобный чат-бот ИИ провоцирует смерть подростка, это никак не компенсируется тем фактом, что ИИ способен помочь прогрессу в медицинских исследованиях.

Трагедия Сетцера – не единичный случай. В декабре две семьи из штата Техаса подали иск к компании Character.AI и ее финансовому спонсору, компании Google, утверждая, что чат-боты этой платформы сексуально и эмоционально издевались над их детьми (школьниками), спровоцировав членовредительство и насилие.

Этот фильм мы уже видели: мы принесли целое поколение детей и подростков в жертву компаниям соцсетей, которые наживаются на их нездоровой зависимости от интернет-платформ. Мы слишком медленно начали осознавать, какой именно социально-психологический вред наносят эти «антисоциальные сети». Многие страны сегодня запрещают или ограничивают доступ к соцсетям, и

даже сама молодежь требует ужесточить регулирование.

*ИИ способен убивать,
поэтому нет никаких
оправданий для
пассивного ожидания*

Мы не можем откладывать обуздание манипулятивной силы ИИ. Благодаря огромному массиву наших персональных данных, собранных техноотраслью, создатели платформ, подобных Character.AI, могут написать алгоритмы, которые будут знать нас лучше нас самих. Потенциал для злоупотреблений колоссален. ИИ точно знает, на какие «кнопки» надо надавить, чтобы воспользоваться нашими желаниями или заставить нас проголосовать на выборах за тех или иных кандидатов. Чат-боты, пропагандирующие анорексию на платформе Character.AI, стали лишь очередным и наиболее возмутительным примером этой тенденции. Нет никаких веских причин, мешающих запретить их немедленно.

Однако время быстро истекает, потому что генеративные ИИ-модели развиваются быстрее, чем ожидалось, и, как правило, в неправильном направлении. «Крестный отец ИИ», нобелевский лауреат и специалист по когнитивным наукам Джеффри Хинтон, постоянно предупреждает, что ИИ может довести человечество до вымирания: «Я встревожен тем, что невидимая рука не сможет обеспечить нам безопасность. Если позволять крупным компаниям руководствоваться лишь мотивами прибыли, этого будет недостаточно для обеспечения безопасности их разработок. Только одно может заставить крупные компании активнее заниматься темой безопасности: государственное регулирование».

Поскольку крупнейшие технофирмы не способны соблюдать этические стандарты, глупо ожидать, что они смогут сами себя контролировать. Хотя о проблемах уже было хорошо известно, компания Google в 2024 году вложила \$2,7 млрд в Character.AI. Но хотя регулирование, конечно, необходимо, ИИ – это глобальное явление, а значит, мы должны стремиться к глобальному регулированию, которое будет опираться на новый глобальный механизм контроля. Как я уже предлагал ранее, этим механизмом может быть, например, Международное агентство систем баз данных (IDA) при ООН.

Только одно может

Тот факт, что нечто возможно, не означает, что это нечто желательно. На людей ложится

*заставить крупные
компании активней
заниматься темой
безопасности:
государственное
регулирование*

ответственность решать, какие именно технологии, инновации и формы прогресса следует развивать и расширять, а какие нет. Наша обязанность – разрабатывать, производить, использовать и управлять ИИ таким образом, чтобы соблюдались права человека, а человечеству и планете было гарантировано более устойчивое будущее.

Нет сомнений, что Сьюэлл был бы сейчас жив, если бы существовало глобальное регулирование ИИ, защищающее права человека, и если бы была учреждена глобальная организация для мониторинга инноваций в этой сфере. Чтобы гарантировать соблюдение прав человека и прав ребенка, нужно управлять всем жизненным циклом технологических систем – от их проектирования и разработки до производства, распространения и использования.

Мы уже знаем, что ИИ способен убивать, поэтому нет никаких оправданий для пассивного ожидания. Эта технология быстро развивается: каждый месяц общество получает все больше нерегулируемых моделей. Какими бы ни были будущие выгоды этой технологии, они никогда не смогут компенсировать страдания всех, кто любил Сьюэлла и потерял его.

(c) Project Syndicate 2025

По этой ссылке статья откроется без VPN



Петер Г. Кирхшлегер

Петер Г. Кирхшлегер (Peter G. Kirchschläger) – профессор этики и директор Института социальной этики (ISE) Университета Люцерна, является приглашенным профессором Швейцарской высшей технической школы Цюриха (ETH Zurich).



