

Від залежності до катастрофи

Ми допустили трагедію – діти стали заручниками соцмереж. Потрібно терміново вжити заходів для їхнього захисту від маніпулятивної сили ШІ.

28 лютого 2024 року Сьюелл Сетцер, 14-річний хлопчик зі штату Флорида, наклав на себе руки після підбурювання комп'ютерного персонажа, керованого штучним інтелектом (ШІ) компанії Character.AI. На цій платформі, як повідомляється, діють також чат-боти ШІ, які пропагують анорексію і заохочують неправильне харчування молоді. Очевидно, що потрібно терміново вжити серйозних заходів для захисту дітей і підлітків від ШІ.

Ні, звичайно, навіть у суто етичному сенсі ШІ має величезний позитивний потенціал: зміцнення здоров'я, підтримка людської гідності, сприяння сталому розвитку, забезпечення освіти для маргіналізованих груп населення. Але всі ці потенційні переваги ніяк не виправдовують применшення або навіть ігнорування етичних ризиків і проблем, що вже виникають у реальному світі. Будь-яке порушення прав людини має вважатися етично неприйнятним. Коли людиноподібний чат-бот ШІ провокує смерть підлітка, це ніяк не компенсується тим фактом, що ШІ здатен допомогти прогресу в медичних дослідженнях.

Трагедія Сетцера – не поодинокий випадок. У грудні дві сім'ї зі штату Техас подали позов до компанії Character.AI та її фінансового спонсора, компанії Google, стверджуючи, що чат-боти цієї платформи сексуально та емоційно знущалися над їхніми дітьми (школярами), спровокувавши каліцтво та насильство.

Цей фільм ми вже бачили: ми принесли ціле покоління дітей і підлітків в жертву компаніям соцмереж, які наживаються на їхній нездоровій залежності від інтернет-платформ. Ми занадто повільно почали усвідомлювати, яку саме соціально-психологічну шкоду завдають ці «антисоціальні мережі». Багато країн сьогодні забороняють або обмежують доступ до соцмереж, і навіть сама молодь вимагає посилити регулювання.

*ШІ здатен вбивати,
тому немає ніяких
виправдань для
пасивного очікування*

Ми не можемо відкладати приборкання маніпулятивної сили ШІ. Завдяки величезному масиву наших персональних даних, зібраних техногалуззю, творці платформ, подібних до Character.AI, можуть написати алгоритми, які будуть знати нас краще за нас самих. Потенціал для зловживань колосальний. ШІ точно знає, на які «кнопки» треба натиснути, щоб скористатися нашими бажаннями або змусити нас проголосувати на виборах за тих чи інших кандидатів. Чат-боти, що пропагують анорексію на платформі Character.AI, стали лише черговим і найбільш обурливим прикладом цієї тенденції. Немає жодних вагомих причин, які заважають заборонити їх негайно.

Однак час швидко спливає, тому що генеративні моделі ШІ розвиваються швидше, ніж очікувалося, і, як правило, в неправильному напрямку. «Хрещений батько ШІ», нобелівський лауреат і фахівець з когнітивних наук Джеффри Хінтон, постійно попереджає, що ШІ може довести людство до вимирання: «Я стурбований тим, що невидима рука не зможе забезпечити нам безпеку. Якщо дозволяти великим компаніям керуватися лише мотивами прибутку, цього буде недостатньо для забезпечення безпеки їхніх розробок. Тільки одне може змусити великі компанії активніше займатися темою безпеки: державне регулювання».

Оскільки найбільші технофірми не здатні дотримуватися етичних стандартів, нерозумно очікувати, що вони зможуть самі себе контролювати. Незважаючи на те, що про проблеми вже було добре відомо, компанія Google в 2024 році вклала \$2,7 млрд в Character.AI. І хоча регулювання, звичайно, необхідне, ШІ – це глобальне явище, а значить, ми повинні прагнути до глобального регулювання, яке буде спиратися на новий глобальний механізм контролю. Як я вже пропонував раніше, цим механізмом може бути, наприклад, Міжнародне агентство систем баз даних (IDA) при ООН.

*Тільки одне може
змусити великі компанії
активніше займатися
темою безпеки: державне
регулювання*

Той факт, що щось є можливим, не означає, що це щось є бажаним. На людей покладається відповідальність вирішувати, які саме технології, інновації та форми прогресу слід розвивати і розширювати, а які ні. Наш обов'язок – розробляти, виробляти,

використовувати і керувати ШІ таким чином, щоб дотримувалися права людини, а людству і планеті було гарантовано більш стає майбутнє.

Немає сумнівів, що Сьюелл був би зараз живий, якби існувало глобальне регулювання ШІ, що захищає права людини, і якби була створена глобальна організація для моніторингу інновацій у цій сфері. Щоб гарантувати дотримання прав людини і прав дитини, потрібно керувати всім життєвим циклом технологічних систем – від їхнього проектування і розробки до виробництва, поширення і використання.

Ми вже знаємо, що ШІ здатен вбивати, тому немає жодних виправдань для пасивного очікування. Ця технологія швидко розвивається: щомісяця суспільство отримує дедалі більше нерегульованих моделей. Якими б не були майбутні вигоди цієї технології, вони ніколи не зможуть компенсувати страждання всіх, хто любив Сьюелла і втратив його.

(c) Project Syndicate 2025



Петер Г. Кірхшлегер

Петер Г. Кірхшлегер (Peter G. Kirchschläger) – професор етики та директор Інституту соціальної етики (ISE) Університету Люцерна, є запрошеним професором Швейцарської вищої технічної школи Цюриха (ETH Zurich).



