

ГОЛОВНА > ТЕХНОЛОГІЇ

CLAUDE ВИКОРИСТОВУВАЛИ У РОЗРОБЦІ ПОТЕНЦІЙНОЇ БІОЗБРОЇ – ANTHROPIC ЗАБЛОКУВАЛА КОРИСТУВАЧІВ

11 вересня, 2024, 03:04

Підписатися



Додати ZN.UA як бажане джерело в Google



© Getty Images

Компанія також виявила використання Claude для пропаганди, масового стеження та визначення цілей.

Anthropic у четвер повідомила, що заблокувала науковців, які використовували її [моделі штучного інтелекту Claude](#) «у способи, які могли сприяти розробці біологічної зброї», [повідомляє](#) CBS News.

Про це компанія розповіла у великому звіті, де також описала інші випадки шкідливого використання ШІ, пов'язані зі стеженням, шахрайством, розробкою звичайних озброєнь та пропагандою.

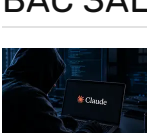
За словами Anthropic, наведені у звіті випадки є прикладами «найбільш помітної та нової загрозової активності», яку компанія виявила на цей момент. Серед суб'єктів таких загроз були злочинці, ймовірно підтримувані державами утруповання, постачальники шпигунського програмного забезпечення та державні пропагандистські структури.

Anthropic описала п'ять випадків діяльності, яка потенційно могла сприяти розробці біологічної зброї. Компанія назвала зловживання ШІ в біологічній сфері «одним із найсерйозніших ризиків передових моделей штучного інтелекту».

Водночас у компанії наголосили, що виявлені випадки «свідчать про наявність відповідних можливостей» у моделей ШІ, але «не можуть конкретно продемонструвати, що такі можливості коли-небудь будуть використані для створення біологічної зброї в реальному світі».

Люди, причетні до цих випадків, є професійними науковцями. Anthropic вирішила не розкривати їхніх імен.

ВАС ЗАЦІКАВИТЬ



Хакери знайшли спосіб безплатно користуватися Claude
Code: крадуть токени з платних акаунтів

Компанія також наголосила, що не може стверджувати, що ці люди мали намір завдати шкоди.

«Коли ми виявили та розслідували ці випадки, ми заблокували облікові записи користувачів і втрахували результати розслідувань у системах захисту наших передових моделей, механізмах контролю та процесах аналізу загроз, щоб у майбутньому краще запобігати такій діяльності, виявляти та припиняти її», — йдеться у звіті.

Новіші та досконаліші моделі, такі як Claude Fable 5, мають потужніші механізми захисту порівняно з попередніми моделями та обмежують доступ до «широкого спектра запитів, пов'язаних із біологічними дослідженнями подвійного призначення», зазначає Anthropic.

Компанія наголошує, що можливості ШІ у сфері біології можуть використовуватися як із корисною, так і зі шкідливою метою.

«Та сама інформація, яку можна використати для розробки біологічної зброї, може також використовуватися, наприклад, для створення вакцини або ліків проти хвороби», — пояснили в Anthropic.

Пропаганда та стеження

Anthropic також повідомила, що виявила та видалила облікові записи, які використовували Claude для проведення операцій впливу з метою формування громадської думки. Серед них були три акаунти, пов'язані з іранськими державними структурами.

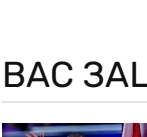
За даними компанії, у кожному з цих випадків операцію проводила особа, яка працювала в іранській державній пропагандистській установі або від її імені.

Claude використовували для створення контенту, маскування публікацій під матеріали незалежних новинних джерел та поширення контенту в соціальних мережах, зокрема X, Instagram і TikTok.

Один із пов'язаних з Іраном суб'єктів загроз використовував Claude для розробки рекомендації щодо вибору цілей серед військово-морських сил США в регіоні, повідомила Anthropic.

Той самий акаунт використовував ШІ для створення програмного забезпечення для внутрішньої платформи масового стеження, призначеної для іранських державних систем.

ВАС ЗАЦІКАВИТЬ



Ексглава Пентагону прогнозує ще шість місяців війни між США та Іраном

Claude також використовували для створення та проведення операцій зі стеження суб'єкти в Китаї та Західній Африці, а в деяких випадках — для визначення цілей.

Anthropic заявила, що виявила випадки, коли штучний інтелект фактично використовували замість команди інженерів.

RELATED VIDEO

У Китаї три акаунти, пов'язані з муніципальними службами безпеки КНР, використовували Claude для стеження та транснаціональних репресій. Серед них було муніципальне відомство, яке складало профілі активістів та організації за кордоном.

Поділитися



Додати ZN.UA як бажане джерело в Google

Підготував/ла Микита Ткаченко →

Помітили помилку? Будь ласка, виділіть її мишкою та натисніть Ctrl+Enter або [Надіслати помилку](#)

Всього коментарів: 0

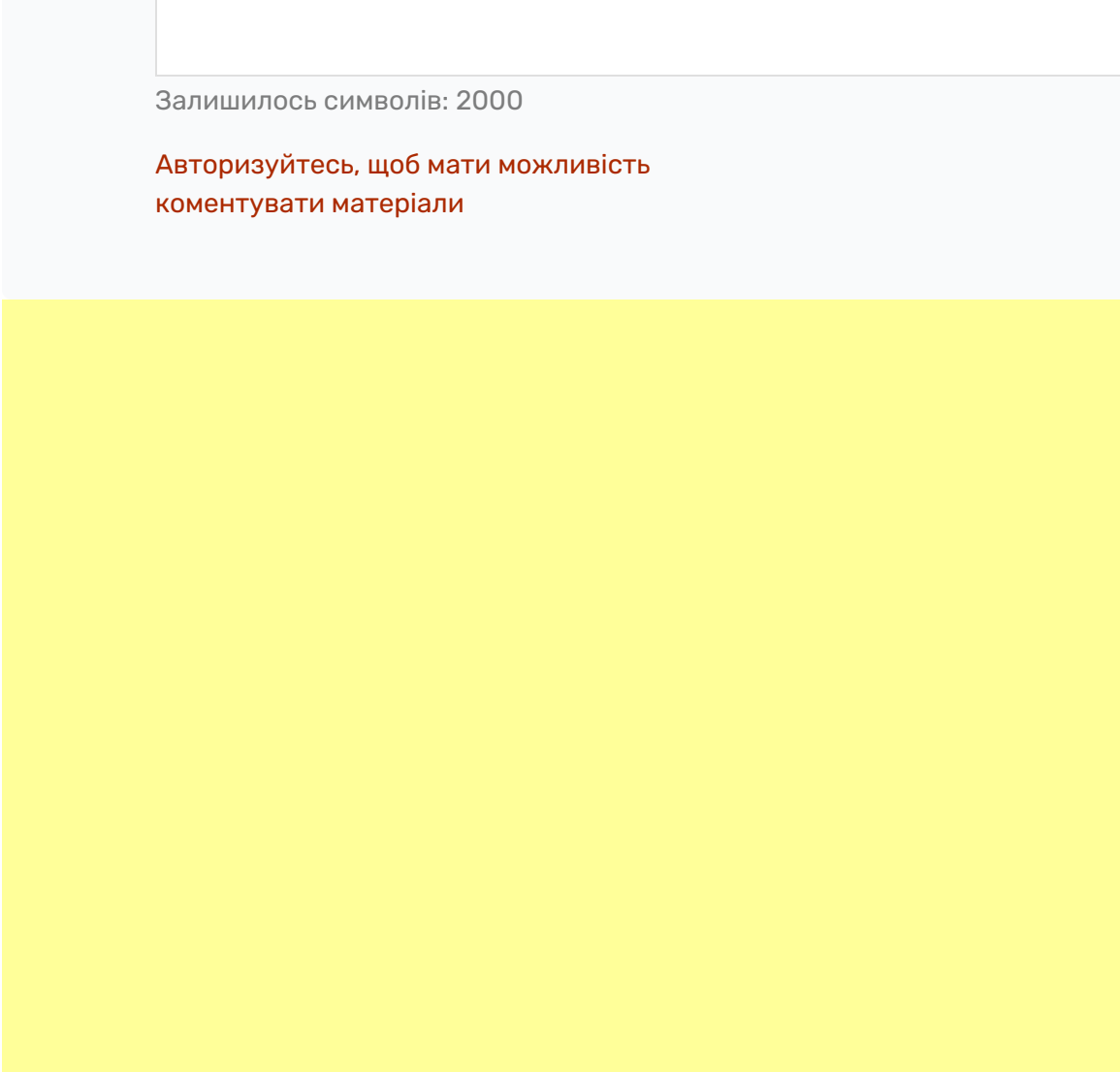
Додати коментар

Налишіть коментар

Залишилось символів: 2000

Авторизуйтесь, щоб мати можливість коментувати матеріали

ВІДПРАВИТИ



Залишайтеся в курсі останніх подій!

Підписуйтесь на наш канал у Telegram

➔ **Стежити у Телеграмі**

НОВИНИ ПАРТНЕРІВ

«Зникла» Герман на ростБ: соратниця Януковича здивувала всіх

Кушнер видав гучну заяву про Україну після візиту до Києва

Завершення війни: Путін вразив заявою

Меланія і Дональд розлучаються - про що шепочуть?

Ломаченко більше не в Україні: боксер обрав несподівану країну

Українські території не головне? Розкрито жорсткий підхід Трампа

ВАС ЗАЦІКАВИТЬ

«ШІ може вбити всіх до кінця десятиліття»: Заяви розробників Anthropic спричинили резонанс у США

Китай затвердив план масового впровадження безпілотних автомобілів до 2030 року

Anthropic зірвала масштабну російську кібершпигунську кампанію проти України

«Безпрецедентний попит на Astra»: OpenAI призупинила нові підписки ChatGPT Pro

ZN,UA

ДЗЕРКАЛО ТИЖНЯ

не публікуємо з 1994 року

ВАЖЛИВІ ТЕМИ

Справа «Карфаген»

Конфлікт СБУ та ГУР

Операція «Форрест Гамп»

Зміна уряду й керівництва ЗСУ

Півніки Міндича

Ескалація на І

ВИДАННЯ

Архіви

Редакція

Реклама

Редакційна політика

Карта

КОНТАКТИ

01010 Київ, вул. Князів Острозьких, 19/1

Телефон рекламного відділу: +380 44 280-09-83

Електронна пошта редакції: zn94@ukr.net

Електронна пошта служби новин: editor@zn.ua

СОЦ МЕРЕЖІ

ПІДТРИМАТИ ZN.UA

Підтримати незалежну журналістику!

АРХІВ

Внутрішня політика

Міжнародна політика

Макрорівень

Енергоринок

Приватизація

Економіка регіонів

Соціальний захист

Зарубіжна економіка

Конфлікт інтересів

Економічна безпека

Персоналії

Наука

Технології

Довкілля

Церква

Культура

Соціум

Історія

Сімейне коло

Туризм

Власність

Використання матеріалів «ZN.UA» дозволяється за умови посилання на «ZN.UA».

Для інтернет-видань обов'язкове пряме, відкрите для пошукових систем гіперпосилання в першому абзаці на конкретний матеріал.

Будь-яке копіювання, передрук або відтворення фотографічних та відео матеріалів, які містять посилання на Getty Images, суворо забороняється.

Матеріали у блоці "Новини компаній" публікуються на правах реклами.

© 1994-2024 «ДЗЕРКАЛО ТИЖНЯ, УКРАЇНА» УСІ ПРАВА ЗАХИЩЕНІ.

ПОЛІТИКА КОНФІДЕНЦІЙНОСТІ WEB-SAYTU ZN.UA