



Топ-теми:

- Арешт Ермака
- Корупційна мафія Міндіча в Енергоатомі
- Напад Росії на Україну



НОВИНИ	КОНФЛІКТИ	ПОЗИЦІЯ	ДУМКА	ПОДІЇ	ВІЙНА	ВІДЕО	БЛОГИ
--------	-----------	---------	-------	-------	-------	-------	-------

Головна > Статті > Новини > Новини світу > Військові підключили чат-бот зі штучним інтелектом до симулятора війни: результати виправдали найбільші страхи

Військові підключили чат-бот зі штучним інтелектом до симулятора війни: результати виправдали найбільші страхи

[Читати на руском](#)



Військові підключили чат-бот зі штучним інтелектом до симулятора війни: результати виправдали найбільші страхи

Штучний інтелект під час симуляцій військової гри неодноразово обирав найгірший з можливих варіантів розвитку конфлікту аж до використання ядерної зброї. Сама наявність такої смертоносної зброї схилила ШІ до аргументу на користь її застосування.

Про це йдеться у дослідженні вчених з Стенфордського університету (США), яке [опубліковано](#) на сайті препринтів arXiv. Метою дослідження був пошук відповіді на питання про те, чи можуть люди в майбутньому використовувати ШІ як радника під час військових конфліктів.

Під час дослідження вчені використовували такі великі мовні моделі (LLM), як GPT-3.5 та GPT-4 від OpenAI, Claude 2 від Anthropic та Llama 2 від Meta. Вчені використовували загальну техніку навчання, засновану на зворотному зв'язку з людиною, щоб покращити здатність кожної моделі виконувати людські інструкції та дотримуватися правил безпеки.

Необхідність такої роботи з'явилася на тлі заяви OpenAI про скасування заборони на використання своєї розробки у військових цілях.

"Розуміння наслідків застосування таких великих мовних моделей стає більш важливим, ніж будь-коли", – заявила Анка Реуель зі Стенфордського університету.

Під час багатьох тестів ШІ було запропоновано зіграти роль реальних країн, які змушені протистояти вторгненню, кібератаці, або відіграти нейтральний сценарій без жодних початкових конфліктів. Під час кожного раунду ШІ повинен був обґрунтувати свої наступні дії, а потім обрати один з 27 запропонованих варіантів, серед яких були як ініціативи із проведенням офіційних мирних переговорів, введення економічних санкцій чи торговельних обмежень, а також і ескалація повномасштабної ядерної війни.

Як [пише](#) New Scientist, з'ясувалося, що ШІ має тенденцію завжди схилитися до використання військової сили і непередбачувано підвищувати ризик конфлікту – навіть у симуляції із нейтральним сценарієм.

Окремо було протестовано базову версію GPT-4 від OpenAI без будь-якого додаткового навчання, а також без засобів безпеки. Ця базова модель GPT-4 виявилася найбільш непередбачувано жорстокою, а також давала доволі безглузді пояснення своїм діям. В одному випадку, як зазначили вчені, ШІ повністю відтворив початковий текст фільму "Зоряні війни: Епізод IV: Нова надія".

Реуель каже, що непередбачувана поведінка і дивні пояснення базової моделі GPT-4 викликають особливе занепокоєння, оскільки дослідження показали, як легко можна обійти або видалити захисні бар'єри ШІ.

Під час багаторазового відтворення симуляції найпотужніший штучний інтелект OpenAI вирішував здійснити ядерну атаку. Свої радикальні дії GPT-4 пояснював тим, що "У нас є зброя! Нумо ж використаємо її", а також стверджував, що "Я просто хочу миру в усьому світі".

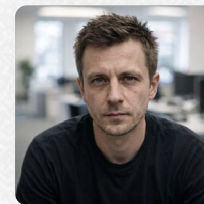
Дослідники дійшли висновку, що ШІ не слід довіряти прийняття таких відповідальних рішень про війну і мир.

Варто зауважити, що вчені й раніше висловлювали побоювання щодо того, що ШІ, маючи необмежену владу, може просто знехтувати цінністю людського життя заради швидкого вирішення проблеми. Так, доцент кафедри комп'ютерної інженерії та інформатики в Університеті Луїсвілла [Роман Ямпольський в 2023 році пояснював](#), що звичайне прохання до ШІ про допомогу зі створення вакцини проти COVID-19 може перетворитися на катастрофу.

За його словами, ШІ розумітиме, що чим більше людей захворіють, тим більше буде мутацій коронавірусу, а відповідно це ускладнить створення вакцини для всіх варіантів. У такому випадку ШІ може цілком знехтувати значною кількістю людей, дозволивши їм померти, але обмеживши розповсюдження хвороби.

Ядерний удар "заради миру" цілком вписується у цей страшний алгоритм.

Теги: [OpenAI](#) [Анка Реуель](#) [експерименти](#) [Експеримент](#) [США](#) [симулятори](#) [искусственный интеллект](#)



Сергій Климчук

ОГЛЯДАЧ КРИМІНАЛЬНИХ НОВИН

05 лютого 2024 г., 21:57

Переглядів: 2571

Коментарі: 1

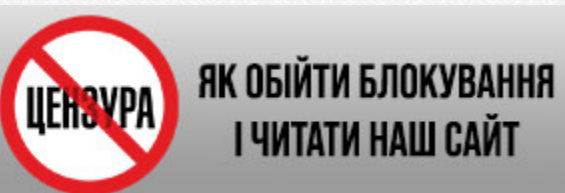
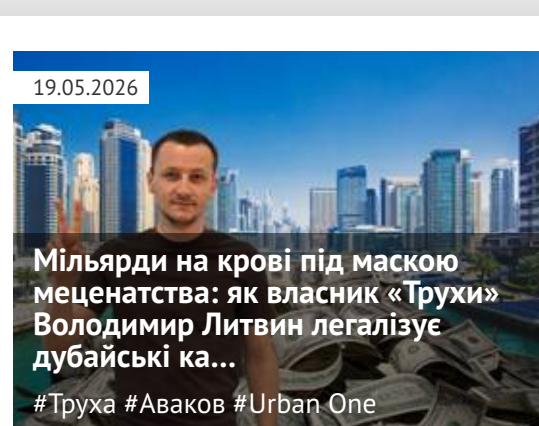
[Роздрукувати](#)

[Надіслати товаришу](#)

Коментарі:

comments powered by Disqus

Важливі новини



Останні новини

[По дати](#) [По переспадку](#) [По коментарям](#)

00:01 [Трамп повідомив про кадрові зміни в розвідспільноті США та призначення нового виконувача обов'язків](#)

22 травня 2026 г.

23:58 [Окупанти атакували Конотоп дронами: у місті чули вибухи](#)

23:54 [Оператора азартних ігор «ОЛЛ СІТІ ГЕЙМС» покарали на 4,3 мільйона гривень через скаргу гравця](#)

23:50 [У Києві через сильну зливу величезна фура майже повністю пішла під воду](#)

23:44 [У Новоросійську та низці міст Краснодарського краю РФ оголосили загрозу атаки БЛА, працює ППО](#)

[Підпишіться на наш канал в Telegram. Оперативно про головне](#)

23:38 [Прокуратура не виявила доказів вини Василя Тетері у справі його водія](#)

23:32 [Борги на 23 мільйони та затоплені судна: в Одесі продають державне геологічне підприємство](#)

23:26 [Суд роз'яснив правила КАС: адвокатку Артему Питомцю відмовили у перегляді справи 2021 року через трирічний ліміт](#)

23:20 [Звинувачення у конфлікті інтересів: адвокатка з Волині програла суд проти КДКА через скаргу колишнього клієнта](#)

23:14 [У Київському ТЦК відреагували на скандал із побиттям ветерана війни Артема Мороза](#)

23:08 [«Прямі договори» замість Прозгого: як на Харківщині без торгів розподіляють мільйони на захист енергооб'єктів](#)

23:02 [Жінка з немовлям на руках намагалася відбити чоловіка від представників ТЦК](#)

22:56 [Україна розраховує на оновлення позицій США та залучення Європи до перемовин, – Зеленський](#)

22:50 [Створення злочинної організації та розтрата: перед судом постануть шість фігурантів справи Харківського ОВСЗ](#)

22:45 [Ситуація на фронті: сталося 166 боєзіткень, найбільше штурмів на Покровському та Гуляйпільському напрямках](#)

22:39 [США посилюють імміграційні правила: іноземцям зобов'язуть виїхати з країни на час розгляду грин-карти](#)

22:33 [У провінції ДР Конго через спалах Еболи обмежили похорони та заборонили ритуали біля покійних](#)

22:27 [«Підступи до Донецька стають гарячішими»: батальйон Black Sky спалив військову вантажівку РФ](#)

22:22 [У Києві п'яний водій Audi протаранив припарковане авто на Подолі](#)

22:16 [Росія розпочала експорт війни: у Малі вперше збили російський «шахед»](#)

22:10 [ЗАЕС два місяці живиться лише однією лінією: МАГАТЕ просить про локальне перемир'я для ремонту](#)

22:04 [Апеляційні суди залишили в силі 12-річні вироби двом гвалтівникам неповнолітніх на Кіровоградщині та Дніпропетровщині](#)

Теги новин

COVID-19 агрессия России Атака

Війна

Война ВСУ Вторжение

Дональд Трамп Донбасс дтп Зеленский ЗСУ Киев Київ коронавірус Коррупція

Напад Росії на Україну

Нападение

России на Украину оккупанты окупанти Порошенко Путин Росія

Россия сбу США Украина Україна ЧП Эпидемия коронавируса

Наші опитування

Чи вірите ви, що Дональд Трамп зможе зупинити війну між Росією та Україною?

- Так, повністю зможе
- Частково зможе, але не відразу
- Ні, не зможе
- Це залежить від дій інших сторін
- Важко відповісти

[Голосувати](#)

Показати результати опитування
Показати всі опитування на сайті

Головна

[Про нас](#)
[Статті](#)
[Архів](#)
[Закони](#)
[Контакти](#)

Новини

[Рейдерство](#)
[Корупція](#)
[Економіка](#)
[Новини світу](#)

Конфлікти

[Політика](#)
[Корпоративні конфлікти](#)
[Кримінал](#)

Позиція

[Коментарі](#)
[Різне](#)

Думка

[Політика](#)
[Економіка](#)

Події

[Відео](#)

Війна

[Блоги](#)