**Szczyt w Davos: ChatGPT demokratyzuje hakowanie – czy można to powstrzymać?**

**Sztuczna inteligencja to przełomowa technologia, która może radykalnie poprawić nasze życie dzięki spersonalizowanej medycynie, bezpieczniejszemu transportowi i wielu innym możliwościom. Ma również ogromny potencjał, aby pomóc branży bezpieczeństwa przyspieszyć rozwój nowych narzędzi ochronnych i bezpiecznego kodowania. Jednak wprowadzenie technologii AI niesie ze sobą również potencjalne zagrożenia - ostrzega Dr Dorit Dor, dyrektor ds. produktów, Check Point Software podczas konferencji World Economic Forum w Davos.**

W listopadzie 2022 r. OpenAI wypuściła na rynek nowy model sztucznej inteligencji o nazwie ChatGPT (Generative Pre-Training Transformer) – umożliwiając ludziom zadawanie pytań i otrzymywanie odpowiedzi. Technologia stała się niezwykle popularna, zdobywając ponad [milion użytkowników](https://www.weforum.org/agenda/2022/12/chatgpt-ai-bot-intelligent-conversation/) w ciągu pięciu dni od jego uruchomienia. Wiele osób używało jej do pisania poezji lub tworzenia nowych przepisów. Ale inni mieli bardziej nikczemne pomysły. Eksperci szybko [odkryli](https://research.checkpoint.com/2022/opwnai-ai-that-can-save-the-day-or-hack-it-away/), że przestępcy używają ChatGPT do tworzenia złośliwych wiadomości e-mail i kodu, które można wykorzystać do hakowania organizacji. I w rzeczywistości - zaledwie kilka tygodni od jego wydania – ChatGPT jest już używany do tego właśnie celu. Zasadniczo jest to demokratyzacja hakowania — umożliwianie nawet nowicjuszom tworzenia złośliwych plików i narażanie nas wszystkich na ryzyko.

Dlaczego to ma znaczenie? Świat doświadczył [38 % wzrostu cyberataków w 2022 roku](https://blog.checkpoint.com/2023/01/05/38-increase-in-2022-global-cyberattacks/) (w porównaniu do 2021 roku). Przeciętna organizacja była atakowana 1168 razy tygodniowo. Edukacja i opieka zdrowotna były dwiema najbardziej atakowanymi branżami, co spowodowało przestój w [szpitalach](https://www.darkreading.com/attacks-breaches/cyberattack-shuts-down-french-hospital) i [szkołach](https://www.npr.org/2022/09/07/1121422336/a-cyberattack-hits-the-los-angeles-school-district-raising-alarm-across-the-coun). Lekarze nie byli w stanie leczyć pacjentów, a szkoły zamknięto, a dzieci odesłano do domu. Niestety, teraz jesteśmy świadkami wzrostu cyberataków z powodu ChatGPT i innych generatywnych modeli sztucznej inteligencji.

OpenAI dokłada starań, aby powstrzymać nadużywanie swojej technologii AI, [pisząc](https://openai.com/blog/chatgpt/): „Chociaż dołożyliśmy starań, aby model odrzucał nieodpowiednie żądania, czasami będzie reagował na szkodliwe instrukcje lub zachowywał się stronniczo”. Ale niestety ChatGPT zmaga się powstrzymywaniem generowania niebezpiecznego kodu.

Zespół ekspertów Check Point Research odkrył, [że niektórzy cyberprzestępcy wykorzystują ChatGPT do tworzenia złośliwych narzędzi](https://research.checkpoint.com/2023/opwnai-cybercriminals-starting-to-use-chatgpt/) – w niektórych przypadkach hakerzy całkowicie polegali na sztucznej inteligencji, podczas gdy inni po prostu używali sztucznej inteligencji do znacznego skrócenia czasu pisania złośliwego kodu. Od złośliwego oprogramowania, przez phishing, aż po skrypt, który może automatycznie zaszyfrować narzędzie ofiary bez interakcji użytkownika — z niepokojem obserwuje się, jak szybko cyberprzestępcy używają ChatGPT w środowisku naturalnym do swoich destrukcyjnych celów.

W tym miesiącu na dorocznym spotkaniu World Economic Forum mamy okazję, by liderzy z obszaru rządowego, sztucznej inteligencji, biznesu i cyberbezpieczeństwa spotkali się i pilnie omówili skrzyżowanie sztucznej inteligencji, cyberbezpieczeństwa i ogólnego bezpieczeństwa. Nie powinniśmy dążyć do tłumienia innowacji, ale raczej do zapewnienia bezpiecznych warunków ich wdrożenia. Razem z pewnością możemy stawić czoła temu zagrożeniu dla bezpieczeństwa publicznego, infrastruktury krytycznej i naszego świata.