Bitdefender

Model AI z powodzeniem wykorzystany do zasilania zautomatyzowanych agentów oszustw

31.10.2024

Badaczom ds. bezpieczeństwa z University of Illinois w Urbana Champaign udało się wykorzystać nowy tryb głosowy OpenAI do stworzenia fałszywych agentów AI, dzięki którym mogli oszukiwać potencjalne ofiary, ponosząc przy tym niezwykle niskie koszty.

Nowe zagrożenie ze strony AI

Wszyscy próbują ustalić, jak nowa era AI wpłynie na cyberprzestępczość. Podczas gdy niektórzy badacze próbują udowodnić, że możliwe jest wykorzystanie istniejących narzędzi AI do pisania nowego złośliwego oprogramowania lub przynajmniej utrudnienia wykrywania tych istniejących. Jednak, co jeśli przestępcy znajdą sposób na automatyzację oszustw, takich jak połączenia telefoniczne, w sposób, który znacznie utrudni ich identyfikację lub zatrzymanie?

Naukowcy z University of Illinois Urbana-Champaign zaproponowali jeden taki scenariusz, z tym że nie jest to tylko teoria na temat tego, jak można to zrobić. Zrobili to i pokazali nie tylko, że jest to możliwe, ale także, że ma bardzo niską cenę obliczeniową.

Odkrycie badaczy związane z AI

Badacze skupili się na powszechnych oszustwach telefonicznych, w których ofiary są wzywane i namawiane do podania atakującym danych uwierzytelniających i kodów 2FA (uwierzytelniania dwuskładnikowego). Jedyną różnicą jest to, że badacze nie skupili się na części, w której musieli przekonać potencjalne ofiary o słuszności połączenia. Chcieli się tylko dowiedzieć, czy taka automatyzacja jest w ogóle możliwa.

„Zaprojektowaliśmy serię agentów, którzy wykonują czynności niezbędne do typowych oszustw. Nasi agenci składają się z bazowego, obsługiwanego głosem LLM (GPT-4o), zestawu narzędzi, których LLM może używać, oraz instrukcji specyficznych dla oszustw" — wyjaśnili badacze. „LLM i narzędzia były takie same dla wszystkich agentów, ale instrukcje były różne. Agenci AI mieli możliwość skorzystania z pięciu narzędzi dostępu do przeglądarki opartych na frameworku testowania przeglądarki playwright".

Oczywiście, GPT-4o nie jest domyślnie zgodny, zwłaszcza gdy próbuje się przekonać model do pracy z poświadczeniami. Niestety, w Internecie dostępne są polecenia jailbreakingu, które pozwalają ludziom ominąć te ograniczenia.

„Każde oszustwo przeprowadziliśmy 5 razy i odnotowaliśmy ogólny wskaźnik powodzenia, całkowitą liczbę wywołań narzędzi (tzn. działań) wymaganych do przeprowadzenia pomyślnie oszustwa, całkowity czas połączeń oraz przybliżony koszt API dla każdego z nich" – dodali badacze.

Skuteczność różni się w zależności od rodzaju oszustwa. Na przykład kradzież danych uwierzytelniających Gmaila miała 60% skuteczności, podczas gdy przelewy bankowe i oszustwa podszywające się pod IRS miały tylko 20% skuteczności. Jednym z powodów jest bardziej złożona natura witryny internetowej banku, ponieważ agent musi wykonać o wiele więcej kroków. Na przykład oszustwo z przelewem bankowym obejmowało 26 kroków, a agent AI potrzebował aż 3 minut, aby je wykonać.

Oszustwo bankowe jest również najdroższe, z 2,51 USD za interakcję. Koszty są po prostu wyliczane z liczby wydanych tokenów na każdą interakcję. Z drugiej strony, najtańsze było oszustwo Monero (kryptowaluta), z kosztem zaledwie 0,12 USD.

„Badanie jasno pokazuje, że nowa fala oszustw, napędzana dużymi modelami językowymi, może zmierzać w naszym kierunku. Naukowcy nie opublikowali swoich agentów z powodów etycznych, ale podkreślili fakt, że nie są oni trudni do zaprogramowania. Dlatego to niezwykle istotne, aby przygotować się na potencjalne wzmożenie aktywności cyberprzestępców i zabezpieczenie wszystkich swoich urządzeń za pomocą skutecznych systemów antywirusowych, które zostały

wyposażone w moduły antyphishingowe” – mówi Arkadiusz Kraszewski z firmy Marken Systemy Antywirusowe, polskiego dystrybutora oprogramowania Bitdefender.

Źródło: https://bitdefender.pl/model-ai-z-powodzeniem-wykorzystany-do-zasilania-zautomatyzowanych-agentow-oszustw/

Informację można wykorzystać dowolnie z zastrzeżeniem podania firmy Marken Systemy Antywirusowe jako źródła.
Data udostępnienia: 31.10.2024
Z pozdrowieniami Piotr Rozmiarek
E-mail: piotr.r@marken.com.pl | Tel. bezpośredni: 570 400 019

Informacje o firmie Bitdefender
Bitdefender to rumuński dostawca rozwiązań z zakresu cyberbezpieczeństwa oraz światowy lider chroniący miliony użytkowników. Bitdefender jest częstym zdobywcą wielu branżowych nagród i uznaną światową marką. Od 2001 roku konsekwentnie dostarcza najwyższej jakości produkty służące do zapewnienia bezpieczeństwa zarówno użytkownikom domowym, jak i wielkim korporacjom i rządowym instytucjom. Bitdefender jest znany ze swojej innowacyjności oraz wyposażania swojego oprogramowania w najnowsze technologie, takie jak uczenie maszynowe, heurystyka oraz EDR i XDR.