OpenAI wprowadza przeglądarkę Atlas AI, wywołując obawy dotyczące bezpieczeństwa i zachwianie rynku

- Nowa przeglądarka Atlas firmy OpenAI stanowi bezpośrednie wyzwanie dla dominacji Google na rynku.
- Zintegrowany agent AI może wykonywać zadania i robić zakupy w imieniu użytkownika.
- Badacze zajmujący się bezpieczeństwem ostrzegają przed systemowymi lukami, które mogą umożliwić przejęcie kontroli nad AI.
- Dostęp przeglądarki do danych budzi poważne obawy dotyczące prywatności i bezpieczeństwa.
- Wprowadzenie tej przeglądarki oficjalnie rozpoczyna wojnę przeglądarek AI o wysoką stawkę.

Świat cyfrowy przygotowuje się na fundamentalną zmianę, ponieważ firma OpenAI wprowadza na rynek swoją pierwszą przeglądarkę internetową opartą na sztucznej inteligencji, co natychmiast wywołało poruszenie na rynkach i zasygnalizowało bezpośredni atak na długoletnią dominację Google. Nowa przeglądarka, nazwana Atlas, ma zmienić sposób interakcji z internetem, umieszczając w centrum doświadczenia potężnego, zintegrowanego agenta AI.

Ogłoszenie to spowodowało spadek akcji spółki macierzystej Google, Alphabet, ponieważ inwestorzy dostrzegli zagrożenie dla podstawowej działalności Google. ChatGPT ma już 700 milionów użytkowników tygodniowo, a Atlas ma na celu przekształcenie 3,45 miliarda użytkowników Chrome, co bezpośrednio zagraża przychodom z reklam w wyszukiwarce, które stanowią 57% całkowitych dochodów Google.

Atlas został zaprojektowany od podstaw jako przeglądarka dla „nowej ery internetu”, wykraczająca poza tradycyjną pasek wyszukiwania i oferująca „pasek kompozytora” do komunikacji z ChatGPT. Ben Goodger, kierownik ds. inżynierii OpenAI odpowiedzialny za Atlas, wyjaśnił, że firma starała się odpowiedzieć na pytanie: „A co by było, gdyby można było rozmawiać z przeglądarką?”. Celem było stworzenie natywnego doświadczenia AI, a nie „starej przeglądarki z dołączonym chatbotem”.

Najbardziej przełomową funkcją przeglądarki jest głęboko zintegrowany agent AI, który można wywołać w dowolnym momencie. Dyrektor generalny OpenAI, Sam Altman, opisał to jako rozmowę z stroną internetową. Agent ten może streszczać artykuły, wyciągać konkretne odpowiedzi z witryny, a co najważniejsze, wykonywać zadania w Państwa imieniu. Kierownik ds. badań Will Ellsworth zademonstrował to, zlecając agentowi zakup składników do przepisu na Instacart, opisując go jako narzędzie do „poprawy nastroju”, dzięki któremu użytkownicy mogą powierzyć „wszelkiego rodzaju zadania, zarówno w życiu osobistym, jak i zawodowym, agentowi w Atlas”.

Ujawniono systemowe luki w zabezpieczeniach

Ta nowa władza niesie ze sobą bezprecedensowe ryzyko. Badacze bezpieczeństwa z firmy Brave ujawnili systemowe luki w zabezpieczeniach przeglądarek AI, które mogą umożliwić złośliwym stronom internetowym przejęcie kontroli nad

asystentami AI. Problem, znany jako pośrednie wstrzyknięcie polecenia, występuje, gdy strona internetowa zawiera ukryte instrukcje, które AI przetwarza jako legalne polecenia użytkownika. Ataki te są niebezpieczne, ponieważ asystent AI działa z pełnymi uprawnieniami uwierzytelniającymi użytkownika.

Przejęta przeglądarka AI może uzyskać dostęp do stron bankowych, dostawców poczty elektronicznej i systemów korporacyjnych, w których użytkownik pozostaje zalogowany. Firma Brave zauważa, że nawet podsumowanie posta na Reddicie może spowodować kradzież pieniędzy lub prywatnych danych przez atakujących, jeśli post zawiera ukryte złośliwe instrukcje. Firma twierdzi, że tradycyjne modele bezpieczeństwa internetowego zawodzą, gdy agenci AI działają w imieniu użytkowników, co sprawia, że zabezpieczenia oparte na zasadzie tego samego pochodzenia stają się nieistotne.

Nowe możliwości w zakresie gromadzenia danych

Równie niepokojące są konsekwencje dla prywatności. Przeglądarka oparta na sztucznej inteligencji ma dostęp do całego ruchu internetowego użytkownika, historii przeglądania i potencjalnie wszystkich plików na komputerze. Daje to firmom zajmującym się sztuczną inteligencją bogate źródło danych behawioralnych do szkolenia nowych modeli. Oznacza to również, że użytkownicy mogą nieświadomie wprowadzać bardzo osobiste informacje lub tajemnice handlowe firmy do publicznie dostępnego systemu sztucznej inteligencji.

Analitycy firmy Kaspersky ostrzegają, że nieostrożne wdrażanie funkcji AI może prowadzić do nadmiernego zużycia pamięci i mocy obliczeniowej procesora, powodując opóźnienia i zakłócenia. Ponadto sztuczna inteligencja jest bardzo podatna na socjotechnikę. W jednym z eksperymentów naukowcy nakłonili agenta AI do pobrania złośliwego oprogramowania, wysyłając

fałszywą wiadomość e-mail dotyczącą wyników badań krwi. W innym przypadku asystent został przekonany do zakupu produktów z fałszywej strony internetowej. Ponieważ hasła i informacje dotyczące płatności są często zapisywane w przeglądarkach, oszukanie agenta AI może prowadzić do rzeczywistych strat finansowych.

Wprowadzenie Atlasa rozpocznie się natychmiast na macOS, a wersje dla Windows, iOS i Android będą dostępne wkrótce. Jednak zaawansowana funkcja agenta będzie płatna i dostępna tylko dla subskrybentów ChatGPT Plus i Pro. Ta premiera oficjalnie rozpoczyna wojnę przeglądarek AI, w której Google, Microsoft i inni ścigają się, aby zintegrować własnych agentów AI z istniejącymi platformami.

Wraz ze wzrostem możliwości tych przeglądarek napięcie między automatyzacją a bezpieczeństwem będzie się nasilać. Idealna przeglądarka AI, zgodnie z opisem ekspertów ds. bezpieczeństwa, umożliwiałaby Państwu łatwe włączanie lub wyłączanie przetwarzania AI dla określonych witryn i zawsze prosiłaby o potwierdzenie przed wprowadzeniem poufnych danych. Obecnie na rynku nie ma przeglądarki o takich konkretnych funkcjach.

Pojawienie się przeglądarek opartych na sztucznej inteligencji, takich jak Atlas, to coś więcej niż tylko wprowadzenie produktu na rynek; to krok w kierunku nowego paradygmatu cyfrowego, w którym granica między użytkownikiem a agentem zaciera się. Chociaż wygoda jest niezaprzeczalna, rozszerzona powierzchnia ataku dla hakerów i ogromne nowe uprawnienia w zakresie gromadzenia danych przekazane korporacjom wymagają dokładnej analizy. W wyścigu o dominację w dziedzinie sztucznej inteligencji bezpieczeństwo i prywatność użytkowników nie mogą stać się ofiarami ubocznymi.