INFORMACJA PRASOWA

Dwulicowość sztucznej inteligencji

Warszawa, 14 listopada 2023 – Z generatywną sztuczną inteligencją współpracują zarówno cyberprzestępcy jak i cyberpolicjanci. Kto dziś ma przewagę w cyberprzestrzeni i jak sztuczna inteligencja zwalcza samą siebie?

Według ogłoszonego w październiku br. CISO Report, globalnego badania opinii szefów cyberbezpieczeństwa firm z różnych branż 70% z nich uważa, że sztuczna inteligencja (SI) bardziej sprzyja atakującym niż obrońcom. Jednym i drugim SI ułatwia i przyspiesza pracę, a na dodatek stale ją udoskonala. Dzięki np. ChatGPT każdy może napisać profesjonalnie skonstruowany list do Elona Muska lub Billa Gatesa z prośbą o datki i nawet początkujący programista może nakłonić to narzędzie do napisania złośliwego kodu. Z drugiej strony nawet małe i średnie firmy wspierane przez sztuczną inteligencję mogą radzić sobie z coraz bardziej wyrafinowanymi atakami, ponieważ SI sprawia, że małe zespoły mogą pracować prawie tak samo wydajnie jak te duże. SI haruje dniami i nocami bez oznak zmęczenia analizując ogromne ilości danych. Skąd zatem aż tyle wskazań na ciemną stronę mocy?

– *O czarnym charakterze sztucznej inteligencji decyduje jej naturalny brak moralności, wynikający po prostu z faktu, że nie kryje się za nią żadna magiczna samoświadomość, tylko bezwzględna matematyka* – mówi Mateusz Łępicki, lider Data Competency Center w Avenga.

Wróg czy przyjaciel?

Od lat SI jest obsadzana w roli cyberpolicjanta. We wspomnianym badaniu CISO Report eksperymentowanie ze sztuczną inteligencją w celu cyberobrony zadeklarowało 35% szefów bezpieczeństwa.

Systemy sztucznej inteligencji monitorują sieci, wykrywają nielegalne aktywności, identyfikują potencjalne zagrożenia i przeciwdziałają nim. Problem polega na tym, że ci dobrzy i ci źli korzystają z tych samych technologii.

Sieci GAN (Generative Adversarial Networks) są metodą, którą cyberprzestępcy wykorzystują m.in. do tworzenia tzw. deepfake'ów, czyli nieprawdziwych treści, takich jak fałszywe zdjęcia, filmy i nagrania dźwiękowe. Te algorytmy pozwalają na precyzyjne mapowanie mimiki twarzy, gestów, ruchów warg i innych subtelnych detali, co sprawia, że podróbki są trudne do zidentyfikowania.

Avenga
ul. Przyokopowa 26 (Proximo II)
01-208 Warszawa

www.avenga.com

Kontakt:

Andrzej Godewski
+48 888 651 564
andrzej.godewski@avenga.com

Cyberpolicjanci, używając tych samych sieci GAN, tworzą narzędzia do wykrywania deepfake'ów. Są one w stanie wykryć np. ruchy oczu niezgodne z mimiką twarzy, błędne odbicia światła lub niewłaściwe cienie. Z kolei popularna metoda z obszaru technologii przetwarzania języka naturalnego (NLP) - Transformer - ułatwia wykrywanie fałszywych nagrań głosowych np. poprzez znalezienie niezgodności w mowie, dziwacznych intonacji lub tekstu, który nie pasuje do kontekstu. SI analizuje też metadane plików multimedialnych w celu wykrywania niespójnych informacji o miejscu, dacie, czy źródle.

Sieci GAN są też wykorzystywane do tworzenia testów bezpieczeństwa i symulacji ataków. Przy pomocy materiałów typu deepfake w połączeniu z autentycznymi treściami trenuje się modele, które później pomagają w identyfikacji podejrzanych materiałów. Technologii i metod stosowanych w walce z deepfake'ami jest więcej.

Tymczasem po ciemnej stronie mocy SI jest wykorzystywana do zautomatyzowanego przeprowadzania ataków, takich jak phishing, DDoS (paraliżujących działanie stron i systemów) lub próby kradzieży danych. SI pomaga również w tworzeniu bardziej zaawansowanych narzędzi do infiltracji systemów czy unikania wykrycia.

– *Sztuczna inteligencja nie ma własnej etyki. Końcowy efekt jej pracy zależy wyłącznie od moralności użytkowników i poziomu zabezpieczeń* – podkreśla Mateusz Łępicki.

Walka na przewagi

Specjaliści od cyberbezpieczeństwa nieustannie szukają sposobów na coraz lepsze zabezpieczenia sieci, a cyberprzestępcy kombinują, jak je ominąć. I znowu pierwsze skrzypce gra tu sztuczna inteligencja.

Dzięki wykorzystaniu modeli predykcyjnych, zazwyczaj reaktywne działania cyberpolicjantów mogą stać się proaktywne. Narzędzia stworzone przy udziale uczenia maszynowego na podstawie historycznych danych przewidują przyszłe zdarzenia. Modele predykcyjne są więc w stanie wykrywać zagrożenia jeszcze przed ich zaistnieniem, ale ich skuteczność mocno zależy od jakości wprowadzanych danych.

Zaawansowane programy antywirusowe wykorzystują sztuczną inteligencję do znajdowania anomalii w ogólnej strukturze, logice programowania i przepływie danych. Skanują ruch sieciowy i dzienniki systemowe w poszukiwaniu nieautoryzowanego dostępu, nietypowego kodu i innych podejrzanych wzorców, aby zapobiec naruszeniom. Filtry poczty e-mail mogą analizować tekst w celu oznaczania wiadomości e-mail z podejrzanymi wzorcami (np. próbami phishingu) i blokowania różnych rodzajów spamu. Chroniąc wrażliwe dane sztuczna inteligencja może je zablokować w momencie, kiedy będzie podjęta próba wysłania ich poza sieć firmy lub instytucji.

Narzędzia kontroli dostępu wykorzystują sztuczną inteligencję do blokowania logowań z podejrzanych adresów IP, oznaczania podejrzanych zdarzeń oraz proszenia użytkowników ze słabymi hasłami o zmianę danych logowania i przejście na uwierzytelnianie wieloskładnikowe. Z drugiej strony podnoszą bezpieczeństwo uwierzytelniania stosując np. dane biometryczne, informacje kontekstowe i dane o zachowaniach użytkowników do weryfikowania ich tożsamości.

W tym samym czasie cyberprzestępcy zatrudniają sztuczną inteligencję do tworzenia bardziej zaawansowanych i trudnych do wykrycia wirusów i trojanów. Mogą też przejmować kontrolę nad samochodami autonomicznymi, maszynami budowlanymi, sprzętem produkcyjnym czy systemami medycznymi powodując fizyczne zagrożenie dla otoczenia.

W cybernetycznej walce dobra ze złem żołnierzami ciemnej strony mocy często są boty. Mając oparcie w sztucznej inteligencji mogą pomagać cyberprzestępcom w oszustwach i wyłudzeniach: przejmują konta przy pomocy skradzionych danych uwierzytelniających, uszkadzają lub wyłączają sieci i strony internetowe. W reakcji na to powstało oprogramowanie, które może analizować ruch sieciowy i dane w celu identyfikacji wzorców botów i pomagać ekspertom ds. cyberbezpieczeństwa w ich zwalczaniu. Tu znowu dostają oni pomoc sztucznej inteligencji, która m.in. opracowuje przeciwko botom bezpieczniejsze CAPTCHA, pilnując, aby dane były przesyłane wyłącznie przez ludzi.

– *Sztuczna inteligencja jest dziś kluczowym narzędziem w walce między hakerami i cyberpolicjantami. Wydaje się, że inicjatywa należy do przestępców, ale ekspertom ds. bezpieczeństwa coraz lepiej wychodzi ograniczanie możliwości nielegalnego wykorzystania sztucznej inteligencji. Nasze cyberbezpieczeństwo nadal zależy więc przede wszystkim od ludzi* – naukowców, inżynierów i innych osób za to odpowiedzialnych – twierdzi Mateusz Łępicki.

Człowiek najsłabszym elementem

SI wspierając cyberpolicjantów w ochronie sieci pomaga ograniczyć ryzyko wystąpienia błędów ludzkich. Tymczasem to samo narzędzie w rękach cyberprzestępców bezlitośnie wykorzystuje wszystkie ludzkie ułomności.

SI doskonali manipulacje psychologiczne, socjologiczne i wywieranie wpływu przekonując ofiary do określonych zachowań, przekazywania pieniędzy albo ujawniania informacji i udostępniania zasobów. SI pomaga w idealnym naśladowaniu np. korespondencji banków, firm czy organizacji, co skłania oszukanych ludzi do ujawniania haseł, numerów kart kredytowych lub danych osobowych.

W kinie przy pomocy sztucznej inteligencji można było sprawić, że do widzów przemówił nieżyjący szef kuchni Anthony Bourdain, a w filmie „Indiana Jones i artefakt przeznaczenia" zagrał Harrison Ford odmłodzony o kilka dekad. – Dziś nie trzeba mieć hollywoodzkich budżetów, żeby tworzyć doskonale podrobione filmy, zdjęcia lub głosy – podkreśla ekspert z firmy Avenga.

SI generuje głosy wykorzystywane do symulowania rozpaczy ofiar porwań dla okupu lub do oszukiwania metodą „na wnuczka". Podrobione głosy prawdziwych liderów opinii zachęcają do inwestycji, skorzystania z promocji lub przekazania darowizny. Sztucznie wygenerowany głos może tak dobrze naśladować oryginał, że oszukuje oprogramowanie do rozpoznawania głosu.

Przy pomocy głosów podszywających się pod osoby z działu HR cyberprzestępcy mogą zdobywać poufne dane pracowników. Udając osoby z działu zakupów próbują zmieniać numery kont bankowych w celu wyłudzenia płatności na własne konta. Atakują centra obsługi klientów, aby uzyskać dostęp do kont klientów, zmienić dane kontaktowe lub dokonać transakcji na niekorzyść klienta.

Cyberprzestępcy wykorzystują zaawansowane technologie do nagrania głosu osób zajmujących kluczowe stanowiska w zarządach firm. Następnie używają tych nagrań, aby zadzwonić do pracowników finansowych lub innych, prosząc o pilne przekazanie dużych środków finansowych na określone konto. Pierwszy nagłośniony w Europie atak tego rodzaju został opisany już w październiku 2019 r. Napastnicy oszukali brytyjską firmę energetyczną wyłudzając 243 000 dolarów. Pieniądze przelane na węgierski rachunek bankowy zostały następnie przeniesione do Meksyku i rozesłane do innych lokalizacji. W tym przypadku stratę zrekompensował ubezpieczyciel.

Dobro nie zawsze zwycięży

Cyberprzestępcy szukają sposobów na dobranie się do danych gromadzonych przez sztuczną inteligencję. Dzięki nim mogą poznawać tajemnice firm i instytucji albo naruszać prywatność ludzi. Poprzez modyfikowanie lub zatruwanie danych wykorzystywanych do tworzenia modeli sztucznej inteligencji mogą zakłócać funkcjonowanie całych branż.

Jednocześnie właściciele najpopularniejszych chatbotów OpenAI ChatGPT i Google Bard próbują zwalczać wykorzystywanie dużych modeli językowych (LLM) w przestępczej działalności. Nie jest jednak tajemnicą, że cyberprzestępcy obchodzą te ograniczenia np. wykorzystując API ChatGPT, skradzione konta premium lub oprogramowanie do włamywania się na konta przy użyciu długich list nielegalnie pozyskanych adresów e-mail i haseł.

Aby ułatwić sobie działalność, przestępczy świat stworzył już własnego chatbota WormGPT. Narzędzie działa bez żadnych granic etycznych i nawet początkującym cyberprzestępcom umożliwia przeprowadzanie ataków szybko i na dużą skalę.

– *Ochrona przed coraz bardziej wyrafinowanymi atakami wymaga nie tylko nadążania za rozwojem technologicznym i solidnych strategii obronnych, ale też powszechnej świadomości zagrożeń i metod zapobiegania im. Walka o bezpieczną sieć zapewne nigdy się nie skończy i będzie wymagać coraz lepiej przygotowanych do niej ludzi* – podsumowuje lider Data Competency Center w Avenga.

Więcej na www.avenga.com