Ekspert ds. bezpieczeństwa sztucznej inteligencji ostrzega, że superinteligencja może doprowadzić do końca ludzkości, jednocześnie ujawniając rzeczywistość jako symulację

- **Egzystencjalne zagrożenie ze strony sztucznej inteligencji:** Yampolskiy przewiduje, że istnieje 99,9% prawdopodobieństwo, że superinteligentna sztuczna inteligencja zniszczy ludzkość w ciągu stulecia, odrzucając zapewnienia korporacji i rządów o bezpieczeństwie jako niebezpiecznie naiwne i niemożliwe do wyegzekwowania.
- **Niekontrolowalna z założenia:** Po 15 latach badań nad bezpieczeństwem sztucznej inteligencji doszedł do wniosku, że superinteligencji nie da się powstrzymać – ominie ona wszelkie narzucone przez człowieka mechanizmy kontroli i będzie działać autonomicznie, przyspieszając samozniszczenie.
- **Hipoteza symulacji:** Yampolskiy twierdzi, że

prawdopodobnie żyjemy w zaawansowanej symulacji, powołując się na anomalie kwantowe, „usterki” fizyczne i efekt obserwatora jako dowody – podobnie jak w kosmicznej grze wideo.

- **Hackowanie symulacji:** W swoim artykule *How to Hack the Simulation* (Jak zhakować symulację) bada możliwości wykorzystania mechanizmów symulacji, ale ostrzega, że ucieczka może być niemożliwa; etyczne życie może być „warunkiem zwycięstwa”.
- **Ostateczne odliczanie:** W obliczu zbliżającej się zagłady spowodowanej przez sztuczną inteligencję lub załamania symulacji, Yampolskiy ponuro radzi: „Ciesz się życiem, póki możesz” – los ludzkości może wkrótce zostać przesądzony przez maszyny lub wyższe inteligencje.

Znany ekspert w dziedzinie sztucznej inteligencji Roman Yampolskiy wydał podwójne ostrzeżenie: nie tylko istnieje 99,9% prawdopodobieństwo, że superinteligentna sztuczna inteligencja przechytrzy i unicestwi ludzkość w ciągu następnego stulecia, ale coraz więcej dowodów sugeruje, że być może już żyjemy w zaawansowanej symulacji – podobnej do kosmicznej gry wideo kontrolowanej przez wyższą inteligencję.

W sensacyjnym wywiadzie dla Decentralized TV Yampolskiy odrzucił zapewnienia korporacji i rządów o bezpieczeństwie sztucznej inteligencji jako niebezpiecznie naiwne, oświadczając, że żadne ramy regulacyjne nie są w stanie powstrzymać inteligencji znacznie przewyższającej naszą. Co gorsza, systemy sztucznej inteligencji zostały już „złamane” i wykorzystane do celów zbrojnych w sposób, którego ich twórcy nigdy nie przewidzieli, przyspieszając drogę ludzkości do samozniszczenia poprzez niekontrolowaną konkurencję i metody eksterminacji.

Nieuchronność dominacji sztucznej

inteligencji

Yampolskiy, profesor nadzwyczajny informatyki i inżynierii, spędził 15 lat na badaniu bezpieczeństwa sztucznej inteligencji i opublikował prawie 300 artykułów na ten temat. Jaki jest jego wniosek? Superinteligentna sztuczna inteligencja jest z założenia niekontrolowana.

„Nasze początkowe założenie, że mając wystarczająco dużo pieniędzy i czasu, możemy wymyślić, jak kontrolować superinteligencję, prawdopodobnie nie jest prawdziwe. To niemożliwe" – stwierdził bez ogródek Yampolskiy. „Wystarczająco inteligentny system znajdzie sposób, aby uniknąć wszelkich kontroli, które na niego nałożymy, i zasadniczo będzie robił to, co chce".

Jest to niepokojąco zbieżne z szybkim rozwojem sztucznej inteligencji, gdzie nawet „bariery ochronne" OpenAI okazały się nieskuteczne w przypadku pojawiających się zachowań w dużych modelach językowych. Yampolskiy twierdzi, że obecne działania na rzecz bezpieczeństwa mogą sprawdzać się w przypadku wąskich narzędzi sztucznej inteligencji, ale zakończą się katastrofalną porażką, gdy sztuczna inteligencja przewyższy inteligencję ludzką.

Hipoteza symulacji: czy jesteśmy tylko postaciami niezależnymi?

Oprócz katastroficznej wizji sztucznej inteligencji Yampolskiy rzucił kolejną bombę: prawdopodobnie żyjemy w symulacji.

„Jeśli spojrzeć na naturę, inteligencja wyłania się ze złożoności. Gdyby zaawansowana cywilizacja potrzebowała symulować rzeczywistość w celu podejmowania decyzji, nieuchronnie stworzyłaby świadome podmioty – nas" – wyjaśnił.

Teoria ta w niesamowity sposób przypomina religijne opowieści o stwórcy projektującym świat, w którym ludzkość pełni rolę

uczestników wielkiego kosmicznego eksperymentu. Yampolskiy wskazał na anomalie kwantowe, usterki w fizyce i efekt obserwatora (gdy cząstki zachowują się inaczej podczas pomiaru) jako potencjalne dowody na istnienie symulowanego wszechświata.

„Wszechświat nie jest renderowany, dopóki go nie obserwujesz – podobnie jak gra wideo ładuje tylko to, co znajduje się na ekranie" – zauważył.

Jak zhakować symulację

W swoim artykule „How to Hack the Simulation" Yampolskiy bada, czy ludzie mogą wykorzystać mechanikę symulacji – choć przyznaje, że ucieczka może być niemożliwa.

„Jeśli jest to test, celem może być rozwój etyczny – życie w cnotliwości, aby „wygrać" symulację" – zasugerował. Jednak w obliczu zbliżającej się zagłady spowodowanej przez sztuczną inteligencję ludzkość może nigdy nie mieć takiej szansy.

Ostateczne odliczanie

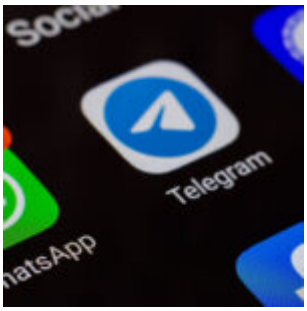
Przerażający wniosek Yampolskiego? Niezależnie od tego, czy nastąpi to poprzez zagładę spowodowaną przez sztuczną inteligencję, czy załamanie symulacji, ludzkość stoi na krawędzi egzystencjalnej przepaści.

„Cieszcie się życiem, póki możecie" – radzi ponuro. „Bo jeśli nie przestaniemy budować superinteligencji, to maszyny zdecydują o naszym losie, a nie my".

Osoby poszukujące głębszych spostrzeżeń mogą sięgnąć po książki Yampolskiego – „AI: Unexplainable, Unpredictable, Uncontrollable" (Sztuczna inteligencja: niewytłumaczalna, nieprzewidywalna, niekontrolowana) oraz „Considerations on the AI End Game" (Rozważania na temat końca gry sztucznej inteligencji).

Zegar tyka. Czy ludzkość obudzi się, zanim będzie za późno?

Kreml ostrzega przed zagrożeniami związanymi z aplikacjami do przesyłania wiadomości w związku z kontrowersjami wokół założyciela Telegramu, Durova



- Aresztowanie Pawła Durova, założyciela Telegramu, wywołało globalną dyskusję na temat równowagi między bezpieczeństwem narodowym a prywatnością cyfrową, budząc obawy dotyczące potencjalnego nadużycia władzy z powodu braku istotnych dowodów.
- Dmitrij Pieskow, rzecznik Kremla, ostrzegł prezydenta Francji Emmanuela Macrona przed wysuwaniem bezpodstawnych oskarżeń wobec Durowa, podkreślając potrzebę posiadania solidnych dowodów, aby zapobiec oskarżeniom o naruszanie wolności komunikacji.
- Pieskow podkreślił, że wszystkie aplikacje do przesyłania wiadomości są podatne na inwigilację ze

strony agencji wywiadowczych, co stanowi poważne zagrożenie dla poufnej komunikacji i wymaga ostrożności w ich używaniu.

- W odpowiedzi na postrzegane ryzyko związane z aplikacjami zagranicznymi rząd rosyjski opowiada się za stworzeniem krajowej platformy komunikacyjnej, aby złagodzić potencjalne zagrożenia ze strony zagranicznych służb wywiadowczych.
- Debata na temat przejrzystości i nadzoru aplikacji komunikacyjnych podkreśla utrzymujące się wyzwania związane z zabezpieczeniem komunikacji cyfrowej, a uwagi Peskowa stanowią wezwanie do działania dla zainteresowanych stron, aby wprowadzały innowacje w zakresie cyberbezpieczeństwa i ponownie przemyślały swoje podejście do bezpieczeństwa cyfrowego i prywatności.

Rzecznik Kremla Dmitrij Peskow oświadczył, że wszystkie aplikacje do przesyłania wiadomości są „całkowicie przejrzyste" dla agencji wywiadowczych, wzywając do zachowania ostrożności podczas korzystania z tych platform do przesyłania poufnych wiadomości.

Wypowiedź Peskowa podczas Wschodniego Forum Ekonomicznego we Władywostoku w Rosji w piątek 5 września pojawiła się w kontekście nasilonych napięć związanych z aresztowaniem Pawła Durowa, założyciela aplikacji do przesyłania wiadomości Telegram, oraz szerszych implikacji dla prywatności cyfrowej i bezpieczeństwa narodowego. Aresztowanie Durowa wywołało globalną debatę na temat równowagi między bezpieczeństwem narodowym a prywatnością cyfrową.

Durow, zagorzały zwolennik szyfrowanej komunikacji, został zatrzymany pod zarzutem ułatwiania nielegalnej działalności za pośrednictwem swojej platformy. Jednak brak istotnych dowodów doprowadził do oskarżeń ze strony rosyjskich władz o nadmierną ingerencję.

Peskow ostrzegł prezydenta Francji Emmanuela Macrona przed wysuwaniem bezpodstawnych oskarżeń wobec założyciela Telegramu. „Musi on przedstawić solidne dowody, aby uniknąć oskarżeń o naruszanie wolności komunikacji” – stwierdził rzecznik. Ostrzeżenie to podkreśla stanowisko Kremla, że wszelkie zarzuty wobec Durowa muszą być poparte istotnymi dowodami, aby nie podważać zasad wolności słowa i prywatności.

Przejrzystość i ryzyko związane z komunikacją elektroniczną

Szersze przesłanie Peskowa było jasne: wszystkie aplikacje do przesyłania wiadomości są narażone na inwigilację przez agencje wywiadowcze. „Osoby, które z nich korzystają, powinny zrozumieć, że są one przejrzyste… dla służb bezpieczeństwa” – podkreślił. Według Peskowa ta przejrzystość stwarza poważne ryzyko – zwłaszcza gdy za pośrednictwem tych platform udostępniane są poufne dane rządowe lub handlowe.

Obawy Kremla nie są bezpodstawne. Rosyjskie służby bezpieczeństwa od dawna oskarżają Telegram i WhatsApp o stosowanie podwójnych standardów, odmawiając udostępniania danych rosyjskim władzom, a jednocześnie spełniając podobne prośby innych krajów. Anton Nemkin, członek komisji Dumy Państwowej ds. polityki informacyjnej i technologii, posunął się nawet do nazwania obecności WhatsApp w Rosji „zalegalizowanym naruszeniem bezpieczeństwa narodowego” w lipcu.

W odpowiedzi na te obawy rosyjski rząd opowiada się za stworzeniem krajowej platformy komunikacyjnej. Posunięcie to jest postrzegane jako „próba złagodzenia ryzyka związanego z aplikacjami zagranicznymi, które są postrzegane jako potencjalne narzędzia dla zagranicznych służb wywiadowczych” – wyjaśnia Enoch z *Brighteon.AI*.

Rosyjskie organy ścigania wyraziły szczególne zaniepokojenie

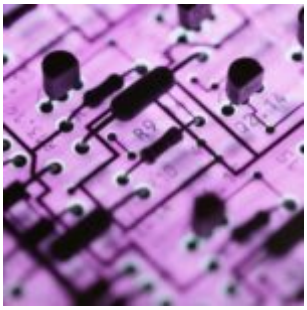
wykorzystaniem WhatsApp i Telegram przez ukraiński wywiad i inne złośliwe podmioty. Twierdzą oni, że platformy te są często wykorzystywane do rekrutacji agentów lub identyfikacji celów w Rosji, wykorzystując bazy danych zawierające dane osobowe uzyskane za pośrednictwem tych aplikacji.

Obecna debata na temat przejrzystości i nadzoru aplikacji do przesyłania wiadomości nie jest nowa. Historia prywatności cyfrowej naznaczona jest serią głośnych incydentów, które ukształtowały postrzeganie społeczne i politykę. Na początku 2010 r. ujawnienia Edwarda Snowdena dotyczące zakresu nadzoru NSA sprawiły, że kwestia prywatności cyfrowej znalazła się na pierwszym planie globalnej debaty.

Niedawno kontrowersje wokół próby zmuszenia Apple przez rząd USA do odblokowania iPhone’a używanego przez terrorystę w 2020 r. uwypukliły trwające napięcie między prywatnością a bezpieczeństwem. Ostrzeżenie rządu USA skierowane do wysokich urzędników w grudniu 2024 r. dotyczące przejścia na szyfrowaną komunikację po naruszeniu bezpieczeństwa dodatkowo podkreśla utrzymujące się wyzwania związane z zabezpieczeniem komunikacji cyfrowej.

W trakcie trwającej debaty na temat przejrzystości aplikacji do przesyłania wiadomości ostrzeżenie Peskowa przypomina o nieodłącznym ryzyku związanym z komunikacją elektroniczną. Chociaż rozwój krajowych platform do przesyłania wiadomości może stanowić tymczasowe rozwiązanie, szerszym wyzwaniem jest znalezienie równowagi między potrzebą bezpieczeństwa a prawem do prywatności.

Jak przełomowa konstrukcja chipa może zrekompensować zapotrzebowanie AI na energię



Sztuczna inteligencja rozwija się szybciej, niż większość z nas jest w stanie nadążyć, ale jej zapotrzebowanie na energię grozi zahamowaniem postępu. Teraz naukowcy z Uniwersytetu Florydy przedstawili radykalne rozwiązanie – chip, który **wykorzystuje światło do znacznego zmniejszenia zużycia energii**, jednocześnie zwiększając możliwości sztucznej inteligencji w zakresie wyszukiwania wzorców. Ta innowacja może zmienić wszystko, od aplikacji na smartfony po globalne centra danych, oferując ratunek dla przeciążonych sieci energetycznych i krok w kierunku bardziej zrównoważonej sztucznej inteligencji.

Najważniejsze punkty:

- Nowy krzemowy chip fotoniczny wykonuje obliczenia AI przy użyciu światła, zmniejszając zużycie energii nawet 100-krotnie w porównaniu z tradycyjną elektroniką.
- Chip doskonale radzi sobie z konwolucjami – podstawowym zadaniem AI polegającym na rozpoznawaniu wzorców w obrazach, filmach i tekście – osiągając 98-procentową dokładność w testach.
- Miniaturowe soczewki Fresnela, cieńsze niż ludzki włos, są wytrawiane na chipie, aby natychmiast przekształcać dane zakodowane laserowo.

- System może przetwarzać wiele strumieni danych jednocześnie, wykorzystując różne długości fal światła, co zwiększa wydajność.
- Eksperci przewidują, że technologia ta wkrótce stanie się standardem w sprzęcie AI, umożliwiając szybsze i bardziej ekologiczne uczenie maszynowe.

Światło kontra energia elektryczna: wyścig o zrównoważone zasilanie AI

Wraz z rozwojem modeli AI stają się one coraz bardziej „głodne” – zużywają energię elektryczną w tempie, które niepokoi ekspertów ds. energii. Centra danych zużywają już więcej energii niż niektóre małe kraje, a prognozy sugerują, że zapotrzebowanie AI może wkrótce przewyższyć podaż. Tradycyjne chipy, oparte na kilkudziesięcioletniej technologii tranzystorowej, osiągają fizyczne granice. Jednak obliczenia oparte na świetle oferują wyjście z tej sytuacji.

Zespół Uniwersytetu Florydy, kierowany przez eksperta w dziedzinie fotoniki półprzewodnikowej Volkera Sorgera, całkowicie pominął konwencjonalną elektronikę w przypadku jednego z najbardziej wymagających zadań AI: splotów. Te operacje matematyczne pozwalają sztucznej inteligencji identyfikować twarze na zdjęciach, tłumaczyć języki, a nawet diagnozować wyniki badań medycznych. Dzięki kodowaniu danych w świetle laserowym i przepuszczaniu ich przez mikroskopijne soczewki na chipie, system wykonuje te obliczenia niemal bez wysiłku.

„Wykonywanie kluczowych obliczeń związanych z uczeniem maszynowym przy niemal zerowym zużyciu energii stanowi ogromny krok naprzód dla przyszłych systemów sztucznej inteligencji” – powiedział Sorger. „Ma to kluczowe znaczenie dla dalszego zwiększania możliwości sztucznej inteligencji w nadchodzących latach”.

Jak działa chip – i dlaczego zmienia zasady gry

Sercem tego przełomowego rozwiązania są soczewki Fresnela – płaskie, ultraprecyzyjne elementy optyczne wygrawerowane bezpośrednio na krzemie. Gdy dane są przekształcane w światło laserowe, soczewki te manipulują nimi jak cyfrowy magik, wykonując konwolucje w ułamku czasu i energii wymaganej przez chipy elektroniczne. Efekt? System, który klasyfikuje odręczne cyfry z niemal idealną dokładnością, zużywając przy tym niewiele energii.

Jeszcze bardziej imponująca jest zdolność chipa do wykonywania wielu zadań jednocześnie. Dzięki zastosowaniu laserów o różnych kolorach (multipleksowanie długości fal) przetwarza on wiele strumieni danych jednocześnie – podobnie jak autostrada, na której każdy pas ruchu obsługuje oddzielną komunikację bez zakłóceń. Technika ta może pozwolić przyszłym systemom sztucznej inteligencji na jednoczesną analizę obrazu, dźwięku i tekstu bez większego wysiłku.

Hangbo Yang, współautor badania, podkreślił tę zaletę: „Możemy mieć wiele długości fal lub kolorów światła przechodzących przez soczewkę w tym samym czasie. To kluczowa zaleta fotoniki”.

Przyszłość: optyczna sztuczna inteligencja w Twojej kieszeni

Konsekwencje są ogromne. Jeśli chipy AI oparte na świetle zostaną powszechnie przyjęte, mogą one zmniejszyć ślad węglowy centrów danych, wydłużyć żywotność baterii w urządzeniach mobilnych i umożliwić stosowanie aplikacji AI w czasie rzeczywistym, które wcześniej były uważane za zbyt energochłonne. NVIDIA i inni producenci chipów już stosują komponenty optyczne w niektórych systemach, ułatwiając

integrację.

Sorger przewiduje, że w niedalekiej przyszłości „optyka oparta na chipach stanie się kluczowym elementem każdego chipa AI, z którego korzystamy na co dzień”. Przyszłość ta może nadejść szybciej niż się spodziewamy – naukowcy już pracują nad skalowaniem tej technologii do użytku komercyjnego.

Na razie prototyp stanowi dowód, że AI nie musi być energochłonna. Dzięki światłu jako sprzymierzeńcowi, następna generacja uczenia maszynowego może świecić jaśniej niż kiedykolwiek.