

Delikatna osobliwość

13 czerwca 2025

Przekroczyliśmy horyzont zdarzeń; start już się rozpoczął. Ludzkość jest bliska stworzenia cyfrowej superinteligencji i przynajmniej na razie jest to znacznie mniej dziwne, niż mogłoby się wydawać.



Roboty nie chodzą jeszcze po ulicach, a większość z nas nie rozmawia przez cały dzień z AI. Ludzie nadal umierają z powodu chorób, nadal nie możemy łatwo polecieć w kosmos, a o wszechświecie wciąż nie wiemy wielu rzeczy.

A jednak niedawno stworzyliśmy systemy, które pod wieloma względami są inteligentniejsze od ludzi i potrafią znacznie zwiększyć wydajność osób, które z nich korzystają. Najtrudniejsza część pracy już za nami; naukowe odkrycia, które doprowadziły nas do stworzenia systemów takich jak GPT-4 i o3, były bardzo trudne do osiągnięcia, ale pozwolą nam zajść bardzo daleko.

Sztuczna inteligencja przyczyni się do rozwoju świata na wiele sposobów, ale korzyści dla jakości życia wynikające z szybszego postępu naukowego i zwiększonej produktywności będą ogromne; przyszłość może być znacznie lepsza niż

teraźniejszość. Postęp naukowy jest największym motorem ogólnego postępu; niezwykle ekscytujące jest myślenie o tym, jak wiele jeszcze możemy osiągnąć.

W pewnym sensie ChatGPT jest już potężniejszy niż jakikolwiek człowiek, jaki kiedykolwiek żył. Setki milionów ludzi korzystają z niego codziennie i to do coraz ważniejszych zadań; niewielka nowa funkcja może mieć ogromny pozytywny wpływ, a niewielka niedoskonałość pomnożona przez setki milionów ludzi może spowodować ogromne negatywne skutki.

W 2025 roku pojawiły się agenty, które potrafią wykonywać prawdziwą pracę poznawczą; pisanie kodu komputerowego już nigdy nie będzie takie samo. W 2026 roku prawdopodobnie pojawią się systemy, które będą w stanie wyciągać nowatorskie wnioski. W 2027 roku mogą pojawić się roboty, które będą w stanie wykonywać zadania w świecie rzeczywistym.

Znacznie więcej osób będzie mogło tworzyć oprogramowanie i dzieła sztuki. Jednak świat potrzebuje znacznie więcej zarówno jednego, jak i drugiego, a eksperci prawdopodobnie nadal będą znacznie lepsi od nowicjuszy, o ile tylko zaakceptują nowe narzędzia. Ogólnie rzecz biorąc, możliwość wykonania przez jedną osobę w 2030 r. znacznie większej ilości pracy niż w 2020 r. będzie stanowić uderzającą zmianę, z której wiele osób będzie chciało skorzystać.

Pod najważniejszymi względami lata 30. XXI wieku mogą nie różnić się zbytnio od dzisiejszych czasów. Ludzie nadal będą kochać swoje rodziny, wyrażać swoją kreatywność, grać w gry i pływać w jeziorach.

Jednak pod wieloma nadal bardzo ważnymi względami lata 30. XXI wieku będą prawdopodobnie diametralnie różnić się od wszystkich poprzednich okresów. Nie wiemy, jak daleko możemy się posunąć w rozwoju inteligencji wykraczającej poza poziom ludzki, ale wkrótce się o tym przekonamy.

W latach 30. XXI wieku inteligencja i energia – pomysły oraz

zdolność do ich realizacji – staną się niezwykle dostępne. Przez długi czas były one podstawowymi ograniczeniami postępu ludzkości, ale dzięki ich obfitości (oraz dobrym rządcom) teoretycznie możemy osiągnąć wszystko.

Już teraz żyjemy w otoczeniu niesamowitej inteligencji cyfrowej i po początkowym szoku większość z nas dość dobrze się do tego przyzwyczaiła. Bardzo szybko przechodzimy od zdumienia, że sztuczna inteligencja potrafi stworzyć pięknie napisany akapit, do zastanawiania się, kiedy będzie w stanie napisać piękną powieść; od zdumienia, że potrafi postawić diagnozę ratującą życie, do zastanawiania się, kiedy będzie w stanie opracować lekarstwo; od zdumienia, że potrafi stworzyć mały program komputerowy, do zastanawiania się, kiedy będzie w stanie stworzyć zupełnie nową firmę. Tak właśnie przebiega singularność: cuda stają się rutyną, a następnie podstawą.

Naukowcy już teraz twierdzą, że dzięki sztucznej inteligencji są dwa lub trzy razy bardziej produktywni niż przed jej wprowadzeniem. Zaawansowana sztuczna inteligencja jest interesująca z wielu powodów, ale chyba nic nie jest tak istotne jak fakt, że możemy ją wykorzystać do przyspieszenia badań nad sztuczną inteligencją. Być może uda nam się odkryć nowe podłoża obliczeniowe, lepsze algorytmy i kto wie, co jeszcze. Jeśli uda nam się przeprowadzić badania warte dziesięciu lat w ciągu roku lub miesiąca, tempo postępu będzie oczywiście zupełnie inne.

Od tego momentu narzędzia, które już stworzyliśmy, pomogą nam zdobyć dalszą wiedzę naukową i stworzyć lepsze systemy sztucznej inteligencji. Oczywiście nie jest to to samo, co system sztucznej inteligencji całkowicie samodzielnie aktualizujący swój kod, ale jest to jednak wstępna wersja rekurencyjnego samodoskonalenia.

Istnieją również inne samonapędzające się pętle. Tworzenie wartości ekonomicznej uruchomiło koło zamachowe, które napędza rozbudowę infrastruktury niezbędnej do działania coraz

potężniejszych systemów sztucznej inteligencji. Roboty potrafiące budować inne roboty (a w pewnym sensie również centra danych, które mogą budować inne centra danych) nie są już tak odległą przyszłością.

Jeśli musimy wyprodukować pierwszy milion humanoidalnych robotów w starym stylu, ale potem będą one mogły obsługiwać cały łańcuch dostaw – wydobywać i przetwarzać minerały, prowadzić ciężarówki, obsługiwać fabryki itp. – aby zbudować więcej robotów, które z kolei będą mogły budować więcej zakładów produkujących chipy, centra danych itp., to tempo postępu będzie oczywiście zupełnie inne.

Wraz z automatyzacją produkcji centrów danych koszty inteligencji powinny ostatecznie zbliżyć się do kosztów energii elektrycznej. (Ludzie często są ciekawi, ile energii zużywa zapytanie ChatGPT; średnie zapytanie zużywa około 0,34 wata-godziny, czyli mniej więcej tyle, ile piekarnik zużywa w nieco ponad sekundę lub energooszczędna żarówka w ciągu kilku minut. Zużywa również około 0,000085 galona wody, czyli mniej więcej jedną piętnastą łyżeczki).

Tempo postępu technologicznego będzie nadal przyspieszać, a ludzie będą w stanie dostosować się do niemal wszystkiego. Pojawiają się bardzo trudne kwestie, takie jak zanik całych grup zawodów, ale z drugiej strony świat stanie się tak bogaty w tak krótkim czasie, że będziemy mogli poważnie rozważać nowe pomysły polityczne, które wcześniej były nie do pomyślenia. Prawdopodobnie nie przyjmimy nowej umowy społecznej od razu, ale gdy spojrzymy wstecz za kilka dziesięcioleci, stopniowe zmiany będą miały ogromne znaczenie.

Jeśli historia jest dla nas wskazówką, to wymyślimy nowe rzeczy do robienia i nowe rzeczy, których będziemy pragnąć, a także szybko przyswoimy nowe narzędzia (dobrym przykładem z ostatnich czasów jest zmiana pracy po rewolucji przemysłowej). Oczekiwania będą rosły, ale możliwości będą rosły równie szybko i wszyscy będziemy mieli lepsze rzeczy. Będziemy

budować dla siebie coraz wspanialsze rzeczy. Ludzie mają długoterminową, ważną i ciekawą przewagę nad sztuczną inteligencją: jesteśmy genetycznie zaprogramowani, aby troszczyć się o innych ludzi, o to, co myślą i robią, a nie przejmujemy się zbytnio maszynami.

Rolnik żyjący z uprawy roli tysiąc lat temu patrzyłby na to, co robi wielu z nas, i powiedziałby, że mamy fałszywe zawody, i pomyślałby, że po prostu bawimy się dla rozrywki, ponieważ mamy pod dostatkiem jedzenia i niewyobrażalne luksusy. Mam nadzieję, że za tysiąc lat spojrzemy na zawody wykonywane w przyszłości i pomyślimy, że są one bardzo fałszywe, i nie mam wątpliwości, że będą one niezwykle ważne i satysfakcjonujące dla osób, które je wykonują.

Tempo osiągania nowych cudów będzie ogromne. Trudno sobie dziś nawet wyobrazić, co odkryjemy do 2035 roku. Być może w ciągu jednego roku uda nam się rozwiązać zagadnienia fizyki wysokich energii, a w następnym rozpoczniemy kolonizację kosmosu. Albo w jednym roku dokonamy przełomowego odkrycia w dziedzinie materiałoznawstwa, a w następnym stworzymy prawdziwe interfejsy mózg-komputer o wysokiej przepustowości. Wiele osób zdecyduje się żyć w podobny sposób, ale przynajmniej niektórzy prawdopodobnie zdecydują się „podłączyć”.

Patrząc w przyszłość, trudno to sobie wyobrazić. Jednak prawdopodobnie przeżycie tego będzie imponujące, ale możliwe do zniesienia. Z perspektywy relatywistycznej osobliwość następuje stopniowo, a połączenie przebiega powoli. Wspinamy się po długim łuku wykładniczego postępu technologicznego; patrząc w przyszłość, zawsze wygląda on na pionowy, a patrząc wstecz – na płaski, ale w rzeczywistości jest to jedna płynna krzywa. (Wróćmy pamięcią do roku 2020 i zastanówmy się, jak brzmiałaby wizja osiągnięcia czegoś zbliżonego do AGI do 2025 roku w porównaniu z tym, jak wyglądało ostatnie pięć lat).

Wraz z ogromnymi korzyściami pojawiają się poważne wyzwania. Musimy rozwiązać kwestie bezpieczeństwa, zarówno pod względem

technicznym, jak i społecznym, ale biorąc pod uwagę konsekwencje gospodarcze, niezwykle ważne jest również zapewnienie szerokiego dostępu do superinteligencji. Najlepszym rozwiązaniem może być coś w rodzaju:

1. Rozwiąż problem dostosowania, co oznacza, że możemy solidnie zagwarantować, że systemy sztucznej inteligencji będą się uczyć i działać zgodnie z tym, czego naprawdę chcemy jako społeczeństwo w perspektywie długoterminowej (przykładem niedostosowanej sztucznej inteligencji są kanały mediów społecznościowych; algorytmy, które je napędzają, są niesamowicie skuteczne w zachęcaniu użytkowników do dalszego przewijania treści i zrozumieniu ich krótkoterminowych preferencji, ale robią to, wykorzystując coś w mózgu, co przeważa nad długoterminowymi preferencjami).

2. Następnie należy skupić się na tym, aby superinteligencja była tania, powszechnie dostępna i nie była zbyt skoncentrowana w rękach jednej osoby, firmy lub kraju. Społeczeństwo jest odporne, kreatywne i szybko się dostosowuje. Jeśli uda nam się wykorzystać zbiorową wolę i mądrość ludzi, to mimo że popełnimy wiele błędów i niektóre rzeczy pójdą naprawdę źle, będziemy się szybko uczyć i dostosowywać, a także będziemy w stanie wykorzystać tę technologię, aby uzyskać maksymalne korzyści i minimalne straty. Bardzo ważne wydaje się zapewnienie użytkownikom dużej swobody w ramach szerokich granic, które musi wyznaczyć społeczeństwo. Im szybciej świat rozpocznie dyskusję na temat tych szerokich granic i sposobu definiowania zbiorowej zgodności, tym lepiej.

My (cała branża, nie tylko OpenAI) budujemy mózg dla świata. Będzie on niezwykle spersonalizowany i łatwy w użyciu dla każdego; ograniczać nas będą tylko dobre pomysły. Przez długi czas technicy z branży start-upowej wyśmiewali „ludzi od pomysłów” – osoby, które miały pomysł i szukały zespołu, który go zrealizuje. Teraz wydaje mi się, że wkrótce nadejdzie ich czas.

OpenAI to obecnie wiele rzeczy, ale przede wszystkim jesteśmy firmą zajmującą się badaniami nad superinteligencją. Przed nami wiele pracy, ale większość drogi jest już wytyczona, a ciemne obszary szybko się zmniejszają. Jesteśmy niezwykle wdzięczni, że możemy robić to, co robimy.

Inteligencja, której nie da się zmierzyć, jest w zasięgu ręki. Może to zabrzmieć szalenie, ale gdybyśmy w 2020 roku powiedzieli, że będziemy w miejscu, w którym jesteśmy dzisiaj, brzmiałoby to prawdopodobnie bardziej szalenie niż nasze obecne prognozy na rok 2030.

Obyśmy mogli płynnie, wykładniczo i bez zakłóceń przejść przez etap superinteligencji.

Autorstwo: Sam Altman

Tłumaczenie: Mora

Ilustracja: WolneMedia.net (CC0)

Źródło zagraniczne: [blog.SamAltman.com](https://blog.samaltman.com)

Źródło polskie: [Nieznane.info](https://nieznane.info)