Liderzy branży sztucznej inteligencji ostrzegają przed „katastrofalnymi skutkami” w związku z pojawieniem się sztucznej inteligencji ogólnej

- Sztuczna inteligencja ogólna (AGI) może pojawić się w ciągu dziesięciu lat, powodując katastrofalne skutki, takie jak cyberataki, broń autonomiczna i zagrożenia egzystencjalne dla ludzkości – ostrzega Demis Hassabis, dyrektor generalny Google DeepMind.
- Sztuczna inteligencja (AI) jest już wykorzystywana do cyberataków na infrastrukturę krytyczną (np. systemy energetyczne, wodociągowe), dezinformację typu deepfake, oszustwa i wypieranie miejsc pracy, a FBI ostrzega przed oszustwami generowanymi przez AI i manipulacją polityczną.
- Ponad 350 ekspertów w dziedzinie AI, w tym Sam Altman z OpenAI oraz pionierzy AI Yoshua Bengio i Geoffrey Hinton, podpisało oświadczenie, w którym porównują ryzyko związane z AI do pandemii i wojny nuklearnej, wzywając do nadania globalnego priorytetu środkom bezpieczeństwa AI.
- Dyrektor generalny DeepMind, Demis Hassabis, wzywa do

międzynarodowego zarządzania AI na wzór traktatów o nierozprzestrzenianiu broni jądrowej, chociaż napięcia geopolityczne utrudniają współpracę. Tymczasem wyścig zbrojeń w dziedzinie sztucznej inteligencji między Stanami Zjednoczonymi a Chinami wyprzedza wysiłki regulacyjne, co grozi niekontrolowaną eskalacją.
- Chociaż sztuczna inteligencja oferuje przełomowe korzyści (wydajność, przełomy naukowe), niekontrolowany rozwój grozi tym, że AGI przekroczy kontrolę człowieka. Kluczowe pytanie brzmi: czy ludzkość wprowadzi zabezpieczenia, czy też pozwoli, aby sztuczna inteligencja stała się zagrożeniem egzystencjalnym?

Szybki rozwój sztucznej inteligencji (AI) wywołał zarówno entuzjazm, jak i głębokie zaniepokojenie wśród liderów branży, którzy ostrzegają, że sztuczna inteligencja ogólna (AGI) – AI dorównująca lub przewyższająca ludzkie zdolności poznawcze – może pojawić się w ciągu najbliższej dekady.

Dyrektor generalny Google DeepMind, Demis Hassabis, ostrzegł, że AGI może przynieść „katastrofalne skutki", w tym cyberataki na infrastrukturę krytyczną, autonomiczną broń, a nawet zagrożenie egzystencjalne dla ludzkości. Przemawiając podczas szczytu *Axios* AI+ Summit w San Francisco, Hassabis opisał AGI jako system wykazujący „wszystkie zdolności poznawcze" ludzi, w tym kreatywność i zdolność rozumowania.

Ostrzegł jednak, że obecne modele AI pozostają „nierównymi inteligencjami" z lukami w długoterminowym planowaniu i ciągłym uczeniu się. Niemniej jednak zasugerował, że AGI może stać się rzeczywistością dzięki „jednemu lub dwóm kolejnym wielkim przełomom".

Hassabis podkreślił, że niektóre zagrożenia związane ze sztuczną inteligencją już się materializują, szczególnie w zakresie cyberbezpieczeństwa. „Prawdopodobnie dzieje się to już teraz… może jeszcze nie przy użyciu bardzo zaawansowanej

sztucznej inteligencji” – powiedział, wskazując cyberataki na systemy energetyczne i wodociągowe jako „najbardziej oczywisty wektor podatności”.

Jego obawy odzwierciedlają ostrzeżenia całej branży. Ponad 350 ekspertów w dziedzinie sztucznej inteligencji, w tym dyrektor generalny OpenAI Sam Altman, dyrektor generalny Anthropic Dario Amodei oraz pionierzy sztucznej inteligencji Yoshua Bengio i Geoffrey Hinton, podpisali oświadczenie Centrum Bezpieczeństwa Sztucznej Inteligencji, w którym stwierdzono: „Ograniczanie ryzyka wyginięcia spowodowanego przez sztuczną inteligencję powinno być priorytetem na całym świecie, podobnie jak inne zagrożenia na skalę społeczną, takie jak pandemie i wojna nuklearna”.

Niewłaściwe wykorzystanie sztucznej inteligencji: deepfake’i, utrata miejsc pracy i zagrożenia dla bezpieczeństwa narodowego

Oprócz ataków na infrastrukturę, sztuczna inteligencja jest już wykorzystywana do dezinformacji, oszustw i manipulacji za pomocą deepfake’ów. Federalne Biuro Śledcze (FBI) ostrzegło przed oszustwami głosowymi generowanymi przez sztuczną inteligencję, w których oszuści podszywają się pod urzędników państwowych, a jednocześnie mnożą się deepfake’i pornograficzne i dezinformacja polityczna.

Enoch z *BrightU.AI* zauważa, że sztuczna inteligencja stała się technologią transformacyjną, rewolucjonizującą różne sektory, od opieki zdrowotnej po finanse. Jednak, podobnie jak w przypadku każdego potężnego narzędzia, potencjał sztucznej inteligencji do nadużyć i wykorzystania jako broń budzi poważne obawy.

Zdecentralizowany silnik zauważa, że wykorzystanie sztucznej

inteligencji jako broni odnosi się do użycia technologii AI w celu wyrządzenia szkody, uzyskania nieuczciwej przewagi lub manipulowania systemami i ludźmi. Może to przybierać różne formy, w tym broń autonomiczna, deepfake'i i dezinformacja, ocena społeczna i inwigilacja, cyberataki oparte na sztucznej inteligencji oraz rozwój broni biologicznej.

Hassabis przyznał, że chociaż sztuczna inteligencja może wyeliminować wiele miejsc pracy – zwłaszcza stanowiska dla początkujących pracowników umysłowych – nadal bardziej martwi go możliwość wykorzystania sztucznej inteligencji przez złośliwe podmioty do destrukcyjnych celów. „Złośliwy podmiot może wykorzystać te same technologie do szkodliwych celów" – powiedział.

Raport z 2023 r. zlecony przez Departament Stanu USA stwierdza, że sztuczna inteligencja może stanowić „katastrofalne" zagrożenie dla bezpieczeństwa narodowego, i wzywa do wprowadzenia bardziej rygorystycznych kontroli. Jednak w sytuacji, gdy kraje takie jak Stany Zjednoczone i Chiny rywalizują o dominację w dziedzinie sztucznej inteligencji, regulacje prawne pozostają w tyle za postępem technologicznym.

Jak prawdopodobna jest katastrofa związana ze sztuczną inteligencją?

Wśród badaczy zajmujących się sztuczną inteligencją dyskusje często koncentrują się wokół „P(doom)" – prawdopodobieństwa spowodowania przez sztuczną inteligencję katastrofy egzystencjalnej. Hassabis ocenił to ryzyko jako „niezerowe", co oznacza, że nie można go lekceważyć. „Warto bardzo poważnie się nad tym zastanowić i podjąć działania zapobiegawcze" – powiedział, ostrzegając, że zaawansowane systemy sztucznej inteligencji mogą „przekroczyć granice", jeśli nie będą odpowiednio kontrolowane.

Hassabis opowiada się za międzynarodowym porozumieniem w sprawie bezpieczeństwa sztucznej inteligencji, podobnym do traktatów o nierozprzestrzenianiu broni jądrowej. „Oczywiście w obecnej sytuacji geopolitycznej wydaje się to trudne” – przyznał, ale podkreślił, że współpraca jest niezbędna, aby zapobiec nadużyciom.

Tymczasem giganci technologiczni nadal dążą do integracji sztucznej inteligencji z codziennym życiem. Google wyobraża sobie „agentów” sztucznej inteligencji pełniących rolę osobistych asystentów, zajmujących się zadaniami od planowania harmonogramów po rekomendacje. Hassabis ostrzegł jednak, że społeczeństwo musi dostosować się do zmian gospodarczych spowodowanych przez sztuczną inteligencję, sprawiedliwie redystrybuując zyski z wydajności.

Potencjał sztucznej inteligencji jest niezaprzeczalny – zwiększa wydajność, przyspiesza odkrycia i transformuje branże. Jednak ryzyko z nią związane jest równie poważne. Jak ostrzegają Hassabis i inni eksperci, bez pilnych zabezpieczeń AGI może wymknąć się spod kontroli człowieka, a konsekwencje tego mogą być porównywalne z pandemią lub wojną nuklearną.