

Global Research, Otvorený list vedcov: Pozastavte GIGANTICKY NEBEZPEČNÉ EXPERIMENTY S UMELOU INTELIGENCIOU!

- CZ24 News | 31. března 2023

Vyzývame všetky laboratóriá AI, aby okamžite prerušili aspoň na 6 mesiacov tréning systémov umelej inteligencie (AI) výkonnejších ako GPT-4.

Systémy AI s ľudskou konkurenčnou inteligenciou môžu predstavovať vážne riziká pre spoločnosť a ľudstvo, ako ukázal rozsiahly výskum ^[1] a potvrdili to špičkové laboratóriá AI. ^[2]

Ako sa uvádza v široko uznávaných [princípoch umelej inteligencie Asilomar](#), pokročilá umelá inteligencia by mohla predstavovať hlbokú zmenu v histórii života na Zemi a mala by byť plánovaná a riadená s primeranou starostlivosťou a zdrojmi.

Žiaľ, takáto úroveň plánovania a riadenia sa neuskutočňuje, aj keď v posledných mesiacoch boli laboratóriá AI zamknuté v pretekoch, ktoré sa vymkli kontrole, aby vyvinuli a nasadili stále výkonnejšie digitálne mysle, ktorým nikto – dokonca ani ich tvorcovia – nerozumie. predvídať alebo spoľahlivo kontrolovať.

Súčasný systémy AI sa teraz stávajú konkurencieschopnými pre ľudí vo všeobecných úlohách ^[3] a musíme si položiť otázku: *Mali by sme nechať stroje zaplaviť naše informačné kanály propagandou a nepravdou?*

Mali by sme automatizovať všetky úlohy, vrátane tých, ktoré naplňajú?

Mali by sme rozvíjať neľudské mysle, ktoré by nás mohli časom prevýšiť, prekabátiť, zastarať a nahradiť nás?

Mali by sme riskovať stratu kontroly nad našou civilizáciou? Takéto rozhodnutia sa nesmú delegovať na nevolených technologických lídrov.

Výkonné systémy umelej inteligencie by sa mali vyvíjať len vtedy, keď sme si istí, že ich účinky budú pozitívne a ich riziká budú zvládnuteľné.

Táto dôvera musí byť dobre odôvodnená a musí sa zvyšovať s veľkosťou potenciálnych účinkov systému. Nedávne vyhlásenie Open AI [týkajúce sa umelej všeobecnej inteligencie](#) uvádza, že „V určitom bode môže byť dôležité získať nezávislú kontrolu pred začatím tréningu budúcich systémov a pre čo najpokročilejšie úsilie dohodnúť sa na obmedzení tempa rastu výpočtovej techniky používanej na vytváranie nových modely.“ Súhlasíme. Ten bod je teraz.

Musia byť dôveryhodné a lojálne

Preto **vyzývame všetky laboratóriá AI, aby okamžite prerušili aspoň na 6 mesiacov tréning systémov AI výkonnejších ako GPT-4.**

Táto pauza by mala byť verejná a overiteľná a mala by zahŕňať všetkých kľúčových aktérov. Ak sa

takáto pauza nedá uzákoníť rýchlo, vlády by mali zasiahnuť a zaviesť moratórium.

Laboratória AI a nezávislí experti by mali využiť túto pauzu na spoločný vývoj a implementáciu súboru zdieľaných bezpečnostných protokolov pre pokročilý návrh a vývoj AI, ktoré sú prísne kontrolované a kontrolované nezávislými externými odborníkmi.

Tieto protokoly by mali zabezpečiť, že systémy, ktoré sa k nim pridržajú, sú bez akýchkoľvek pochybností bezpečné. ^[4] *Neznamená to pauzu vo vývoji AI vo všeobecnosti, iba krok späť od nebezpečných pretekov k stále väčším nepredvídateľným modelom čiernej skrinky s novými schopnosťami.*

Výskum a vývoj AI by sa mal preorientovať na to, aby boli dnešné výkonné, najmodernejšie systémy presnejšie, bezpečnejšie, interpretovateľné, transparentné, robustné, zosúladené, dôveryhodné a lojálne.

Súčasne musia vývojári AI spolupracovať s tvorcami politiky, aby dramaticky urýchlili vývoj robustných systémov riadenia AI.

Mali by zahŕňať minimálne:

- nové a schopné regulačné orgány venované AI;
- dohľad a sledovanie vysoko výkonných systémov AI a veľkých zdrojov výpočtových schopností;
- systémy proveniencie a vodoznaku, ktoré pomáhajú rozlíšiť skutočné od syntetických a sledovať úniky modelov;
- robustný audítorský a certifikačný ekosystém;
- zodpovednosť za škody spôsobené AI; r
- ozsiahle verejné financovanie výskumu technickej bezpečnosti AI;
- a dobre vybavené inštitúcie na zvládanie dramatických ekonomických a politických narušení (najmä demokracie), ktoré AI spôsobí.

Ludstvo sa môže tešiť z prosperujúcej budúcnosti s AI. Po úspešnom vytvorení výkonných systémov umelej inteligencie si teraz môžeme užiť „leto umelej inteligencie“, v ktorom žneme odmeny, vytvárame tieto systémy pre jasný prospech všetkých a dávame spoločnosti šancu prispôbiť sa. Spoločnosť sa pozastavila nad ďalšími technológiami s potenciálne katastrofálnymi dopadmi na spoločnosť. ^[5] Môžeme tak urobiť tu. Užime si dlhé AI leto, neponáhľajme sa nepripravení na jeseň.

AUTOR: Future of Life Institute

[ZDROJ](#)

Překlad: S.S.