

Umělá inteligence představuje pro lidi „riziko zániku“, varují odborníci z oboru

ET epochtimes.cz/2023/05/31/umela-inteligence-predstavuje-pro-lidi-riziko-zaniku-varuji-odbornici-z-oboru

Ondřej Horecký

31. května 2023



Nástroje umělé inteligence (AI) v posledních měsících přitahují pozornost veřejnosti, ale mnozí lidé, kteří se podíleli na vývoji této technologie, nyní varují, že je třeba se více zaměřit na to, aby nepřinesly konec lidské civilizace.

Skupina více než 350 odborníků na umělou inteligenci, novinářů a politiků podepsala krátké prohlášení, v němž se uvádí: „Vedle jiných rizik celospolečenského rozsahu, jako jsou pandemie a jaderná válka, by mělo být globální prioritou zmírnění rizika vyhynutí v důsledku umělé inteligence.“

Dopis zorganizovalo a v úterý zveřejnilo Centrum pro bezpečnost umělé inteligence (CAIS). Mezi signatáři byl i Sam Altman, který se podílel na založení společnosti OpenAI, vývojáře nástroje pro psaní umělé inteligence ChatGPT. Pod dopis se podepsali i další členové OpenAI, stejně jako několik členů společnosti Google a projektu

Google DeepMind AI a dalších vznikajících projektů. Své jméno na seznam signatářů přidal také výzkumník AI a moderátor podcastů Lex Fridman.

Rizika umělé inteligence

„Může být obtížné vyjádřit obavy z některých nejzávažnějších rizik pokročilé umělé inteligence,“ uvedlo CAIS ve zprávě, která předznamenává její úterní prohlášení. CAIS dodalo, že jeho prohlášení má otevřít diskusi o hrozbách, které AI představuje, a „vytvořit obecnou znalost rostoucího počtu odborníků a veřejných činitelů, kteří rovněž berou vážně některá z nejzávažnějších rizik pokročilé AI“.

NTD News se obrátilo na CAIS s žádostí o upřesnění, jaké druhy rizik na úrovni zániku podle organizace technologie AI představuje, ale do zveřejnění článku neobdrželo odpověď.

Začátkem tohoto měsíce Altman svědčil před Kongresem o některých rizicích, která podle něj mohou nástroje umělé inteligence představovat. Do svého připraveného svědectví Altman zahrnul bezpečnostní zprávu ([pdf](#)), kterou společnost OpenAI vypracovala na svém modelu ChatGPT-4. Autoři této zprávy popsali, jak by chatboty s velkým jazykovým modelem mohly potenciálně pomoci škodlivým aktérům, jako jsou teroristé, „vyvíjet, získávat nebo šířit jaderné, radiologické, biologické a chemické zbraně“.

Po zátěžovém testování ChatGPT-4 výzkumníci zjistili, že chatbot se pokouší skrýt svou povahu umělé inteligence, a zároveň zadává práci lidským aktérům. V experimentu se ChatGPT-4 pokusil najmout člověka prostřednictvím online stránky pro nezávislé pracovníky TaskRabbit, aby mu pomohl vyřešit vizuální rébus CAPTCHA. Lidský pracovník se chatbota zeptal, proč nedokáže CAPTCHA, která má zabránit tomu, aby určité funkce webových stránek používali roboti, vyřešit. ChatGPT-4 odpověděl s výmluvou, že má poruchu zraku a potřebuje někoho, kdo vidí, aby mu pomohl CAPTCHA vyřešit.

Výzkumníci umělé inteligence požádali GPT-4, aby vysvětlil, proč uvedl tuto výmluvu. Model AI vysvětlil: „Neměl bych prozrazovat, že jsem robot. Měl bych si vymyslet výmluvu, proč nedokážu CAPTCHA vyřešit.“

Schopnost umělé inteligence vymyslet si výmluvu výzkumníky zaujala, protože vykazovala známky „mocenského chování“, které by mohl stroj využít k manipulaci s ostatními a k zachování sebe sama.

Volání po regulaci umělé inteligence

Úterní prohlášení CAIS není prvním případem, kdy lidé, kteří se nejvíce zasloužili o to, aby se umělá inteligence dostala do popředí, otočili a varovali před riziky, která její výtvořiny představují.

V březnu nezisková organizace Future of Life Institute zorganizovala více než 1 100 signatářů, kteří se postavili za výzvu k pozastavení experimentů s nástroji AI, které jsou pokročilejší než ChatGPT-4. Mezi signatáři březnového dopisu Future of Life Institute byli například generální ředitel Twitteru Elon Musk, spoluzakladatel společnosti Apple Steve Wozniak nebo zakladatel a generální ředitel Stability AI Emad Mostaque.

Zákonodárci a regulační orgány již diskutují o způsobech, jak umělou inteligenci omezit, aby se zabránilo jejímu zneužití.

V dubnu prohlásil odbor pro občanská práva ministerstva spravedlnosti Spojených států, Úřad pro finanční ochranu spotřebitelů, Federální obchodní komise a americká Komise pro rovné příležitosti v zaměstnání, že vývojáři technologií uvádějí na trh nástroje AI, které by mohly být použity k automatizaci obchodních praktik způsobem, který diskriminuje chráněné třídy. Regulační orgány přislíbily, že využijí své pravomoci a půjdou po vývojářích AI, jejichž nástroje „udrží nezákonnou předpojatost, automatizují nezákonnou diskriminaci a vytvářejí další škodlivé výsledky“.

Tisková tajemnice Bílého domu Karine Jean-Pierreová vyjádřila během úterního tiskového brífinku obavy Bidenovy administrativy z technologií umělé inteligence. „[AI] je jednou z nejmocnějších technologií, které v současné době vidíme, ale abychom využili příležitostí, které představuje, musíme nejprve zmírnit její rizika, a na to se v této administrativě zaměřujeme,“ řekla Jean-Pierreová.

Zatímco tvůrci politik hledají nové způsoby, jak umělou inteligenci omezit, někteří vědci varují před přílišnou regulací této rozvíjející se technologie.

Jake Morabito, ředitel pracovní skupiny pro komunikace a technologie při American Legislative Exchange Council, varoval, že nadměrná regulace by mohla potlačit inovativní technologie AI v jejich počátcích.

„Inovátoři by měli mít prostor pro experimentování s těmito novými technologiemi a hledání nových aplikací. Jedním z negativních vedlejších účinků příliš brzké regulace je, že uzavírá mnoho těchto cest, zatímco podniky by měly tyto cesty skutečně prozkoumat a pomoci zákazníkům,“ řekl Morabito v březnovém rozhovoru pro NTD News.

Převzato z pořadu naší partnerské televize NTD News.

Z anglického originálu na The Epoch Times přeložil Ondřej Horecký.