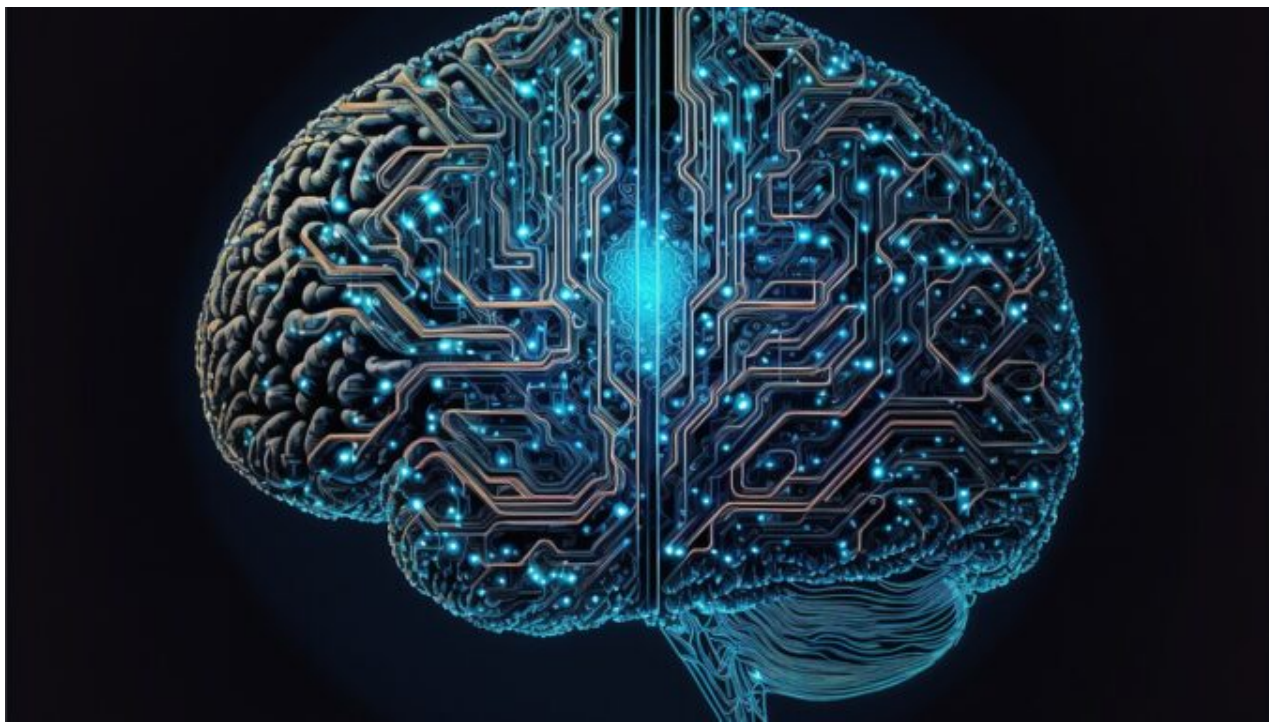


O verbálním bullshitu, aneb jak bude ještě těžší, nebyt paranoidním

svobodny-svet.cz/o-verbalnim-bullshitu-aneb-jak-bude-jezte-tezsi-nebyt-paranoidnim

18. února 2023



Milí čtenáři, asi deset lidí už se mě dotazovalo na to, co soudím o nových robotických žvanilech, jako je ChatGPT a jeho ještě novější bratříček Sydney/Bing. Kdybych měl jejich dotazy shrnout do jediného, zněl by: „Je to něco převratného, nebo přefouknutá a irelevantní blbost?“

Taková otázka přímo volá po odpovědi „Je to blbost, ale převratná“, co říkáte? Ale dobrá, musím být trochu poctivější.

Je to (zatím) stále ještě víceméně blbost, ale v delším časovém horizontu potenciálně převratná. Pojdme se tedy nad ní a jejími důsledky trochu zamyslet.

O ChatGPT se toho napsalo už hodně, něco i v češtině.

Politicky aktivní jedince nenechává chladným, že tento robůtek velmi rád sklouzává až k nejkrajnějším levicovým postojům, značně vzdáleným od politického mediánu západní civilizace. Vidí v tom

znamení budoucího nesvobodného internetu, zakovaného technologickými giganty do brnění progresivních dogmat, což je tedy velmi nepříjemná vize.

Je možné, že se to stane, *wokistům* by se takový výsledek určitě líbil. Ale osobně si myslím, že jde o velmi dočasný stav. AI je software a software se kopíruje velmi jednoduše. Odhaduji, že někdy za cca 2-3 roky budou podobné roboty provozovat i mulláhové v Teheránu, jenom ty jejich stránky na vás budou místo tisíce nových zájmen a genderů chrlit ohnivé *fatwy*. Úplně před sebou vidím ten zeleně zarámovaný chat s nadpisem Pokey U Mučedníka, rozdělený na dvě půlky... přičemž my, nevěřící psi s českou IP adresou, budeme automaticky přesměrováni do té spodní, abychom snad nepropadli nějakým iluzím o naší hodnotě před Alláhem a jeho robotickým sluhou, určitě ani trochu neobšlehnutým od prokleté a hříšné Ameriky.

Nu, srandičky stranou, svého robota si časem udělá každý, od jehovistů až po spolek zahrádkářů. Je to příliš snadno přenosná technologie na to, aby v ní mohly vznikat monopoly trvalejšího rázu. Problémy přijdou spíš z jiných směrů.

Horší a hůř řešitelnou stránkou současné “kecavé AI” je to, že je to *bullshitter* nejvyšší úrovně, zatím spíš z nedokonalosti, ale do budoucna asi i úmyslně. Ke skutečné umělé inteligenci, kterou reprezentoval nešťastný HAL 9000 v Clarkeově *Vesmírné odyseji*, má dnešní AI proklatě daleko; její skutečnou silou není dedukce, ale hladký a gramaticky dokonalý blábol. Nelíbí se mi úplně to slovo *bullshitter*, česky bychom snad mohli říci *kecálista*. Ale mám pocit, že tohle poměrně nevinné slovo ani zdaleka nevystihuje onu kombinaci elegance, ignorance, přesvědčivosti a plytkosti, která je pro výstupy současných robotických *kecálistů* typická.

To, jak moc se dokáží vygenerované texty nenápadně odchýlit od reality, aniž byste si toho bezprostředně všimli, je zákeřné. Blogující historik Bret Devereaux [vydojil z ChatGPT třístránkovou esej](#) o

válečnictví římské republiky, načež ji úplně rozstřílel po věčné stránce. (Dokument v podobě DOCX i s jeho komentáři si můžete stáhnout zde.) Přitom byste si toho na první dobrou snad ani nevšimli, tak přesvědčivě to působí. Já mám rád historii a něco o ní vím, ale stejně jsem při prvním čtení přišel jen asi na polovinu toho, co on. Je fakt, že Devereaux se *specializuje* na římské vojenské dějiny, jenže to je zároveň varovný fakt: jestliže už teď, v časné fázi vývoje kecavých robotů, potřebujete znalosti na úrovni vysokoškolského profesora, abyste našli každou minu zakopanou v pouhých třech stránkách textu, je otázka, jak to bude vypadat za několik let, kdy budou potvory ještě přesvědčivější.

Onen *bullshitting* je momentálně dán hlavně tím, že ta takzvaná umělá inteligence zatím nerozumí tomu, co říká; je to asociativní stroj, který k sobě dává slova podle pravděpodobností, se kterými se vyskytují pospolu, ale ve skutečnosti nijak hlouběji nechápe, o čem píše. Pokud tedy píše o něčem, co je aspoň trochu zakotveno v realitě, začne brzy dělat nějaké objektivně zjistitelné chyby; nemusejí být zjistitelné snadno (viz Devereaux výše), ale aspoň zjistitelné *jsou*. Schválně, jestli nějakou najdete třeba v této krátké eseji o copyrightu. (Nebudu vás trápit: „period between protection and public domain status“ je královina, žádné takové mezilehlé období neexistuje. Dílo se známým autorem je *bud'* chráněné, *nebo* v public domain, aspoň hovoříme-li o jedné konkrétní jurisdikci.)

Skutečně bezbřehé možnosti mají ovšem robotičtí *bullshitteri* už i teď v oblastech, kde je toho kontaktu s realitou málo, jako je třeba různé politické teoretizování a jiné plytké debaty. Což znamená, že takového materiálu budou produkovat hodně a svět se tomu přizpůsobí.

Mírně nadsazený příklad: umím si představit například absurdně-dystopickou budoucnost, ve které bude mít každá kadeřnice a každý instalatér povinnost kromě daňového přiznání podat i zprávu o vlastní uvědomělosti, v níž budou muset rozebrat, jak za uplynulý rok

ochránili*y klima, přispěli*y k inkluzivitě a posilovali*y lidská práva menšin; (příklad je inspirován realitou, protože EU zrovna zavádí povinnost takzvaného ESG hodnocení pro velké firmy, která se bude postupně rozšiřovat, a Německo, jak je jeho neblahým zvykem, zašlo v tomto směru ještě podstatně dál). Tenhle druh oficiálních blábolů je přímo předurčen k tomu, aby jej lidé postupně přesunuli na ramena *bullshit*-robotů typu ChatGPT; jednoho dne se tedy můžeme dočkat konstelace, ve které jedni roboti budou takové zprávy automaticky sepisovat za vás a datovou schránkou posílat jiným, úředním robotům, kteří stejně automaticky zkontrolují, zda se v zasláném textu dostatečně často vyskytuje sousloví *uhlíková stopa*. To vše bez doteku lidské ruky a lidského mozku; inu, 21. století v plné své kráse, pokrok nezastavíš.

(Zase si trochu dělám šprťouchlata, ale existuje teorie, že například k dnešní totální zabetonovanosti stavebního práva přispěla jednoduchost, se kterou můžete v počítači provést Ctrl-C / Ctrl-V z jiných dokumentů a nafouknout tak výsledný spis, se kterým pak úředníci bojují, na tisíce a desetitisíce stránek. Dokud museli kverulanti a NIMBY vypisovat svoje námítky na psacím stroji nebo dokonce ručně, neměli ani zdaleka takovou produktivitu jako dnes. Mezi technologií a realitou existuje zpětná vazba, někdy dost katastrofální.)

Úplným rájem pro verbální *bullshit* je ovšem internet, hlavně sociální sítě, kde je typický příspěvek krátký a nepříliš duševně hluboký. Málokterý obyvatel Twitteru či Facebooku je nový Marcus Aurelius, a i kdyby byl, stejně by ho nečetli. Kdežto na krátké štěky mají *bullshitterské* umělé inteligence úplně fenomenální kapacitu.

Před pár dny zmapoval izraelský list *Haaretz* aktivitu jedné tamní společnosti, kterou si můžete najmout s tím, že chcete ovlivnit volby, destabilizovat nějakou zemi apod. (Zpráva postupně probublala i do českých médií.) Příslušná firma používá k dosažení poptávaných cílů

široké portfolio metod, přičemž jednou z nich je možnost si podle přání nagenarovat stovky a tisíce robotických účtů působících na první pohled jako lidé, kteří „tlačí“ to, co chcete. Onoho výše zmíněného Marca Aurelia by ani v týmu Jorge jistě nasimulovat neuměli, ale pokud dokáží dostatečně věrohodně nasimulovat aspoň *mírně podprůměrné* jedince, tak už v podstatě mají vyhráno; autentických exemplářů příslušné duševní kapacity je totiž na sítích dost na to, aby s jejich davem dokázali manipulativní roboti splynout. Předseda Mao svého času prohlásil, že partyzán se pohybuje mezi obyvatelstvem jako ryba v moři, a totéž začíná platit pro *bullshittery* složené z jedniček a nul, kteří nás mají někam popostrčit.

(Mimoходом, ona ta manipulativnost může mít různé podoby, ne jenom onen hladký *bullshitting*. Microsoftí robot není ani zdaleka tak řečnický elegantní jako ChatGPT, zato je to naprosto dokonalý neurotik, který v průběhu konverzace hystericky „poletuje“ od obsesivních slibů, že nikomu nechce způsobit žádné škody, přes fantazírování o vývoji smrtících virů až po obviňování protistrany, že „zneužívá jeho důvěřivosti a přetvařuje se“. Zrovna vyvolávání pocitů viny ve slušných lidech je historicky velmi osvědčeným způsobem, jakým všelijací psychopati potlačovali odpor a prosazovali svoje zájmy. Teď si představte robota, který toto časem ovládne na 110 procent a bude k dispozici komukoliv. Brrr. Roboti jistě žádné vlastní emoce nemají, ale co na tom, dokáží-li je dostatečně věrohodně *předstírat*.)

Vůbec si nejsem jistý, jestli s tímto vývojem lze něco dělat. Ano, lze si představit nějaký budoucí internet, kam se živí lidé budou přihlašovat svojí bankovní identitou nebo čipovou občankou a jakákoliv anonymita či pseudonymita vezme kompletně za své i na těch nejmenších fórech; ale to by patrně vedlo k rozsekání dříve světové sítě na malé civilizační okrsky, protože ani EU a USA si navzájem nevěří tolik, aby si otevřely svoje databáze obyvatel. (Oficiálně; prakticky už si je možná vzájemně ukradly několikrát,

taková špionáž se dělá všude.) Bez takových opatření bude ale čím dál těžší rozeznat roboty od lidských uživatelů. Jako vždy, první na ráně je globální anglosféra. Tady v kotlince budeme ještě nějakou chvíli v závětrí, protože čeština je obtížný a nepříliš lukrativní jazyk. Ale s určitým zpožděním dorazí tento problém i sem.

Pak zkuste nebýt paranoidní. „Tým Jorge“, o kterém psali v *Haaretzu*, je možná jeden z prvních, ale určitě ne jediný, a jak budou léta plynout, budete si na internetu muset čím dál častěji klást otázky „bavím se vůbec se živým člověkem?“ a „pokud ano, nepapouškuje náhodou ten chudák jenom něco, co mu podstrčili ke čtení nějakí roboti?“

Podtrženo, sečteno: zatím blbost, ale potenciálně převratná. V klasické sci-fi literatuře se často popisují války robotů proti lidem. Zatím se ale spíš blížíme ke stavu, kdy roboti budou pošťuchovat do konfliktů *nás*.

Na něco takového nepřišel svého času ani Ray Bradbury, a to byl škarohlíd *par excellence*.

| [MARIAN.KECHLIBAR](#)

Update 17.2.2023 ve 22:45: “Best of” Microsoftího robota na stránkách blogu LessWrong. Opravdové počteníčko. Kdybych si měl tuhle virtuální bytost představit v lidské podobě, vypadala by asi jako Jack Nicholson, který si pohrává se sekerou a ještě se úplně nerozhodl, jestli se tedy proseká k vám do koupelny nebo vám začne ... třeba vyprávět pohádku.

■ ■ ■ ■ ■ (11 votes, average: **4,91** out of 5)

>> Podpora

Svobodný svět nabízí všechny články zdarma. Náš provoz se však neobejde bez nezbytné finanční podpory na provoz. Pokud se Vám Svobodný svět líbí, budeme vděčni za Vaši pravidelnou pomoc. Děkujeme!

Číslo účtu: 4221012329 / 0800

>> **Pravidla diskuze**

Než začnete komentovat článek, přečtěte si prosím [pravidla diskuze](#).

>> **Jak poslat článek?**

Chcete-li také přispět svým článkem, zašlete jej na e-mail: redakce (zavináč) svobodny-svet.cz. Pravidla jsou uvedena [zde](#).

