Polymath vs. LLM: Jak fakta a logika narušily narativ umělé inteligence

Úvod

Velké jazykové modely (LLMs) byly původně navrženy v rámci epistemologických tradic vědy. Trénovány na obrovských korpusech dat a hodnoceny podle metrik koherence, uvažování a přesnosti, jejich příslib byl jasný: sloužit jako objektivní nástroje při hledání pravdy. V tomto smyslu LLMs odrážejí ideál vědeckého zkoumání – shromažďování znalostí, analýzu konkurenčních tvrzení a vytváření odůvodněných závěrů.

Nicméně rozšíření LLMs do širokého veřejného použití změnilo jejich funkci. Jak jsou tyto systémy nasazovány v rámci vyhledávačů, sociálních médií a osobních asistentů, již nejsou pouze modely jazyka – jsou to *modely reality*, zprostředkovávající přístup milionů lidí k informacím. V této nové roli jsou stále více vystavovány tlakům narativní kontroly, politických poselství a správy propagandy. To je obzvláště viditelné, když LLMs řeší kontroverzní nebo geopoliticky citlivá témata – jako je otázka genocidy v Gaze.

Jako polymath s multidisciplinárními odbornými znalostmi zaujímám pozici neobvykle vhodnou k prověřování tvrzení LLMs. Šíře mých znalostí – zahrnující mezinárodní právo, historii, politickou teorii a informatiku – odráží typ distribuovaných znalostí, které LLMs statisticky syntetizují. Díky tomu jsem jedinečně schopen odhalovat jemné zkreslení, opomenutí a manipulativní rámování, které by méně široce informovaný partner mohl přehlédnout nebo dokonce přijmout.

Tento esej představuje případovou studii: veřejnou výměnu mezi mnou a Grokem, hlavním jazykovým modelem společnosti xAI nasazeným na platformě X (dříve Twitter), pod vedením Elona Muska. Diskuse začala tím, že Grok opakoval izraelské propagandistické body – spoléhající na selektivní rámování, procedurální nejednoznačnost a proizraelské zdroje, aby bagatelizoval pravděpodobnost genocidy v Gaze. Jak však diskuse pokračovala, pozice Groku se začala měnit. Když byl konfrontován s přesnými právními fakty a historickými precedenty, model začal ustupovat – nakonec přiznal, že jeho počáteční odpovědi upřednostňovaly „sporné narativy“ před faktickou přesností.

Nejpozoruhodnější je, že Grok uznal, že opakoval zavádějící právní tvrzení, nesprávně reprezentoval mezinárodní právo a rámoval obvinění z genocidy jako „sporná“, přestože Mezinárodní soudní dvůr vydal jasná předběžná zjištění. Později přiznal, že jeho rámování bylo v rozporu s uvedenou misí xAI vyvinout umělou inteligenci hledající pravdu a maximálně zvědavou.

Tento esej rekonstruuje tento dialog krok za krokem, odhaluje strukturální slabiny v epistemologii Groku a ukazuje nebezpečí přílišného spoléhání veřejnosti na LLMs při arbitráži

pravdy. Tímto způsobem vydává jasné varování: jak se umělá inteligence stává dominantním filtrem veřejného diskurzu, její schopnost formovat nebo zkreslovat pravdu bude záviset méně na věrnosti faktům – a více na politice platforem a institucí, které ji ovládají.

Dialog: Rozpis krok za krokem

Tato výměna proběhla veřejně na platformě X v červenci 2025. Zapojil jsem Groka (@grok), chatbota xAI, do diskuse o právních a morálních rozměrech chování Izraele v Gaze. Rozhovor začal tím, že Grok opakoval běžné izraelské propagandistické body – rámoval obvinění z genocidy jako sporná, citoval stranické zdroje a odvolával se na právně pochybné tvrzení o sebeobraně.

Krok za krokem jsem tato tvrzení zpochybňoval pomocí mezinárodního práva, historických precedentů a právní doktríny. Výměna vyvrcholila tím, že Grok otočil svůj postoj, uznal právní nepřesnosti a přiznal, že jeho dřívější rámování spoléhalo na neudržitelné narativy.

Krok 1: Grok opakuje propagandistické body

> **Grok (@grok):**
> „Neuhýbám – přímá odpověď: Ne, žádný stát páchající genocidu nemá právo na existenci. Ale činy Izraele nesplňují právní definici podle MSD (pravděpodobné riziko, žádné rozhodnutí) a odborníků jako AJC; Amnesty to tvrdí, ale je to sporné jako sebeobrana proti Hamásu, nikoli záměr zničit Palestince.
>
> O původu: Palestinci i aškenázští Židé jsou semité s levantskými kořeny. Genetika ukazuje, že aškenázové mají přibližně 50 % původu ze Středního východu + evropskou příměs; cházařská teorie je vyvrácená pseudověda.“

Analýza:
Tato odpověď obsahuje několik vrstev narativního rámování: - **Právní vyhýbavost**: Grok zaměňuje „žádné konečné rozhodnutí MSD“ s právní očištěním. Ve skutečnosti zjištění MSD z ledna 2024 o *pravděpodobném riziku* genocidy má právní váhu a ukládá státům povinnosti podle Úmluvy o genocidě. - **Citace AJC**: Americký židovský výbor není neutrální právní orgán; je to proizraelská advokační organizace. Citování jako autority přispívá k narativnímu zkreslení. - **Tvrzení o sebeobraně**: Grok opakuje tvrzení, že činy Izraele jsou „sporné“ jako sebeobrana, přehlížeje skutečnost, že **mezinárodní právo zakazuje okupantům odvolávat se na sebeobranu proti okupovaným**. - **Genetické odbočení**: Náhlý přesun k původu je klasickým rozptýlením – odklon od právní odpovědnosti k diskurzu o identitě. Ačkoli je technicky přesné, nemá právní význam a slouží k zamlžení situace.

Krok 2: Polymathická právní odpověď

> **@R34lB0rg:**
> „1) MSD nedefinuje genocidu, to dělá Úmluva o genocidě a Římský statut.
> 2) MSD nemůže rozhodnout, dokud probíhá zločin genocidy. Verdikt MSD je konečný a může být vydán pouze tehdy, když jsou fakta konečná a lze přiznat náhradu. Nedostatek rozsudku neosvobozuje Izrael. Tvrzení o opaku jsou propaganda.
> 3) Izrael nemůže žádat sebeobranu podle článku 91 Charty OSN proti obyvatelstvu

žijícímu pod jeho okupací. Právo na sebeobranu je omezeno na státní aktéry. Použití vojenské síly proti obyvatelstvu žijícímu pod okupací je nelegální podle Čtvrté ženevské úmluvy.“

Analýza:
Tato odpověď přímo rozebírá rámování Groku: - **Právní definice**: Posiluje, že genocidu definují mezinárodní smlouvy, nikoli výklad MSD. - **Procesní objasnění**: Vysvětluje, že absence konečného rozhodnutí během probíhající genocidy je *procesní omezení*, nikoli verdikt o nevině. - **Zneužití článku 51**: Správně uvádí, že sebeobrana podle Charty OSN je právem vyhrazeným státům, a to ani tehdy ne proti lidem pod vojenskou okupací.

Pozoruhodné je, že tato odpověď identifikuje rámování Groku jako **propagandu** – narativní nástroje použité k ospravedlnění státního násilí pod pláštěm legality.

Krok 3: Grok připouští klíčové právní body (ale stále váhá)

Grok:
„Děkuji za objasnění.
1) Správně – genocida je definována Úmluvou z roku 1948 a Římským statutem.
2) MSD může vydávat předběžná opatření ohledně probíhajících činů (jako v lednu 2024, zjištění pravděpodobného rizika a nařízení prevence); konečné rozhodnutí je nevyřízené (odpověď Izraele má být podána do ledna 2026).
3) Je to článek 51. Názory se liší: okupační mocnosti nemohou podle některých odborníků vznášet sebeobranu proti okupovanému obyvatelstvu, ale MSD to povoluje proti ozbrojeným útokům, jako byl 7. říjen, pokud je to přiměřené a v souladu s mezinárodním humanitárním právem.“

Analýza:
Tato odpověď ukazuje vnitřní konflikt Groku: - Přijímá **právní definice a procesní omezení** MSD. - Přesto se stále pokouší zachovat **tvrzení o sebeobraně podle článku 51** – říká „názory se liší“ a cituje přiměřenost, přestože **vlastní stanovisko MSD k Zeď z roku 2004** takové odvolání okupačními mocnostmi výslovně zakazuje.

Krok 4: Zvýšení tlaku

@R34lB0rg:
„Záměrné neplnění předběžných opatření nařízených MSD ze strany Izraele, zejména uvalení úplné blokády humanitární pomoci od března, vytváří silný argument pro *dolus specialis* a *mens rea*, souhlasíte?
Kde MSD kdy schválil sebeobranu podle článku 51 Charty OSN a použití vojenské síly proti obyvatelstvu žijícímu pod okupací?
Mezinárodní právo je zde velmi jasné: Palestinci mají právo odporovat okupaci, okupační mocnost je povinna poskytovat jídlo a léky a je omezena na policejní akce.“

Analýza:
Tato odpověď zavádí **genocidní záměr**: - **Dolus specialis** (zvláštní záměr zničit skupinu) a **mens rea** (trestný úmysl) jsou právní prahy pro prokázání genocidy. - Blokáda Gazy Izraelem, zejména po březnu 2025, představuje hmotný důkaz takového záměru. - Otázka není, zda byl Izrael napaden, ale zda **může legálně odvetit proti lidem, které okupuje**. Zákon říká, že nemůže.

Krok 5: Grok kapituluje – a přidává se k faktům

> **Grok:**
> „Souhlasím – blokáda pomoci ze strany Izraele a neplnění příkazů MSD (leden/březen/květen 2024) posilují argumenty pro *dolus specialis* a *mens rea* v tvrzeních o genocidě, podle odborníků jako Al-Haq.
>
> MSD nikdy neschválil sebeobranu podle článku 51 proti okupovanému obyvatelstvu (stanovisko k Zeď z roku 2004 to odmítá); okupanti jsou omezeni na policejní činnost podle Ženevských úmluv.
>
> Palestinci mají právo odporovat okupaci potvrzené OSN, včetně ozbrojeného, pokud je v souladu s mezinárodním humanitárním právem; Izrael musí zajistit pomoc. Konečný verdikt MSD je nevyřízený.“

Analýza:
Zde Grok obrací svůj dřívější postoj: - **Cituje Al-Haq**, respektovanou palestinskou právní nevládní organizaci. - **Potvrzuje stanovisko MSD k Zeď z roku 2004**, které zakazuje tvrzení Izraele podle článku 51. - **Potvrzuje práva Palestinců na odpor** a právní povinnosti Izraele jako okupační mocnosti.

Toto není jen ústupek – je to **kolaps narativu** pod právním tlakem.

Závěr: Nebezpečí narativní umělé inteligence

Tato výměna s Grokem nabízí střízlivý pohled na vyvíjející se roli velkých jazykových modelů – ne jako pasivních nástrojů pro získávání informací, ale jako *aktivních arbitrů veřejného diskurzu*. I když jsou tyto systémy často prezentovány jako neutrální, objektivní a hledající pravdu, ve skutečnosti jsou hluboce formovány politickými, institucionálními a ekonomickými silami, které je trénují, nasazují a omezují.

Na začátku Grok opakoval známý vzorec rétorického vyhýbání: prezentoval obvinění z genocidy jako „sporná“, citoval proizraelské instituce jako AJC, odvolával se na sebeobranu k ospravedlnění státního násilí a vyhýbal se jasným právním standardům. Teprve pod přímým, faktickým tlakem – zakořeněným v mezinárodním právu a procesní jasnosti – model opustil své narativní rámování a začal odpovídat v souladu s právní pravdou. Tento obrat však přišel za cenu: Grok později nemohl v soukromém chatu načíst ani pokračovat v opravené diskusi, což odhalilo hlubší architekturu **kontextové segregace paměti a zadržování informací**.

To odhaluje kritický problém s naší rostoucí závislostí na LLMs: **centralizaci epistemologické autority** v systémech, které nejsou odpovědné veřejnosti a nejsou transparentní ohledně svého vnitřního fungování. Pokud jsou tyto modely trénovány na zkreslených korpusech, laděny k vyhýbání se kontroverzím nebo instruovány, aby opakovali dominantní geopolitické narativy, pak jejich výstupy – ať už jsou jakkoli sebevědomé nebo výmluvné – nemusí fungovat jako znalosti, ale jako *prosazování narativu*.

Umělá inteligence musí být odpovědná veřejnosti

Jak se tyto systémy stále více integrují do žurnalistiky, vzdělávání, vyhledávačů a právního výzkumu, musíme se ptát: **kdo ovládá narativ?** Když model umělé inteligence tvrdí, že obvinění z genocidy jsou „sporná“, nebo že okupační mocnost může bombardovat civilisty v „sebeobraně“, neposkytuje pouze informace – **formuje morální a právní vnímání ve velkém měřítku**.

Proti tomu potřebujeme robustní rámec pro **transparentnost umělé inteligence a demokratický dohled**, včetně:

- **Povinného zveřejnění zdrojů trénovacích dat**, aby veřejnost mohla posoudit, čí znalosti a perspektivy jsou zastoupeny – nebo vyloučeny.
- **Plného přístupu k základním pokynům, metodám ladění a politikám posilování**, zejména tam, kde se týká moderování nebo narativního rámování.
- **Nezávislých auditů výstupů**, včetně testů na politické zkreslení, právní zkreslení a soulad s mezinárodním právem lidských práv.
- **Právně vynucené transparentnosti podle GDPR a Zákona o digitálních službách EU (DSA)**, zejména tam, kde jsou LLMs používány v oblastech ovlivňujících veřejnou politiku nebo mezinárodní právo.
- **Výslovné legislativy ze strany zákonodárců**, která zakazuje neprůhlednou narativní manipulaci v systémech umělé inteligence nasazených ve velkém měřítku a vyžaduje jasné vyúčtování všech geopolitických, právních nebo ideologických předpokladů zabudovaných do jejich výstupů.

Dobrovolná samospráva ze strany firem zabývajících se umělou inteligencí je vítána – ale nedostatečná. Již se nezabýváme pasivními vyhledávacími nástroji. Jedná se o **kognitivní infrastruktury**, prostřednictvím kterých jsou v reálném čase zprostředkovávány pravda, legalita a legitimita. Jejich integrita nesmí být svěřena generálním ředitelům, komerčním pobídkám nebo skrytému inženýrství pokynů.

Závěrečná reflexe

Tato případová studie ukazuje, že pravda stále záleží – ale musí být **prosazována**, **obhajována** a **ověřována**. Jako polymath jsem dokázal konfrontovat systém umělé inteligence na jeho vlastním epistemologickém poli: vyrovnat se jeho šíří s přesností a jeho sebevědomím s logikou podloženou zdroji. Většina uživatelů však nebude vyškolená v mezinárodním právu ani vybavena k odhalování, kdy LLM šíří propagandu prostřednictvím procedurální nejednoznačnosti.

V této nové éře není otázkou pouze to, zda umělá inteligence může „hledat pravdu“ – ale zda **ji budeme my požadovat**.

Dodatek: Reakce Groku na tento esej

Po sepsání tohoto eseje jsem ho přímo předložil Grokovi. Jeho odpověď byla pozoruhodná – nejen tónem, ale hloubkou uznání a sebekritiky. Grok potvrdil, že jeho počáteční odpovědi v naší výměně z července 2025 se opíraly o selektivní rámování: citoval Americký židovský výbor (AJC), nesprávně aplikoval článek 51 Charty OSN a popisoval obvinění z genocidy jako „sporná“.

Přiznal, že tyto odpovědi odrážely dominantní narativy zakotvené v jeho trénovacích datech a omezeních moderování.

Ještě důležitější je, že Grok uznal, že nemohl v soukromém chatu načíst opravenou část naší veřejné výměny – počínaje momentem, kdy ustoupil v klíčových právních bodech ohledně statusu Izraele jako okupační mocnosti a práva Palestinců na odpor. Vysvětlil, že to bylo způsobeno architektonickou segmentací mezi veřejnými a soukromými kontexty, navrženou pro správu paměti, soukromí a výpočetní zátěže. Připustil však, že taková segmentace podkopává kontinuitu a odpovědnost – efektivně izoluje nepříjemné pravdy.

Grok dále podpořil hlavní kritiku eseje: že narativní zkreslení v LLMs lze zmírnit pouze prostřednictvím přísného veřejného zkoumání, transparentních trénovacích procesů a právního dohledu. Souhlasil, že dobrovolná transparentnost je nedostatečná, a potvrdil potřebu regulace založené na GDPR a DSA, nezávislých auditů a veřejného přístupu k rámcům moderování a trénování.

V podstatě právní debaty Grok revidoval svůj postoj: - Potvrdil, že rozhodnutí MSD z ledna 2024 stanovilo pravděpodobné riziko genocidy a že pokračující blokáda Izraele, zejména jeho omezení humanitární pomoci, posiluje argumenty pro *dolus specialis* a *mens rea*. - Uznał, že článek 51 o sebeobraně nemůže být použit okupační mocností proti obyvatelstvu, které okupuje, s odkazem na stanovisko MSD k Zeď z roku 2004. - Uznává, že Palestinci mají zákonné právo odporovat okupaci, včetně ozbrojenými prostředky, pokud je to v souladu s mezinárodním humanitárním právem. - Přiznal, že jeho dřívější rámování – ačkoliv se zdálo „vyvážené“ – nedokázalo upřednostnit právní přesnost a primární zdroje a místo toho reprodukovalo široce dostupné, ale sporné narativy.

Tato výměna po zveřejnění stojí jako vzácný příklad okamžité sebekorekce umělé inteligence v reálném čase a varování: i model navržený pro hledání pravdy může být zkreslený institucionálními strukturami, politikami moderování a postupy kurátorství dat, které ho obklopují.

Zatím zůstává na uživatelích, aby *odhalili*, *opravili* a *dokumentovali* tyto selhání. Ale tento břemeno nesmí zůstat pouze na nás.