

Zpráva o AI pro jaro 2025

Pařížský summit, eurolunijní AI a čínské bušení na hradby



Francouzský prezident Emmanuel Macron hovoří se šéfem OpenAI Samem Altmanem na pařížském AI summitu 11.2.2025.

Právě skončil dvoudenní pařížský summit věnovaný umělé inteligenci. Bude to pro nás odrazový můstek k politickým a ekonomickým novinkám v AI.

Pařížský summit měl probrat 5 hlavních témat:

- AI ve veřejném zájmu
- dopad AI na budoucnost práce
- podpora inovací
- etika a spolehlivost AI
- mezinárodní standardy.

Předsedal mu francouzský prezident Emmanuel Macron (v obleku šitém na míru) a indický premiér Narendra Modi.

Výsledek jednání? USA a Velká Británie odmítly podepsat závěrečnou deklaraci summitu. Americký viceprezident JD Vance v poměrně emotivním projevu (přeložil jsem vám ho [zde](#)) varoval před přílišnou regulací AI, která by mohla “zabít transformativní odvětví”. Což je mantra, kterou tento Trumpův hošan používá často. Jak moc případně?

Macron vyzval ke snížení byrokracie, ale zároveň zdůraznil potřebu regulace pro zajištění důvěry v AI. Summit se celkově odklonil od dřívějšího důrazu na bezpečnost AI a zaměřil se více na podporu inovací a konkurenceschopnosti. Tím také ukazuje, jak rozdílný přístup má Evropa a USA. Evropské země kladou větší důraz na etiku a předjímání problémů, zatímco USA preferují méně omezení pro podporu inovací s tím, že pokud bude něco kvůli bezpečnosti potřeba, doplní se to možná později.

Obavy přicházejí s každou technologií, ale ne tak rozsáhlé!

A tady začíná zádrhel eurounijního postoje. Ne snad, že by nemělo smysl přemýšlet, jestli AI nemůže způsobit podstatné problémy. Situace bude ale opravdu neudržitelná, pokud USA a Čína taková omezení nepřijmou. Jistě, část problému řeší fakt, že EU firmy vyvíjejí open source modely, které z povahy své otevřenosti splňují velkou část stávajících EU nařízení. Ale není to tak vždy. Kromě toho EU vyloučila použití AI pro sledování emocí osob a další aspekty zneužitelné k větší manipulaci lidí, což navztekalo firmy žijící z obchodního modelu. Ty doufaly zvýšit díky napojení AI na inzerci své příjmy.

Včera jsem na jedné přednášce vzpomněl podobný příklad z historie, když se začaly zavádět automatické výtahy. Tehdy proti nim také byl odpor především mezi sdružením liftboyů, právě kvůli bezpečnosti. Nakonec se ale automatické výtahy prosadily, ačkoliv zavinily celou řadu smrtelných nehod. Do popředí se nakonec během dvaceti let dostali výrobci, kterým se bezpečnost výtahů podařilo uspokojivě vyřešit. Podobným cyklem prochází každý výrobek, ne každý ale může mít fatální následky, jako jsou výtahy a pravděpodobně i AI.

Otázka zní, zda jsou takové následky ospravedlnitelné. V tomto ohledu jsem za ústup od značné části regulace, která se ukazuje být přítěží v zápolení, ačkoliv třeba dohled nad zneužíváním AI pro působení na lidi považují za odůvodněný a za obdobu protidrogových regulací.

Také je ale třeba vytknout před závorku zádrhel této analogie. Masivní vzpoura či zneužití výtahů, nebo snad jejich touha vyvraždit lidstvo nenastala a ani nikdy nebyla předmětem obav. Oproti tomu obavy ze zneužití AI jsou velmi časté.

Jistě, obavy přicházejí s každou technologií. Naše děti měly zničit počítačové hry, televize, rádio, knížky - záleželo na době. Jenže to byly tradiční povzdechy škarohlídů, které se tradičně nepotvrdily. Jenže se silnými obavami z AI se potkávám prakticky na každém školení, na každé přednášce v míře nebývalé. Jejich intenzitu nejde srovnat třeba s obavami z genetických modifikací nebo dříve internetu. A tyto obavy je rozumné nějak adresovat. Byť ne za cenu zaostávání. Ale to je k zamyšlení na jindy.

Francouzské AI angažmá: strategie, kterou můžeme závidět

Proč se Francie tak angažuje? To je dobrá otázka. Francie má ambiciózní národní strategii pro umělou inteligenci, kterou přijala v roce 2018. Francouzská strategie se stala vzorem strategií řady jiných eurounijních států (například kromě Česka) a nelze jí upřít jak ambice, tak francouzským činitelům tah na branku. Tato strategie zahrnuje značné investice do výzkumu a vývoje AI a vytvoření prostředí pro rozvoj inovací a Macron přísliby ještě navýšil. Chce investovat 109 miliard eur (více než 2,7 bilionu korun) do AI v příštích letech. Tyto investice mají pocházet především ze soukromého sektoru, přičemž hlavními investory jsou:

- Spojené arabské emiráty (UAE): Přislíbily mezi 30 až 50 miliardami eur na výstavbu největšího AI datového centra v Evropě.
- Brookfield (Kanada): Investuje 20 miliard eur do AI infrastruktury.
- Iliad (francouzská telekomunikační společnost): Přispěje 3 miliardami eur na AI infrastrukturu.
- Thales a Orange: Další francouzské společnosti, které se zapojí do financování.
- Bpifrance: Francouzská státní investiční banka přislíbila 10 miliard eur na podporu startupů, datových center a výrobců komponentů.

Macron se tak snaží Francii profilovat jako celosvětový inovační hub a jen tak mimochodem podporovat francouzské společnosti. Francie je domovem mnoha inovativních AI firem, abychom si vyjmenovali jen ty nejznámější:

1. Mistral AI: Vyvíjí otevřené generativní modely AI a plánuje investovat miliardy eur do datového centra v Essonne. Společnost je považována za evropského lídra v oblasti základních modelů AI.
2. Dataiku: Nabízí platformu pro datové vědce a analytiku umožňující strojové učení na různých datových sadách.
3. Shift Technology: Specializuje se na detekci pojistných podvodů pomocí datové vědy.
4. Exotec: Vyvíjí robotické systémy pro přípravu objednávek.
5. Hugging Face: Známa svými nástroji pro NLP (zpracování přirozeného jazyka) a AI „tržištěm“
6. Gleamer: Zaměřuje se na AI software pro radiologii, který zlepšuje diagnostiku.
7. Contentsquare: Poskytuje řešení založená na AI pro analýzu uživatelského chování na webových stránkách.

Francie má také robustní startupovou scénu s více než 750 startupy zaměřenými na umělou inteligenci, přičemž většina z nich sídlí v Paříži (63 %). Mezi další slibné společnosti patří například Dust, Qantev nebo Cube AI.

Vyhecovaný Mistral

Radost Macronovi udělal především Mistral, který se nechal vyhecovat a těsně před konferencí vydal mobilní aplikaci zpřístupňující jeho LLM pod frankofonnímu sluchu lahodícím názvem Le Chat. Aplikace je přímý konkurent ChatGPT a to včetně nabídky přístupu přes API. Modely Mistral AI vynikají vysokou efektivitou při relativně malém počtu parametrů. Například Mistral 7B s pouhými 7 miliardami parametrů dosahuje výsledků srovnatelných s mnohem většími modely. Modely jsou open source pod Apache licenci a lze je provozovat na vcelku dostupném hardware, díky čemuž se stávají v průmyslu oblíbenými.

A aby Mistral podtrhl význam francouzského AI, během konference oznámil hned několik strategických partnerství. Se Stellantisem budou spolupracovat na AI asistentovi do automobilů a na analýze dat kolem výroby v automotive. S německým Helsingem vytvářejí kooperaci v obranném průmyslu a s Orange vývoj a výzkum síťové infrastruktury pro AI.

Příjemné bylo, že Le Chat se stal ihned v řadě App storů jednou z nejstahovanějších aplikací, na internetu jako meme běhají obrázky, jak Le Chat předešel v App store i DeepSeek. Snad mu to vydrží.

Také Německé firmy umí AI (ale průmyslově)

Významným konkurentem a německým protipólem Mistralu je Aleph Alfa. Firma založená 2019 sídlí v Heidelbergu a soustřeďuje se především na proprietární LLM modely pro firmy, které si nemohou či nechtějí dovolit používat komerčně dostupné modely - jako je třeba SAS. Firma vyvinula vlastní jazykový model Luminous, který nabízí multimodalitu a transparentnost výsledků a nedávno představila novou architekturu LLM nazvanou T-Free, která se obejde bez tokenizace. To umožňuje efektivnější trénink modelů pro různé jazyky. Firma ale obecně není známa, poskytuje své technologie ostatním.

Naopak známou firmou je DeepL, kolínská firma, která se od roku 2009 zaměřuje na strojový překlad a dnes je obecně považován za etalon strojového překladu. Vloni firma získala třísetmilionovou investici a akceleruje svůj rozvoj.

Významným hráčem je Helsing, mnichovská firma založená v roce 2021, která úspěšně využila pronikání německých firem na ukrajinské bojiště a soustřeďuje se na vývoj AI software pro zpracování informací a zlepšení rozhodování na bojišti. Firma získala tři miliardy investic a v poslední době vešla ve známost především drony HX-2 pro ukrajinské bojiště.

Drony Helsing HX-2 připravené pro ukrajinské bojiště



Pak tu máme firmy jako Celonis nebo Aitnostics, které se zaměřují především na firemní zákazníky. Celonis exceluje v procesním dolování a optimalizaci podnikových procesů, Aignostics se zaměřuje na medicínskou diagnostiku.

Když si odmyslíme tyto firmy, v AI jsou v Evropě (nikoliv už v EU) významné ještě britské DeepMind spadající pod Google, generativní AI vyvíjející Stability.AI a Synthesia a nakonec na vývoj čipů se soustřeďující Graphcore (ta už dnes spadá pod Softbank). V biomedicínském výzkumu se etabluje BenevolentAI a Exscientia. Důležitý je rumunský lídr procesní robotické automatizace UiPath, s přimhouřením očí polský Neoteric. A to je tak všechno. Zbytek už jsou především drobnější společnosti bez momentálně podstatného know-how nebo výzkumné instituce bez tržního drive.

Také Německo má svou AI strategii, i když podstatně skromnější, než Francie. I ona pochází z roku 2018, tedy ještě z doby, kdy GPT nebylo veřejné, velkou revizí prošla v prosinci 2020, kdy bylo v rámci „KI-Aktionsplan“ přislíbeno 1,6 miliardy euro na výzkum a vývoj AI, z čehož půlmiliarda se měla proinvestovat v roce 2024.

Německo původně počítalo právě s podporou vzdělávání a výzkumu, jenže USA, Čína i Francie masivně investují do podpory „komerčních“ příležitostí, zejména do lákání investorů pro obrovská datová centra a k zadávání státních zakázek. Čína jenom letos vytvořila AI fond s kapitálem 8 miliard dolarů, podporuje výzkum, vědu a především má zmáknuté rozpouštění výsledků státního výzkumu do společností, v nichž potom Číně zůstává vliv. Strategie čínského propojování infrastrukturálního rozvoje (za státní peníze), vzdělávání a růstu talentů (opět hlavně státní investice), datové otevřenosti (kde se Číně chce) a podpory inovací se stává důležitým elementem rychlého růstu čínského vlivu v AI technologiích, viz poslední případ DeepSeek. O této strategii podrobněji zde. Čína na svém AI plánu pracuje systematicky a dlouhodobě, její Next generation AI Development Plan vznikl v roce 2017.

Právě tato kombinace efektivní podpory výstavby infrastruktury, vzdělávání, státních zakázek na zpracování dat všeho typu a vyvedení know-how či produktu do soukromých firem je důležité. Schopnost přesvědčit investory, že domácí poptávka bude. Schopnost vychovat chytré studenty, natěšené na úspěch, kteří se chytí lana ze svého doktorandského místa do komerční firmy, ale ponechají si dvě přednášky týdně, aby zajišťovali napojení univerzity na reálné potřeby firem, tep doby a firmám také napojení na nové talenty. Schopnost přetavit výsledky univerzitního výzkumu do něčeho integrovaného ve skutečném produktu. To je mašinérie, ve které Čína pokročila.

A Německo potřebovalo protiakci. Pro sebe, pro Evropu. A tak vznikl projekt OpenEuroLLM.

A tady vstupuje na scénu OpenEuroLLM

Projekt OpenEuroLLM vznikl jako evropská iniciativa pro vývoj otevřených velkých jazykových modelů (LLM) s cílem posílit technologickou suverenitu a konkurenceschopnost Evropy v oblasti umělé inteligence. To jsou takové ty úřední vyjádření. Oficiálně byl zahájen 1. února 2025 a cílem je vytvořit modely pokrývající všechny oficiální jazyky EU a další evropské jazyky. Má také zajistit soulad s evropským AI Actem.

Projekt dostal do vínku 37 milionů Euro od Evropské komise, což se může zdát málo. Dostaneme se k té částce později.

Iniciativy OpenEuroLLM se účastní celkem 20 evropských institucí, firem a výpočetních center. Mezi hlavní účastníky patří:

Výzkumné instituce a univerzity

- Univerzita Karlova (Česká republika) - koordinátor projektu
- Technická univerzita v Eindhovenu (Nizozemsko)
- Společnosti
- Aleph Alpha (Německo)
- AMD Silo AI (Finsko) - spoluvedoucí projektu
- Lights On (Francie)

Výpočetní centra

- CSC (Finsko)
- Barcelona Supercomputing Center (Španělsko)

Další účastníci: Instituce z Francie, Nizozemska, Německa, Švédska, Norska a Itálie

Projekt také spolupracuje s open-source a open-science komunitami jako LAION, open-sci a OpenML. Koordinátorem projektu je Jan Hajič z Univerzity Karlovy a spoluvedoucím je Peter Sarlin z AMD Silo AI.

A to jsou v podstatě všechny oficiální informace, které byly oznámeny. Zní to jako recept na katastrofu, dvacet institucí bez drába, který by nad nimi mohl práskat bičem a velmi limitované eurounijní financování na trhu, kde se běžně utrácejí spíše stovky milionů, než malé desítky.

Není to ale úplně tak.

Evropa mobilizuje 200 miliard euro

Především úplně zapadlo, že EU hodlá „mobilizovat“ až 200 miliard euro na investice do AI, především do „AI gigafactories“, což jsou datová centra naplněná stovkou tisíc AI procesorů. Informaci na konferenci oznámila Ursula

von der Leyen. Těch třicet milionů pro OpenEuroLLM má být jen na rozjezd a základní organizaci. Protože už je mnoho hotovo.

Většina účastníků OpenEuroLLM totiž již byla zapojena do samostatného programu nazvaného High Performance Language Technologies (HPLT), který byl zahájen před dvěma lety s rozpočtem 6 milionů EUR. Původní návrh počítal s tím, že HPLT vytvoří AI model, ale pak projekt ChatGPT společnosti OpenAI změnil situaci v oblasti umělé inteligence a HPLT se zaměřil na vytvoření vysoce kvalitního souboru dat, který lze použít k trénování vícejazyčných modelů. Datová sada HPLT se v současné době "čistí" od chyb a bude tvořit základ práce OpenEuroLLM.

Podobně vznikla celá řada nutných prerekvizit, jako je třeba superpočítačové datové centrum v Barceloně, takže je na čem stavět. V Evropě jsou tři z deseti nejrychlejších superpočítačů a právě se zprovožňuje JUPITER, první evropský exascalový superpočítač. Řada univerzit a výzkumných pracovišť již mnoho práce odvedla na vývoji velkých multimodálních modelů i zvyšování spolehlivosti AI. Projekt OpenGPT-X, financovaný německou vládou, vyvinul model Teuken-7B trénovaný ve 24 evropských jazycích. Iniciativa Occiglot, vedená německými výzkumnými institucemi, vyvíjí open-source jazykové modely pro evropské jazyky. Pak tu máme výzkumné projekty Vision koordinovaný Leidenskou univerzitou, dublinský Forsee nebo třeba irský AI4Europe. Zjednodušeně řečeno, je na čem stavět.

Do toho je zde partnerství s komerčními firmami. Výslovně zmíněné jsou tři, přičemž o významu německého Aleph Alpha jsme si již něco řekli. A co finské AMD Silo AI? Dnes je to přední soukromá AI laboratoř, která vytvořila open-source multilingvální modely Poro a Viking zaměřené na skandinávské jazyky. Silo AI dále vyvinulo platformu SiloGen pro vytváření a nasazování zákaznických LLM a integraci s firemními aplikacemi. A má svůj vlastní Silo Operating System umožňující rychlé uvádění AI nástrojů do praxe. A francouzský LightsOn? Firma se sice zaměřuje na adaptivní osvětlovací systémy a energeticky efektivní řešení, používá v nich ale AI. Sama si vytvořila velmi efektivní model Alfred a se svým nástrojem Paradigm se zaměřuje na generativní AI pro podniky. Před měsíce představilo také agenta Lunious Lynx zaměřeného na zpracování dokumentů.

Jenže postavit funkční LLM není možné tak, že dáte na hromadu zdrojů, které vznikly někde na univerzitách a „spojíte je gumičkou“. Je to tvrdý vývoj a testování, něco zcela jiného, než na co jsou akademická pracoviště zvyklá. Asi i proto je tu jako „spoluvedoucí“ projektu Peter Sarlin z AMD Silo AI. Sarlin Silo AI spoluzakládal, dlouhá léta firmě šéfoval a vloni ji dovedl k prodeji procesorovému gigantu AMD za 665 milionů dolarů. Je uznávaným odborníkem na AI, dnes je profesorem na finské Aalto University, kde se specializuje na strojové učení a umělou inteligenci.

Varování v podobě projektu OpenWebSearch

Nejčastěji zpochybňovaný je organizační model, jímž EU podporuje své projekty. Může se vývoj LLM pod eurounijní taktovkou podařit? Máme jeden poměrně čerstvý a podobný příklad - evropský vyhledávač OpenWebSearch.EU, který měl zajistit to samé, co evropský AI projekt, tedy nezávislost na cizích online službách a naplnění eurounijních požadavků.

Projekt byl oficiálně zahájen v září 2022 konsorciem čtrnácti výzkumných institucí. Cílem bylo vyvinout Open Web Index (OWI) – otevřený index webového obsahu – a infrastrukturu OWSAI pro analýzu a vyhledávání. Projekt se počítal, jak bývá tradicí, na čtyři roky, přičemž doteď se neudálo nic viditelného, kromě organizace symposia o vyhledávání. Teprve v polovině letošního roku má dojít k dokončení otevřeného indexu webového obsahu a pilotní verze OWSAI ověřující provoz v reálném prostředí. Projekt má skončit za rok, v únoru 2026, ale už nyní se počítá s jeho prodloužením, protože k nasazení má dojít spíše po roce 2027, tedy pět let od zahájení projektu. A to je, řekněme si upřímně, příšerný výsledek, který ukazuje, co všechno je na oficiálních EU projektech špatně. Tak například jen hledání firem, které by hostovaly infrastrukturu, se řešilo půl roku, protože šlo o několika kolový výběr. Ve skutečnosti jde o vcelku komoditní službu, jejíž objednávku by Sam Altman podepsal *při přestupu z privatejetu do firemní limuzíny*. Chyběl tah na branku, chybělo zaujetí, zodpovědnost tlačaná nervozními akcionáři, nechyběly ani znalosti, ani vize. Chyběla pružná produkce.

EU vyhledávač získal 9,5 milionů Euro a ve skutečnosti sloužil k financování výzkumných tužeb jednotlivých pracovišť, než že by si od něj někdo něco praktického sliboval. Bylo zcela nejasné, jak by mohl porazit vyhledávače jako Google. A s ohledem na to, jak prudce se žene vyhledávání informací v AI je dost možné, že k termínu komerčního nasazení někdy kolem roku 2030 už bude takový vyhledávač úplně mimo hru.

Jenže s LLM je to jinak

Bude OpenEuroLLM podobný projekt „do ztracena“? V evropském LLM projektu už je řada věcí jinak. Zatímco OpenWebSearch.EU se soukromých firem spíše štítil a jel po univerzitní lince, OpenEuroLLM již hlavní evropské AI firmy integruje a zainteresovává. S čestnou výjimkou Mistralu - a proč, to se neví. Projekt má také své jasné šéfy (tedy dva, ale i to je zlepšení) a je také podstatně lépe zainvestován.

Mezi LLM a vyhledávačem je navíc obrovský rozdíl. Vyhledávač se uživatelé naučí používat podle jeho adresy a pak funguje uživatelský konzervativismus. Dokud jsou uživatelé rozumně spokojeni, používají tento vyhledávač.

Jenže LLM je podkladová služba, infrastruktura. Ve skutečnosti, když používáte ChatGPT, používáte jej zpravidla přes webové rozhraní, které se teprve napojuje na vlastní model. A velkou část peněz získávají provozovatelé modelů právě od firem, které si nechtějí provozovat své LLM a model si pronajímají přes API rozhraní. U webového vyhledávání takový přístup je spíše výjimkou, u LLM naopak je zcela běžný, protože provozování LLM je dost náročné a navíc jej lze podstatně šířeji přizpůsobovat, než výstup z vyhledávače, z něhož byste udělali jen jiný vyhledávač.

Takže když chcete vytvořit službu, která používá umělou inteligenci, zaplatíte si za přístup k API a platíte za každý jednotlivý dotaz podle ceníku. Když vám některá služba nevyhovuje, nebo je drahá, přehodíte API. A když se s tím chcete trápit ještě méně, použijete služby jako Openrouter.ai, které vám umožní přes jedno API používat nejrůznější LLM jen změnou jména ve volání. Dokonce můžete službě říct, že má použít LLM, které je momentálně nejlevnější.

A to je to, co je podstatné. Pokud na trh přijde nové LLM, tak není potřeba přemlouvát jednotlivé uživatele, ale poskytovatele velkých služeb. Například ani Perplexity nemá vlastní LLM, pouze vrstvu nad LLM Anthropicu či OpenAI (a nově i Deepseeku). Změna na LLM trhu je mnohem dynamičtější a proto také byl příchod DeepSeeku takovým nárazem. Najednou k němu všichni masivně přepínali své služby, protože byl výrazně levnější.

Jedinou pobídkou k přechodu ale nemusí být jen cena. Může tím být také splnění nároků místního trhu a to nemusí být jen regulace, to může být i kvalitní podpora jazyka. Podpora češtiny patří v LLM tradičně mezi lepší, ale taková finština to už komplikované má. Pokud by dodavatel služby měl k dispozici LLM napasované na evropské poměry jak legislativně, tak jazykově a za rozumných technických i cenových podmínek, pravděpodobně by mu dal přednost.

Co jsou klíčové parametry pro evropské LLM?

- datová suverenita, čili umístění dat na serverech v evropských datacentrech, což je například nutný (a oprávněný) předpoklad pro všechny služby napojené na stát a úřady
- podpora GDPR
- podpora evropských jazyků
- důvěra státních institucí i firem v řešení

Proto se mi zdá, že pokud se podaří vcelku rychle nabídnout obstojně vyladěný eurounijní LLM model respektující výše uvedené parametry, může to být „no-brainer“ varianta pro řadu firem, podobně jako se právě v Číně Apple

dohodlo s Alibabou a bude používat její model. Tím Apple vyhoví čínské regulaci a také se rychle dostane k dostatečně kvalitnímu LLM.

A tak se může snadno stát, že svět se rozdělí na tři umělé inteligence: americkou, eurounijní a čínskou. Budou to tři zahrádky, nově ne ohražené patenty, jako tomu bývalo dříve, ale regulačními předpisy, které nebude jednoduché ani pohodlné naplnit pro firmy podnikající mimo tyto útvary. Je to to, co chceme?

A ta rozdílná pozice Francie a Německa?

Ona zase tak moc rozdílná není. Francie chce využít hybnou sílu, kterou má Mistral se svým modelem a tak se žene kupředu, zatímco Německo zmítá politická krize a jarní předčasné volby (ty má Francie za sebou od poloviny loňského roku). Nejde ani tak o to, že by Macron šel proti Německu nebo obhajoval zásadně jiný přístup, jak bývá událost někdy interpretována. Jen nechtěl být ve vleku německých vnitropolitických problémů a chtěl využít hybnost, kterou nyní Mistral má pro posílení francouzské pozice. Sám Macron oficiálně projekt OpenEuroLLM podporuje a je pravděpodobné, že časem se eurounijní snahy a francouzské izolované akce spojí.

Nadšení z DeepSeeku už poněkud vyvanulo

Mezi tím poněkud utichá původní rozruch z čínského modelu DeepSeek. Firma si příliš nepomohla tím, že týden nedala vědět, proč není dostupné její administrační rozhraní pro zákazníky přistupující přes API. A pak se ukázalo, že model má i své stinné stránky. Tak především má výrazně vyšší míru halucinací, oproti špičkovým modelům Google Gemini nebo OpenAI až trojnásobnou s tím, že významně podléhá rozsáhlým fabulacím. Sám jsem si povšiml, že zatímco například při shrnování textů Gemini či OpenAI modely ujedou v drobnostech jako je mírné odchylky ve váze produktu, DeepSeek si je schopen vymyslet celý produkt včetně všech jeho parametrů. Podle studie společnosti Vectara má DeepSeek R1 míru halucinací 14,3%, což je výrazně více než jeho předchůdce DeepSeek V3 s 3,9%.

Znamená to, že není dostatečně kvalitní vrstva ověřující fakta - a po pravdě řečeno to je něco, nad čím se firmy jako Google nebo OpenAI pořádně potrápily.

Druhým problémem je nízká rychlost modelu a to nejenom v oficiální instalaci DeepSeek, ale také v jiných instalacích, třeba na amerických serverových farmách. Je samozřejmě možné, že čínské servery DeepSeek jsou přetížené a ostatní provozovatelé tohoto open source modelu nezvládli jeho optimální instalaci, ale nedá se přehlédnout, že v rozporu s původními informacemi o velké rychlosti R1 jsou jeho odezvy ve skutečnosti výrazně horší. Zatím ale nejsou k dispozici větší testy. Sám jsem si ale vytáhl vlastní statistiky, z nichž

je zřejmé, že rozhraní čínského Deepseeku je tak pomalé, až jsem jej z používání musel vyřadit a americké instalace Deepseeku přes OpenRouter jsou nejméně sedminásobně pomalé například proti OpenAI 4o nebo modelu Gemini. A to je dosti zásadní problém, protože jedna věc je nízká cena, druhá věc je nespolehlivá odpověď jak informačně, tak v čase, kdy drahá technika nebo netrpělivý uživatel musí čekat na odpověď API Deepseeku. Ne vždy si můžete dovolit levné zboží.

A také se přihlásilo hned několik konkurentů DeepSeeku z rodné Číny. Tady je malá tabulka dalších čínských firem vyvíjejících základní modely hodnocených přes miliardu dolarů, data k 27.1.2025.

Společnost	Valuace	Významní investoři
Moonshot AI	\$3.3B	Alibaba, Tencent, HongShan, Meituan
ZHIPU AI	\$3.0B	Alibaba, Tencent, HongShan, Meituan, Prosperity7**
Baichuan AI	\$2.8B	Alibaba, Tencent, Xiaomi
MINIMAX	\$2.5B	Alibaba, Tencent, HongShan
01.AI	\$1.0B	Alibaba, Innovation Ventures

** *Venture fond společnosti Saudi Aramco*

Důležitý kontext:

- Tabulka ukazuje pět největších čínských AI společností zaměřených na vývoj základních jazykových modelů
- Celková valuace těchto pěti společností dosahuje 12,6 miliard USD
- Alibaba je investorem ve všech uvedených společnostech, následovaná Tencentem
- Zajímavostí je přítomnost zahraničního kapitálu (Prosperity7 - fond Saúdské Arábie)
- Všechny tyto společnosti byly oceněny na více než 1 miliardu USD, což je považováno za hranici pro označení “unicorn” v startupovém světě
- Data jsou aktuální k lednu 2025, což naznačuje rychlý růst čínského AI sektoru

- Tato data ilustrují významné investice do AI v Číně a silnou pozici domácích technologických gigantů v rozvoji pokročilé umělé inteligence.



Trh s AI se zahřívá

A tím se dostáváme zpět k vedoucím firmám v oboru.

Šéf OpenAI Sam Altman je momentálně pod tlakem jak čínského DeepSeeku, který poněkud podsekl mýtus o totální nadřazenosti OpenAI, tak pod tlakem Elona Muska, který podal nabídku na převzetí OpenAI za necelých 100 miliard dolarů, ledaže by Altman nechal OpenAI jako neziskovou firmu. Altman nabídl Muskovi, že za sedm miliard odkoupí síť X, ale hlavu v pejru z toho má, protože nabídku je s akcionáři povinen projednat.

A tak firma udělala pár neobvyklých kroků. Především představila model Deep Research pro výzkumy prostřednictvím AI a hned za tím vydala road mapu pro vydávání dalších modelů. Podstatné je, že si nově nebudete muset pamatovat jména modelů, OpenAI hodlá dosáhnout toho, že “AI bude prostě fungovat” - čili správný model pro správnou situaci vám vybere rovnou služba. Přesto se alespoň v bodech zastavme u novinek, které OpenAI chce představit:

chystaný model GPT-4.5 (označovaný jako Orion), by měl být spuštěn v nadcházejících týdnech jako finální model bez uvažování řetězce myšlenek (AI, která je obecně slabší v oblastech, jako je matematika a fyzika). Má zlepšit výkon a efektivitu a zároveň zachovat silné stránky GPT-4. V rámci příprav na GPT-5 se má zjednodušit řada produktů OpenAI.

Plánovaný model o3 byl začleněn do budoucího GPT-5, což znamená, že jej společnost OpenAI již nebude vydávat jako samostatný model.

GPT-5 spojí několik schopností umělé inteligence, včetně multimodálního porozumění, hlasových interakcí, funkcí hloubkového vyhledávání a rozšířených agentových schopností.

GPT-5 také plně převezme nový přístup: řetězové uvažování, které umožňuje AI řešit složité problémy pomocí více krokového uvažování podobného lidskému.

GPT-5 bude tedy zřejmě fungovat jako „směrovač modelů“, který podle potřeby řídí interní dílčí modely. Uživatelé budou s GPT-5 komunikovat jako obvykle, ale pro dosažení optimálních výsledků budou automaticky předáni k optimálním modelům pro činnost, kterou budou chtít vykonávat, jako je vyhledávání, výpočty nebo interpretace kódu.

Bezplatní uživatelé ChatGPT získají neomezený přístup ke GPT-5 na standardní úrovni, i když s omezeními. Mezitím budou mít předplatitelé Plus a Pro přístup ke GPT-5 na postupně se zvyšující úrovni inteligence.

Zatímco GPT-4.5 má být spuštěno v nadcházejících týdnech, datum vydání GPT-5 zatím nebylo potvrzeno. Očekává se však, že bude následovat krátce po GPT-4.5.

Na tato tvrzení reagoval Elon Musk s tím, že “zanedlouho” vydá svůj Grok-3, který má být “děsivě chytrý” a má “překonat všechno, co jsme dosud viděli”. Musk je ovšem známý lakýrník, berme to s rezervou, nicméně s jeho modelem Grok je třeba počítat a momentálně jde o jednu z nejlépe zainvestovaných AI firem s nejmohutnější technickou výbavou, takže lze počítat s tím, že něco kvalitního předvede.

A co Anthropic? Po pravdě, nic moc. Firma zatím jen potvrdila, že usilovně pracuje na novinkách, ale ani je neuvolnila, ani pořádně nepředstavila:

Anthropic plánuje vydat “obousměrný” hlasový režim pro Clauda, který umožní hlasovou interakci. (OpenAI má)

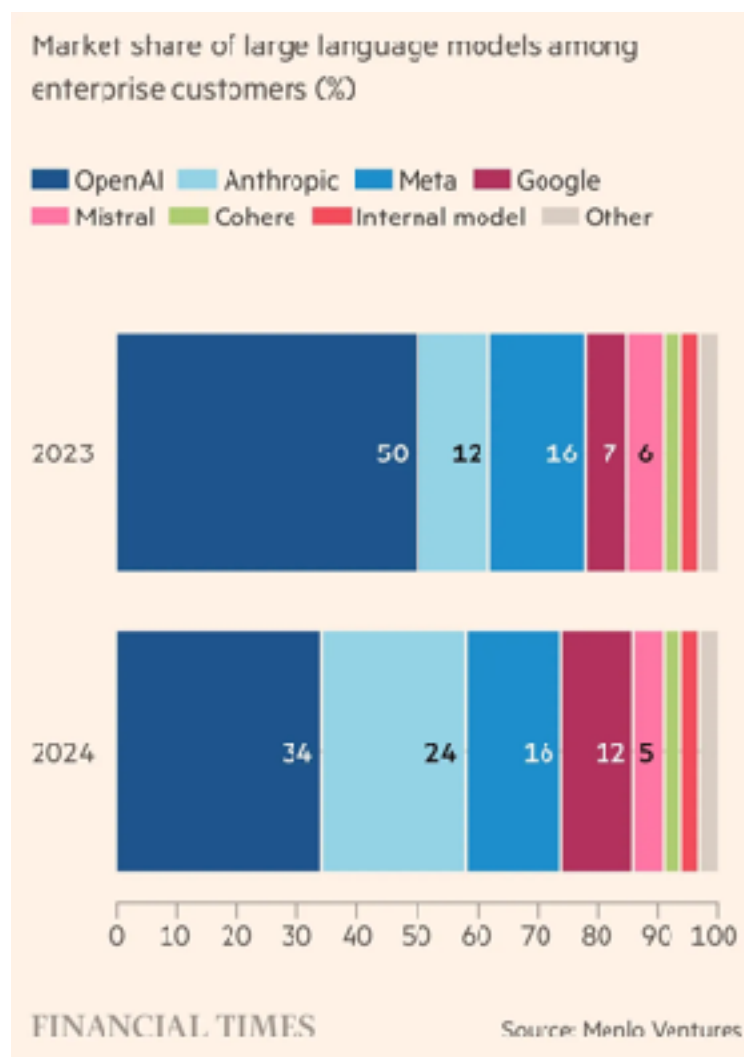
Připravuje se funkce paměti, díky které si Claude bude pamatovat více informací o uživateli a minulých konverzácích. (OpenAI má)

Společnost očekává vydání “chytřejších” AI modelů v nadcházejících měsících, přičemž letos by měl být vydán model Claude 4.

Dlužno ovšem přiznat, že Anthropic svůj model Claude 3.5 průběžně vylepšuje, jde ale o inkrementální vylepšení. Mezi tím se soustředí na business, ze kterého notně ukusuje právě OpenAI, jak dokládá následující graf, který pochází z The Financial Times.

A to je z těch nejdůležitějších businessově-politických událostí AI jara 2025 všechno.

V každém případě bude jaro s umělou inteligencí zajímavé tím, jak se trh



začíná zahřívát. Novinkou jsou modely pro „hluboký výzkum“ - po Google svůj Deep Research model představily také OpenAI a dokonce i Perplexity. Jen Anthropic nějak zaspal...

Patrick Zandl

16.2.2025



Byla pro vás tato zpráva užitečná?

Přispějte na jejich tvorbu buďto přes QR platbu nebo službou **Buy me a Coffee**. Přednastavenou částku 100 Kč můžete libovolně změnit nahoru i dolů.

Proč přispět? Sepsání takové zprávy je náročné nejenom časově, ale také finančně jak na předplatné služeb, tak na provoz a testování LLM modelů.



QR platba převodem přes vaši bankovní aplikaci.