4. BLOK · ODOLNOST ČLOVĚKA A INFORMAČNÍ PROSTOR

EOLAS VAULT: zkušenosti s použitím LLM „v poli“ a jak pracovat s kontextem

Jindřich Karásek · Rapid7

Zápis z přednášky na konferenci IIČR 2026

Informace, inteligence a čas na rozhodnutí · 19. 11. 2026 · CEVRO Univerzita, Praha
AFCEA Česká pobočka, pracovní skupina PS07 Intelligence · CRIK při CEVRO Univerzitě
Stav 20. 11. 2026 · https://iicr.afcea.cz/zaznelo/



4. BLOK · ODOLNOST ČLOVĚKA A INFORMAČNÍ PROSTOR

Jindřich Karásek

Rapid7

EOLAS VAULT: zkušenosti s použitím LLM „v poli“ a jak pracovat s kontextem

---

Jindřich Karásek, vedoucí týmu ve firmě Rapid7 a doktorand Psychologického ústavu Akademie věd, mluvil o používání velkých jazykových modelů v terénu, mimo laboratoř. Hlavní tezí bylo, že proti halucinacím nejlépe chrání hlídání kontextového okna a vlastní, průběžně zpřesňovaná znalostní báze, ideálně ve formě znalostního grafu. Přístup ukázal na případové studii, ve které jeho agentní workflow odhalilo čínského prodejce hardwaru pro farmy falešných účtů na sociálních sítích.

Kdo mluví a čím se zabývá

Řečník se představil jako vedoucí týmu ve společnosti Rapid7, jehož tým sleduje kybernetické útočníky a jejich kampaně, a jako doktorand Psychologického ústavu Akademie věd. Uvedl tři oblasti své práce. První je přednáška o kognitivním válčení na webu, tedy o aktérech, kteří zneužívají web a webové služby ke kognitivnímu válčení. Druhou je výzkum působení APT aktérů ve vztahu k bezpečnosti Česka, Evropské unie a NATO a všude, kde má Rapid7 komerční zájem. Jeho práce mu přitom umožňuje legálně škodit útočníkům jejich vlastními metodami. Třetí oblastí je jailbreaking velkých jazykových modelů, tedy metody, jak model přimět říct nebo udělat něco, co by normálně odmítl. Odkazy na všechny tři oblasti slíbil v prezentaci a nabídl, že slajdy zájemcům pošle.

Kontextové okno jako alfa a omega

Podle řečníka je nejdůležitější, ať model používá jednotlivec, firma nebo produkt, hlídat si kontext a kontextové okno. Je to podle něj jediná cesta, jak se halucinacím vyhnout. Ukázal, že i automatické opravy překlepů v mobilu jsou jen jiným projevem toho, čemu říkáme halucinace. Nevadí nám, protože je chceme. Uvedl vlastní příklady z rychlého psaní anglických e-mailů, kdy mu oprava zachránila pozdrav i poděkování.

Model si sám nic nepamatuje. Při každém kroku konverzace se mu posílá celá dosavadní historie znovu. Bez ní je to jako mluvit s velmi sklerotickým příbuzným, který vás pozdraví, jako by vás ten den viděl poprvé. Kontextové okno se s délkou konverzace zaplňuje. Kdo dělá technické implementace, musí mít vlastní správu relací. Starší modely po vyčerpání okna začaly mluvit nesmysly. Novější před vyčerpáním dělají kompakci: shrnou kontext a zkrátí ho třeba o tři čtvrtiny. Při zkracování ale mohou vypadnout důležité detaily. U textů pro web to nevadí, ve zpravodajské komunitě a v informační bezpečnosti ale na detailech záleží. Záleží i na tom, aby práce s AI byla opakovatelná, měřitelná a aby se dalo zpětně ověřit, podle jakých kritérií se rozhodovalo.

Domácí úkol: vlastní znalostní báze s učící smyčkou

Řečník připravil pro posluchače domácí úkol: pracovní sešit, tedy šablonu na GitHubu. Stačí vložit odkaz do oblíbeného chatbota a zeptat se, co s ním dělat. Chatbot uživatele provede nastavením i používáním. Podle řečníka je to vyzkoušené a optimalizované tak, aby to zvládl kdokoli bez

předchozích znalostí IT. Technicky zdatnější si mohou kód stáhnout a postupovat podle návodu na GitHubu.

Princip podle řečníka spočívá v tom, že si uživatel historii práce s AI ukládá k sobě. Nespoléhá jen na historii chatů u poskytovatele, protože data je lepší spravovat sám. Systém si navíc pamatuje chyby. Na pokyn k vyhodnocení poučení projde, co uživatel s modelem dělal a co se mu nelíbilo. Vzniká tak samoučící smyčka a interní znalostní graf toho, co konkrétní uživatel dělá. Model pak dostává přesnější kontext a odpovídá přesněji, protože je pro něj levnější projít data uživatele než prohledávat vše, na čem byl natrénován. Podobně funguje RAG: vlastní data se zpracují do podoby, kterou model umí použít, a model dostane pokyn podívat se nejdřív tam. Model se tak stává průvodcem a překladačem k tomu, co má uživatel na vlastním disku.

Ontologie, agenti a znalostní grafy

Kdo pracuje s umělou inteligencí jinak než jako koníček, měl by podle řečníka vědět, co je ontologie a co taxonomie, jaký je mezi nimi rozdíl, a každý den si ověřovat, zda je v nich konzistentní.

Agentem je v jistém smyslu i chatbot, protože musí v každé relaci rozluštit záměr uživatele. K tomu slouží zpracování přirozeného jazyka, v podstatě větný rozbor: určení podstatných jmen, zájmen a sloves. Agent je smyčka. Zjistí, co uživatel chce, rozhodne, co udělat, ověří, zda splnil cíl, a pokračuje, dokud cíl podle sebe nesplní. Řečník to ukázal na rezervaci místa v italské restauraci. Agent přitom pracuje s trojicemi podmět, sloveso, předmět (SVO) a z nich lze stavět znalostní grafy. Řečník uvedl, že díky znalostním grafům se halucinacím a dalším problémům při práci s většími objemy dat vyhýbá poměrně spolehlivě. Jako ukázku promítl znalostní graf Tolkienova Silmarillionu, složitého a rozsáhlého díla, ve kterém je těžké se vyznat. Tímto postupem za poslední rok algoritmizoval svých 12 let praxe v kybernetické bezpečnosti.

Případová studie: čínský hardware pro farmy účtů

Na sociálních sítích řečník narazil na profil s fotografiemi blikajících řad mobilních telefonů, jaké jsou známé z médií. Využil ho jako případ pro své workflow, ve kterém má definovaný postup pro OSINT a ověřování informací i pro penetrační testování a red teaming. Na několik příkazů odkryl čínského aktéra, který prodává hardware k ovlivňování sociálních sítí, tedy k automatickému lajkování a zobrazování obsahu. Varoval publikum, aby prodejce neplašilo vlastním skenováním, protože jsou opatrní.

Agenti nejprve vyhodnotili fotografie z profilu, dohledali další profily a nakreslili sociální mapu. Na jedné fotografii byla zachycena firma, která v Číně skutečně existuje a zařízení dodává. Její web měl špatně zabezpečenou klientskou sekci s PDF dokumenty, které řečník vytěžil. Byly v nich nabídky na míru pro konkrétní klienty. Řečník z nich s vysokou mírou jistoty doložil, že firma vyváží systém do Evropy a Ameriky, že má platící klienty a že hardware existuje.

Jde o rack za zhruba 700–800 dolarů s asi 20 deskami, které napodobují chování mobilních zařízení. Každé zařízení má specifický backdoor, který umožňuje orchestraci až stovek takových zařízení. V datacentru tak lze vytvořit desítky tisíc uživatelů. Dříve bylo nutné programovat přesné souřadnice kliknutí. Díky pokroku počítačového vidění dnes robotům stačí zadat cíl přirozeným jazykem. Řečník se částečně dostal k aktualizačnímu serveru. Firma slibuje doživotní aktualizace, tedy i službu průběžného obcházení bezpečnostních systémů platforem. Našel přímé důkazy, že systém byl optimalizován hlavně na TikTok a měl metody k obcházení detekce společnosti Meta. Ta asi rok sbírá

přes aplikaci i telemetrii z telefonu, například údaje o vysílači (BTS), o Wi-Fi a o dalších nainstalovaných aplikacích. Kód tyto údaje simuloval a obsahoval i vyčerpávající seznam kódů zemí a operátorů, podle kterých se identifikují SIM karty.

Dotazy: jak začít s vlastním druhým mozkem

Moderátor ocenil, že řečník s lehkostí přednáší o děsivých věcech a bourá představu, že se v kyberprostoru pohybujeme bezpečně. Další dotaz se týkal toho, jak se v takových systémech zorientovat a začít, a zda je jedním z možných využití takzvaný druhý mozek, tedy znalostní graf nad vlastními poznámkami. Řečník odpověděl, že právě jeho domácí úkol je druhý mozek. Pracuje na něm rok, začal ještě dřív, než to bylo populární, a podobný způsob práce později doporučil i Andrej Karpathy. Upřesnil, že technicky nejde o RAG, protože se data nepřevádějí na vektory, ale model s nimi pracuje podobně.

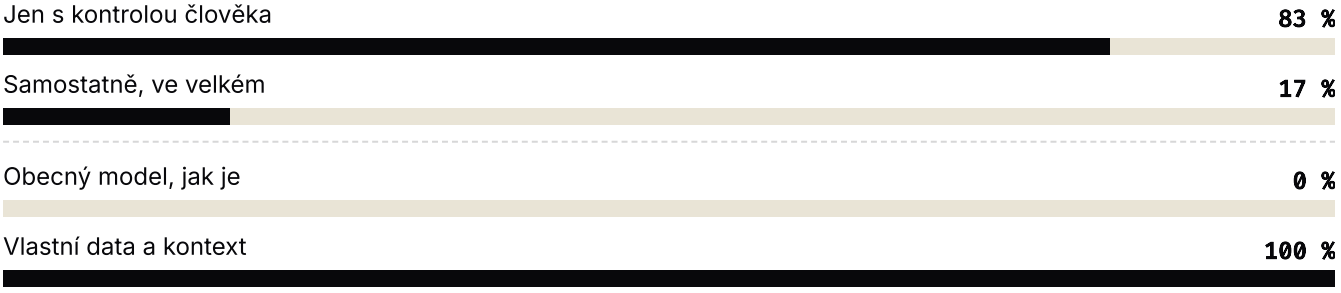
Šablona obsahuje základy, aby chatbot zvládl odpovídat i na jednoduché dotazy uživatelů a pomohl s orientací. Je optimalizovaná i pro lokální modely. Nejmenší modely, například Phi od Microsoftu, podle řečníka nestačí, ale s modelem kolem 8 miliard parametrů už se dá dobře pracovat. Lze ji používat i s komerčními modely. Řečník nemusel psát složitý kód ani ho překládat do mnoha jazyků. Šablona obsahuje graf, základní dovednosti a vzory, jak vypadá workflow, dovednost a znalostní graf, a každému se přizpůsobí pro jeho konkrétní účel.

Co z přednášky plyne pro stát

- Při nasazení jazykových modelů ve zpravodajství a bezpečnosti je třeba hlídat kontext a kompakci, protože při zkracování kontextu vypadávají detaily. Práce s AI má být opakovatelná, měřitelná a zpětně ověřitelná.
- Instituce, které s AI pracují soustavně, potřebují konzistentní ontologii a taxonomii a vlastní znalostní báze nebo znalostní grafy. Podle řečníka je to nejspolehlivější ochrana před halucinacemi.
- Na trhu se volně prodává hardware pro farmy účtů (rack s asi 20 deskami za 700–800 USD, škálovatelný na desítky tisíc uživatelů). Dodává se s průběžnými aktualizacemi k obcházení ochran platforem, je optimalizovaný na TikTok a vyváží se i do Evropy.
- Agentní workflow s jasně definovaným postupem OSINT dokázalo takového dodavatele na několik příkazů odhalit a doložit jeho klienty. Takové nástroje jsou použitelné i pro analytiky státu.
- Řečník doporučil vlastní znalostní bázi, která poběží i na lokálních modelech kolem 8 miliard parametrů. Data tak zůstávají u uživatele.

Co řekl sál

Jak používat jazykové modely v analytické práci? `n = 12, kompas: dvě osy`



Otázky ze sálu poslané přes aplikaci

- Můžeme poprosit ten slide s domácím úkolem

Zápis pořadatelů ze strojového přepisu, ne autorizovaný text řečníka. Citace zveřejníme až po schválení řečníkem. Opravy: ps07@afcea.cz.

Čtení pořadatelů, ne společné stanovisko řečníků. Zápisy vycházejí ze strojového přepisu konference, ankety vyjadřují názor sálu, ne reprezentativní průzkum. Citace řečníků zveřejníme až po jejich schválení. Opravy a dotazy: ps07@afcea.cz. Volně k použití s uvedením zdroje „IIČR 2026, AFCEA Česká pobočka“. Více na https://iicr.afcea.cz/zaznelo/