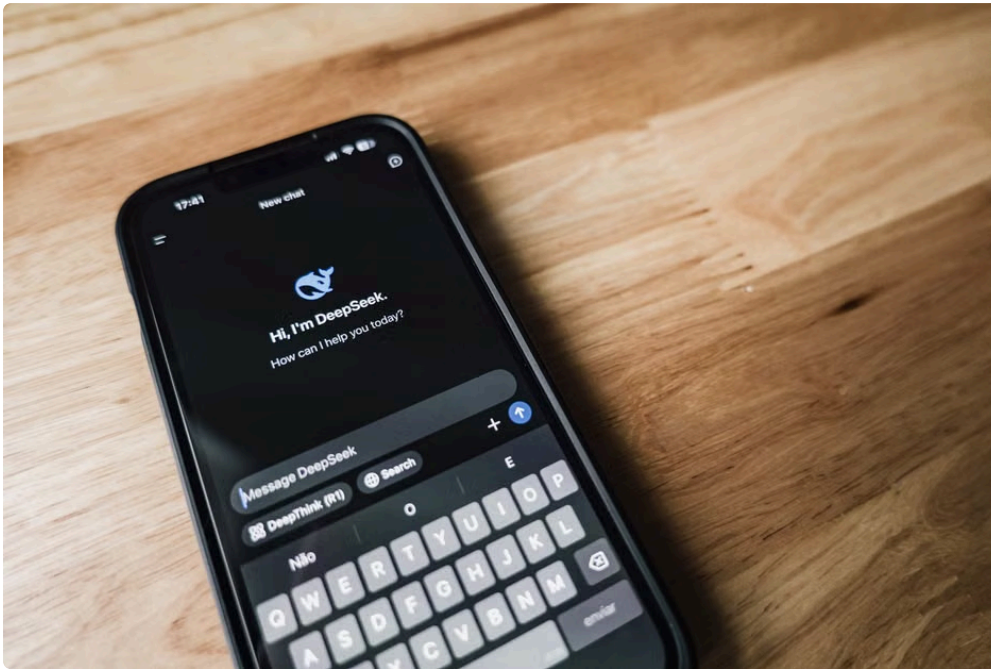


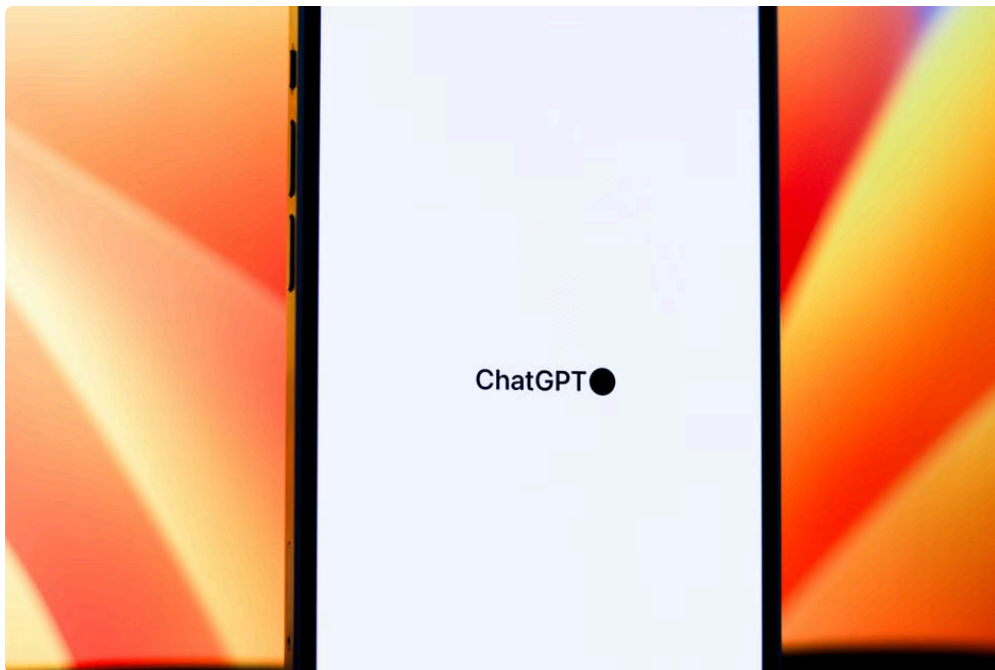
V posledných rokoch sa umelá inteligencia (AI) v kombinácii s veľkými jazykovými modelmi (LLM) začala intenzívne využívať v právnom sektore. Právnické kancelárie hľadajú spôsoby, ako zrýchliť a zefektívniť prípravu dokumentov, vyhľadávanie súdnej judikatúry či tvorbu právnych analýz. No právo je disciplína, kde sú **povinné citácie** zákonov a presnosť informácií nevyhnutné: chyba v citácii môže viesť k fatálnym dôsledkom vrátane právnej neplatnosti dokumentov alebo reputačným škodám.

V tomto blogu sa pozrieme na to, čo sú *AI halucinácie*, prečo LLM halucinujú, kedy je riziko najvyššie práve v právnej oblasti, a ako moderné metódy ako **Retrieval-Augmented Generation (RAG)** či **fine-tuning na doménových dátach** pomáhajú minimalizovať riziko zlých citácií zákonov. Zároveň zdôrazníme dôležitosť konceptu *human-in-the-loop* pre overenie výstupov AI.



Čo sú AI halucinácie a prečo pri LLM vznikajú?

AI halucinácie označujú jav, keď model vyprodukuje informácie, ktoré nie sú pravdivé alebo ktoré nie sú zakotvené v žiadnych dátach, na ktorých bol trénovaný. Ide o falošné fakty, „vymyslené“ citácie, nesprávne dátumy alebo nesprávne interpretácie zákonov.



Prečo halucinácie vznikajú?

1. **Pravdepodobnostný model namiesto databázy faktov:** LLM sú trénované ako pravdepodobnostné predpovedné modely textu. Neobsahujú databázu "skutočných faktov", ale snažia sa odhadovať, aké slová alebo vety budú najpravdepodobnejšie nasledovať za predchádzajúcim textom.
2. **Nedostatok presných referenčných dát:** Ak LLM nemá v tréningových dátach dostatok právnych dokumentov alebo aktuálnych legislatívnych textov, môže vytvárať všeobecné alebo nepresné odpovede.
3. **Zložitosť viackrokových právnych otázok:** Právne otázky často vyžadujú precízne sledovanie viacerých paragrafov, precedentov a právnych princípov. LLM sa môže pri viacstupňovom dôvode „stratiť“ alebo zlúčiť informácie nesprávne.

Kedy je riziko zlých citácií najvyššie?

Vo všeobecnosti platí, že riziko halucinácií a nesprávnych faktov u AI modelov je najvyššie v oblastiach, kde ide o:

- **Právo:** Povinné presné citácie zákonov, platné paragrafy, súdne rozhodnutia a ich správna interpretácia.
- **Medicína:** Diagnózy, liečebné postupy, farmakologické informácie.
- **Financie:** Regulácie, daňové zákony, finančné produkty.
- **Multi-step výskum:** Kde je potrebná analýza, syntéza a správne kombinovanie viacerých zdrojov.

Právnické kancelárie sa preto musia obzvlášť zamerať na elimináciu potenciálnych halucinácií, pretože poškodenie reputácie, zlyhanie v súdnom procese alebo sankcie za nesprávne informácie môžu viesť k závažným následkom.

RAG (Retrieval-Augmented Generation) – spojenie znalostnej bázy s AI

Retrieval-Augmented Generation (RAG) predstavuje moderný prístup, kde je LLM doplnený o externú *knowledge base* – teda databázu overených dokumentov, zákonov a iných relevantných zdrojov, z ktorých model čerpá pri generovaní odpovedí.

Ako RAG funguje?

1. Užívateľ zadá otázku alebo požiadavku.

2. Prvým krokom je vyhľadanie relevantných dokumentov v overenej knowledge base podľa kľúčových slov a dotazu.
3. Extrahované pasáže sa poskytnú LLM ako kontext na generovanie odpovede.
4. Model vytvorí odpoveď, ktorá je do veľkej miery založená na reálnych a overených dátach.

Tento prístup významne znižuje riziko halucinácií, pretože odpovede sú "kotvené" v reálnych právnych textoch namiesto toho, aby model vytváral odpovede len na základe štatistických pravdepodobností v jazyku.

Kvalita knowledge base

- **Dôveryhodnosť:** Zdroje musia byť oficiálne (napr. štátne právne databázy, lexikóny, známe právne portály).
- **Aktualizovanosť:** Právne normy sa neustále menia, preto je kľúčové pravidelné aktualizovanie databázy.
- **Štruktúrované a indexované údaje:** Pre rýchly a presný retrieval.

Fine-tuning na doménových dátach – špecializácia modelu

Fine-tuning znamená dodatočné tréningovanie LLM na špecifických, doménových dátach právnickej kancelárie. Tieto dáta môžu obsahovať interné dokumenty, typické formuláre, vzory zmlúv, ale aj anotované právne texty a verifikované citácie.

Výhody fine-tuningu

- Model lepšie "rozumie" slovníku a špecifickým formuláciám v právnej oblasti.
- Redukuje sa nesprávne generované fakty, pretože model má "skutočné" príklady správnych citácií.
- Zvyšuje sa konzistentnosť a štandardizácia výstupov.

Limitácie a riziká

- Fine-tuning vyžaduje kvalitné, v dostatočnej miere anotované a aktuálne dáta.
- Riziko "overfittingu": model sa príliš prispôsobí doménovým dátam a môže stratiť schopnosť generalizácie.
- Vyžaduje neustálu kontrolu a validáciu výstupov.

Human-in-the-loop: Kde je zdroj a ako to otestujeme?

V právnom kontexte nie je možné plne automatizovať generovanie dokumentov bez zásahu človeka. Preto koncept **human-in-the-loop (HITL)** zodpovedá za:

- **Overenie správnosti citácií:** Právnik ručne preverí, či sú zákonné odkazy správne a aktuálne.
- **Testovanie kvality AI výstupov:** Simulovať reálne otázky a prípady, testovať odpovede na overených príkladoch.
- **Zber spätnej väzby:** Vyhodnocovať, kde AI halucinovala a učiť sa z chýb pri ďalšom tréningu fond-tuningu alebo update knowledge base.

Mini-checklist pre právnickú kanceláriu nasadzujúcu AI:

1. Založiť robustnú a pravidelne aktualizovanú knowledge base z overených zdrojov zákonov.
2. Implementovať RAG systém na generovanie odpovedí s referenčným retrieval krokom.
3. Fine-tunovať model na špecifických doménových dátach kancelárie.

4. Zaviest' jasný proces human-in-the-loop overovania každej AI generovanej citácie.
5. Nastaviť metriky a monitoring výstupov pre včasné odhalenie halucinácií a chýb.
6. Pravidelne testovať model na reálnych právnych dotazoch a dokumentoch.

Záver

Nasadenie AI v právnych službách otvára nové možnosti efektivity a dostupnosti právnej pomoci. Avšak bez adekvátnej kontroly a správnych technológií, ako sú RAG a fine-tuning na doménových dátach, hrozí riziko **AI halucinácií** – nesprávnych alebo nepravdivých citácií zákonov. Preto je nevyhnutné budovať systémy **povinné citácie v AI** so silnou knowledge base, implementovať human-in-the-loop procesy a testovať AI riešenia na reálnych prípadoch právnej praxe.

Právnická kancelária, ktorá vie využiť AI zodpovedne, dokáže znížiť čas potrebný na vyhľadávanie, zlepšiť presnosť dokumentov a minimalizovať riziko právnych chýb.

Zdroj a testovanie: Pre budovanie knowledge base využívajte oficiálne zdroje ako Slov-Lex, pre RAG systémy použite knižnice ako Hugging Face Transformers s integračným frameworkom (napr. Haystack). Testovanie realizujte na historických právnych prípadoch a validujte pravé citácie manuálne v spolupráci s právnikmi.