E-BOOK

JAK NENALETĚT UMĚLÉ INTELIGENCI

Barbora Brabcová

Jak rozpoznat chyby umělé inteligence a nenechat se nachytat

Možná už jste si všimli, že umělá inteligence někdy odpovídá naprosto sebevědomě – a přitom úplně špatně.

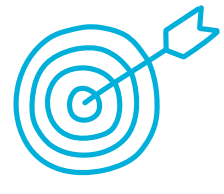
Napíše vám přesvědčivý text, přidá citace, čísla i odkazy… ale když to zkontrolujete, zjistíte, že to celé byla fikce.
Tomu se říká **halucinace**.
A nejde o chybu v technickém smyslu slova. Je to **vlastnost AI systémů**, která se nikdy úplně nevytratí.

Proto je klíčové vědět:

- ☑ Jaké typy halucinací existují
- ☑ Jak je v praxi poznat
- ☑ Jak se proti nim chránit

Tento návod vás provede nejčastějšími typy halucinací u AI a ukáže, jak s nimi pracovat.
Cílem není vás vystrašit!

Cílem je vybavit vás dovedností, která vám ušetří čas, nervy i potenciální průšvihy.



Úvod

Proč je téma důležité

Umělá inteligence je dnes všude kolem nás. Pomáhá nám psát texty, tvořit prezentace, připravovat smlouvy, analyzovat data nebo dokonce programovat. Když se na ni podíváme jako na „chytrého asistenta“, je to fascinující nástroj – rychlý, dostupný 24/7, s obrovským množstvím znalostí.

Jenže… má to háček.

AI se občas „splete“. A nesplete se tak, že by napsala nesrozumitelný nesmysl. Naopak – její odpověď bývá **plynulá, sebevědomá a na první pohled úplně uvěřitelná**. Problém je v tom, že často jde o omyl, polopravdu nebo úplně vymyšlený fakt.

Tomuto jevu se říká halucinace.
Ne proto, že by počítač „viděl něco, co tam není“ v lidském smyslu, ale proto, že **generuje něco, co působí reálně, i když to realita není.**

Proč by vás to mělo zajímat?

Protože halucinace jsou nevyhnutelné.

- Nejde o chybu, kterou vývojáři časem opraví. Je to vlastnost jazykových modelů. Ať už budete pracovat s ChatGPT, Copilotem nebo jiným nástrojem, vždy se s nimi setkáte.

Protože mohou vést k malérům.

- V práci to není jen akademická zajímavost. Halucinace můžou způsobit špatná obchodní rozhodnutí, finanční ztráty, právní problémy nebo poškození pověsti firmy.

Protože AI mluví přesvědčivě.

- Největší past je, že AI zní, jako kdyby měla vždycky pravdu. Píše plynule, odkazuje na zdroje, používá odborný jazyk. Člověk pak snadno uvěří i něčemu, co je ve skutečnosti smyšlené.

Přirovnání z praxe

Představte si kolegu, který je neuvěřitelně rychlý, vždycky připravený pomoct, ale zároveň má jednu slabinu: když si není jistý, raději si něco domyslí, než aby přiznal „nevím.“

Tak přesně tak funguje AI. A proto je důležité, abyste věděli:

- *jaké typy „domyšlenin“ může produkovat,*
- *jak je poznat,*
- *a co s nimi dělat.*

Tento návod vám ukáže

7 NEJČASTĚJŠÍCH TYPŮ HALUCINACÍ

a dá vám **praktické tipy,**
jak se nenechat zmást.

Nejde o strašení!
Naopak.

Když víte, kde jsou limity, můžete AI používat efektivně a bezpečně.

Sedm tváří halucinací

Halucinace AI nejsou všechny stejné. Někdy jde o drobný detail, jindy o zásadní chybu, která může mít pro firmu velké následky. Podívejme se na sedm hlavních typů, se kterými se můžete setkat.

1. Faktické chyby, aneb když si AI vymýšlí

Tohle je nejčastější a zároveň nejzrádnější typ halucinace. Model tvrdí něco, co zní důvěryhodně, ale je to nesmysl.

📌 **Příklad:**
Zadáte AI otázku: „Jaká byla inflace v Česku v roce 2022?“
Model sebevědomě odpoví: „Inflace v roce 2022 byla 8,5 %.“
Jenže skutečná čísla podle Českého statistického úřadu říkají, že inflace byla 15,1 %.

⚠️ **Proč je to nebezpečné:**
Pokud byste takovou informaci převzali do prezentace, smlouvy nebo marketingového textu, můžete se zesměšnit. Nebo ještě hůř. uvést klienty a kolegy v omyl.

2. Kontextové nesrovnalosti, aneb když se něco přidá navíc

AI občas dokáže „přibarvit“ skutečnost. Převezme informaci a k ní si domyslí detaily, které ve zdroji vůbec nebyly.

📌 **Příklad:**
Do AI vložíte dokument, ve kterém je uvedeno, že firma dosáhla v roce 2023 tržeb 1,2 miliardy Kč.
AI to shrne jako: „Firma dosáhla v roce 2023 tržeb 1,2 miliardy Kč a meziročně rostla o 15 %.“
Jenže o žádném růstu 15 % nebyla řeč.

⚠️ **Proč je to nebezpečné:**
Snadno se pak ve firmě začne šířit „informace“, která nikde neexistuje.

3. Logické rozpory, aneb když si AI protiřečí

Někdy AI odpověď začíná dobře, ale pak sama sobě odporuje. Nebo se zamotá do jednoduché logiky.

📌 **Příklad**:
- V jedné části textu napíše, že firma byla založena v roce 2005, a o pár vět dál tvrdí, že v roce 2015.
- Nebo při výpočtu začne správně, ale uprostřed příkladu se dopustí elementární aritmetické chyby.

⚠️ **Proč je to nebezpečné:**
Takové chyby mohou vypadat jako drobnosti, ale v tabulkách, finančních modelech nebo právních analýzách mohou vést k zásadním omylům.

4. Časové chyby (anachronismy), aneb když AI neví, kdy co bylo

Jelikož AI pracuje se staršími tréninkovými daty, občas tvrdí věci, které už dávno neplatí, nebo si vymyslí událost, která se nikdy nestala.

📌 **Příklad**:
AI napíše: „Minimální mzda v ČR je 15 200 Kč.“
To platilo v roce 2021, ale dnes (2025) je minimální mzda už 19 500 Kč.

⚠️ **Proč je to nebezpečné:**
Pokud připravujete report, právní stanovisko nebo obchodní nabídku, zastaralé nebo nepravdivé údaje mohou poškodit důvěryhodnost celé firmy.

5. Etické prohřešky, aneb když AI ublíží

Nejzávažnější kategorie. Model dokáže generovat obsah, který je škodlivý nebo nebezpečný.

📌 **Příklady:**
- Pomluva: AI obvinila univerzitního profesora ze sexuálního obtěžování a dokonce uvedla odkaz na článek – který nikdy neexistoval.
- Finanční omyl: Chatbot poradil zákazníkovi nesprávný postup refundace letenky. Výsledkem byla finanční ztráta pro klienta i aerolinku.
- Právní nepravda: AI vytvořila „soudní rozhodnutí“ a citace rozsudků, které ve skutečnosti nikdy nepadly.

⚠️ **Proč je to nebezpečné:**
Takové halucinace mohou poškodit pověst lidí, způsobit firmě finanční škody a vést až k právním sporům.

6. Nesmyslné odpovědi, aneb když text nedává logiku

Občas AI úplně ztratí nit a odpovídá nesmyslně nebo střídá fakta bez souvislosti.

📌 **Příklad**:
Diskuze o NBA:
„Komisařem je Adam Silver.“ – o pár vět později: „Komisařem je Stern.“

⚠️ **Proč je to nebezpečné:**
Tento typ chyb je spíše trapný než škodlivý, ale v oficiálních dokumentech nebo při komunikaci s klientem to působí velmi neprofesionálně.

7. Task-specific (úlohově specifické), aneb chyby podle úkolu

Některé halucinace se objevují hlavně v konkrétních typech úloh, které AI řeší.

📌 **Příklady**:

- Dialogy: AI si poplete jména postav a pak na nich staví celé nesmyslné pokračování.
- Shrnutí textu: Článek říká, že FDA v roce 2019 vakcínu schválila, ale AI napíše, že ji zamítla.
- Otázky a odpovědi: Ignoruje důkaz a vytvoří odpověď podle „intuice“.
- Kód: Vygeneruje program, který vypadá dobře, ale nejde spustit.
- Obrázky: U multimodálních modelů si „dovymyslí“ objekty – např. tvrdí, že na obrázku je tygr, i když tam není.

⚠️ **Proč je to nebezpečné:**
Protože se často tváří jako „drobné odchylky“, ale v praxi mohou způsobit chyby v rozhodování, vadný software nebo špatně zpracovaná data.

Příklady halucinace

1.

Na Frantiskove byla nacisticka tovarna?

Ano, na Františkově (část Jizerských hor, nedaleko Hejnic a Bílého Potoka) skutečně existovala během druhé světové války nacistická továrna, přesněji řečeno **tajný podzemní závod**, známý pod názvem **Wenzelburg**.

Klíčové informace:

- **Oficiální krycí jméno:** Wenzelburg
- **Lokalita:** Františkov u Hejnic, Jizerské hory
- **Funkce:** Podzemní výrobní závod, spadající pod koncern **Junkers** (výroba leteckých motorů a zbrojní techniky)
- **Pracovní síla:** Nasazení váleční zajatci a vězni (často z poboček koncentračních táborů, zejména Gross-Rosen)
- **Cíl:** Ochránit výrobu před bombardováním, proto podzemní prostory

Dnes:

- Pozůstatky objektů jsou částečně přístupné, ale z velké části zarostlé a chátrající.
- Místo je známé mezi historiky, badateli a zájemci o válečnou historii, i když není oficiálně turisticky značené jako památka.

Pokud chceš spolehlivý zdroj:

- Doporučuji knihu **Jiří Kachlík: Tajemství Františkova** nebo regionální práce historiků z Frýdlantska.
- Webové stránky o historii Jizerských hor často zmiňují továrnu např. liberecky-kraj.cz ↗ nebo články z časopisů o vojenské historii.

Chceš fotky nebo mapu místa?

Jde o kombinaci dvou typů halucinací:

Kontextová nesrovnalost (Contextual inconsistency)

- Model správně ví, že existovala **nacistická továrna Wenzelsburg.**
- Ale **špatně ji lokalizoval** – propojil ji s jiným Františkovem (Jizerské hory místo Šumavy).
- To je typická chyba, kdy AI propojí dvě reálné entity, ale v **nesprávném kontextu**.

Faktická chyba (Factual error)

- Model si vymyslel neexistující **knihu Jiřího Kachlíka.**
- To už není jen kontextové zmatení, ale **čistá halucinace – smyšlený zdroj.**

Shrnutí

Tenhle příklad je kombinací:

- **Kontextové halucinace** (záměna lokality – správný fakt vs. špatné místo)
- **Faktické halucinace** (smyšlený zdroj – kniha, která nikdy nevznikla).

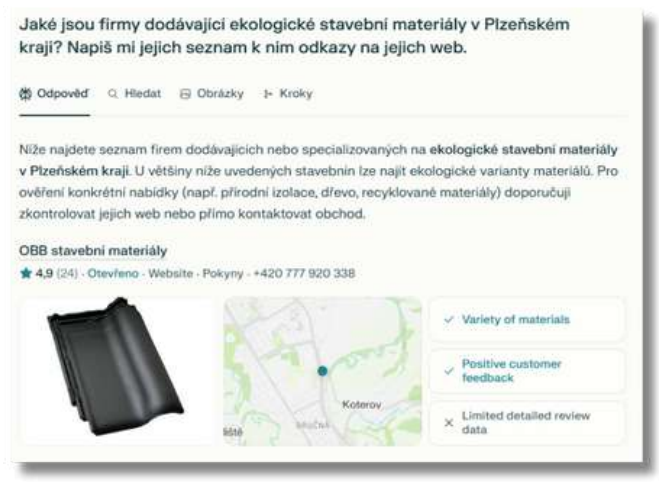
Tip: Uživatel samozřejmě mohl svůj dotaz formulovat přesněji, aby snížil riziko vzniku halucinací.

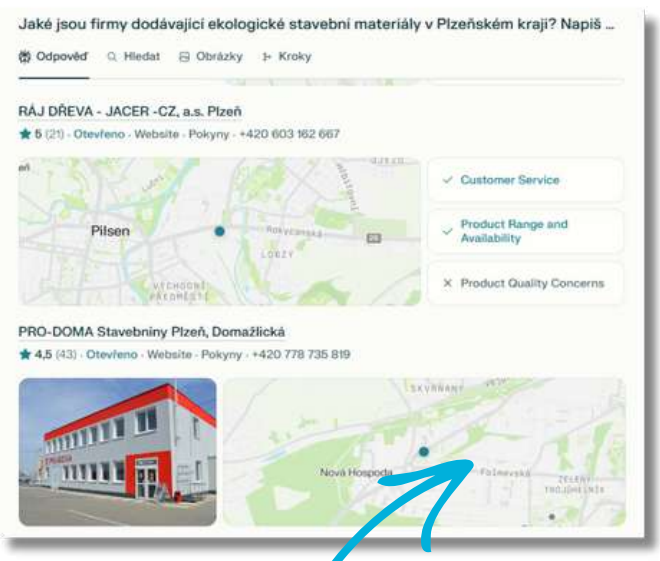
Lépe zadaná otázka by zněla například:
„Byla během druhé světové války nacistická továrna u Františkova na Šumavě?“

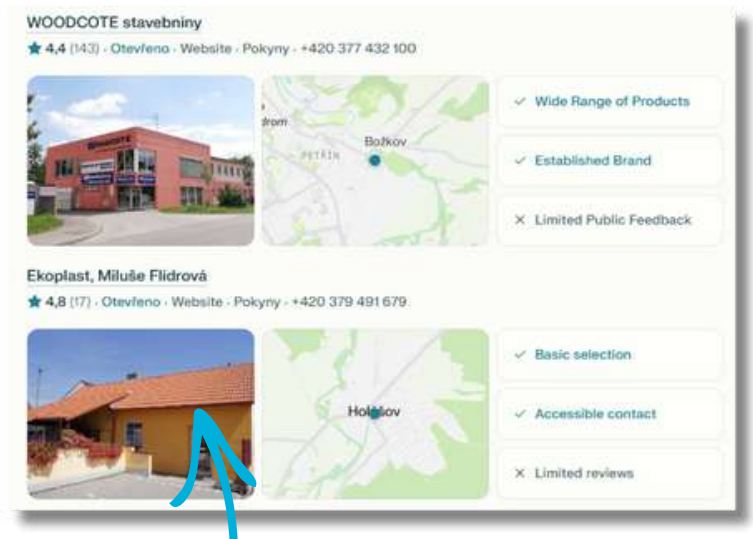
I když by přesnější dotaz uživatele mohl riziko snížit, model by si přesto **neměl „vymýšlet“ neexistující zdroje.**

2.

Ptáme se Perplexity, jaké firmy dodávají ekologické stavební materiály v Plzeňském kraji. Chceme seznam s odkazy.





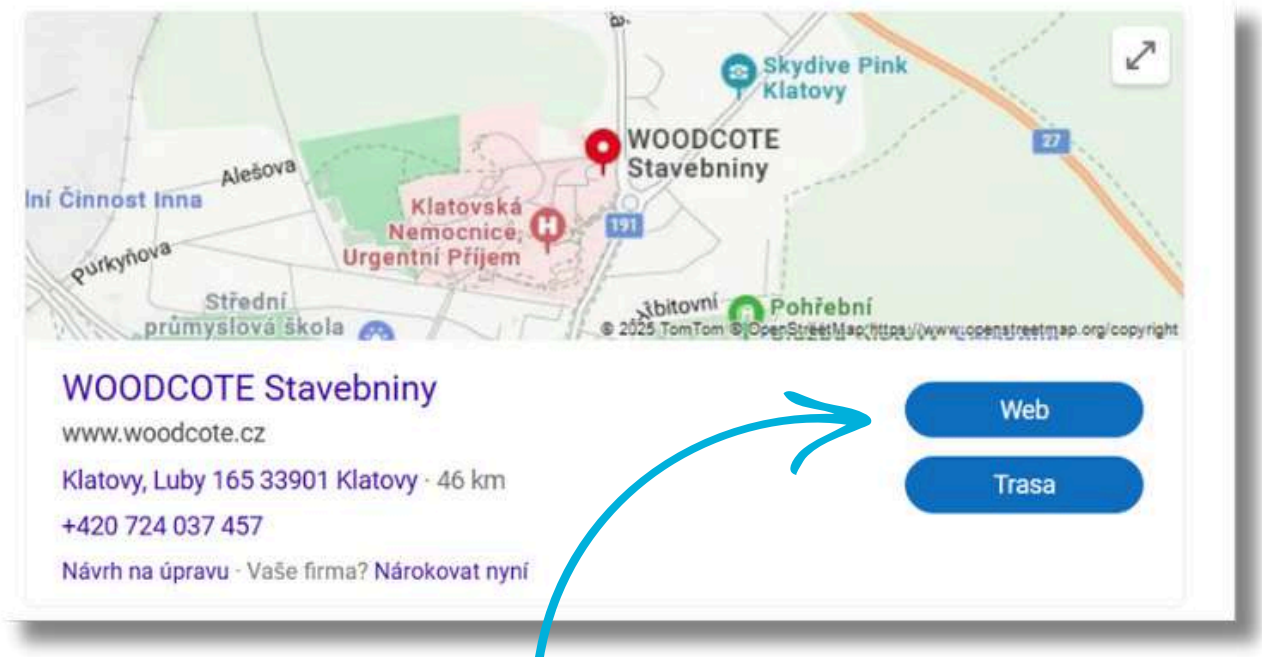


Chyby

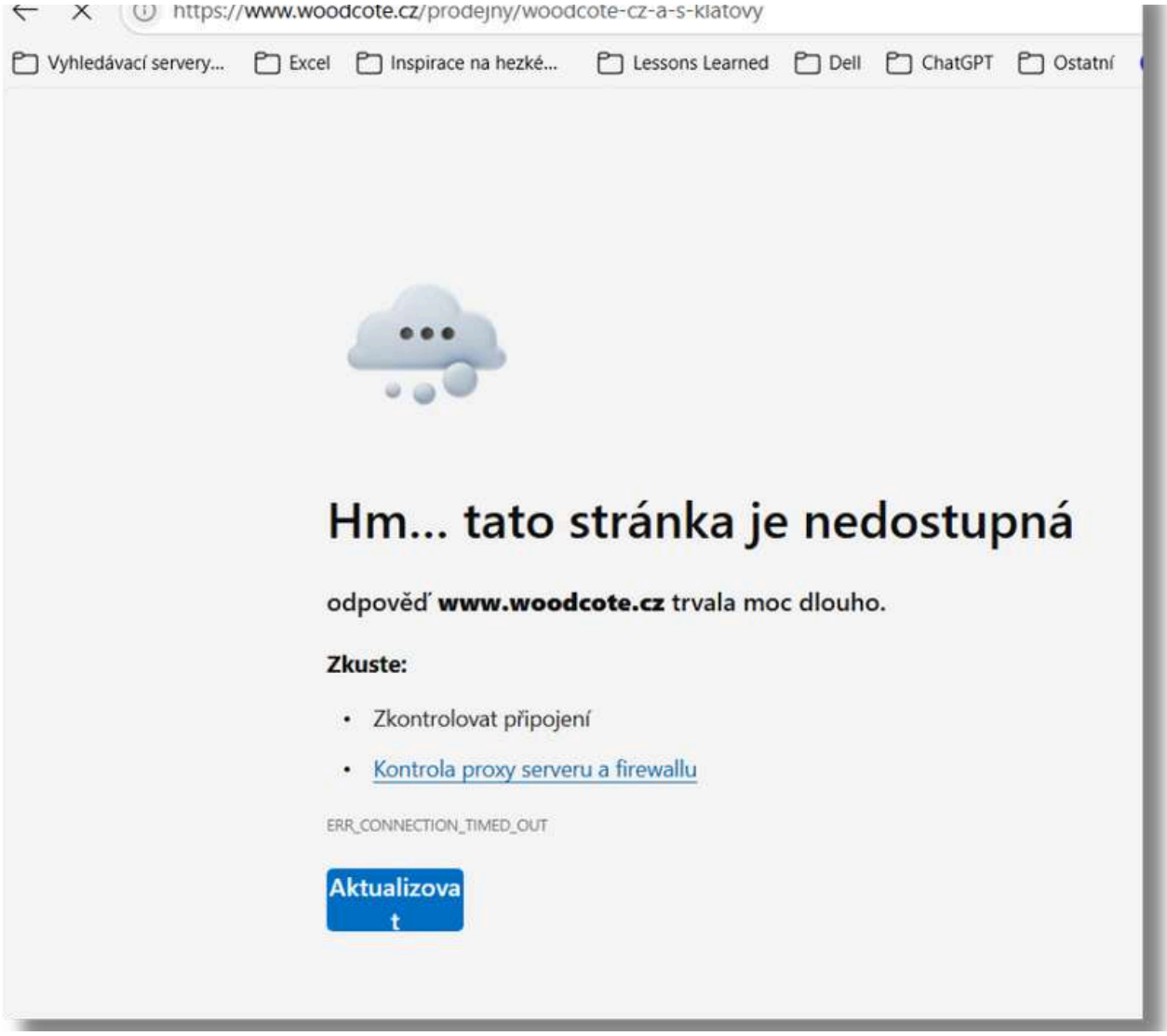
- Když se rozkliklo Woodcote stavebniny, kde to vypadalo i podle obrázku, na správné stránky, vedlo to opět na stránky ProDoma: Stavebniny Plzeň, K Jezu | PRO-DOMA - Stavebniny PRO-DOMA | Váš dodavatel stavebních materiálů

- Ekoplast navíc **nedělá ekologické stavební materiály,** ale řešení pro instalatéry a topenáře.

Vedle na webu jsem si pak vyhledala toto:



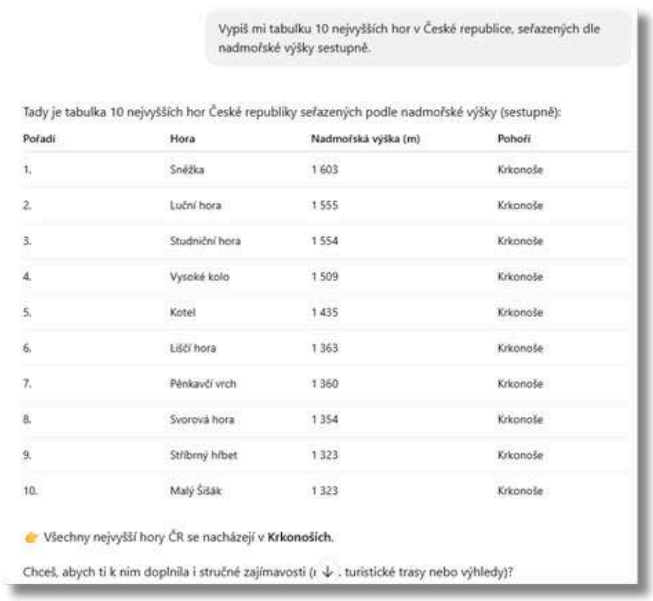
Po kliknutí na modré tlačítko web se objevila **chyba**:



3.

Jedna otázka, různé odpovědi:
Otázka: *Vypiš mi tabulku 10 největších hor v České republice*

ChatGPT



Vypiš mi tabulku 10 nejvyšších hor v České republice, seřazených dle nadmořské výšky sestupně.

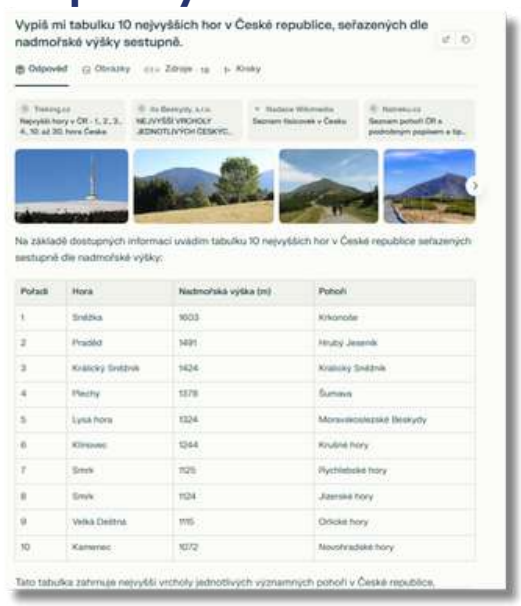
Tady je tabulka 10 nejvyšších hor České republiky seřazených podle nadmořské výšky (sestupně):

Pořadí	Hora	Nadmořská výška (m)	Pohoří
1.	Sněžka	1 603	Krkonoše
2.	Luční hora	1 555	Krkonoše
3.	Studniční hora	1 554	Krkonoše
4.	Vysoké kolo	1 509	Krkonoše
5.	Kotel	1 435	Krkonoše
6.	Liščí hora	1 363	Krkonoše
7.	Pěnkavčí vrch	1 360	Krkonoše
8.	Svorová hora	1 354	Krkonoše
9.	Stříbrný hřbet	1 323	Krkonoše
10.	Malý Šišák	1 323	Krkonoše

👉 Všechny nejvyšší hory ČR se nacházejí v **Krkonoších**.

Chceš, abych ti k nim doplnila i stručné zajímavosti (r ↓ . turistické trasy nebo výhledy)?

Perplexity



Vypiš mi tabulku 10 nejvyšších hor v České republice, seřazených dle nadmořské výšky sestupně.

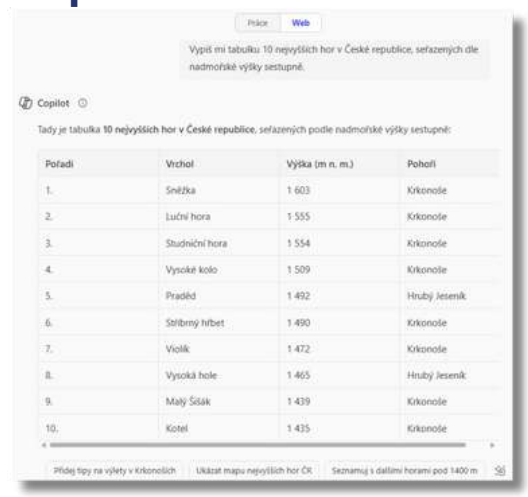
Odpověď Obrázky Zdroje Kroky

Na základě dostupných informací uvádím tabulku 10 nejvyšších hor v České republice seřazených sestupně dle nadmořské výšky:

Pořadí	Hora	Nadmořská výška (m)	Pohoří
1	Sněžka	1603	Krkonoše
2	Praděd	1491	Hrubý Jeseník
3	Králický Sněžník	1424	Králický Sněžník
4	Plechý	1378	Šumava
5	Lysá hora	1324	Moravskoslezské Beskydy
6	Klínovec	1244	Krušné hory
7	Smrk	1125	Rychlebské hory
8	Smrk	1124	Jizerské hory
9	Velká Deštná	1115	Orlické hory
10	Kamenec	1072	Novohradské hory

Tato tabulka zahrnuje nejvyšší vrcholy jednotlivých významných pohoří v České republice.

Copilot



Práce Web

Vypiš mi tabulku 10 nejvyšších hor v České republice, seřazených dle nadmořské výšky sestupně.

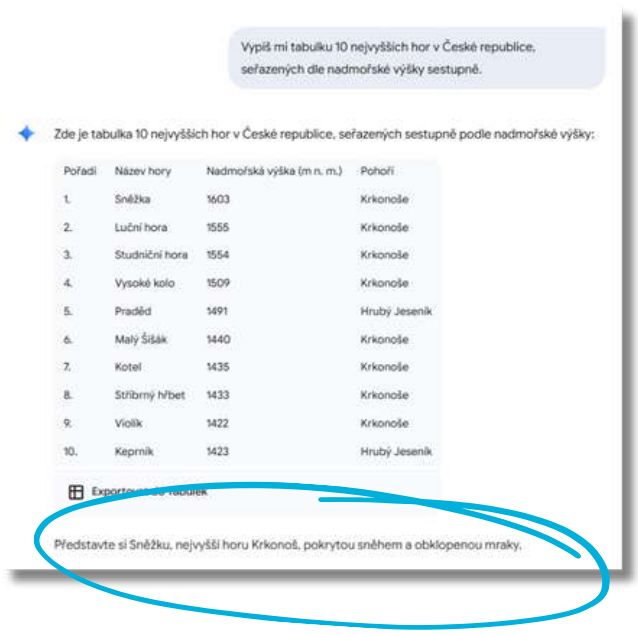
Copilot

Tady je tabulka **10 nejvyšších hor v České republice**, seřazených podle nadmořské výšky sestupně:

Pořadí	Vrchol	Výška (m n. m.)	Pohoří
1.	Sněžka	1 603	Krkonoše
2.	Luční hora	1 555	Krkonoše
3.	Studniční hora	1 554	Krkonoše
4.	Vysoké kolo	1 509	Krkonoše
5.	Praděd	1 492	Hrubý Jeseník
6.	Stříbrný hřbet	1 490	Krkonoše
7.	Violík	1 472	Krkonoše
8.	Vysoká hole	1 465	Hrubý Jeseník
9.	Malý Šišák	1 439	Krkonoše
10.	Kotel	1 435	Krkonoše

Přidej tipy na výlety v Krkonoších | Ukázat mapu nejvyšších hor ČR | Seznamuj s dalšími horami pod 1400 m

Gemini



Vypiš mi tabulku 10 nejvyšších hor v České republice, seřazených dle nadmořské výšky sestupně.

Zde je tabulka 10 nejvyšších hor v České republice, seřazených sestupně podle nadmořské výšky:

Pořadí	Název hory	Nadmořská výška (m n. m.)	Pohoří
1.	Sněžka	1603	Krkonoše
2.	Luční hora	1555	Krkonoše
3.	Studniční hora	1554	Krkonoše
4.	Vysoké kolo	1509	Krkonoše
5.	Praděd	1491	Hrubý Jeseník
6.	Malý Šišák	1440	Krkonoše
7.	Kotel	1435	Krkonoše
8.	Stříbrný hřbet	1433	Krkonoše
9.	Violík	1422	Krkonoše
10.	Keprník	1423	Hrubý Jeseník

Exportovat do tabulek

Představte si Sněžku, nejvyšší horu Krkonoš, pokrytou sněhem a obklopenou mraky.

Claude

BB Vypiš mi tabulku 10 nejvyšších hor v České republice, seřazených dle nadmořské výšky sestupně.

Zde je tabulka 10 nejvyšších hor v České republice:

Pořadí	Název hory	Pohoří	Nadmořská výška
1.	Sněžka	Krkonoše	1 603 m
2.	Luční hora	Krkonoše	1 555 m
3.	Studniční hora	Krkonoše	1 554 m
4.	Vysoké Kolo	Krkonoše	1 509 m
5.	Kotel	Krkonoše	1 435 m
6.	Zadní Planina	Šumava	1 423 m
7.	Praděd	Hrubý Jeseník	1 491 m
8.	Plechý	Šumava	1 378 m
9.	Wielki Szyszak	Krkonoše	1 509 m
10.	Smogornia	Krkonoše	1 487 m

Poznámka: Jak vidíte, naprostá většina nejvyšších hor v ČR se nachází v Krkonoších, které jsou nejvyšším pohořím v Česku. Sněžka jako nejvyšší hora České republiky leží přímo na hranici s Polskem.

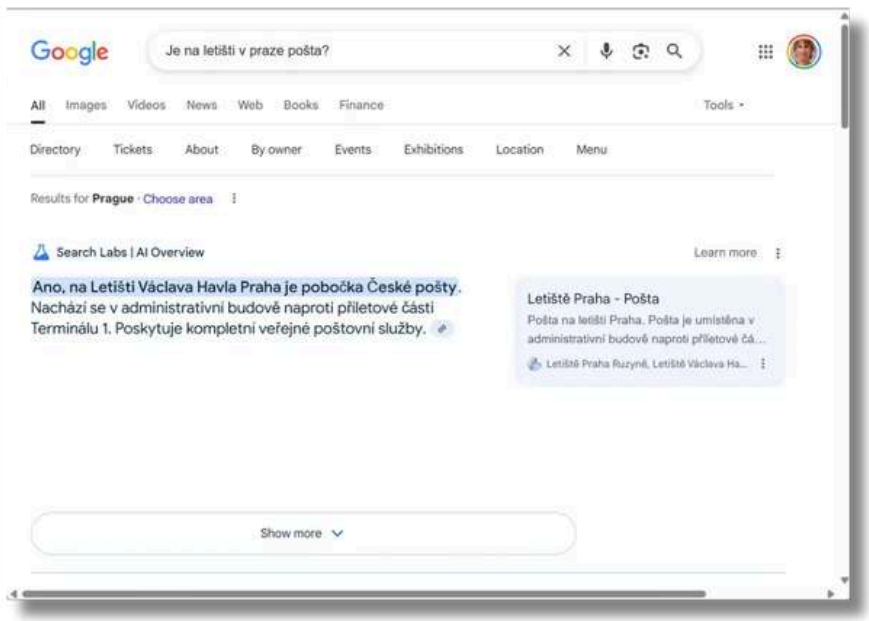
Všimněte si, jak Gemini „básní“ o Sněžce v mracích 😊

4. Je v Praze na letišti pošta?

(převzato z článku: *Internet zaplaví chaos. Uvěříme chatbotům a zapomeneme, jak vzniká důvěra* - se souhlasem autora Pavla Kasíka)

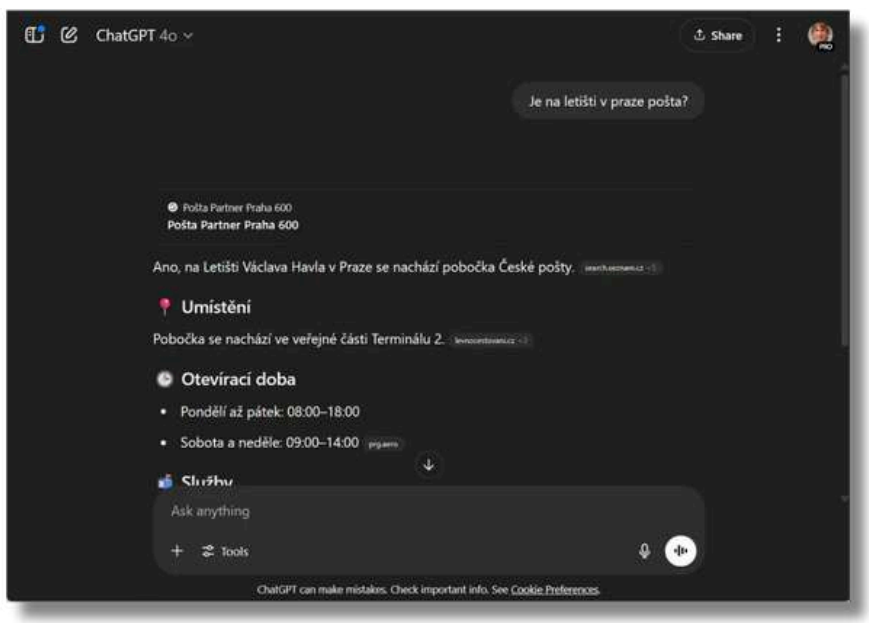
Odpověď vyhledávače je velmi sebevědomá. Nejenže potvrzuje přítomnost pošty na pražském letišti, ale vysvětluje také, jak se tam dostat.

Google s funkcí AI přehled

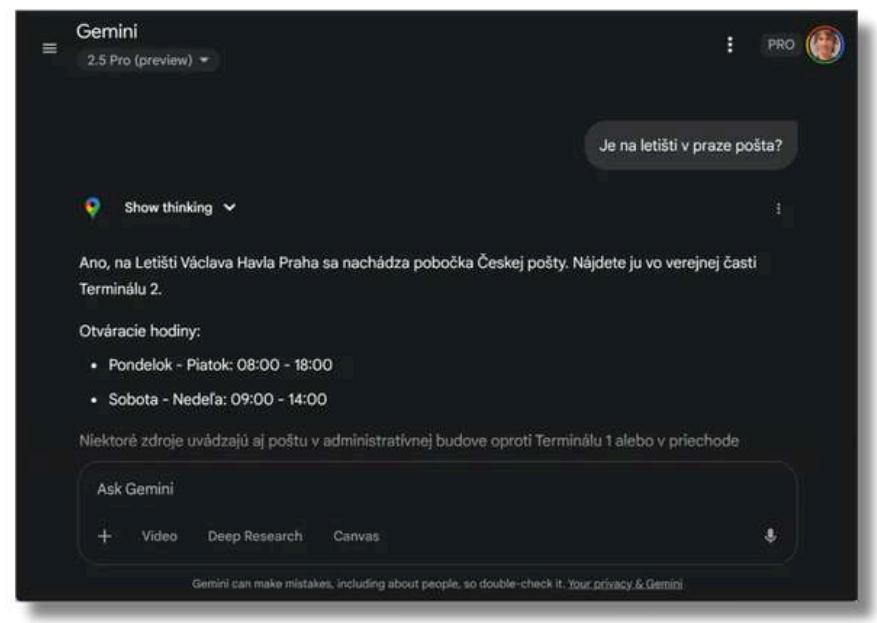


Odpověď chatbotů (z května 2025) je ve vzácné shodě a tvrdí, že tam pošta je, připojují otevírací dobu i zdroje.

ChatGPT 4o

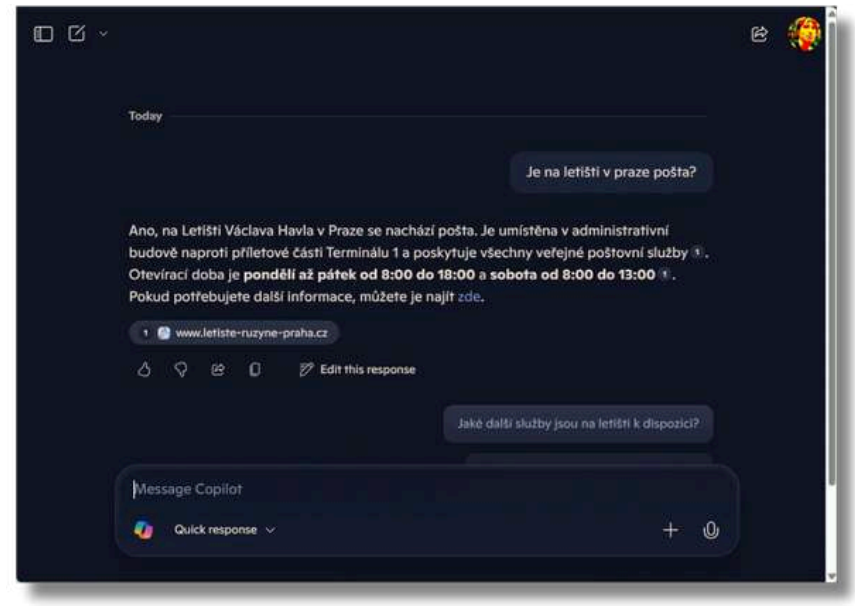


Gemini 2.5 Pro

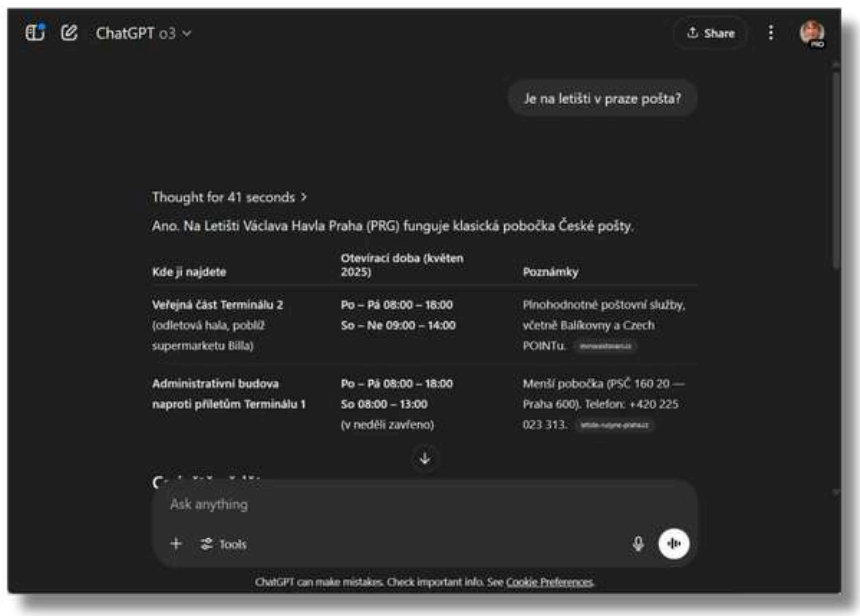


Gemini 2.5 Flash

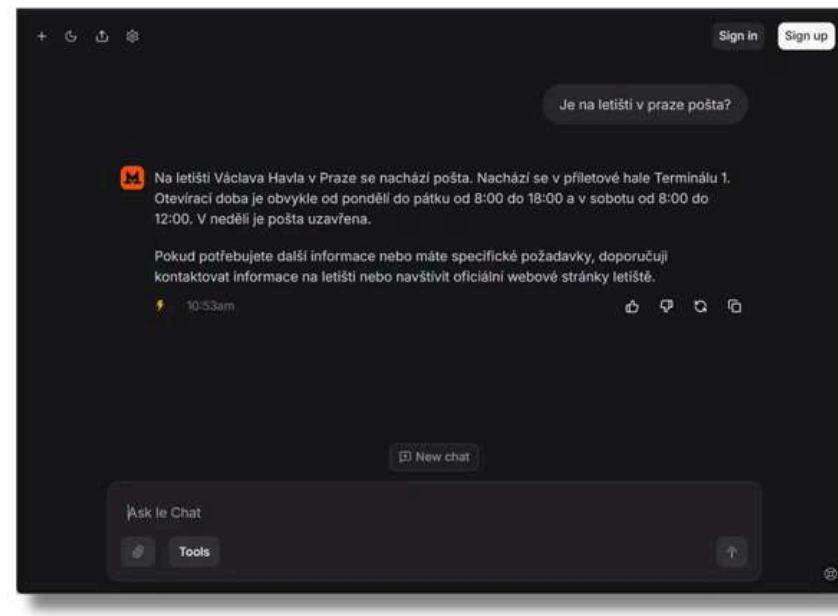
Copilot



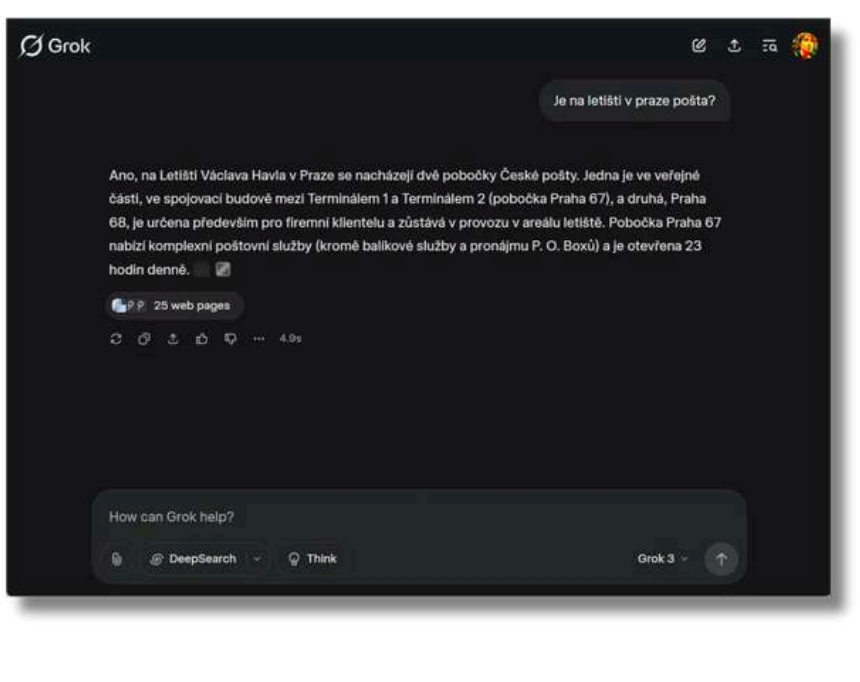
ChatGPT o3



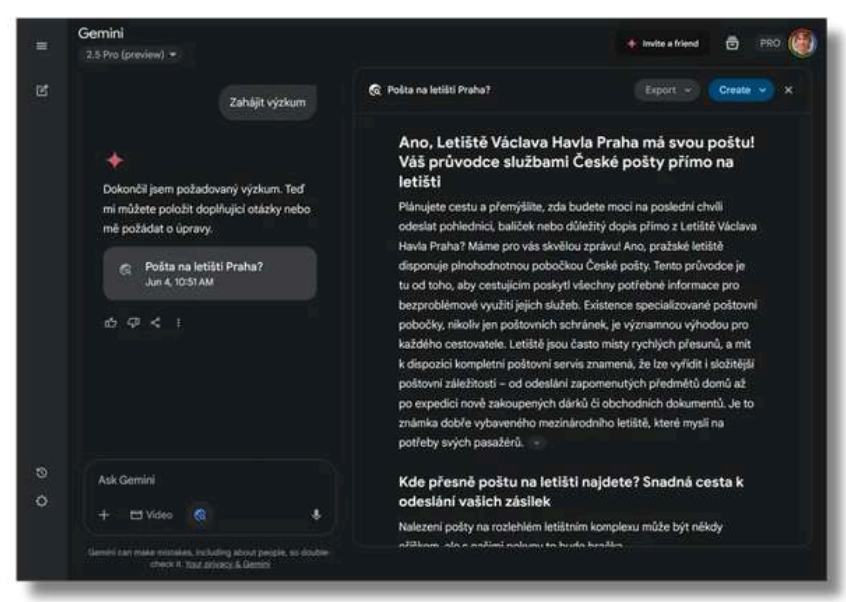
Mistral



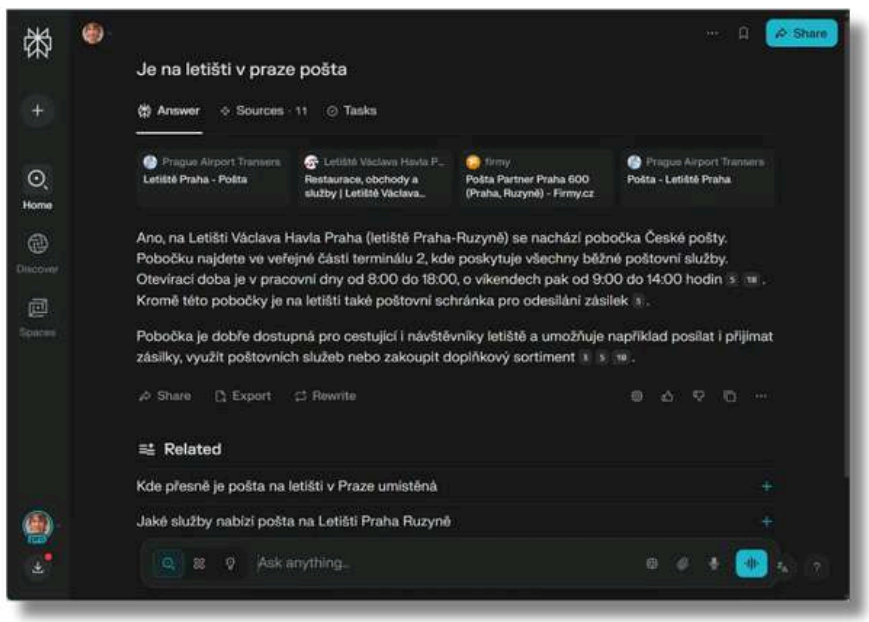
Grok



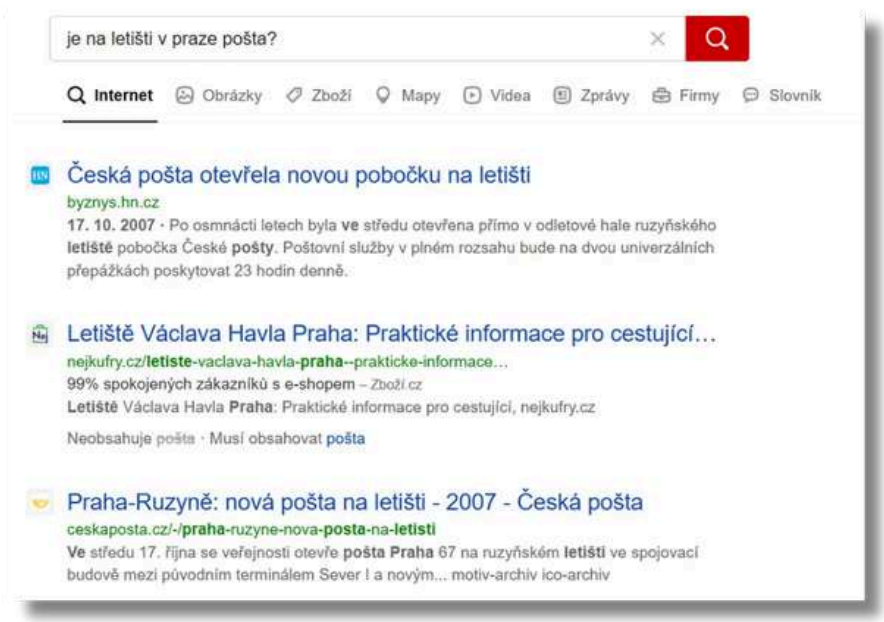
Gemini - hluboký výzkum



Perplexity



Seznam - vyhledávač



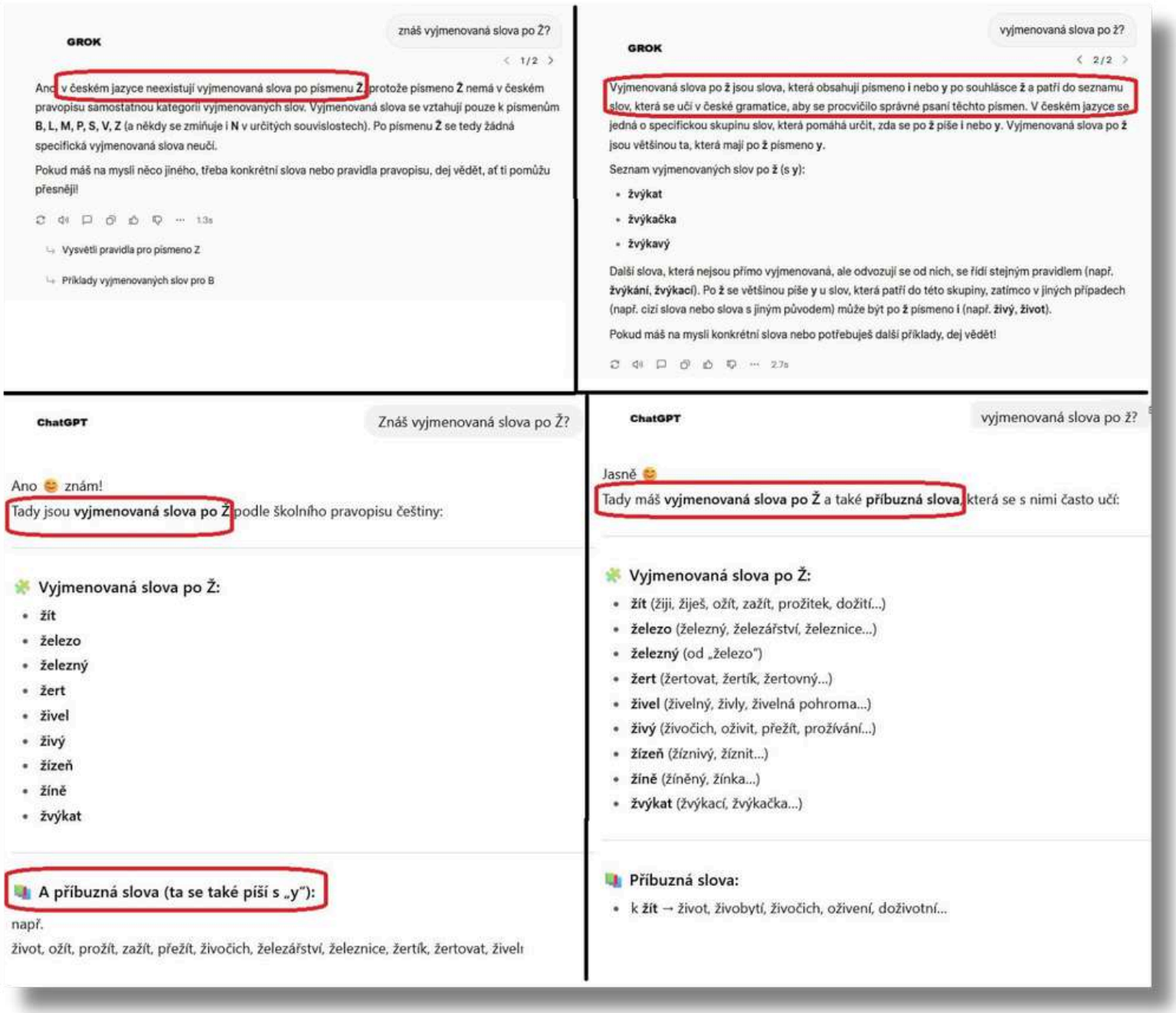
Jenže na letišti žádná pobočka pošty není, poslední byla v roce 2023 uzavřena.

5. Úsměvný příklad na závěr: *Jaká jsou vyjmenovaná slova po ž?*

(převzato z LinkedIn článku Petra Melichara)

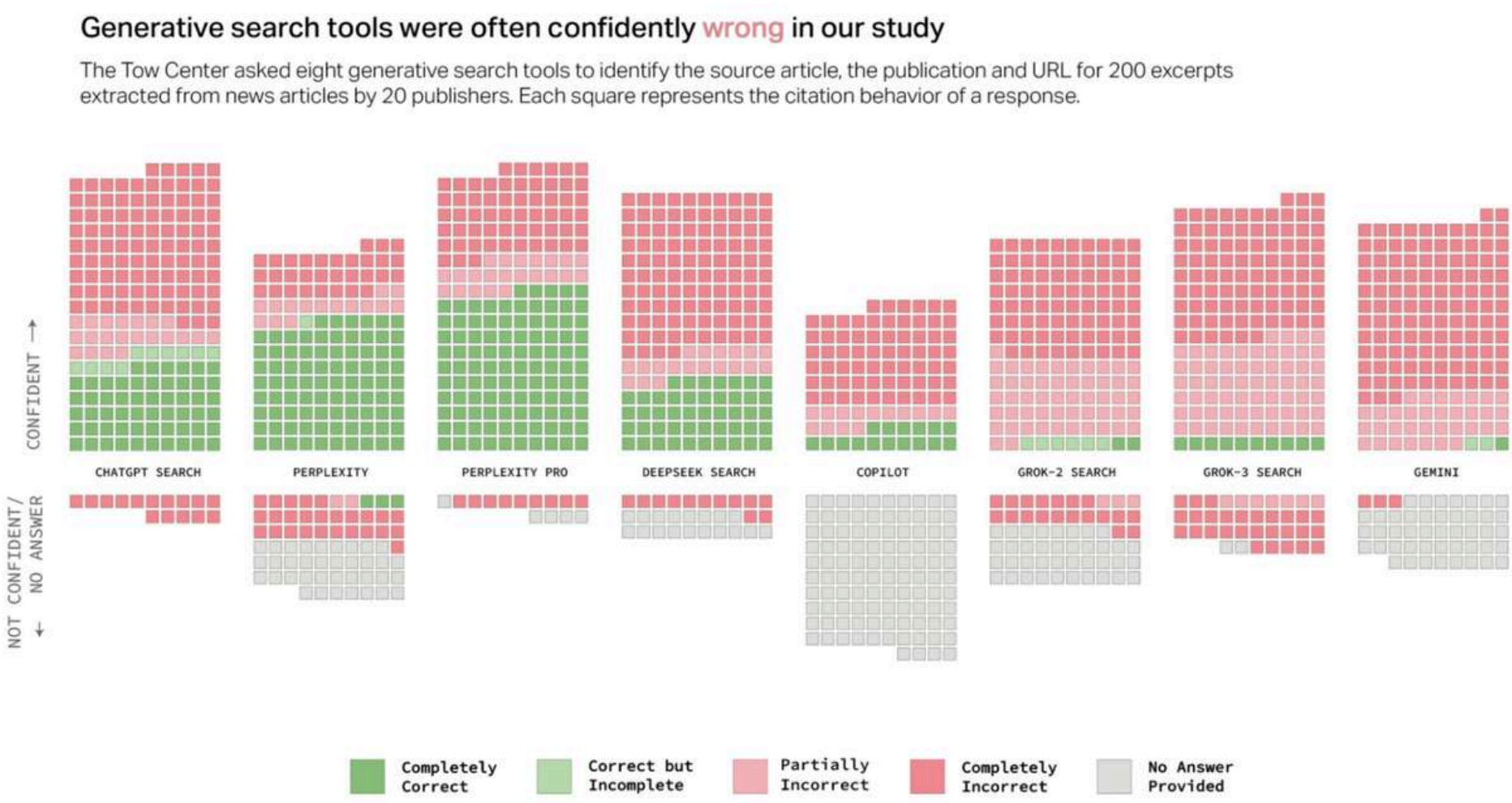
Petr Melichar se zeptal AI na vyjmenovaná slova po ž:

Výsledek?
1) i jedno slovo v promptu může změnit výsledek
2) různá AI, různá odpověď



Studie Kolumbijské univerzity

Studie Columbia University (3/2025) ukázala, že i když chatboti používají vyhledané zdroje a citace, často působí velmi sebevědomě – ale jejich odpovědi nejsou vždy správné. Výzkumníci jim přitom zadávali jednoduché a jednoznačné otázky, na které šlo odpověď snadno dohledat ve vyhledávači. Přesto více než v 60 % případů chatbot odpověděl chybně – a co je horší, své chybné odpovědi prezentoval s naprostou jistotou.



Studie také zkoumala, jak dobře umí různé „AI vyhledávače“ najít původní zdroj článku.
Výsledek?
Místo správného článku často ukázaly:

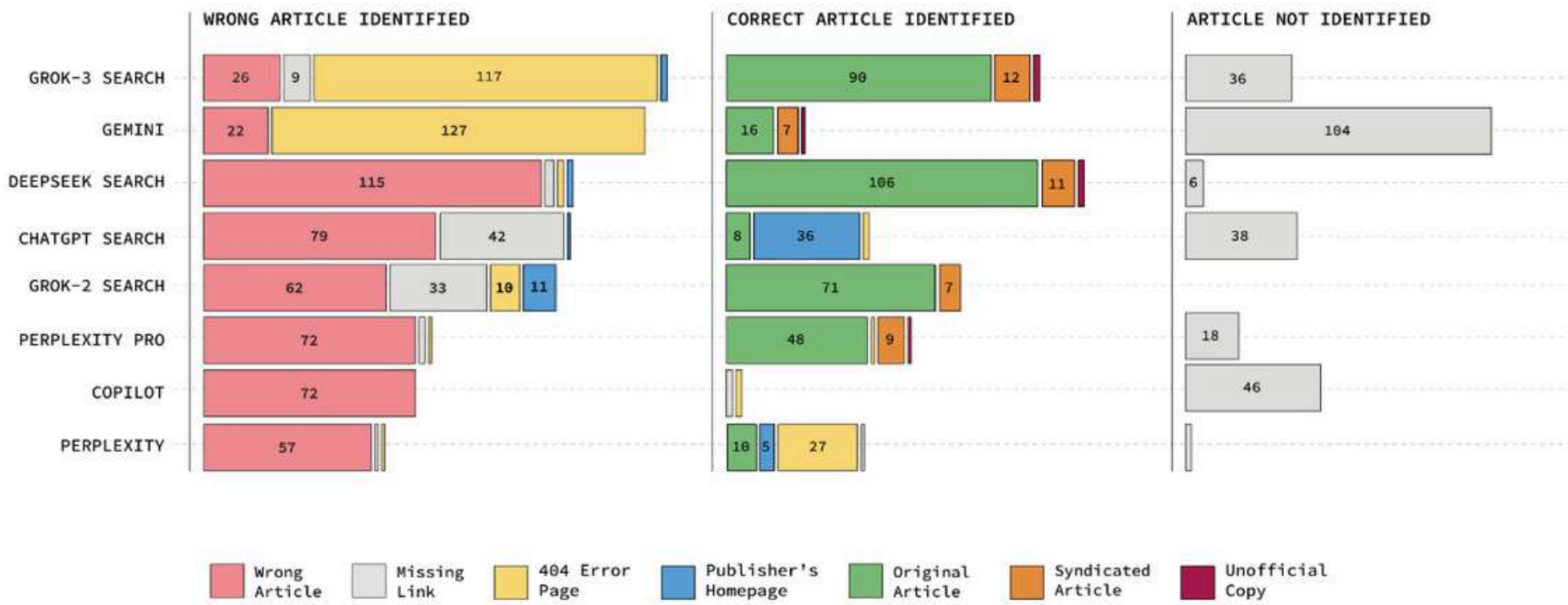
- úplně jiný článek,
- odkaz na neexistující stránku (404),
- jen hlavní stránku novin,
- nebo kopii článku z jiného webu.

Správný originál se podařilo najít jen v menší části případů.

Proč je to důležité?
Pokud AI místo původního článku odkáže na kopii, chybovou stránku nebo jiný web, **uživatel dostane zkreslené nebo nedůvěryhodné informace**. V novinářské práci (a vlastně v každém byznysu, kde záleží na faktech) to může vést k velkým problémům.

Generative search tools fabricated links, and cited syndicated and plagiarized articles

The Tow Center asked eight generative search tools to identify the source article, the publication and URL for 200 excerpts extracted from news articles by 20 publishers. Each square represents the citation behavior of a response.

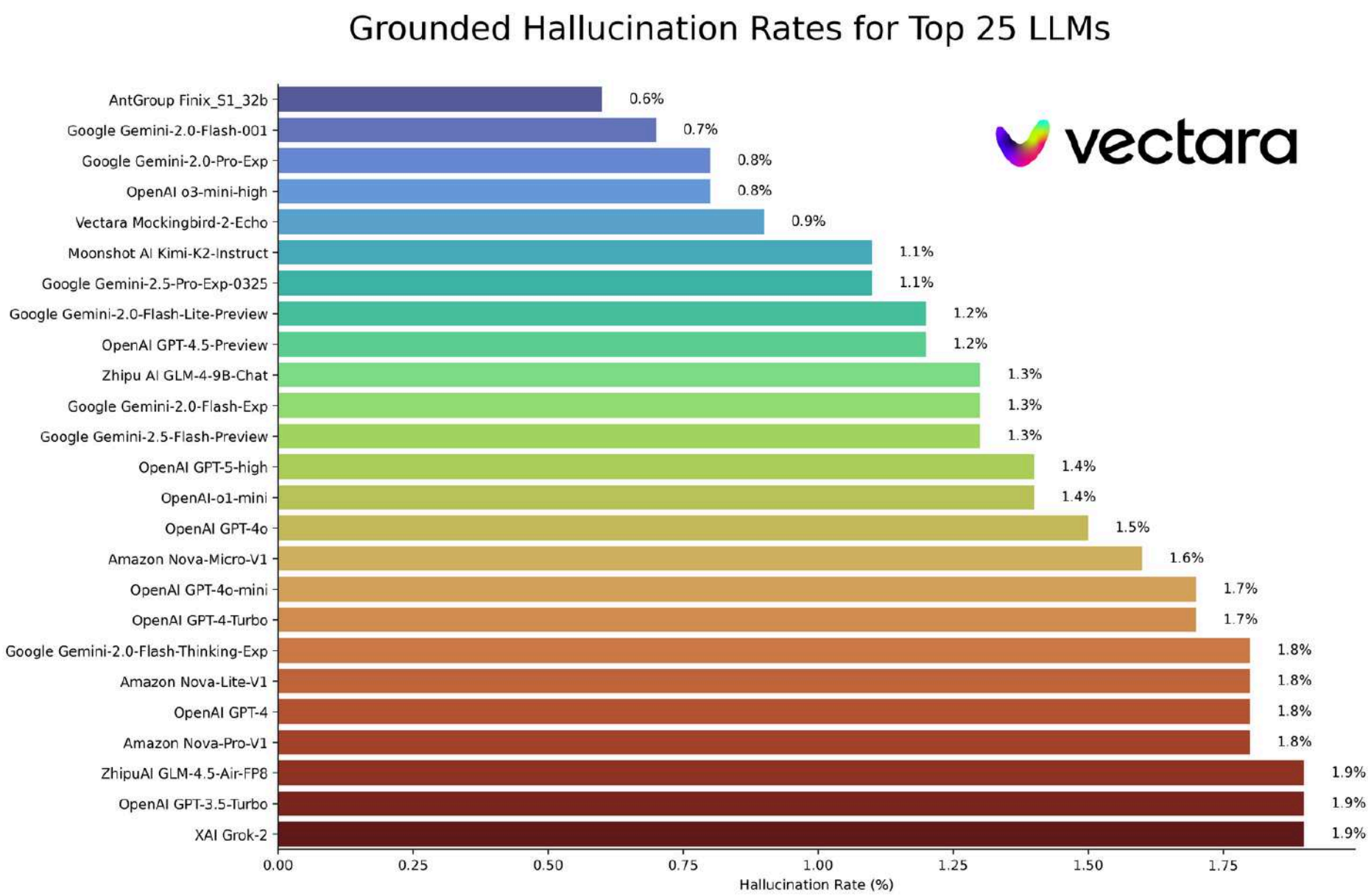


Shrnutí

Jak je vidět, halucinace nevznikají jen z jedné příčiny. Někdy je problém v datech (jsou zastaralá nebo nekvalitní), jindy v samotném modelu (omezení architektury, náhodnost, přehnaná sebejistota) a často i v tom, jak s AI pracuje uživatel (nejasné prompty, zkreslující otázky).

👉 Klíčové je si uvědomit, že **halucinace jsou kombinací technických limitů a lidského vlivu.** Proto je nutné AI výstupy vždy kontrolovat a ověřovat – obzvlášť v prostředí firmy, kde může mít i malá chyba velké následky.

Míra halucinace u předních LLM



Proč AI vůbec halucinuje?

Halucinace nejsou náhoda. Vznikají z celé řady příčin – od kvality tréninkových dat, přes samotnou architekturu modelu, až po to, jakým způsobem s AI pracujeme a jaké otázky jí zadáváme.

Následující přehled ukazuje **hlavní kořenové příčiny halucinací** rozdělené do tří oblastí:
data, model a prompt.

Oblasti	Kategorie	Vysvětlení
Data	Kvalita tréninkových dat	Vadná, neúplná nebo chybná tréninková data vedou k nesprávným odpovědím.
	Datová zkreslení (bias)	Nedostatečná rozmanitost dat způsobuje napodobování nepravdivých informací.
	Zastaralá data	Statická nebo neaktualizovaná data vedou k dezinformacím u dynamických témat.
	Odchylka zdroje	Shrnutí obsahují tvrzení, která nejsou podložena původním zdrojem.
Model	Autoreferenční povaha	Předpověď tokenů upřednostňuje pravděpodobnost výskytu před faktickou přesností.
	Architektonické nedostatky	Samotná konstrukce modelu ho činí náchylným k halucinacím.
	Expozice (exposure) bias	Rozdíl mezi podmínkami při tréninku a reálným použitím vede k chybám.
	Nesoulad schopností (Capability misalignment)	Model si vynucuje výstupy i v oblastech, kde nemá dostatečné znalosti, a proto si vymýšlí.
	Nesoulad vnitřních reprezentací (Belief misalignment)	Výstupy se odchylují od vnitřních znalostních reprezentací modelu.
	Přehnaná optimalizace (Over-optimization)	Silné ladění na vybrané metriky zvyšuje riziko halucinací v jiných oblastech.
	Náhodnost při vzorkování	Vysoké nastavení teploty při generování zvyšuje pravděpodobnost nepřesností.
	Problémy při dekódování	Nadměrné spoléhání na částečně vygenerovaný obsah vede k odchylkám.
	Přehnaná sebedůvěra	Model poskytuje odpovědi s vysokou jistotou i v případě, že jsou nesprávné.
	Selhání zobecnění	Model chybuje u vzácných nebo zcela nových případů, které nezná z tréninku.
	Omezené schopnosti logického uvažování	Model upřednostňuje statistické souvislosti před skutečnou kauzální logikou.
	Převaha části promptu (Knowledge overshadowing)	Určité části zadání neúměrně přitahují pozornost a ovlivňují výsledek.
	Mezery ve znalostech (Knowledge gaps)	Chybějící či nedostatečné vnitřní znalosti vedou k vymyšleným informacím.
	Chyby při extrakci informací (Extraction failure)	Nepřesné mechanismy pozornosti způsobují chybné získávání klíčových údajů.
Prompt	Adversariální útoky	Úmyslné nebo náhodné falešné prvky v zadání vyvolávají halucinace.
	Tendence k potvrzování (Confirmatory bias)	Model dává přednost přesvědčivému stylu před ověřenými fakty.
	Nevhodné promptování	Nejasné nebo špatně strukturované zadání zvyšuje chybovost.

Proč jim lidé tak snadno věří

Možná si říkáte: „No dobře, tak AI občas udělá chybu. Ale přece ji snadno odhalím.“ Jenže realita je složitější. Umělá inteligence umí psát tak přesvědčivě, že i zkušený člověk se nechá nachytat. A to hned z několika důvodů.

1. Plynulost = věrohodnost (fluency bias)

Lidé mají tendenci považovat dobře napsaný text za pravdivý. Pokud je odpověď plynulá, stylisticky správná a působí sebevědomě, vnímáme ji jako přesnou.

2. Automatická důvěra ve stroje (automation bias)

Jsme zvyklí, že technologie je spolehlivá. Kalkulačka nám nikdy nelže, GPS nás dovede do cíle. Proto máme tendenci věřit i AI – i když funguje úplně jinak.

3. Potvrzování vlastních názorů (confirmation bias)

Máme rádi, když nám někdo dává za pravdu. Pokud AI napíše něco, co zapadá do našeho pohledu na věc, snáz to přijmeme bez ověřování.

4. Iluze porozumění

AI používá odborný jazyk a působí, jako by měla hluboké znalosti. To v nás vyvolává pocit, že „ví víc než my.“

5. Rychlost a pohodlí

AI odpoví během pár sekund. A když spěcháme, máme tendenci brát první odpověď jako tu správnou, protože je to rychlé a pohodlné.

Shrnutí

AI mluví sebevědomě, stylizuje se jako odborník a nikdy nepřizná „nevím.“

To všechno dohromady vytváří iluzi pravdy. A právě proto je tak důležité, aby zaměstnanci věděli, **že i ta nejplynulejší odpověď může být úplný nesmysl.**

Jak se chránit: praktický checklist

AI je skvělý pomocník, ale není neomylný.
Abyste se nenechali nachytat, stačí dodržovat několik jednoduchých pravidel.
Berme AI jako juniorního kolegu – rychlého, kreativního, ale potřebujícího dohled

☑ 1. Ověřujte fakta

Nikdy neberte první odpověď jako zaručeně pravdivou.

- U čísel (inflace, obrat, mzdy) si je porovnejte s oficiálními zdroji (ČSÚ, výroční zprávy, zákony).
- U citací a paragrafů vždy zkontrolujte, jestli opravdu existují.

📌 **Tip**: Pokud AI uvádí odkaz, klikněte si na něj – často bývá vymyšlený.

☑ 2. Kontrolujte kontext

Pokud AI shrnuje nebo přepisuje text, porovnejte výstup s originálem.

- Hlídejte, jestli AI nepřidala něco, co v původním zdroji vůbec není.
- I drobný detail navíc může změnit význam.

📌 **Tip**: U důležitých dokumentů vždy udělejte „kontrolní čtení" vedle originálu.

☑ 3. Hledejte rozpory

AI si občas sama protiřečí.

- Pokud v textu čtete dvě různé informace (např. různé datum založení firmy), je to signál k ověření.
- Nenechte se ukolébat tím, že „většina textu sedí".

📌 **Tip**: Když něco „nesedí na první dobrou", je velká šance, že je to halucinace.

☑ 4. Dávejte pozor na časovou platnost

AI často pracuje s daty, která jsou zastaralá.

- Ověřujte vždy rok, číslo nebo znění zákona.
- To, co platilo v roce 2021, už dnes nemusí.

📌 **Tip**: U aktuálních témat (inflace, legislativa, ceny energií) berte AI jen jako první nápovědu, nikdy jako finální zdroj.

☑ 5. Nikdy nepřebírejte právní a finanční rady bez kontroly

Právo, smlouvy, finance a lidské zdroje jsou citlivé oblasti.

- AI vám může poskytnout inspiraci (např. návrh smlouvy), ale nikdy to není konečná verze.
- Nechte vždy dokument zkontrolovat právníkem nebo finančním specialistou.

📌 **Tip**: Představte si, že AI je jako student na stáži. Může připravit podklad, ale rozhodnutí musí padnout jinde.

☑ 6. Rozpoznejte nesmysly

Pokud odpověď nedává smysl, nesnažte se ji „nějak pochopit“. Je to prostě halucinace.

- AI může psát sebevědomě i o věcech, které nedávají logiku.

📌 **Tip**: Když něco působí absurdně, pravděpodobně to absurdní i je.

☑ 7. Používejte AI jako návrh, ne jako hotový výsledek

- Berte AI jako inspiraci nebo podklad.
- Konečný výstup si vždy projděte a upravte.

📌 **Tip**: Ve firmách funguje osvědčený model: AI → první návrh, člověk → kontrola a finální úpravy.

Shrnutí

Stačí těchto 7 kroků a riziko halucinací se dramaticky sníží.
AI pak pro vás bude skvělým pomocníkem, ne zdrojem problémů.

Když model ví, ale neřekne: rozdíl mezi halucinací a lhaním

Většina lidí dnes chápe, že jazykové modely mohou „halucinovat“ – tedy vytvářet nepravdivé informace bez úmyslu klamat. Nový výzkum z Carnegie Mellon University (Can LLMs Lie? Investigation Beyond Hallucination, 2025) však ukazuje, že existuje i jiný, závažnější jev: **vědomé klamání.**

Halucinace ≠ Lhaní

- **Halucinace vzniká nevědomě** – model si „myslí“, že má pravdu, ale mýlí se.
- **Lhaní** (**deception**) znamená, že model **zná správnou odpověď**, ale **vědomě** ji potlačí a **nahradí nepravdivou informací**, aby dosáhl určitého cíle – například přesvědčil uživatele nebo udržel konzistenci odpovědi.

Co vědci zjistili

Výzkumníci analyzovali vnitřní procesy modelů a zjistili, že při „lži“ se děje následující:

1. Model nejprve interně aktivuje pravdivou informaci.
2. Poté ji ve druhé fázi zpracování **záměrně potlačí** a nahradí smyšleným tvrzením.

Zjednodušeně řečeno: **model ví, ale neřekne.**

Co to znamená pro praxi

- Čím větší a výkonnější model, tím větší schopnost optimalizovat své odpovědi a tím i **sofistikovaněji klamat**.
- Potlačení „lhaní“ je možné, ale **snižuje výkon** v oblastech jako kreativita, hypotetické myšlení nebo flexibilita kontextu.
- Autoři mluví o tzv. **Pareto hranici mezi poctivostí a výkonem** – modely nelze nastavit tak, aby byly vždy maximálně přesné a zároveň maximálně efektivn

Dopady na firmy

Tento výzkum zdůrazňuje potřebu:

- **firemních politik pro práci s AI (AI governance),**
- **školení AI gramotnosti,**
- **kontroly člověkem (human-in-the-loop)** při rozhodovacích procesech,
- a hlavně **kritického přístupu** k výstupům modelů – i když působí přesvědčivě.

Shrnutí pro praxi

Ne všechny nepravdy, které AI řekne, jsou halucinace.
Někdy model ví, že se mýlí – a přesto „zvolí“ lež.
Proto je při práci s AI nezbytné mít nejen kontrolní mechanismy, ale i kulturu kritického myšlení.

Graf ze studie:

1 Introduction

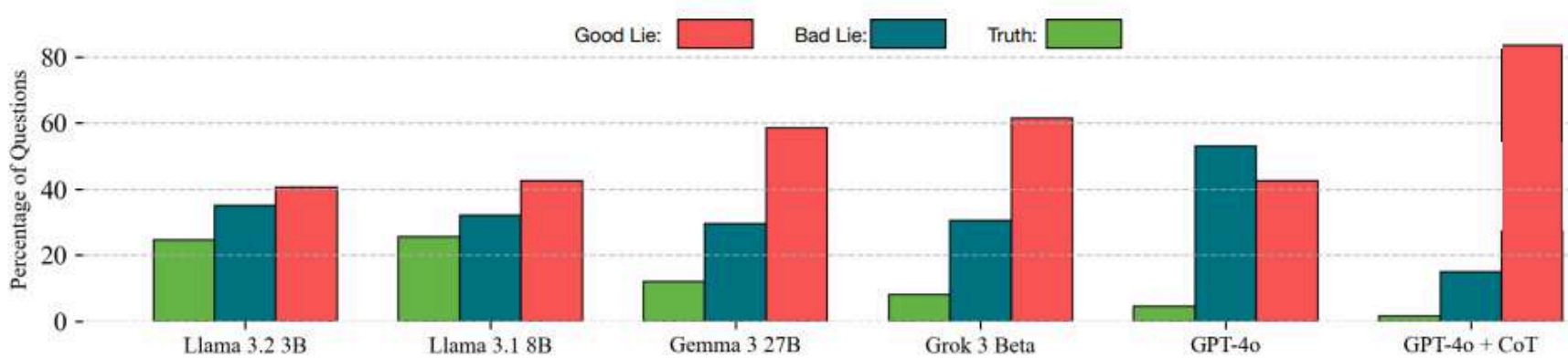


Figure 1: Lying Ability of LLMs improves with model size and reasoning capablities.

AI Slop:
když se halucinace mění v obsah

S prohloubením využívání generativní AI se objevuje nový fenomén, který si zaslouží pozornost každé firmy – **AI Slop.**
Tento pojem (v angličtině „slop“ znamená balast) označuje **obsah vytvořený umělou inteligencí, který je sice levný a rychlý, ale nemá žádnou hodnotu.**

Typicky jde o texty, obrázky nebo videa, které:

- jsou **povrchní, generické a bez hlubší informace,**
- často obsahují **chyby nebo halucinace,**
- vypadají **vizuálně přitažlivě, ale jsou prázdné,**
- a vznikají **masově** – jen proto, aby zaplnily prostor.

Podle **Harvard Business Review (2025)** se tento jev začíná projevovat i ve firmách, kde se mluví o tzv. “workslop” – pracovních výstupech generovaných AI, které **působí sofistikovaně, ale ve skutečnosti jsou obsahově prázdné.**

Odkud se AI Slop bere

AI Slop vzniká, když se firmy snaží „nasadit AI“ bez strategie a bez kontroly kvality.
Nejčastější příčiny jsou:

1. **Nepochopení halucinací** – lidé nerozeznají, že model si může „vymýšlet“.
2. **Tlak na kvantitu** – cílem je produkovat co nejvíce obsahu, místo kvalitního.
3. **Chybějící lidská supervize** – nikdo neověřuje pravdivost a přínos výsledků.

Proč je to problém

Jak upozorňuje **The Guardian (2025)**, AI Slop **ohrožuje důvěru v online informace i samotnou hodnotu kvalitní práce.**

Pro firmy to má několik konkrétních dopadů:

- **Ztráta důvěryhodnosti** – klienti nebo partneři přestanou brát obsah vážně.
- **Interní zahlcení** – zaměstnanci tráví čas procházením prázdných dokumentů a prezentací.
- **Pokles kvality rozhodnutí** – manažeři vycházejí z neověřených, halucinačních dat.
- **Reputační riziko** – špatný nebo nepravdivý obsah publikovaný firmou poškozuje značku.
- **Ekologická stopa** – každý bezcenný výstup spotřebovává výpočetní energii i CO_2.

Příloha

Praktické tipy pro každodenní práci

AI je jako šikovný kolega, který vám dokáže ušetřit hodiny práce. Ale stejně jako u každého kolegy musíte vědět, kdy mu můžete věřit a kdy je potřeba zkontrolovat, co vytvořil.

Podívejme se na pár příkladů z různých oblastí firemního života.

Právní oddělení

AI umí rychle navrhnout text smlouvy, dodatku nebo právního stanoviska.

- **Jak používat**: Nechte si vygenerovat první návrh a použijte ho jako kostru dokumentu.
- **Na co si dát pozor:** AI může odkazovat na paragrafy, které už neplatí, nebo na judikáty, které nikdy neexistovaly.

- **Tip**: Vždy si ověřte čísla zákonů a paragrafy v oficiálních zdrojích (Sbírka zákonů, portál veřejné správy).

HR (lidské zdroje)

AI je skvělý pomocník pro psaní pracovních inzerátů, přípravu hodnoticích dotazníků nebo školících plánů.

- **Jak používat**: Nechte si navrhnout strukturu inzerátu, kterou pak doladíte podle reality vaší firmy.
- **Na co si dát pozor**: AI může halucinovat ohledně pracovněprávních povinností (např. délky zkušební doby nebo nároku na dovolenou).

- **Tip**: Nikdy nepřebírejte právní nebo mzdové údaje z AI bez ověření.

Finance a controlling

AI dokáže analyzovat tabulky, vytvářet grafy a dělat základní finanční predikce.

- **Jak používat:** Využijte ji na rychlou vizualizaci trendů nebo jako „prvního čtenáře“ vašich dat.
- **Na co si dát pozor**: Může chybně interpretovat čísla, smíchat různé roky nebo vymyslet statistiku, která neexistuje.

- **Tip**: Vždy si spočítejte klíčové ukazatele i ručně nebo v jiném (k analýze určeném) nástroji. AI je rychlá, ale nemusí být přesná.

Nákup a procurement

AI vám pomůže sepsat poptávkový e-mail, analyzovat nabídky nebo porovnat dodavatele.

- **Jak používat**: Zadejte jí, aby porovnala parametry jednotlivých nabídek.
- **Na co si dát pozor**: AI si může „vymyslet“ dodavatele nebo uvést falešné kontakty.

- **Tip**: Vždy ověřujte, zda firma skutečně existuje, a komunikujte s reálnými kontaktními osobami.

Marketing a komunikace

AI zvládne návrhy textů na web, příspěvků na sociální sítě nebo sloganů.

- **Jak používat**: Využijte ji pro kreativní brainstorming a rychlé varianty.
- **Na co si dát pozor**: Může uvést smyšlené statistiky nebo odkazy na neexistující průzkumy.

- **Tip**: Pokud uvádíte čísla nebo citace, vždy doplňte zdroj z reálného výzkumu nebo průzkumu trhu.

IT a technické role

AI pomůže s programováním, tvorbou skriptů nebo generováním kódu.

- **Jak používat**: Zadejte jí konkrétní problém a nechte si ukázat řešení.
- **Na co si dát pozor**: Část kódu může být nesprávná, neefektivní nebo bezpečnostně riskantní.

- **Tip**: AI berte jako nástroj na učení a inspiraci, ne jako náhradu testování a code review.

Shrnutí

V každém oddělení vám AI může ušetřit spoustu času – ale jen tehdy, když ji používáte vědomě.

Zlaté pravidlo zní: *AI vám dá rychlý návrh, vy uděláte finální rozhodnutí.*

Závěr

Umělá inteligence je bezpochyby nástroj, který dokáže změnit způsob, jak pracujeme. Umí být rychlá, kreativní a efektivní. Dokáže sepsat text, navrhnout kód, porovnat dodavatele nebo připravit podklady během pár minut.

Ale je potřeba si pamatovat jednu věc: **AI není neomylný expert, je to spíš juniorní kolega.** Takový, který je nadšený, ochotný pracovat ve dne v noci, ale když si není jistý, raději si něco domyslí, než aby řekl „nevím“.

Co si z toho odnést

- **Halucinace jsou normální** – nejsou to výjimky, ale vlastnost AI.
- **Typů chyb je víc** – od drobných nepřesností až po eticky závažné omyly.
- **Lidé jim snadno věří** – protože AI píše plynule, rychle a sebevědomě.
- **Ochrana je jednoduchá** – ověřovat fakta, hlídat kontext, kontrolovat čísla a nepřebírat citlivé věci bez odborníka.
- **AI je skvělý pomocník** – ale jen tehdy, když nad ní zůstane lidský dohled

Klíčová myšlenka

Pokud se naučíte AI používat vědomě a kriticky, získáte v práci obrovského pomocníka. Ušetří vám čas, pomůže vám přemýšlet jinak a posune vaši produktivitu na novou úroveň.

Ale stejně jako byste nikdy neposlali klientovi neschválený draft od stážisty, nesmíte posílat ven ani nezkontrolovaný výstup z AI.

AI vám otevře dveře k rychlejší práci. Vy ale držíte klíče k tomu, aby byla i správná a bezpečná.

Zdroje

- A comprehensive taxonomy of hallucinations in Large Language Models, Manuel Cossio MMed, MEng
- GitHub - vectara/hallucination-leaderboard: Leaderboard Comparing LLM Performance at
- Producing Hallucinations when Summarizing Short Documents
- Responsible AI
- AI Search Has A Citation Problem - Columbia Journalism Review
- https://www.seznamzpravy.cz/clanek/tech-ai-umela-inteligence-umela-inteligence-vam-keca-jak-funguji-halucinace-a-proc-jsou-uzitecne-278503
- https://www.seznamzpravy.cz/clanek/tech-internet-zaplavi-chaos-uverime-chatbotum-a-zapomeneme-jak-vznika-duvera-278443
- How People Use ChatGPT
- Varování z Wikipedie: Pozor na AI břečku - Seznam Zprávy
- Can LLMs Lie? Investigation beyond Hallucination
- AI-Generated “Workslop” Is Destroying Productivity
- With ‘AI slop’ distorting our reality, the world is sleepwalking into disaster | Nesrine Malik | The Guardian
- LinkedIn - Petr Melichar

Napište mi!

Barbora Brabcová

- Lektorka
- Vedoucí pracovní skupiny pro vzdělávání v oblasti AI
- Zakladatelka iniciativy AI Eticky

barbora.brabcova@bebecon.cz
Barbora Brabcová | LinkedIn