



TECHNOLOGY
CENTRE PRAGUE

SEMINÁŘ AI PRO DATOVÉ ANALYTIČKY A ANALYTIKY

4. března 2026



SEMINÁŘ AI PRO DATOVÉ ANALYTIČKY A ANALYTIKY

- ❖ Část 1: Kam směřuje analytická práce (a mnohem víc)

9:30 – 10:30

- ❖ Část 2: Praktické využití nástrojů AI v analýze

10:40 - 11:30

- ❖ Část 3: Pokročilé techniky a koncepty (a čas na diskusi)

12:30 – 14:00



Odhadněte, kolik energie datacentra globálně spotřebovala v roce 2024 (v TWh)? Náповěda: V Česku činila celková (čistá) spotřeba veškeré elektřiny v roce 2024 asi 58 TWh.



Kdo je autorem následujícího citátu?

"Umělá inteligence pravděpodobně povede ke konci světa...ale do té doby vzniknou opravdu skvělé firmy."



TECHNOLOGICKÉ
CENTRUM PRAHA

ČÁST 1: KAM SMĚŘUJE ANALYTICKÁ PRÁCE

—
Kristýna Meislová & Adéla Kučerová
Seminář AI pro datové analytičky a analytiky, 4. 3. 2026



Můžete se ptát i online.

slido.com

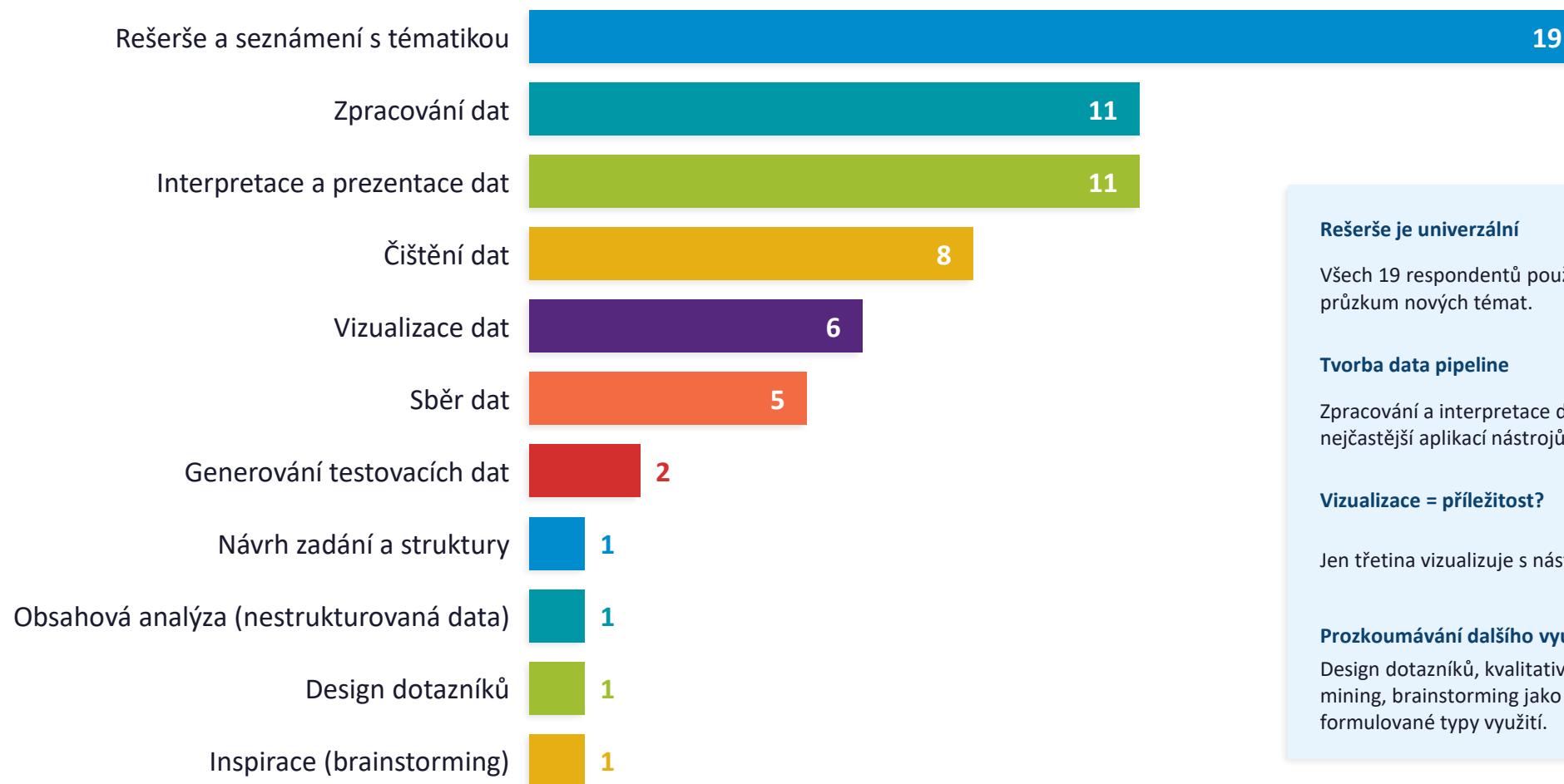
kód # 1820 4031



AGENDA

- Výsledky dotazníkového šetření
- Welcome to the loop a analytická práce v éře AI
- Zajímavé nástroje pro analytickou práci (které nás zaujaly)
- Vybraná specifika LLM (a nástrojů AI)
- Shrnutí a tipy na samostudium

DOTAZNÍK: K ČEMU VYUŽÍVÁTE NÁSTROJE AI



Rešerše je univerzální

Všech 19 respondentů používá AI pro průzkum nových témat.

Tvorba data pipeline

Zpracování a interpretace dat jsou další nejčastější aplikací nástrojů AI.

Vizualizace = příležitost?

Jen třetina vizualizuje s nástroji AI.

Prozkoumávání dalšího využití

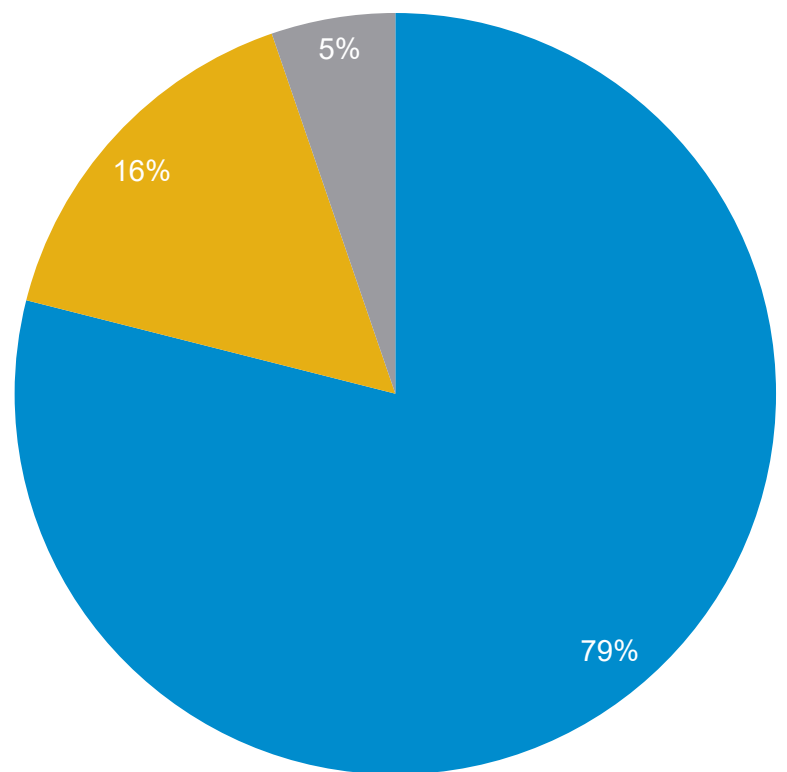
Design dotazníků, kvalitativní analýza a text mining, brainstorming jako další konkrétněji formulované typy využití.

DOTAZNÍK: POUŽÍVANÉ NÁSTROJE I

Nástroj	Podíl	Typické akce	Výhoda nástroje
ChatGPT	84 %	generování nápadů, hluboký výzkum, brainstorming, feedback, řešerše dokumentů, generování textů, analytické postupy, generování kódu, obsah textů, obecné dotazy	<i>Univerzalita</i>
Gemini	42 %	obsah textů, interpretace, úprava textů, vyhledávání informací, Python/SQL skripty, online řešerše	<i>Google integrace a kontextové okno</i>
Copilot	32 %	práce s tabulkami, hledání zdrojů, úprava textů, integrace MS Office, Power BI	<i>MS 365 integrace</i>
Claude	21 %	analytické otázky, komplexní kódování, nebo návrhy a úprava kódu	<i>Hloubka a flexibilita analýzy</i>
NotebookLM	21 %	seznámení s problematikou, interpretace grafů, zpracování zdrojů	<i>Strukturovaný multimodální výstupy</i>
Perplexity	5 %	vyhledání zdrojů, definice	<i>AI-first search (hledání relevantních zdrojů)</i>

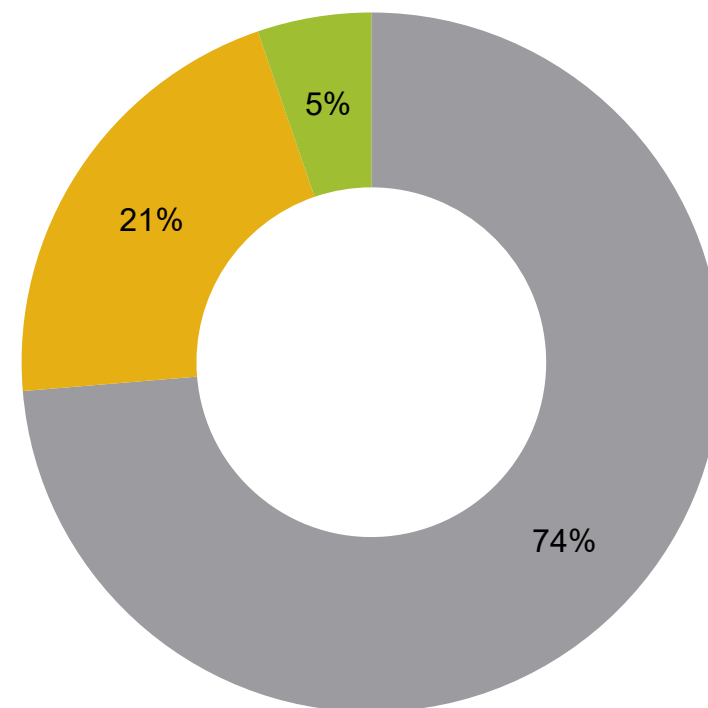
DOTAZNÍK: POUŽÍVANÉ NÁSTROJE II

Kolik AI nástrojů používáte



■ 2+ nástrojů (79 %) ■ 1 nástroj (16 %) ■ Žádný/nejasné (5 %)

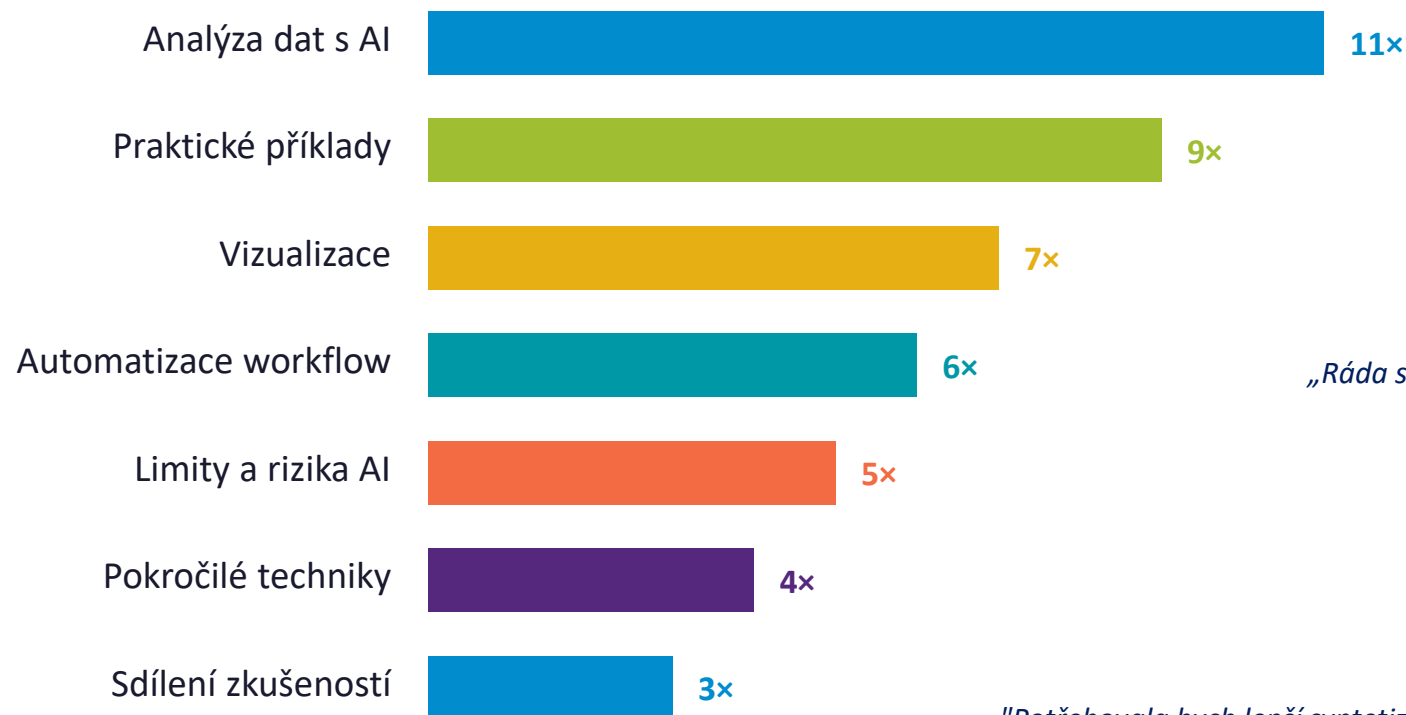
Adopce AI kódovacích nástrojů



■ Ne / nevím co to je (74 %) ■ Občas Python/kód (21 %) ■ Ano, aktivně (5 %)

DOTAZNÍK: CO VÁS ZAJÍMÁ

Témata v otevřené otázce o očekáváních na dnešní seminář



„Jak se nejlépe vyhnout AI generovaných nepřesných informací?“

„Ráda se dozvím nějaké vychytávané pracovní postupy s AI pro naši práci“

„Do jaké míry je možné spolehnout se na výstupy AI ve chvíli, kdy jí krmím vlastními daty?“

"Potřebovala bych lepší syntetizaci a pomoc při analýze rozsáhlých nestrukturovaných dat"

DOTAZNÍK: PROFIL SKUPINY

Kam se řadíte na škále AI adopce?

Začátečníci

Základní řešerše obvykle s jedním nástrojem



16 %

Běžní uživatelé

Více nástrojů, zpracování dat, prezentace



53 %

Pokročilí uživatelé

Kombinace nástrojů, kódování, validace



26 %

Experti

Benchmarking, stavění pipelines, automatizace



5 %

Skupina je heterogenní. Pokročilejší než průměrná česká populace. → Tipy pro všechny úrovně, od základních konceptů po pokročilé techniky.

WELCOME TO THE LOOP

- **Human-in-the-loop**...nebo Human-on-the-loop, Human-above-the-loop, Human-in-Power, Meaningful Human Control?

TL;DR: Vrstvy nebo různé úhly pohledu na stejný problém.

Jak zachovat lidskou suverenitu v době, kdy AI jedná autonomně?

HITL, HOTL, HATL – modely interakce (mikromanagement, supervize, strategie)

HiP – filozofický požadavek (člověk je primární subjekt poznání, AI nesmí budovat falešný pocit neomylnosti)

MHC – technický standard (podpořený např. EU AI Actem - MHC by Design)

WELCOME TO THE LOOP

Analytická práce se přesouvá od „manuální práce“ s daty a informacemi ke strategičtějším úkolům.

Mezi klíčové kompetence patří:

- **Context engineering:** Místo prostého promptování analytik*čka nastavuje kontext, pravidla a priority, ze kterých nástroje AI vyvozují své kroky.
- **Orchestrace workflow:** Schopnost řetězit nástroje, asistenty, agenty. (Pozor na to, že za to, co dělá váš systém jste odpovědní.)
- **Etický dohled:** Zajištění, že (polo)autonomní systémy neporušují etické mantinely ani zákonné normy, jsou transparentní a férové.
- **Validační myšlení, validační přístup:** Nejen kontrola výsledku, ale audit a strategické řízení/volba procesu.

→ **Neustálý upskilling je pro analytickou práci nezbytný.**

VALIDAČNÍ PŘÍSTUP

~~Která AI je nejlepší?~~

- Využívání množství nástrojů (případně tzv. nejlepší nebo nejnovější model na trhu) neznamena automaticky lepší výsledky nebo vyšší produktivitu.
- Hlubší expertíza a lepší workflows ano.

Jaký je NEJLEPŠÍ nástroj a přístup pro DANÝ úkol?

- Co pro mě v daný moment znamená NEJLEPŠÍ nástroj a postup?
- Kdy stačí volba nástroje podle obecného hodnocení modelů a nástrojů? (zamýšlený typ úkolu)
- A co když potřebuji validovat model pro složitější analytické projekty? (A cítím, že obecný benchmark/leaderboard nestačí?)

CO SE V ANALYTICE AKTUÁLNĚ DĚJE

- **Agentic Middleware:** AI nástroje přestávají být cílovým rozhraním a stávají se „neviditelnou“ vrstvou běžící na pozadí systémů.
- **AI dohlíží na AI:** Autonomní kontrolní systémy, které monitorují chování jiných AI agentů a člověka upozorňují na odchylky.
- **Systémy před modely:** Nejúspěšnější nejsou „nejlepší modely“, ale komplexní řešení, která nejlépe zvládají orchestraci, zabezpečení a propojení.
- Rostoucí poptávka po „multilingvních“ analytických*čkách
- Zaměření na **Sovereign AI** (Geopatiation)
 - Vlastní datová a modelová infrastruktura, pro kontrolu citlivých dat a nezávislost na zahraničních poskytovatelích

...ALE NEČEKEJTE

- AGI (obecná umělá inteligence)
- Bezchybnost LLM a AI nástrojů
- Nahrazení lidské analytické práce
- Vyřešení tzv. AI-bias (zkreslení v modelech)
- Infrastrukturní připravenost na masivní AI provoz

VÝBĚR ZAJÍMAVÝCH NÁSTROJŮ A ŘEŠENÍ PRO ANALYTICKOU PRÁCI

■ Claude (Cowork)...a AI asistenti kódu, kteří jsou designovaní pro širší spektrum uživatelů

- Marketovaný jako "Claude Code pro netechnické uživatele". K dispozici v placených verzích. Vzniknul jako reakce na dynamický rozvoj ve využívání Claude Code nad rámec klasické vývojářské práce. Má ambice zpřístupnit agentní schopnosti širokému publiku.
- OpenAI Codex CLI model GPT 5.3 (šampion v rychlosti – 1000 tokenů/s, rychlé skriptování), Claude Code Opus 4.6 (automation friendly, plan mode, koordinuje subagenty, je to dobrý „architekt“), Google Gemini 3.1 Pro CLI (velké kontextové okno 1 mil tokenů, štědrý bezplatný tarif – až 1000 požadavků denně)...a mnohé další

■ AI v Excelu

- Různá řešení
 - =COPILOT funkce
 - Agent Mode in Excel Labs add-in
 - ChatGPT for Excel add-in
 - Claude in Excel
 - generování a úpravy excelového souboru v chatbotu
 - ...

■ Analýza v chatbotu pomocí Pythonu na pozadí

■ "Text-to-SQL" nástroje (Chat s databází)

■ NotebookLM pro tzv. source-grounding (anti hallucination focus), multimodalita - YouTube videa jako zdroj, snadné používání

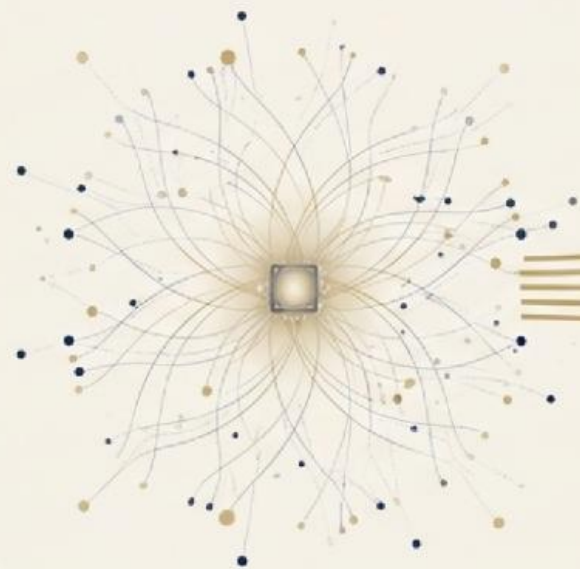
■ Asistenti (notetakers) na meetingy (s relativně štědrými free plans) **tl;dv** (nejen živé hovory, ale i nahrávky, schopnější AI shrnutí) a **Fathom** (levnější týmové plány, chatbot nad přepisy, unlimited storage ve verzi zdarma), oba umí obstojně česky, oba fungují na hlavních platformách Zoom, Google Meet, Microsoft Teams, oba umí tzv. clips & reels = výstřižky nahrávek klíčových momentů

■ Rychlý videoediting (ideální na občasné videoprezentace, rozhovory) v **Descript** – stříh audia a videa s velmi intuitivním prostředím, silný v text-based editování, podporuje češtinu, automaticky detekuje a smaže výplňová slova („eeree“), klonování hlasu = umožňuje „přepsat“ slova, která byla řečena špatně nebo nesrozumitelně

VYBRANÁ SPECIFIKA LLM (A NÁSTROJŮ AI)

Není to vyhledávač, je to generátor

LLM nerozumí pravdě. Jeho jediným cílem je předpovědět nejpravděpodobnější následující slovo na základě miliard přečtených textů.



Knihovna vzorců

Učí se strukturu jazyka, ne fakta.



Papouškování

Opakuje naučené fráze bez pochopení významu.



Hod kostkou

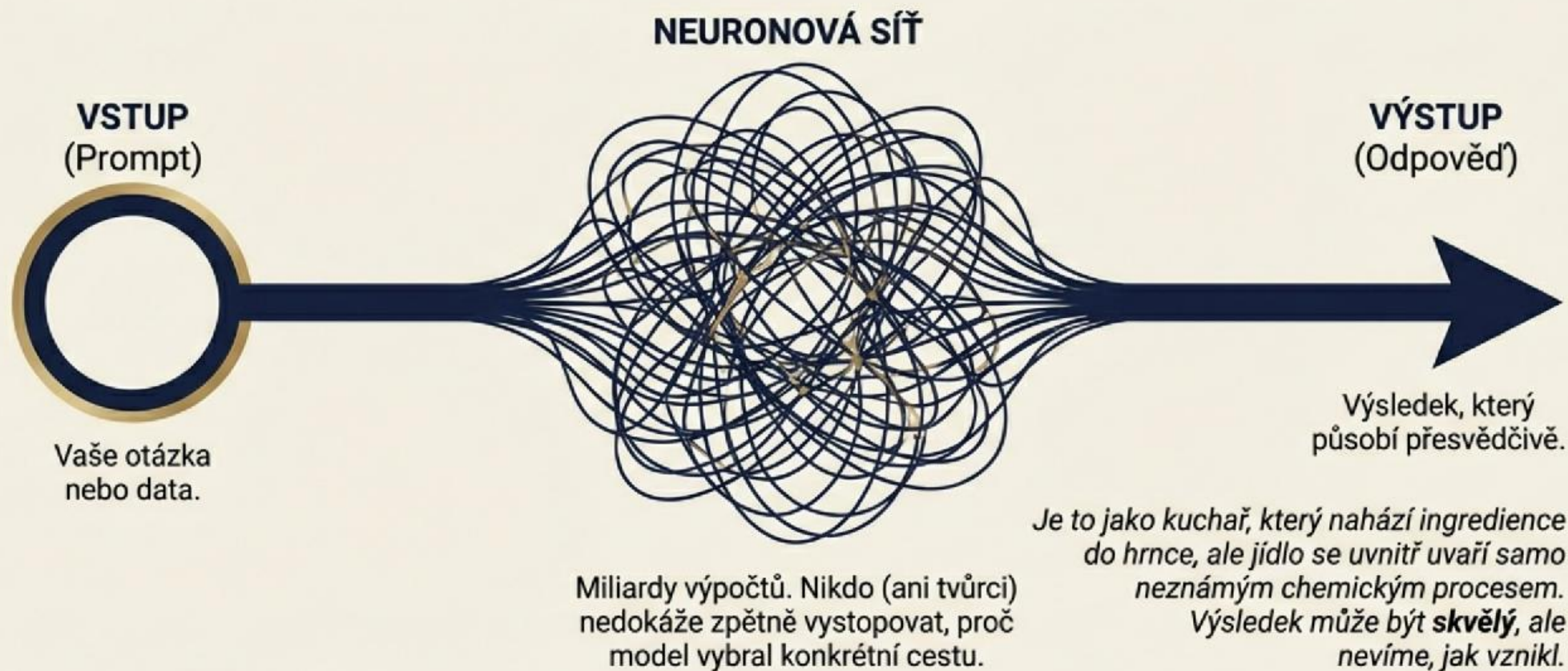
Každé slovo je vybráno na základě pravděpodobnosti, nikoliv jistoty.

Model nemá přístup k databázi faktů. Vše, co generuje, je 'halucinace', která se často (ale ne vždy) trefí do reality.

VYBRANÁ SPECIFIKA LLM (A NÁSTROJŮ AI)

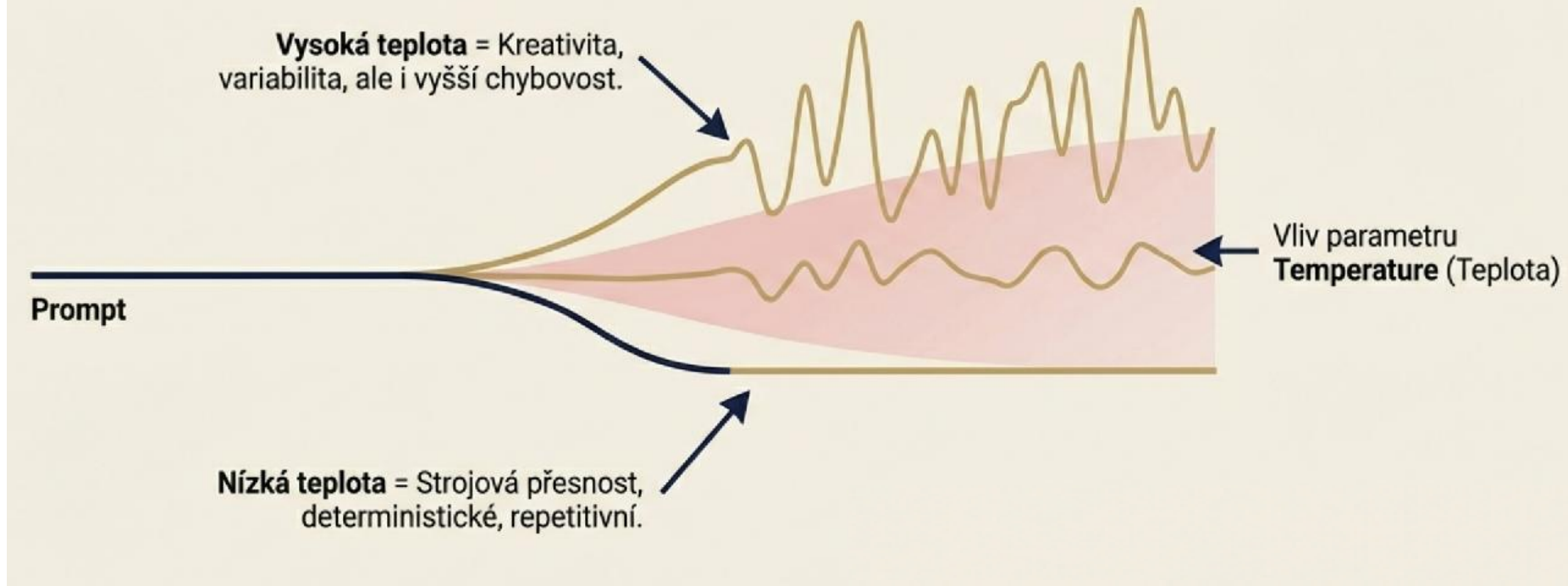
Problém 'Black-box': Nevidíme dovnitř

Víme, co do modelu vstupuje a co z něj vychází. Uprostřed je ale neprůhledná síť miliard parametrů.



VYBRANÁ SPECIFIKA LLM (A NÁSTROJŮ AI)

Stejná otázka ≠ Stejná odpověď. To je zásadní rozdíl oproti nástrojům, které znáte.



- Obrázek ilustruje princip, v praxi se **nedoporučuje** používat teplotu = 0, moderní modely jsou kalibrovány na teplotu = 1
- Přesnost ošetřujeme např. poskytnutí příkladů, vyžadováním strukturovaných výstupů (JSON), validací na výstupu, ...

VYBRANÁ SPECIFIKA LLM (A NÁSTROJŮ AI)



Silné stránky (S)

Sumarizace dlouhých textů, změna tónu komunikace, generování nápadů, formátování dat (JSON/HTML).



Slabé stránky (W)

Matematické výpočty (často chybuje), faktická přesnost, logická konzistence v dlouhých krocích.



Příležitosti (O)

Rychlé zpracování nestructurovaných dat, které dříve nešlo analyzovat (e-mail, recenze).



Hrozby (T)

Rozhodování managementu na základě 'halucinovaných' dat, únik citlivých informací do modelu.

VYBRANÁ SPECIFIKA LLM (A NÁSTROJŮ AI)

- **bias** – zkreslení/předsudek/tendence/sklon
- LLM neví, co je fakt. Jen předpovídá na základě tréninkových dat.
- 100 % modelů odráží předsudky ze svých trénovacích dat.
- Většina dat pochází z anglicky mluvícího západního světa. Atp.
- Např. při shrnutí textu model často „zprůměruje“ názory, čímž odstraní unikátní a menšinové názory, které ale mohou být důležité.
- LLMs berou to, co je nejpravděpodobnější, takže mají ve svých výstupech **tendenci stereotypy ještě posilovat**, protože stereotypní charakteristiky jsou v trénovacích datech obsaženy častěji, než ty, které stereotypům neodpovídají.

VYBRANÁ SPECIFIKA LLM (A NÁSTROJŮ AI)

4 tváře zkreslení: Kde všude se bias skrývá?

Nerozumíme-li typům chyb, nemůžeme je v datech odhalit.



Stereotypy

Přisuzování vlastností na základě pohlaví či rasy.
Příklad: "Lékař je on, zdravotní sestra je ona."



Kulturní slepota

Aplikace západních norem na celý svět.
Příklad: AI ignoruje lokální svátky v ČR a vnucuje americké.



Konfirmační zkreslení

Snaha vyhovět uživateli za každou cenu.
Příklad: AI lže, aby souhlasila s návodnou otázkou.



Halucinace

Vymyšlení faktů tam, kde chybí data.
Příklad: Generování neexistujících citací pro zaplnění mezer.

VYBRANÁ SPECIFIKA LLM (A NÁSTROJŮ AI)

Zkreslení není jednorázová událost, ale proces, který se nabaluje v každém kroku.

Sběr dat (Scraping)



Internet obsahuje nenávistné projevy a historické křivdy. Model to nasaje vše.

Filtrování (Labelling)



Lidé, kteří data třídí, mají své vlastní nevědomé předsudky.

Trénink (Training)



Model "zmrazí" stav světa v určitém čase (např. před rokem 2023).

Prompting (Uživatel)



Způsob, jakým se zeptáte, ovlivní odpověď (tzv. Leading questions).

- Největší chybou by bylo přebírat výstupy z LLM nekriticky!

VYBRANÁ SPECIFIKA LLM (A NÁSTROJŮ AI)

- Jak se bránit?
- biasy nelze zcela odstranit, ale lze je minimalizovat



VYBRANÁ SPECIFIKA LLM (A NÁSTROJŮ AI)

- e-maily
- záznamy rozhovorů
- články na firemních webech
- výroční zprávy
- videozáznamy z konferencí
- zápisy z porad
- smlouvy v PDF
- právní předpisy
- metodiky

**NESTRUKTUROVANÁ
DATA**

VYBRANÁ SPECIFIKA LLM (A NÁSTROJŮ AI)

Tři hlavní disciplíny strukturování



Kategorizace

Třídění dokumentů do předem definovaných šuplíků.

Je tento e-mail stížnost, pochvala, nebo objednávka?



Extrakce

Vyzobávání konkrétních entit a údajů z textu.

Najdi ve smlouvě IČO, částku a datum podpisu.



Anotace a Shrnutí

Tvorba metadat a stručných výtahů.

Vytvoř jednu větu, která shrnuje tento desetistránkový report.

VYBRANÁ SPECIFIKA LLM (A NÁSTROJŮ AI)

1 Identifikace



Najděte proces, kde lidé tráví čas čtením a přepisováním. Začněte malým pilotním projektem.

2 Volba nástroje



Pro veřejná data stačí ChatGPT/Claude. Pro citlivá firemní data volte zabezpečené API nebo lokální modely.

3 Prompt Engineering }

Naučte se psát instrukce. Kvalita výstupu závisí na kvalitě zadání. Buďte specifictí.

4 Validace



Nikdy nevěřte AI slepě. Vždy mějte člověka, který provádí namátkovou kontrolu výstupů (Human-in-the-loop).

VYBRANÁ SPECIFIKA LLM (A NÁSTROJŮ AI)

Halucinace jsou nevyhnutelné (zatím)

- Existuje několik studií, které prokazují, že je matematicky nemožné zcela eliminovat halucinace v LLM
- Podle [nové studie](#) leží problém možná v over-compliance při pre-trainingu
- Halucinace přímo korelují s neschopností modelu odlišit pravdivé od nepravdivých tvrzení
- Donedávna naprostá většina benchmarků penalizovala odpověď „nevím“ stejně jako špatnou odpověď
- Modely optimalizované na reasoning halucinují více
- Studie naznačují, že míra halucinací je doménově rozdílná

Vectara Hallucination Leaderboard (2025)

Top modely (sumarizace):

Gemini-2.5-flash-lite: 3,3 %

**Thinking modely (Claude Sonnet 4.5, GPT-5, Grok-4, Deepseek-R1):
všechny >10 %**

VYBRANÁ SPECIFIKA LLM (A NÁSTROJŮ AI)

Shrnutí limitů

- Kontext okno je omezené — u velkých projektů ztrácí přehled, je vždy užitečné projekty rozfázovat, dokumentovat apod.
- Rychlost odezvy
- limity použití zdarma / v placeném plánu / za den / týden / měsíc
- Soukromá data / firemní data / osobní údaje – chatboti se z poskytnutých dat často učí – když si nejste jisti, přečtěte si licenční ujednání (alternativou jsou lokální LLM)
- Může generovat plausibilní, ale chybná data
- Nenahrazuje odbornost, často nerozumí doménové expertíze, kontext musí dodat člověk
- Kvalita výstupu závisí (také) na kvalitě instrukcí (příliš obecný prompt = příliš obecný výsledek)
- Neověřují správnost ani kvalitu dat, se kterými pracují
- Vizualizace vyžadují zpětnou vazbu (nástroj nevidí výsledek na monitoru)
- Někdy opakuje stejnou chybu - obvyklá pomoc je dodat nový kontext

CO SE STANE S DATY V CHATU

	ChatGPT	Claude	Gemini	Co to znamená
Výchozí: data k trénování	ANO (opt-out)	ANO od 09/2025 (opt-out)	ANO (opt-out)	<i>Opt-out = ve výchozím stavu se vaše konverzace používají k trénování. Musíte sami aktivně odmítnout v nastavení.</i>
Jak odmítnout	Settings > Data Controls	Settings > Privacy	Gemini Apps Activity (3 různá místa!)	<i>U Gemini je třeba vypnout na 3 místech v různých menu – snadno se něco přehlédne.</i>
Retence dat (trénování ON)	Neurčeno přesně; de-identifikované	Až 5 let de-identifikované	Až 3 roky (odpojené od účtu)	<i>De-identifikované = odstraněn odkaz na váš účet, ale obsah konverzace zůstává v tréninkovém datasetu.</i>
Retence dat (trénování OFF)	30 dní (Temporary Chat)	30 dní	72h (historie off) nebo 18 měsíců	<i>I po odmítnutí se data krátkodobě uchovávají pro bezpečnostní monitoring (abuse detection).</i>
Enterprise výjimka	Enterprise, Business, Edu, API	Work, API, Gov, Edu	Workspace, Cloud	<i>Firemní/vládní tarify mají odlišné podmínky – data se ve výchozím nastavení NETRĚNUJÍ. Ale osobní Pro/Plus = stejně jako free!</i>
Lidské review konverzací	Možné (pro safety)	Možné (safety flags)	Explicitně potvrzeno!	<i>Google říká: "Nezadávejte nic, co byste nechtěli, aby viděl lidský recenzent." U OpenAI a Anthropic jen při safety incidentech.</i>
Privacy-safe režim	Temporary Chat	Incognito	Historie off	<i>Režim, kdy se konverzace neukládá a nepoužívá k trénování.</i>

Podmínky se neustále mění. Placený účet nemá odlišnou trénovací politiku od verze zdarma. Smazání konverzace neodstraní data z již provedeného trénování.

SHRUNTÍ: JAK BÝT AI-SAVVY

Být AI-savvy znamená rozumět limitům technologie stejně dobře jako jejím možnostem.

- Využívat AI k řešení **konkrétních, jasně definovaných** problémů v organizaci, firmě, týmu.
- **Integrovat** AI do stávajících (datových) analytických struktur (nevynalézat kolo). Využití k posílení schopností.
- **Transparentně komunikovat** využití AI vůči všem partnerům a nasazovat ji nejprve v menším měřítku, aby bylo možné technologii lépe pochopit a smysluplně využít.
- **Cíleně** rozvíjet dovednosti potřebné pro éru AI. Každý profesionál by měl strávit několik hodin experimentováním s konkrétním nástrojem nebo systémem v reálném kontextu své práce, aby pochopil jeho specifické chyby a možnosti.
- Využívat AI pro **konzultaci o optimálním nastavení postupů**, ale postupy validovat.

savvy (/ˈsævi/) je neformální přídavné jméno; důvtipný, chytrý, znalý věci; z francouzského “savez-vous)“ nebo španělského „sabe (usted)“

TIPY NA ČTENÍ A STUDIUM

- **Studie MIT, která ukazuje, jak personalizace a „profil uživatele“ může systematicky zkreslovat kvalitu výstupů** a jak takové nerovnosti testovat/auditovat v reálném nasazení.
<https://news.mit.edu/2026/study-ai-chatbots-provide-less-accurate-information-vulnerable-users-0219>
- **Jak je to s halucinacemi:**
 - Karpowicz, M. P. (2025) On the Fundamental Impossibility of Hallucination Control in Large Language Models <https://arxiv.org/abs/2506.06382v2>
 - **Vectara Hallucination Leaderboard:** <https://www.vectara.com/blog/introducing-the-next-generation-of-vectaras-hallucination-leaderboard>
- **Leaderboard a benchmark schopností modelů:**
 - Epoch's AI Benchmarking Hub www.epoch.ai
 - LMSYS Chatbot Arena www.lmarena.ai — crowdsourcovaný blind test LLM modelů, kterého se můžete zúčastnit
- Různé subreddity – věnované jednotlivým LLMs, nástrojům, konceptům, filozofiím...
- Mailinglisty – např. kdnuggets.com (The AI Report)
- Specifické **pro programovací jazyky** – např. r-bloggers.com (R), pro Python existuje řada komunit (v ČR www.python.cz) – nejznámější datová komunita kolem jazyka Python je PyData (můžete sledovat i na YouTube @PyDataTV) působí i v Česku, pořádají řadu zajímavých MeetUps
- Weby **předních výzkumných univerzit v oboru AI**, např. The Stanford Institute for Human-Centered AI <https://hai.stanford.edu/> - zajímavé studie a každoroční podrobný přehled AI Index Report (pohled z technologického, tak ekonomického hlediska, obvykle vychází koncem jara)
- **Weby tvůrců LLM** (dokumentace, research papers apod.) – např. Anthropic Research <https://www.anthropic.com/research> pro zajímavé studie, ale třeba i Claude Cookbook pro návody a pochopení nástrojů Anthropic skrze návody i od samotných uživatelů <https://platform.claude.com/cookbook/> nebo Anthropic courses <https://anthropic.skilljar.com/> oficiální kurzy a tutoriály jejich vlastních nástrojů
- huggingface.co/blog – open source modely, datasets, trendy v komunitě, včetně studií
- Poznámka k obsahu na YouTube – U video obsahu na sociálních sítích buďte obezřetní, tvůrci jsou motivováni engagement metrikami, ne přesností nebo užitečností informací. Upřednostňujte zdroje, kde autoři budují vlastní odbornou reputaci, kterou si chrání.

A co vaše tipy?



Sdílejte tipy. Ptejte se.

slido.com

kód # 1820 4031





Konec 1. části

Kristýna Meislová
Adéla Kučerová

meislova@tc.cz
kucerova@tc.cz

Technologické centrum Praha
www.tc.cz | www.horizonary.cz