

Ing. Jan Kaiser

**3. interní klinika 1. LF UK a
VFN**



1. LÉKAŘSKÁ
FAKULTA
Univerzita Karlova

AI V AKADEMICKÉ A PRAKTICKÉ SFÉŘE

Průvodce pro výzkum a práci s
daty (bez kódování)

CO SE DNES NAUČÍME?

1. **Demystifikace AI:** Jak to funguje a proč to "kecá"?
2. **Nástroje:** ChatGPT, Claude, Gemini, Perplexity a další.
3. **Etika:** Halucinace, GDPR a akademická integrita.
4. **Praktické využití:**
 - AI jako čtenář a editor.
 - AI jako statistický konzultant.
5. **Agentická AI:** Váš digitální stážista.
6. **Prompting:** Jak se správně ptát (Dos & Don'ts).

1. CO JE TO AI?

(BEZ SLOŽITÝCH POJMŮ A MATEMATIKY)

CO JE VLASTNĚ "GENERATIVNÍ AI"?

Představte si **automatické dokončování textu** (jako v mobilu), které přečetlo celý internet.

- **Není to mozek:** Nemyslí, necítí, nerozumí světu.
- **Je to pravděpodobnostní stroj:** Hádá, jaké slovo bude následovat.
- **Analogie:** Když řeknu "Kdo jinému jámu kopá...", většina z vás doplní "...sám doní padá."
 - AI to nedoplňí proto, že chápe morální poučení, ale protože tato slova jdou statisticky nejčastěji za sebou.

JAK AI "ČTE"? (TOKENIZACE)

Počítače neumí číst písmena, umí jen počítat s čísly. Text se proto musí nejprve rozbit na kousky = **Tokeny**.

- **Token** \approx 4 znaky (v angličtině), nebo 1 slovo.
- V češtině je to složitější (kvůli skloňování).

Příklad: "Pes štěká" → [Pes, ště, ká] → [1052, 4821, 992]

Každý kousek textu dostane své unikátní číslo (ID).

VEKTORY (EMBEDDINGS)

Kde bydlí význam?

Čísla (ID) jsou jen nálepky. Aby AI chápala *význam*, převede si každé číslo na **Vektor**.

- **Vektor** = Dlouhá řada čísel (souřadnice v mračnu významů).
- Představte si obrovskou knihovnu, kde každá kniha (slovo) má své přesné místo.

Kouzlo: Slova s podobným významem "bydlí" blízko sebe.

- `Pes` a `Kočka` jsou sousedé.
- `Pes` a `Mikrovlnná trouba` jsou daleko od sebe.

MATEMATIKA JAZYKA

"Královská rovnice"

Díky vektorům může AI s jazykem **počítat**. Slavný příklad vektorové aritmetiky:

$$K(\text{Král}) - K(\text{Muž}) + K(\text{Žena}) \approx K(\text{Královna})$$

Když od vektoru "Král" odečtete "Mužství" a přičtete "Ženství", dostanete souřadnice velmi blízko vektoru "Královna".

Co je to trénink? Trénink sítě je hledání těch správných souřadnic pro miliardy slov, aby tyto vztahy fungovaly.

KLÍČOVÉ POJMY: KONTEXTOVÉ OKNO (PAMĚТЬ)

Ted' už víme, co je token. **Kontextové okno** je počet tokenů, které AI udrží v "krátkodobé paměti" (jako RAM v počítači).

Představte si **školní tabuli**.

- **Kontextové okno** je velikost této tabule.
- Můžete na ni napsat instrukce, vložit článek nebo data.
- **Malé okno:** Vejde se jen jeden odstavec.
- **Velké okno (dnešní standard):** Vejde se celá kniha (nebo několik PDF).
- **Gemini 1.5 Pro:** Má okno až 2 miliony tokenů (celá knihovna).

Prakticky: Pokud chcete, aby AI pracovala s vaším textem, musí se vejít na tabuli.

2. SOUČASNÁ KRAJINA NÁSTROJŮ

CO POUŽÍT A KDY?

VELKÁ TROJKA (A TI DALŠÍ SE DOTAHUJÍ)

Nástroj	Výrobce	V čem vyniká (Superpower)	Cena
ChatGPT	OpenAI	Univerzální "švýcarák". Skvělý na všechno, má nejvíce funkcí (hlas, vizuál, data).	Free / \$20
Claude	Anthropic	Nejlidštější text a logika. Méně "robotický", skvělý na psaní a analýzu dlouhých textů.	Free / \$20
Gemini	Google	Ekosystém. Napojení na Google Docs/Drive/Gmail. Obří paměť (kontext).	Free / \$20
DeepSeek	DeepSeek	Levný génius. Čínský model, který v roce 2025/26 šokoval svět výkonem zdarma.	Free

MÁM PLATIT 500 KČ MĚSÍČNĚ? (PRO VS. FREE)

Modely zdarma jsou skvělé, ale "ořezané". Co získáte za předplatné (~\$20/měs)?

1. Chytrý vs. Rychlý model:

- *Free*: Rychlý, ale občas "hloupější" (GPT-4o mini, Flash).
- *Paid*: Nejchytrější dostupné modely (GPT-4o/5, Claude 3.5/3.7 Opus, Gemini 1.5 Pro).

2. Kapacita a Limity:

- *Free*: Po pár otázkách vás "vykopne" nebo zpomalí.
- *Paid*: Výrazně vyšší limity (až 5x více zpráv).

3. Soukromí:

- Placené verze často garantují, že vaše data *nepoužijí* pro trénink (nutné zkontolovat v nastavení!).

4. Pokročilé funkce:

- Generování obrázků na vysoké úrovni (DALL-E 3), nahrávání souborů, analýza dat, tvorba vlastních GPTs/Projectů.

SPECIALIZOVANÉ NÁSTROJE PRO VÝZKUM

Tyto nástroje jsou pro studenty často užitečnější než obecný chat:

1. Perplexity AI

- **Co to je:** Vyhledávač na steroidech.
- **Jak funguje:** Prohledá internet a odpoví na otázku s **citacemi**. Ideální na rychlé ověření faktů.

2. NotebookLM (od Google)

- **Co to je:** Váš osobní znalostní asistent.
- **Jak funguje:** Nahrajete svá PDF (skripta, články) a bavíte se **pouze** nad nimi. Halucinuje minimálně.

3. Scite.ai & Elicit (Pro pokročilé)

- **Scite:** Ukazuje, *jak* byl článek citován (podpořen / vyvrácen?). Skvělé pro kritickou rešerši.
- **Elicit/Consensus:** Pomáhá najít "vědecký konsenzus" na danou otázku prohledáním milionů paperů.

3. ETIKA, BEZPEČNOST A "LŽI"

NA CO SI DÁT POZOR

HALUCINACE: KDYŽ AI SNÍ

AI nelže záměrně. Ona **halucinuje**.

- **Co to je?** AI si vymyslí fakt, citaci nebo číslo, které vypadá naprosto věrohodně.
- **Proč se to děje?**
 - AI nerozumí pravdě. Byla trénována, aby text *vypadal* lidsky a pravděpodobně.
 - Pokud se zeptáte na něco, co v datech chybí, AI "dopočítá" nejpravděpodobnější slova tak, aby vás potěšila.
- **Příklad:** AI vymyslí citaci neexistujícího článku od skutečného autora, protože "autor" a "téma" se v datech často vyskytují spolu.

Obrana: Vždy ověřujte fakta. AI je skvělý sluha pro formu, ale špatný pán pro fakta.

ZKRESLENÍ (BIAS) A STEREOTYPY

AI je zrcadlem naší společnosti – i s jejími chybami.

- **Co to je?** AI může (nechtěně) reprodukovat rasové, genderové nebo kulturní předsudky.
- **Proč se to děje?**
 - AI se učila z celého internetu (Reddit, Wikipedia, diskuze).
 - Pokud je na internetu slovo "lékař" v 80 % případů spojeno s mužem a "sestra" s ženou, AI tento vzorec převezme jako "pravidlo".
- **Důsledek:** Při generování obrázků nebo příběhů může AI ignorovat menšiny nebo posilovat stereotypy.
- **Rada:** Buďte si toho vědomi a specificky žádejte o diverzitu, pokud je to potřeba.

AKADEMICKÁ INTEGRITA

AI je nástroj, ne autor.

- **Proč AI nemůže být autorem?**
 - Nemá právní subjektivitu, zodpovědnost ani "úmysl".
 - Jen remixuje a statisticky průměruje lidské texty. Nemůže zaručit originalitu.
- **Plagiátorství vs. AI:**
 - Použití AI bez přiznání je podvod (vydáváte práci stroje za svou).
 - Přiznané použití (např. pro korekturu) je u mnoha škol akceptované.
- **Správný přístup:** "Autorem myšlenky jsem já, AI mi pomohla ji formulovat." (Vždy citujte použití AI v metodice!)

SOUKROMÍ: VAŠE DATA JSOU PLATIDLO

Zlaté pravidlo: Co napíšete do chatu, může být použito pro trénink modelu.

- **Proč je to problém?**

- Firmy používají vaše konverzace k "doučování" modelu (RLHF - Reinforcement Learning from Human Feedback).
- **Riziko:** Pokud nahrajete do chatu tajná data (např. rodná čísla pacientů), model se na nich může "naučit" vzory a teoreticky je (v extrémním případě) vyzradit jinému uživateli.

- **Bezpečný postup:**

- Anonymizujte data ("Pacient A" místo "Jan Novák").
- Vypněte v nastavení "Trénování na datech" (ChatGPT to umožňuje).

4. PRAKTIČKÉ WORKFLOW: ČTENÍ A PSANÍ

AI JAKO VÁŠ EDITOR

AI JAKO "READING ASSISTANT"

Máte před sebou 20 stránek složitého odborného textu?

1. **Sumarizace:** "Vysvětli mi hlavní myšlenku tohoto abstraktu jako laikovi."

- **Proč to funguje:** AI umí identifikovat klíčová "sémantická centra" (nejdůležitější pojmy) a ignorovat omáčku.

2. **Hledání mezer:** "Přečti si úvod. Jaké otázky autoři kladou? Jsou tam logické díry?"

- **Proč to funguje:** AI porovnává text s miliony jiných úvodů a hledá anomálie.

3. **Zjednodušení:** "Tento odstavec je složitý. Přepiš ho pro studenta prváku."

- **Proč to funguje:** AI viděla miliardy párů "odborný text" <-> "populární článek" a umí tento převod napodobit.

Tip: Použijte NotebookLM – halucinuje minimálně, protože se "dívá" jen do vašeho PDF.

DIALOG S TEXTEM (ITERACE)

Google je jednorázový: Otázka → Odpověď → Konec. AI je konverzace: Otázka → Odpověď → **Upřesnění**.

- **První pokus není finální:**
 - *"To je moc dlouhé, zkrať to na polovinu."*
 - *"Ten druhý bod nedává smysl, vysvětli ho jinak."*
 - *"Přidej tam příklad z české praxe."*
- **Proč je to důležité?** AI nemá telepatické schopnosti. První odpověď je "průměrná". Skvělou odpověď získáte až doplňujícími pokyny (= iterací).

⚠️ POZOR: ZNEČIŠTĚNÍ KONTEXTU (CONTEXT POLLUTION)

Pamatujete na "školní tabuli" (kontextové okno)?

- Pokud na ni napíšete příliš mnoho pokynů a oprav, AI se začne "ztrájet" a míchat staré instrukce s novými.
- **Rada:** Když AI začne "bláznit", smažte tabuli → **Začněte nový chat.**

*Potřebujete, aby si AI pamatovala složitý postup a neztrácela se? K tomu slouží **Agenti** (viz sekce 6).*

AI JAKO EDITOR (NE SPISOVATEL)

Nechtějte po AI, aby psala za vás. Text pak bývá "bez duše" a plný klišé ("V dnešní dynamické době...").

Lepší využití:

- **Brainstorming:** "Navrhni 5 názvů pro práci o AI v medicíně."
 - **Proč:** Halucinace jsou zde výhodou – AI "vymýslí" kreativní spojení, která by vás nenapadla.
- **Osnova:** "Vytvoř strukturu kapitoly o diabetu 2. typu."
 - **Proč:** AI zná "tvar" tisíců správně napsaných akademických prací.
- **Korektura:** "Najdi v mému textu stylistické neobratnosti a vysvětli, proč jsou špatně."
- **Překlad:** "Přelož do angličtiny (akademický styl)."

5. PRAKTICKÉ WORKFLOW: DATA SCIENCE BEZ KÓDOVÁNÍ

"NEJSEM PROGRAMÁTOR, ALE POTŘEBUJI DATA"

KÓD? KDEPAK, KONVERZACE.

Zapomeňte na "programování" a děsivé černé obrazovky. Dnes stačí **data a dobrá otázka**.

- **AI jako váš analytik:** Vy zadáváte úkoly (česky), AI je vykonává (píše kód za vás).
- **Cíl:** Získat odpovědi z dat, ne se učit syntaxi Pythonu.
- **Výhoda:** Reprodukovatelnost. AI přesně zapíše, jak data vyčistila a jaký graf vytvořila.

AI JAKO STATISTICKÝ KONZULTANT

Nemusíte umět kódovat, abyste udělali analýzu. Stačí umět popsát problém.

- **Proč to funguje:** AI "přečetla" všechny učebnice statistiky. Zná přesné definice, předpoklady (normalita dat, homoskedasticita) a ví, kdy který test použít.
- **Role:** AI je váš "Senior Statistik", vy jste "Vedoucí výzkumu", který zadává cíle.

Scénář 1: Výběr testu

- *Prompt:* "Mám data o krevním tlaku (spojitá veličina) u dvou skupin pacientů (léčení vs. neléčení). Data mají normální rozdělení. Jaký test mám použít a proč?"
- *Odpověď AI:* "Použijte **Nepárový T-test**, protože porovnáváte průměry dvou nezávislých skupin se spojitými daty..."

NÁSTROJE PRO "DATA SCIENCE" BEZ KÓDOVÁNÍ

Zapomeňte na instalaci Pythonu. Tyto nástroje běží v prohlížeči:

1. Julius AI:

- Specialista na data. Nahrajete Excel → "Vytvoř mi graf..." → Hotovo.
- Umí pokročilé vizualizace i čištění dat.

2. ChatGPT (Data Analyst):

- Integrovaný "Python v chatu".
- Nahrajete soubor a řeknete: "*Zjisti korelace mezi věkem a platem.*" (On si na pozadí napíše skript a spustí ho).

3. Excel Copilot:

- AI přímo v Excelu. "Zvýrazni řádky, kde je hodnota > 100 a seřaď je podle data."

AI A EXCEL / SPSS

Zapomněli jste vzorec v Excelu? Nebo nevíte, kam kliknout v SPSS?

Scénář 2: Excel Helper

- *Prompt:* "Mám ve sloupci A data narození (DD.MM.RRRR). Napiš mi vzorec, který vypočítá věk k dnešnímu dni."
- *Odpověď:* =DATEDIF(A1, TODAY(), "y")

Scénář 3: Interpretace výstupu (SPSS/R)

- *Prompt:* (Vložte fotku/tabulku výsledků) "Tady je výstup logistické regrese. Vysvětli mi jako laikovi, co znamená Odds Ratio 2.5 u proměnné 'Kouření'."
- *Odpověď:* "Znamená to, že kuřáci mají 2,5x vyšší šanci na výskyt onemocnění než nekuřáci..."

6. AGENTICKÁ AI

OD "KECÁNÍ" K "PRÁCI"

CO JE TO "AGENT"? (ŘEŠENÍ ČISTÉ TABULE)

Pamatujete na problém "znečištěné tabule"? **Agenti** ho řeší.

- **Definice:** Agent = LLM + Nástroje + **Plánování**.
- **Jak to funguje:**
 - Hlavní agent ("Manažer") rozbije úkol na podúkoly.
 - Na každý podúkol si zavolá "čerstvého" podřízeného agenta s **čistou hlavou (tabulí)**.
 - Výsledek: Nikdo se neztratí, protože každý řeší jen svůj malý kousek.

EXTERNÍ PAMĚŤ: JAK OBEJÍT "ZAPLNĚNOU TABULI"

Kontextové okno se jednou zaplní. Co s dlouhodobým úkolem (např. psaní knihy)? **Řešení:** Naučte AI psát si "poznámky" bokem.

- **Co je to Markdown (.md)?**

- Obyčejný text ("Poznámkový blok") s "cukrátkem".
- Nadpis se dělá křížkem (`# Nadpis`), tučné písmo hvězdičkami (`**Text**`). Čitelné pro lidi i stroje.

- **Workflow "Nekonečné paměti":**

1. **Chat 1:** "Přečti tento článek a vytáhni klíčové teze do souboru `pamet.md` ." (Stáhněte soubor).
2. **Nový Chat (Čistá tabule):** Nahrajete `pamet.md` .
3. **Prompt:** "Přečti si `pamet.md` jako kontext a napiš úvod."

Takto přenesete jen to důležité a "zapomenete" balast.

NÁSTROJE PRO ORCHESTRACI (KDE SI POSTAVIT AGENTA?)

Nemusíte být programátor, abyste si postavili tým agentů.

1. Custom GPTs (OpenAI):

- Nejjednodušší "agent".
- Nahrajete mu svá PDF a dáte instrukce ("Jsi expert na české právo").

2. Zapier Central / Make.com:

- Propojení AI s vaším e-mailem, kalendářem a Excelem.
- *Příklad:* "Když přijde faktura na e-mail → přečti částku → ulož do Excelu."

3. n8n (Pro odvážné):

- Vizuální "stavebnice" (bod po bodu). Umožňuje stavět složité firemní workflow zdarma (open-source).

PŘÍKLAD AGENTICKÉ PRÁCE

Zadání: "Zjisti aktuální incidenci chřipky v ČR a vytvoř report."

Co udělá Chatbot: Ztratí se v datech nebo si vymyslí číslo. **Co udělá Agent (Workflow):**

1. **Manažer:** Rozdělí úkol.
2. **Agent "Researcher":** Jde na web SZÚ, stáhne PDF. (Přečte, vytáhne číslo, zavře okno).
3. **Agent "Analyst":** Dostane *jen* to číslo. Vypočítá trend.
4. **Agent "Writer":** Dostane *jen* výsledek a napíše e-mail.

Každý krok je izolovaný → nulové riziko halucinace z přetížení.

7. PROMPTING: JAK SE PTÁT

UMĚNÍ KOMUNIKACE SE STROJEM

RÁMEC C-R-E-A-D

Aby AI pochopila, co chcete, buděte specifičtí.

- **C - Context (Kontext):** "Jsem student medicíny, píšu diplomku..."
- **R - Role:** "Chovej se jako zkušený statistik / editor..."
- **E - Explicit (Jasnost):** "Nepoužívej odrážky, chci souvislý text."
- **A - Action (Akce):** "Vytvoř, analyzuj, porovnej..."
- **D - Data:** "Tady jsou data/text, se kterým pracuj..."

MÁM BÝT SLUŠNÝ? (EMOCE V PROMPTINGU)

Mýtus: "Na AI musím být hodný, jinak se vzbouří." Realita: AI nemá city, ale **rozumí naléhavosti**.

- **Vulgarita:** Špatně. Spouští bezpečnostní filtry ("I cannot answer that").
- **Emoční tlak:** Funguje.
 - *"Tohle je kritické pro mou diplomovou práci."* → Model se víc snaží (ověřeno výzkumem).
 - *"Musí to být naprosto přesné, jinak se stane katastrofa."* → Menší chybovost.
- **Poděkování:** Neurazí, nenadchne. Je to zbytečný token navíc, ale udržuje "konverzační flow".

POKROČILÉ TECHNIKY (CO OPRAVDU FUNGUJE)

Když jednoduchá otázka nestačí, vytáhněte těžký kalibr:

1. Chain of Thought (Myšlenkový řetězec):

- Magická věta: "*Let's think step by step*" (Pojďme přemýšlet krok za krokem).
- Nutí AI rozložit problém na kousky → Drasticky snižuje halucinace v logice.

2. Few-Shot Prompting (Ukaž příklad):

- Místo složitého popisu dejte 3 příklady toho, co chcete.
- *Příklad:* "Napiš tweet o AI. Styl:
 - Vstup: Pivo. Výstup: Orosená odměna po práci! 🍺 #pohoda
 - Vstup: Káva. Výstup: Palivo pro ranní start. ☕ #energie
 - Vstup: AI. Výstup: ..."

DOS & DON'TS

✓ ANO (Dos)

- Dávejte příklady ("Chci to přesně jako v tomto textu...").
- Rozdělte složitý úkol na menší kroky ("Chain of Thought").
- Když se AI splete, opravte ji ("Ne, myslí jsem to jinak").
- Ptejte se "Proč?" ("Proč jsi vybral tento test?").

✗ NE (Don'ts)

- Nepoužívejte vágní zadání ("Napiš něco o AI").
- Nevěřte slepě všemu (ověřujte!).
- Neposílejte AI celé zadání seminárky "vypracuj to za mě".
- Nepředpokládejte, že si AI pamatuje něco z minulého týdne (pokud to není v kontextu).

ZÁVĚR

AI je "Kolo pro mysl"

Nejezdí samo, musíte šlapat. Ale dostanete se dál a rychleji.

Klíč k úspěchu: Kritické myšlení + Schopnost klást správné otázky.

DÍKY ZA POZORNOST!

Ing. Jan Kaiser vyuka@jankaiser.cz