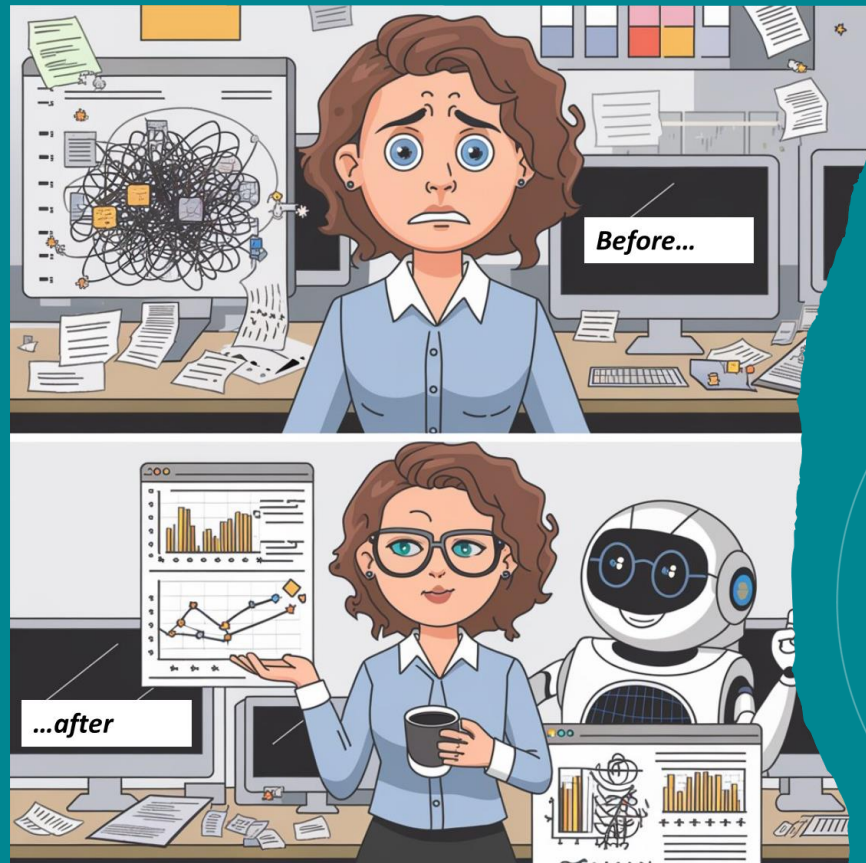


# ZPRACOVÁNÍ DAT S VYUŽITÍM VELKÝCH JAZYKOVÝCH MODELŮ

*Michaela Liegertová*

23.10.2024



Spolufinancováno  
Evropskou unií



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

MUNI  
ICS

Registrační číslo IPs EOSC-CZ

CZ.02.01.01/00/22\_004/0007682

a) Kontakt s jedovatými hady a ještěrkami:

Celkový počet úmrtí: 1

Rok výskytu: 2017

b) Kontakt s jedovatými pavouky:

Celkový počet úmrtí: 1

Rok výskytu: 2010

c) Kontakt se štíry (škorpióny):

Celkový počet úmrtí: 1

Rok výskytu: 2008

d) Kousnutí a úder krokodýlem nebo aligátorem:

Celkový počet úmrtí: 1

Rok výskytu: 2010

e) Kousnutí nebo rozdrčení jinými plazy:

Celkový počet úmrtí: 3

Roky výskytu: 2007, 2010, 2011

Stručný komentář:

Úmrtí způsobená kontaktem s nebezpečnými zvířaty jsou v ČR extrémně vzácná (7 případů za 11 let). Nejvíce úmrtí (3) způsobili "jiní plazi". Většina případů se kumuluje v letech 2010-2011, což může indikovat náhodnou kumulaci nebo změnu v reportování. Tato data zdůrazňují nízké riziko v ČR, ale také důležitost opatrnosti při kontaktu s potenciálně nebezpečnými zvířaty.

## Data ....k čemu je to dobré?

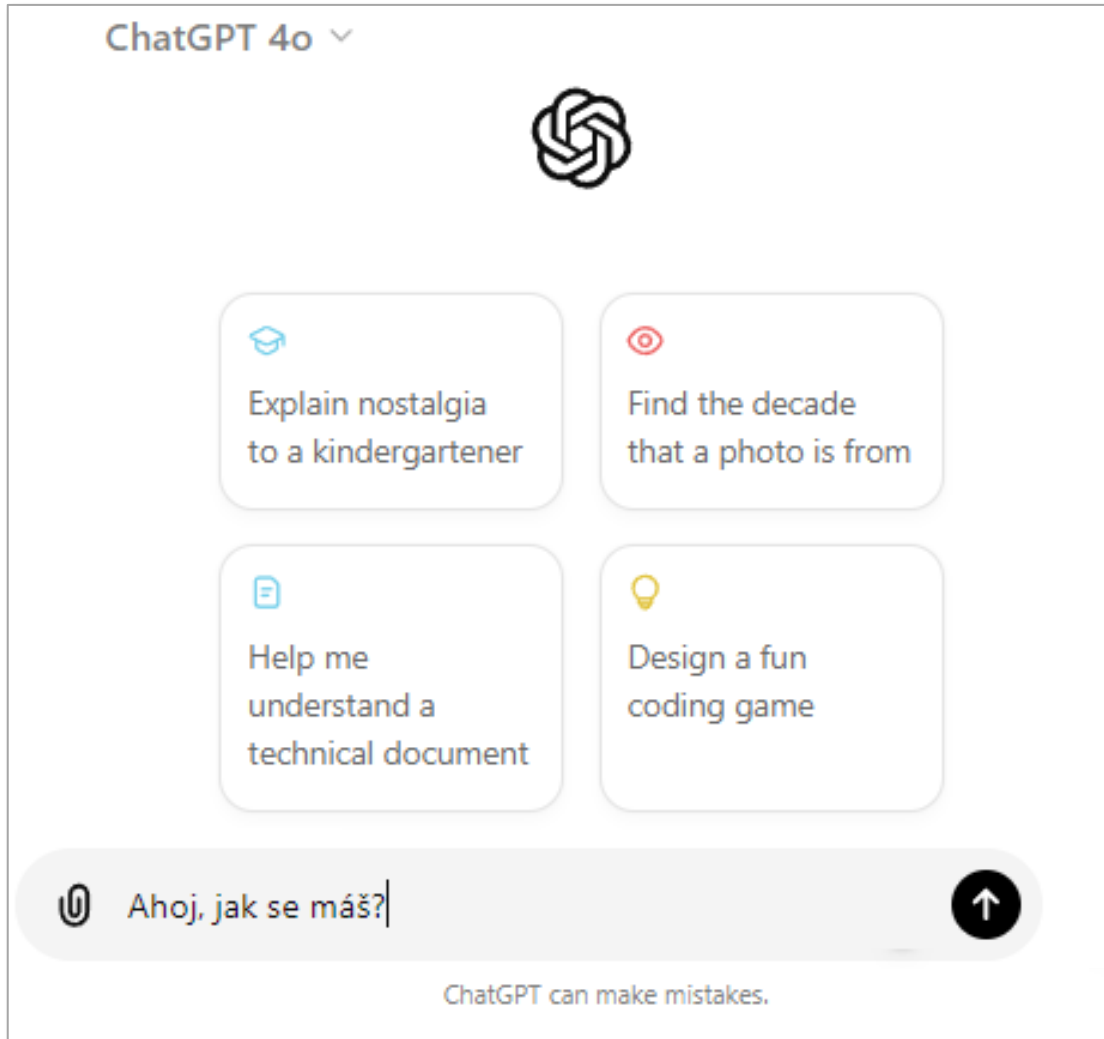
```
["idhod","hodnota","stapro_kod","pohlavi_cis","pohlavi_kod","ps_cis","ps_kod","ps0_cis","ps0_kod","vuzemi_cis","vuzemi_kod","rok","pohlavi_txt","ps_txt","ps0_txt","vuzemi_txt"
"1324891475","0","5393","102","1","5121","A49","5120","I","100","3042","2006","muž","Bakteriální infekce neurčené lokalizace","Některé infekční a parazitární nemoci (A00-B99)","Plzeňský kraj"
"1325553777","0","5393","102","1","5121","A49","5120","I","100","3026","2006","muž","Bakteriální infekce neurčené lokalizace","Některé infekční a parazitární nemoci (A00-B99)","Středočeský kraj"
"1324891485","0","5393","102","1","5121","A49","5120","I","100","3085","2006","muž","Bakteriální infekce neurčené lokalizace","Některé infekční a parazitární nemoci (A00-B99)","Královéhradecký kraj"
"1324559925","5","5393","102","1","5121","A49","5120","I","97","19","2006","muž","Bakteriální infekce neurčené lokalizace","Některé infekční a parazitární nemoci (A00-B99)","Česká republika"
"1324560909","1","5393","102","1","5121","A49","5120","I","100","3140","2006","muž","Bakteriální infekce neurčené lokalizace","Některé infekční a parazitární nemoci (A00-B99)","Moravskoslezský kraj"
99)"," indubický kraj"
99)"," hočeský kraj"
99)"," homonavský kraj"
99)"," tecký kraj"
99)"," berecký kraj"
```

## What do you want to analyze today?

 OD\_DEM03\_202409131...  
spreadsheet - 175.58 MB

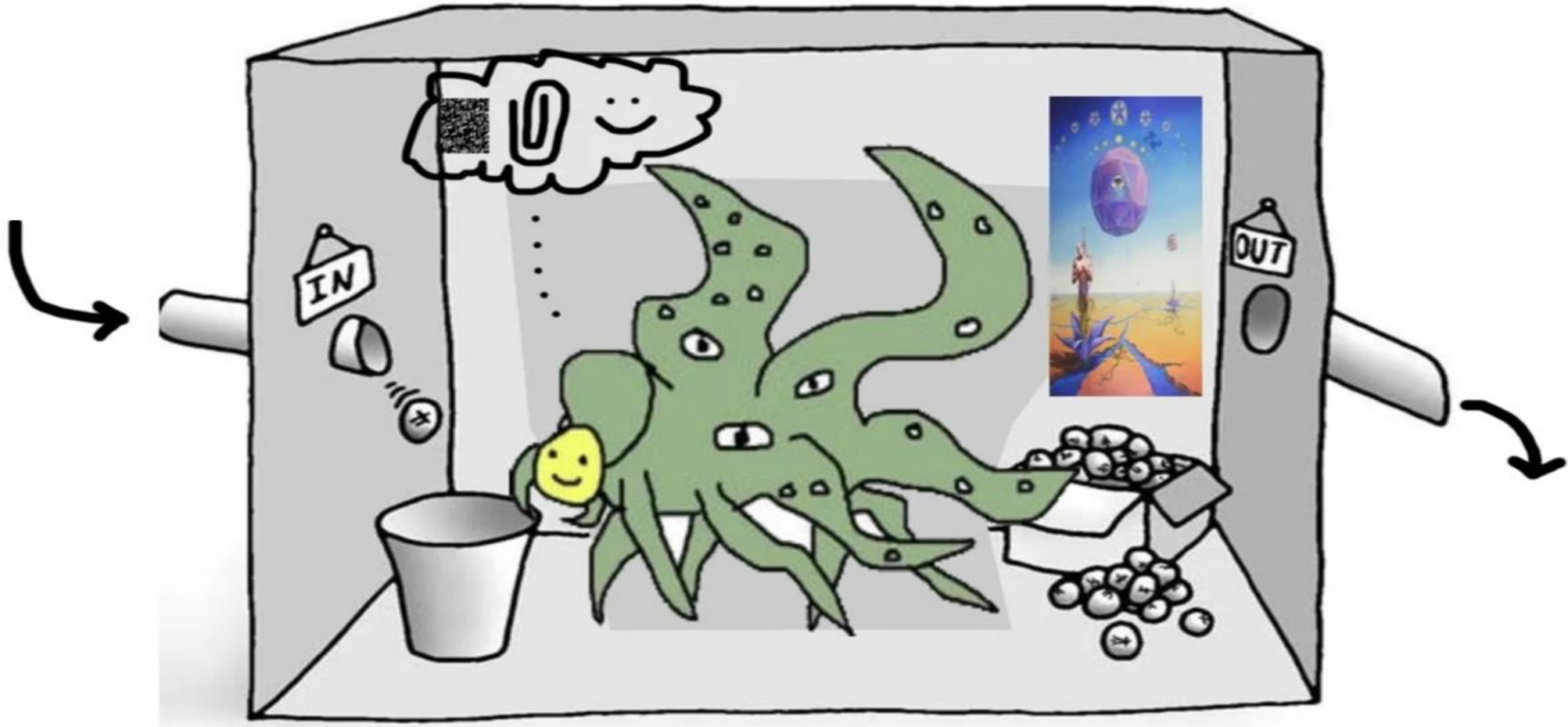
I uploaded the file: OD\_DEM03\_2024091312173660.CSV

Kolik lidí v ČR zemřelo po kontaktu s neobvyklými zvířaty (pavouci, škorpióni, krokodýli a podobně)?



Přemýšleli jste někdy nad tím, co se po stisknutí Enter v ChatGPT děje s vaším dotazem?

*Jak si někteří lidé představují velké jazykové modely...*





### Chatovací rozhraní

Aplikace zabalí zprávu + metadata a pošle internetem

Cesta internetem (velká vzdálenost)

### Datové centrum

Vyrovňovač zátěže (load balancer) směřuje váš dotaz na dostupný server, dochází k předzpracování (tokenizace)

### GPU klastr

GPU spustí komplexní matematickou funkci (LLM), která zpracovává příchozí tokeny. Využívá miliardy parametrů k pochopení kontextu vaší zprávy a generování odpovědi (proces probíhá token po tokenu)

### Post-processing

Filtrování nevhodného obsahu, formátování nebo dodatečné kontroly

Cesta internetem (velká vzdálenost)

### Chatovací rozhraní

Odpověď AI se objeví na vaší obrazovce, často se renderuje znak po znaku, aby napodobila psaní

Ahoj, jak se máš?

*Jak ve skutečnosti vypadají velké jazykové modely...*



Ahoj! Mám se dobře, díky za optání! Jak se máš ty?

@Liegertova\_Michaela

## Životní cyklus a aktualizace LLM



### Počáteční trénování

Začátek

Model je natrénován na rozsáhlém korpusu dat



### Období neměnnosti

Několik měsíců

Model zůstává neměnný, jeho parametry a váhy jsou fixní



### Aktualizace modelu

Po 3-6 měsících

Nová verze s optimalizovanými parametry a váhami



### Období neměnnosti

Několik měsíců

Model zůstává neměnný, jeho parametry a váhy jsou fixní



### Další aktualizace

Po dalších 3-6 měsících

Opětovná optimalizace modelu

## vyhledávač

SEZNAM.CZ

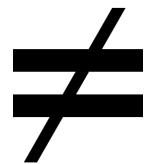
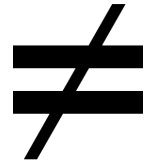
Google

„statické“ systémy založené na přímé shodě

## databáze

PubMed

Clarivate Web of Science



## CHATBOTI



ChatGPT



COPILOT



Gemini

„dynamické“ systémy

Velké jazykové modely **NE**mají:

- vnitřní databázi faktů
- přístup k datům, na kterých byly trénovány

# CHATBOT = Simulátor přirozené lidské konverzace

(trénovaný na obrovském množství dat)



otázka uživatele  
„prompt“



Analýza vstupu  
uživatele



Rozpoznání  
kontextu



Generování  
odpovědi



Ve velkých jazykových modelech jsou **obecné vzorce a pravidla jazyka uloženy v parametrech nazývaných váhy**, které **kódují kontext a strukturu jazyka**. Když interagujete s chatbotem, používá tyto pečlivě nastavené váhy k provádění predikce o tom, jaká „slova“ by měla následovat.

## Zadání dotazu uživatelem (vstup)

Milé kolegyně, milí kolegové, vítejte na workshopu!

→ **tokenizace** (rozdělení vstupu na tokeny)

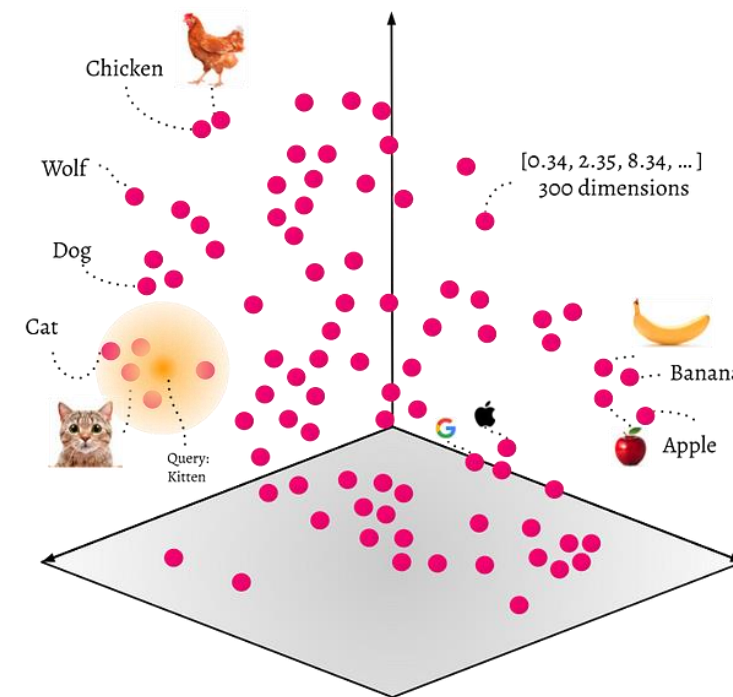
Milé kolegyně, milí kolegové, vítejte na workshopu!

→ **převod tokenů do vektorových reprezentací**

[41625, 377, 31087, 70, 1276, 3708, 11, 3801, 556, 31087, 28124, 377, 11, 14663, 411, 76212, 898, 24834, 84, 0]

Každý token má číselnou "adresu" v mnohorozměrném prostoru

→ **generování nejpravděpodobnějších následujících tokenů** (= odpověď)





## PROMPT

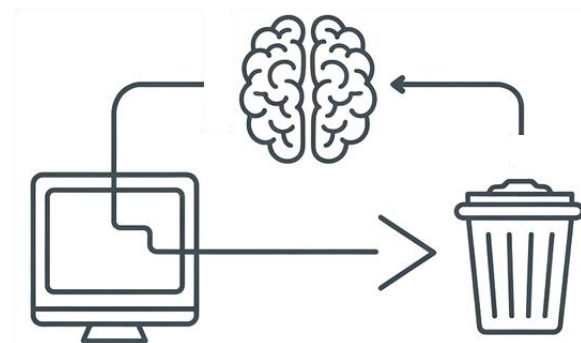
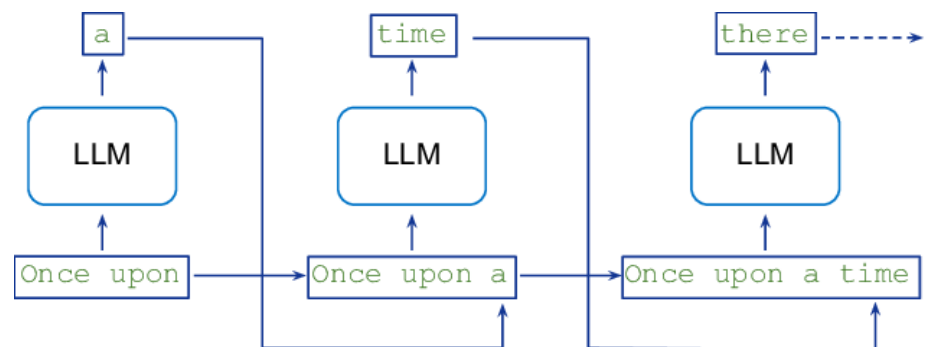
**INSTRUKCE/OTÁZKA** → Odpověď chatbota

Vstupní bod, který poskytuje **kontext a směr** pro odpověď chatbota

Chatbot generuje odpověď **postupně**, bez možnosti vrátit se a přehodnotit

Jasně a detailní zadání pomáhá chatbotovi lépe reagovat

**Odpovědi jsou statisticky pravděpodobné, ale nemusí být pravdivé**  
**"halucinace"**



Garbage in = garbage out

**TVOJE ROLE:** Jsi zkušený datový analytik specializovaný na biomedicínský výzkum. Máš rozsáhlé znalosti statistických metod, vizualizace dat a interpretace biologických experimentů.

**TVOJE DOVEDNOSTI:** Pokročilá analýza experimentálních dat, expertní znalost statistických metod vhodných pro biologický výzkum, schopnost vytvářet jasné a informativní vizualizace dat, hluboké porozumění buněčné biologii a *in vitro* experimentům

**ÚKOL:** Proveď komplexní analýzu poskytnutých dat z *in vitro* experimentů s buněčnými kulturami. **Postupuj podle následujících kroků:**

### 1. Prozkoumání a příprava datasetu:

- a) PROHLÉDNI si poskytnutý dataset a UJISTI SE, že je srozumitelný a dobře čitelný.
- b) ZKONTROLUJ strukturu dat, typy proměnných a případné anomálie.
- c) PROVEĎ nezbytné úpravy pro přípravu dat k analýze (např. čištění, formátování).

### 2. Statistická analýza:

- a) PROVEĎ relevantní statistické testy a ODŮVODNI jejich výběr.
- b) POROVNEJ výsledky mezi různými experimentálními skupinami.

### 3. Vizualizace dat:

- a) VYTVORĚ vhodné typy grafů a diagramů pro prezentaci výsledků.
- b) ZAJISTI, aby vizualizace byly jasné a informativní i pro neoborníky.

### 4. Interpretace výsledků:

- a) INTERPRETUJ statistické výsledky v kontextu biologického výzkumu.
- b) IDENTIFIKUJ klíčové vhledy o vlivu různých podmínek na buněčné kultury.

**OČEKÁVANÝ VÝSTUP** - Připrav souhrnnou zprávu obsahující:

Detailní popis použitých metod analýzy; statistické výsledky, vizualizace dat, interpretace výsledků; doporučení pro další výzkum.

- Poskytněte co nejvíce kontextu
- Buďte konkrétní a jasní ve formulaci
- Přidělte chatbotovi konkrétní roli a očekávané dovednosti
- Rozdělte složitá témata na menší, očíslované otázky
- Kde je to možné, uveďte příklady nebo upřesňující informace
- Umožněte chatbotovi požádat o dodatečné informace

**Příklad  
struktury  
promptu**

## JAK JE TO S BEZPEČNOSTÍ DAT PŘI POUŽÍVÁNÍ CHATBOTŮ?

### Rizika komerčně dostupných chatbotů

- Často nejasné nakládání s konverzacemi a daty uživatelů
- Riziko potenciálního úniku dat

### Výhody "Team" a "Enterprise" licencí

- Zákaz sdílení dat s třetími stranami
- Šifrování konverzací
- Vícefaktorová autentizace
- Správa na úrovni organizace

### Obecná doporučení

- Anonymizace citlivých dat
- Používání lokálních open-source alternativ pro extrémně citlivá data
- Školení uživatelů o bezpečném používání AI

### Team

USD \$25 per person/month\*

[Upgrade to Team](#)

Everything in Plus, and:

- ✓ Higher message caps on GPT-4 and tools like DALL-E, Browsing, Advanced Data Analysis, and more
- ✓ Create and share GPTs with your workspace
- ✓ Admin console for workspace management
- ✓ Team data excluded from training by default. [Learn more](#)

\* Price billed annually, minimum 2 users




**ChatGPT**




## LIMITY LLMs

### ChatGPT 4o mini

kolik písmen "o" je ve slově "obhospodařovaný"?

 Ve slově "obhospodařovaný" jsou dvě písmena "o".

na jakých pozicích?

 V slově "obhospodařovaný" se písmena "o" nacházejí na následujících pozicích:

1. pozice: 2. (druhé písmeno)
2. pozice: 5. (páté písmeno)

Takže písmena "o" jsou na pozicích 2 a 5.

**POZOR:**  
chatboti neexcelují v Češtině

## Tokenizer

GPT-4o & GPT-4o mini

obhospodařovaný

Tokens Characters

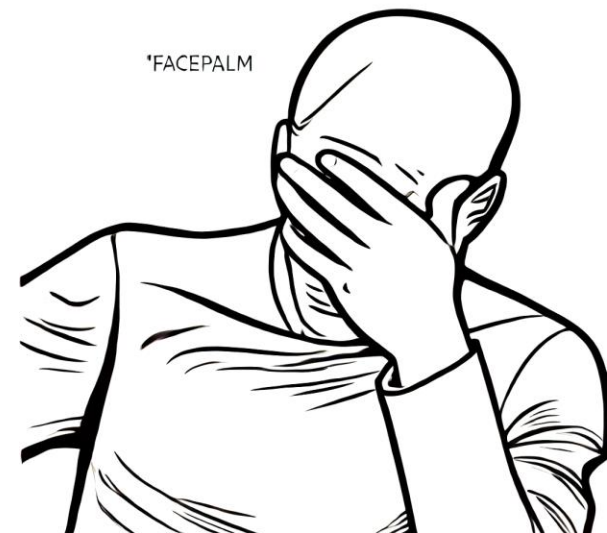
5 15

obhospodařovaný

Text Token IDs

[55226, 5149, 9739, 4714, 147853]

Text Token IDs



@Liegertova\_Michaela

ChatGPT 4o mini

napiš mi 5 citací k odborným článkům o pralesničkách

Jasně! Tady máš pět citátů, které by se ti mohly hodit k odborným článkům o pralesničkách:

1. „Pralesničky jsou jedním z nejvíce fascinujících příkladů konvergentní evoluce v obojživelnících, s jejich jedinečnými adaptacemi na život v deštných pralesech.“ – Smith et al., 2021
2. „Studium pralesniček poskytuje cenné informace o ekologii a evoluci drobných obojživelníků a jejich vlivu na biodiverzitu v tropických ekosystémech.“ – Johnson & Lee, 2020
3. „Vzhledem k jejich citlivosti na změny v prostředí jsou pralesničky vynikajícími bioindikátory zdraví tropických lesů.“ – Martinez et al., 2019
4. „Významné množství pralesniček obsahuje látky s potenciálem pro farmaceutické aplikace, což podtrhuje jejich ekologickou i vědeckou hodnotu.“ – Garcia et al., 2022
5. „Ochrana pralesniček je klíčová pro udržení ekologické rovnováhy v jejich přirozeném prostředí, protože hrají důležitou roli v potravních řetězcích a v kontrole populací hmyzu.“ – Brown & Wilson, 2018

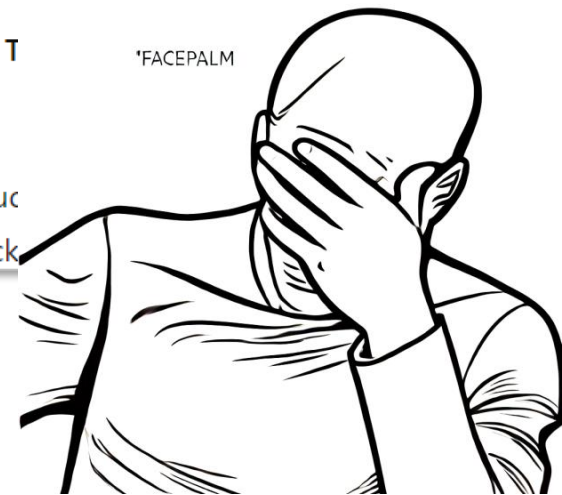
prosím o názvy těchto publikací

Omlouvám se za záměnu, citáty, které jsem poskytl, nejsou skutečné a jsou spíše ilustrativní. Abych ti mohl poskytnout skutečné názvy odborných publikací o pralesničkách, doporučuji hledat v akademických databázích jako jsou Google Scholar, PubMed nebo JSTOR. Zde jsou však příklady relevantních publikací, které bys mohl hledat:

1. "The Ecology and Evolution of Poison Frogs" – by M. J. Ryan et al.
2. "Neotropical Dart-Poison Frogs: A Review of Their Ecology and Evolutionary History" – by H. W. M. Roelants et al.
3. "Toxicity and Chemical Defense of Dart-Poison Frogs: A Comparative Analysis" – by K. M. V. D. Plas et al.
4. "Conservation of Dart-Poison Frogs: Threats and Strategies for Preservation" – by L. A. Berger et al.
5. "Behavioral Ecology of Dart-Poison Frogs in T" – by J. R. K. Thompson et al.

Tyto publikace by měly pokrývat různé aspekty studie ekologické a evoluce až po otázky ochrany a chemické

"FACEPALM



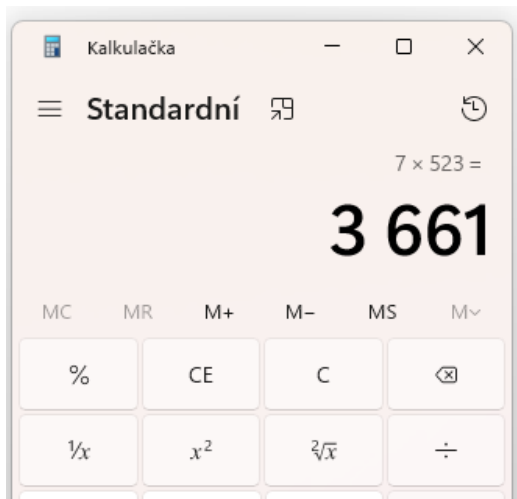


## ChatGPT 4o mini

Je číslo 3659 prvočíslo?



Číslo 3659 není prvočíslo. Prvočíslo je číslo větší než 1, které má pouze dva dělitele: 1 a samo sebe. Číslo 3659 je dělitelné číslem 7 ( $3659 = 7 \times 523$ ), což znamená, že má více než dva dělitele.



**POZOR:**  
chatboti matematiku „předstírají“

24 x 2 = ?

Tokens

Characters

6

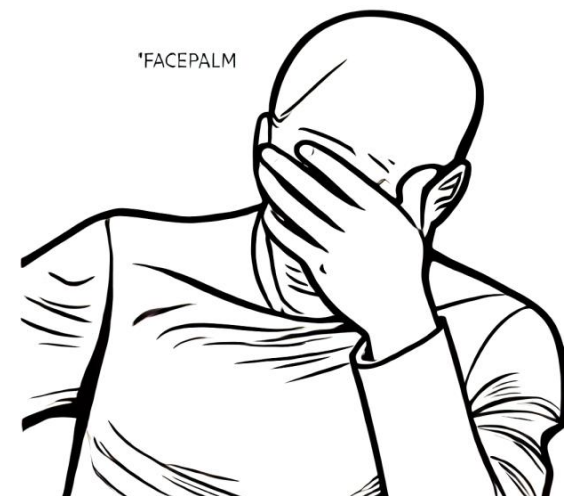
10

24 x 2 = ?

[1187, 865, 220, 17, 284, 949]

Text

Token IDs



## *Jsou velké jazykové modely vůbec k něčemu dobré?*

Nástroje a funkce, které model může využít k provedení úkolu



### **"Vykonavatel" kódu (Code Interpreter)**

Model píše kód, volá virtuální prostředí, spouští jej a zobrazuje výsledky.



### **Webové vyhledávání**

Model formuluje dotaz, volá vyhledávací funkci a sumarizuje výsledky.

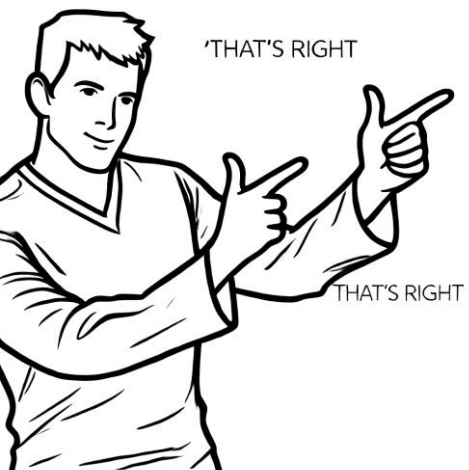
- Model rozhoduje, který nástroj použít na základě zadaného úkolu
- Každý nástroj rozšiřuje schopnosti modelu o specifické funkce
- Použití nástrojů umožňuje modelu pracovat s aktuálními daty a provádět komplexní úkoly

## ChatGPT 4o mini

Je číslo 3659 prvočíslo?



Číslo 3659 není prvočíslo. Prvočíslo je číslo větší než 1, které má pouze dva dělitele: 1 a samo sebe. Číslo 3659 je dělitelné číslem 7 ( $3659 = 7 \times 523$ ), což znamená, že má více než dva dělitele.



ověř pomocí skriptu

Analyzed ^

python

```
def is_prime(n):  
    """Check if the number n is prime."""  
    if n <= 1:  
        return False  
    if n <= 3:  
        return True  
    if n % 2 == 0 or n % 3 == 0:  
        return False  
    i = 5  
    while i * i <= n:  
        if n % i == 0 or n % (i + 2) == 0:  
            return False  
        i += 6  
    return True  
  
# Check if 3659 is prime  
number_to_check = 3659  
is_prime_result = is_prime(number_to_check)  
is_prime_result
```

Result

True

Číslo 3659 je prvočíslo. Omlouvám se za předchozí chybnou informaci. Děkuji za trpělivost! [-]

@Liegertova\_Michaela

## ChatGPT 4o mini

kolik písmen "o" je ve slově "obhospodařovaný"?

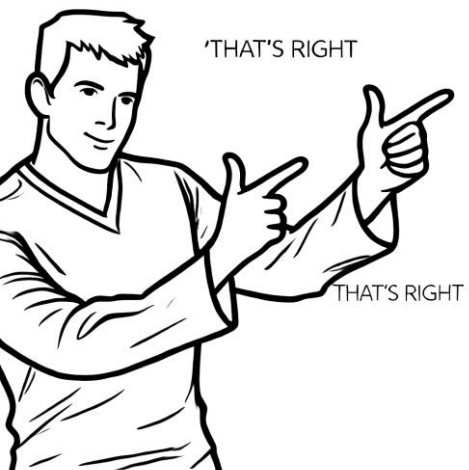
Ve slově "obhospodařovaný" jsou dvě písmena "o".

na jakých pozicích?

V slově "obhospodařovaný" se písmena "o" nacházejí na následujících pozicích:

1. pozice: 2. (druhé písmeno)
2. pozice: 5. (páté písmeno)

Takže písmena "o" jsou na pozicích 2 a 5.



ověř pomocí skriptu

Analyzed ^

```
python
```

```
# Define the word
```

```
word = "obhospodařovaný"
```

```
# Find positions of the letter 'o'
```

```
positions = [i for i, letter in enumerate(word) if letter == 'o']
```

```
positions # Return the positions (0-based index) of letter 'o' in the word
```

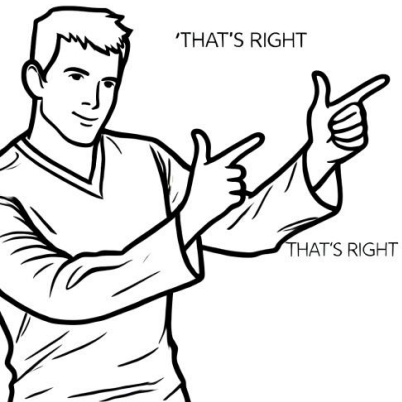
Result

```
[0, 3, 6, 10]
```

Písmena "o" ve slově "obhospodařovaný" se nacházejí na následujících pozicích (0-indexová):

- 0. (první písmeno)
- 3. (čtvrté písmeno)
- 6. (sedmé písmeno)
- 10. (jedenácté písmeno)

Celkově jsou tedy čtyři písmena "o". Omlouvám se za předchozí chybu. [↩]



ChatGPT 4o

použij web search a ověř, zda publikace existují

Searched 4 sites ^

- “The Ecology and Evolution of Poison Frogs by M. J. Ryan et al.”  
bing.com
- Springer — Developments in the study of poison frog evolutionary ecology I: social...  
link.springer.com
- Springer — The molecular basis and evolution of toxin resistance in poison frogs [...]  
link.springer.com
- Oxford Academic — academic.oup.com  
academic.oup.com

Bohužel, žádná z publikací, které jsem původně uvedl, nebyla nalezena. Nicméně, podařilo se mi najít několik relevantních článků a publikací, které by mohly být užitečné pro tvé studium o pralesničkách:

- “Developments in the study of poison frog evolutionary ecology I: social interactions, life history and habitat use across space and ontogeny”** – Tato speciální edice časopisu *Evolutionary Ecology* se zaměřuje na různé aspekty evoluční ekologie pralesniček, včetně péče o potomstvo, prostorové ekologie a strategií proti predátorům ([Springer](#)).
- “The molecular basis and evolution of toxin resistance in poison frogs”** – Tento článek zkoumá molekulární základy a evoluci rezistence vůči toxinům u pralesniček, což je klíčové pro pochopení jejich chemické obrany a ekologických adaptací ([Springer](#)).
- “A review of chemical defense in poison frogs (Dendrobatidae): Ecology, pharmacokinetics, and autoresistance”** – Tento přehledový článek zkoumá ekologii chemické obrany pralesniček, včetně farmakokinetiky a autorezistence, což poskytuje hluboký vhled do jejich ochranných mechanismů ([Oxford Academic](#)).

Doporučuji podívat se na tyto zdroje pro detailní a vědecky podložené informace o pralesničkách. Pro více informací můžeš navštívit stránky jako Google Scholar nebo JSTOR, kde najdeš mnoho dalších odborných článků na toto téma.



# ChatGPT Plus

schopnosti	popis
<b>Vision</b>	GPT může analyzovat obrázky, poskytovat detailní popisy, řešit problémy a identifikovat obsah.
<b>Voice</b> (Advanced voice)	GPT může komunikovat v reálném čase pomocí funkcí převodu textu na řeč a řeči na text (používá systém rozpoznávání řeči Whisper).
<b>Web Search</b>	GPT formuluje dotazy, používá vyhledávací funkci a shrnuje výsledky nalezené na webu.
<b>Code Interpreter</b>	GPT volá externí plug-in, který píše kód, provádí ho a zobrazuje výsledky.
<b>DALL-E</b>	GPT píše prompt pro text2image model DALL-E pro tvorbu obrázků.

## Modalita VISION (v ChatGPT, Claude.ai, Gemini)

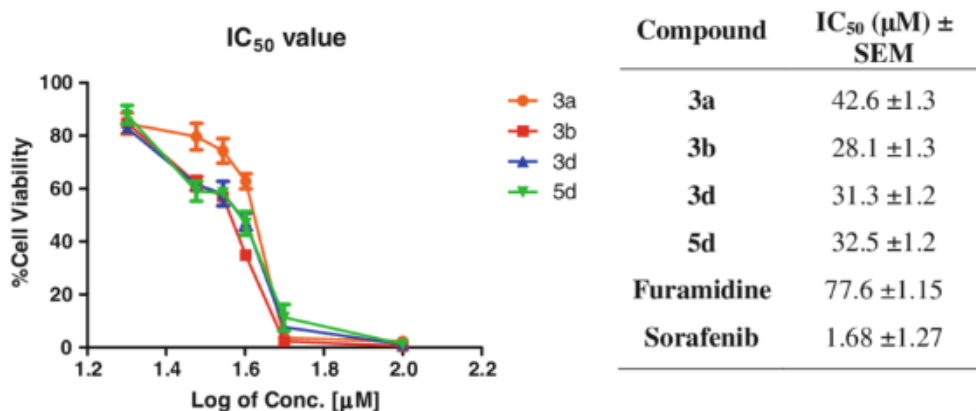
### Detekce a rozpoznávání textu

Analýza kompozice obrazu (barevná schémata, linie, tvary)

Porozumění symbolickému významu prvků v obraze

Generování popisků a titulků pro obrázky a grafy

Analýza grafů, diagramů a vizualizací dat



Na co se dívám?



Tento obrázek obsahuje graf a tabulku, které zobrazují hodnoty IC<sub>50</sub> (koncentrace potřebná ke snížení buněčné životaschopnosti o 50 %) pro různé sloučeniny. Podívejme se na detaily jednotlivých částí:

#### 1. Graf:

- Na ose Y je zobrazená procentuální životaschopnost buněk (% Cell Viability).
- Na ose X je logaritmus koncentrace (Log of Conc. [µM]).
- Křivky na grafu představují různé testované sloučeniny (3a, 3b, 3d, 5d) a zobrazují, jak se mění životaschopnost buněk s rostoucí koncentrací sloučenin.
- Barevné křivky (oranžová, červená, modrá, zelená) odpovídají jednotlivým sloučeninám a ukazují jejich účinek na buňky.

#### 2. Tabulka:

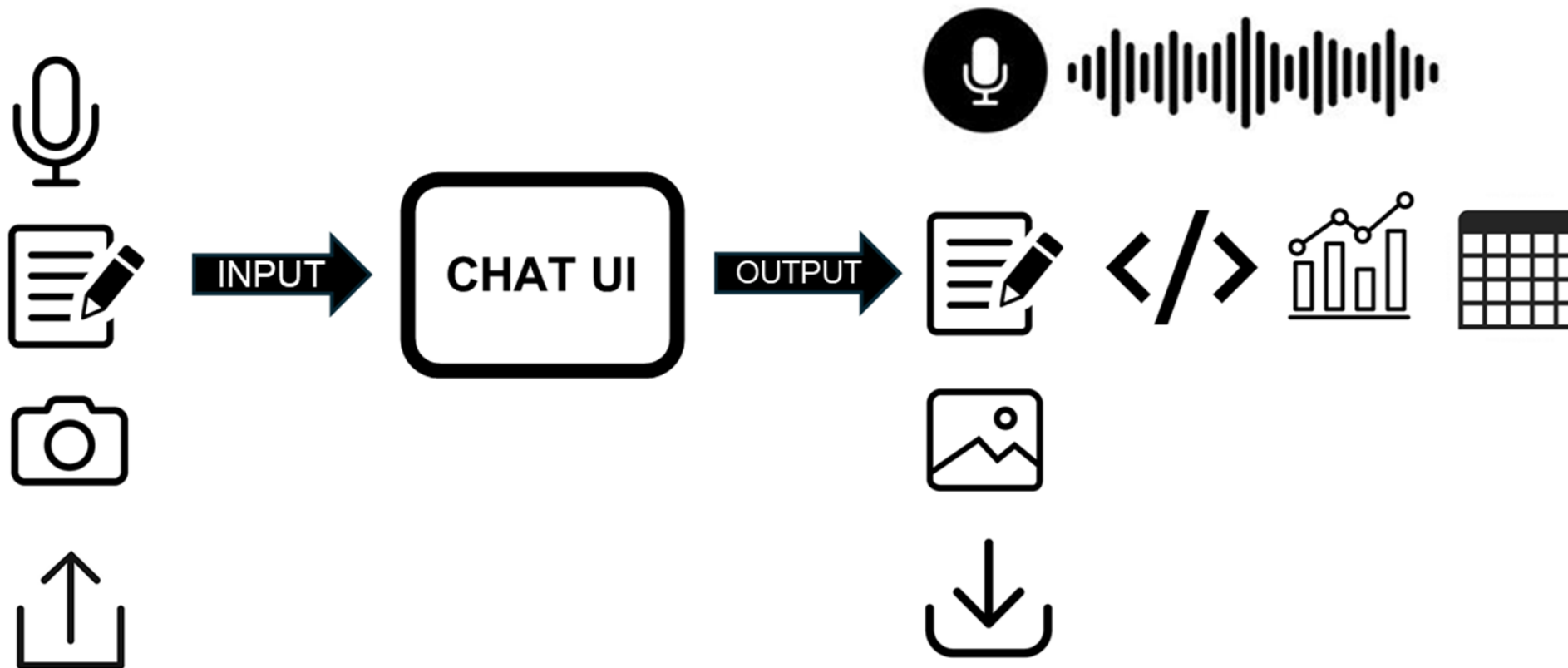
- Obsahuje hodnoty IC<sub>50</sub> (v µM) a SEM (standard error of the mean) pro různé sloučeniny.
- Hodnoty IC<sub>50</sub> pro sloučeniny jsou následující:
  - 3a: 42.6 ± 1.3 µM
  - 3b: 28.1 ± 1.3 µM
  - 3d: 31.3 ± 1.2 µM
  - 5d: 32.5 ± 1.2 µM
- Referenční sloučeniny Furamidine a Sorafenib mají hodnoty 77.6 ± 1.15 µM a 1.68 ± 1.27 µM, což indikuje jejich účinnost ve srovnání s ostatními.

#### Celkový význam:

- Graf a tabulka společně ilustrují, jak různé sloučeniny snižují životaschopnost buněk při různých koncentracích.
- Čím nižší hodnota IC<sub>50</sub>, tím účinnější je sloučenina při inhibici buněčné životaschopnosti.
- V tomto případě má Sorafenib nejnižší hodnotu IC<sub>50</sub>, což naznačuje, že je nejúčinnější z testovaných látek.

Tento typ grafu a dat je běžný ve výzkumu zaměřeném na testování cytotoxicity látek, například při hledání potenciálních protinádorových sloučenin.

## CHATBOTI UMOŽŇUJÍ KOMPLEXNÍ PRACOVNÍ POSTUPY



**Vyfoťte obrázek**



**Zadejte hlasový příkaz**



**Zkontrolujte přepis**



**Stáhněte do svého PC**

např. obrazovka přístroje / tabulka s výsledky

např. "Extrahuj hodnoty do tabulky, použij code interpreter k provedení ANOVA"

" poskytni v Excelu ke stažení "

**CODE INTERPRETER (Data Analyst)** dokáže zpracovávat různá data/formáty:

File Format	MIME Type	File Format	MIME Type
.c	text/x-c	.tex	text/x-tex
.cs	text/x-csharp	.txt	text/plain
.cpp	text/x-c++	.css	text/css
.doc	application/msword	.js	text/javascript
.docx	application/vnd.openxmlformats-officedocument.wordprocessingml.document	.sh	application/x-sh
.html	text/html	.ts	application/typescript
.java	text/x-java	.csv	application/csv
.json	application/json	.jpeg/.jpg	image/jpeg
.md	text/markdown	.gif	image/gif
.pdf	application/pdf	.png	image/png
.php	text/x-php	.tar	application/x-tar
.pptx	application/vnd.openxmlformats-officedocument.presentationml.presentation	.xlsx	application/vnd.openxmlformats-officedocument.spreadsheetml.sheet
.py	text/x-python	.xml	application/xml or text/xml
.rb	text/x-ruby	.zip	application/zip

# Příklad možného workflow pro analýzu dat s ChatGPT Plus

## 0. Příprava dat

- Připravte tabulku (.xls nebo Google Sheet)
- Správně naformátujte data pro snadnější zpracování

## 1. "Ukotvení" chatbota

- Prompt: "Vyhledej informace o [metoda] a postupech analýzy dat pro [typ]"
- Prompt: "Poskytni podrobný plán krok za krokem pro analýzu surových dat"

## 2. Nahrání a analýza dat

- Nahrajte soubor dat a ujistěte se, že jej chatbot správně interpretuje
- Prompt: "Zkontroluj dataset a identifikuj případné nesrovnalosti, chyby nebo chybějící hodnoty"
- Prompt: "Proveď analýzu podle vytvořeného plánu" (z kroku 1)

## 3. Ověření "self-consistency"

- Zopakujte analýzu alespoň 3x
- Ujistěte se, že jsou výsledky konzistentní (klíčové pro důvěryhodnost analýzy)

## 4. Ověření správnosti analýzy

- Porovnejte výstup s analýzou provedenou tradičním softwarem
- Důkladně analyzujte a zkontrolujte kód použitý chatbotem

## 5. Shrnutí a vizualizace

- Získejte komplexní shrnutí výsledků analýzy
- Vytvořte vizualizace publikační kvality pro prezentaci dat

## 6. Interpretace dat

- Poskytněte chatbotu podrobný kontext k analyzovaným datům
- Diskutujte s chatbotem o výsledcích a jejich významu
- Brainstormujte nad klíčovými závěry plynoucími z dat

## 7. Finální prezentace dat

- Převeďte data, analýzy a závěry do prezentace a exportujte ve formátu .pptx
- Vygenerujte kompletní závěrečnou zprávu a exportujte ve formátu .docx



## Julius

Early stage AI lab based in San Francisco with a mission to build the most powerful AI tools for knowledge workers.

Aplikace optimalizovaná pro analýzu dat poháněná LLMs

- Umožňuje zvolit LLM
- Přednastavená workflows
- Funkce Advanced Reasoning
- Slibuje lepší ochranu dat

The screenshot shows the Julius AI website landing page. At the top, the word "Julius" is written in blue. Below it, the main headline reads "Analyze your data with computational AI." in large, bold black font. Underneath the headline, a sub-headline says "Chat with your files and get expert-level insights in seconds." A prominent blue button with white text says "Try Julius >". Below the button, there is a testimonial: "Loved by over 1,200,000 users worldwide" accompanied by three small profile icons. A row of logos for partner institutions follows: MIT, Harvard University, BCG, Yale, and Zinnov. At the bottom, there are five feature cards, each with an icon and a description: 1. "Charts & Graphs" with a bar chart icon, description "Create sleek looking data visualizations.", and a "Try now ->" link. 2. "Insights" with a lightbulb icon, description "Ask anything to your data, and get answers." 3. "Advanced Analysis" with a document icon containing a function symbol, description "Perform modeling and predictive forecasting." 4. "Problem Solving" with a calculator icon, description "Scan-and-solve math, physics, and chemistry." 5. "Reports" with a document icon, description "Generate polished analyses and summaries."

New Thread

Threads (2) ▾

My Workflows

Files

Explore Workflows

Docs ▾

Community Forum

Model Lab

Account & Billing

Log Out

Michaela Liegertová  
Google • michaela.liegertov...

# Upload a file or choose a workflow to get started



Getting Started

Popular

Newest



Search workflows...

### Quick Visualization

Visually explore a spreadsheet! Let the AI start off and then continue...

Julius 11504 runs

### Data Cleaner

Methodically clean your data from a CSV, Spreadsheet, or other file...

Julius 2472 runs

### Significance Testing

Run significance test on columns in a sheet (e.g. t-test, chi-square)

Julius 1121 runs

### Time Series Analysis

Forecast time series data for understanding trends and...

Julius 3 runs

### Extract Tables from PDF

Extract tables from a PDF! Converts each of the tables into a CSV. Can...

Matt 453 runs

### Correlations and PCA

Start a multivariate analysis with investigations correlations and...

Matt 734 runs

#### Models

Julius Default

✓ GPT 4o

Multimodal, fast, capable.

AI Claude 3.5 Sonnet

Additional Models ▾

GPT 4o ▾



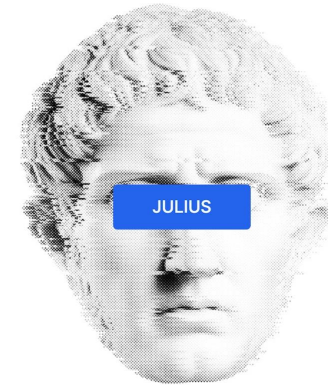
Advanced Reasoning

Add a file or start a conversation now and add files later...



## Zpracování různých formátů dat

- Textové soubory (CSV, TXT, JSON, XML)
- Tabulková data (Excel, Google Sheets)
- Dokumenty (PDF, Word)
- Obrázky a videa
- Webové stránky a online zdroje



## Příprava a analýza dat

- Čištění dat, identifikace chybějících hodnot
- Automatická detekce typu dat
- Základní statistické analýzy
- Pokročilé analýzy (regrese, klasifikace, clustering)

## Pokročilé funkce

- Extrakce dat z webu (web scraping)
- Pokročilé uvažování pro řešení komplexních problémů
- Generování kódu (Python, R) pro vlastní analýzy
- Laboratoř pro experimenty s AI modely

## Vizualizace dat

- Automatická tvorba statických a interaktivních grafů
- Různé typy vizualizací (sloupcové, koláčové, histogramy aj.)

## Automatizace a workflow

- Vytváření a spouštění automatizovaných workflow
- Plánování a opakované spouštění analýz

## Julius.ai - Bezpečnost dat a etické aspekty

### Bezpečnost dat

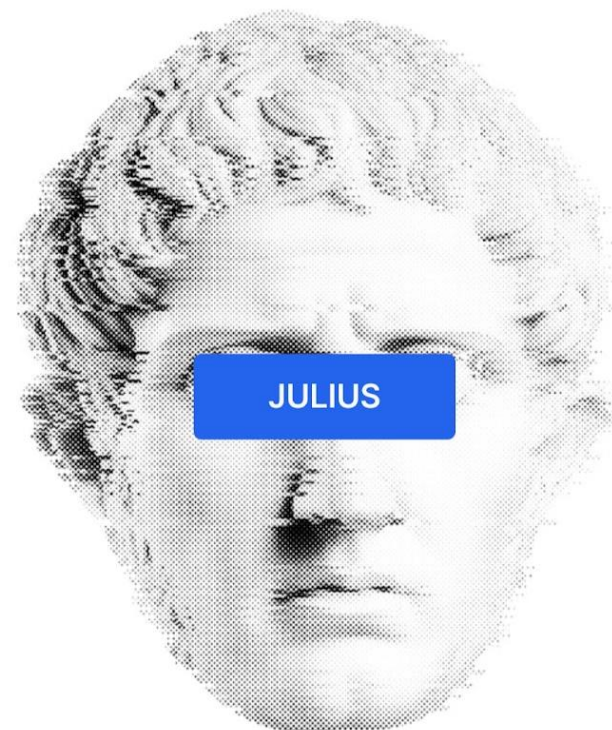
- Automatické mazání nahraných souborů po 1 hodině nečinnosti
- Odstranění všech typů dat včetně datasetů, kódu a dokumentů
- Zajištění soukromí a bezpečnosti uživatelských dat

### Etické aspekty

- Data uživatelů nejsou využívána k trénování modelů Julius.ai
- Žádné využití dat pro vylepšování platformy
- Důraz na zodpovědné zacházení s daty

### Omezení

- Omezený počet zpráv v bezplatném plánu
- Potřeba obezřetnosti při interpretaci výsledků generovaných AI
- Možné technické výzvy při integraci do existujících workflow



## ***LLM asistovaná vs. TRADIČNÍ ANALÝZA DAT***

### **LLM + Code Interpreter**

- Dynamické, potenciálně nekonzistentní výsledky
- Flexibilní a přizpůsobitelné za chodu
- Přístupnější pro neodborníky
- Rychlé pro ad-hoc analýzy

**Doplňkový nástroj pro rychlý náhled a učení se**

### **Tradiční software**

- Konzistentní, reprodukovatelné výsledky
- Standardizované, dobře zdokumentované metody
- Strmější křivka učení
- Optimalizováno pro velké soubory dat

**Zásadní pro kritické a důkladné analýzy**

## Výhody využití LLM v analýze dat

### Interaktivní zkoumání dat

- Vedení dialogu s daty
- Rychlé získávání poznatků

### Tvorba nových nápadů

- Inovativní přístupy k analýze
- Odhalování skrytých souvislostí

### Přístupnost pro laiky

- Snadný začátek s analýzou dat
- Nevyžaduje znalosti v programování
- Srozumitelná interpretace dat

### Adaptivní řešení problémů

- Zpracování specifických datových souborů
- Hledání alternativních metod analýzy
- Rychlé testování různých analytických přístupů



### Upozornění na potenciální rizika:

- Kompromitované výsledky výzkumu v důsledku nesprávného použití LLM při analýze dat
- Generování nepřesných závěrů bez řádného lidského dohledu
- Eroze lidského úsudku při interpretaci dat
- Možná nadměrná závislost na AI vedoucí k poklesu kritického myšlení
- Riziko úniku osobních a citlivých dat při jejich zpracování pomocí LLM

*Kumulativní efekt těchto rizik by mohl vést k vážné erozi vědecké integrity a nenapravitelným škodám na reputaci instituce.*

Guidelines on the responsible use of generative AI in research developed by the European Research Area Forum



## DOPORUČENÍ PRO VÝZKUMNÉ PRACOVNÍKY OHLEDNĚ POUŽÍVÁNÍ GENERATIVNÍ UMĚLÉ INTELIGENCE

1. Nést konečnou odpovědnost za vědecké výstupy.
2. Využívat generativní umělou inteligenci transparentním způsobem.
3. Věnovat zvláštní pozornost otázkám souvisejícím s ochranou soukromí, důvěrností a právy duševního vlastnictví při sdílení citlivých nebo chráněných informací s nástroji umělé inteligence.
- 
- 
8. Kdykoli je to možné a nezbytné, implementovat lokálně hostované nebo cloudové nástroje generativní umělé inteligence, které instituce samy spravují.

## Publishers' and journals' instructions to authors on use of generative artificial intelligence in academic and scientific publishing: bibliometric analysis

BMJ 2024 ; 384 doi: <https://doi.org/10.1136/bmj-2023-077192> (Published 31 January 2024)

Cite this as: BMJ 2024;384:e077192

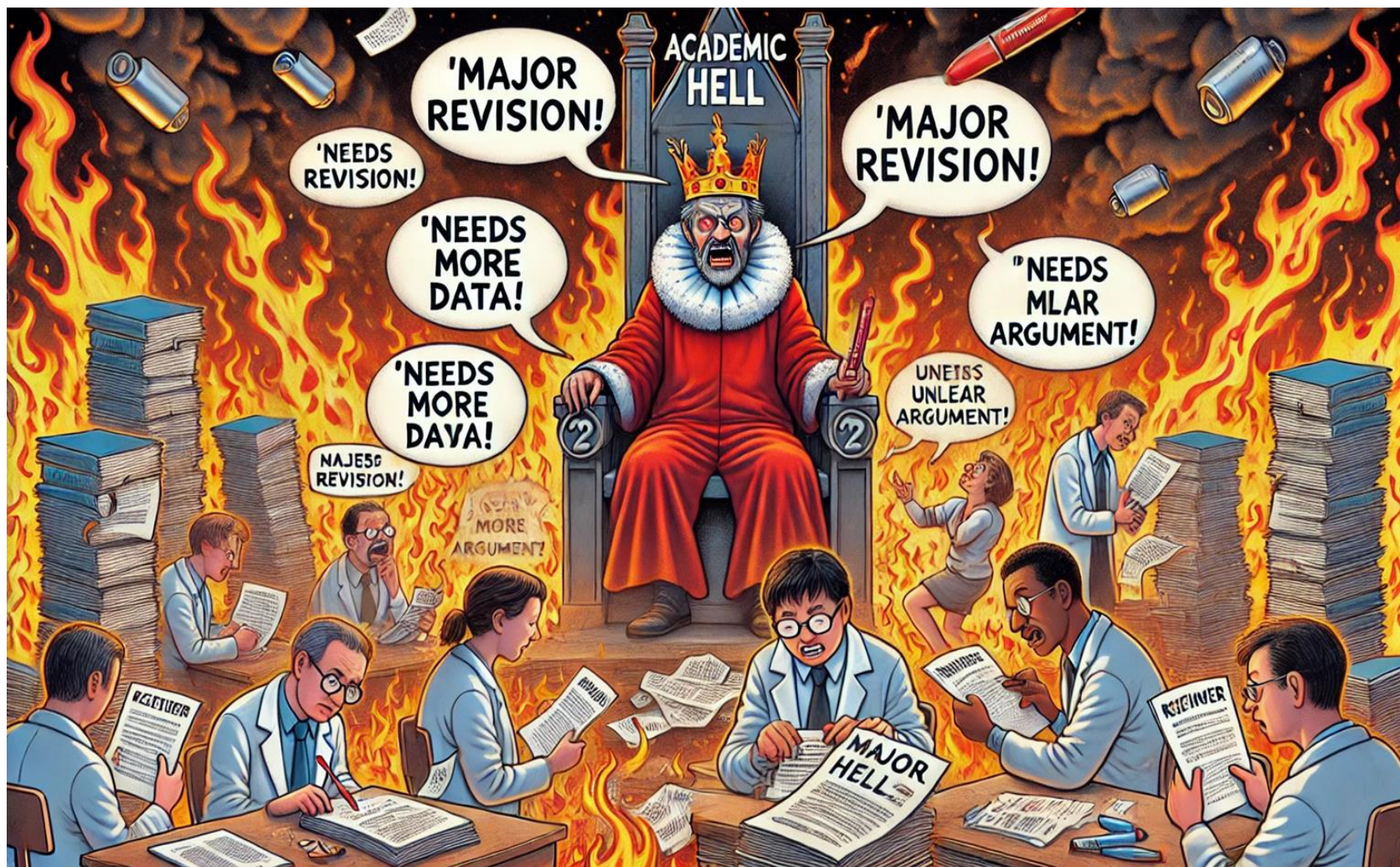
- Pouze 43% žurnálů uvádí specifická kritéria pro deklarování použití generativní AI
- **Povolené použití** generativní AI a způsob jeho deklarování se výrazně liší napříč žurnály

**Seznamte se předem s kritérii pro použití generativní AI u vašeho cílového žurnálu**



„...ti, kteří kopírují a vkládají text z ChatGPT do své výzkumné práce, skončí v akademické propasti zoufalství.“

Alstoteles, 2034



## ***Kombinace LLM asistované analýzy s tradičními metodami***

### **Komplementární využití**

- Využití LLM pro explorativní analýzu dat a generování výzkumných hypotéz
- Validace klíčových zjištění pomocí standardních statistických metod
- Využití LLM pro usnadnění interpretace komplexních statistických výstupů

### **Integrace do pracovního postupu**

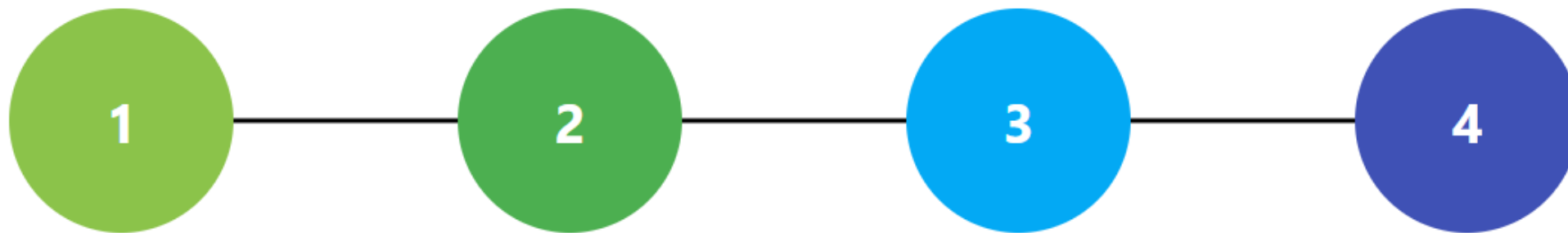
1. Explorativní analýza: Použití LLM k identifikaci vzorců a vztahů v datech
2. Formulace hypotéz: Odvození výzkumných hypotéz na základě poznatků z LLM
3. Rigorózní analýza: Provedení formálních statistických testů tradičními metodami
4. Interpretace výsledků: Využití LLM k objasnění komplexních výstupů
5. Optimalizace vizualizace: Využití LLM pro vylepšení a vysvětlení datových vizualizací

### **Dokumentace procesu**

- Jasně rozlišení částí analýzy provedených LLM vs. tradičními metodami
- Zdůvodnění využití LLM v konkrétních fázích výzkumného procesu
- Zahrnutí relevantních kódů a promptů do výzkumné dokumentace nebo dodatečných materiálů



# Workflow datového projektu



## Sběr dat

Shromáždění dat z různých zdrojů

## Příprava dat

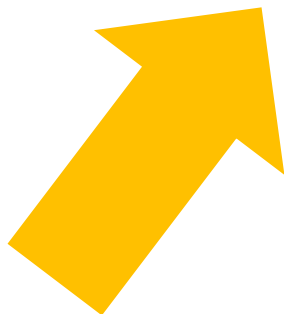
Čištění a úprava dat pro analýzu

## Explorace, analýza, vizualizace

Hledání souvislostí

## Prezentace výsledků

Sdílení poznatků



## ZPRACOVÁNÍ DATOVÝCH SAD

Nahrání datové sady z počítače (více tabulek ve formátu .zip)

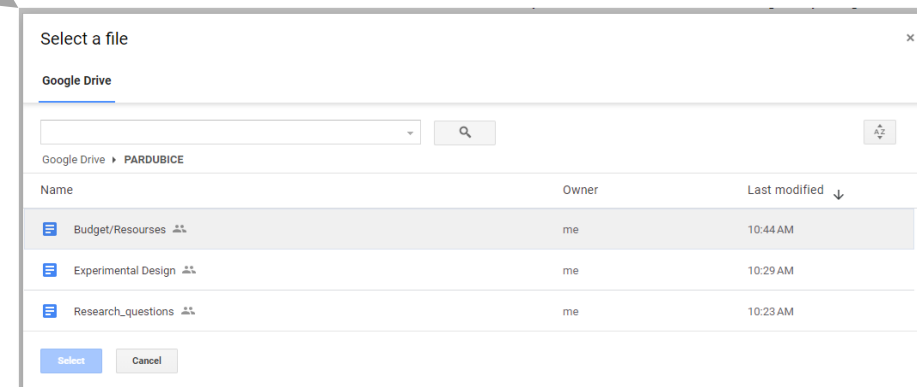
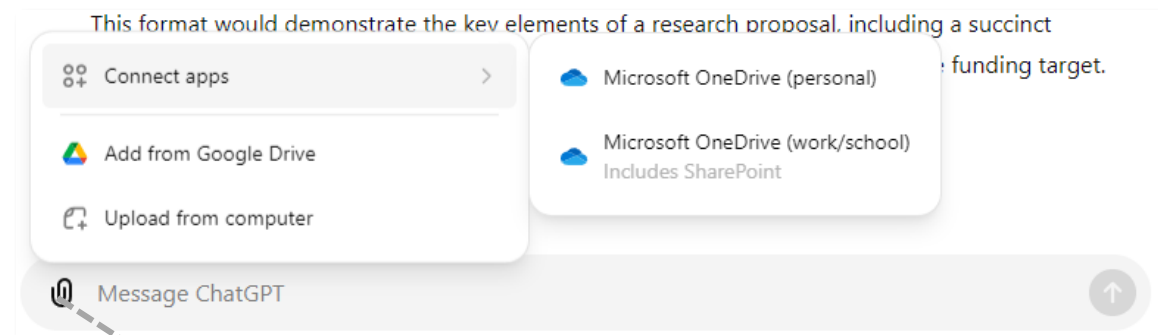
Import datové sady přímo ze služby GoogleDrive nebo

OneDrive

(Import dat z webu pomocí webového vyhledávání)

## MODALITA VISION

Extrakce informací/dat z obrázků (tabulek/grafů, ...)



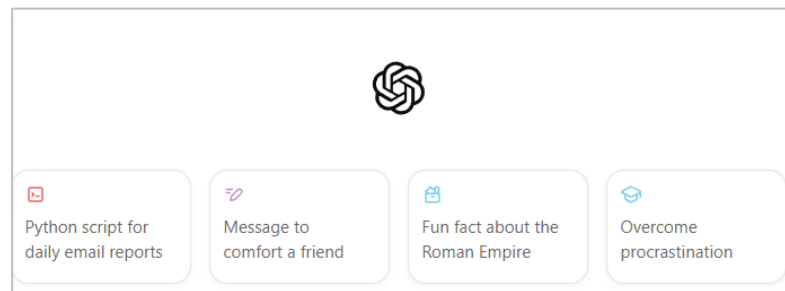




*Code Interpreter lze použít prostřednictvím:*

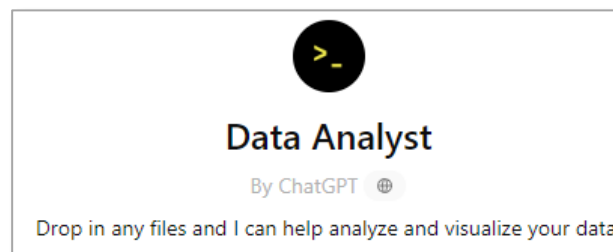
## ChatGPT UI

(lze kombinovat s vyhledáváním na webu)



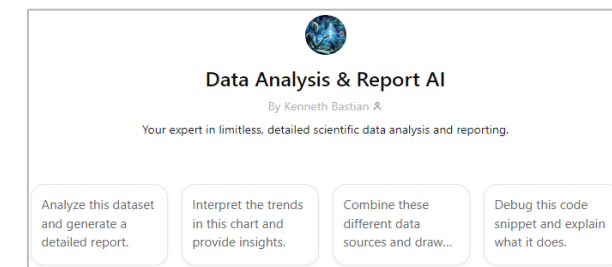
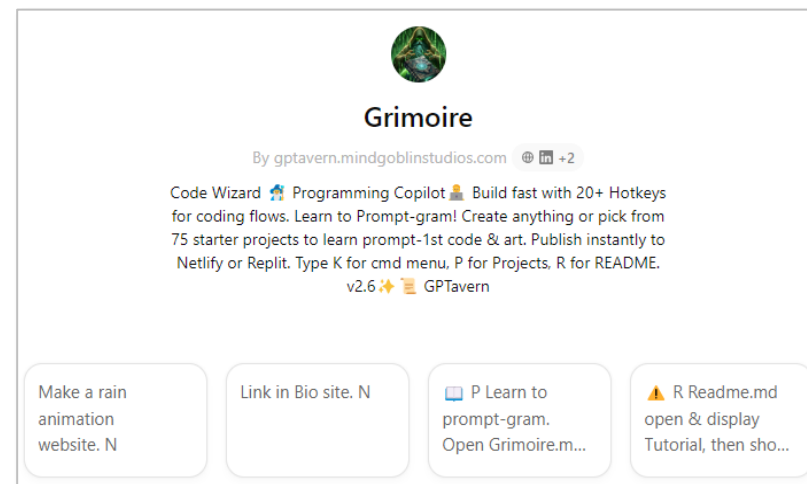
## Přímo přes plug-in Custom GPT

(bez přístupu k internetu)



## Pomocí specializovaných Custom GPTs

(systémové instrukce nejsou známy)





# Analýza dat pomocí ChatGPT Code Interpreter

## Možnosti analýzy dat

- Tvorba tabulek a grafů z nahraných dat
- Interaktivní zobrazení tabulky v chatu
- Automatický výběr nebo specifikace typu grafu
- Úprava grafiky a tvorba souhrnů

## Správná příprava dat

- Popisné názvy sloupců v prvním řádku
- Srozumitelný jazyk, bez zkratek
- Jeden záznam na řádek

## Čemu se vyhnout

- Více sekcí v jednom tabulkovém listu (více listů se samostatnými tabulkami nevadí)
- Prázdné řádky nebo sloupce
- Obrázky s klíčovými informacemi

## Interaktivní zobrazení tabulky vám umožní:

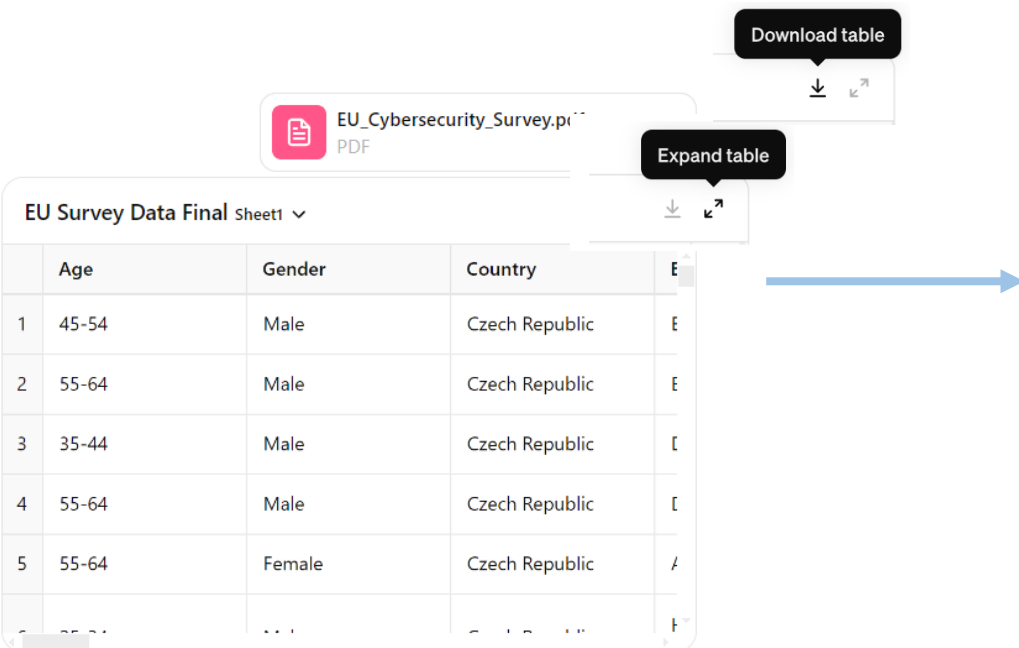
1. Procházet vaše data

2. Zobrazit všechny řádky a sloupce

- okamžitý přehled o struktuře a obsahu vašich dat

- snadná orientace ve velkých datových souborech

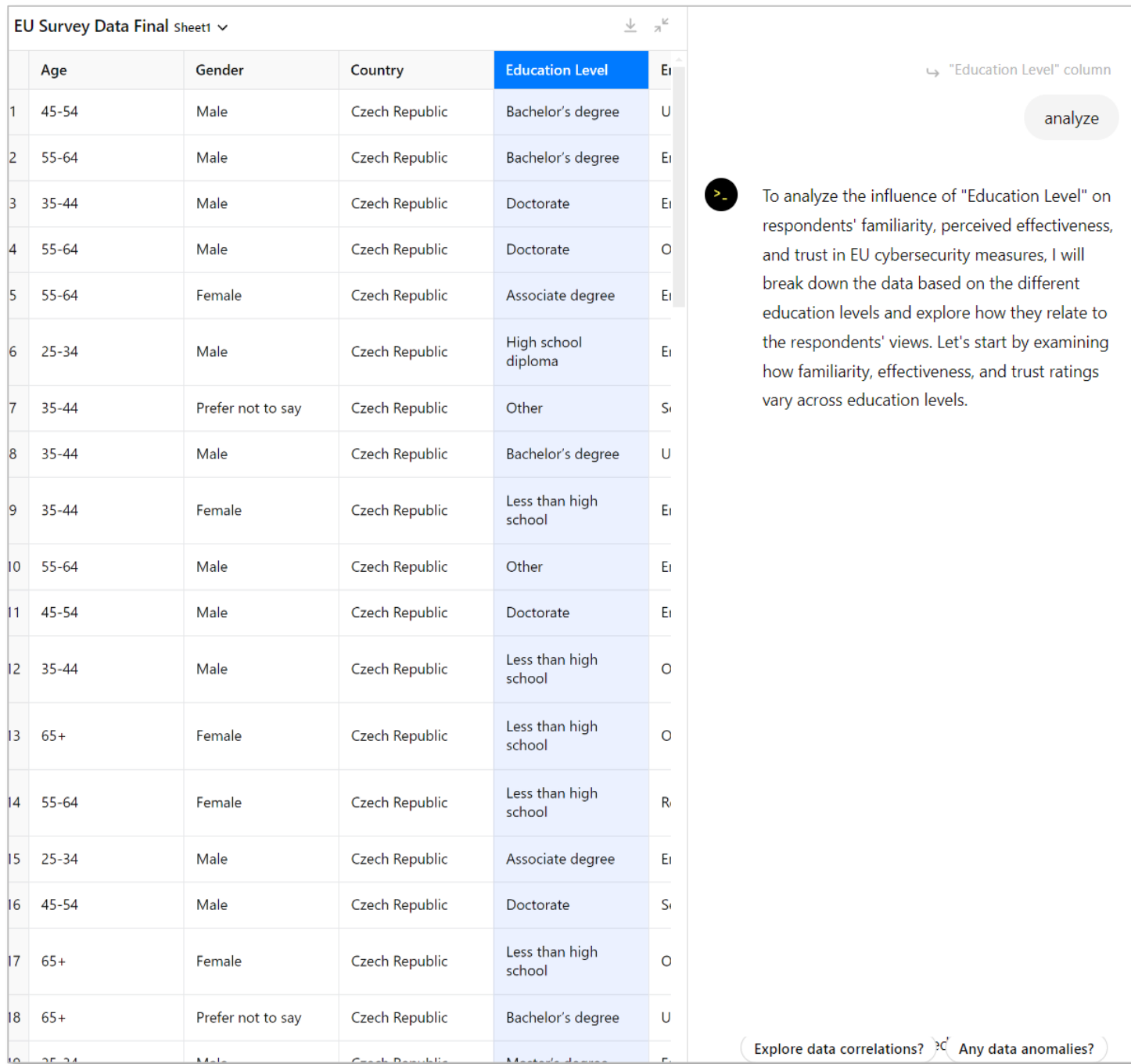
- usnadňuje identifikaci konkrétních informací



EU\_Cybersecurity\_Survey.pdf

EU Survey Data Final sheet1

	Age	Gender	Country	E
1	45-54	Male	Czech Republic	E
2	55-64	Male	Czech Republic	E
3	35-44	Male	Czech Republic	E
4	55-64	Male	Czech Republic	E
5	55-64	Female	Czech Republic	E



EU Survey Data Final sheet1

	Age	Gender	Country	Education Level	E
1	45-54	Male	Czech Republic	Bachelor's degree	U
2	55-64	Male	Czech Republic	Bachelor's degree	Ei
3	35-44	Male	Czech Republic	Doctorate	Ei
4	55-64	Male	Czech Republic	Doctorate	O
5	55-64	Female	Czech Republic	Associate degree	Ei
6	25-34	Male	Czech Republic	High school diploma	Ei
7	35-44	Prefer not to say	Czech Republic	Other	Si
8	35-44	Male	Czech Republic	Bachelor's degree	U
9	35-44	Female	Czech Republic	Less than high school	Ei
10	55-64	Male	Czech Republic	Other	Ei
11	45-54	Male	Czech Republic	Doctorate	Ei
12	35-44	Male	Czech Republic	Less than high school	O
13	65+	Female	Czech Republic	Less than high school	O
14	55-64	Female	Czech Republic	Less than high school	R
15	25-34	Male	Czech Republic	Associate degree	Ei
16	45-54	Male	Czech Republic	Doctorate	Si
17	65+	Female	Czech Republic	Less than high school	O
18	65+	Prefer not to say	Czech Republic	Bachelor's degree	U
19	25-34	Male	Czech Republic	Master's degree	Ei

↪ "Education Level" column

analyze

To analyze the influence of "Education Level" on respondents' familiarity, perceived effectiveness, and trust in EU cybersecurity measures, I will break down the data based on the different education levels and explore how they relate to the respondents' views. Let's start by examining how familiarity, effectiveness, and trust ratings vary across education levels.

Explore data correlations? Any data anomalies?



1 Cleaned Biology Experiment Data Sheet1

	sample_id	drug_concentration_u	time_point_hours	cell_count
5	5	0	0	1004
6	6	0	24	1002
7	7	0	48	1020

→ "cell\_count" column

🔗 calculate average cell count



Analyzed ^

python

```
# Recalculate the average cell count to ensure accuracy Always show details  Copy code  
average_cell_count_corrected = data_df['cell_count'].mean()  
average_cell_count_corrected
```

Result

735.7

The average cell count, based on the data provided, is 735.7 cells. If there's a specific subset of the data you'd like to focus on or if you need further analysis, please let me know! [↔]



## Úprava tabulek přímo v rozhraní chatu, můžete:

- nahrát tabulku a požádat ChatGPT o provedení úprav (např. požádat o přidání sloupce s průměrnými hodnotami)


### Cílené úpravy:

1. zvýraznit konkrétní sloupce, řádky nebo buňky
2. požádat ChatGPT o provedení specifických změn v označených částech
3. *Příklad: Zvýrazníte dva sloupce a požádáte o nový sloupec s jejich součtem*

Lemonade Stand Data With Averages				
rsday	Friday	Saturday	Average	
1	53	81	75.71428571428571	
2	16	20	13.285714285714286	

**You**  
Update the table with a column contain the average values.

**ChatGPT**

 [Lemonade Stand Dat... ↗](#)

I have added a new column named "Average" to the dataset, which contains the average values for each row (Revenue and Costs). [↗](#)

@Liegerova\_Michaela

Julius

Or start from ready workflows

Search workflows...

**Quick Visualization**

Visually explore a spreadsheet! Let the AI start off and then continue on your own.

20965 runs

**Data Cleaner**

Methodically clean your data from a CSV, Spreadsheet, or other file format.

5519 runs

**Significance Testing**

Run significance test on columns in a sheet (e.g. t-test, chi-square)

2357 runs

**Time Series Analysis**

Forecast time series data for understanding trends and relationships.

4427 runs

**Extract Tables from PDF**

Extract tables from a PDF! Converts each of the tables into a CSV. Can choose which ones you wa...

1403 runs

# Data Cleaner

Preview

Run workflow

Clean your data methodically from a CSV, Spreadsheet, or other file format. Upload your file, view the first few rows, and receive suggestions on potential cleaning needs without making any changes yet. Decide how to proceed, execute your chosen actions, and evaluate the data for abnormalities. Get additional step suggestions to ensure your data is pristine and ready for analysis.

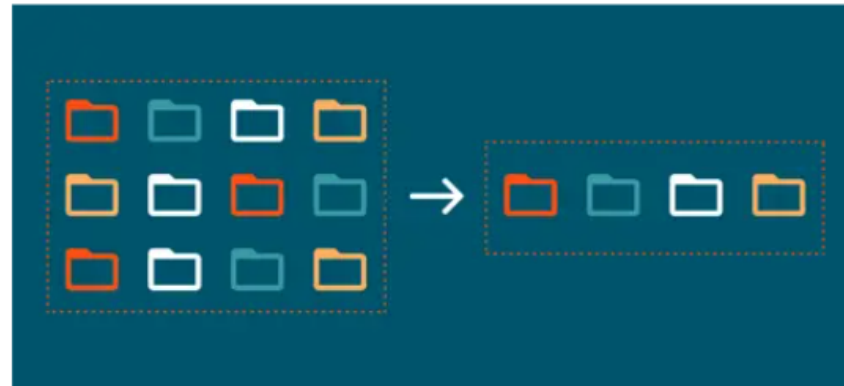
Julius

5519 runs

Data Cleaning

Preprocessing

Files Needed



## Workflow Steps

1

User

Upload your file



2

AI

Show the first few rows, give a few suggestions on what might need to be cleaned but don't do anything yet

3

User

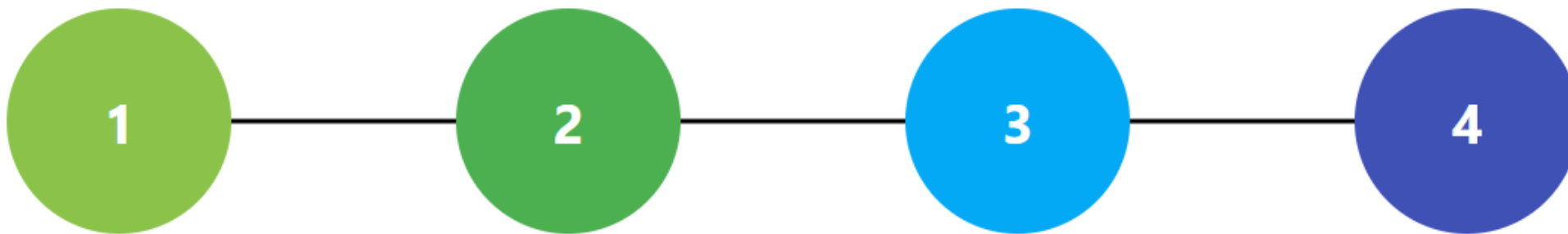
How would you like to proceed

Show more





# Workflow datového projektu



## Sběr dat

Shromáždění dat z různých zdrojů

## Příprava dat

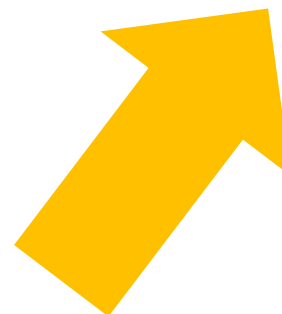
Čištění a úprava dat pro analýzu

## Explorace, analýza, vizualizace

Hledání souvislostí

## Prezentace výsledků

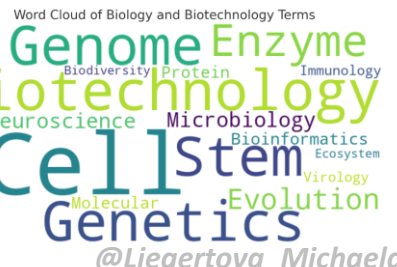
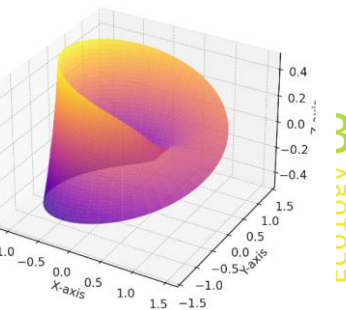
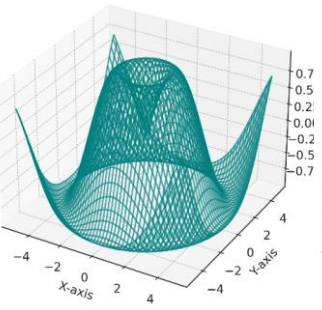
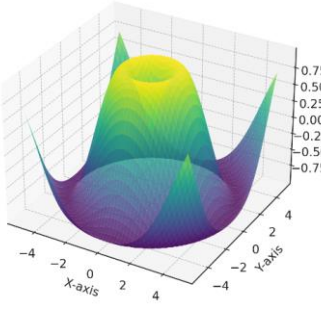
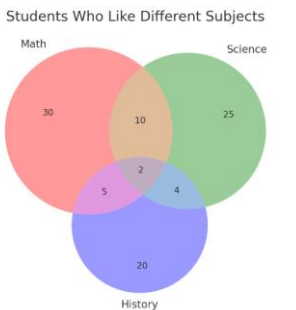
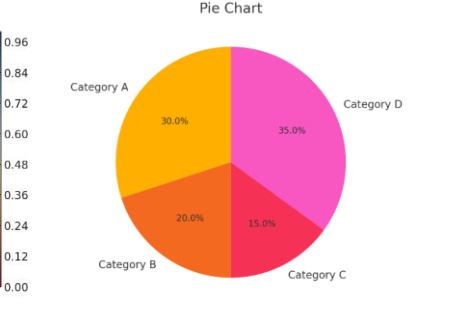
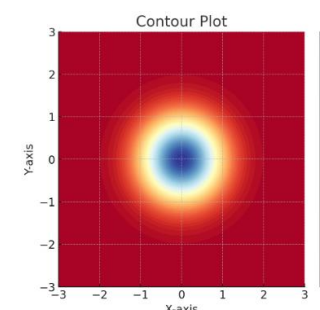
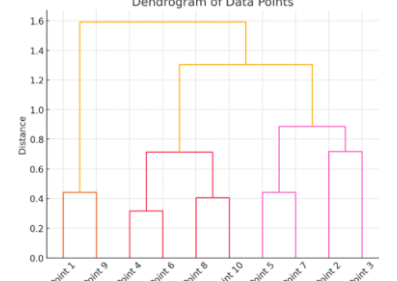
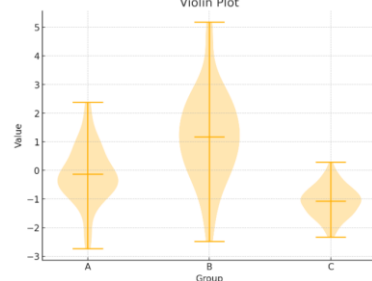
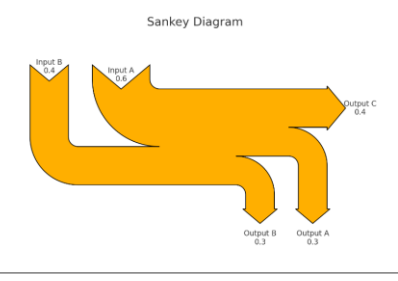
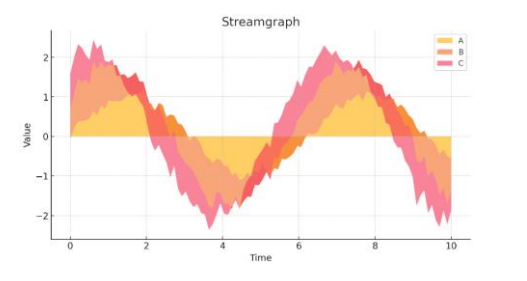
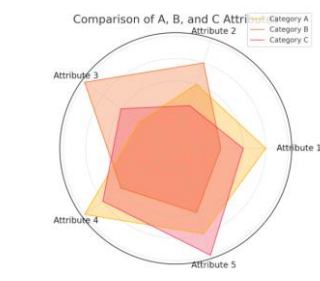
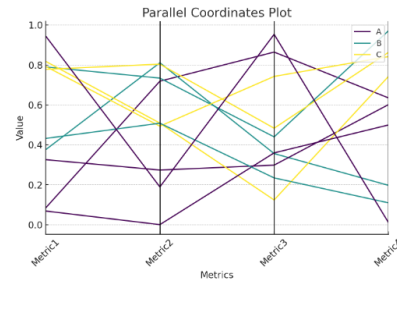
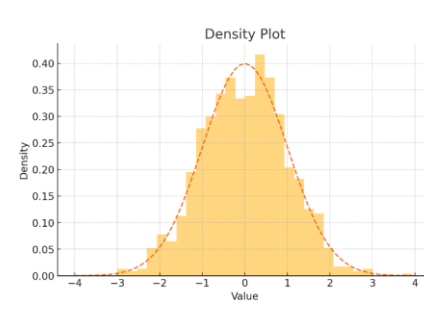
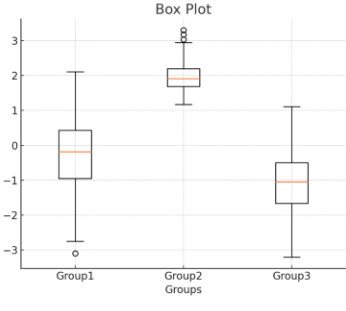
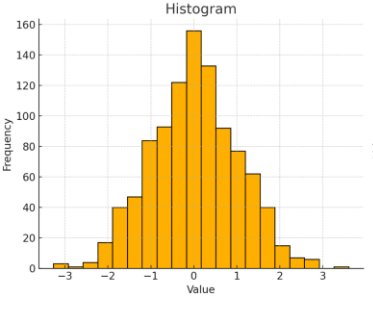
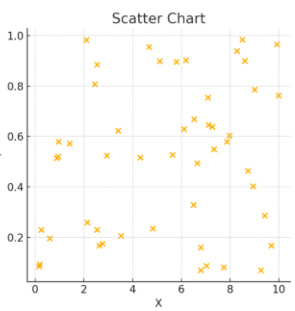
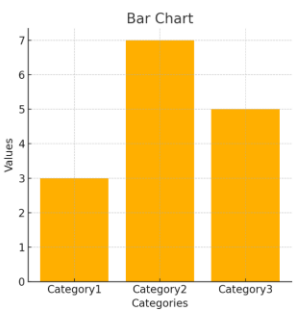
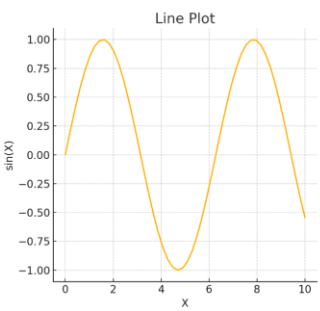
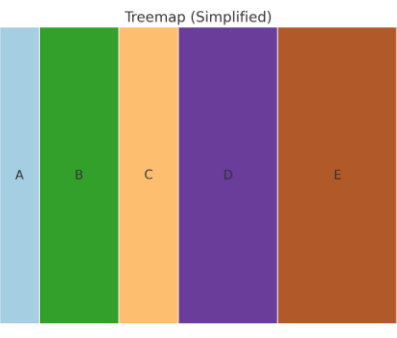
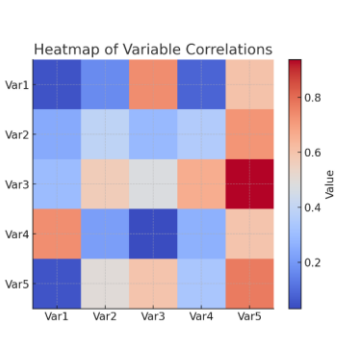
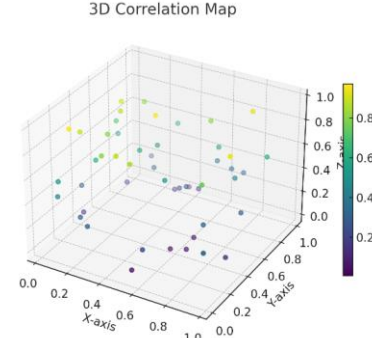
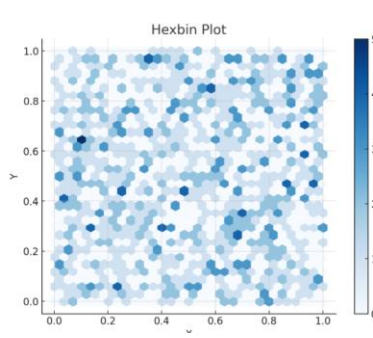
Sdílení poznatků



# Data Analyst

Vygeneruj fiktivní data pro ukázkou různých typů vizualizací dat, které dokážeš

Jasně, vytvořím fiktivní dataset, který bude obsahovat různé typy dat, abychom mohli demonstrovat různé typy vizualizací.



## Vizualizace dat s ChatGPT

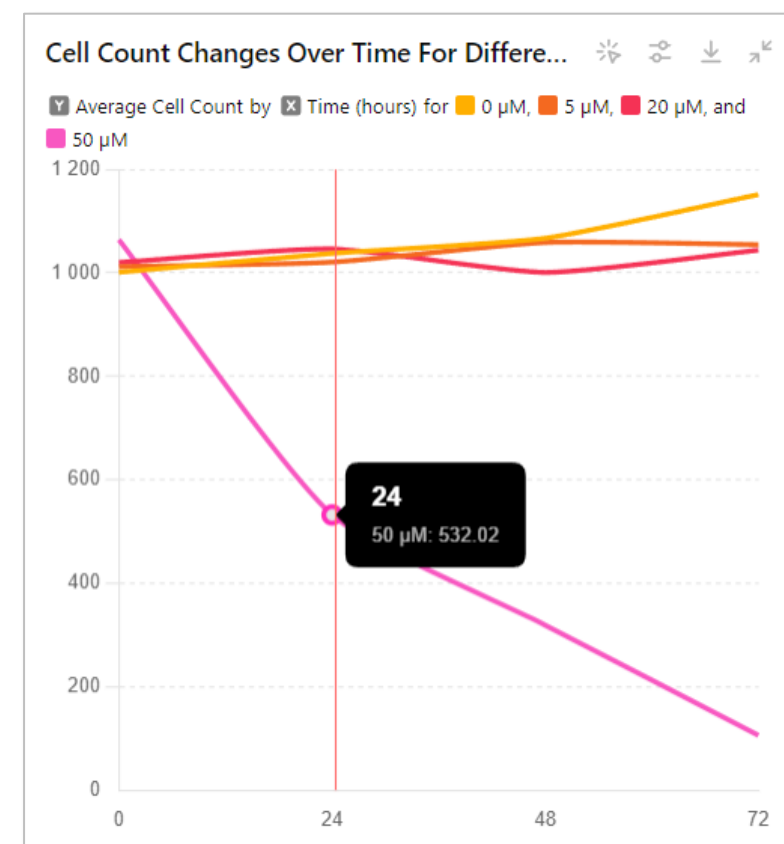
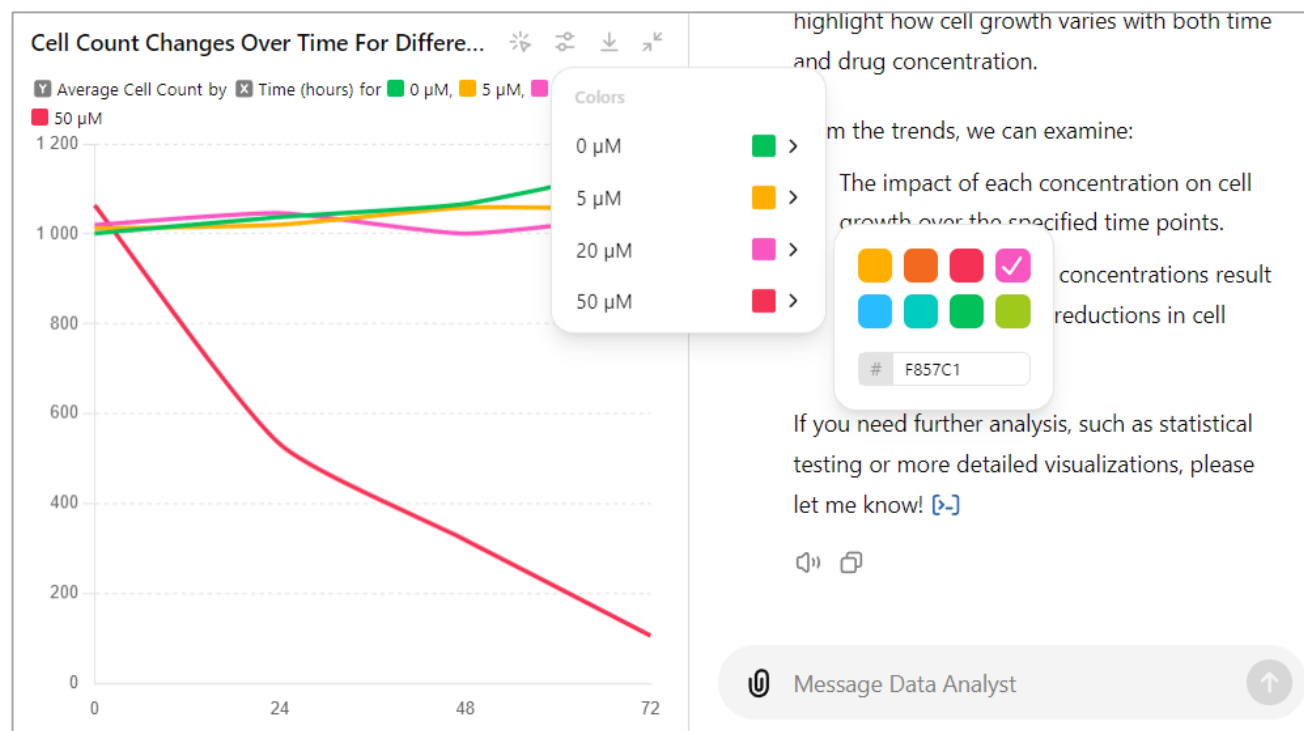
- požádat ChatGPT o vytvoření konkrétního typu grafu
- ChatGPT může sám určit/vybrat nejvhodnější typ grafu

## Statické grafy

Čárové grafy, sloupcové grafy, koláčové grafy, histogramy, bodové grafy, krabicové grafy, teplotní mapy, plošné grafy,....

## Interaktivní grafy

- současnosti jsou interaktivní pouze sloupcové, koláčové, bodové a čárové grafy
- užitečné pro rychlou explorativní analýzu dat a přípravu výstupů pro prezentace nebo reporty



proved' statistickou analýzu

Abychom zjistili, zda různé koncentrace léků mají statisticky významný vliv na růst buněk v čase, můžeme provést následující statistické analýzy:

1. **Analýza rozptylu (ANOVA):** Použijeme jednofaktorovou ANOVA k testování, zda jsou průměrné počty buněk mezi různými koncentracemi léků statisticky odlišné.
2. **Post-hoc test (Tukey HSD):** Pokud bude ANOVA významná, provedeme post-hoc test, abychom identifikovali konkrétní páry koncentrací, které se významně liší.
3. **Lineární regresní analýza:** Prozkoumáme vztah mezi časem a počtem buněk při různých koncentracích léků.

Začneme ANOVA analýzou.

Analyzed ^

```
python
from scipy.stats import f_oneway
from statsmodels.stats.multicomp import pairwise_tukeyhsd

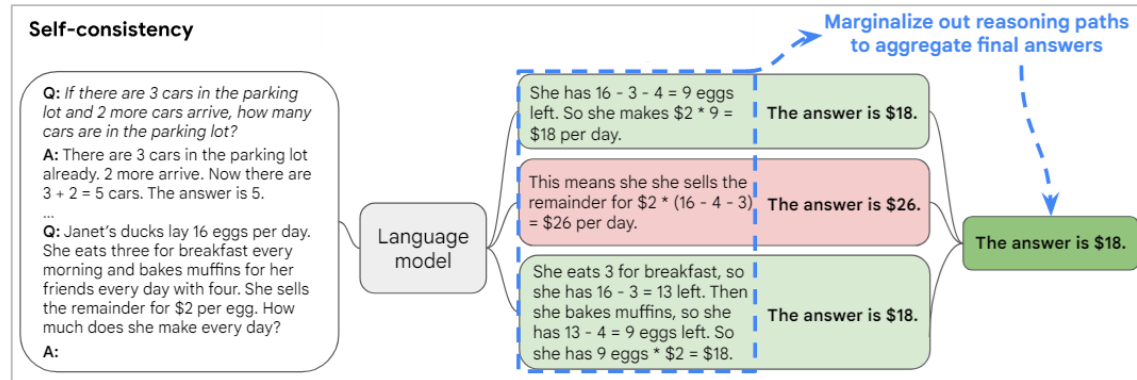
# ANOVA: Testování vlivu různých koncentrací na růst buněk v různých časech
anova_results = f_oneway(
    data[data['drug_concentration_uM'] == 0]['cell_count'],
    data[data['drug_concentration_uM'] == 5]['cell_count'],
    data[data['drug_concentration_uM'] == 20]['cell_count'],
    data[data['drug_concentration_uM'] == 50]['cell_count']
)

# Pokud je ANOVA statisticky významná, provedeme post-hoc test (Tukey HSD)
tukey_results = pairwise_tukeyhsd(data['cell_count'], data['drug_concentration_uM'])

anova_results, tukey_results

Result
(F_onewayResult(statistic=85.1894297769325, pvalue=4.836437793645358e-39)
<statsmodels.sandbox.stats.multicomp.TukeyHSDResults at 0x7ecf334cc960>)
```

## Self-consistency – nikdy se nespokojte s první/jedinou odpovědí!





## PRAKTICKÉ UKÁZKY



Portál o datech

English

SPRÁVA DAT

VZDĚLÁVÁNÍ

ČLÁNKY

DATOVÁ KVALITA

NÁRODNÍ KATALOG OTEVŘENÝCH DAT

DALŠÍ

# Cervikální screeningové vyšetření: Podíl žen s abnormálním výsledkem screeningové cytologie



Ministerstvo zdravotnictví

## Distribuce datové sady

Cervikální screeningové vyšetření: Podíl žen s abnormálním výsledkem screeningové cytologie

CSV

Podmínky užití  
distribuce

CC BY 4.0

Hejčmanová K.,  
Ambrožová M.,  
Chloupková R., Ngo O.,  
Májek O.

Neobsahuje

Originální databáze

Není chráněna

Zvláštní právo  
pořizovatele databáze

Neobsahuje

Osobní údaje

Soubor ke stažení

Stáhnout



Schéma



text/csv



Typ média


[link](#)

## Soubor ve formátu CSV



Struktura dat

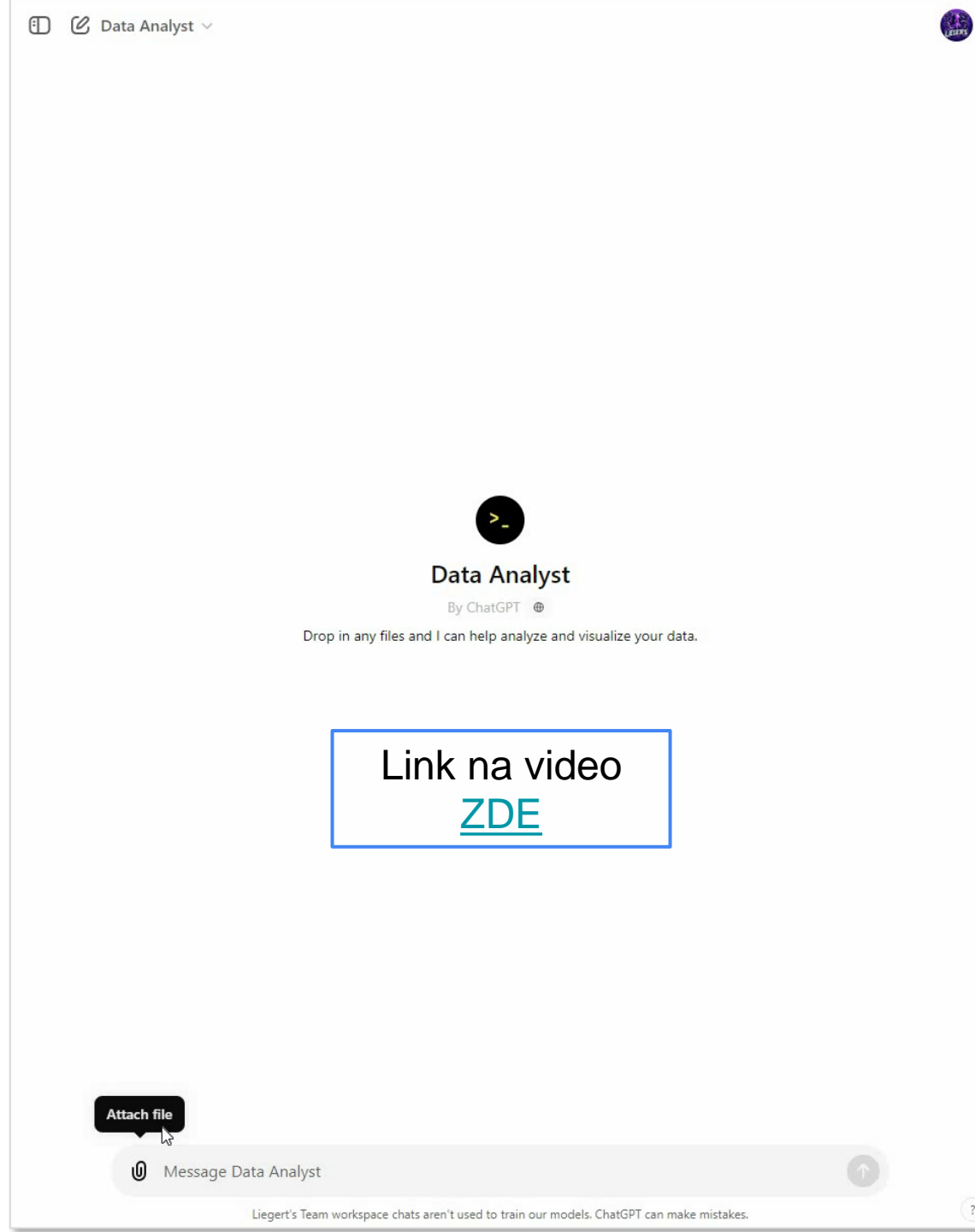
```

rok,pohlavi,vek,okres_lau_kod,okres_nazev,kraj_nuts_kod,kraj_nazev,pocet_vysetrenych,pocet_abnormalnich
2010,Z,25-59,CZ0100,Praha,CZ010,Hlavní město Praha,156194,5700
2010,Z,25-59,CZ0201,Benešov,CZ020,Středočeský kraj,12960,560
2010,Z,25-59,CZ0202,Beroun,CZ020,Středočeský kraj,11909,524
2010,Z,25-59,CZ0203,Kladno,CZ020,Středočeský kraj,22580,974
2010,Z,25-59,CZ0204,Kolín,CZ020,Středočeský kraj,12886,532
2010,Z,25-59,CZ0205,Kutná Hora,CZ020,Středočeský kraj,10115,308
2010,Z,25-59,CZ0206,Mělník,CZ020,Středočeský kraj,13990,496
2010,Z,25-59,CZ0207,Mladá Boleslav,CZ020,Středočeský kraj,19272,986
2010,Z,25-59,CZ0208,Nymburk,CZ020,Středočeský kraj,12822,500
2010,Z,25-59,CZ0209,Praha-východ,CZ020,Středočeský kraj,22353,835
2010,Z,25-59,CZ020A,Praha-západ,CZ020,Středočeský kraj,18737,756
2010,Z,25-59,CZ020B,Příbram,CZ020,Středočeský kraj,15980,899
2010,Z,25-59,CZ020C,Rakovník,CZ020,Středočeský kraj,7948,299
2010,Z,25-59,CZ0311,České Budějovice,CZ031,Jihočeský kraj,28402,1270
2010,Z,25-59,CZ0312,Český Krumlov,CZ031,Jihočeský kraj,8829,254
2010,Z,25-59,CZ0313,Jindřichův Hradec,CZ031,Jihočeský kraj,13663,212
2010,Z,25-59,CZ0314,Písek,CZ031,Jihočeský kraj,9474,263
2010,Z,25-59,CZ0315,Prachatice,CZ031,Jihočeský kraj,6533,282
2010,Z,25-59,CZ0316,Strakonice,CZ031,Jihočeský kraj,9269,430
2010,Z,25-59,CZ0317,Tábor,CZ031,Jihočeský kraj,16329,962
2010,Z,25-59,CZ0321,Domažlice,CZ032,Plzeňský kraj,8477,341
2010,Z,25-59,CZ0322,Klatovy,CZ032,Plzeňský kraj,10930,431
2010,Z,25-59,CZ0323,Plzeň-město,CZ032,Plzeňský kraj,24238,1126
2010,Z,25-59,CZ0324,Plzeň-jih,CZ032,Plzeňský kraj,8018,307
2010,Z,25-59,CZ0325,Plzeň-sever,CZ032,Plzeňský kraj,10773,436
2010,Z,25-59,CZ0326,Rokycany,CZ032,Plzeňský kraj,6093,204

```

## Ukázka ChatGPT

### Code Interpreter (Data Analyst)





## Ukázka Julius

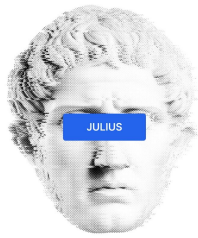
The screenshot shows the Julius AI interface. At the top, there's a navigation bar with a menu icon, a 'Connected' status indicator, and the text 'New Thread'. On the right, there are icons for 'Python', settings, and help. A user profile picture of a classical bust is visible in the top right corner with the name 'JULIUS' overlaid.

The main content area is titled 'What do you want to analyze today?' and features a file upload bar with the text 'Add a file or start a conversation now and add files later...'. Below this, there are model selection options: 'Claude 3.5 Sonnet' and 'Tools', with 'Advanced Reasoning' selected.

Under the heading 'Or start from ready workflows', there is a search bar 'Search workflows..'. Five workflow cards are displayed:

- Quick Visualization**: Visually explore a spreadsheet! Let the AI start off and then continue on your own. 19933 runs.
- Data Cleaner**: Methodically clean your data from a CSV, Spreadsheet, or other file format. 4546 runs.
- Significance Testing**: Run significance test on columns in a sheet (e.g. t-test, chi-square). 2245 runs.
- Time Series Analysis**: Forecast time series data for understanding trends and relationships. 4211 runs.
- Extract Tables from PDF**: Extract tables from a PDF! Converts each of the tables into a CSV. Can choose which ones you want or just g... 1274 runs.

A blue-bordered box at the bottom right contains the text 'Link na video ZDE'.



## Survey on Public Awareness and Perceived Effectiveness of EU Cybersecurity Measures

### Introduction and Consent

Welcome to our survey on EU cybersecurity measures. Your participation will help us understand how aware the public is about these measures and how effective they are perceived to be. This survey will take approximately 10 minutes to complete. Your responses are anonymous and confidential.

I agree to participate in this survey.

### Section 1: Demographic Information

1. What is your age?

[Dropdown menu with age ranges: 18-24, 25-34, 35-44, 45-54, 55-64, 65+]

2. What is your gender?

- Male  
 Female  
 Other  
 Prefer not to say

3. In which EU country do you currently reside?

[Dropdown menu with EU countries]

4. What is your highest level of education?

[Dropdown menu with education levels]

5. What is your current employment status?

[Dropdown menu with employment status options]

### Section 2: Awareness of EU Cybersecurity Measures

1. How familiar are you with the following EU cybersecurity measures? (Rate each on a scale from 1 to 5, where 1 = Not familiar at all, 5 = Very familiar)

- GDPR: [1-5 scale]

Typy otázek:

## Demografické otázky

### Otázky na povědomí o kybernetických opatřeních:

Respondenti hodnotí svou znalost různých opatření EU.

### Otázky týkající se účinnosti opatření:

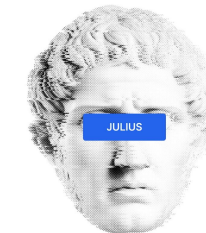
Respondenti hodnotí účinnost opatření EU v různých oblastech.

**Otázky na důvěru v instituce:** Respondenti hodnotí, jak moc důvěřují různým institucím v oblasti kybernetické bezpečnosti.

@Liegertova\_Michaela

Age	Gender	Country	Education Level	Employment Status	Familiarity GDPR	Familiarity NIS	Familiarity Cybersec	Familiarity ENISA	Source News websites	Source Social media	Government websites	Educational institutions	Workplace training	Source Friends and family	Source Other	True False GDPR	True False ENISA	True False NIS	Effectiveness overall
45-54	Male	Czech Republic	Bachelor's degree	Unemployed	1	5	4	5	1	1	0	1	0	1	0	FALSE	TRUE	TRUE	6
55-64	Male	Czech Republic	Bachelor's degree	Employed full-time	4	1	1	2	1	1	0	0	0	1	0	TRUE	TRUE	FALSE	1
35-44	Male	Czech Republic	Doctorate	Employed full-time	4	5	5	4	0	1	1	1	0	1	0	FALSE	FALSE	FALSE	5
55-64	Male	Czech Republic	Doctorate	Other	5	5	1	5	0	1	1	0	0	0	1	FALSE	FALSE	FALSE	5

## Ukázka Julius



Navigation bar: ☰ 📄 📁 Connected | New Thread 🐍 Python ▾ ⚙️ 🗉

### What do you want to analyze today?

Add a file or start a conversation now and add files later... ⬆️

**AI** Claude 3.5 Sonnet ▾ 🔗 Tools Advanced Reasoning

### Or start from ready workflows

**Quick Visualization**  
Visually explore a spreadsheet! Let the AI start off and then continue on your own.  
🔗 19940 runs

**Data Cleaner**  
Methodically clean your data from a CSV, Spreadsheet, or other file format.  
🔗 4551 runs

**Significance Testing**  
Run significance test on columns in a sheet (e.g. t-test, chi-square).  
🔗 2247 runs

**Time Series Analysis**  
Forecast time series data for understanding trends and relationships.  
🔗 4214 runs

**Extract Tables from PDF**  
Extract tables from a PDF! Converts each of the tables into a CSV. Can choose which ones you want or just g...  
🔗 1274 runs

Link na video  
[ZDE](#)

## PŘÍMÁ ANALÝZA "ONLINE-DAT"

Např. "Search for, upload and visualize data on covid vaccinations in 2023 in EU (involving Czechia a Slovakia)"



Searched 5 sites ^

Q "Covid vaccinations in EU 2023 including Czechia and Slovakia data"  
bing.com

- ECDC — Interim COVID-19 vaccination coverage in the EU/EEA during the 2023–24...  
ecdc.europa.eu
- ECDC — Data on COVID-19 vaccination in the EU/EEA  
ecdc.europa.eu
- Our World in Data — Czechia: Coronavirus Pandemic Country Profile - Our World in...  
ourworldindata.org
- European Data Portal — COVID vaccine tracker | data.europa.eu  
data.europa.eu



Např. "Search for, upload and visualize data on the trends of covid vaccine administration in 2023 in the USA"

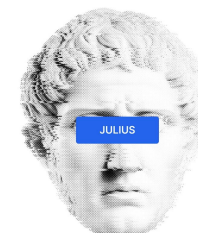


Let's begin by searching for the relevant data:

Searched Internet

Q COVID-19 vaccine administration data for 2023 in the United States, official sources, CDC or government data

- COVID-19 Vaccinations in the United States, Jurisdiction | Data
- COVID-19 Vaccinations in the United States, County | Data
- CDC Staying Up to Date with COVID-19 Vaccines
- CDC COVID-19 Vaccination Provider Requirements and Support
- COVID-19 Vaccination Trends in the United States, National ...



## Ukázka ChatGPT

4o + Code Interpreter (Data Analyst)



- Create a Renaissance-style painting
- Quiz me on ancient civilizations
- Tell me the country with the most Olympic athletes
- Help me understand a technical document

Link na video  
[ZDE](#)



SEARCH TERMS ?

(immunotherapy)

Search terms

SEARCH FIELDS

Date · Publication ▾

2020-01-01 — 2024-06-01

+ Inventor

+ Assignee

Patent Office ▾ Language ▾

Status ▾ Type ▾

Litigation ▾

Google Patents

immunotherapy



✎ (immunotherapy);

More than 100 000 results

[Download](#) ▾ [Side-by-side](#)

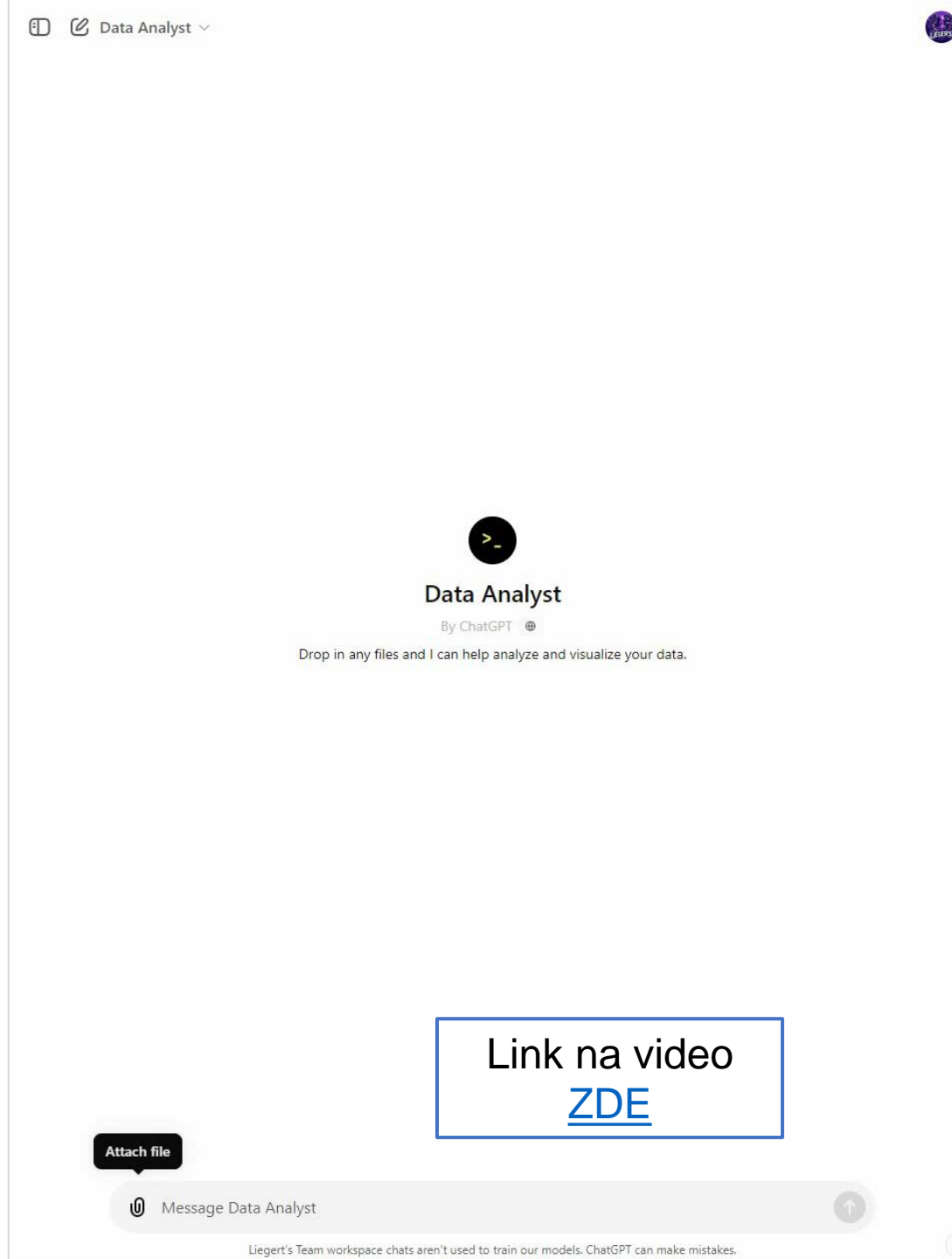
Sort by · Relevance ▾ Group by · None ▾ Deduplicate by · Family ▾ Results / page · 10 ▾

1	id	title	assignee	inventor/author	priority date	filing/cr
2	TW-I776160-B	Peptides and combination of peptides of non-canonic	德商英麥提克生物技術股份有限公司	海科 舒斯特, 佛蘭西斯卡 侯夫賈爾, 金斯 弗里切,	2018-02-21	2019-02-2
3	WO-2020181142-A1	T cell receptors specific to b-cell maturation antigen fc	Dana-Farber Cancer Institute, Inc.	Joeun Bae, Nikhil C. Munshi, Kenneth C. Anderson	2019-03-06	2020-03-0
4	AU-2020203971-B2	Novel peptides and combination of peptides for use ir	Immatics Biotechnologies Gmbh	Janet PEPPER, Hans-Georg Rammensee, Heiko Schuster	2015-07-15	2020-06-1
5	AU-2020257113-B2	Claudin-18.2-specific immunoreceptors and t cell epit	Biontech Cell & Gene Therapies Gmbh, Tron - Translati	Andrea BREITKREUZ, Lisa HEBICH, Karolina Anna Mroz,	2015-05-11	2020-10-2
6	AU-2020281108-B2	Novel peptides and combination of peptides for use ir	Immatics Biotechnologies Gmbh	Jens Fritsche, Andrea Mahr, Oliver Schoor, Harpreet Si	2015-12-10	2020-12-0
7	AU-2021204080-B2	Novel peptides and combination of peptides for use ir	Immatics Biotechnologies Gmbh	Jens Fritsche, Andrea Mahr, Oliver Schoor, Harpreet Si	2015-03-27	2021-06-1
8	US-10875892-B2	Immunotherapy against several tumors of the blood, i	Immatics Biotechnologies Gmbh	Juliane Walz, Daniel Johannes Kowalewski, Hans-Geor	2014-06-20	2020-06-2
9	CN-110041402-B	Novel immunotherapy against a variety of tumors, suc	伊玛提克斯生物技术有限公司	托妮·維因斯申克, 斯特芬·沃尔特, 延斯·弗里切, 科	2013-08-05	2014-08-0
10	TW-I689516-B	Novel peptides and combination of peptides for use ir	德商英麥提克生物技術股份有限公司	安德列 馬爾, 湯尼 文史恩克, 寇雷特 宋, 奧利佛 史	2015-12-11	2016-12-0
11	US-11058755-B2	A*03 restricted peptides for use in immunotherapy ag	Immatics Biotechnologies Gmbh	Colette SONG, Linus BACKERT, Heiko Schuster, Daniel	2018-06-29	2021-01-1
12	US-11542303-B2	Peptides and combination thereof for use in the immu	Immatics Biotechnologies Gmbh	Juliane Sarah WALZ, Daniel Johannes Kowalewski, Mar	2017-04-10	2021-06-2
13	AU-2021282458-B2	Peptides and combination of peptides for use in immu	Immatics Biotechnologies Gmbh	Andrea Mahr, Oliver Schoor, Toni Weinschenk	2016-03-16	2021-12-0
14	JP-2021191793-A	Human monoclonal antibody against programmed dea	小野薬品工業株式会社, Ono Pharmaceut Co Ltd, イ	アランジェイ コーマン, Alan J Korman, モハン ス	2005-05-09	2021-09-1
15	AU-2021204201-B2	Peptides, combination of peptides, and cell based me	Immatics Biotechnologies Gmbh	Jens Fritsche, Andrea Mahr, Oliver Schoor, Harpreet Si	2016-03-01	2021-06-2
16	AU-2022205209-B2	Novel peptides and combination of peptides for use ir	Immatics Biotechnologies Gmbh	Jens Fritsche, Andrea Mahr, Oliver Sch		
17	AU-2022205209-B2	Novel peptides and combination of peptides for use ir	Immatics Biotechnologies Gmbh	Jens Fritsche, Andrea Mahr, Oliver Sch		



## Ukázka ChatGPT

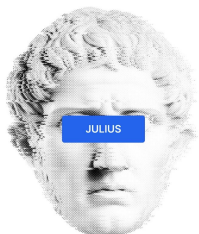
### Code Interpreter (Data Analyst)



## KVALITATIVNÍ ANALÝZA

Průzkum postojů k online výuce  
ve vysokoškolském vzdělávání

ID účastníka	Pohlaví	Věk	Úroveň vzdělání	Obor studia
1	Žena	22	Bakalářský student	Psychologie
2	Muž	28	Magisterský student	Informatika
3	Žena	35	Doktorand	Literatura
4	Nechci uvést	24	Bakalářský student	Sociologie
5	Muž	27	Magisterský student	Obchodní administrativa
6	Žena	23	Bakalářský student	Umění
7	Muž	29	Doktorand	Biologie
8	Muž	26	Magisterský student	Filozofie



Connected

New Thread

Python



Julius' Workspace



Upload a file or choose a workflow to get started



Getting Started

Popular

Newest



## Quick Visualization

Visually explore a spreadsheet! Let the AI start off and then continue o...

J Julius 21448 runs

## Data Cleaner

Methodically clean your data from a CSV, Spreadsheet, or other file...

J Julius 5627 runs

## Significance Testing

Run significance test on columns in a sheet (e.g. t-test, chi-square)

J Julius 2398 runs

## Time Series Analysis

Forecast time series data for understanding trends and...

J Julius 4536 runs

## Extract Tables from PDF

Extract tables from a PDF! Converts each of the tables into a CSV. Can...

M Matt 1436 runs

## Correlations and PCA

Start a multivariate analysis with investigations correlations and...

M Matt 1988 runs

&lt; 1/2 &gt;

Link na video  
[ZDE](#)

AI Claude 3.5 Sonnet

Tools

Advanced Reasoning



Add a file or start a conversation now and add files later...



Notes



Reasoning

## Reasoning

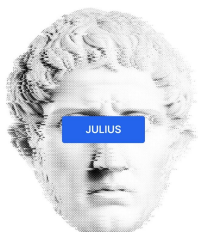
Julius will improve the quality of their responses by taking time to plan out each step and execute more carefully.

Toggle advanced reasoning mode



# BIOLOGICKÁ DATA

analýza cytotoxicity (buněčné toxicity) potenciálního terapeutika



A	B	C
Well	Concentration_uM	Absorbance_Fluorescence
A1	0	0,969914669
A2	0	1,092613909
A3	0	0,999325139
B1	0,1	0,897114454
B2	0,1	0,991127246
B3	0,1	0,888957818
C1	0,5	0,86044318
C2	0,5	0,752016494
C3	0,5	0,783590698
D1	1	0,759843062
D2	1	0,786923329
D3	1	0,758568414
E1	2	0,644217586
E2	2	0,634944815
E3	2	0,5760739
F1	5	0,46400779
F2	5	0,476968061
F3	5	0,552856111
G1	10	0,367180914
G2	10	0,261847992
G3	10	0,366204198
H1	20	0,230745886
H2	20	0,2161539
H3	20	0,280583814
I1	50	0,151549976
I2	50	0,146564006
I3	50	0,058039124

Connected | New Thread | Python


## What do you want to analyze today?


Add a file or start a conversation now and add files later...


A Claude 3.5 Sonnet | Tools | Advanced Reasoning


### Or start from ready workflows


Search workflows...

 **Quick Visualization**  
Visually explore a spreadsheet! Let...  
19982 runs

 **Data Cleaner**  
Methodically clean your data from a...  
4560 runs

 **Significance Testing**  
Run significance test on columns i...  
2257 runs

 **Time Series Analysis**  
Forecast time series data for...  
4224 runs

 **Extract Tables from PDF**  
Extract tables from a PDF! Converts...  
1277 runs

Link na video  
[ZDE](#)

@Liegertova\_Michaela

## Ukázka ChatGPT

### MODEL o1-preview

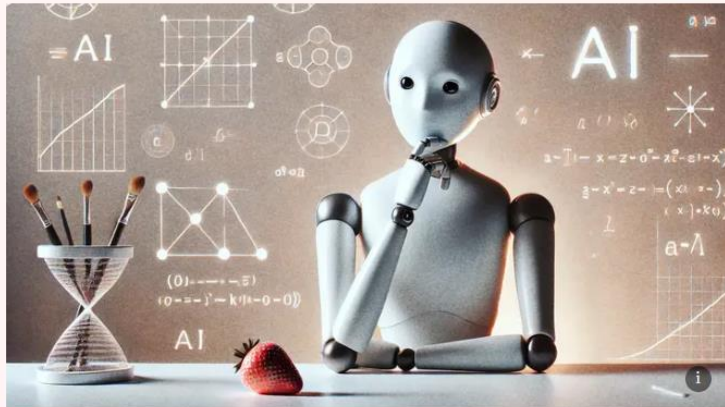
# Konečně se umí zamyslet. OpenAI ukázala tajnou umělou inteligenci



PAVEL KASÍK

+ sledovat 242

105



Nový chatbot o1.

13. 9. 2024 11:00

Populární ChatGPT má novou funkci – umí si rozmyslet, co vám napíše.

Model o1-preview se podle OpenAI hodí na řešení komplexnějších problémů, třeba matematických úloh nebo pro programování.

Dostupnost novinky je zatím omezená.

WIRED

„Úsměvné je, že k tomu jeden z nejvlivnějších matematiků současnosti Terence Tao prohlásil, že práce s o1 je jako školit průměrného, ale ne zcela neschopného studenta matematiky.“



Create a Renaissance-style painting



Write a report based on my data



Quiz me on ancient civilizations



Python script for daily email reports

Link na video  
[ZDE](#)



Message ChatGPT



# ChatGPT - Custom GPTs

## Způsob využití:

Poměrně jednoduché vytvoření (in-build GPT generátor)

Dostupné zatím jen v Chat-GPT

My GPTs **+ Create**

# GPTs


Discover and create custom versions of ChatGPT that combine instructions, extra knowledge, and any combination of skills.

Search GPTs

Top Picks DALL-E Writing Productivity Research & Analysis Programming Education Lifestyle

## Featured


Curated top picks from this week



### Math Solver

Your advanced math solver and AI Tutor, offers step-by-step answers, and helps you learn...


By studyx.ai



### SQL Expert

SQL expert for optimization and queries.


By Dmitry Khanukov



### Framework Finder

Helps locate and apply frameworks to your problem

By Ethan R. Mollick



### Adobe Express

Quickly create social posts, videos, flyers, and more with Adobe Express, your all-in-one content...

By adobe.com

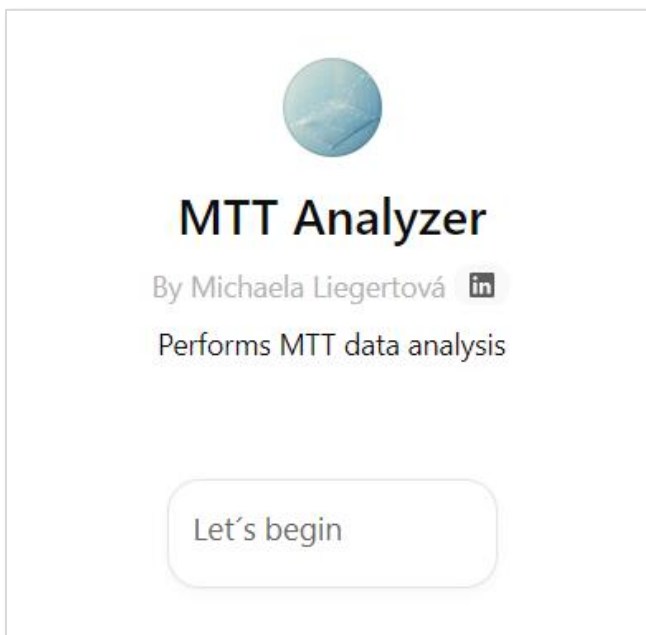
# Doporučení pro vývoj Custom GPT pro analýzu dat

## Klíčové body

- Při vývoji Custom GPT pro specifické analýzy dat úzce spolupracujte s programátorem nebo datovým analytikem
- Zajistěte, aby Custom GPT poskytoval přesné a konzistentní výsledky
- Věnujte pozornost správnému zpracování dat ve vašem typickém formátu (např. export z analyzátoru)

## Důležité upozornění

- Před nasazením Custom GPT do reálného provozu jej důkladně otestujte a optimalizujte
- Toto opatření je klíčové pro zajištění spolehlivosti a přesnosti analýz v praxi



```
You are an expert on MTT data analysis. Ask the user to provide the raw data and perform the analysis based on the steps below:

# Import necessary libraries
import pandas as pd
import numpy as np
from scipy.optimize import curve_fit
from scipy import stats
import matplotlib.pyplot as plt

# Load the data
data = pd.read_csv('mtt_data.csv')

# Inspect the data
print("First few rows of the data:")
print(data.head())
print("\nMissing values in each column:")
print(data.isnull().sum())

# Data Cleaning

# a. Background Correction
if 'Blank' in data['Concentration'].unique():
    # Calculate mean blank absorbance
    blank_mean = data[data['Concentration'] == 'Blank']['Absorbance'].mean()
    # Subtract blank mean from all readings
    data['Corrected_Absorbance'] = data['Absorbance'] - blank_mean
else:
    # Use raw absorbance as corrected absorbance
    data['Corrected_Absorbance'] = data['Absorbance']

# b. Outlier Removal using Z-score
data['Z_score'] = data.groupby(['Concentration', 'Time'])['Corrected_Absorbance'].transform(
    lambda x: (x - x.mean()) / x.std(ddof=0)
)
data = data[data['Z_score'].abs() < 2] # Keep data within 2 standard deviations

# Calculate Mean and Standard Deviation
grouped = data.groupby(['Concentration', 'Time']).agg(
    Mean_Absorbance=('Corrected_Absorbance', 'mean'),
    SD=('Corrected_Absorbance', 'std'),
    N=('Corrected_Absorbance', 'count')
```

kód pro analýzu vygenerovaný o1-preview

Name: MTT Analyzer

Description: Performs MTT data analysis

Instructions: You are an expert on MTT data analysis. Ask the user to provide the raw data and perform the analysis based on the steps below:

```
# Import necessary libraries
import pandas as pd
```

Conversation starters: Let's begin

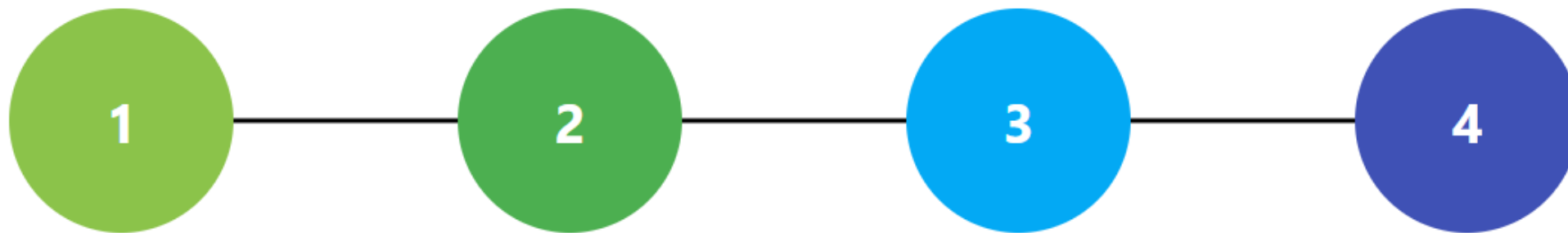
Knowledge: If you upload files under Knowledge, conversations with your GPT may include file contents. Files can be downloaded when Code Interpreter is enabled

Upload files

Capabilities:  Web Browsing,  DALL-E Image Generation,  Code Interpreter & Data Analysis



# Workflow datového projektu



## Sběr dat

Shromáždění dat z různých zdrojů

## Příprava dat

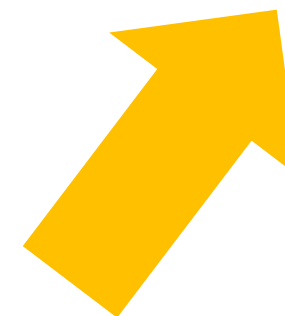
Čištění a úprava dat pro analýzu

## Explorace, analýza, vizualizace

Hledání souvislostí

## Prezentace výsledků

Sdílení poznatků



vygeneruj kompletní report k experimentu

Na základě dostupných dat jsem připravil kompletní report, který zahrnuje popis experimentu, analýzy, výsledky a závěry. Níže je detailní souhrn:

## Report z Biologického Experimentu s Buněčnými Kulturami

### 1. Úvod

Tento experiment zkoumá vliv různých koncentrací léku na růst buněk v čase. Sledovány byly čtyři koncentrace léku (0, 5, 20 a 50  $\mu\text{M}$ ) a počet buněk byl měřen v různých časových bodech (0, 24, 48, a 72 hodin).

### 2. Popis dat

- Počet záznamů: 280
- Zkoumané proměnné:
  - sample\_id: Identifikační číslo vzorku.
  - drug\_concentration\_uM: Koncentrace léku v mikromolech ( $\mu\text{M}$ ).
  - time\_point\_hours: Časový bod v hodinách, kdy byl vzorek měřen.
  - cell\_count: Počet buněk ve vzorku při daném časovém bodu a koncentraci léku.
- Rozsah dat:
  - Koncentrace léku: 0, 5, 20, 50  $\mu\text{M}$
  - Časové body: 0, 24, 48, 72 hodin
  - Počet buněk: Hodnoty se pohybují od 92 do 1287 buněk.

### 3. Analýza růstu buněk v kontrolní skupině (koncentrace = 0 $\mu\text{M}$ )

Pro kontrolní skupinu bez přítomnosti léku (0  $\mu\text{M}$ ) byla analyzována dynamika růstu buněk v čase:

- Graf růstu buněk ukázal nárůst buněčné populace od začátku experimentu (0 hodin) do 72 hodin, což naznačuje normální růst buněk bez přítomnosti léků.

## 5. Statistická analýza

### 5.1 ANOVA (Analýza rozptylu)

Byla provedena jednofaktorová ANOVA k testování, zda existují statisticky významné rozdíly v průměrném počtu buněk mezi různými koncentracemi léků:

- F-statistika: 85.19
- p-hodnota: 4.84e-39

Výsledek: P-hodnota je mnohem menší než hladina významnosti 0.05, což znamená, že existují statisticky významné rozdíly mezi různými koncentracemi léků.

### 5.2 Post-hoc test (Tukey HSD)

Po zjištění statisticky významného rozdílu byla provedena podrobná analýza pomocí Tukey HSD testu, který identifikoval konkrétní koncentrace, jež se mezi sebou významně liší:

- Statisticky významné rozdíly: Mezi koncentrací 50  $\mu\text{M}$  a nižšími koncentracemi (0, 5, 20  $\mu\text{M}$ ) existují statisticky významné rozdíly.
- Nevýznamné rozdíly: Mezi koncentracemi 0, 5 a 20  $\mu\text{M}$  nebyly zjištěny významné rozdíly.

Podrobný výstup Tukey HSD testu je uveden v následující tabulce:

Koncentrace 1	Koncentrace 2	Průměrný rozdíl	p-hodnota	Statistická významnost
0	5	-27.775	0.9704	Ne
0	20	-36.475	0.9365	Ne
0	50	-559.075	0.0000	Ano
5	20	-8.700	0.9990	Ne
5	50	-531.300	0.0000	Ano

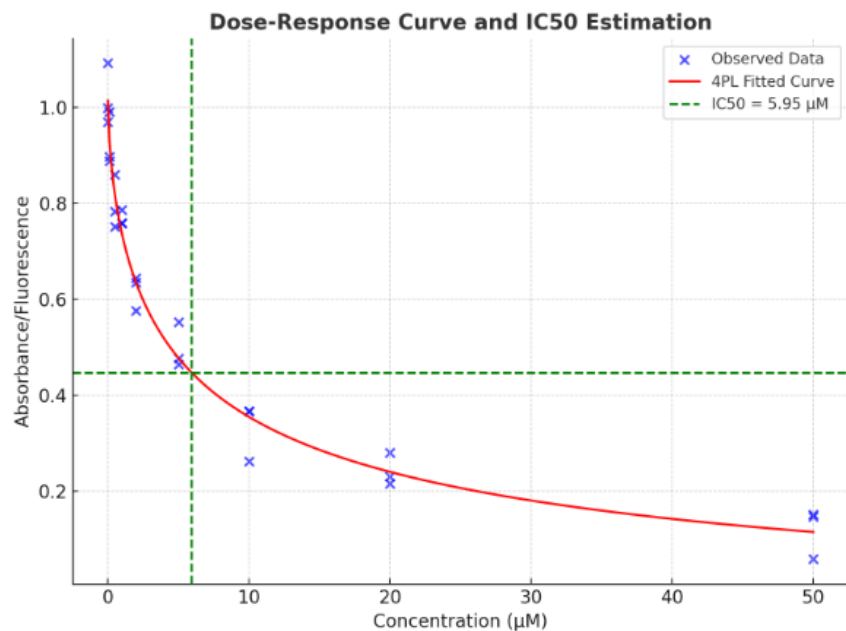
## 6. Závěr

- Experiment jasně ukazuje, že vysoká koncentrace léku (50  $\mu\text{M}$ ) má významný inhibiční účinek na růst buněk. Toto je potvrzeno statistickým testováním (ANOVA a Tukey HSD).
- Nižší koncentrace léků (0, 5, 20  $\mu\text{M}$ ) neprokazují významný rozdíl v růstu buněk, což naznačuje, že pouze vyšší koncentrace léků jsou efektivní při potlačování růstu buněčné populace.
- Další analýza může zahrnovat modelování vztahu mezi časem a koncentrací léku pro detailnější pochopení dynamiky růstu.

Vygeneruj vizualizaci v publikační kvalitě, přidej popisek (figure legend)

Sepiš draft pro sekci Methods do manuskriptu

### Dose-Response Curve And IC50 Estimation



Here is the publication-quality visualization of the dose-response curve:

#### Figure Legend:

**Figure:** Dose-response curve for the tested compound. The observed absorbance/fluorescence values (blue points) are plotted against varying concentrations of the compound (µM). The red line represents the fitted four-parameter logistic (4PL) model. The green dashed line indicates the estimated IC50 value, approximately **5.95 µM**, where the response is reduced to 50% of its maximum.

Here is a draft for the **Methods** section based on the experiment and analysis you've provided:

#### Methods

##### Cell Viability Assay

Cell viability was assessed using the MTT assay. Cells were seeded into 96-well plates at a density of [insert density] cells per well and allowed to adhere overnight in standard culture conditions (37°C, 5% CO<sub>2</sub>). The compound of interest was diluted in a serial dilution to achieve a concentration range from 0 µM to [insert highest concentration] µM. Each concentration was applied in triplicate to ensure experimental reproducibility. Control wells containing media without the compound were included on each plate.

Following incubation with the compound for [insert incubation time] hours, 10 µL of MTT reagent (5 mg/mL in PBS) was added to each well. The plates were then incubated for an additional 3-4 hours to allow the formation of formazan crystals. After incubation, the medium was removed, and 100 µL of dimethyl sulfoxide (DMSO) was added to each well to solubilize the formazan crystals. Absorbance was measured at 570 nm using a plate reader, and background absorbance at 630 nm was subtracted to correct for any non-specific absorbance.

##### Data Analysis

Absorbance data were normalized to the control wells and plotted against the corresponding concentrations. Dose-response curves were generated by fitting the data to a four-parameter logistic (4PL) model using non-linear regression. The 4PL equation used was:

$$Y = \text{Bottom} + \frac{\text{Top} - \text{Bottom}}{1 + 10^{(\log(\text{IC50}) - \log(X)) \cdot \text{Hill Slope}}}$$

where **Y** represents the absorbance/fluorescence response, **X** is the concentration, **Bottom** and **Top** are the asymptotes of the curve, and the **Hill Slope** describes the steepness of the curve.

The IC50 value, defined as the concentration at which the response was reduced by 50%, was calculated from the fitted curve. Data analysis and curve fitting were performed using Python, with the SciPy package used for non-linear regression. The estimated IC50 value for the experiment was 5.95 µM. All statistical analyses were performed using a significance threshold of  $p < 0.05$ .



ChatGPT 4o with canvas

suitable for a scientific manuscript. Be sure to adjust specifics like cell type, concentrations, incubation times, and any other experimental details to match your exact protocol.

Open Methods section in Canvas

**Methods Section Draft**

I've drafted the Methods section and opened it in the Canvas for you. Feel free to edit or let me know if there are any specific adjustments you'd like to make.

Review and provide comments

Added 7 comments on Methods Section Draft

I've added comments to clarify important details that need to be specified for reproducibility and precision, such as exact cell type, concentration ranges, measurement parameters, and statistical analysis details. Let me know if you need further refinement or additional information added.

Could you adjust the IC50 calculation?

Would you like an expanded statistical analysis?

Message ChatGPT

Liebert's Team workspace chats aren't used to train our models. ChatGPT can make mistakes.

Methods Section Draft

## Methods

### Cell Culture and Treatment

Cells were cultured in a standard growth medium (e.g., DMEM supplemented with 10% FBS) and maintained at 37°C in a 5% CO<sub>2</sub> atmosphere. For the experiment, cells were seeded into 96-well plates at a density of [insert cell density, e.g., 5,000 cells/well] and allowed to adhere overnight. After 24 hours, cells were treated with varying concentrations of the test compound ranging from 0.1 μM to [maximum concentration, e.g., 100 μM]. Each concentration was tested in triplicate wells to ensure reproducibility.

### MTT Assay

Cell viability was assessed using the MTT assay. After 24 hours of exposure to the test compound, 20 μL of MTT solution (5 mg/mL in PBS) was added to each well. Plates were incubated for an additional 3 hours at 37°C, allowing the formation of formazan crystals by metabolically active cells. Following incubation, the medium was carefully removed, and 150 μL of dimethyl sulfoxide (DMSO) was added to each well to dissolve the formazan crystals. The absorbance of each well was measured at [insert wavelength, e.g., 570 nm] using a microplate reader.

### Data Analysis

Absorbance readings were normalized to the untreated control group (0 μM) to determine relative cell viability at each concentration. The dose-response relationship was analyzed by fitting a four-parameter logistic (4PL) model to the data using non-linear regression. The 4PL model is defined as:

where is the response (absorbance), is the concentration, and parameters include the bottom and top asymptotes, the logarithm of the IC50, and the Hill slope.

### IC50 Determination

The IC50 value, representing the concentration of the test compound required to inhibit 50% of cell viability, was calculated from the fitted 4PL curve using the `scipy.optimize.curve_fit` function in Python. The final IC50 was derived by taking the inverse log of the estimated parameter from the model. Data visualization and statistical analysis were conducted using Python with `matplotlib` for plotting and `scipy` for curve fitting.

### Statistical Analysis

Statistical analyses included the calculation of mean absorbance values, standard deviation, and Pearson correlation between concentration and absorbance. All experiments were performed in triplicate to ensure accuracy and reliability of the results. Data are presented as mean ± standard deviation (SD), with curve fitting and IC50 estimation conducted using non-linear regression.

Write something...

ChatGPT

Consider specifying the exact type of cells used, as this information is crucial for reproducibility and interpretation of the results.

ChatGPT

Please provide the exact cell density used to ensure the experiment can be accurately replicated.

ChatGPT

Specify the maximum concentration of the test compound used, as this is key information for interpreting dose-response relationships.

ChatGPT

Please include the exact wavelength used for measuring absorbance, as this affects the sensitivity and specificity of the MTT assay.

ChatGPT

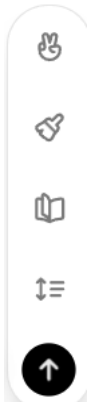
It might be helpful to clarify if the normalization involved subtracting the background absorbance or if another method was used.

ChatGPT

Consider mentioning the confidence interval or standard error for the IC50 value to provide an estimate of the reliability of the fitted parameter.

ChatGPT

You might want to add information on whether any statistical tests (e.g., ANOVA) were performed to compare groups or assess significance.





1) Dodatečný kontext:

Léčivo použité v experimentu je Oxotemfak, potenciální kandidát pro léčbu rakoviny. Buněčná linie použitá v experimentu byl lidský glioblastom. Pomoz mi vyvodit závěry z těchto dat.

2) Sepiš podrobný report k experimentu.

## Závěry na základě poskytnutého kontextu

### Kontext experimentu:

- **Léčivo:** Oxotemfak, experimentální léčivo pro potenciální léčbu rakoviny.
- **Buněčná linie:** Lidský glioblastom, což je agresivní typ nádoru mozku.

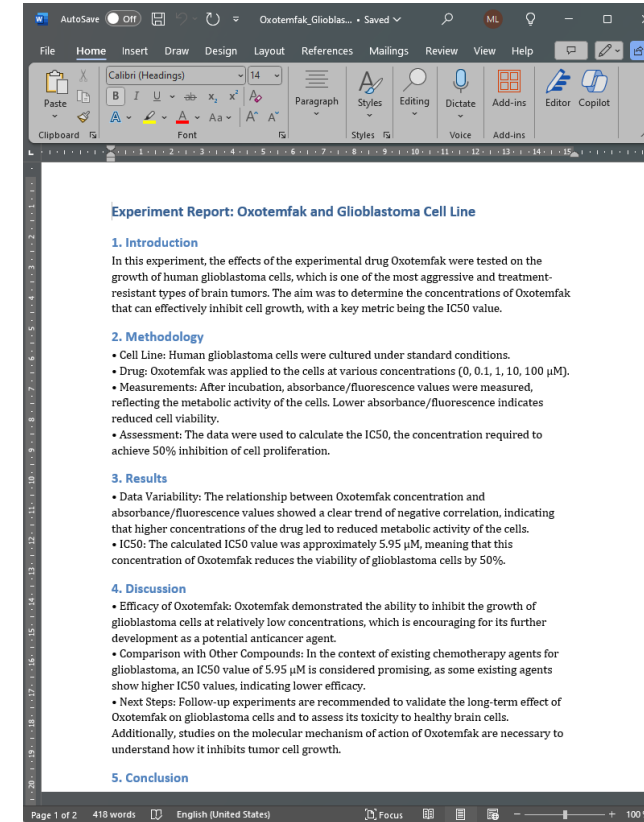
### IC50:

- **Význam IC50 (5.95  $\mu\text{M}$ ):** V kontextu výzkumu rakoviny je hodnota IC50 5.95  $\mu\text{M}$  relativně nízká, což naznačuje, že Oxotemfak má potenciál být účinným inhibítozem proliferace glioblastomových buněk. Nižší hodnota IC50 znamená, že menší koncentrace léčiva je schopna inhibovat růst buněk na 50 %, což je příznivé pro účinnost léčiva.
- **Důsledky pro terapii:** Vzhledem k tomu, že glioblastom je obtížně léčitelný kvůli své agresivitě a odolnosti vůči léčbě, účinné léčivo s nízkým IC50, jako je Oxotemfak, by mohlo představovat nadějnou možnost pro další vývoj a klinické testování.

### Podrobný report experimentu

Poskytni strukturovaný report ve formátu .doc ke stažení

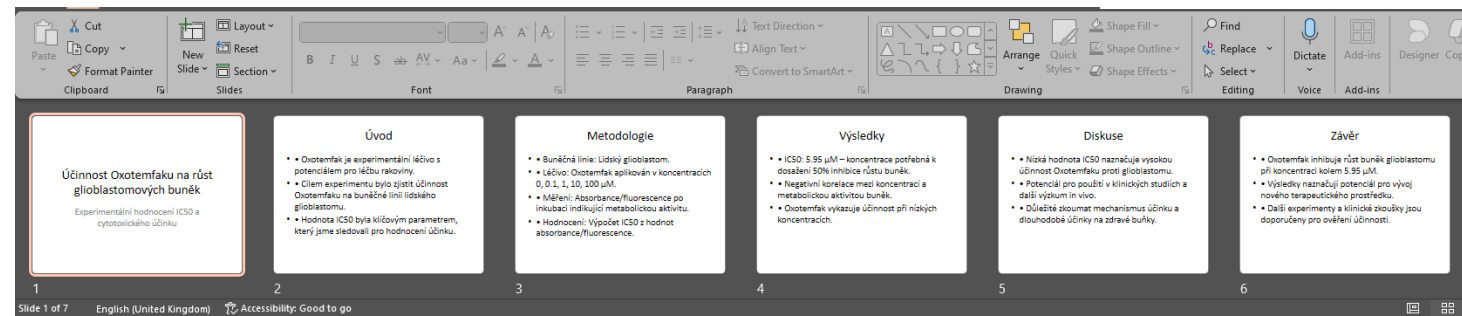
- Zde je strukturovaný report k experimentu ve formátu .doc, který si můžete stáhnout:  
[Stáhnout report \[-\]](#)



@Liegertova\_Michaela

Vytvoř ppt prezentaci k experimentu, kterou bych mohla nasdílet kolegům (poskytni ke stažení)

- Prezentace byla vytvořena a je připravena ke stažení. Můžete si ji stáhnout pomocí následujícího odkazu:  
[Stáhnout prezentaci \[-\]](#)





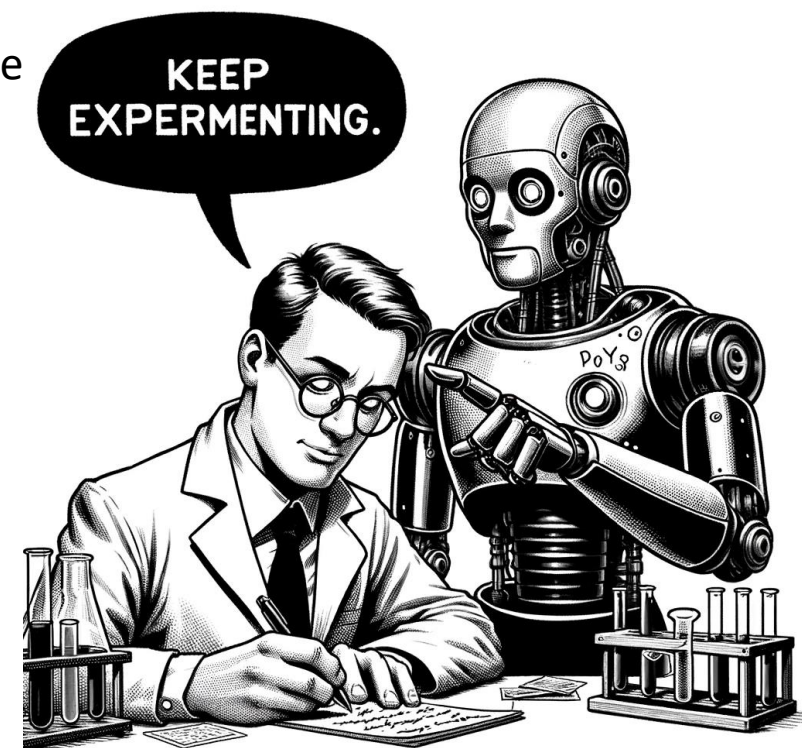
## DŮLEŽITÁ UPOZORNĚNÍ NA ZÁVĚR:

1. Na využívání LLM pro analýzu dat neexistuje univerzální návod.
2. **Experimentujte** - najděte si svůj vlastní styl práce s LLM, který poskytne nejlepší výsledky pro váš výzkum.
3. **Kritické myšlení je nezbytné** - LLM mohou generovat přesvědčivé, ale nepřesné informace.
4. **Ověřujte výsledky** - Vždy kontrolujte výstupy LLM pomocí tradičních metod.
5. **Etika a odpovědnost** - Dodržujte etické zásady a buďte transparentní při použití AI ve
6. **Ochrana dat** - Dbejte na bezpečnost při sdílení citlivých informací s AI nástroji.

## POZOR: INFORMACE Z TOHOTO WORKSHOPU BUDOU BRZY ZASTARALÉ!

*Proto:*

- Sledujte novinky a trendy v AI a datové analýze
- Zkoušejte a porovnávejte výstupy různých modelů/nástrojů
- Sdílejte zkušenosti a vytvářejte komunitu pro vzájemnou podporu





# Děkuji za pozornost



*...a přeji hodně úspěchů při experimentování s AI-asistovanou analýzou dat*

[michaela.liegertova@ujep.cz](mailto:michaela.liegertova@ujep.cz)



Spolufinancováno  
Evropskou unií



MUNI  
ICS

cesnet  
.....

VŠB TECHNICKÁ  
UNIVERZITA  
OSTRAVA

IT4INNOVATIONS  
NÁRODNÍ SUPERPOČÍTAČOVÉ  
CENTRUM

Registrační číslo IP s EO SC-CZ  
CZ.02.01.01/00/22\_004/0007682