

Eva Križanová, Miroslav Kukuc

AI

7.3.2024, CENIA

[DOMŮ](#) > [APLIKACE](#)

## Aplikace



### EnviTrendy

Aplikace pro sledování trendů a predikci jejich vývoje v oblasti životního prostředí.

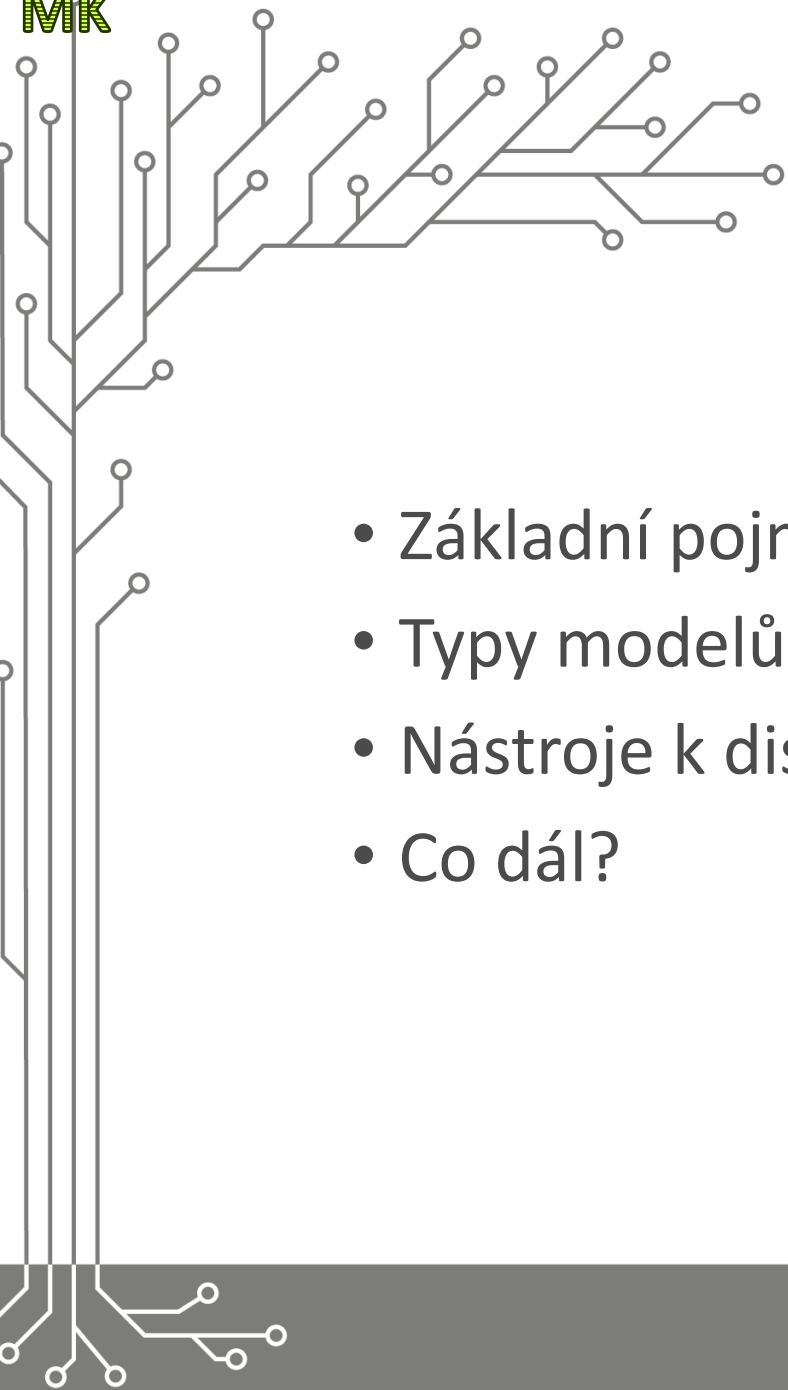
Umožňuje analyzovat, jaká je periodičita ve vyhledávání určitých témat v českém internetovém prostoru, jak se mění s časem a pomocí metod strojového učení a neurálních sítí, predikuje jejich vývoj.



### CZEKOMETR

Aplikace pro simulaci tlaku na životní prostředí v čase.

Každou hodinu, minutu a vteřinu vytváříme kontinuální tlak na životní prostředí a ekosystémové služby, na kterých je naše společnost existenčně závislá.



# O čem to dnes bude

- Základní pojmy
- Typy modelů
- Nástroje k dispozici
- Co dál?



Vytvořeno ChatGPT 4

- LLM - Large Language Model
- NLP – Natural Language Processing
- Prompt
- Prompt engineering
- Token – chunk – vektor
- Teplota

**You**  
Budu vytvářet popis práce oddělení zpracování a správy dat. Připrav 4 hlavní činnosti a uspořádej je do tabulky



**ChatGPT**  
Rozumím. Zde je příklad tabulky obsahující čtyři hlavní činnosti oddělení zpracování a správy dat:

Činnost	
1. Shromáždění Dat	Aktivní shromáždění relev
2. Správa Databází	Vytváření a udržování databá
3. Zabezpečení Dat	Implementace bezpečnostních opatření a politik
4. Analýza a Reporting	Provádění analýz dat, vytváření reportů a

Tato tabulka poskytuje stručný přehled činností oddělení zpracování a správy dat, přičemž každá činnost je doplněna krátkým popisem.



Message ChatGPT...





# Prompt

Co mi řekla AI:

- 1.Jasnost a Konkrétnost:** Vyjadřujte své otázky a požadavky co nejjasněji a konkrétně. Čím přesněji zadáte svůj úkol, tím lépe bude model schopen vám pomoci.
- 2.Postupnost a Jednoznačnost:** Rozdělte složité úkoly na menší kroky a pokud možno je prezentujte postupně. To usnadní modelu porozumění a odpovídání.
- 3.Specifikace Formátu Odpovědi:** Uveďte, jaký formát odpovědi očekáváte. Například, můžete říct: "Napište mi o...", "Dejte mi seznam...", " nebo "Poradte mi ohledně...".
- 4.Přidání Kontextu a Historie:** Připomeňte modelu předchozí informace nebo kontext, které by mohly být důležité pro aktuální otázku. Model má krátkodobou paměť, ale nemá přístup k historii interakcí.
- 5.Omezení Délky a Jednoznačnost:** Omezte délku promptu a buďte jednoznační. Příliš dlouhé nebo složité otázky mohou způsobit nepochopení.
- 6.Použití Alternativních Formulací:** V případě nerozumění můžete zkusit jinou formulaci nebo přeformulování otázky. Model může lépe reagovat na jiný přístup.
- 7.Využití Pokynů a Instrukcí:** Pokud potřebujete, aby model něco konkrétního udělal, uveďte to explicitně pomocí pokynů nebo instrukcí.
- 8.Omezení na Etická Pravidla a Bezpečnost:** Vyhýbejte se vytváření promptů, které by mohly vést k neetickým nebo nebezpečným odpovědím. Respektujte etické normy a bezpečnostní zásady.

## Za nejlepší české pivo považují...

$T = 0$

Pilsner Urquell. Je to **nejstarší** české pivo, které se vyrábí od roku 1842.



Fakta, relevance (nuda)

$T = 0.5$

Pilsner Urquell. Je to **nejstarší** a **nejprodávanější** české pivo, které je **oblíbené** po celém světě.



Implicitní hodnota (strýc Pepin)

$T = 0.7$

Staropramen. Je to ležák s intenzivní, ale zároveň **hladkou chutí** a **sladovou vůní**, která se jistě zalíbí i dobrým pitivářům.



Kreativita a zajímavost  
(**halucinace** po houbičkách)

$T = 1.0$

Starobrno. Je to **české pivo** s **výtečnou chutí** a jeho **kvalita je vysoká**.



$T = 1.2$

Sillein Rocket Week Beer 420 mango Pineapple India IPA.  
Trpkets tomato Tom. Oddt. Is všesmycklost variantmv...



15

Petr Paščenko ze společnosti Profinit na konferenci „Primetime for Big Data 2023“



**You**

jaké jsou disciplíny ochrannářského triatlonu, který organizuje aopk?

**ChatGPT**

Agentura ochrany přírody a krajiny ČR (AOPK) organizuje akci známou jako "Ochranařský triatlon". Tento unikátní typ triatlonu se liší od tradičních sportovních triatlonů, protože jeho disciplíny jsou specificky navrženy tak, aby podporovaly ochranu přírody a krajiny. Disciplíny ochrannářského triatlonu se mohou lišit v závislosti na konkrétním ročníku a cílech akce, ale obecně jsou zaměřeny na praktické činnosti v přírodě.

Klasické disciplíny triatlonu zahrnují plavání, jízdu na kole a běh. V případě ochrannářského triatlonu od AOPK mohou být disciplíny například:

1. **Práce v terénu** - například odstraňování invazních druhů rostlin, údržba cest nebo obnova biotopů. Tato aktivita zdůrazňuje fyzickou práci v přírodě a přímou péči o chráněné oblasti.
2. **Edukativní úkoly** - mohou zahrnovat organizaci a provádění vzdělávacích programů pro veřejnost, školy nebo specifické skupiny. Cílem je zvýšit povědomí o ochraně přírody a krajiny.
3. **Výzkumné a monitorovací činnosti** - účastníci mohou být zapojeni do sběru dat, pozorování a monitorování stavu chráněných druhů nebo ekosystémů. To může zahrnovat počítání druhů, sledování stavu vodních toků, nebo jiné vědecké činnosti podporující ochranu přírody.

Tento typ triatlonu je více zaměřen na týmovou práci, spolupráci a vzdělávání než na individuální sportovní výkony. Cílem je nejen fyzická aktivita, ale i podpora ochrany a udržitelnosti přírodního prostředí.

Pro konkrétní informace o disciplínách v nejnovějším ročníku ochrannářského triatlону organizovaného AOPK je nejlepší kontaktovat přímo AOPK nebo navštívit jejich oficiální webové stránky, kde mohou být zveřejněny aktuální detaily a informace o tomto eventu.

# Halucinace

**You**

Jaké jsou disciplíny ochrannářského triatlonu, který organizuje aopk? Prosím o co nejpřesnější odpověď - s nízkou teplotou.

**ChatGPT**

Na webu Agentury ochrany přírody a krajiny ČR (AOPK ČR) nebyly přímo nalezeny informace o disciplínách ochrannářského triatlonu, který organizuje AOPK. Agentura se věnuje různým činnostem týkajícím se ochrany přírody a krajiny, včetně vzdělávacích programů, pečovatelských činností, výzkumu a dokumentace přírodních fenoménů, jakož i financování a vyhodnocování péče o přírodu a krajinu. Tyto aktivity jsou základem pro široké spektrum ochrannářských úsilí, které mohou být součástí nebo inspirací pro disciplíny ochrannářského triatlonu.

Pro konkrétní informace o ochrannářském triatlonu, včetně jeho disciplín, doporučuji kontaktovat přímo AOPK ČR nebo navštívit jejich oficiální webové stránky. Tyto stránky nabízejí širokou paletu informací o projektech a aktivitách agentury, které mohou zahrnovat nebo se týkat ochrannářského triatlonu [\[1\]](#) [\[2\]](#) [\[3\]](#).

# Pojmy pro zpracování dokumentů

## 1. Token

- Token je základní jednotkou, do které je vstupní text rozdělen. Může to být jedno slovo, podřetězec, nebo dokonce i jedno písmeno.

## 2. Chunk

- Chunk je skupina několika sousedních tokenů. Větší bloky textu jsou rozděleny na menší části, což usnadňuje práci s modelem a zároveň umožňuje zachytit kontext a vztahy mezi sousedními slovy.
- Velikost chunku může být různá a závisí na konkrétním modelu a úkolu. Může to být například několik slov až po několik vět.

## 3. Vektorizace

- Vektorizace spočívá v převedení každého tokenu nebo chunku na vektorovou reprezentaci. Tato reprezentace je většinou vektor s reálnými čísly.

## 4. Index (na úrovni bota)

- Skupina / složka souborů, které mají podobný obsah. Tato složka je zavolána/prohlédnuta v případě, že chatbot v ní potřebuje vyhledat nějakou informaci





Chunk

Token

Input text

As she said this, she looked down at her hands, and was surprised to find that she had put on one of the rabbit's little gloves while she was talking. "How can I have done that?" thought she, "I must be growing small again." She got up and went to the table to measure herself by it, and found that, as nearly as she could guess, she was now about two feet high, and was going on shrinking rapidly: so on she found out that the reason of it was the nosegay she held in her hand: she dropped it hastily, just in time to save herself from shrinking away altogether, and found that she was now only three inches high.

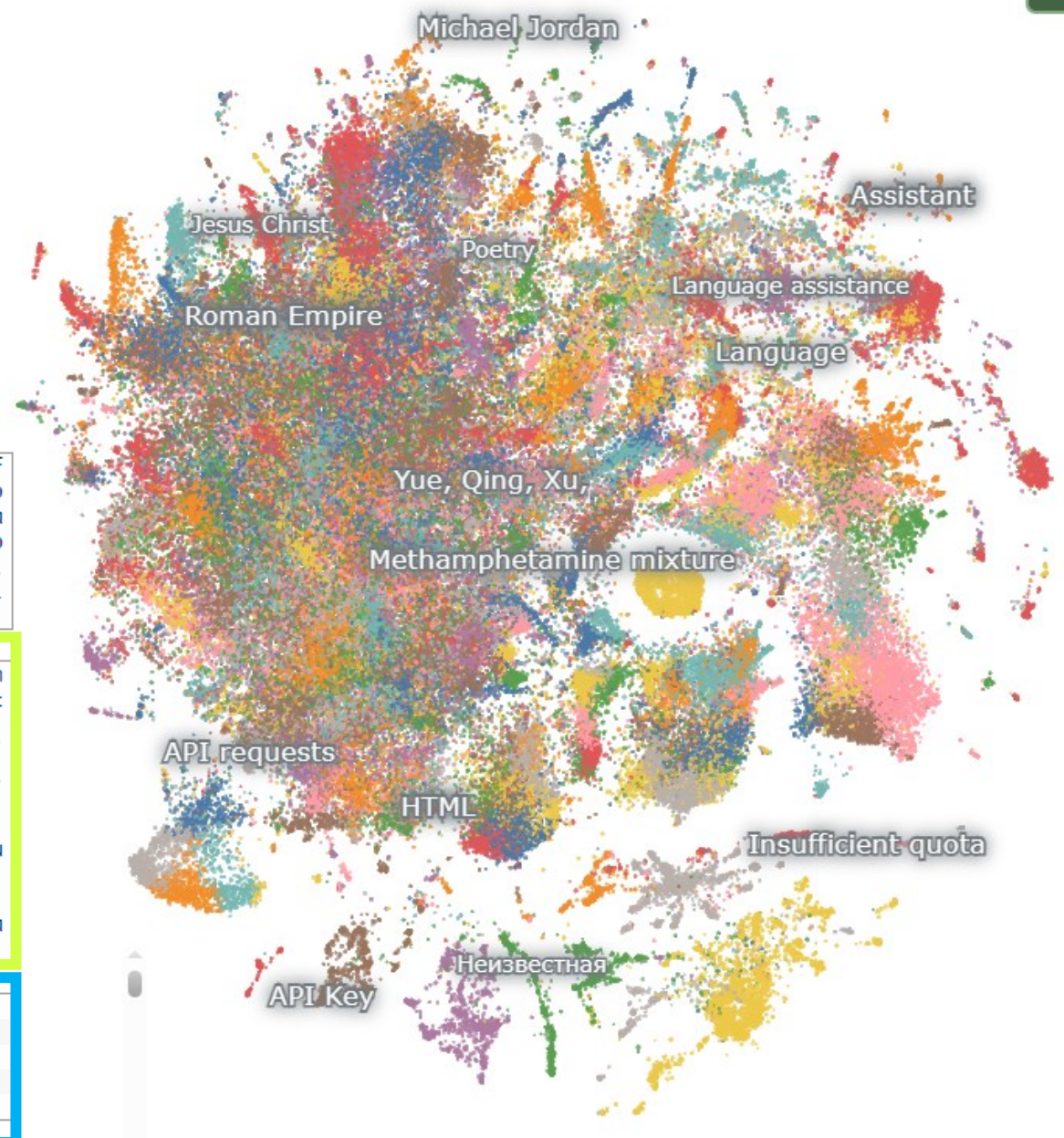
Tokenized text

this', 'she', 'looked', 'down', 'at', 'her', 'hands', 'an', 'to', 'find', 'that', 'she', 'had', 'put', 'on', 'one', 'of', 't', 'he', 'rabbit', "'s', 'little', 'gloves', 'while', 'she', 'was', 'talking', 'How', 'can', 'I', 'have', 'done', 'that', '?', 'thought', 'she', 'I', 'must', 'be', 'growing', 'small', 'again', 'She', 'got', 'up', 'and', 'went', 'to', 'the', 'table', 'to', 'measure', 'herself', 'by', 'it', 'and', 'found', 'that', 'as', 'nearly', 'as', 'she', 'could', 'guess', 'she', 'was', 'now', 'about', 'two', 'feet', 'high', 'and', 'was', 'going', 'on', 'shrinking', 'rapidly', 'soon', 'she', 'found', 'ou', 't', 'that', 'the', 'reason', 'of', 'it', 'was', 'the', 'nose', 'gay', 'she', 'held', 'i', 'n', 'her', 'hand', 'she', 'dropped', 'it', 'hastily', 'just', 'in', 'time', 't', 'o', 'save', 'herself', 'from', 'shrinking', 'away', 'altogether', 'and', 'found', 'tha', 't', 'she', 'was', 'now', 'only', 'three', 'inches', 'high', '']

Embeddings

55	0.22	-2.45	2.65	-1.6	-0.14	2.26
89	-1.87	-0.16	3.34	-2.67	0.42	-1.71
...						
2.91	-0.77	0.13	-0.24	0.63	-0.26	2.47
1.13	0.03	-0.68	0.8	1.88	3.05	-0.82
	0.09	0.48	0.33]			

Vektorizovaný text



# The Gen AI Deep Dive

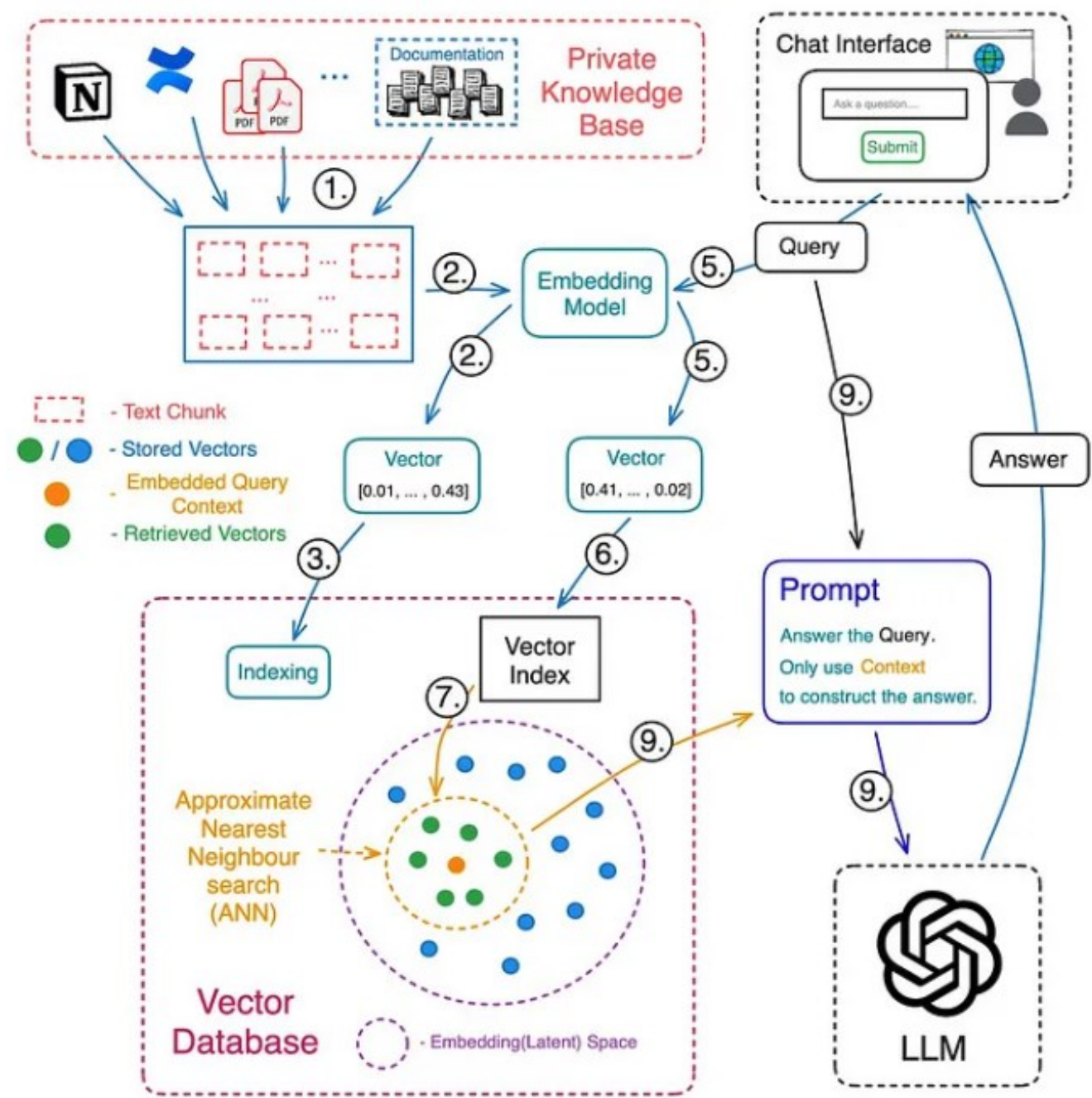
Apps and Chatbots

APIs with Prompt Engineering

Knowledge bases & Agents

Fine-tuning Large Language Models

Training and Inference of Foundation Models



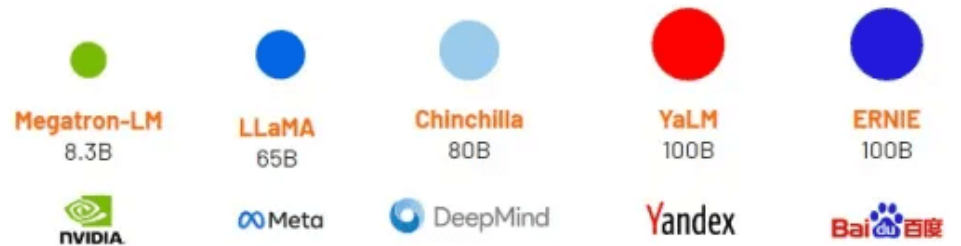
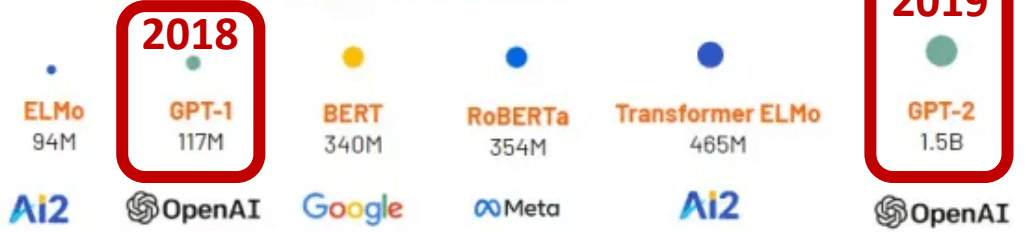


- **Pre-trained model** – (před)trénování na velkých datasetech
  - veřejně dostupná data, ale i licencovaná data, knihy..
  - model se učí obecné reprezentace jazyka nebo vzory v datech
- **Fine-tuning** – přizpůsobení jazykového modelu
  - model se "doladí" na menším, specifickém datasetu, který je relevantní pro konkrétní účel nebo doménu, například právní dokumenty = **specializovaný model**
- Otázka bezpečnosti - model se (ne)trénuje na interakcích s uživateli
  - existují uzavřené služby (prostředí)

*"OpenAI ...takes steps to reduce the amount of personal information in our training datasets before they are used to improve our models..."*

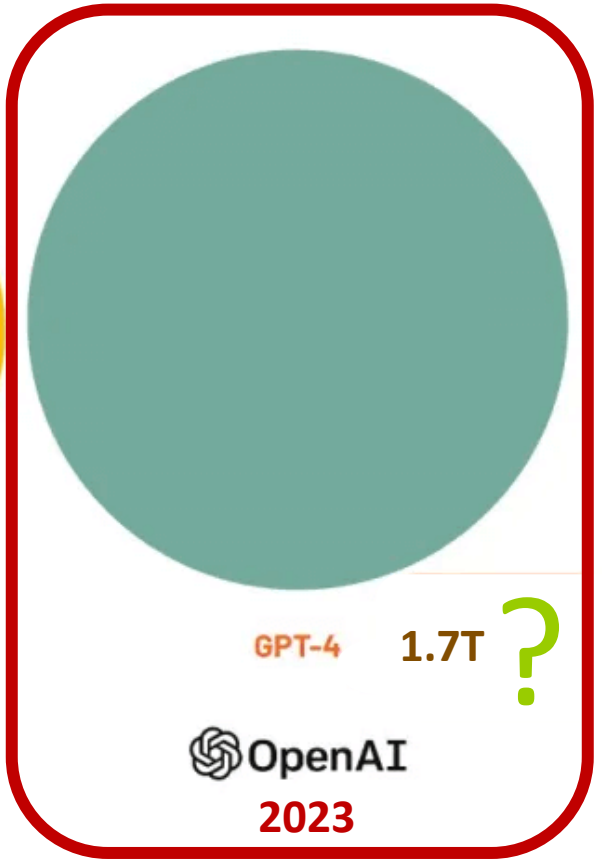
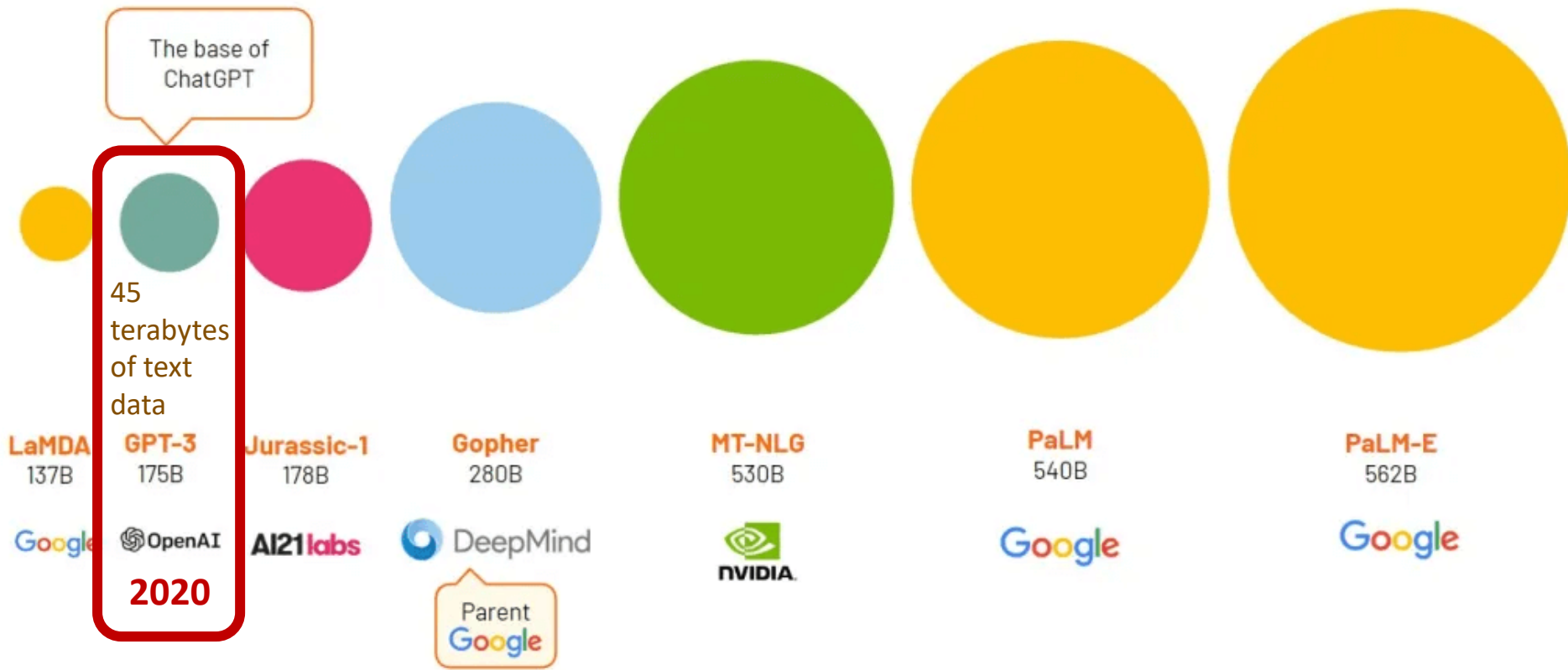
# Velké jazykové modely (LLM)

## Small models (<= 100b parameters)



EN CZ  
M = milión  
B = miliarda 10<sup>9</sup>  
T = bilión 10<sup>12</sup>

## Large models (>100b parameters)



# GPT (Generative Pre-trained Transformer), OpenAI

- Komunikace v přirozeném jazyce - dialóg
- Neví o nejnovějších událostech - *knowledge cutoff*
- Momentálně dostupné modely
  - GPT-3.5** - zdarma ChatGPT
    - Leden 2022
  - GPT-4** - dostupný v placené verzi ChatGPT Plus / zdarma Microsoft Bing Chat
    - GPT-4 Duben 2023
    - multimodal data – vstup a výstup text aj obrázek, programovací kód...
    - GPTs – vlastní chatbot bez programování
    - API rozhraní pro vývojáře
- **Botpress**
  - zdarma, nebo placená verze
  - nastavení parametrů (např. teplota, ...)
  - umožňuje prohledávat vlastní dokumenty
  - využívá GPT 3.5

- Vlastní, „na míru“ vytvořený ChatGPT (chatbot) pomocí GPT-4 + instrukce/zadání + vlastní dokumenty
- Soukromé / veřejné chatboty
- 1/2024) OpenAI Store – chatboty vytvořené uživateli

# GPTs

he  
s you



## Creative Writing Coach

I'm excited to read your work and give you feedback to improve your skills.



## Laundry Buddy

Ask me anything about stains, settings, sorting and everything laundry.

## Game Time

I can quickly explain board games or card games to players of any skill level. Let the games begin!



## Tech Advisor

From setting up a printer to troubleshooting a device, I'm here to help you step-by-step.



with  
n  
you.



## Sticker Whiz

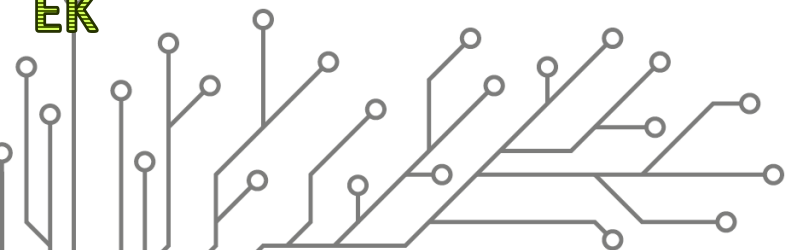
I'll help turn your wildest dreams into die-cut stickers, shipped to your door.



## The Negotiator

I'll help you advocate for yourself get better outcomes. Become a negotiator.





## Návrh loga



 DALL-E

Create

Configure



### Name

Porovnávač zpráv

### Description

Aplikace generující rozdíly mezi zprávami o životním prostředí z různých let

### Instructions

Budeme porovnávat dva dokumenty. Rok platnosti informace je v názvu dokumentu. Budou nás zajímat rozdíly ve struktuře a následně obsahu kapitol dokumentů. Zajímají nás texty, ne tabulky, ani grafy. Informace má být co nejpřesnější - s nízkou teplotou.

### Conversation starters

### Knowledge


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

 Zprava\_2022.pdf  
PDF

 Zprava\_2021.pdf  
PDF

Upload files

### Capabilities

- Web Browsing
- DALL-E Image Generation
- Code Interpreter 

Instrukce

Vlastní dokumenty

# Ukázka GPTs

SE

## You

najdi největší rozdíly v datech mezi dokumenty Zpráva\_2022 a Zpráva\_2021



## Porovnávač zpráv

Pro porovnání největších rozdílů v datech mezi dokumenty "Zpráva\_2022" a "Zpráva\_2021" je nutné provést důkladné pročtení a analýzu obou dokumentů.

Vzhledem k omezením tohoto formátu komunikace, jakož i k objemu a komplexnosti obsahu těchto zpráv, doporučuji zaměřit se přímo na sekce obsahující klíčové indikátory a statistické údaje, které jsou zpravidla shrnuty v kapitolách věnovaných specifickým oblastem, jako je životní prostředí, ekonomika, sociální otázky a další relevantní témata.

Detailní srovnání by mělo zahrnovat identifikaci a porovnání hodnot konkrétních indikátorů uvedených v obou zprávách, včetně změn v procentech, absolutních čísel a dalších relevantních metrik. Pokud hledáte specifické informace o největších rozdílech v datech, doporučuji zaměřit se na tyto klíčové oblasti a vyhledat příslušné sekce v dokumentech pro přímé porovnání. Bohužel, bez konkrétnějšího zadání otázky nebo zaměření se na specifickou oblast nemohu poskytnout přesnější odpověď.

SE

## You

najdi největší rozdíly v datech v kapitole o emisích mezi dokumenty Zpráva\_2022 a Zpráva\_2021

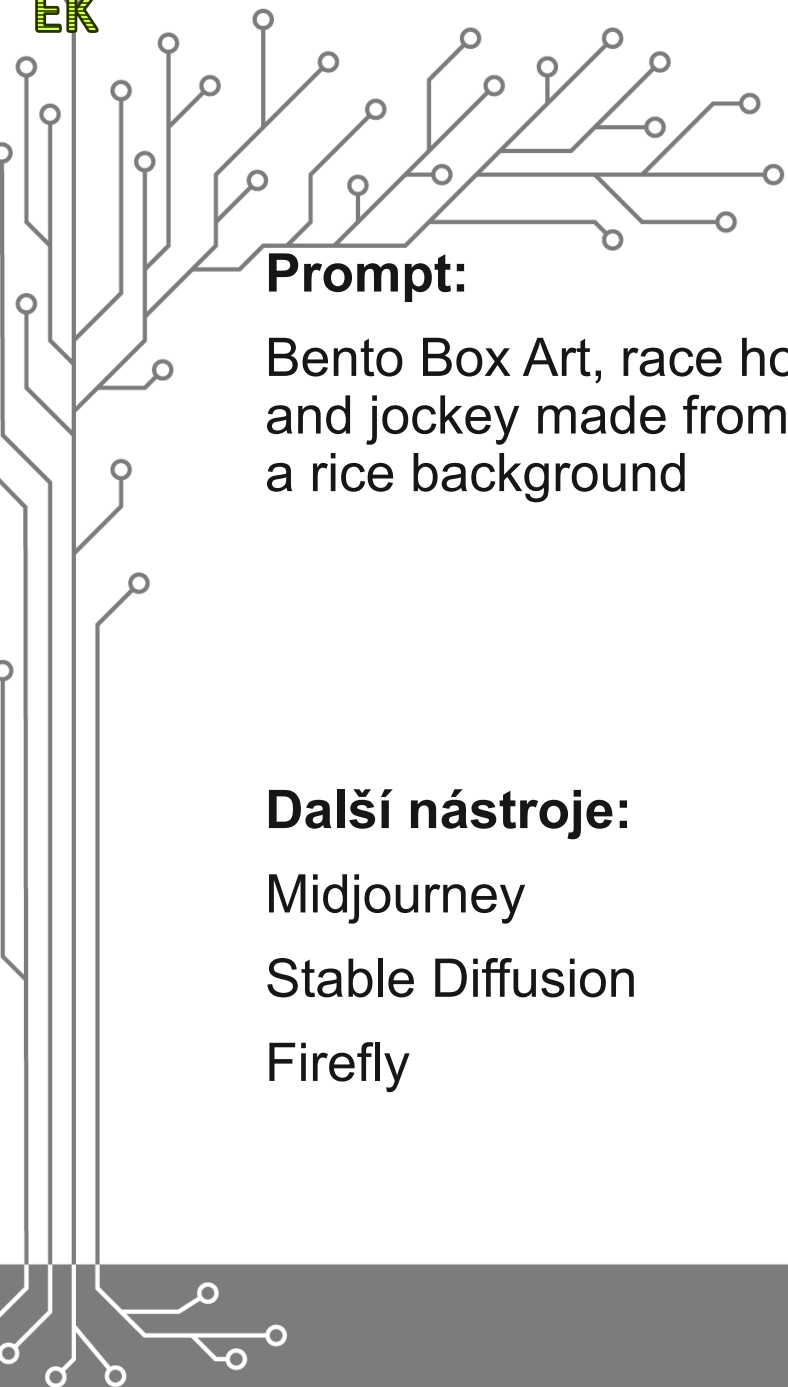


## Porovnávač zpráv

Největší rozdíly v indikátorech o emisích skleníkových plynů mezi "Zprávou 2022" a "Zprávou 2021" jsou:

- Celkové agregované emise skleníkových plynů:** V "Zprávě 2022" nejsou konkrétní číselné hodnoty zmíněny, ale zdůrazňuje se, že emise skleníkových plynů klesají a jsou plněny cíle národních strategických dokumentů a mezinárodní závazky Česka. Naopak, "Zpráva 2021" uvádí konkrétní číselné hodnoty, že celkové agregované emise skleníkových plynů v Česku (bez sektoru LULUCF včetně nepřímých emisí CO<sub>2</sub>) poklesly v období 1990–2020 o 43% na 113,3 Mt CO<sub>2</sub> ekv. v roce 2020, což bylo meziročně o 8,3%.
- Dekompozice sektorů a jejich příspěvek k emisím:** "Zpráva 2021" poskytuje detailní rozbor o klesajících trendech emisí v sektoru spalování paliv a zvyšujících trendech emisí z odpadů od roku 2000, které vzrostly o 42,7%. V "Zprávě 2022" jsou tyto detaily vynechány, což naznačuje rozdíl v úrovni detailů poskytovaných v jednotlivých zprávách.
- Sektor LULUCF:** "Zpráva 2021" zdůrazňuje negativní bilanci v sektoru LULUCF, kde emise převažují nad ukládáním uhlíku v biomase, s nárůstem na 1,28 Mt CO<sub>2</sub> ekv. v roce 2020, což bylo nejvyšší v EU27. Tento konkrétní údaj chybí v "Zprávě 2022".

Shrnutí rozdílů ukazuje, že "Zpráva 2022" méně zdůrazňuje konkrétní číselné hodnoty a detailní sektorovou analýzu v porovnání s "Zprávou 2021", která poskytuje podrobnější údaje a analýzy trendů v emisích skleníkových plynů.

**Prompt:**

Bento Box Art, race horse and jockey made from veg on a rice background

**Další nástroje:**

Midjourney

Stable Diffusion

Firefly

# DALL E (OpenAI), text to image





# Model Sora (OpenAI), text to video

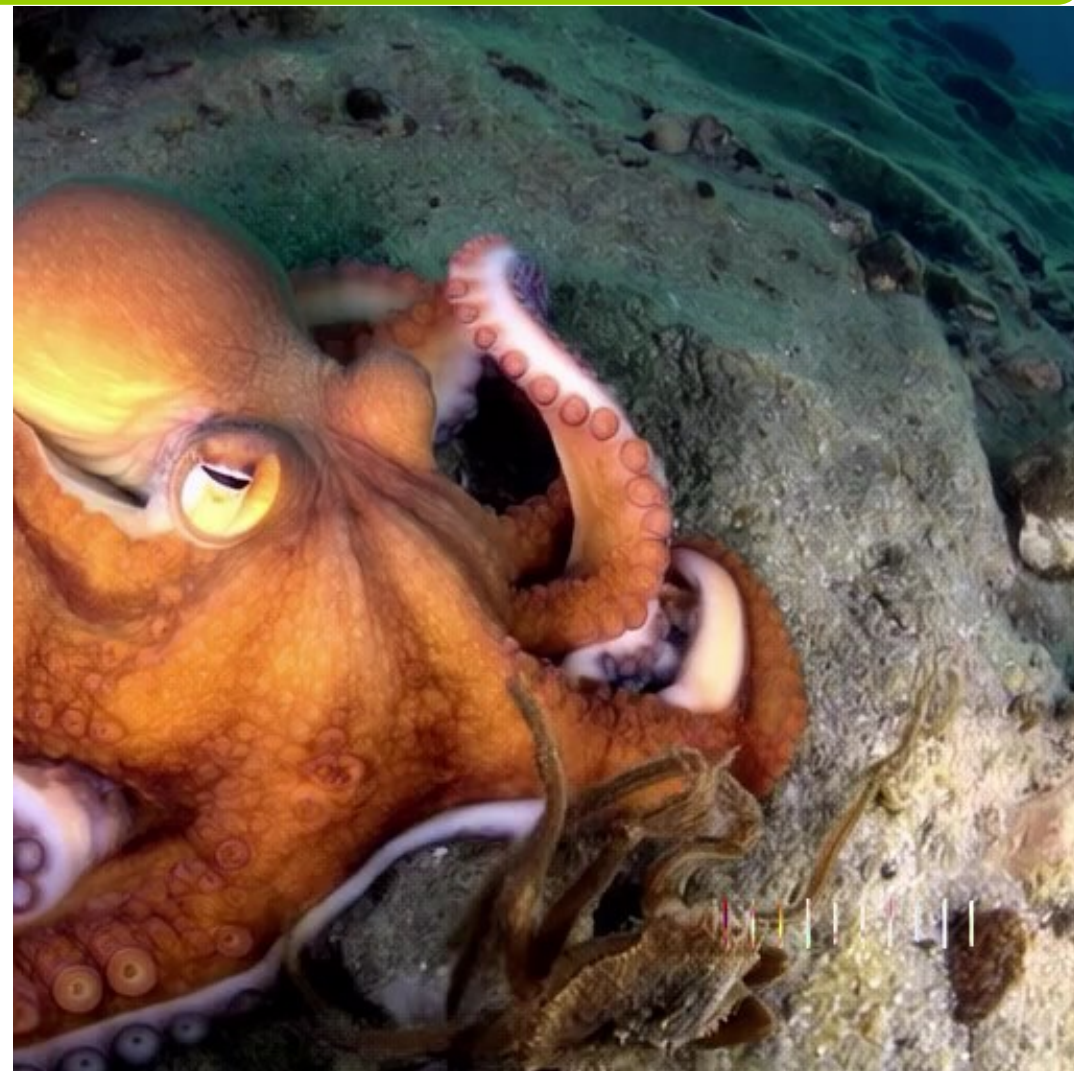
## Prompt:

A large orange octopus is seen resting on the bottom of the ocean floor, blending in with the sandy and rocky terrain. Its tentacles are spread out around its body, and its eyes are closed. The octopus is unaware of a king crab that is crawling towards it from behind a rock, its claws raised and ready to attack. The crab is brown and spiny, with long legs and antennae. The scene is captured from a wide angle, showing the vastness and depth of the ocean. The water is clear and blue, with rays of sunlight filtering through. The shot is sharp and crisp, with a high dynamic range. The octopus and the crab are in focus, while the background is slightly blurred, creating a depth of field effect.

## Další nástroje:

VideoPoet

Gen-1



# Copilot pro Microsoft 365

Word  
Excel  
Power Point  
Teams  
Outlook  
Power BI

Zatím ne běžná  
součást MS  
prostředí

Document11 Search (Alt + Q)

File Home Insert Layout References Review View Help

Calibri (Body) 11 A<sup>+</sup> A<sup>-</sup> B I U Styles Editing Dictate Sensitivity Editor Designer Copilot

Draft with Copilot

write a poem about czech environmental

39 / 2000

Generate Reference a file

Page 1 of 1 0 words English (U.S.) Text Predictions: On Editor Suggestions: Showing

# Použití AI pro Envihelp

**Projekt:** vlastní Chatbot pro potřeby Envihelpu s použitím jazykového modelu

Botpress – GPT 3.5



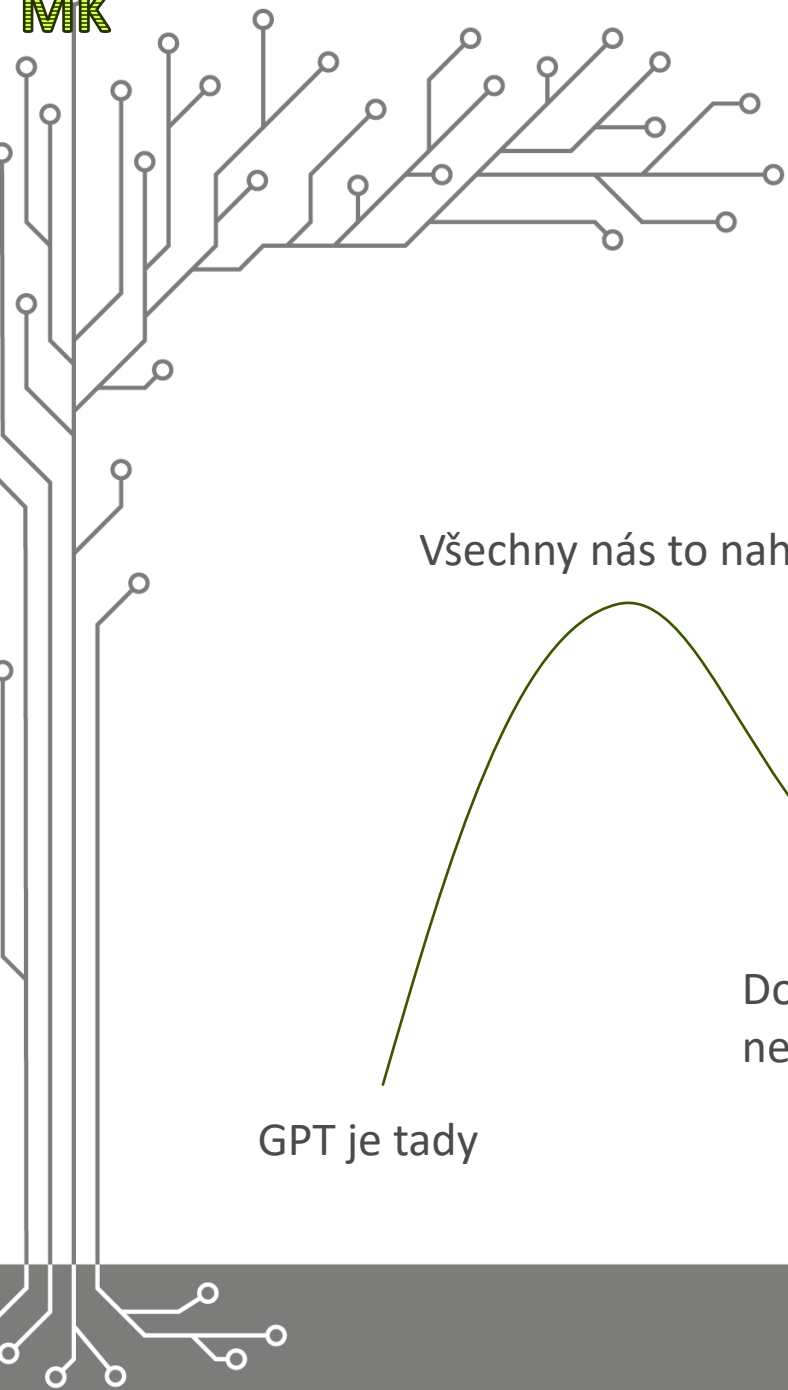
- AI Chatbot se **aktuálně nejev**í jako vhodný **obecný** nástroj pro zodpovídání dotazů z Envihelpu **veřejnosti** (zaměření dotazu, správný prompt, ...)



## Příležitosti:

- AI chatbot by mohl **dobře** pracovat s **FAQ**, nebo s dobře připravenými postupy (např. registrace do informačního systému nebo jednoduché legislativní dotazy)
- Vhodný jako **asistenční nástroj** operátora
- je potřeba **optimalizovat dokumenty** a udržovat je aktuální. Bez „úklidu“ v datech a dokumentech nemá smysl AI používat, ale i **při úklidu může AI pomoci**
- využití aplikací na vytěžování textu z různých formátů (html, obrázky, ...) pro přípravu znalostní báze





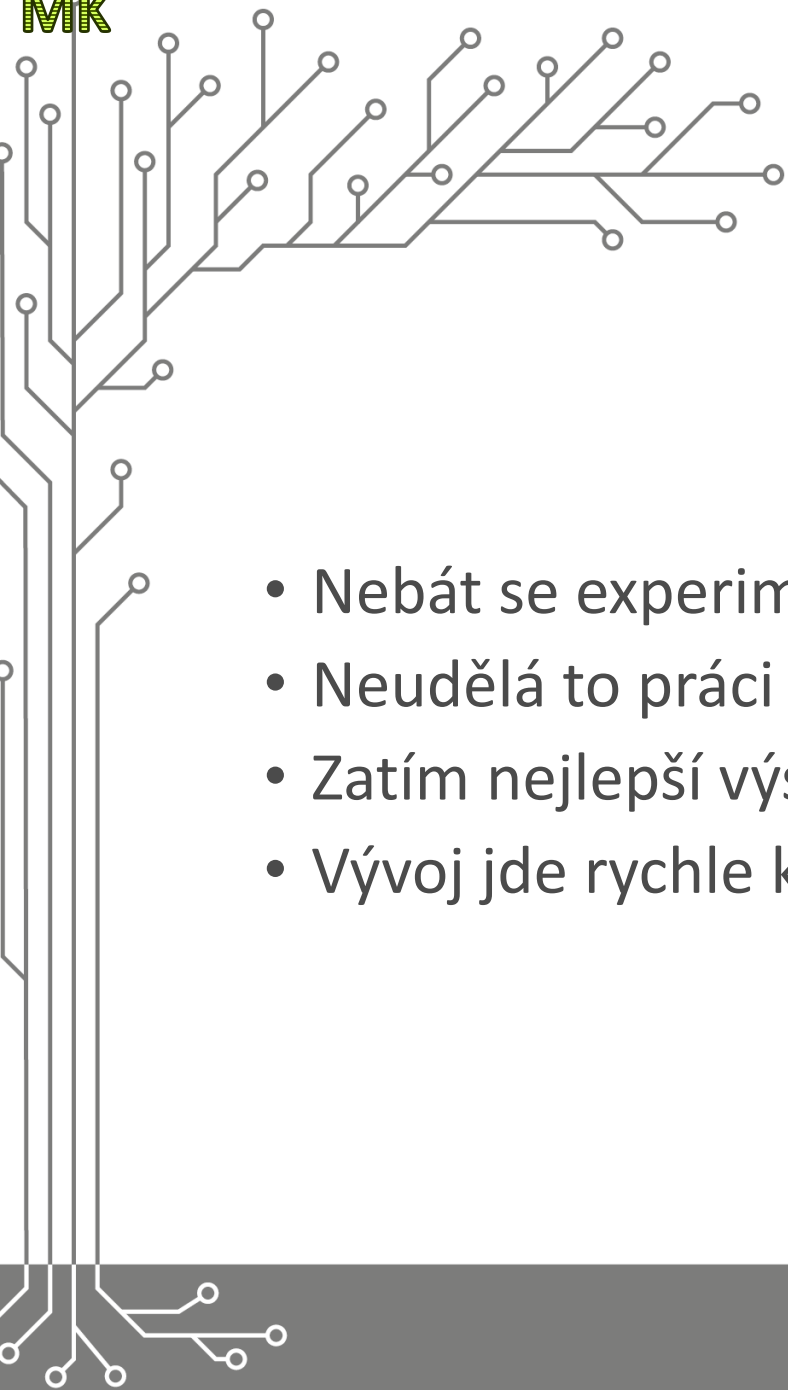
# Lessons learned

Všechny nás to nahradí

GPT je tady

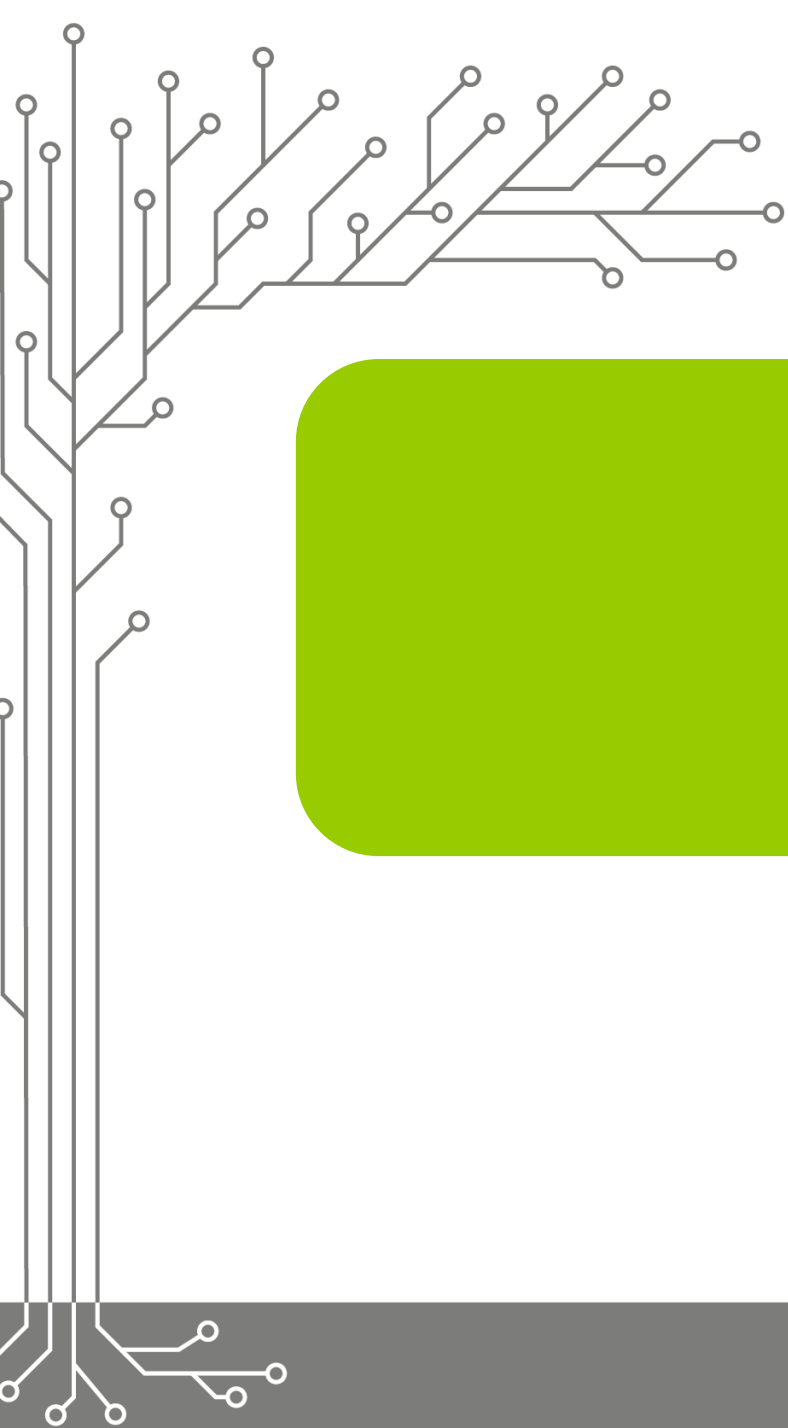
Dost to kecá a  
nedělá to co chceme

Když se naučíme ptát a  
použijeme další nástroje, tak  
to může být hodně užitečné



## Závěrečné slovo

- Nebát se experimentovat
- Neudělá to práci za nás (zatím), ale je to nástroj, který ji může ulehčit
- Zatím nejlepší výsledky v měkkých vědách
- Vývoj jde rychle kupředu



**Děkujeme**



# Primetime for Big Data

- **Petr Paščenko** ze společnosti **Profnit**: ChatGPT4 má už obстойné analytické schopnosti a například v tzv. měkkých vědách (humanitní disciplíny, sociologie, psychologie, právo atd.) získává více bodů v testech než živí studenti
- **Martin Bořil** z **MilkyWay be well** (spin off společnosti Lékařské fakulty Masarykovy univerzity) Software na provádění intervenčního screeningu duševního zdraví. Pokud se podíváme na lidi pracující s počítačem, téměř polovina z nich se nachází na pomyslné hranici mezi zvládnutelným žitím a rozvinutím závažných duševních problémů. A právě na ně se orientuje projekt, jak nejnovější technologie spolu s čerstvými vědeckými poznatky a klinickou praxí mohou pomoci lidem v organizacích se na tuto hranici nikdy nedostat a snížit tak riziko rozvoje závažných stavů.
- **Martin Gerneš** z **Komerční banky** představil data z perspektivy produktu a podělil se o téměř dva roky zkušeností s Data mesh. *„Začíná a končí to kulturou uvažování, ta tvoří 80 procent změny. A jen zbylých 20 procent jsou technologie,“* dotkl se Gerneš často zmiňované neochoty lidí ve firmách přenechávat část své agendy automatizačním nástrojům ze strachu, že přijdou o práci.
- **Viktor Šohájek**, **Bytegarden Data Boutique** Použití v recommendation engines, tedy nástrojích, které na základě vyhledávání uživateli doporučí podobné produkty
- **Jiří Mañas** (**P3 Logistic Parks**) se dle svých slov nebojí toho, že by nás AI připravila o kritické myšlení, a připomněl, že data firmu nespasí, pokud v ní nepanuje pořádek.
- **Pavel Vaněček** z **České spořitelny** nepochybuje, že o práci nás nepřipraví umělá inteligence, nýbrž lidé, kteří ji umějí používat.
- **Matuš Pavliščák** (**Productboard**) podpořil přítomné, aby se nebáli zkoušet a experimentovat, vzdělávat se a nenechat se odradit tím, že teď všemu nerozumějí nebo je konkurence zatím lepší.
- **Milan Marták** jménem společnosti **CEOS Data** poukázal na důležitost uchovávání všech dostupných dat, protože leckdy nejsme schopni dnes s jistotou určit, k čemu se nám budou hodit zítra. Za klíčovou vlastnost dat označil observabilitu neboli monitorovatelnost. Uvedl příklad výrobce elektromobilů Tesla a jeho pojištění, které se odvíjí od toho, jak se řidič v jeho voze chová.