

# Krocení umělé inteligence v praxi

pro CACIO 30.10.2024

**Ondřej Michalák**  
Principal Consultant, KPMG Lighthouse  
omichalak@kpmg.cz

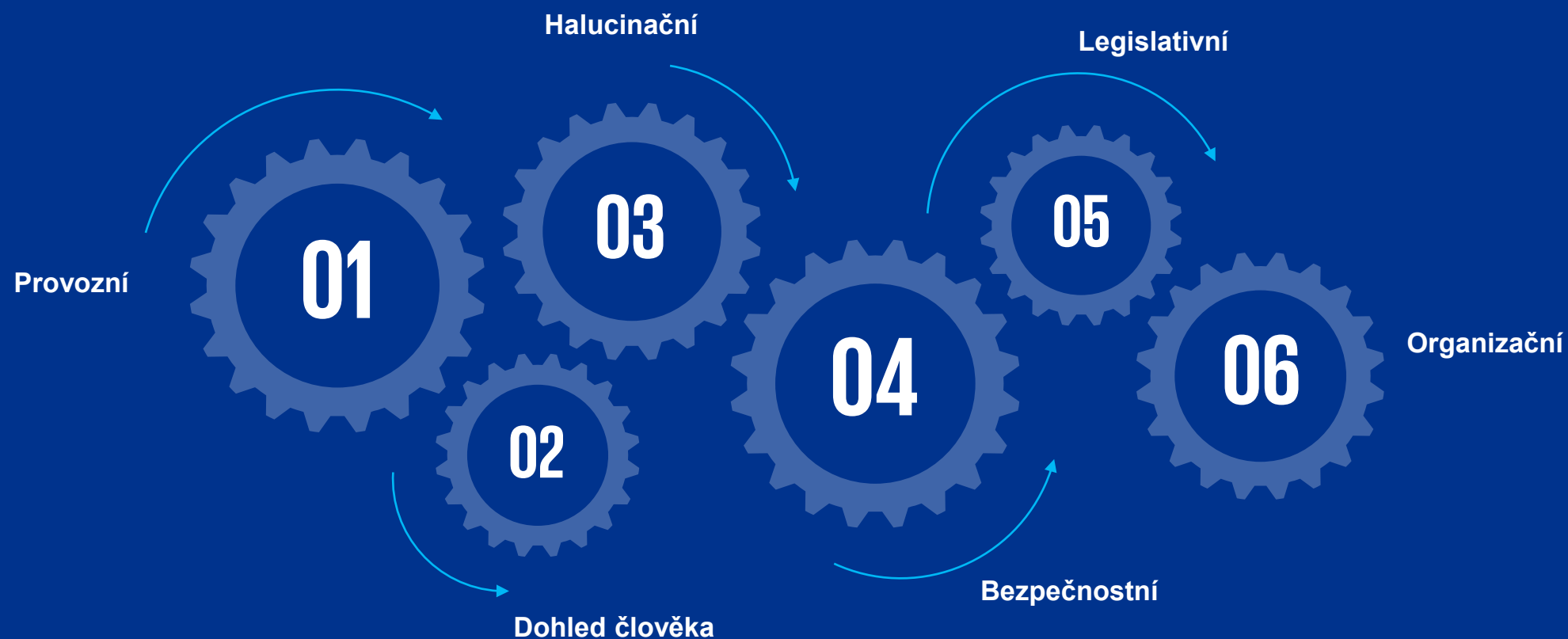


© [year] [legal member firm name], a [jurisdiction] [legal structure] and a member firm of the KPMG global organization of independent member firms affiliated with KPMG International Limited, a private English company limited by guarantee. All rights reserved.

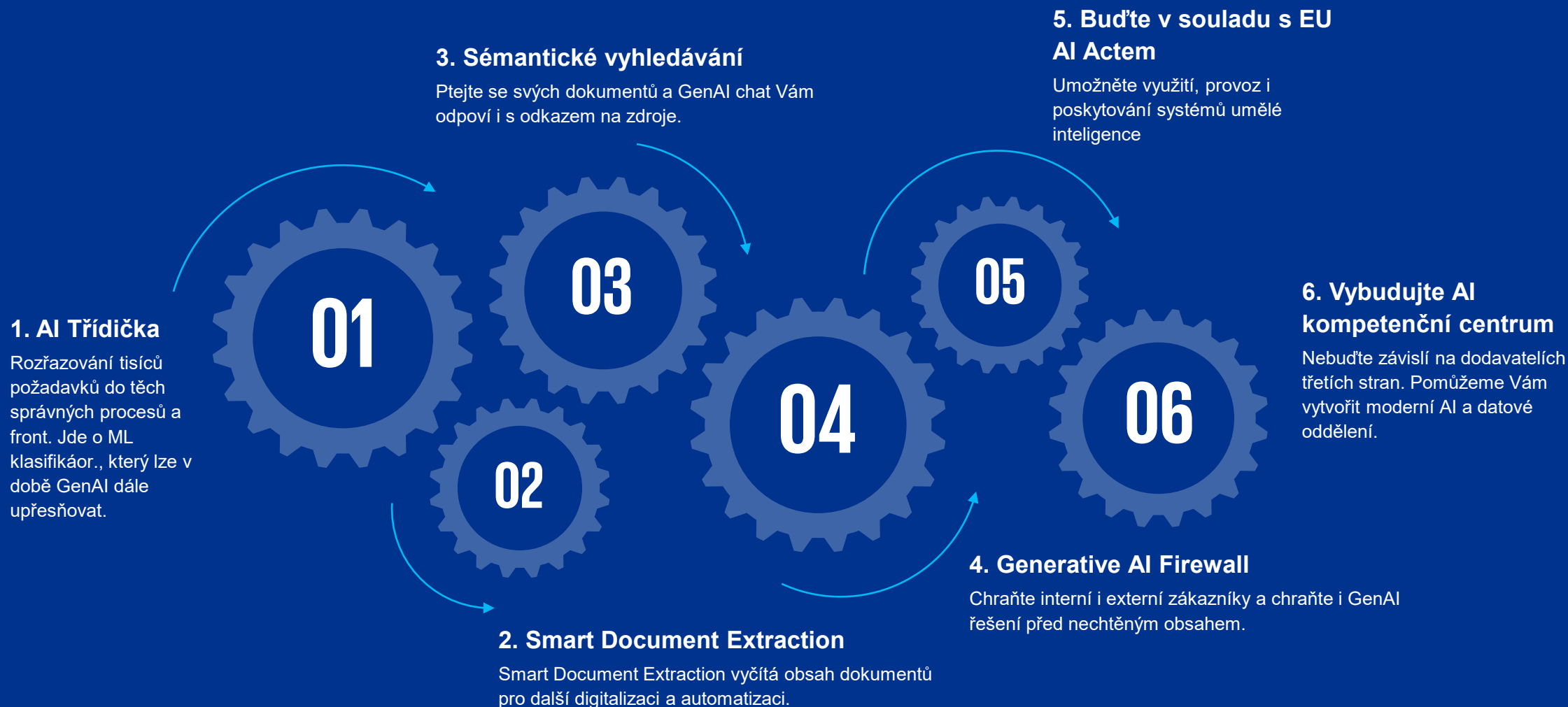
Document Classification: KPMG Public

1

# Krocení AI – Oblasti krocení



# Krocení AI – Přehled vybraných řešení



# 1. AI Třídíčka

**Představte si, že dostanete denně tisíce mailů a  
musíte rozhodnout, jak s nimi naložit?**

# Třídící asistent - Popis problému

## Nových požadavků denně

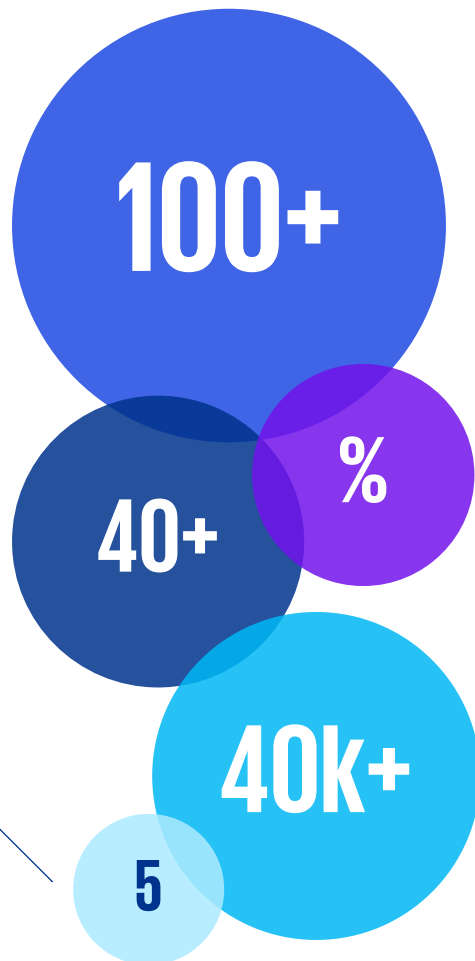
Na email info@... přichází denně stovky až tisíce požadavků, které musí člověk pročíst a rozhodnout, co s nimi dál. Tedy především zařadit do správné kategorie a vyhodnotit správný typ akce.

## Počet kategorií

Každý požadavek operátor ručně zařazuje do příslušné kategorie a odtamtud jej přebírá pracovník příslušného oddělení.

## Počet operátorů

Problém průběžně zaměstnává několik lidí.



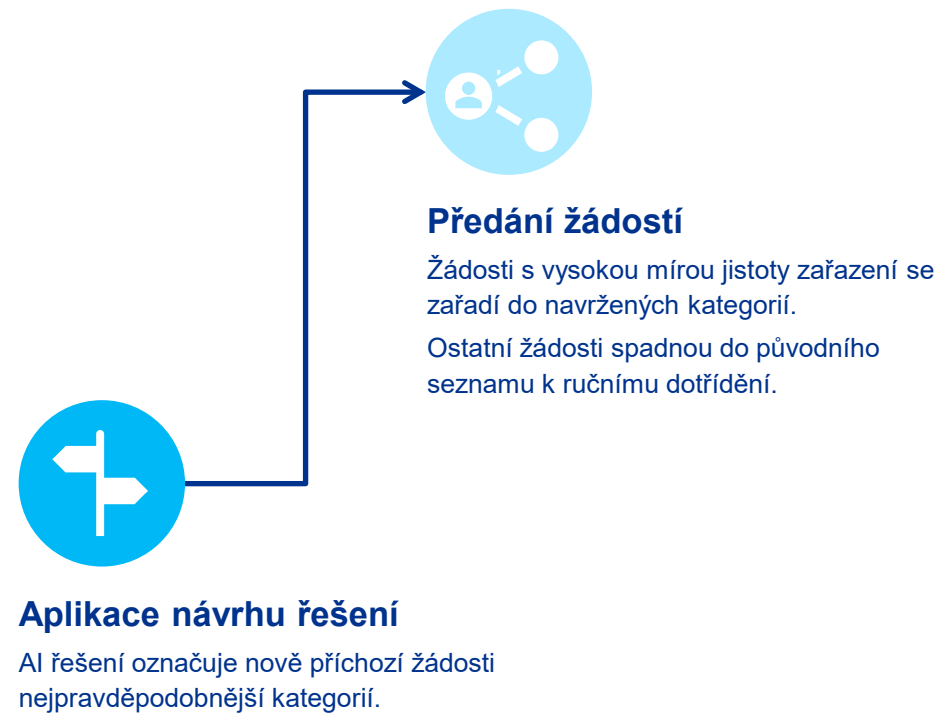
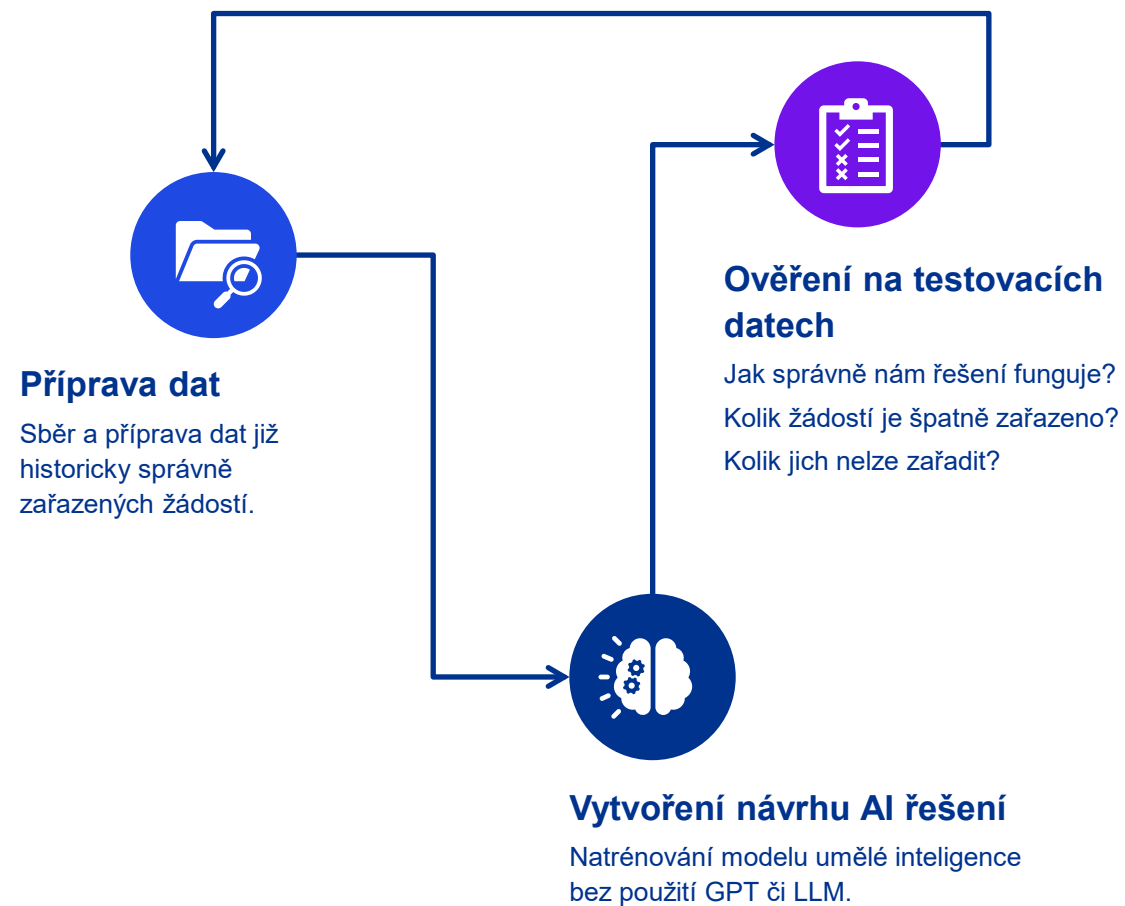
## Část se musí řešit složitě

Není jasné, jak dlouho zařazení operátorem trvá, ani operátor nemusí znát odpověď a zároveň se může obměňovat a ztrácet tak „historickou paměť“.

## Již zařazených žádostí

Máme se z čeho učit, k dispozici je historie správně zařazených požadavků.

# Třídící asistent – Příprava řešení



# Třídící asistent – Dlouhodobé vylepšování

## Kolik žádostí se nezatřídilo?

Pokud se žádost nepodaří s velkou mírou spolehlivosti zařadit, skončí v kategorii „Ostatní“. Tam jich ale může končit velmi mnoho.

## Měření úspěšnosti

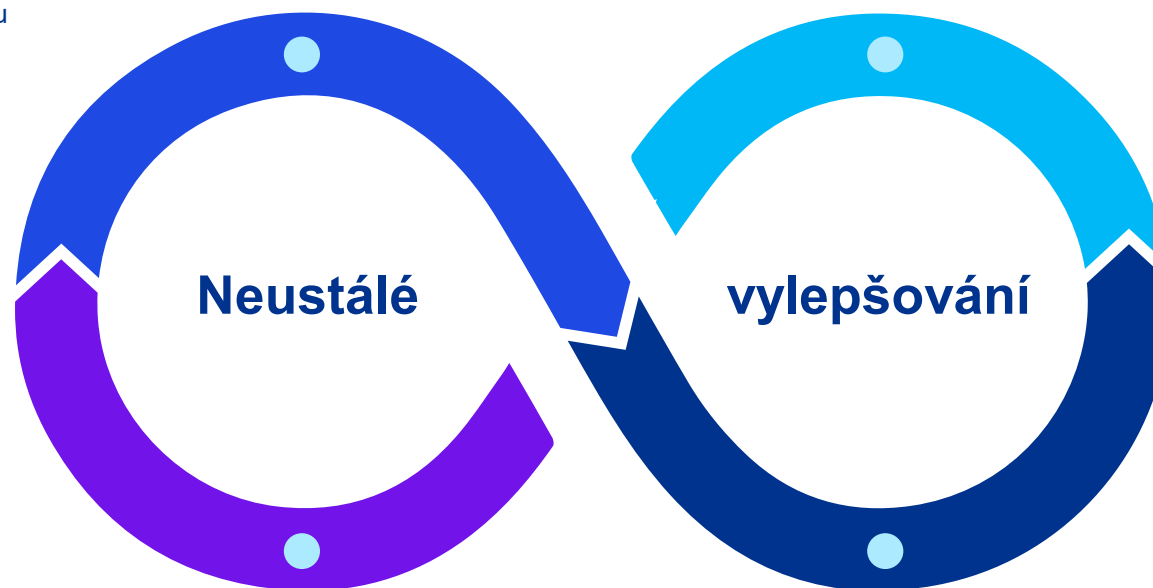
Vyhodnocení vylepšeného agenta. Možnost vrátit se k předchozím verzím, pokud se něco nedaří.

## Vylepšení Asistenta

Asistent se může přeučit na základě nových podkladů.

Doplnění pravidel.

Musí se otestovat na starších i novějších datech.



## Identifikace neúspěšného zařazení

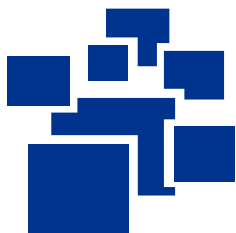
- Špatně zatříděné žádosti
- Žádosti, pro které není kategorie
- Nestandardní podklady
- Cíl: Zkrátit „Ostatní“



## 2. Smart Document Extraction (SDE)

# Vytěžujte dokumenty s dohledem člověka

Společně od prototypu k robustnímu řešení ve 4 krocích:



## Příprava vstupních dat

Extrakce historických dokumentů z prostředí klienta, analýza kvality skenů a typů šablon.



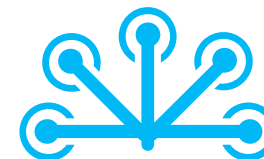
## Vytvoření prototypu

Vytvoření prototypového workflow a modelu pro extrakci dat z dokumentů



## Extrakce metadat ze dokumentů

Extrakce a uložení dat z dokumentů do strukturované databáze.



## *Návrh zpracování nových dokumentů*

Návrh procesu extrakce a archivace dat z nově přijatých dokumentů.

# 3. Sémantické vyhledávání

# Sémantické vyhledávání – AI chat s přístupem k dokumentům

- Tradiční GPT chat je obohacený dalšími dokumenty, které nebyly v trénovací sadě LLM technologií RAG.
- Takto obohacený chat může fungovat buď jako obecný dotazovací nástroj nad dodanými dokumenty nebo jako specializovaný „odborník“ na danou problematiku či Autonomní úzce zaměřený agent.
- Demo na obrázku je obohaceno několika soudními rozhodnutími.

KPMG AI Search Chat Zeptejte se Demo

Vymazat chat Developer settings

## Vyhledávač v datech

Zeptejte se svých dat, například...

Jaký byl účel emise listinných korunových dluhopisů vydávaných žalobcem dne 29. 2. 2012?

Vypiš v tabulce rozhodnutí, ve kterých se řeší problematika daně z příjmu

Uplatňoval nějakém v rozsudku týkající se kasační stížnosti stěžovatel námitku nepřezkoumatelnosti?

Zeptejte se svých dat...

# Místo odkazu na odpověď dostaneme odpověď s odkazem


KPMG AI Search Chat Zeptejte se Demo

[Vymazat chat](#) [Developer settings](#)

Jaký byl účel emise listinných korunových dluhopisů vydávaných žalobcem dne 29. 2. 2012?

Účelem emise listinných korunových dluhopisů vydávaných žalobcem dne 29. 2. 2012 byla "novace předchozího krátkodobého úročeného závazku vůči [redacted] z titulu akvizice společnosti [redacted] s.r.o." <sup>1</sup>. Tato emise byla upsána a uhrazena formou zápočtu s výše popsaným závazkem panem [redacted] <sup>1</sup>.

Zdroje: [1. 1 Afs 208\\_2023- 43 - html.html](#)

Zeptejte se svých dat... 

Jaký byl účel emise listinných korunových dluhopisů vydávaných žalobcem dne 29. 2. 2012?



Účelem emise listinných korunových dluhopisů vydávaných žalobcem dne 29. 2. 2012 byla "novace předchozího krátkodobého úročeného závazku vůči [redacted] Endlovi z titulu akvizice společnosti [redacted] 1. Tato emise byla upsána a uhrazena formou zápočtu s výše popsaným závazkem panem [redacted] 1.

Zdroje: [1. 1 Afs 208\\_2023- 43 - html.html](#)

Zeptejte se svých dat...



Supporting content

Citation

43

1 Afs 208/2023 -

## U S N E S E N Í

proti žalovanému: **Odvolací finanční ředitelství**, se sídlem Masarykova 31, Brno, proti rozhodnutí žalovaného ze dne 6. 4. 2021, č. j. 11567/21/5200-11434-701858, v řízení o kasační stížnosti žalobce proti rozsudku Krajského soudu v Ostravě – pobočky v Olomouci ze dne 13. 9. 2023, č. j. 60 Af 24/2021 - 101,

t a k t o :

Věc se p o s t u p u j e rozšířenému senátu.

O d ů v o d n ě n í :

# 4. Generative AI Firewall

**Již před dvaceti lety o vás vyhledávač věděl více, než si dokážete představit a to jen a jen na základě vyhledávaných slov a kliků na odkazy.**

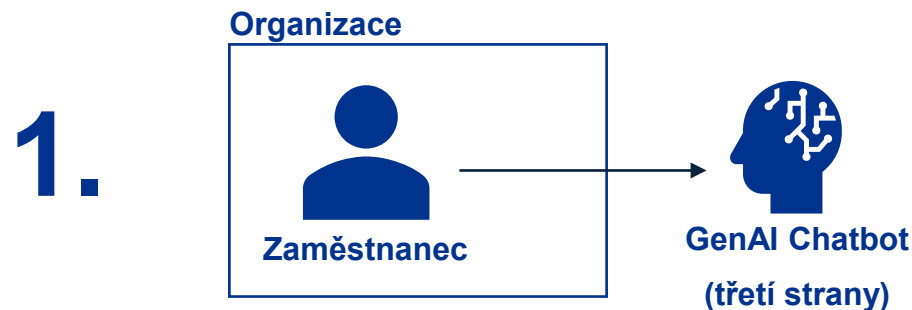


**Co všechno o vás dokáže zjistit chat dle otázek,  
které mu kladete a dokumentů, které do něj vkládáte?**

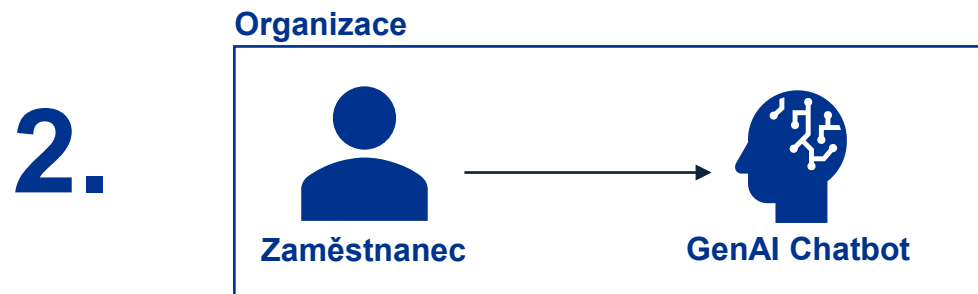
# Provozujete interní nebo externí GenAI chat řešení?

## Jste si vědomi rizik?

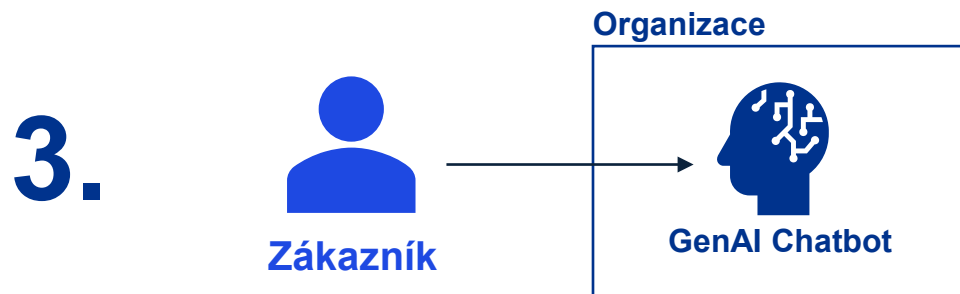
# Proč GenAI Firewall?



- Mít kontrolu nad komunikací a daty přicházejícími i odcházejícími
- Chránit zaměstnance před škodlivým nebo zavádějícím obsahem
- Auditovat interakce

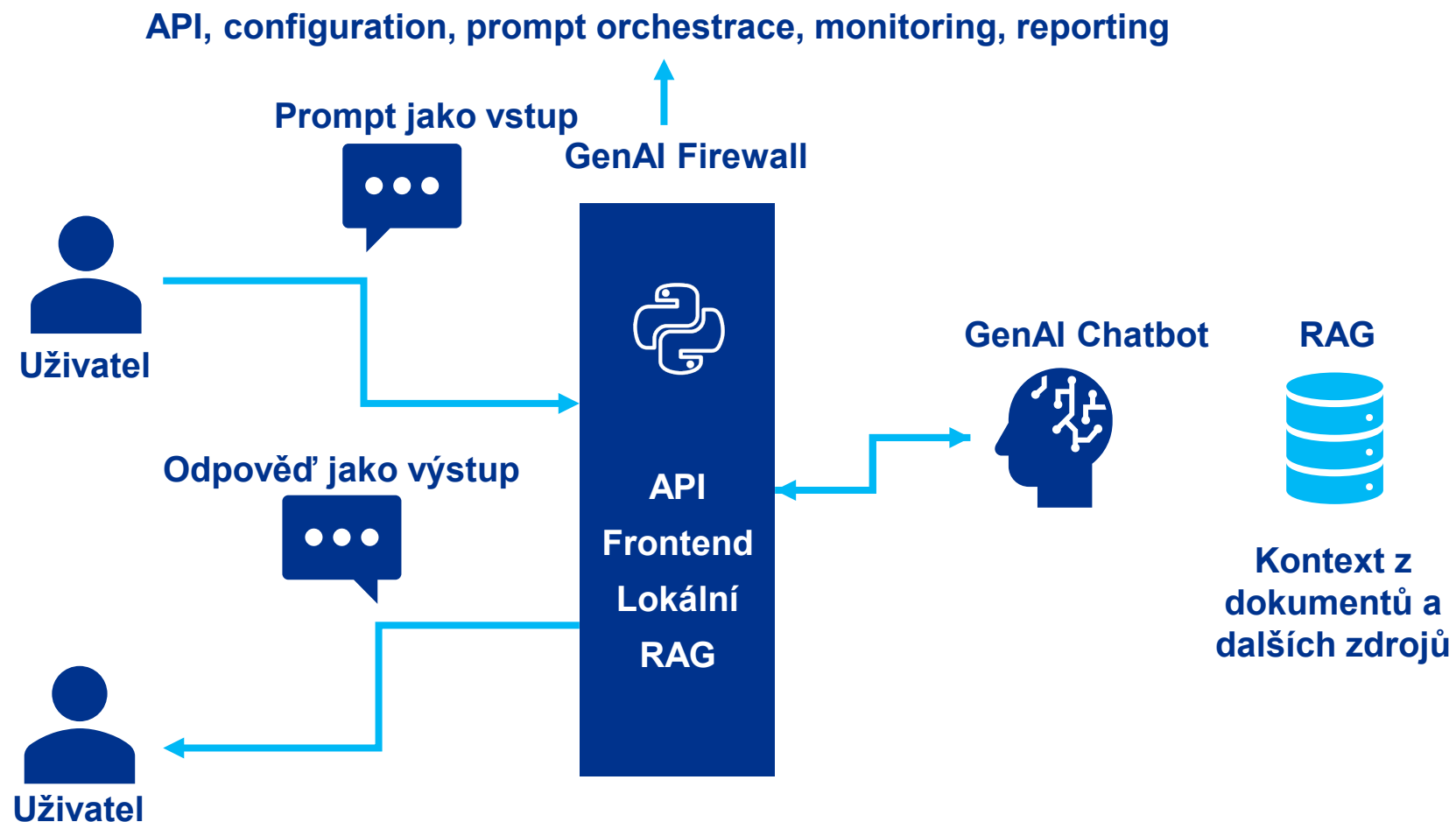


- Učit se z uživatelských potřeb z konverzací
- Chránit systém před škodlivým úmyslem
- Auditovat interakce









- Zvýšení zákaznické spokojenosti
- Chránit uživatele před zavádějícím obsahem
- Chránit systém před škodlivým úmyslem
- Auditovat interakce

# Generative AI pod kontrolou – Koncept



# Snižujeme riziko známých nebezpečí Generativní AI a LLM

## OWASP Top 10 for LLM Applications

<p><b>LLM01</b> </p> <h3>Prompt Injection</h3> <p>This manipulates a large language model (LLM) through crafty inputs, causing unintended actions by the LLM. Direct injections overwrite system prompts, while indirect ones manipulate inputs from external sources.</p>	<p><b>LLM02</b> </p> <h3>Insecure Output Handling</h3> <p>This vulnerability occurs when an LLM output is accepted without scrutiny, exposing backend systems. Misuse may lead to severe consequences like XSS, CSRF, SSRF, privilege escalation, or remote code execution.</p>	<p><b>LLM03</b> </p> <h3>Training Data Poisoning</h3> <p>This occurs when LLM training data is tampered, introducing vulnerabilities or biases that compromise security, effectiveness, or ethical behavior. Sources include Common Crawl, WebText, OpenWebText, &amp; books.</p>	<p><b>LLM04</b></p> <h3>Model Denial of Service</h3> <p>Attackers cause resource-heavy operations on LLMs, leading to service degradation or high costs. The vulnerability is magnified due to the resource-intensive nature of LLMs and unpredictability of user inputs.</p>	<p><b>LLM05</b></p> <h3>Supply Chain Vulnerabilities</h3> <p>LLM application lifecycle can be compromised by vulnerable components or services, leading to security attacks. Using third-party datasets, pre-trained models, and plugins can add vulnerabilities.</p>
<p><b>LLM06</b> </p> <h3>Sensitive Information Disclosure</h3> <p>LLMs may inadvertently reveal confidential data in its responses, leading to unauthorized data access, privacy violations, and security breaches. It's crucial to implement data sanitization and strict user policies to mitigate this.</p>	<p><b>LLM07</b> </p> <h3>Insecure Plugin Design</h3> <p>LLM plugins can have insecure inputs and insufficient access control. This lack of application control makes them easier to exploit and can result in consequences like remote code execution.</p>	<p><b>LLM08</b></p> <h3>Excessive Agency</h3> <p>LLM-based systems may undertake actions leading to unintended consequences. The issue arises from excessive functionality, permissions, or autonomy granted to the LLM-based systems.</p>	<p><b>LLM09</b> </p> <h3>Overreliance</h3> <p>Systems or people overly depending on LLMs without oversight may face misinformation, miscommunication, legal issues, and security vulnerabilities due to incorrect or inappropriate content generated by LLMs.</p>	<p><b>LLM10</b></p> <h3>Model Theft</h3> <p>This involves unauthorized access, copying, or exfiltration of proprietary LLM models. The impact includes economic losses, compromised competitive advantage, and potential access to sensitive information.</p>

# 5. Připravte se na EU AI Act

**Používáte, provozujete nebo poskytujete řešení s prvky umělé  
inteligence. Kolik takových systémů asi máte?  
Jste připraveni na EU AI Act?**

**Až 94 %  
společností věří,  
že umělá  
inteligence může  
být konkurenční  
výhodou,  
nicméně 92 %  
společností  
zpochybňuje  
důvěryhodnost  
svých dat  
a obává se ztráty  
dobré pověsti  
své firmy**

# Zodpovědná AI

KPMG Zodpovědná AI (Responsible AI) je rámec pro hodnocení a řízení integrity, odolnosti, důvěryhodnosti a vysvětlitelnosti modelů strojového učení a umělé inteligence. KPMG experti Vám poradí co (a jak) změnit, abyste používali AI zodpovědně, a zároveň Vám pomůžou připravit se na soulad s připravovanou unijní regulací, zejména s nařízením o umělé inteligenci, tzv. „AI akt“

Naše služba pomáhá odpovědět na tyto otázky:

- Důvěřujete AI aplikacím, datům a jejich monetizaci?
- Je to legální?
- Je to etické?

## KPMG nabídka služeb

### Zhodnocení

Současného stavu využití aplikací AI

### Design

Dokreslení chybějících komponent do efektivního AI procesu

### Implementace

Pomoc při implementaci AI systémů



## Klíčové oblasti a přístup KPMG

Zajistit vysokou úroveň důvěryhodnosti a zvýšit transparentnost strojového rozhodování a systémů AI. Zajistit, že jsou tyto systémy zároveň legální a etické

### Důvěryhodnost a transparentnost AI



- Spolehlivý systém řízení jakosti
- Požadovaný postup posuzování shody
- Příslušná dokumentace a spolehlivý systém monitorování po uvedení na trh
- Informovanost uživatelů o tom, že komunikují s AI systémy

### Právní aspekty



- Identifikace příslušných zákonů
- Analýza vnitřních předpisů a jejich soulad se zákony
- Analýza smluv s dodavateli a zákazníky
- Řízení společnosti z pohledu práva

### Etické aspekty



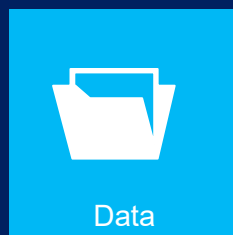
- Etický kodex
- Řízení společnosti z pohledu etiky
- Znalost etických problémů
- Etická kontrola budoucích produktů / služeb / projektů



# Zodpovědná AI v souladu s EU nařízením „AI akt“

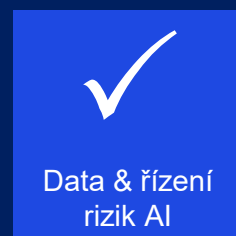
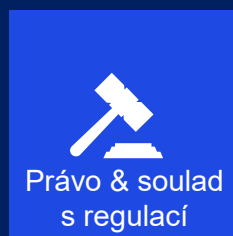
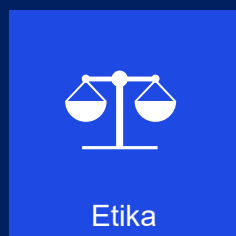
Připravované EU nařízení, kterým se stanoví harmonizovaná pravidla pro umělou inteligenci, má zavést povinnosti pro výrobce, dovozce, distributory, uživatele nebo jakékoli jiné třetí strany. Za nesoulad s nařízením bude možné uložit pokutu až ve výši 7% z celosvětového obrátu společnosti\*

## 9 oblastí pro Zodpovědnou AI



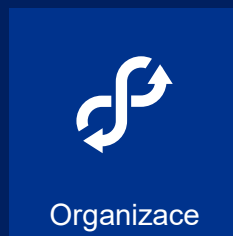
### Aplikace řízené daty

V rámci těchto témat se zaměřujeme na správu a řízení životního cyklu algoritmů, dále na kvalitu používaných dat a technologií



### Rizika

Dále se specializujeme na etiku AI, soulad AI systémů s právními předpisy a EU regulací, a zároveň identifikujeme potenciální rizika včetně návrhu jejich mitigace



### Základy

V rámci organizace a řízení daného podniku analyzujeme jaké jsou předpoklady (základy) společnosti pro úspěšný a odpovědný vývoj a řízení algoritmů

## Společnosti využívající Zodpovědnou AI dosáhly:

- Zvyšování důvěry zainteresovaných stran a pozitivní pověsti
- Lepšího přijetí AI řešení
- Možnosti pokročilé analytiky
- Souladu s regulací a interními pravidly
- Jistotu v oblasti umělé inteligence
- A – v konečném důsledku – měřitelné konkurenční výhody prostřednictvím datové analýzy

# 6. Vytvořte si AI kompetenční centrum

# Začlenění AI, digitálního a datového centra do organizace

*Klíčovým krokem transformace bylo jasné vymezení kompetencí datového centra v rámci stávající organizace.*

## 01

Podpora při realizaci projektu

*Přímá podpora všech AI, digitálních, technologických a datových projektů v rámci společnosti.*

## 02

Minimální životaschopné produkty

*Dodávání analytických a datových MVP a koordinace komunikace pro všechna zúčastněná oddělení.*

## 03

Sdílení znalostí

*Centralizace informací a znalostí z různých projektů, systematické sdílení ve skupině.*

## 04

Dlouhodobý vývoj

*Internalizace uvedených kompetencí umožňuje sledování výsledků, plánování a rozvoj vyspělosti datového centra.*



# Shrnuti



# Krocení AI – Oblasti krocení a jak jsme je pokryli



# Provedeme Vás procesem úspěšné adopce AI.

# KPMG



© 2024 KPMG Česká republika, s.r.o., společnost s ručením omezeným založená dle právních předpisů České republiky a členská společnost globální organizace nezávislých členských společností KPMG, přidružených ke KPMG International Limited, soukromé anglické společnosti s ručením omezeným. Všechna práva vyhrazena.