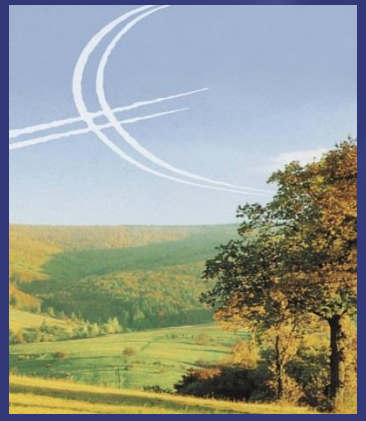
# Bezpečné zpracování dat pomocí AI

Únor 2025



Petr Pavlinec, Kraj Vysočina

Kraj Vysočina

## Co je EU AI Act?

- **Evropské nařízení o umělé inteligenci (AI Act) je první komplexní regulace AI na světě.**
- **Cíl: Zajistit bezpečnost, transparentnost a odpovědné používání AI v EU.**

- Přístup založený na **hodnocení rizik** – čím vyšší riziko, tím přísnější pravidla.
- **Klasifikace AI systémů podle rizika:**

  - **Neakceptovatelné riziko** (zakázané AI – např. manipulativní techniky, sociální scoring)
  - **Vysoké riziko** (kritické oblasti jako zdravotnictví, finance, zaměstnanost)
  - **Omezené riziko** (např. chatboty – nutnost informovat uživatele)
  - Minimální riziko (např. AI pro hry, doporučovací algoritmy – bez regulace)

# Klíčové požadavky a dopady EU AI Act

**Klíčové povinnosti pro AI vývojáře a provozovatele:**

- **Transparentnost –** uživatelé musí vědět, že interagují s AI.
- **Odpovědnost –** firmy musí zajistit bezpečnost AI před uvedením na trh.
- **Lidský dohled – AI nesmí plně nahrazovat rozhodování v kritických oblastech – klíčové pro veřejnou správu**
- **Ochrana soukromí –** omezení biometrického sledování na veřejnosti.

- **Časový plán implementace:**
  - 2024: Přijetí nařízení
  - 2025: Platnost pro zakázané systémy AI
  - 2026: Povinnost pro vysokorizikové AI systémy

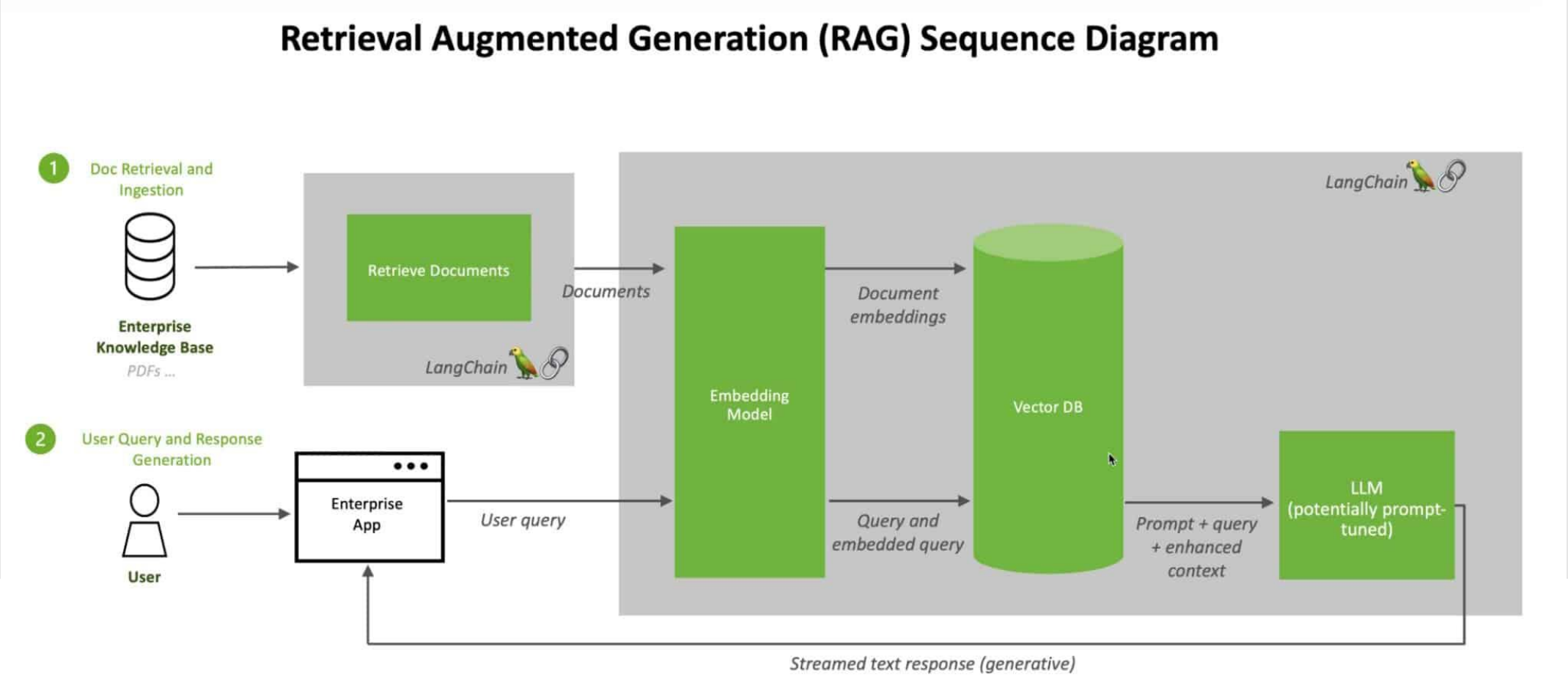
# AI – právní problémy

- **Právní problémy cloudových LLM**

**OpenAI Europe privacy policy:** By using our Services, you understand and acknowledge that your Personal Data will be processed and stored in our facilities and servers in the United States and may be disclosed to our service providers and affiliates in other jurisdictions. **We will transfer** your Personal Data to recipients outside of the EEA, Switzerland and the UK for the purposes described in this Privacy Policy. If you are based in the EEA, Switzerland or the UK and **your Personal Data is transferred to a third country, that third country may not offer the same level of data protection as your home country.**

# Automatická Vyjadřovna – využití AI

- Testujeme a trénujeme vlastní GPT modely a API
- Zvažujeme model RAG (nejen pro Vyjadřovnu)



# Výhody RAG architektury

- **Oddělení generativního modelu od datového úložiště**
  - RAG pracuje tak, že jazykový model (např. GPT) generuje odpovědi na základě relevantních informací získaných z externích zdrojů.
  - To umožňuje uchovávat **citlivá data v bezpečném** a izolovaném úložišti (např. lokální databáze nebo zabezpečený cloud), které není přímo součástí generativního modelu.
  - Citlivá data tak **nejsou "naučena" nebo pevně integrovaná do jazykového modelu**, čímž se snižuje riziko jejich úniku.

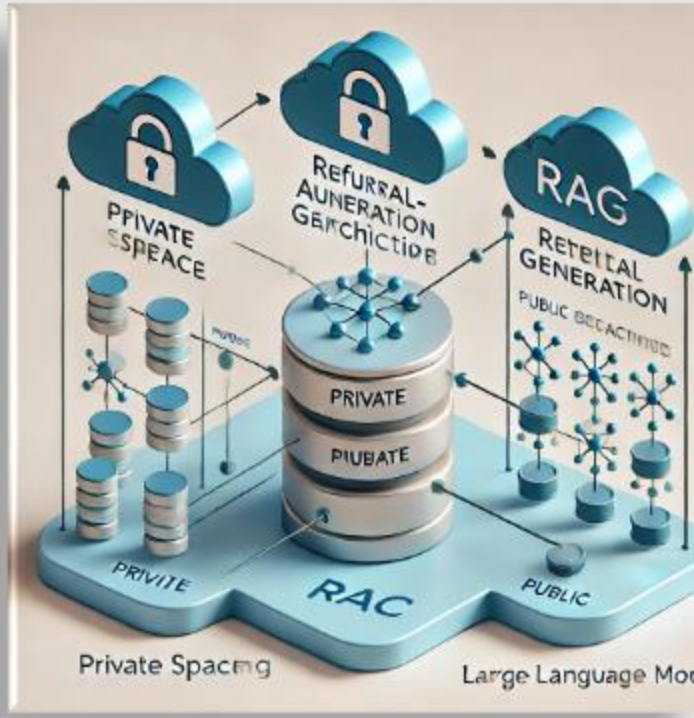

- **Minimalizace přenosu citlivých dat**
  - Jazykový model **nemusí** přistupovat k celému datasetu, ale pouze k malým a relevantním částem dat na vyžádání.
  - Tento přístup snižuje pravděpodobnost přenosu nebo **zpracování nepotřebných citlivých dat**, což je důležité z pohledu ochrany osobních údajů (např. GDPR nebo HIPAA).

# Výhody RAG architektury

- **Flexibilita a aktualizovatelnost dat**
  - Externí databázi nebo znalostní bázi lze **průběžně aktualizovat**, aniž by bylo nutné znovu trénovat jazykový model.
  - To je klíčové pro zpracování citlivých dat, protože umožňuje správu přístupových práv a **aktualizaci obsahu v reálném čase**.

- **Lepší kontrola nad odpověďmi**
  - Organizace má **větší kontrolu** nad tím, jaké informace jsou poskytovány. To je zvláště důležité při práci s citlivými nebo regulovanými daty.
  - Organizace může implementovat filtry nebo ověření, aby zajistila, že **pouze relevantní a povolené informace** budou použity při generování odpovědí.

- **Auditovatelnost a transparentnost**
  - RAG umožňuje **zpětně sledovat**, které zdroje dat byly použity při generování odpovědí. Tato auditovatelnost je zásadní pro zajištění souladu s právními a etickými požadavky při práci s citlivými daty.



# Jak soutěžit IS s podporou AI

- **Rozdíly vůči dodávce/vývoji tradičních IS**
  - Velmi rychlé změny technologií v čase (např. nové verze LLM, růst velikosti modelů, rychlý pokles nároků na HW)
  - Závislost na konkrétní architektuře a modelu, vendor-lock vůči třetí osobě
  - Dodatečné blokace některých typů modelů a architektur – státní regulace (EU AI Act, licence, auto-cenzura)

- **Agilní vývoj**
  - Možnost rychlé reakce na změny podmínek, transparentní náklady
  - Velmi důležitá analytická práce, řízení vývoje, vhodné definice milníků
  - Soutěžíme vývojový tým, jeho kvalitu (zkušenost, znalosti), případně cenu za MD

- **Value-based zakázkový model**
  - Skandinávský vzor postavený na definici skutečné přidané hodnoty řešení
  - Definice pomocí KPI; model bonusů a malusů
  - Možnost úpravy KPI v čase
  - Nutné dohoda obou (případně i více) stran v kontraktu

# Kontakt

## Krajský úřad Kraje Vysočina – odbor informatiky:

**Žižkova 57, Jihlava 587 33**

**www.kr-vysocina.cz/it**

**www.kr-vysocina.cz/ict - v angličtině**

**Vedoucí odboru IT – Ing. Petr Pavlinec**

**pavlinec.p@kr-vysocina.cz, tel.: 564 602 114**