OKAI-302

Privátní AI pod kontrolou

Popis:

Tento intenzivní jednodenní kurz vás naučí, jak si postavit a provozovat vlastní „soukromou AI“ přímo ve vaší firmě nebo domácí kanceláři. Ukážeme si, že k provozu moderních jazykových modelů nepotřebujete drahé serverové farmy – díky chytré optimalizaci a nástroji **Ollama** využijeme výkon běžných procesorů a paměti RAM k dosažení plynulého a bezpečného provozu.

Během kurzu si prakticky vybudujete kompletní AI ekosystém. Od instalace v prostředí **Docker**, přes nasazení profesionálního uživatelského rozhraní **Open WebUI**, až po pokročilé techniky, jako je chatování s vlastními PDF dokumenty (**RAG**), aniž by jediný bajt dat opustil lokální síť. Na závěr se naučíte, jak k této privátní AI bezpečně přistupovat i z terénu pomocí technologie **Cloudflare Tunnels**.

Absolvent kurzu bude umět:

- Nainstalovat a spravovat lokální LLM modely pomocí **Ollama**.
- Provozovat profesionální chatovací rozhraní v **Dockeru**.
- Zprovoznit lokální **RAG systém** pro chatování nad vlastními dokumenty.
- Bezpečně vystavit lokální službu do internetu pomocí **Cloudflare Tunnels**.

Požadavky pro absolvování kurzu:

- Základní orientace v příkazové řádce (příkazy typu cd, ls, docker run).
- Základní povědomí o operačním systému Ubuntu (jiný Linux).
- Analytické myšlení a základní znalost síťových konceptů (co je to IP, port, API).
- *Poznámka:* Není nutné umět programovat, stačí nebát se konfiguračních souborů.

Kurz určen pro:

Tento kurz je určen primárně pro ty, kteří vnímají umělou inteligenci jako strategický nástroj, ale odmítají slevit z přísných požadavků na kybernetickou bezpečnost a suverenitu svých dat. Ideálními účastníky jsou:

- IT administrátoři a systémoví inženýři, kteří chtějí do firemní infrastruktury nasadit robustní řešení AI bez závislosti na cloudových poskytovatelích a měsíčních poplatcích.
- Bezpečnostní manažeři a specialisté na ochranu dat (DPO), kteří hledají technologickou cestu, jak moderní AI nástroje zpřístupnit kolegům v souladu s nejpřísnějšími vnitřními předpisy a legislativou (např. GDPR, bankovní nebo advokátní tajemství).
- Profesionálové z vysoce regulovaných odvětví, jako je právní sféra, zdravotnictví, finanční sektor nebo státní správa, kde je práce s citlivými informacemi každodenním standardem a jejich únik do veřejných cloudů představuje nepřípustné riziko.
- Techničtí manažeři a vizionáři, kteří hledají alternativy k uzavřeným ekosystémům a chtějí mít plnou kontrolu nad výkonem, dostupností a výběrem modelů, které jejich firma používá.

Literatura:

Všichni účastníci školení obdrží materiály společnosti OKsystem.

Technické vybavení:

Všechny učebny jsou vybaveny nadstandardními počítači připojenými k Internetu, učebny jsou prostorné, klimatizované, bezbariérové a s připojením na Wi-Fi. V případě zájmu lze školení absolvovat online live.

Osnova:

1. blok: Architektura a lokální ekosystém (Základy)

- **Proč lokální AI?** Soukromí, nezávislost na předplatném, práce s citlivými firemními daty offline.
- **Ollama – motor pro modely:** Instalace a správa modelů přes příkazovou řádku. Jak funguje knihovna modelů (Llama 3, Phi-3, Mistral).
- **Docker jako standard:** Proč používáme kontejnery pro AI aplikace a jak zajistit, aby vše běželo izolovaně a čistě.
- **Praxe:** Zprovoznění prvního modelu v terminálu a pochopení parametrů (System Prompts).

2. blok: Open WebUI – rozhraní pro AI

- **Instalace Open WebUI:** Nasazení moderního rozhraní (přes Docker), které vypadá a funguje jako ChatGPT, ale běží u vás v učebně.
- **Propojení s Ollama:** Konfigurace backendu a nastavení víceuživatelského přístupu.
- **Správa modelů z UI:** Stahování, mazání a přepínání modelů bez psaní kódu.
- **Praxe:** Vytvoření uživatelských účtů a první interakce v prohlížeči.

3. blok: Hybridní AI a RAG (Vaše data v AI)

- **Externí modely:** Jak do Open WebUI připojit „bleskové“ externí API (např. OpenAI a jak vybrat správnou variantu) a mít vše v jednom okně.
- **RAG (Retrieval-Augmented Generation):** Nahrávání vlastních PDF a dokumentů do lokální databáze. Jak nechat AI odpovídat pouze na základě vašich soukromých souborů.
- **Vektory a Embeddingy:** Jednoduché vysvětlení, jak AI „čte“ vaše dokumenty, aniž by se je učila (soukromí zůstává zachováno).
- **Praxe:** Analýza interní směrnice v lokálním offline modelu.

4. blok: Cloudflare a bezpečné sdílení (Connectivity)

- **Cloudflare Tunnel:** Jak bezpečně zpřístupnit vaši lokální AI z internetu (např. z mobilu) bez nutnosti otevírat porty na routeru nebo mít veřejnou IP.
- **Zero Trust:** Základy zabezpečení – jak zařídit, aby k vaší AI mohl jen autorizovaný uživatel (e-mailová verifikace).
- **Hardware limity a optimalizace:** Jak z počítače vytáhnout maximum pomocí kvantizace (GGUF formáty).
- **Shrnutí:** Vytvoření „soukromého ChatGPT“, který je dostupný odkudkoliv, ale data nikdy neopustí váš počítač (nebo firmu).

Kontaktujte nás
OKsystem a.s., Na Pankráci 1690/125, 140 00 Praha 4
(+420) 236 072 111 skoleni@oksystem.cz www.okskoleni.cz

OKškolení
by OKsystem