SOITRON

Ako využívať AI bezpečným spôsobom

BRUNCH: KÁVA, KOLÁČ A KAPITOLA “AI”

Košice, Bratislava

Prečo vôbec riešiť „bezpečnú AI“ v podniku?

AI dnes používajú:

- zamestnanci (ChatGPT, Copilot, Gemini...)
- oddelenia (HR, marketing, IT, právne)
- vedenie (reporty, analýzy, rozhodovanie)

Riziká bez pravidiel:

- únik citlivých dát
- porušenie zmluvných a právnych povinností (GDPR, NDA)
- reputačné škody (halucinácie, predsudky, diskriminácia)

Únik interných a osobných údajov

Zamestnanec vloží do verejného AI nástroja:

- zmluvu s kľúčovým klientom
- zdrojový kód, architektúru systému
- internú finančnú tabuľku

Následky:

- porušenie NDA, GDPR
- údaje sa môžu použiť na trénovanie modelu
- riziko pri úniku zo strany poskytovateľa

Halucinácie a nepresnosti

AI odpoveď môže byť:

- presvedčivo napísaná, ale fakticky nesprávna
- čiastočne správna, ale doplnená vymyslenými detailmi (napr. neexistujúci paragraf zákona)

Následky:

- chybné rozhodnutia manažmentu
- právne chyby
- zavádzajúce informácie pre klientov

Predsudky (bias) a diskriminácia

Modely sú trénované na dátach z internetu

- AI LLM vedia preberať stereotypy

Následky:

- diskriminácia v HR (výber kandidátov)
- nevyvážené marketingové kampane
- nespravodlivé automatizované rozhodovanie

Ako na to?

Klasifikácia informácií
Zakázané typy údajov
Špecifiká pre HR, FIN, Marketing, IT, Obchod

Pravidlá

Vybrané Enterprise AI

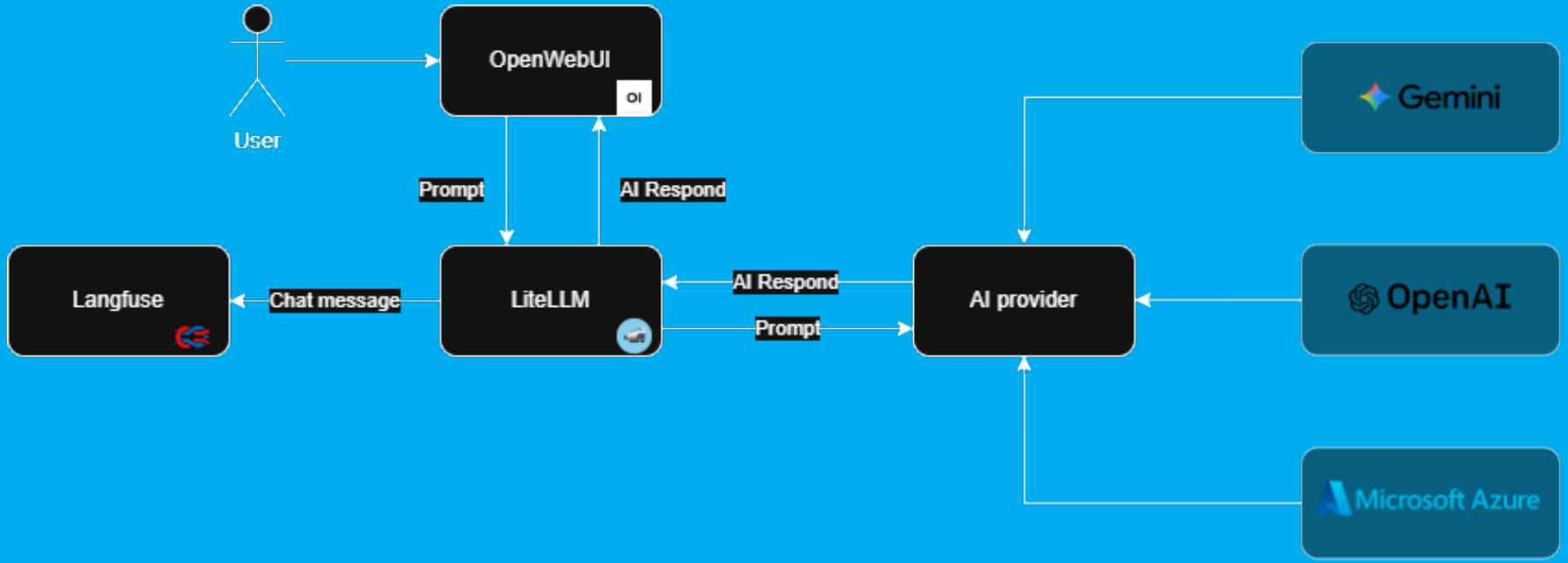
Voľba dôveryhodných partnerov pre enterprise AI služby
Vytvorenie ekosystému pre citlivé interné údaje
Zablokovanie ostatných možností
Založené na dôvere

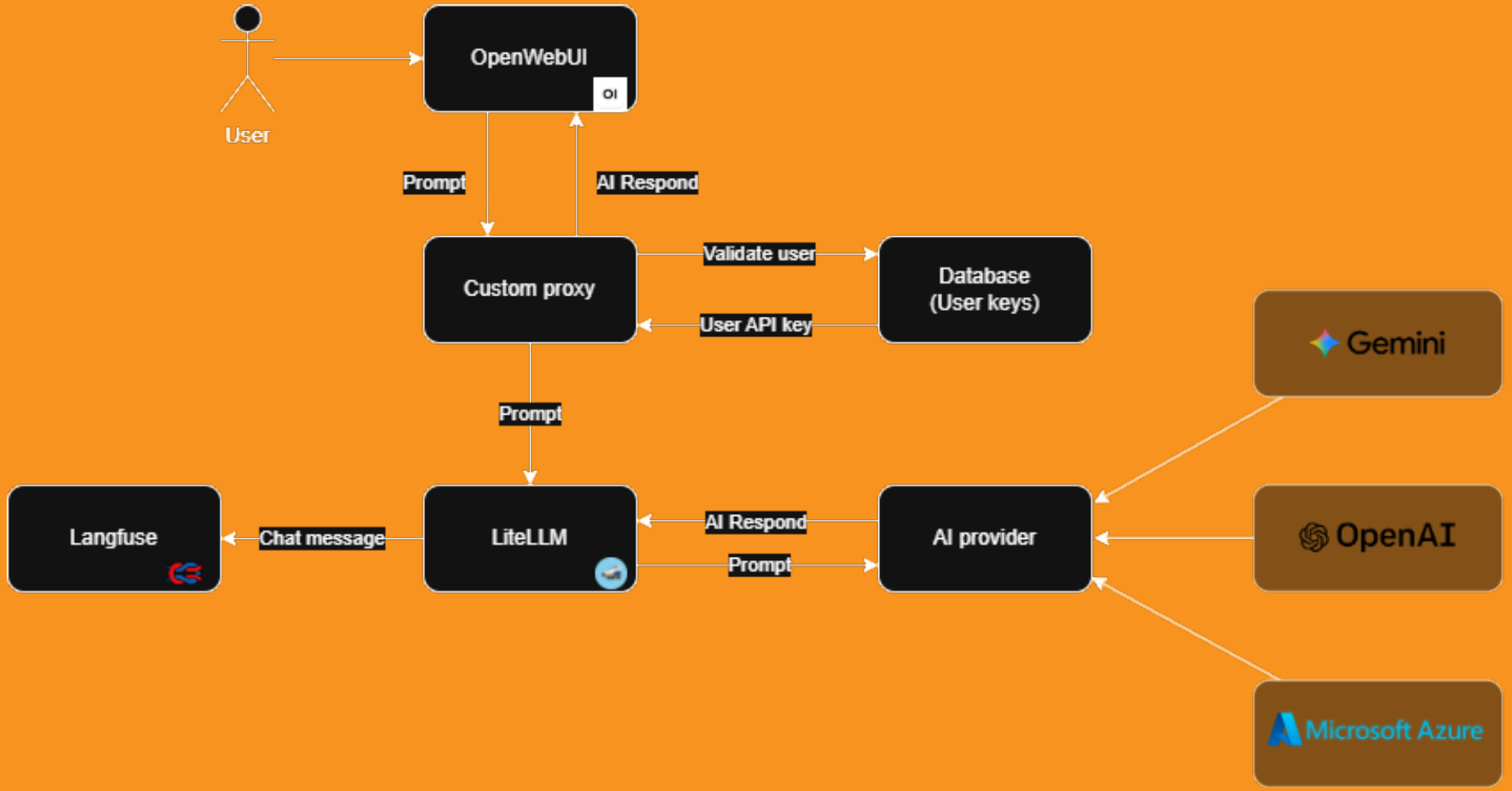
Orchestrácia enterprise aj public modelov
Optimalizácia nákladov
Maskovanie citlivých dát

AI Proxy

AI vo vlastnej réžii

Vysoké náklady na compute
Obmedzený výber modelov
Dáta plne pod kontrolou
Model do značnej miery pod kontrolou (trénovanie, konfigurácia)


User
OpenWebUI
Prompt
AI Respond
Langfuse
Chat message
LiteLLM
AI Respond
Prompt
AI provider
Gemini
OpenAI
Microsoft Azure


User
OpenWebUI
Prompt
AI Respond
Custom proxy
Validate user
Database
(User keys)
User API key
Prompt
Gemini
Langfuse
Chat message
LiteLLM
AI Respond
AI provider
Prompt
OpenAI
Microsoft Azure

Ako zamedziť využívaniu iných AI služieb?

AI policy
Informovanosť
Vzdelávanie

Pravidlá

Firemné počítače

Blokovanie URL + AI category na proxy/firewalle
DNS filtering
Always-on-VPN / ZTNA
DLP

MDM - blokovanie
- AI apps
- AI pluginov
- AI URLs

Firemné mobily

BYOD

MAM – mobile app mgmt
VDI – prístup z domácich PC

	Bezpečnosť	Kontrola	Flexibilita	Náklady	Komplexita	Vhodné pre
Pravidlá	Nízka	Veľmi obmedzená	Vysoká	Nízke	Nízka	Malé firmy Nízke nároky na bezpečnosť
Enterprise AI	Vysoká	Vysoká	Stredná	Stredné až vyššie	Stredná	Vyššie nároky na bezpečnosť a kontrolu
AI Proxy	Vysoká	Veľmi vysoká	Vysoká	Stredné až vyššie	Vysoká	Väčšie firmy s nárokmi na flexibilitu
Vlastná AI	Veľmi vysoká	Veľmi vysoká	Technicky náročná	Stredné až vyššie	Veľmi vysoká	Korporácie Firmy s AI produktmi

AKO ĎALEJ?

„Najväčšie riziko nie je AI, ale firma, ktorá ju používa bez pravidiel.“

- AI nezakazujte – nastavte pravidlá
  - Definujte, čo do AI môže a čo nie a ľudí to naučte.
- Začnite jednoducho, potom pridávajte komplexitu
  - Najprv politika a školenie, potom enterprise nástroje / proxy / vlastné riešenia.
- AI je poradca, nie rozhodca
  - Dôležité rozhodnutia a kontrola výstupov musia zostať na človeku (aspoň zatiaľ).
- Bezpečné používanie AI je konkurenčná výhoda.
  - Firma, ktorá s ňou vie pracovať, bude rýchlejšia, ale aj dôveryhodnejšia

SOITRON*