

GEOPOLITICKÁ RIZIKA AI MODELŮ

MIROSLAV NEČAS

DILEMA: SUPERIORITA VS SUVERENITA

bez AI nebudeme konkurenceschopní, AI nás učiní závislými

Model

hodnoty,
trénovací kriteria,
použitá data.

Hardware

85 % AI čipů
používaých v Evropě
dodala NVIDIA

Poskytovatel

může službu vypnout,
omezit nebo změnit
na pokyn státu

Zdroje: Ekonomický deník, The Next Web, Institute for Security and Technology

ZÁVISLOST NA POSKYTOVATELI

Otevřené LLM lze provozovat na vlastním hardware

128 GB

na stole

Modely do zhruba 120 mld. parametrů. Pilot a malý tým, příkon asi 140 W

2,1 TB

v serverovně

Modely do zhruba 800 mld. parametrů. Provoz jednoho úřadu, asi 14 kW

20 TB

v datacentru

Bilionové modely a sdílená služba pro více institucí, asi 140 kW

Referenční hardware NVIDIA (DGX Spark, DGX B300 a NVL72), limit velikosti modelu je odhadnut pro 4bitovou kvantizaci

JAK DOBRÉ JSOU OTEVŘENÉ MODELY?

Pro většinu agend to bohatě stačí

4 měsíce

Odstup nejlepšího otevřeného modelu v kybernetických úlohách za LLM špičkou podle CAISI

25 bodů

Odstup nejlepšího otevřeného modelu za nejlepším uzavřeným škála 0-100 v žebříčku BenchLM

Zdroje: AI Policy Bulletin k hodnocení GLM-5.3 (září 2026), BenchLM (13. 11. 2026)

MODELY KTERÉ STOJÍ ZA POZORNOST

Původ	Na stůl	Na server	Do datacentra
USA	gpt-oss-120b Gemma 4 Nemotron Super	Muse Spark Open Nemotron Ultra	Muse Spark Open
Evropa a západ	Mistral Small 4 EuroLLM Apertus	Mistral Large 3 Cohere Command A+	Mistral Large 4
Čína	Qwen3.8-27B	GLM-5.3 DeepSeek V4.1 Flash	Kimi K3.1 Qwen3.8-Max

GEOPOLITICKÁ RIZIKA MODELŮ

USA

Cloud Act, uzavřenost, bias, pochlebování, cena

nejvyšší celkový výkon, nejlepší guardrails, komerční ekosystém

EU

výkon a kapacita, bias, pochlebování, ekosystém

EU jurisdikce a kompetence, otevřené váhy, AI Act compliance

ČÍNA

cenzura, slabé guardrails, sankce, reputační riziko

nejvyšší výkon open weight, rychlý vývoj, open ekosystém

BEZ AI

nižší efektivita kognitivních úloh (15-40%) zaostávání a odliv talentů, shadow AI

zachování lidských kompetencí, bez nákladů na infrastrukturu a tokeny, bezpečnost a kontrola

Zdroje: Manhattan Institute k hodnocení CAISI, ITPro, arXiv 2609.35544, Brynjolfsson aj. 2025

OSM OTÁZEK NA KAŽDÝ MODEL

Původ a jurisdikce

Kdo může poskytovateli co nařídit?

Licence

Smíme model provozovat, upravit a sdílet? Kolik stojí?

Schopnosti a čeština

Ověřili jsme to na vlastních datech?

Nároky na hardware

Pojede to na našem hardware?

Odolnost proti jailbreak

Odmítne model škodlivý požadavek i pod tlakem?

Zaujatost a podlézavost

Je model objektivní a řekne nám to, co nechceme slyšet?

Transparentnost

Víme jak model myslí a na čem byl trénován?

Exit strategie

Lze model vyměnit bez přestavby systému?

Princip řízení rizik: žádný model není dokonalý, co je důležité je třeba určit podle plánovaného užití a rizik

MODELOVÉ SCÉNÁŘE

Strategické rozhodování

Vláda, zpravodajská služba, krizový štáb, armáda

Hlavní rizika: podlézavost modelu, klapky na očích

Čínský model: oponent v rámci multi-modelového systému

Asistent pro správní agendu

Třídění podání, shrnutí podnětu, návrh rozhodnutí, odůvodnění

Hlavní riziko: přehlednutí důkazu, halucinace, diskriminace, slabá lidská kontrola

Čínský model: ?

Operativní analytika

OSINT týmy, auditoři a vyšetřovatelé

Hlavní rizio: podstrčené pokyny v obsahu staženém z internetu, únik dat

Čínský model: rešerše v čínských a asijských zdrojích

Asistent pro občany

Dotazy na webu úřadu

Hlavní riziko: anonymní útočník, jailbreak modelu, exploitace modelu, ztráta reputace

Čínský model: ne

MULTI-MODELOVÁ ARCHITEKTURA

Rozpor mezi modely může být cennější signál než shoda

Hlavní model

Evropský, jediný s přístupem k neveřejným dokumentům

Kontrolní model

Jiný západní dodavatel, nezávislá druhá odpověď na totéž zadání

Čínský oponent

Oddělený perimetr, jen otevřené zdroje, ďáblův advokát

NEZAPOMÍNEJTE NA OBECNÁ RIZIKA

Ne/přímý prompt injection (web, e-mail, zpracovávaný dokument…)

Nadměrná autonomie a oprávnění agentů (přístup k datům a nástrojům)

Únik interních dat (prompty, logy, telemetrii, cache, RAG a nástroje)

Kompromitace dodavatelského řetězce (model, knihovna, plugin, aktualizace…)

Odcizení modelu, funkční replikace, rekonstrukce trénovacích sad

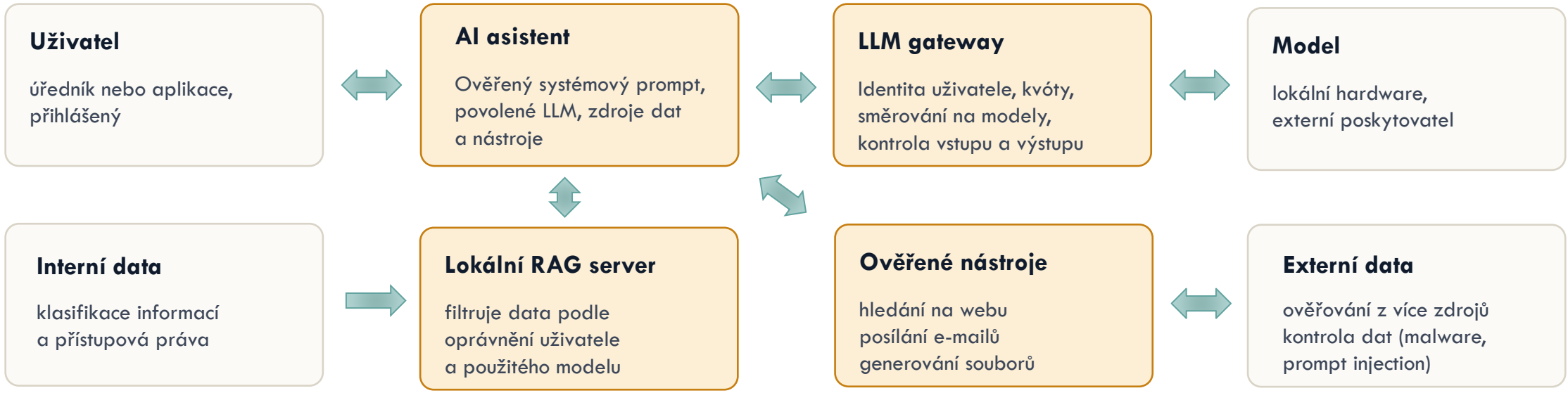
Otrávení modelu, znalostní báze nebo paměti asistenta

DoS a ekonomické útoky (token flooding, agentní smyčky – náklady/vyčerpání)

Lidský faktor (jailbreak, předpojatost, kognitivní disonance)

LINIE OBRANY: NEPOUŠTĚJTE MODEL PŘÍMO K UŽIVATELI, NÁSTROJŮM A DATŮM

AI dozor: řízení rizik, schválené užití, testy před uvedením do provozu, řízení změn, řešení incidentů, uchovávání záznamů



Analýza požadavků: uživatelské dotazy, prompty do LLM, inference (CoT modelu), rozhodnutí guardrails, sdílené dokumenty, nástroje

GEOPOLITIKCÁ RIZIKA NELZE IGNOROVAT, ALE MŮŽEME JE ŘÍDIT



Vícevrstvá architektura

Vlastní hardware, LLM gateway, RAG server a analýza logů

Pilotní provoz a testování

Testovací sady blízké reálným datům, testovací prompty a benchmarky

Řízení rizik

Zvažte rizika vyplývající z původu modelu v kontextu plánovaného užití

Ing. Miroslav Nečas, Ph.D.

miroslav.necas@ctu.gov.cz