

Vertex AI

Vertex AI je ucelená cloudová platforma společnosti Google pro strojové učení (Machine Learning), která je součástí **Google Cloud Platform (GCP)**. Jejím účelem je sjednotit všechny nástroje pro vývoj, trénování a nasazování modelů umělé inteligence do jednoho prostředí.

Umožňuje vývojářům využívat jak vlastní modely, tak i pokročilé modely od Googlu (např. [Gemini](#), PaLM 2, Imagen).

Hlavní pilíře platformy

Vertex AI se skládá z několika klíčových komponent, které pokrývají celý životní cyklus AI (tzv. **MLOps**):

1. Model Garden

Katalog hotových modelů. Uživatelé si zde mohou vybrat:

- **First-party modely:** Přimo od Googlu (celá rodina [Gemini](#)).
- **Open-source modely:** Například Llama od Mety nebo modely od Mistral AI.
- **Specifické modely:** Modely pro generování kódu, obrázků nebo analýzu genomu.

2. Generative AI Studio

Nástroj pro rychlé experimentování. Zde mohou vývojáři testovat prompty, nastavovat parametry modelu (teplota, top-k) a ladit odpovědi bez nutnosti psát složitý kód.

3. Vertex AI Pipelines

Nástroj pro automatizaci. Pomáhá spravovat celý proces od sběru dat, přes trénování na čipech TPU (vyvíjených ve spolupráci s [Broadcomem](#)), až po nasazení do produkce.

Klíčové funkce pro firmy

Funkce	Popis
Fine-tuning	Doškolení modelu na vlastních datech firmy (např. právní dokumenty nebo interní kód).
Grounding	Propojení modelu s reálnými daty (např. Google Search), aby se snížilo riziko halucinací.
Vector Search	Vyhledávání informací na základě významu, nikoliv jen klíčových slov.
AutoML	Automatizovaný vývoj modelů pro uživatele, kteří nejsou experty na datovou vědu.

Propojení s hardwarem

Vertex AI je unikátní tým, že běží na vlastní infrastruktuře Googlu. Využívá:

- **TPU (Tensor Processing Units):** Specializované čipy optimalizované pro hluboké učení, vyráběné v [TSMC](#).
- **GPU NVIDIA:** Pro specifické úlohy vyžadující architekturu CUDA.

Proč je Vertex AI důležitý?

Před příchodem Vertex AI musely firmy kombinovat desítky různých nástrojů. Vertex AI tento proces „demokratizuje“ - umožňuje i menším firmám využívat stejnou sílu, jakou má Google, a stavět na ní vlastní řešení se zárukou ochrany soukromí (data nahraná do Vertex AI se nepoužívají k trénování veřejných modelů Gemini).

— *Související pojmy:* [Gemini](#), [Hluboké učení](#), [Broadcom](#), [TSMC](#), [OpenAI](#)

From:

<https://serviceit.cz/> - IT ENCYKLOPEDIE

Permanent link:

https://serviceit.cz/doku.php?id=vertex_ai

Last update: **2026/01/05 19:34**

