

Hardware pro AI v roce 2026

Výběr hardwaru pro umělou inteligenci se v roce 2026 dělí na tři hlavní scénáře: lokální inference (běh modelů), trénování (fine-tuning) a edge computing (IoT a mobilní zařízení).

1. Grafické karty (GPU) - Srdce AI

GPU zůstávají dominantním prvkem díky své schopnosti paralelních výpočtů. V roce 2026 jsou standardem architektury **NVIDIA Blackwell (RTX 50/60)** a **AMD RDNA 4/5**.

Klíčové parametry pro AI:

- VRAM (Video RAM):** Nejdůležitější parametr. Pro běh kvalitních SLM je minimum 16 GB, pro střední modely 24 GB+.
- Tensor Cores:** Specializovaná jádra pro maticové operace. Verze 2026 podporují nativně formáty FP4 a INT4 pro extrémní zrychlení.
- Propustnost paměti:** Standardem je GDDR7 s rychlostí přes 1.5 TB/s.

2. NPU (Neural Processing Unit) - Nový standard v CPU

V roce 2026 už každý procesor (Intel Core Ultra 3, AMD Ryzen AI) obsahuje dedikovanou **NPU**.

- Účel:** Přebírá lehké AI úlohy (rozostření pozadí, rozpoznávání hlasu, predikce textu) od GPU, čímž drasticky snižuje spotřebu.
- Výkon:** Udává se v **TOPS** (Tera Operations Per Second). Standard pro „AI PC“ v roce 2026 je minimálně 50 TOPS.

3. Úložiště a RAM

Rychlost přenosu dat je kritická pro eliminaci „bottlenecku“ (úzkého hrdla).

Komponenta	Standard 2026	Význam pro AI
RAM	DDR5-8400+ / DDR6	Rychlé načítání vah modelů do paměti.
SSD	NVMe PCIe 5.0 / 6.0	Okamžitý start modelů a swapování dat.
Kapacita	64 GB+ RAM	Nutné pro běh více agentů a velký kontext (Context Window).

4. Specialized AI Hardware (ASICs)

Pro specifické nasazení se objevují specializované čipy:

- Groq LPU:** Extrémně rychlá inference pro textové modely s minimální latencí.
- Apple Silicon (M5/M6):** Díky unifikované paměti (Unified Memory) umožňují Macům používat

až 128 GB+ paměti jako VRAM, což je ideální pro obří modely.

5. Doporučené konfigurace pro rok 2026

Nadšenec / Vývojář SLM

- **CPU:** AMD Ryzen 9 nebo Intel Core i9 (s NPU 50+ TOPS).
- **GPU:** NVIDIA RTX 5090 (24 GB VRAM - GDDR7).
- **RAM:** 64 GB DDR5.
- **Disk:** 2TB NVMe PCIe 5.0.

Firemní AI stanice (Local Inference)

- **CPU:** Workstation grade (Threadripper / Xeon).
- **GPU:** 2x NVIDIA RTX A-series nebo RTX 6000 Ada.
- **Konektivita:** 10Gb Ethernet pro synchronizaci s firemními daty.

TIP: Pokud plánujete stavbu AI stanice, zaměřte se primárně na **kapacitu VRAM**. I slabší karta s více pamětí je pro AI často užitečnější než výkonná karta s malou pamětí.

— Související témata:

- [Průvodce Small Language Models](#)
- [Standardy PCIe 5.0 a 6.0](#)
- [Architektura čipsetů](#)

— **Autor:** @AI_Hardware_Guru **Poslední revize:** 04. 01. 2026

From:

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

Permanent link:

https://www.serviceit.cz/doku.php?id=it:hardware_pro_ai

Last update: **2026/01/04 16:55**

