Versie: februari 2026

Generative AI: Hardware & VRAM

Benodigde infrastructuur en grafisch geheugen

De motor van AI: GPU en VRAM.
Het draaien van AI-modellen vereist enorme rekenkracht, voornamelijk geleverd door grafische kaarten (GPU's). De beperkende factor is vaak het grafisch geheugen (VRAM). Hieronder een overzicht van de hardware-eisen voor verschillende schalen van AI-modellen.
Let op: Waarden variëren sterk per model en kwantisatie. Deze cijfers zijn indicatieve gemiddelden.

Lokaal / Klein

Gemini Nano, Mistral 7B

8 - 16 GB

Past op laptop/smartphone

Pro / Lokaal

Llama 3 70B, Mixtral

48 - 80 GB

1 à 2 Professionele GPU's

Big Tech (Cloud)

GPT-4, Claude 3, Gemini

> 1000 GB

Cluster van supercomputers

Bronnen: Hardware specs (NVIDIA/Meta), Mistral AI, OpenAI.
Schattingen gebaseerd op technische specificaties van gangbare LLM's.