FREE & OPEN SOURCE

VRAM Budget

Plug in your GPU. See what actually fits.

Pick your graphics card and see which language models fit in its memory, with the math shown: weights, KV cache, overhead, and a safety margin.



01 The real math
VRAM ≈ parameters × bits ÷ 8, plus cache and overhead.

02 Before you download
Know what runs before pulling a 40 GB model.

03 No sign-up
Open it and type in your card.

WHO IT'S FOR
Anyone running AI on their own computer or server.

PRICING
Free
Free to use

GOOD TO KNOW
- Made in Texas