DAY 53 · Part 5 · Looking Inside

How did graphics cards made for gaming become the heart of AI? — GPUs

Five-Minute AI Lessons · Oct 29, 2026

AI didn't pick graphics cards. The shape of the math was the same all along.

In a nutshell

1. A graphics card (GPU) is a part built to do **a huge number of easy calculations all at once**. To color in the millions of dots on a game screen at the same time, that was the only way.

2. It became AI's heart because AI training is exactly that kind of math. A card widely used for training packs **about 80 billion** tiny switches inside.

3. That one card does about 1,000 trillion calculations a second. Even that isn't enough, so Llama 3 was made by tying **16,000 cards** together into one giant cluster and running them for 54 days.

In one picture

A few smart people vs thousands of simple ones

CPU · a few smart people

· Finishes hard jobs on its own
· A few things at a time, one after another
· Great at tasks with complicated rules

→

GPU · thousands of simple people

· Only good at easy multiplying and adding
· Thousands of things all at once
· Perfect for drawing screens and AI math

AI training isn't about solving one hard problem. It's about solving trillions of easy multiplications at the same time.

Key idea: AI didn't pick graphics cards, the shape of the math was the same all along

Getting a feel for graphics cards by order of magnitude

About 80 billion

Switches in one card

Switches in one card **About 80 billion**

One card's math per second About 1,000 trillion

Cards used for Llama 3 16,000

Cards at one Memphis site 100,000

A single card called the H100 has 80 billion. That's like fitting ten times Earth's population...

About 1,000 trillion

One card's math per second

Switches in one card About 80 billion

One card's math per second **About 1,000 trillion**

Cards used for Llama 3 16,000

Cards at one Memphis site 100,000

If a person did one calculation every second, catching up with just one second of this card wou...

16,000

Cards used for Llama 3

Switches in one card About 80 billion

One card's math per second About 1,000 trillion

Cards used for Llama 3 **16,000**

Cards at one Memphis site 100,000

It ran for 54 days, and in that time it stopped 466 times. Even so, they reported putting more...

Use the slider to guess roughly what power of 10 it is, then tap to check the answer. The goal isn't to nail it...

Try it yourself in the web lesson · next-token.net/en/learn/day-53/

Quick check

Just three questions

Q1

Why are graphics cards such a good fit for AI math?

① They do tons of easy calculations all at once
② They solve hard problems all on their own
③ They have fast internet
④ They have huge storage space

Q2

About how many tiny switches are in one card used for AI training?

① About 8,000
② About 80 billion
③ About 800,000
④ About 80 quadrillion

Q3

How many cards were tied together to train Llama 3?

① 16 cards
② 160 cards
③ 16,000 cards
④ 1,600 cards

Know the why, too

Q1

① They were built that way to color in millions of dots on a game screen at the same time. AI training is also trillions of easy multiplications at the same time, so the shape w...

Q2

② The H100 has 80 billion. That one card does about 1,000 trillion calculations a second.

Q3

③ 16,000 cards were tied together into one cluster and ran for 54 days. Even with each card doing 1,000 trillion calculations a second, one big model needs this many.

AI didn't pick graphics cards.
The shape of the math was the same all along.

Tomorrow: Data centers and electricity — how much power does one building full of those cards eat up?