Beginner · short, with analogies

DAY 01–10 · Part 1 · What AI Is

Part 1 Review — What AI Is

All ten days of Part 1 at a glance · short, with analogies

No test — a review to read through

Pick a piece, add it, pick again

1. AI doesn't make up a whole sentence at once. It picks just the next piece, adds it, then picks again.

2. Repeat that pick hundreds of times, and you get the long answer we read.

3. So AI isn't fetching an answer stored somewhere — it builds a new one every time.

That piece has a name: token

1. The piece AI picks has a name. It's called a token.

2. A token isn't a letter or a word — common chunks stay whole, and unfamiliar ones get chopped up small.

3. To you, 'What's for lunch today?' is four words. To an AI, it's about six tokens.

The probability table and the temperature knob

1. To pick the next piece, AI scores every candidate, then holds a raffle where higher scores get more tickets.

2. Temperature is the knob that sets how adventurous that raffle gets.

3. Turn it down and #1 wins every time; turn it up and lower candidates pop up more often.

Training is tiny fixes, repeated

1. AI training isn't a sudden 'aha' — it's 'tiny fixes,' repeated.

2. Hide part of a sentence, have it guess the next piece, then nudge the numbers inside by how much it missed.

3. Repeat that an enormous number of times, and grammar and common sense soak into those numbers.

Parameters: how many numbers get tuned

1. How many numbers training tunes — that's the parameter count.

2. The more there are, the bigger the container for patterns.

3. Today's big models run from tens of billions to around 1 trillion — not a number you could ever count.

Confidently making things up — hallucination

1. When AI calmly makes up facts that don't exist, it's called hallucination.

2. It isn't trying to lie. It comes naturally with a setup that picks a plausible next piece.

3. AI has no separate built-in way to flag 'I don't know this.'

Not remembering — rereading

1. AI doesn't save your chat. Every time, it rereads the whole conversation so far, then answers.

2. There's a limit to how much it can read at once. That limit is called the context length.

3. As the chat gets longer, the early part gets pushed out of the window.

New question, new probability table

1. A prompt is the whole input you hand to AI.
2. Change the words, and the probability table for the next piece changes too — so the answer changes a lot.
3. You're not changing the AI — you're changing the question.

Size means parameter count

1. A 'big' model just means it has lots of parameters — the ones from DAY 05.

2. Big models know a lot and can handle hard problems, but they're slow, pricey and only run on servers.

3. Small models fit on your own device, so they're fast and work even without internet.

Where the core lives

1. A closed model's core stays on the company's servers. We can only talk to it through the front desk.

2. With an open model, you can download that core — the weights — in full and run it on your own computer.

3. You ask questions the same way — the difference is where the model lives.

Ten days, one story.
A machine that picks the next piece.



In Part 2: where this thing came from — all the way back to the neuron that learned on its own, in 1958.