DAY 14 · Part 2 · How We Got Here

The words in a sentence started looking at each other — 2017: the Transformer

Five-Minute AI Lessons · Sep 20, 2026

Once words could look at each other, AI started understanding whole sentences.

In a nutshell

1. In 2017, one paper from a Google research team set the basic design of the AI you use today. The design is called the **Transformer**, and the paper's title is **'Attention Is All You Need'**.

2. Before that, AI read a sentence one word at a time, from the start. The Transformer let **every word in a sentence look at all the others at the same time** — that's called 'attention.'

3. Looking at everything at once means **it can all be calculated in one go**. That made it a perfect fit for graphics cards, and GPT, Claude and Gemini all grew up on top of this design.

In one picture

Reading in order vs. looking all at once

Before 2017 · in order

· Words read one by one, from the start
· Memory of early words slowly fades
· One at a time — you have to wait

→

Transformer · all at once

· Every word looks at every word
· Grabs far-away words right away
· All calculated at once — great for GPUs

Even reading the same sentence, the right side works out how every pair of words relates as one table, in one go.

Key idea: drop reading in order, let words look at each other — and AI got faster and smarter

What does 'it' point to? Watch the words look at each other

Focus on 'it'

...because it was too tired.

The animal didn't cross
the street because **it**
was too **tired.**

'Tired' is right there, so 'it' must be the animal. A street can't get tired. The Transformer figures this out...

Change the sentence

...because it was too wide.

The animal didn't cross
the street because **it**
was too **wide.**

Only one word changed ('wide'), but where 'it' looks jumped to the street. Reading one word at a time from the...

The old way

Word by word — early words fade

The animal didn't cross
the street because **it**
was too **wide.**

The old way read in order, so the longer the sentence got, the more the early words faded. That's why it often...

Pick a word, and the colors show which words in the sentence it looks at, and how hard. Bright means a hard loo...

Try it yourself in the web lesson · next-token.net/en/learn/day-14/

Quick check

Just three questions

Q1

What's the biggest difference between the Transformer and the way before it?

① Words in a sentence look at each other at once
② It memorizes words faster
③ It searches the internet to answer
④ People type in rules by hand

Q2

Why did 'looking at once' matter so much?

① It made sentences prettier
② All-at-once math is perfect for graphics cards
③ It used less memory
④ It was easier for people to read

Q3

In 'it was too tired,' how does attention know 'it' means the animal?

① It's the first word
② A dictionary says so
③ Scores how well words fit
④ It picks at random

Know the why, too

Q1

① Instead of reading in order from the start, every word looks at every other word at once. That's 'attention' — and it's why the paper was called 'Attention Is All You Need.'

Q2

② Reading in order, you have to finish one word before you can look at the next. Looking at once means one big calculation, and the graphics cards you met yesterday are great at...

Q3

③ Each word scores 'which words should I look at?' What goes with 'tired' is the animal, not the street, so the scores pile up there.

Once words could look at each other, AI started understanding whole sentences.



Tomorrow: 2020: GPT-3 — what happens when you grow this design to 175 billion numbers.