



How Transformers Changed AI Forever: The Story Behind Modern Language Models



Vipul Lohani

Follow

4 min read · Jul 21, 2025



Introduction

In 2017, a group of researchers published a paper with a bold title: “Attention Is All You Need.” This paper introduced something called the Transformer, and it completely changed how computers understand and generate human language. Today, every major AI system you’ve heard of — ChatGPT, Claude, BERT, and many others — uses this Transformer technology. It’s the secret ingredient that makes these AI systems so good at understanding what we say and responding in ways that feel natural and intelligent. Think of Transformers as the engine that powers modern AI. Just like how the invention of the wheel revolutionized transportation, Transformers revolutionized artificial intelligence.

What Made Transformers So Special?

The Old Way Was Slow

Before Transformers, AI systems that worked with text had to read one word at a time, just like how you might read a book from left to right. This made them slow and bad at understanding connections between words that were far apart in a sentence. Imagine trying to understand a mystery novel by only being allowed to read one word every hour, and you had to forget earlier words as you went along. That's basically how the old AI systems worked.

The Transformer's Breakthrough

The Transformer did something revolutionary: it could look at all the words in a sentence at the same time. Instead of reading "The cat sat on the mat" one word at a time, it could see the whole sentence instantly and understand how each word related to every other word. This was like giving someone the ability to see an entire page of text at once instead of looking through a tiny keyhole that only showed one word at a time.

How Transformers Actually Work

The Magic of Attention

The key innovation in Transformers is something called "attention." Think of attention like a spotlight that can focus on the most important parts of a sentence when trying to understand what it means. When you hear the sentence "The bank can guarantee deposits will eventually cover future tuition costs because the annual percentage rate..." your brain automatically knows that "bank" probably refers to a financial institution, not the side of a river. That's because you're paying attention to words like "deposits," "percentage rate," and "guarantee." Transformers do the same thing. They learn to pay attention to the right words that help them understand what's going on.

Multiple Spotlights Working Together

Transformers don't just use one spotlight — they use many spotlights at the same time, each looking for different things. One spotlight might focus on who's doing what in a sentence. Another might look for emotional tone. A third might track the timeline of events. This is called “multi-head attention,” and it's like having a team of specialists all working together to understand a piece of text.

Understanding Word Order

Since Transformers look at all words at once, they need a way to understand that “Dog bites man” is different from “Man bites dog.” They solve this by adding special markers that tell them the position of each word in the sentence. It's like numbering the words: “Dog¹ bites² man³” versus “Man¹ bites² dog³” so the system knows the order matters.

From Research Lab to Real World

The Big Discovery

Researchers discovered something amazing: the bigger they made these Transformer models, the smarter they became. It wasn't just that they got a little better — they started showing completely new abilities that smaller models didn't have. Small models could maybe complete simple sentences. Medium models could write paragraphs. But huge models started showing reasoning skills, creativity, and the ability to solve complex problems. This was like discovering that making a car engine bigger didn't just make it faster — it made it able to fly.

Get Vipul Lohani's stories in your inbox

Join Medium for free to get updates from this writer.

Enter your email

Subscribe

Training on the Internet

These large models are trained on massive amounts of text from the internet — books, articles, websites, and more. They learn patterns in how people write and communicate, which helps them understand and generate human-like text. Think of it like someone reading millions of books and conversations and gradually learning how language works, except the Transformer can do this much faster than any human ever could.

Beyond Just Text

Transformers Everywhere

The success of Transformers in language led researchers to try them in other areas, and they worked amazingly well there too:

- *Images*: Transformers can now look at pictures and understand what's in them, just like they understand text. They treat parts of an image like words in a sentence.
- *Videos*: Some of the latest AI systems use Transformers to create and understand videos, seeing how scenes change over time.
- *Science*: Researchers are using Transformers to understand DNA sequences, treating genetic code like a language that can be read and interpreted.

Why Transformers Work So Well

Speed and Efficiency

Because Transformers can process all words at once instead of one at a time, they can be trained much faster on powerful computers. This speed allowed researchers to build much larger and more capable models than ever before.

Learning Transfer

Transformers are great at taking what they learned from one task and applying it to another. A model trained to translate languages might also become good at writing poetry or answering questions, even without specific training on those tasks.

Understanding Context

Transformers excel at understanding context — they can keep track of what's being discussed over long conversations or documents. This is why modern AI assistants can have coherent conversations that stay on topic.

Conclusion

Transformers took AI from being a cool research project to being something that affects our daily lives. They solved fundamental problems with how computers understand language, and in doing so, they opened the door to the AI revolution we're experiencing today. Understanding Transformers helps us understand not just how current AI works, but also where it might go next. As these systems become more capable and more common, this understanding becomes valuable for everyone, not just computer scientists. The next time you use an AI writing assistant, language translator, or smart search engine, you'll know that there's a Transformer working behind the scenes — paying attention to your words, understanding the context, and generating responses that feel surprisingly human. That's the power of attention, and it really is all we needed to change the world of artificial intelligence forever.

Transformers

LLM

AI



Written by Vipul Lohani

1 follower · 4 following

Software Engineer at Intuit

Follow

No responses yet



Write a response

What are your thoughts?

More from Vipul Lohani