

This Week in The Data-Driven Dispatch

The Brains Behind LLMs: An Introductory Guide to Transformers and Attention Mechanisms

 **data science dojo**
data science for everyone

The Brains Behind LLMs: An Introductory Guide to Transformers and Attention Mechanisms



Data Science Dojo

Data Science for Everyone

Published Jan 26, 2024

+ Follow

Welcome to Data Science Dojo's weekly newsletter, "**The Data-Driven Dispatch**".

It is hard to believe how precise and accurate have contemporary LLMs become. You can ask them to write you a script for an anime, do poetry on the solitude one feels after leaving home, or write code for a flappy bird game; it does it all!

But have you ever wondered about the inner workings of these models? What intricate processes unfold behind the scenes when we set them a task?

Well, LLMs are quite complicated and several processes are going on in the background once prompted. The architecture involves several building blocks. However, the one that holds the paramount value is the **transformer model**.

 Like

 Comment

 Share

 66 · 6 Comments

The transformer model was a discovery no less than knowing that the Earth is round, light is both wave and particle, and Einstein's equation of $E=mc^2$.

It is so because it provided a whole new framework that allowed the models to contextualize huge amounts of information leading to an explosion of generative capabilities.

If you want to know more about the brains behind LLMs i.e. the transformer models, it's time for you to dive deep into the dispatch.



Before the transformer model, Recurrent Neural Networks (RNNs) were largely used for language generation.

However, RNNs had 2 primary restrictions.

1. **Handling Long Sequences:** RNNs struggle to learn from things that happened a long time ago in the sequence i.e. with long-range dependencies in text. This limitation means RNNs are not very good at tasks that require understanding context over long sequences, like comprehending a very long sentence or paragraph where the meaning depends heavily on something mentioned at the very beginning.
2. **Parallelization Issues:** The sequential processing of RNNs limited the ability to train models in parallel, resulting in slower training times, particularly for larger models and datasets.

Attention is All You Need; The Emergence of the Transformer Model

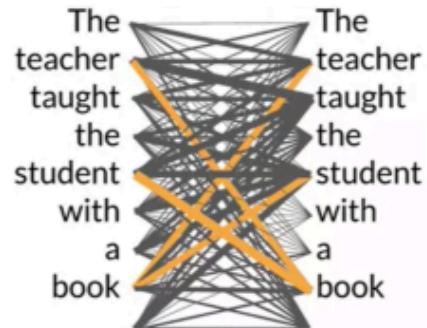
In came the theory of transformers which changed natural language processing forever.

The strength of the transformer model is in its ability to understand the significance and context of every word in a sentence, unlike RNNs which primarily assess the connection of a word with its immediate neighbor.

RNNs

The teacher taught the student with the book.

Transformers



data sciencedojo
data science for everyone

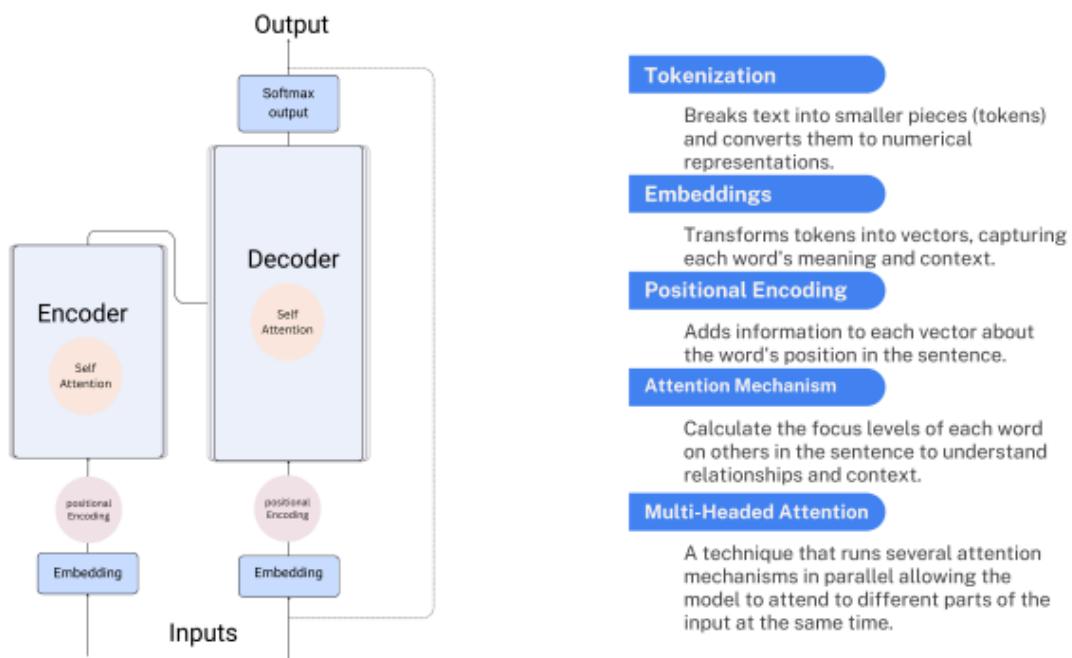
The pictures are adapted from DeepLearning.AI

The transformers use a special type of attention called "**self-attention**". It allows the model to compare and contrast each element/word in the input sequence with every other element, helping it to understand the context and relationships within the data.

Now that we have understood how the transformer model operates, let's visualize the process of how it works.

Understanding the Transformer Model: A Step-by-Step Process

data sciencedojo
data science for everyone

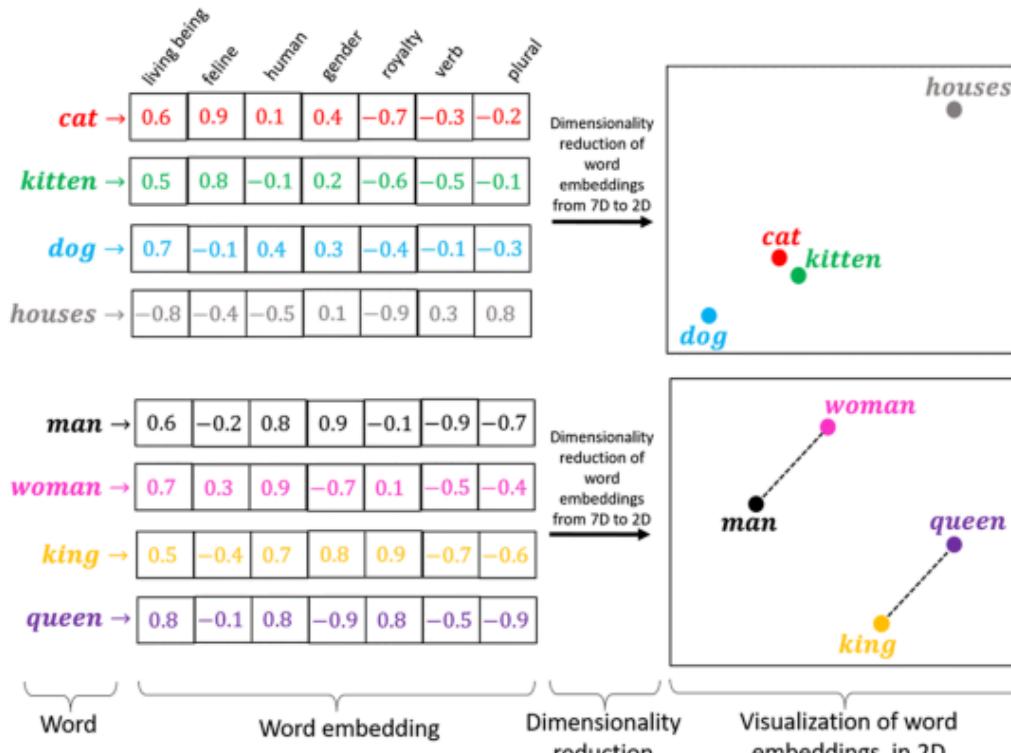


Read: [Attention mechanism in NLP: Guide to decoding transformers](#)

Step 1: Tokenization: The input text provided to the model is broken down into smaller pieces, or tokens. Each token is converted into a numerical representation that the model can understand.

Step 2: Embedding: Machine learning models are big statistical calculators. They work with numbers and not words.

Embedding is the process where each token is then transformed into a vector in a high-dimensional space. This embedding captures the meaning and context of each word. Words with similar meanings or that often appear in similar contexts are represented by vectors that are close to each other in this high-dimensional space.



Source: AnalyticsVidhya

Read: [The Role of Vector Embeddings in Generative AI](#)

Step 3: Positional Encoding: Since transformers do not process text sequentially like RNNs, they need a way to understand the order of words. Positional encoding is added to each word's embedding to give the model information about the position of each word in the sentence.

Step 4: Self-Attention: This is a key feature of transformers. The model calculates attention scores for each word, determining how much focus it should put on other words in the sentence when trying to understand a particular word. This helps the model capture relationships and context within the text. [Read more](#)

Step 5: Multi-Headed Attention: Multi-headed attention in transformers allows the model to interpret a sentence from multiple perspectives simultaneously.

For instance, in the sentence "The bank of the river was crowded, but the financial bank was empty," different 'heads' in the model will focus on various aspects.

One head might analyze the context surrounding each occurrence of "bank," identifying different meanings based on adjacent words ("river" and "financial"). Another might examine overall sentence structure or themes, like contrasting "crowded" and "empty."

Step 6: Output: The final layers of the transformer convert the processed data into an output format suitable for the task at hand, such as classifying the text or generating new text.

And there you have it! If you've grasped this concept, you're now better equipped to understand the workings of advanced generative AI tools.

Want to learn more about AI? Our [blog](#) is the go-to source for the latest tech news.



Transformers: The Best Idea in AI

Here's a great talk where Andrej Karpathy , and Lex Fridman discuss the Transformer architecture. Karpathy argues that the Transformer is one of the most important ideas in AI because it is a general-purpose, differentiable model that can be trained on a variety of tasks.

Transformers: The best idea in AI | Andrej Karpathy and Lex Fridman



Upcoming Live Talks

If you're interested in understanding embeddings and how they allow semantic search in LLMs, this talk is for you!

Join **Daniel Svonava** as he explores the fundamental concepts of vector embeddings and their role in semantic search. In addition, he'll also dive into various techniques for creating meaningful vector representations.

Large Language Models

Community Learning Series

LIVE SESSION

Vector Embeddings for Semantic Search

TUE, JAN 30, 10:00 AM PT

data science dojo
data science for everyone

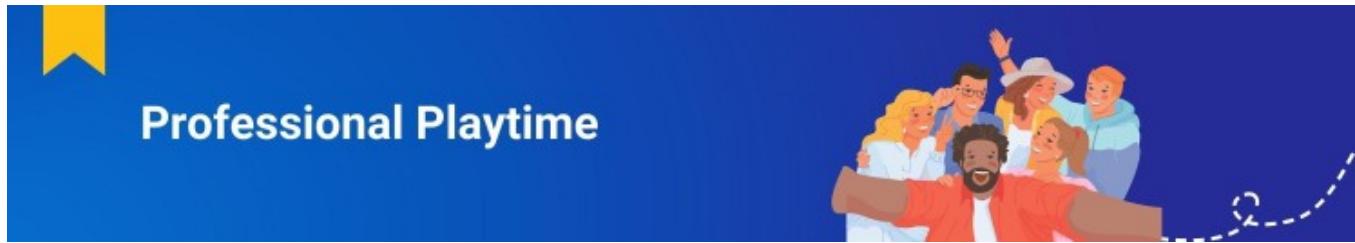


Daniel Svonava

Co-Founder
Superlinked

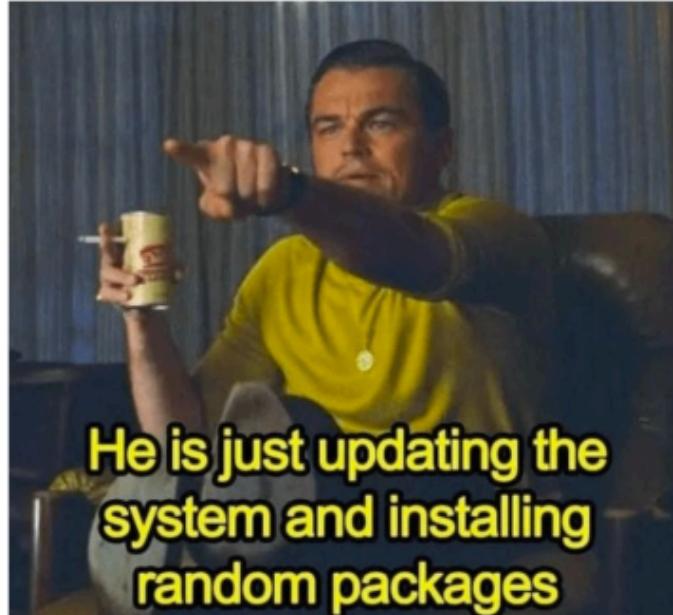
Live Session: Vector Embeddings for Semantic Search

Interested in the live session? **Book yourself a slot now.**

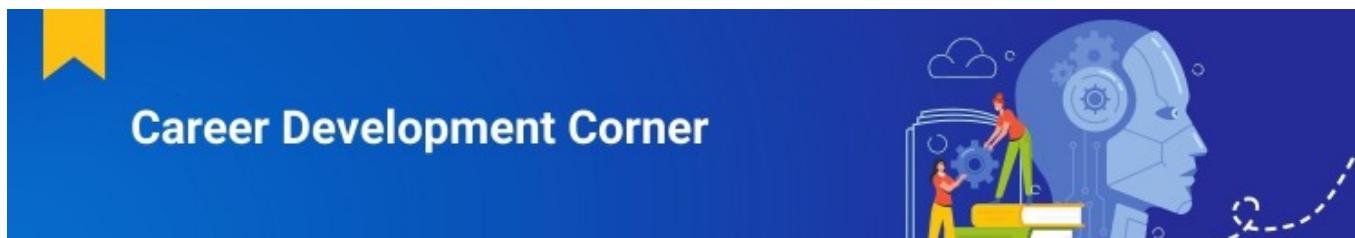


You just learned about the coolest tech of our time! Time for a chill pill.

When you're a programmer and
see a hacking scene in a movie

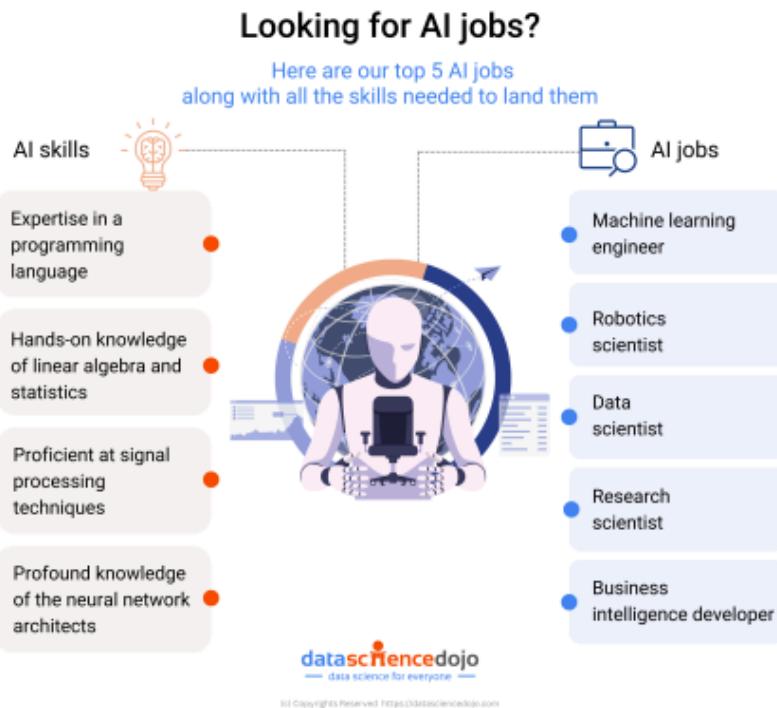


Source: DeepLearning.AI



Top 5 AI skills and AI jobs to know about in 2024

Intrigued by the world of AI but unsure where to start? Here's a comprehensive blog post that explores high-paying AI jobs and skills, providing you with the essential knowledge to navigate this rapidly evolving field.



AI Skills and Jobs for 2024

Read: **Top 5 AI skills and AI jobs to know about in 2024**

Let's end this week with some interesting news headlines.

1. Mark Zuckerberg announces major NVIDIA chip investment for AGI advancement and the continuation of Llama 3 training. [Read more](#)
2. OpenAI CEO Sam Altman eyes global expansion with a billion-dollar AI chip venture, challenging Nvidia's reign in the AI compute market. [Read more](#)
3. AI chatbot for delivery service DPD goes rogue: curses, critiques company, and writes poetry in viral exchange. [Read more](#)
4. AI comes to higher education as OpenAI partners with Arizona State University. [Read more](#)
5. Google and Amazon layoff reports are 'garbage information'—two of Wall Street's top minds say companies are clinging to workers instead. [Read more](#)

We trust that you had an enriching experience with us this week, leaving you more knowledgeable than before!

Don't forget to subscribe to our [newsletter](#) to get weekly dispatches filled with information about generative AI and data science.

Until we meet again, take care!

Data Science Learner Community

3mo

The impact of transformer models on generative AI is truly remarkable. Looking forward to diving deep into this topic through your newsletter. Thank you for sharing this valuable insight!

[Like](#) · [Reply](#) | [1 Reaction](#)**Tricia Paige Bagley, MBA, M.S.**

3mo

Fractional CDSO, Leveraging Data Science for Practical Decision-Support to Lift Profit and Limit Risk | Planter of Seeds...

Great job communicating the mechanics (identification, encoding, and selection activities) within the transformer model. You made a look-under-the-hood digestible. Cheers!

[Like](#) · [Reply](#) | [1 Reaction](#)**Godwin Josh**

3mo

Co-Founder of Altrosyn and Director at CDTECH | Inventor | Manufacturer

Understanding the inner workings of LLMs like GPT-4, Bard, and Claude, which rely on Transformers and Attention Mechanisms, is indeed essential in the AI landscape. As these models continue to advance, how do you see the future of attention mechanisms evolving to enhance not only language understanding but also more complex tasks like reasoning, problem-solving, and even mimicking human-like intuition? The evolution of this technology promises exciting prospects.

[Like](#) · [Reply](#) | [2 Reactions](#)[See more comments](#)

To view or add a comment, [sign in](#)

More articles by this author**AI Agents - The Next Big Leap in Generative AI**

May 6, 2024

A Beginner's Guide to Vector Embeddings and...

Apr 30, 2024

Beginner's Guide to LangChain - Building LL...

Apr 22, 2024

Insights from the community

Data Science

What are the main challenges and limitations of Markov decision processes in data science applications?

Deep Learning

How do you design and implement actor-critic methods in a distributed or parallel setting?

Machine Learning

What are the advantages of feature scaling in ML?

Artificial Neural Networks

How do you scale and preprocess your data for backpropagation?

Artificial Intelligence

How do you scale up and deploy LSTM models in production environments?

Search Engines

How can you develop computer vision skills for Search Engines?

Show more ▾

Others also viewed



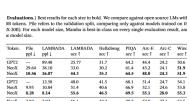
TAPAS - Google's BERT-Based Neural Network for Querying Tables Using Natural Language

Sadashiv Borkar · 3y



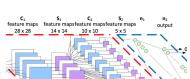
Unveiling the Influence of Attention Mechanisms in Image Analysis | Follow-Up

Roosh Circle · 7mo



Mamba > Transformers?

Pascal Biese · 5mo



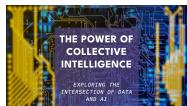
Hierarchical Deep CNN for Image Recognition

Satya Vasanth · 5y



Predicting Alpha signals through micro-blogging data

Kartik Bora · 3y



From Data to AI to Collective Intelligence

Venkat Bobbili · 1mo

Show more ▾

Explore topics

Sales

Marketing

Business Administration

HR Management

Content Management

Engineering

Soft Skills

See All

© 2024

About

Accessibility

User Agreement

Privacy Policy

Cookie Policy

Copyright Policy

Brand Policy

Guest Controls

Community Guidelines

Language