Member-only story

The Transformer Architecture From a Top View

Exploring the encoder-decoder magic in NLP behind LLMs

Dimitris Effrosynidis · Follow

Published in Towards AI · 7 min read · 3 days ago

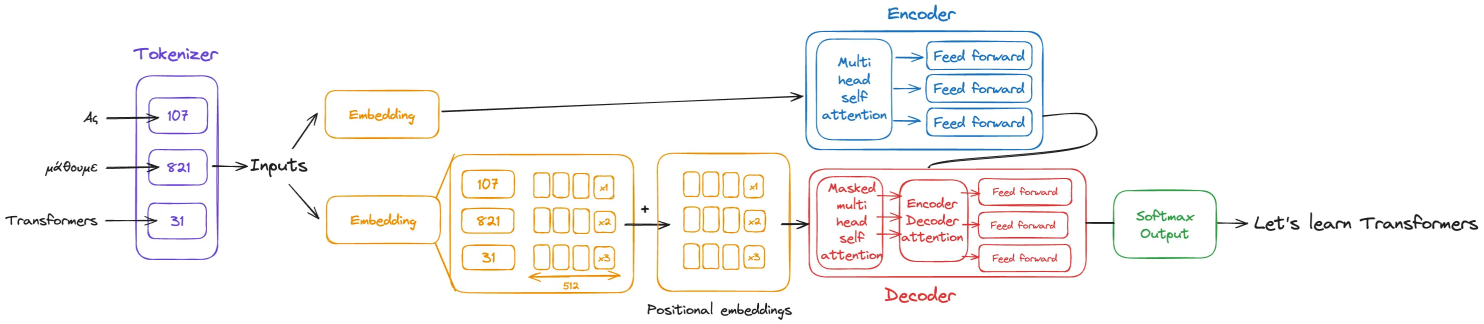


Image created by the author.

The state-of-the-art Natural Language Processing (NLP) models used to be Recurrent Neural Networks (RNN) among others.

And then came Transformers.

Transformer architecture significantly improved natural language task performance compared to earlier RNNs.

Developed by Vaswani et al. in their 2017 paper “Attention is All You Need,” Transformers revolutionized NLP by leveraging self-attention mechanisms, allowing the model to learn the relevance and context of all words in a sentence.

Unlike RNNs that process data sequentially, **Transformers analyze all parts of the sentence simultaneously.** This parallel processing capability allows Transformers to learn the context and relevance of each word about every other word in a sentence or document, overcoming limitations related to long-term dependency and computational efficiency found in RNNs.

But let’s explore the architecture step by step.

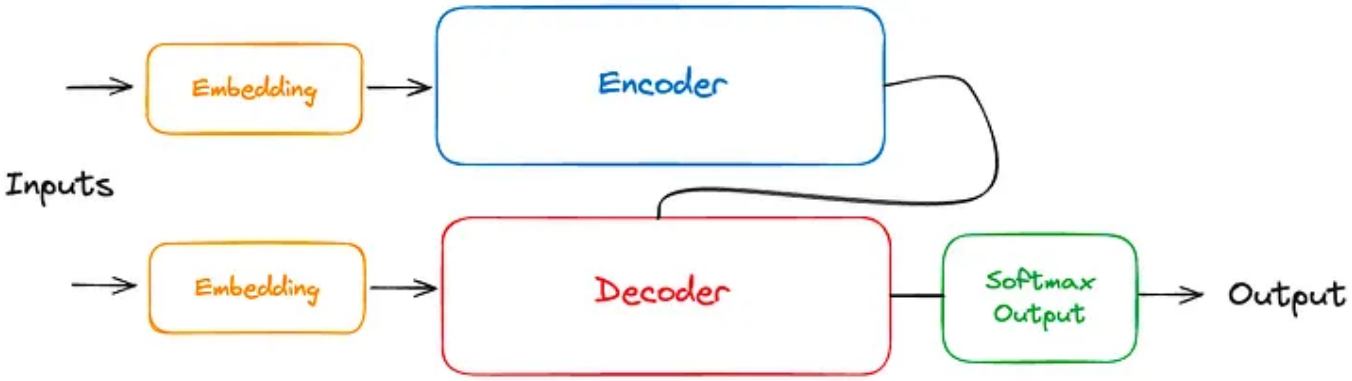


Image created by the author.

- There are two components in a Transformer Architecture: the Encoder and the Decoder.
- These components work in conjunction with each other and they share several similarities.

- **Encoder:** Converts an input sequence of tokens into a rich, continuous representation that captures the context of each token within the sequence. Its output is a sequence of embedding vectors, often called the hidden state or context.
- **Decoder:** Uses the encoder's hidden state to iteratively generate an output sequence of tokens, one token at a time.

Although both the Encoder and the Decoder exist in the Transformer

Written by Dimitris Effrosynidis

Follow

513 Followers · Writer for Towards AI

I read and learn every day. | Data Science • Personal Finance • Self-Improvement | Ph.D.
https://www.linkedin.com/in/dimitrios-effrosynidis/

More from Dimitris Effrosynidis and Towards AI

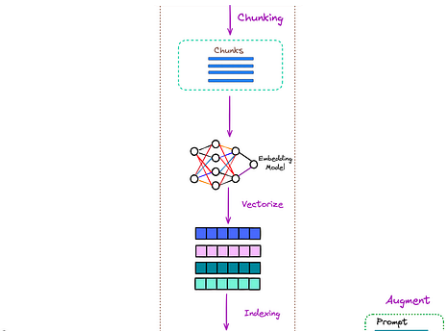
Dimitris Effrosynidis in DataDrivenInvestor

195 Data Science Libraries You Should Reconsider Using

It has been over a year since their last update

· 20 min read · Feb 3, 2024

106 2



Florian June in Towards AI

Advanced RAG 02: Unveiling PDF Parsing

Including key points, diagrams, and code

· 13 min read · Feb 3, 2024

1.1K 11

IVAN ILIN in Towards AI

Advanced RAG Techniques: an Illustrated Overview

A comprehensive study of the advanced retrieval augmented generation techniques...

19 min read · Dec 17, 2023

4.6K 29

Dimitris Effrosynidis in DataDrivenInvestor

Introduction to Hugging Face Datasets

A short tutorial with everything you need to know to get started

· 5 min read · Feb 8, 2024

35

See all from Dimitris Effrosynidis | See all from Towards AI

Recommended from Medium

Benedict Neo in bitgrit Data Science Publication

Roadmap to Learn AI in 2024

A free curriculum for hackers and programmers to learn AI

11 min read · 2 days ago

1.93K | 24

Ryota Kiuchi, Ph.D. in Towards Data Science

How OpenAI's Sora is Changing the Game: An Insight into Its Core...

A masterpiece of state of the art technologies

· 12 min read · 5 days ago

333 | 4

Lists

Predictive Modeling w/ Python

20 stories · 928 saves

Practical Guides to Machine Learning

10 stories · 1091 saves

Natural Language Processing
1223 stories · 702 saves

General Coding Knowledge
20 stories · 940 saves

Liu Zuo Lin in Level Up Coding

7 Things I Wish I Knew Earlier About Python Classes

Day 90 of experimenting with video content:

· 5 min read · 6 days ago

1K 6

Cris Velasquez

Automating Scientific Knowledge Retrieval with AI in Python

End-to-End Guide for Developing a Research Chatbot with OpenAI functions Capable of...

· 13 min read · 4 days ago

485

Tessa Rowan

How a Simple Countdown Timer Website Making $10,000 Per...

From Idea to Income

· 4 min read · Feb 10, 2024

BoredGeekSociety

Google Finally Did It! Gemini 1.5: Best AI Model! Everything you nee...

Gemini 1.5 might be the tipping point changing the AI game in 2024! Since the...

· 6 min read · 5 days ago