8.4

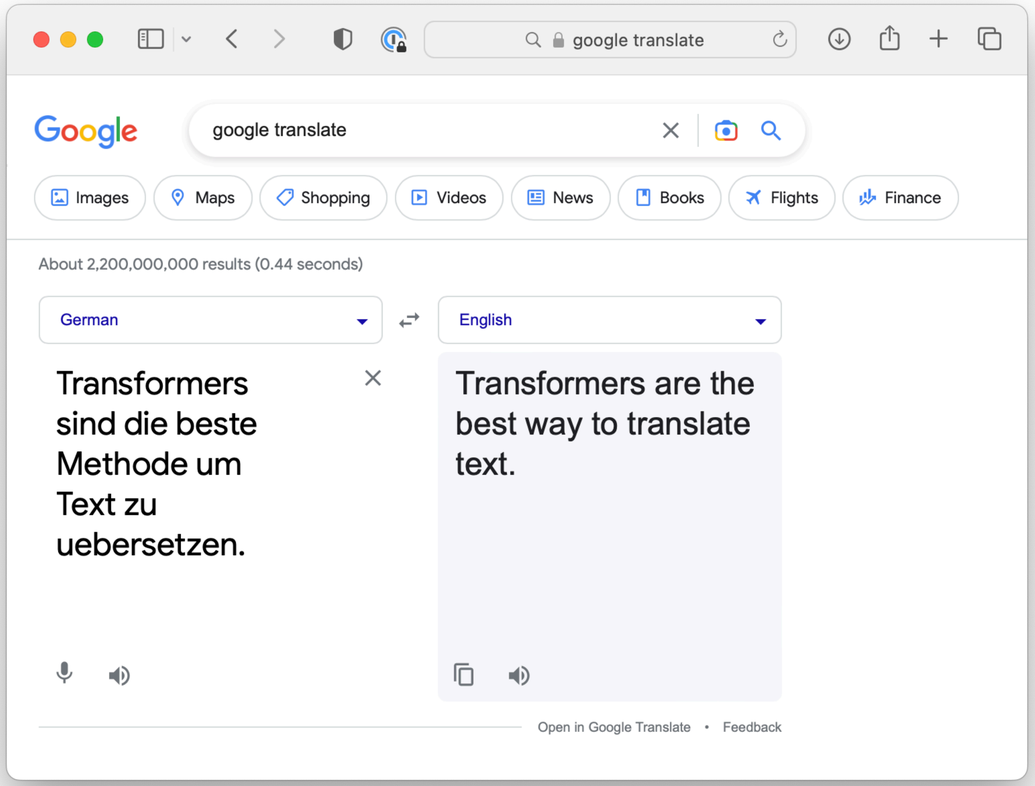
From RNNs to the Transformer Architecture

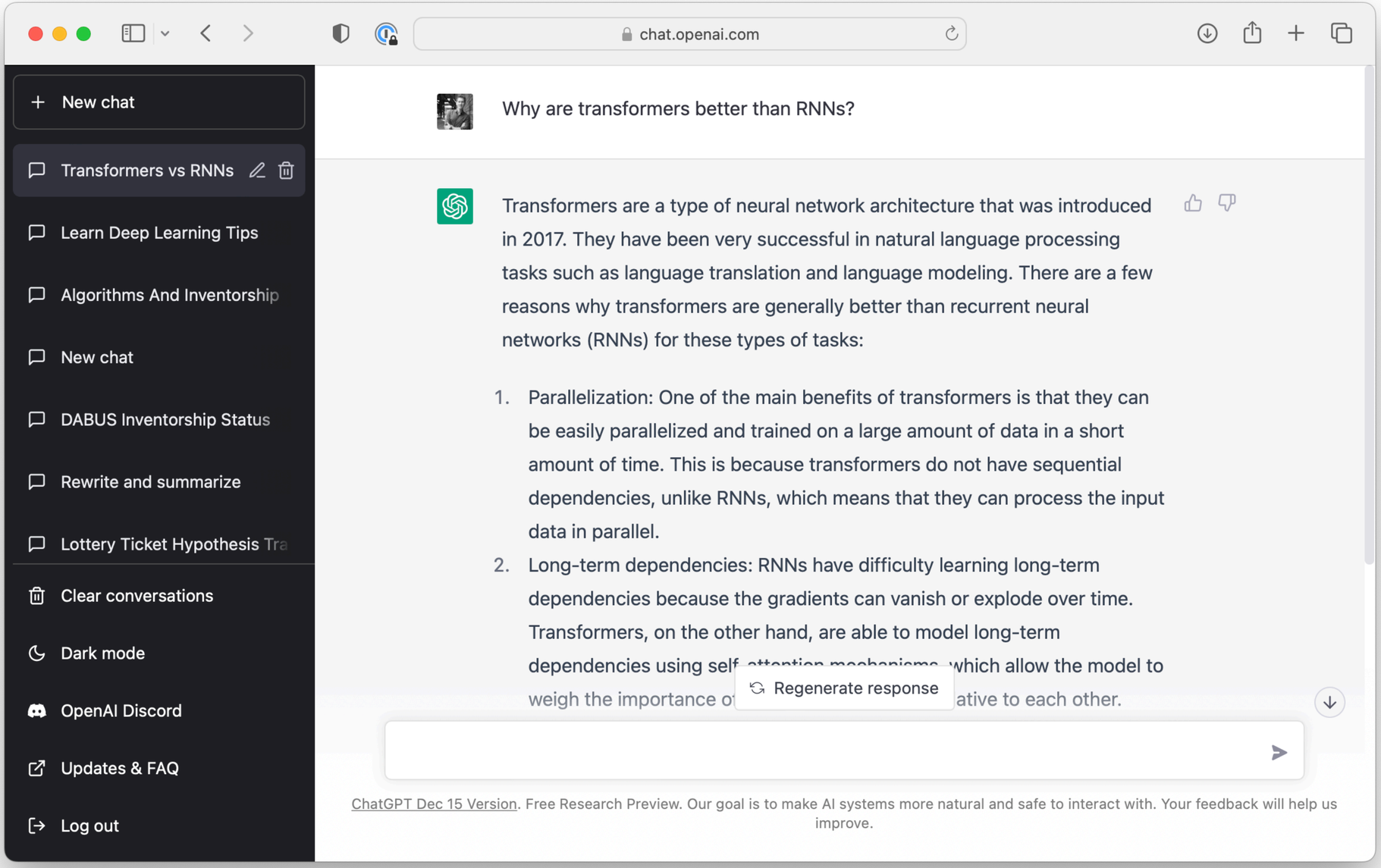
Part 1: Introducing Transformers

Sebastian Raschka and the Lightning AI Team

Recurrent Neural Networks for text are now obsolete.

So, we are not spending much more time on the RNN architecture (except for explaining attention).


google translate
Google
google translate
Images
Maps
Shopping
Videos
News
Books
Flights
Finance
About 2,200,000,000 results (0.44 seconds)
German
English
Transformers sind die beste Methode um Text zu uebersetzen.
Transformers are the best way to translate text.
Open in Google Translate • Feedback


chat.openai.com
New chat
Transformers vs RNNs
Learn Deep Learning Tips
Algorithms And Inventorship
New chat
DABUS Inventorship Status
Rewrite and summarize
Lottery Ticket Hypothesis Tra
Clear conversations
Dark mode
OpenAI Discord
Updates & FAQ
Log out
Why are transformers better than RNNs?
Transformers are a type of neural network architecture that was introduced in 2017. They have been very successful in natural language processing tasks such as language translation and language modeling. There are a few reasons why transformers are generally better than recurrent neural networks (RNNs) for these types of tasks:
1. Parallelization: One of the main benefits of transformers is that they can be easily parallelized and trained on a large amount of data in a short amount of time. This is because transformers do not have sequential dependencies, unlike RNNs, which means that they can process the input data in parallel.
2. Long-term dependencies: RNNs have difficulty learning long-term dependencies because the gradients can vanish or explode over time. Transformers, on the other hand, are able to model long-term dependencies using self
which allow the model to weigh the importance of
ative to each other.
Regenerate response
ChatGPT Dec 15 Version. Free Research Preview. Our goal is to make AI systems more natural and safe to interact with. Your feedback will help us improve.

Download PDF

Article | Open Access | Published: 15 July 2021

Highly accurate protein structure prediction with AlphaFold

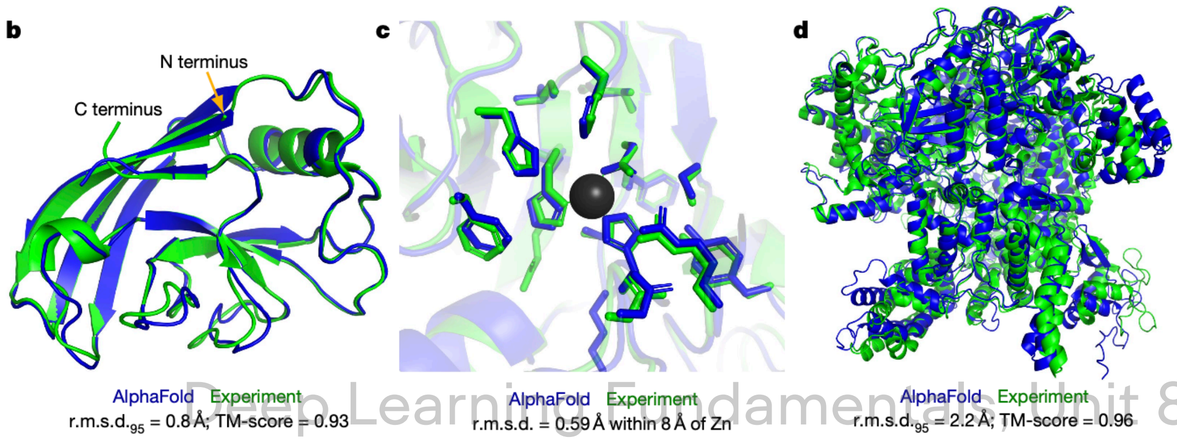
John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, ... Demis Hassabis + Show authors

Nature **596**, 583–589 (2021) | Cite this article

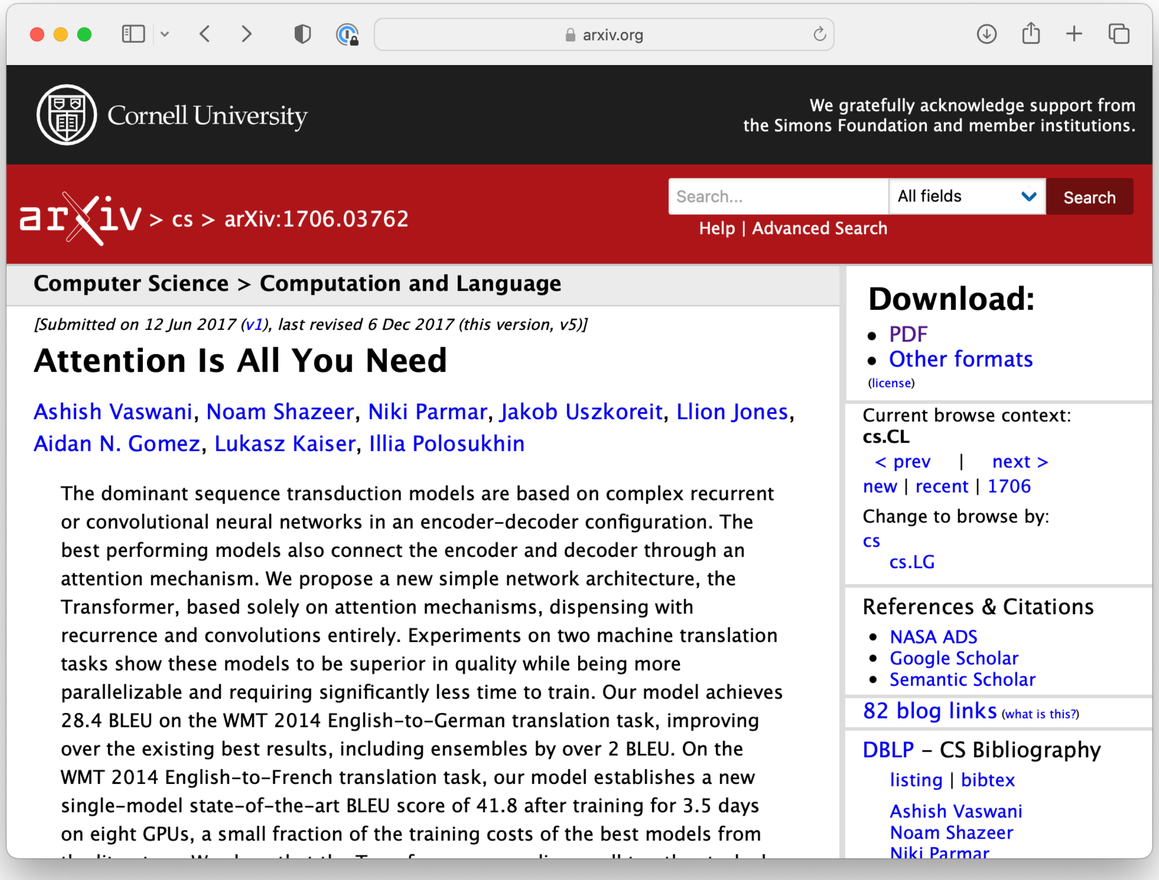
923k Accesses | **4944** Citations | **3393** Altmetric | Metrics

Abstract

Proteins are essential to life, and understanding their structure can facilitate a mechanistic understanding of their function. Through an enormous experimental effort[1,2,3,4], the structures of around 100,000



Large language models go back to the original transformer architecture proposed in 2017



arxiv.org

Cornell University

We gratefully acknowledge support from the Simons Foundation and member institutions.

arXiv > cs > arXiv:1706.03762

Search... All fields Search

Help | Advanced Search

Computer Science > Computation and Language

[Submitted on 12 Jun 2017 (v1), last revised 6 Dec 2017 (this version, v5)]

Attention Is All You Need

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks in an encoder-decoder configuration. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from

Download:
- PDF
- Other formats

(license)

Current browse context: cs.CL

< prev | next >

new | recent | 1706

Change to browse by: cs, cs.LG

References & Citations
- NASA ADS
- Google Scholar
- Semantic Scholar

82 blog links (what is this?)

DBLP - CS Bibliography

listing | bibtex

Ashish Vaswani
Noam Shazeer
Niki Parmar

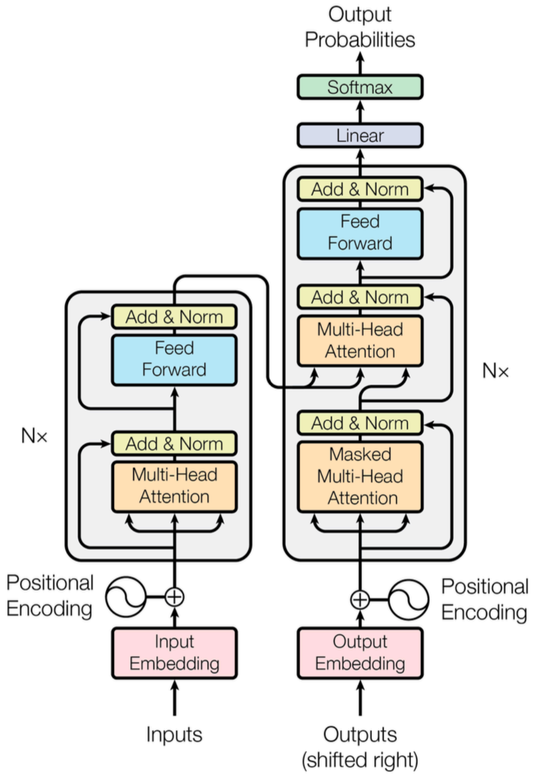


Figure 1: The Transformer - model architecture.

It may look complicated, but we should recognize most of its components

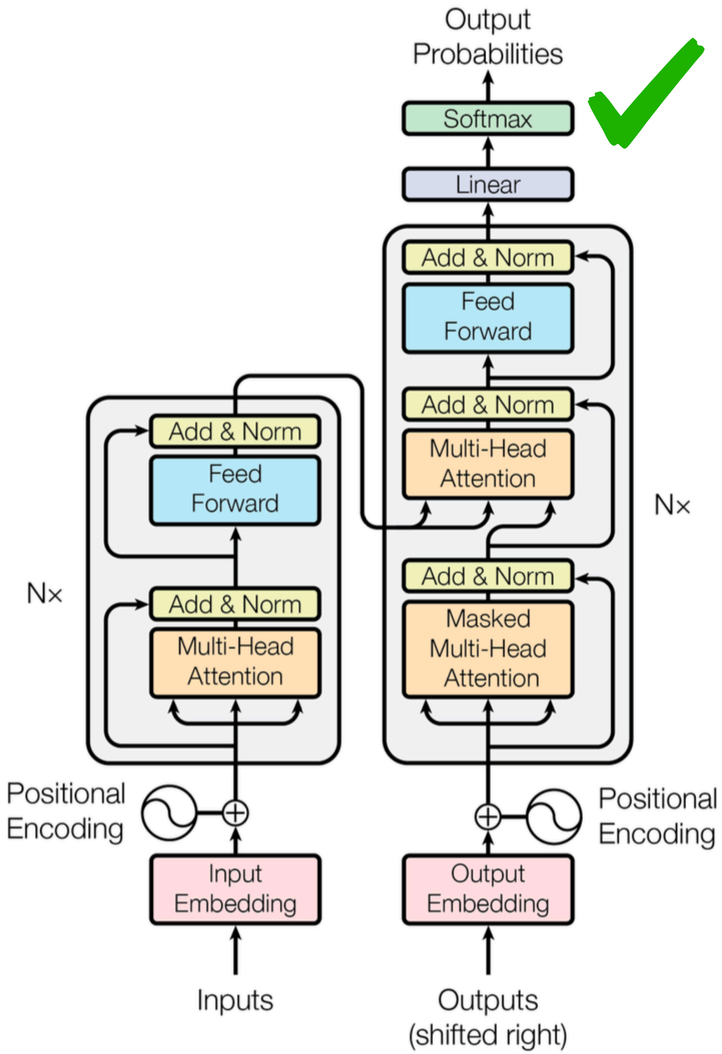


Figure 1: The Transformer - model architecture.

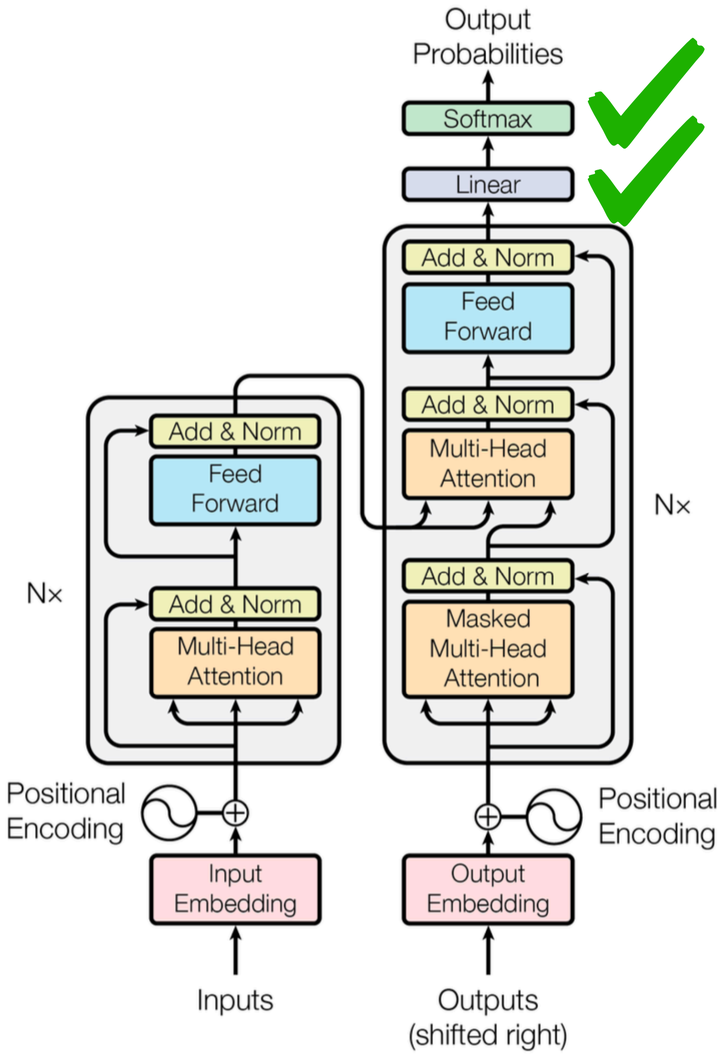


Figure 1: The Transformer - model architecture.

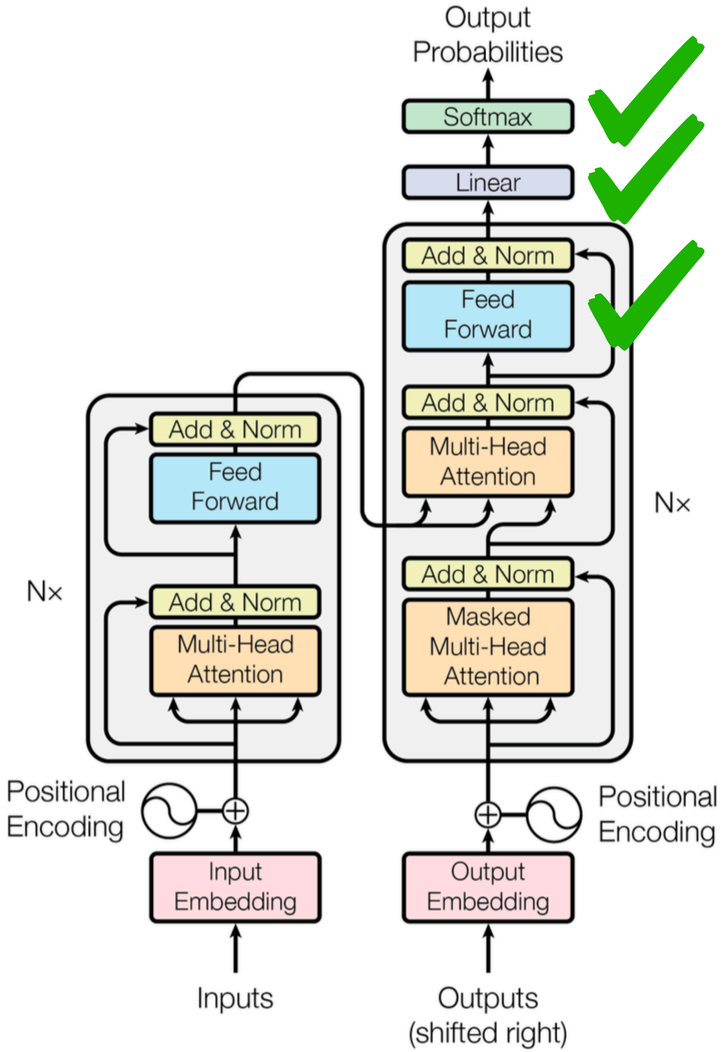


Figure 1: The Transformer - model architecture.

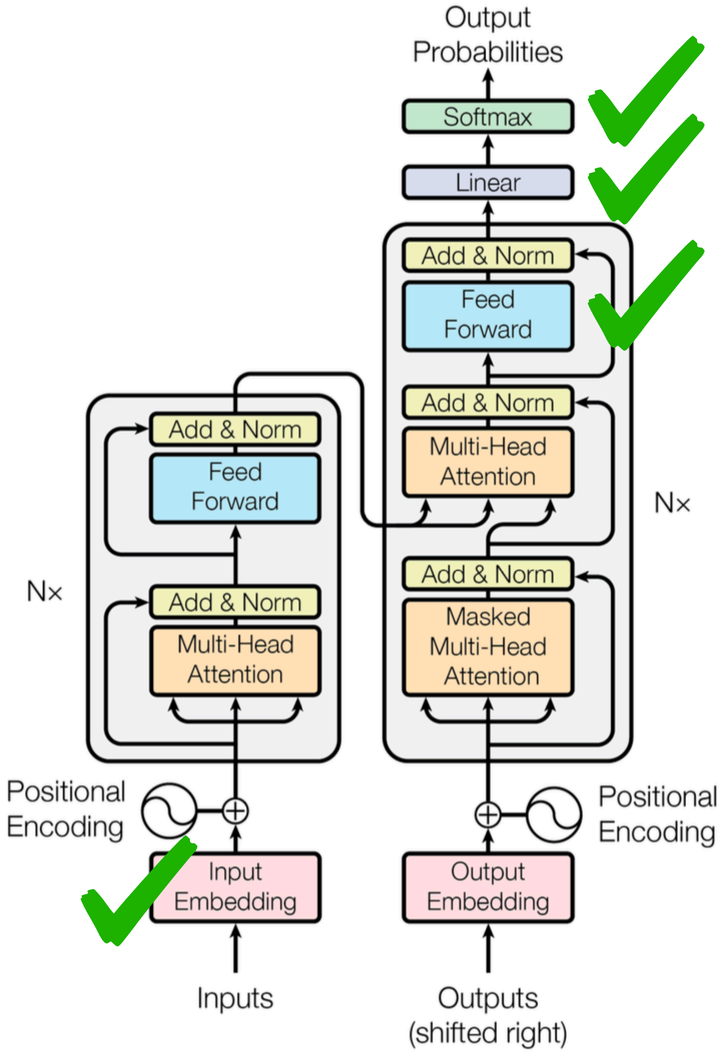


Figure 1: The Transformer - model architecture.

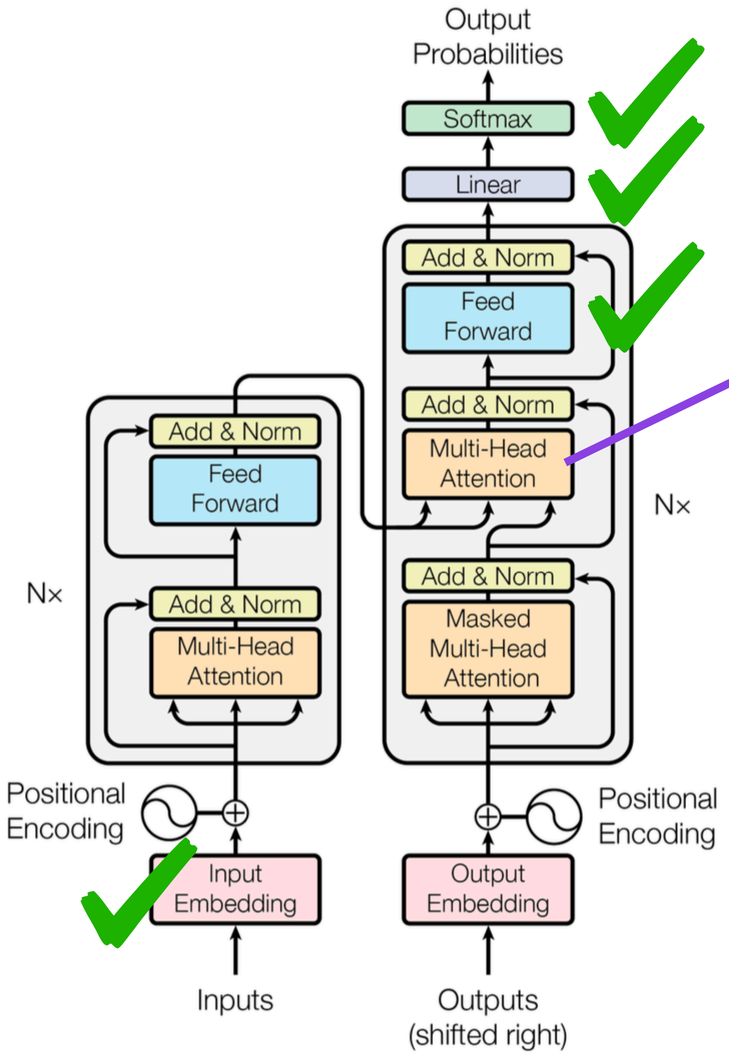


Figure 1: The Transformer - model architecture.

Next: What is “attention” and why do we need it?