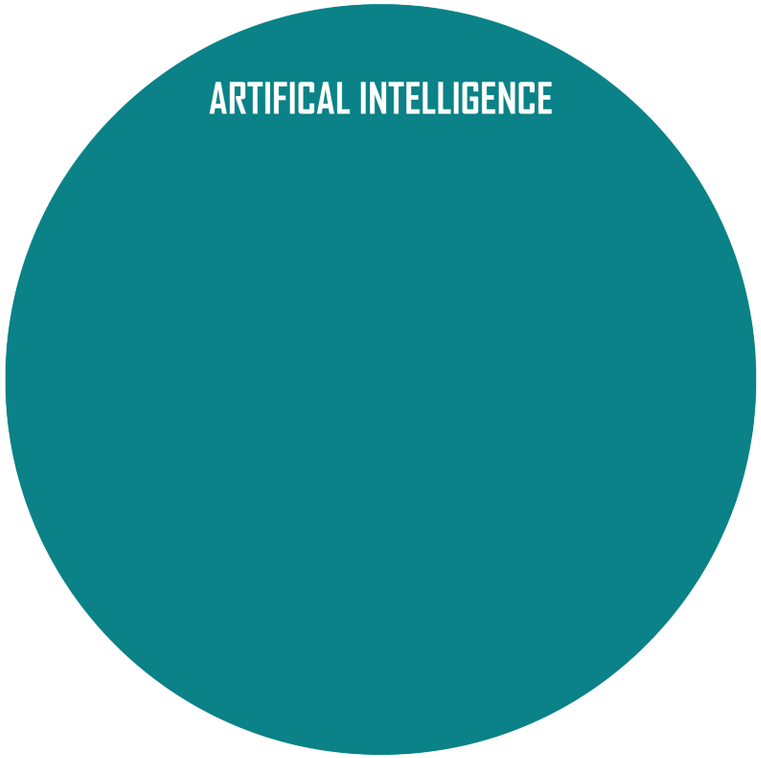
AI-Powered Public Health:
How Labs Can Prepare for the Road Ahead

Microsoft Copilot: DALL-E 3

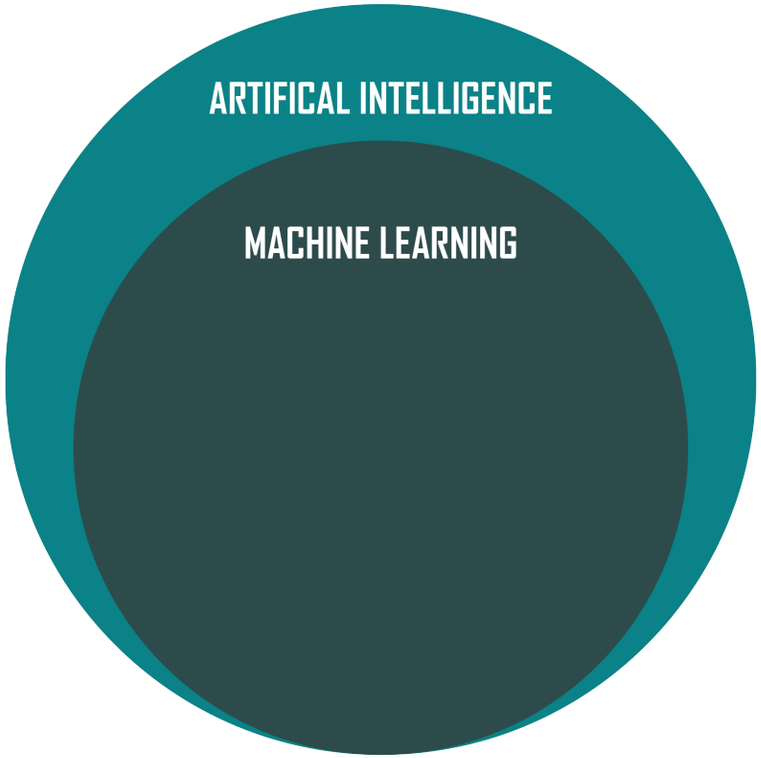
Artificial Intelligence

Software that allows machines or computer systems to perceive their environment and use learning and intelligence to achieve a defined goal.



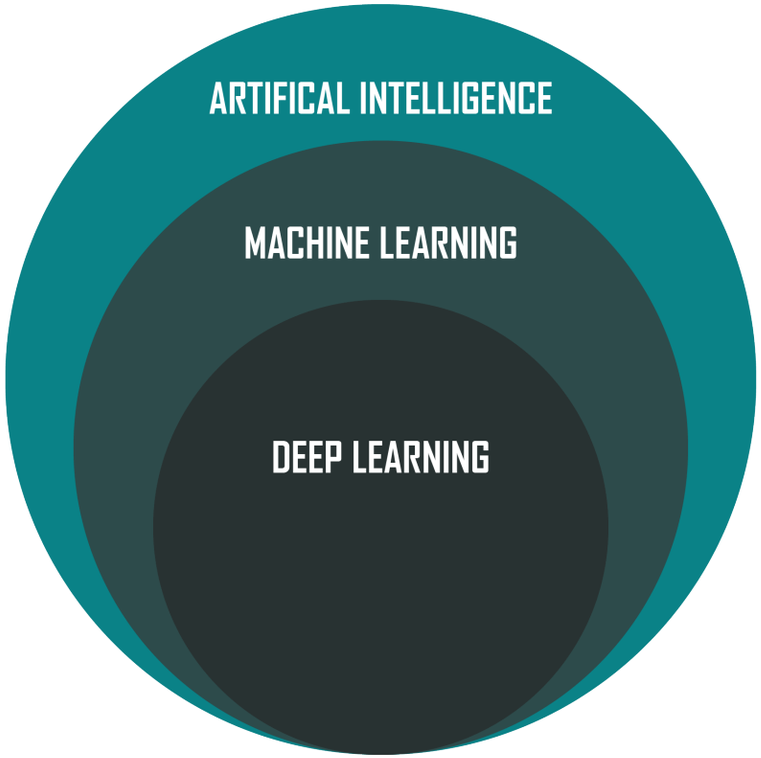
Machine Learning

An area in artificial intelligence with a focus on statistical algorithms that can learn from data and generalize to unseen data.



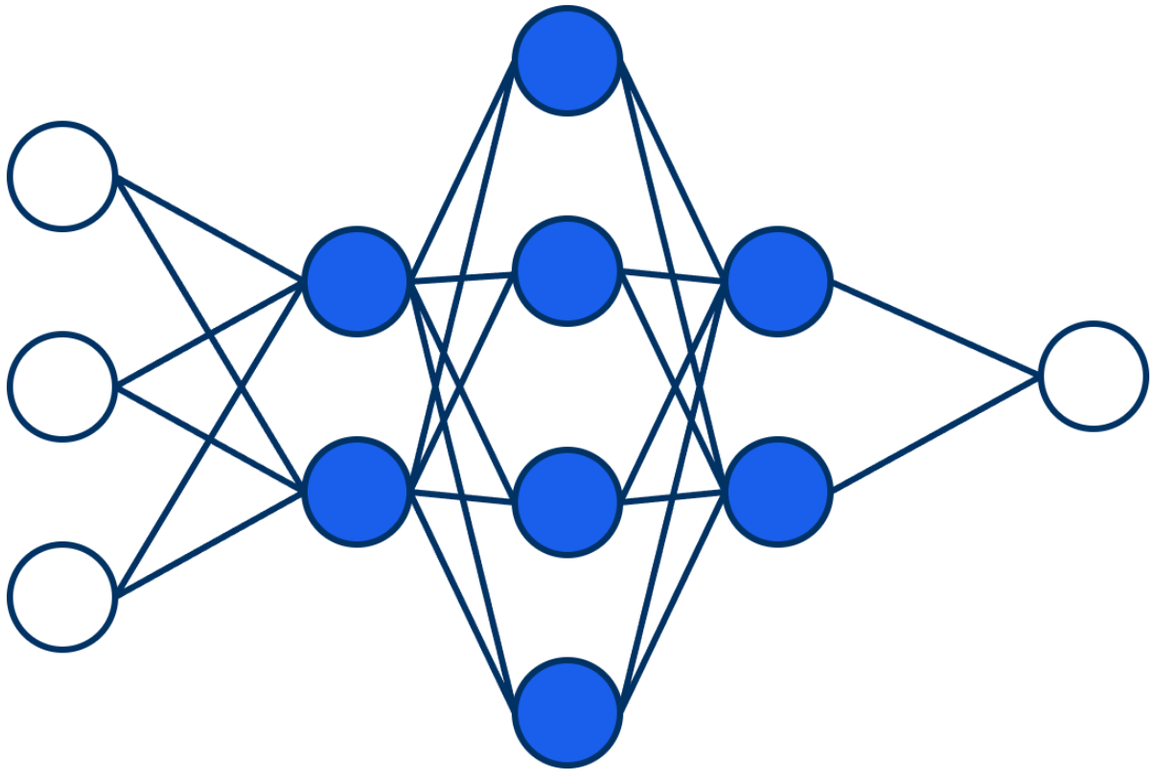
Deep Learning

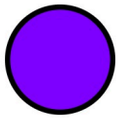
A subset of machine learning methods that are based on neural networks, with deep implying multiple layers.



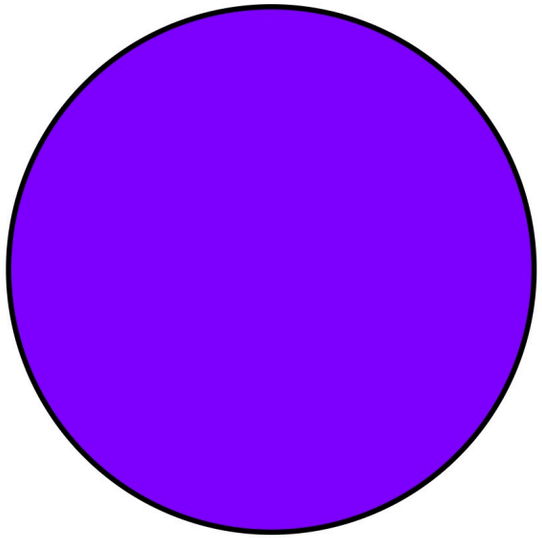
Artificial Neural Networks

A machine learning model inspired by the structure and function of biological neural networks.





GPT-3
175 billion parameters



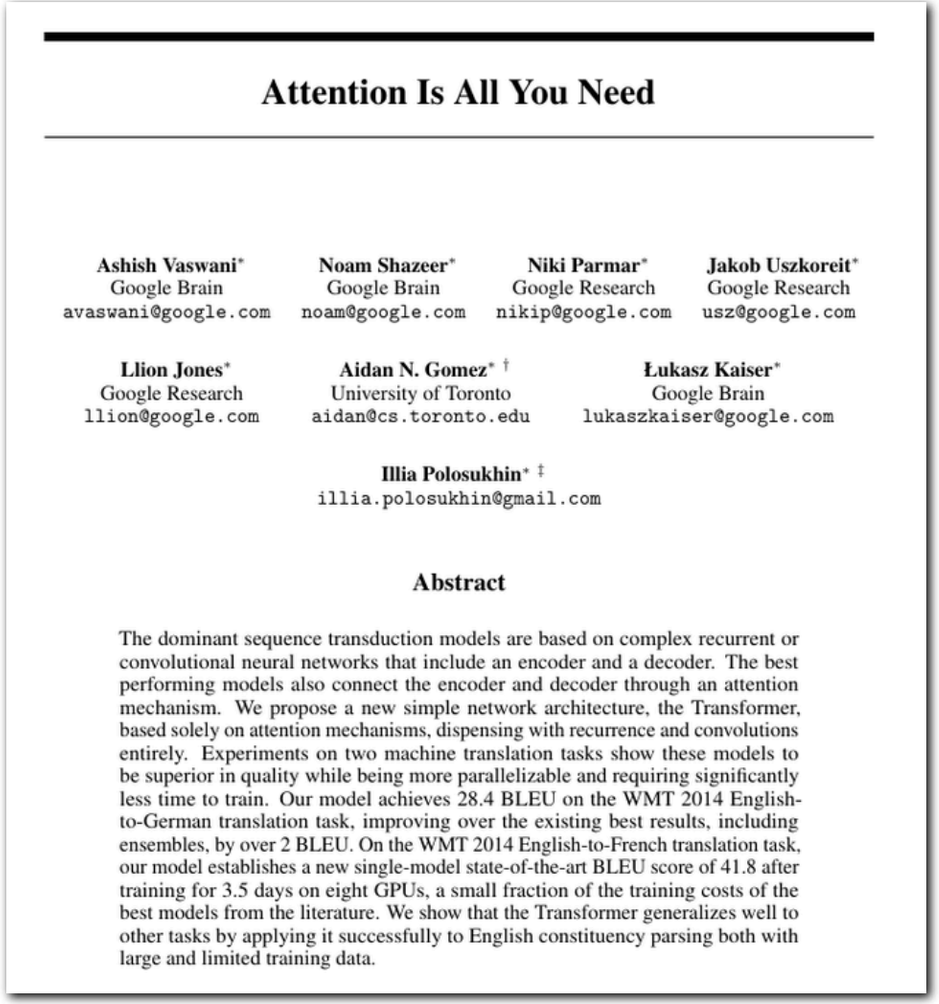
GPT-4
~1.8 trillion parameters

Large Language Model (LLM)

A model that has been trained on a massive amount of text data to mimic human language patterns and structures.

Transformers

A type of neural network architecture that is well equipped for tasks that consist of sequences of data. Most commonly a sentence.



Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. 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 the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

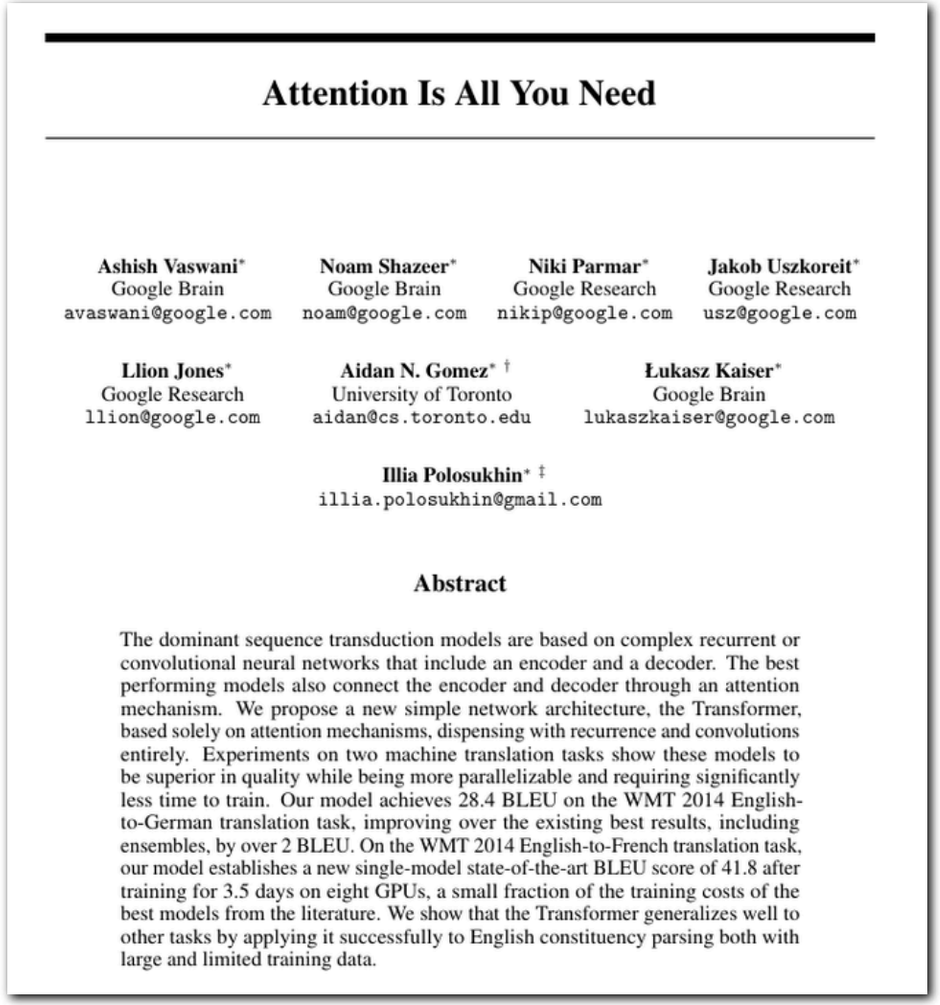
Vaswani, A. et al. 2017. “Attention Is All You Need.“ arXiv:1706.03762

Towards Data Science - De-coded: Transformers explained in plain English

Transformers

A type of neural network architecture that is well equipped for tasks that consist of sequences of data. Most commonly a sentence.

GPT-4: **G**enerative **P**re-trained **T**ransformer - 4



Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez* †
University of Toronto
aidan@cs.toronto.edu

Łukasz Kaiser*
Google Brain
lukaszkaiser@google.com

Illia Polosukhin* ‡
illia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. 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 the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

Vaswani, A. et al. 2017. “Attention Is All You Need.“ arXiv:1706.03762
Towards Data Science - De-coded: Transformers explained in plain English

AI-Powered Public Health:
How Labs Can Prepare for the Road Ahead

Microsoft Copilot: DALL-E 3