

# Generative Pretrained Transformer

A **Generative Pretrained Transformer (GPT)** is a type of [artificial intelligence model](#) designed for [natural language processing tasks](#), including text [generation](#), [translation](#), [summarization](#), and question answering. The model is based on the Transformer architecture, which allows it to handle long-range dependencies in text efficiently.

Key features of GPT include:

- 1. Pretraining and Fine-tuning:** GPT models are initially pretrained on large amounts of text data, where they learn the statistical relationships between words and sentences. After pretraining, they can be fine-tuned on specific tasks or datasets to improve performance in a particular domain.
- 2. Generative Capability:** As the name suggests, GPT models are generative, meaning they can create new content based on the input they receive. They predict the next word or sequence of words, making them effective at generating coherent and contextually relevant text.
- 3. Transformers:** The Transformer architecture relies on self-attention mechanisms, which enable the model to process all parts of the input sequence simultaneously, improving both training efficiency and the ability to understand complex relationships between different parts of the text.

GPT has evolved through several versions, each with improvements in model size, capabilities, and performance, such as GPT-2, GPT-3, and the current GPT-4, which is the model you're interacting with right now.

From:  
<https://neurosurgerywiki.com/wiki/> - **Neurosurgery Wiki**

Permanent link:  
[https://neurosurgerywiki.com/wiki/doku.php?id=generative\\_pretrained\\_transformer](https://neurosurgerywiki.com/wiki/doku.php?id=generative_pretrained_transformer)

Last update: **2025/01/14 18:52**

