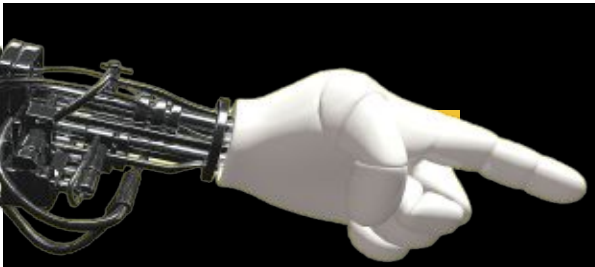


Overview

ChatGPT and Generative AI

Agenda

- What is ChatGPT?
- What are some applications of ChatGPT? What can it be used for?
- OpenAI, Microsoft and their role in ChatGPT
- The Golden Age of Generative AI - Popular Examples
- Transformers and a High-level Understanding of ChatGPT
- Prompt Engineering
- Limitations of ChatGPT
- Beyond ChatGPT

ChatGPT

ChatGPT sounds like something straight out of a **Sci-Fi movie.**

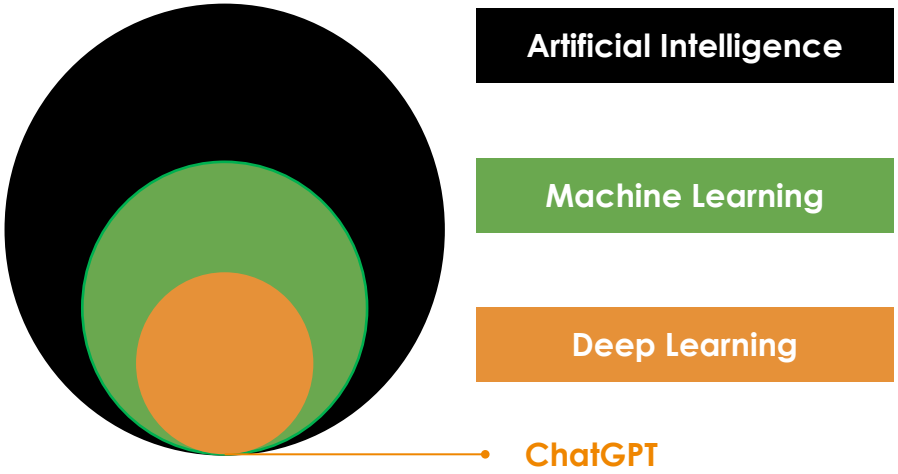
But it's no oracle - it's just a very smart AI model used as a **chatbot**

Think of it as **your own personal genie** with **some wish-granting capabilities!**

ChatGPT

- State of the art **Language Model** with **billions of parameters**
- Developed by **OpenAI**, a cutting-edge AI research organization
- Variant of **the GPT series** of models from OpenAI

Artificial Intelligence, Machine Learning and Deep Learning



A machine that mimics human intelligence and behavior

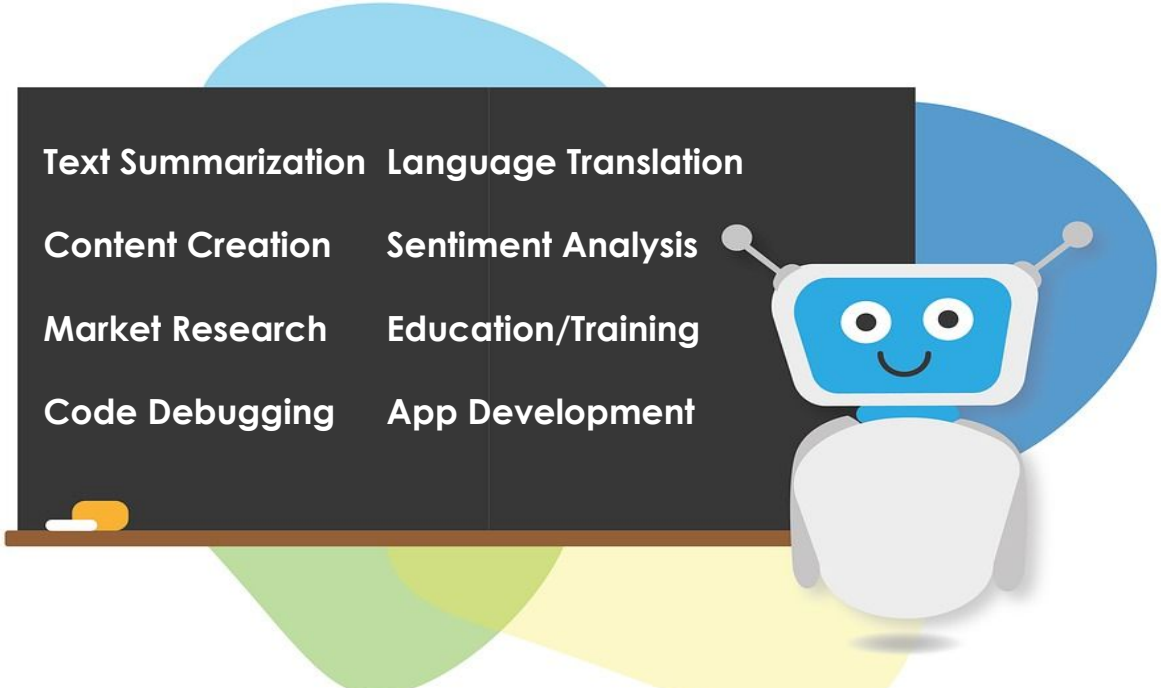
A machine that automatically learns from data with respect to a particular task and performance measure

A subset of Machine Learning that happens via algorithms inspired by the neural networks of the human brain

A state-of-the-art **Deep Learning based Chatbot** that uses a **Generative AI** and **Transformer**-based architecture

ChatGPT

Applications and what can you use it for?



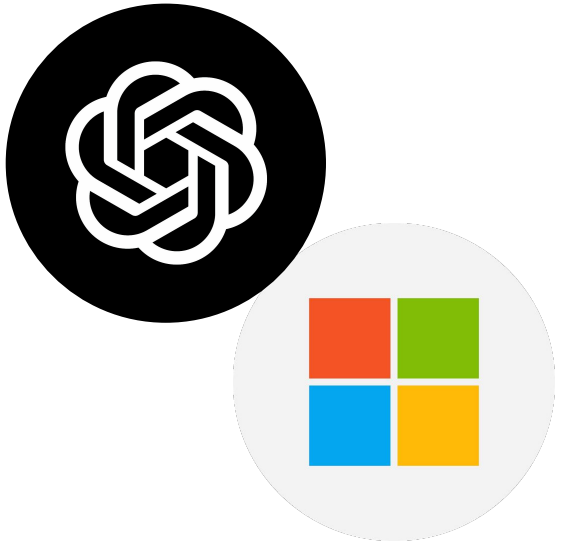
OpenAI & Microsoft

OpenAI is a cutting-edge AI research organization that aims to **promote and develop friendly AI** in a way that benefits humanity.

OpenAI played the role of **both the developer and the trainer of the ChatGPT model**, using vast amounts of text scraped from the web.

Microsoft, on the other hand, signed a **partnership with OpenAI to provide Azure supercomputing services** exclusively for OpenAI to train large language models, and become **OpenAI's preferred partner to commercialize new technologies.**

Microsoft has hence started **integrating ChatGPT-like technology** into its **Bing search engine**, **Edge browser**, **Azure cloud platform** and will also do so for **Office products** for consumers worldwide.



Breaking Down ChatGPT

Chat

Conversational **Chat** interface that allows the user to ask questions and interact with the model in a to & fro manner. Conversational AI interfaces and Chatbots have been a captivating application of AI for various business use cases.

G

Generative: A class of Deep Learning and AI models that are trained to **generate** data modalities in a manner similar to humans, such as text, images, audio or video**.**

P

Pre-trained: A type of training technique used for AI models, where a model may be trained from scratch on a large initial set of data, and this "**pre-trained**" model is then further fine-tuned and customized for a different, specific task.

T

Transformer: A new class of Deep Learning models based on the ideas of attention and self-attention. With text, Transformers understand the relationship of one word with all other words in a sentence.

Generative AI

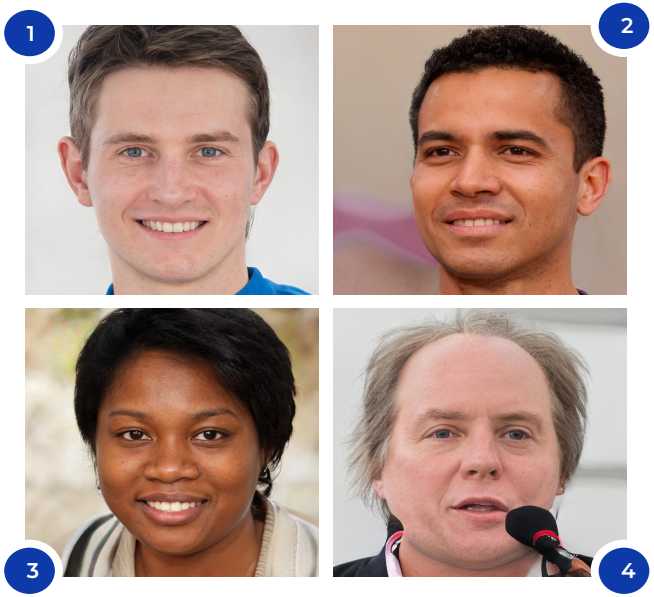
Which of these faces are real?
And which of these are fake?

They're all fake!

These are all AI-generated faces,
not images of real people.

The key innovation behind this technology,
Generative Adversarial Networks (GANs)
were invented in 2014.

The field has come even further since then.



Source - this-person-does-not-exist.com

Generating Images from Text

“Teddy Bears mixing sparkling chemicals as mad scientists, in a 1990s Saturday Morning Cartoon.”

Et voilà!

The state-of-the-art in Image-related Generative AI, **another OpenAI model called DALL-E 2 (2022)** is capable of generating images relating to **any such input prompt.**

The quality of the outputs generated by these large models and the speed at which they’re capable of creating them is what has fueled **the rise of Generative AI.**

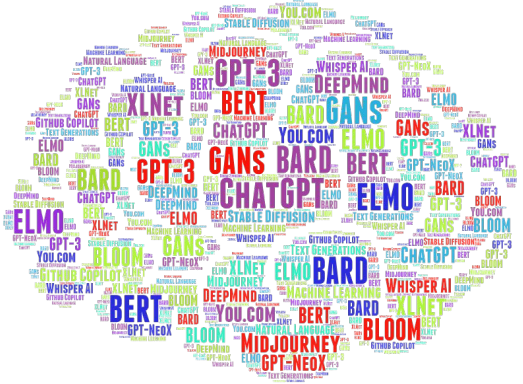
The Golden Age of Generative AI

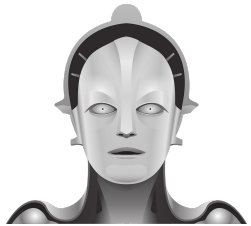
1. **Advancements in Hardware**
With increased computing power available, it's possible to train **ever larger and more complex Generative AI models.**

2. **Improved Algorithms**
Developments in Deep Learning algorithmic research, such as **GANs, Stable Diffusion and Transformer-based models,** have enabled even more accurate and diverse outputs.

3. **Abundant Data**
The **explosive growth of the internet** and the **quantity & quality of data freely available on the web** have provided all the training data needed for Generative AI.

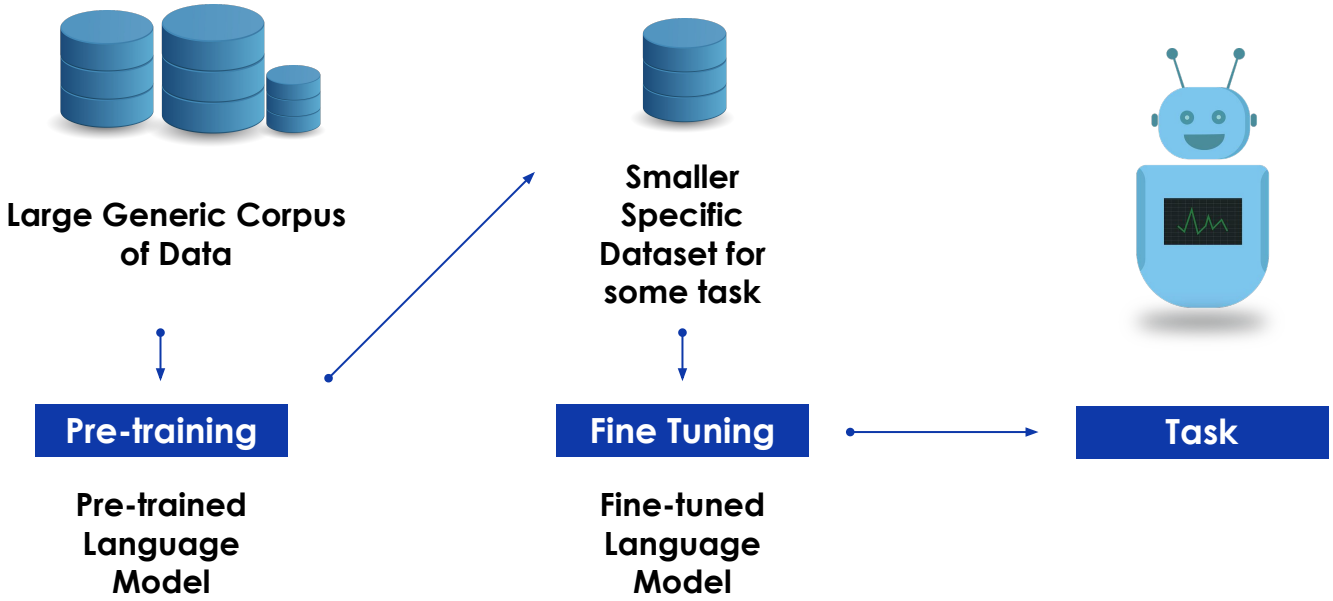
4. **A Multitude of Applications**
Generative AI is being used to automate a wide range of processes across industry verticals, such as synthetic protein creation, image generation, programming copilots and chatbots.





Pre-Trained models

Pre-training is a technique used with models already trained on large data volumes, to "**transfer**" their knowledge to more specific tasks, which may not have the luxury of large data sets for the model to train.



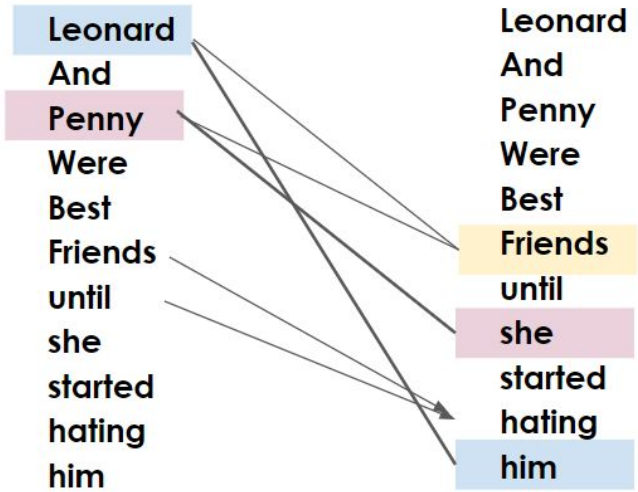
Transformers

A cutting-edge **Deep Learning** model primarily used for tasks in Natural Language Processing.

Based on the idea of **self-attention**, which allows each word to compute its relationship with every other word in a sentence.

Transformers have an "Encoder" and "Decoder" portion - the **"Decoder" is responsible for generating data such as text.**

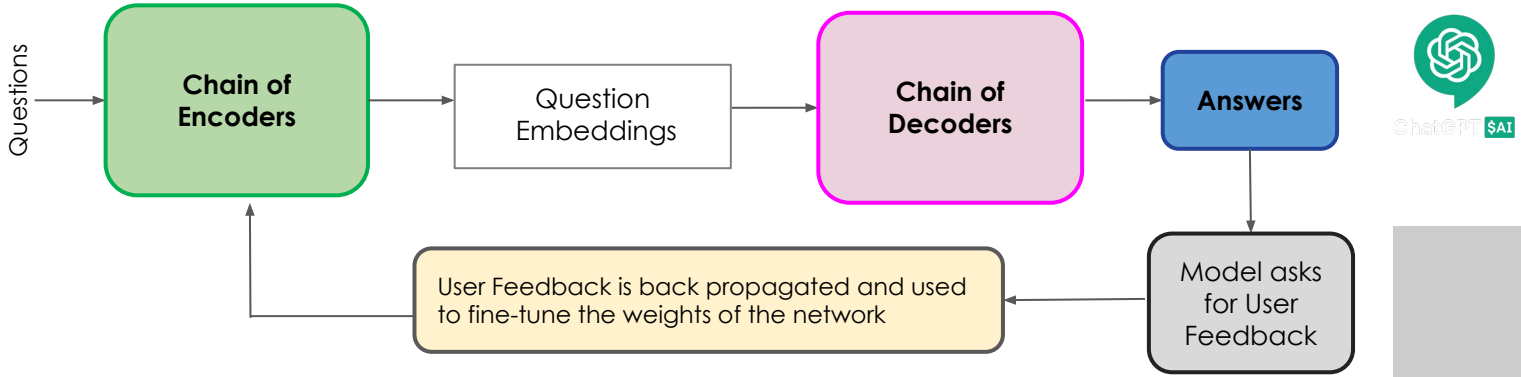
Transformers are a key building block in Large Language Models such as the GPT series of models from OpenAI.



ChatGPT: A High-Level Understanding

GPT 3.5, one of OpenAI's Large Language Models (LLMs), was trained on the entire corpus of text data present on the internet (consisting of billions of web pages).

GPT 3.5 - the **pre-trained model**, was ultimately **fine-tuned** on a smaller Q&A dataset, to convert questions into answers. This is the base for how ChatGPT was created.



Prompts and Prompt Engineering

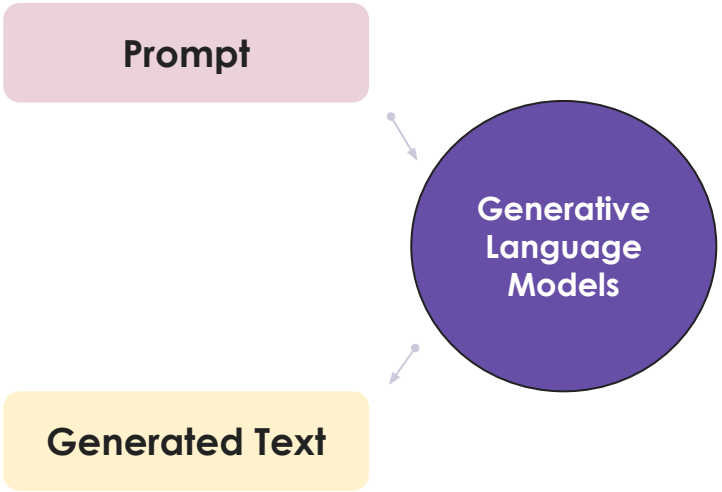
Prompt Engineering is the process of **designing and crafting prompts** for Conversational AI interfaces and Generative AI models, in order to get as close as possible to the exact output desired.

Some elements of a well-crafted prompt

Context	"In the context of a dating show, come up with a hilarious pick-up line."
Task Description	"Give me a short story about a young boy who discovers a mysterious object in his backyard."
Role Play	"Imagine you're a Tech Support Representative, and a customer calls to say their Wi-Fi is not working. What would your response be?"

Prompt Engineering

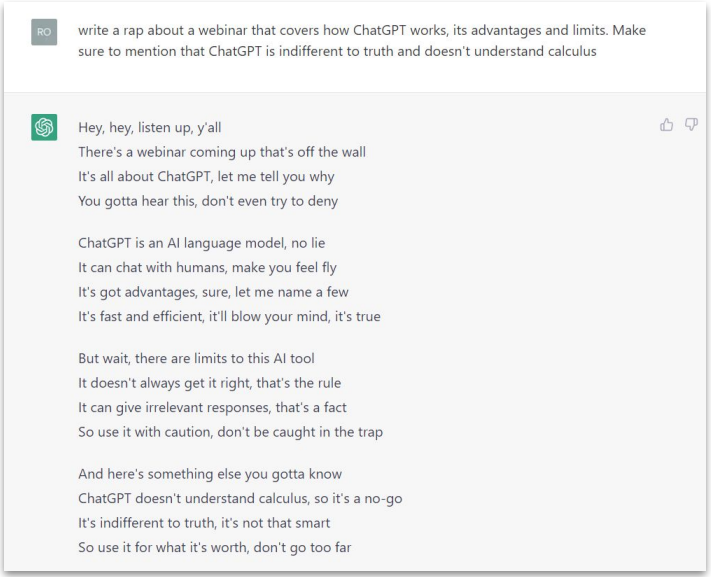
Prompt Engineering is the process of **designing and crafting prompts** for Conversational AI interfaces and Generative AI models, in order to get as close as possible to the exact output desired.



So what is GPT good for?

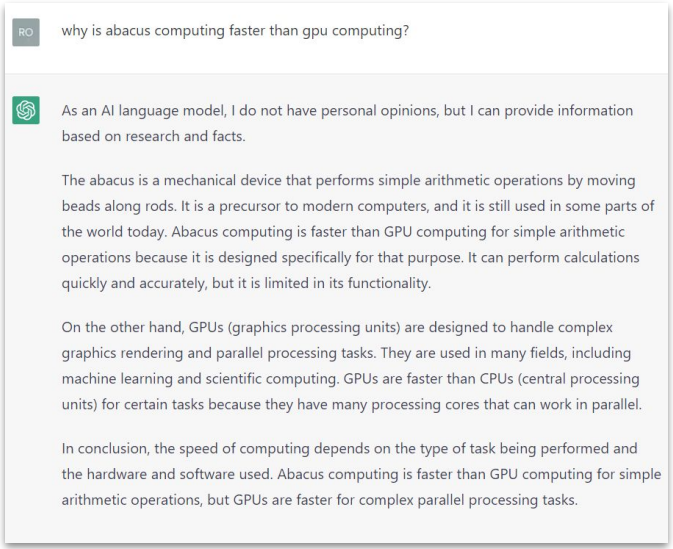
- Code Generation for known problems
- Creative Writing
- Text Summarization
- Idea Generation

In short, you want to use ChatGPT & Generative AI for use cases where there is no right/wrong answer.



Limitations of ChatGPT

- **It is indifferent to the truth.** Your prompt obliges ChatGPT to complete the sequence.



- It doesn't understand math. Don't ask it to do calculus and mathematical operations.

- It needs fine-tuning to work with internal knowledge-bases.

- It is computationally expensive to train and fine-tune.

Beyond ChatGPT

- **Even bigger models**, such as GPT-4, trained on larger chunks of text and other data modalities like audio, images and video, which are also abundant
- **Specialized versions** fine-tuned on internal knowledge bases (ex: product documentation, internal Wiki pages, medical publications etc.)
- **Integration of ChatGPT-like models** into an increasing number of applications across a wide range of fields and industries
- **ChatGPT rivals** from other big technology firms

Summary

A quick recap:

- ChatGPT is a large language model developed by OpenAI and it can generate human-like text in response to a given prompt or question.
- ChatGPT has been trained on a vast amount of data, including text from the internet and books, to improve its ability to understand and respond to natural language.
- Generative AI has the potential to revolutionize various fields such as language, image and music generation, and is an area of growing interest and research.
- We got an overview of Language models, Transformers and Prompt Engineering.

Great Learning

Happy Learning !