

## Introduction to Large Language Models (LLMs)



### Value Quantified in Action

In the end, it's the resulting actions that prove the worth of any analytics engagement.

With our end-to-end expertise, we're with you through that last mile.



### The Right Approach

We know Advanced Analytics, and we know how to make it work for you.

Our experience means our clients get successful, scalable transformations, not hype and buzzwords.

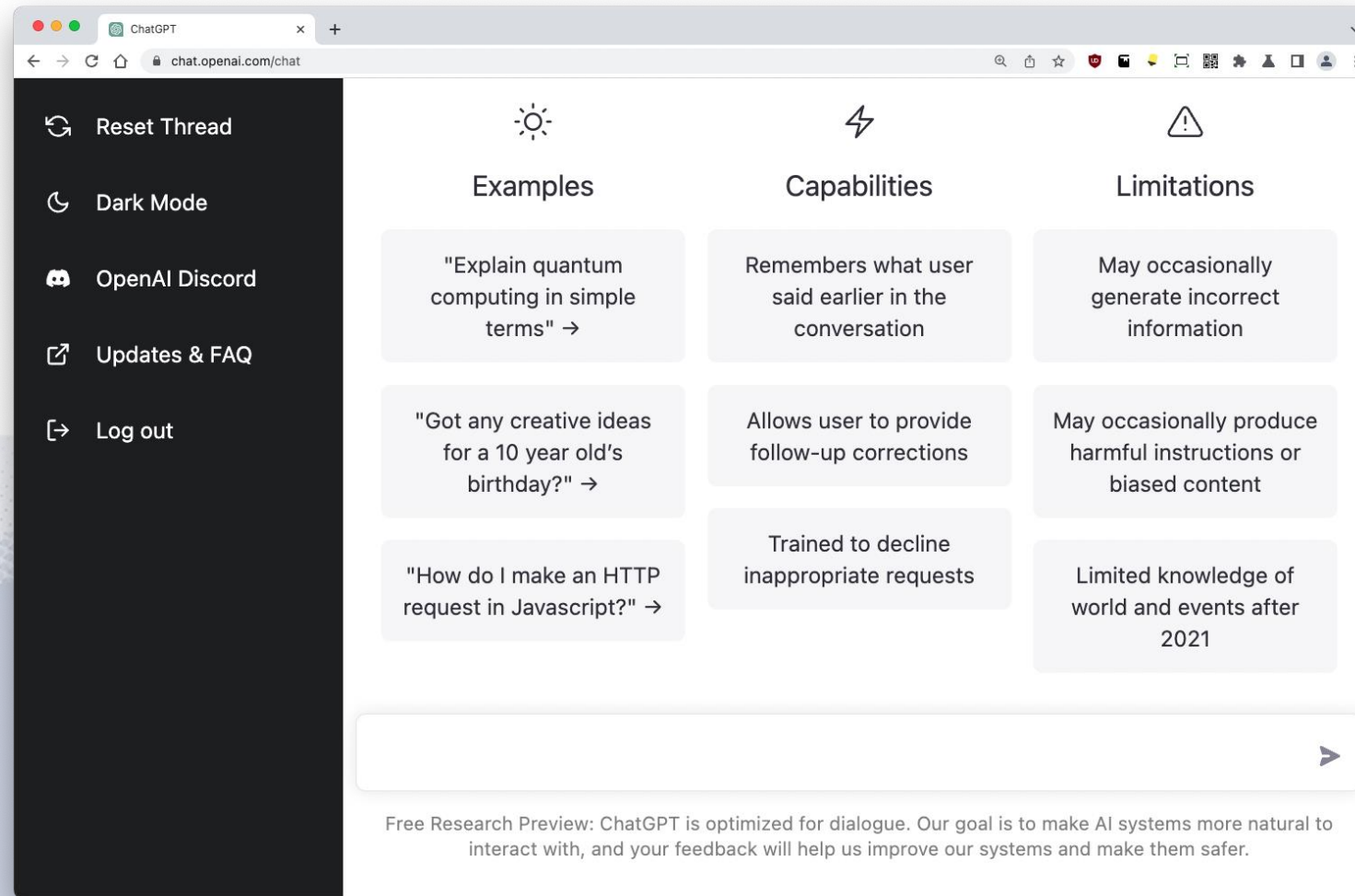


### Your Secret Data Weapon

Any Analytics initiative is only as successful as the data engineering behind it.

That's why many organizations and technology partners trust us to do their heavy lifting.

# ChatGPT – December 2022



Powered by GPT 3.5

# Meet Eliza 1966



Welcome to

```
EEEEEE LL      IIII  ZZZZZZ  AAAAA
EE      LL      II    ZZ    AA   AA
EEEEEE LL      II    ZZZ   AAAAAA
EE      LL      II    ZZ    AA   AA
EEEEEE LLLLLL  IIII  ZZZZZZ  AA   AA
```

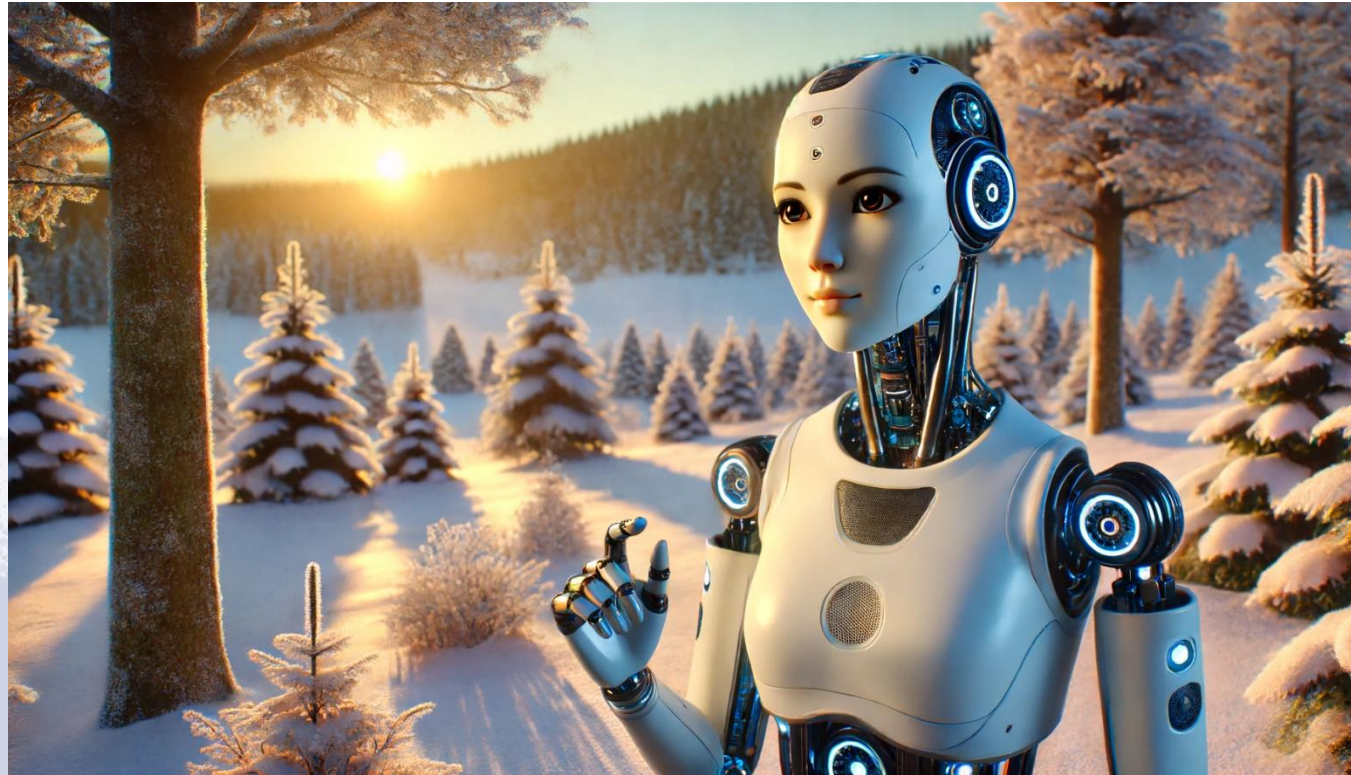
Eliza is a mock Rogerian psychotherapist.  
The original program was described by Joseph Weizenbaum in 1966.  
This implementation by Norbert Landsteiner 2005.

```
ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```



**Joseph  
Weizenbaum**

# AI Winters

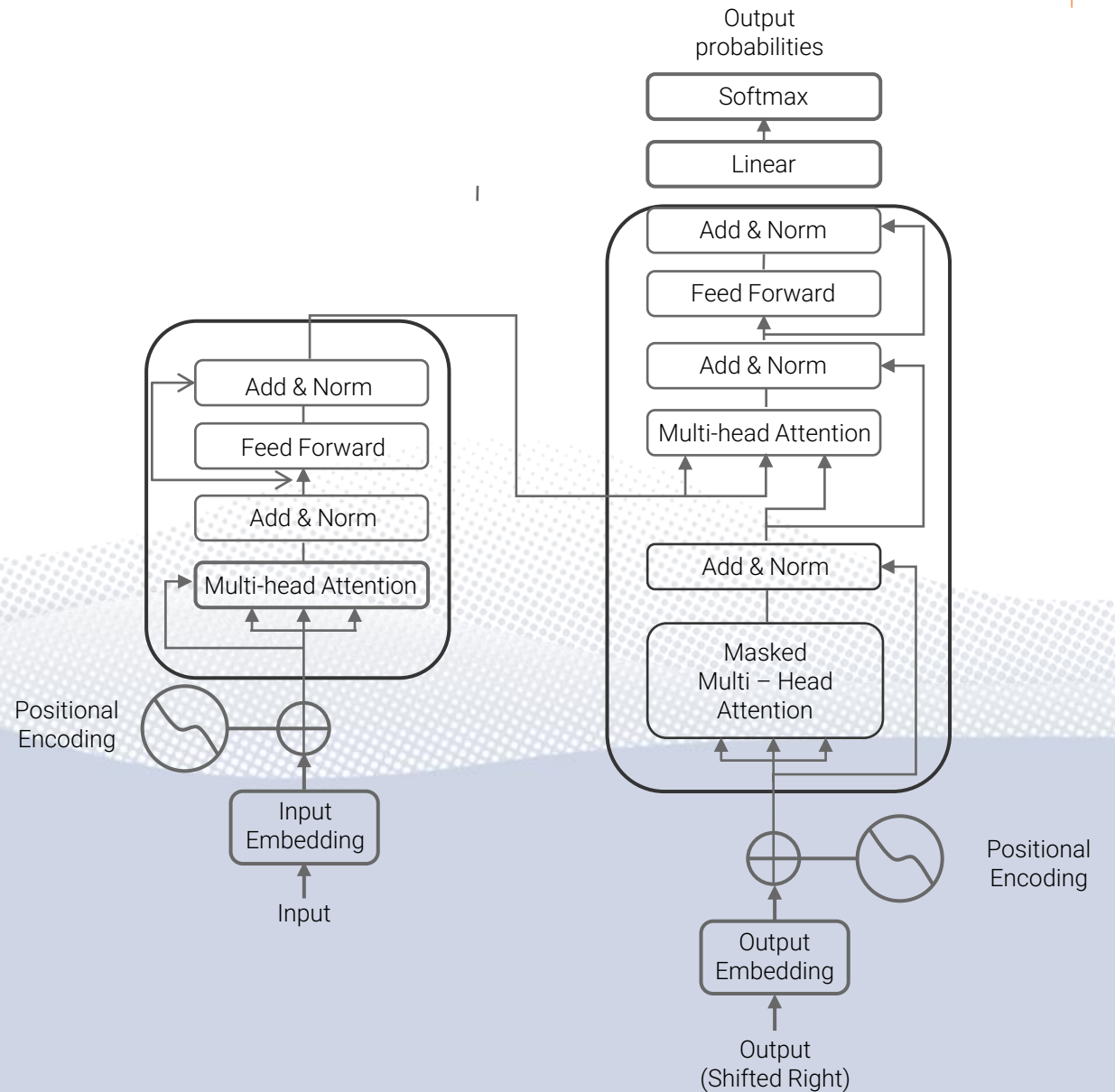


1974-1980 & 1987-2000

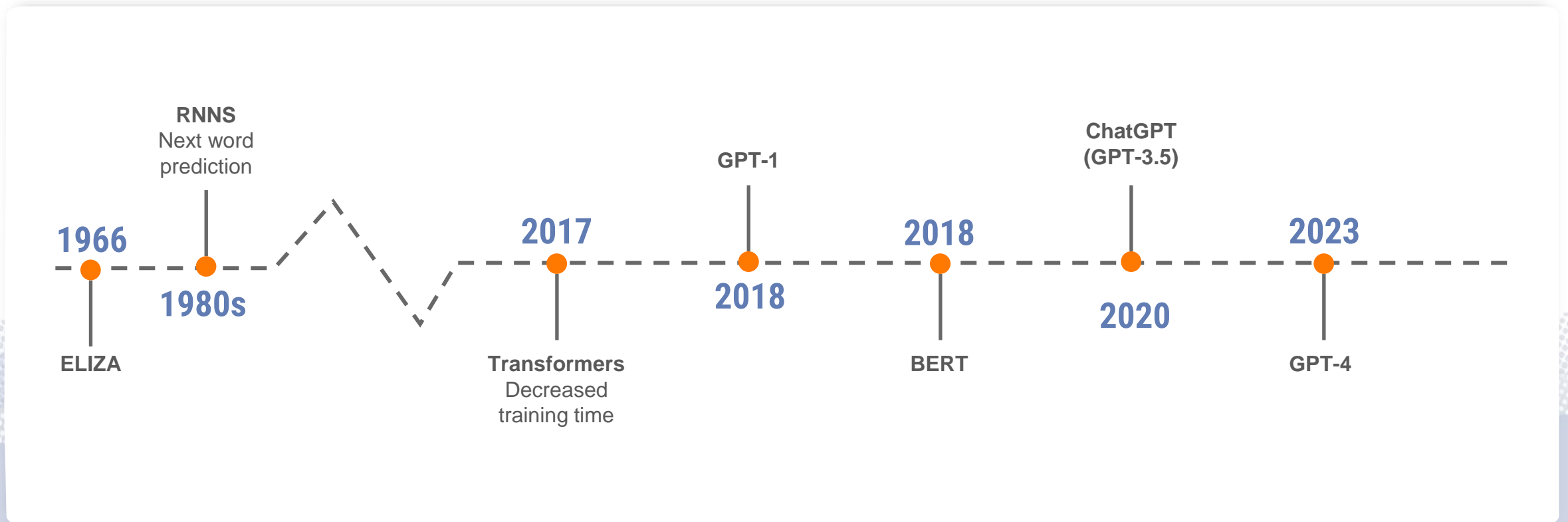
# Attention is all you Need



- Released by Google in 2017
- Transformers architecture
- Massively reduced compute requirements for training AI models
- Made Generative Pre-Trained Models Possible



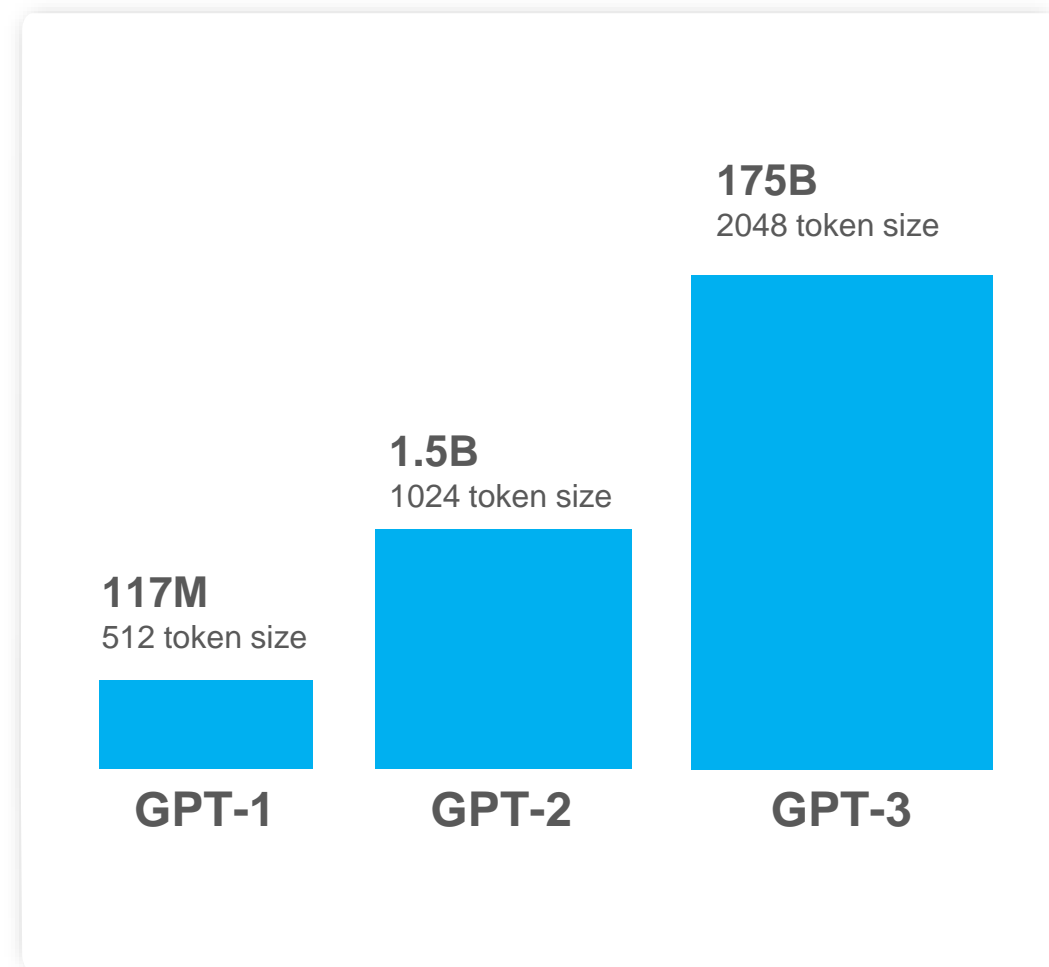
# GenAI Time-Lines



# Scaling the Models



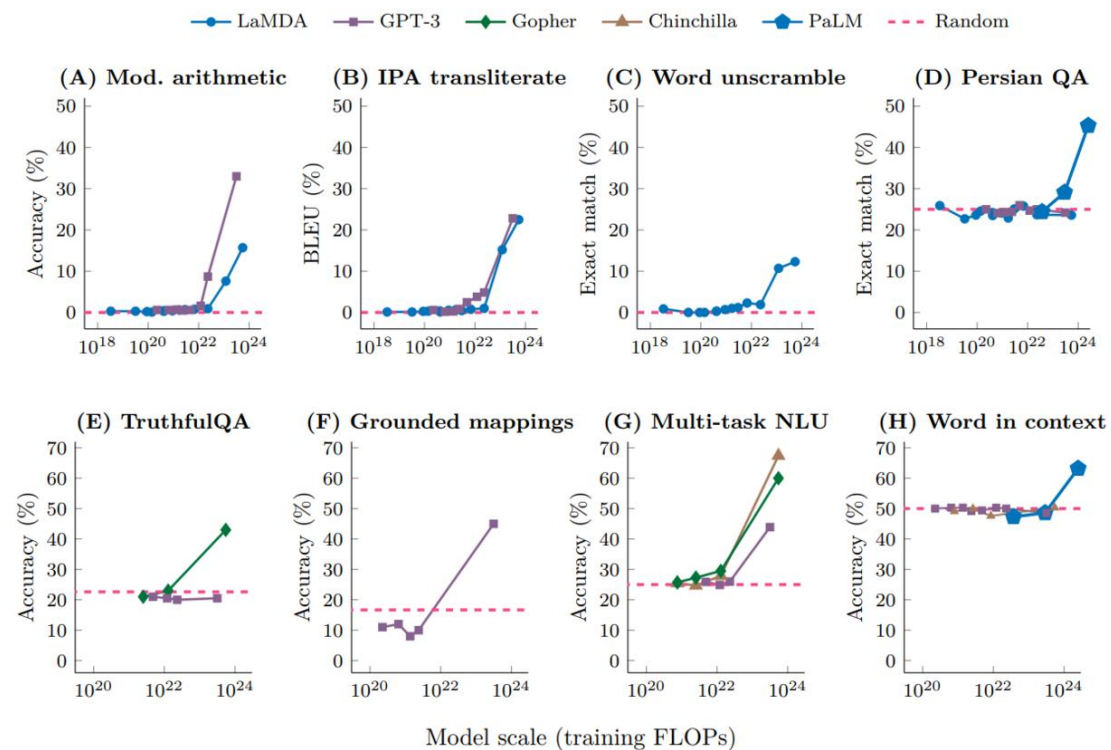
**Parameters** - the weights and biases in the neural network which change as the network “learns”



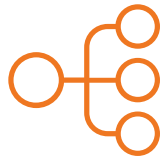
# Emergent Abilities



The capabilities of LLMs that appear suddenly and unpredictably as model size, computational power, and training data scale up.



# How it works



## Data

Massive Amounts of  
Data – often  
petabytes



## Architecture

The transformers  
architecture



## Training

Compute Intensive



## Training vs Inference

# So, what are LLMs again?



LLMs are a type of Foundational models pretrained on extensive amount of unlabeled and self-supervised data.

This training allows them to identify patterns and produce outputs that are both generalizable and adaptable.

LLMs are foundation models applied to **text** and **text-based** applications like coding.

# Tokenization



Pear, Hello, Bye => single token

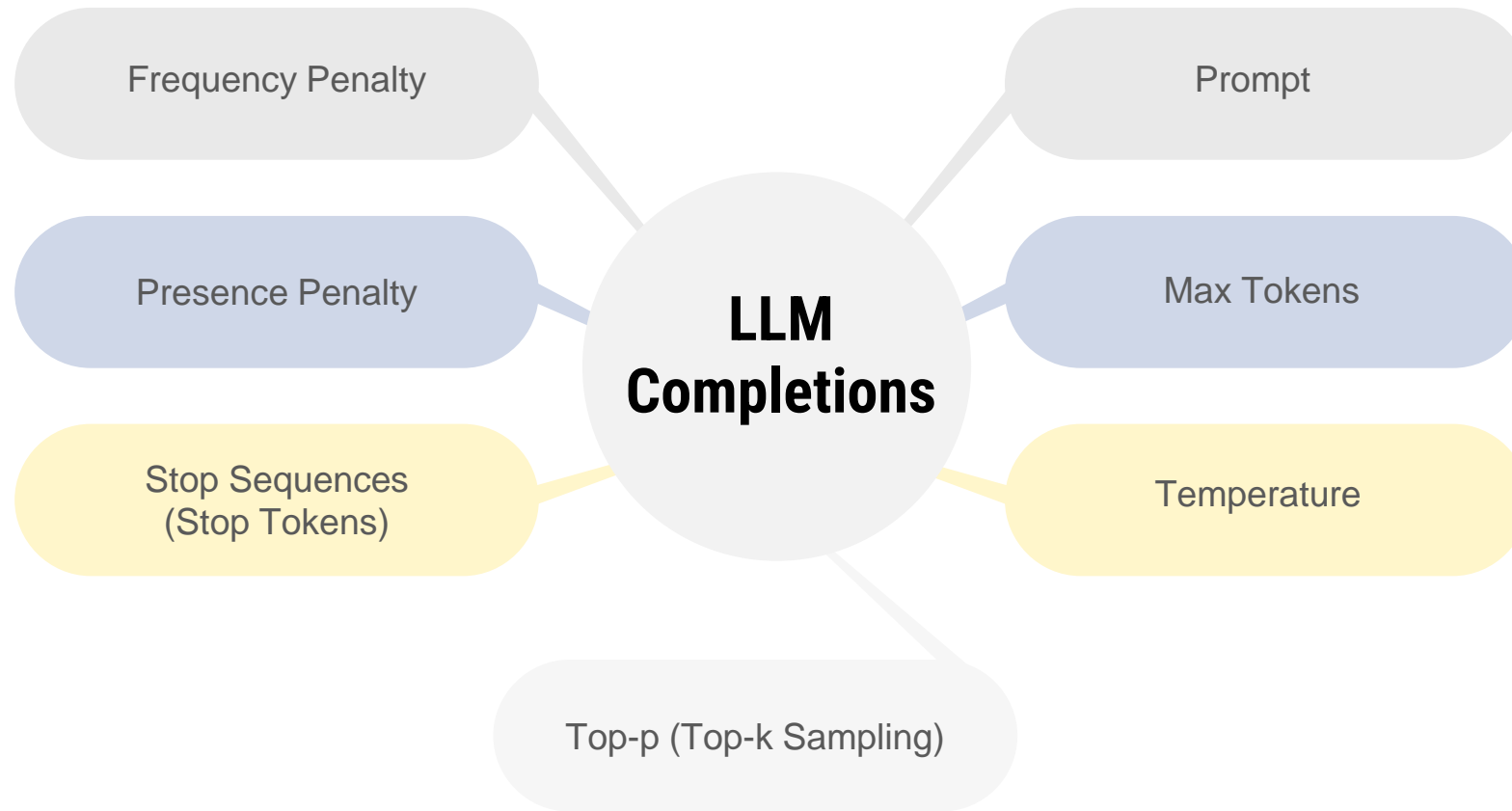
Hamburger => ham bur ger => 3 tokens

Images => 512-pixel square tiles => 170 tokens

## Language Models

Models	Context	Input (Per 1,000 tokens)	Output (Per 1,000 tokens)
GPT-4o Global Deployment	128K	\$0.005	\$0.015
GPT-4o Regional API	128K	\$0.005	\$0.015
GPT-3.5-Turbo-0125	16K	\$0.0005	\$0.0015
GPT-3.5-Turbo-Instruct	4K	\$0.0015	\$0.002

# LLM Completions Parameters



# Benchmarking and Leaderboards



Top 6 in Multitask Reasoning (MMLU)

Name	Score
GPT4o	88.7
Claude 3 Opus	86.8
GPT-4T 2024-04-09	86.5
GPT-4	86.4
Gemini Ultra	83.7
Gemini 1.5 Pro	81.9

Top 6 in Math (MATH)

Name	Score
GPT4o	76.6
GPT-4T 2024-04-09	72.2
Claude 3 Opus	60.1
Gemini 1.5 Pro	58.5
Gemini Ultra	53.2
GPT-4	52.9

Top 6 in coding (HumanEval)

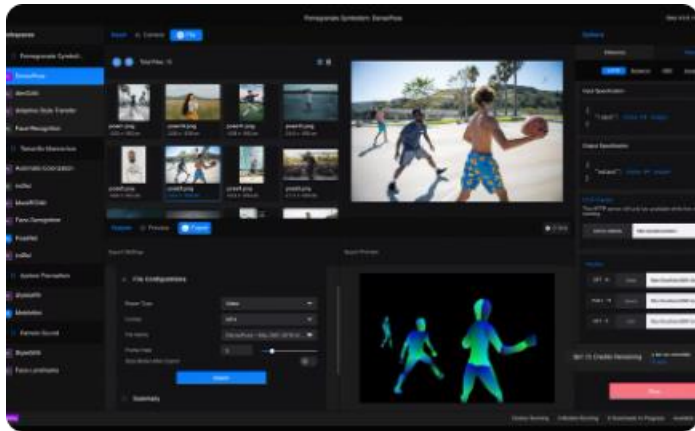
Name	Score
GPT4o	90.2
GPT-4T 2024-04-09	87.6
Claude 3 Opus	84.9
Llama 3 70B	82
Claude 3 Haiku	75.9
Gemini Ultra	74.4

MMLU - Multi-task language understanding

# Multimodality of AI

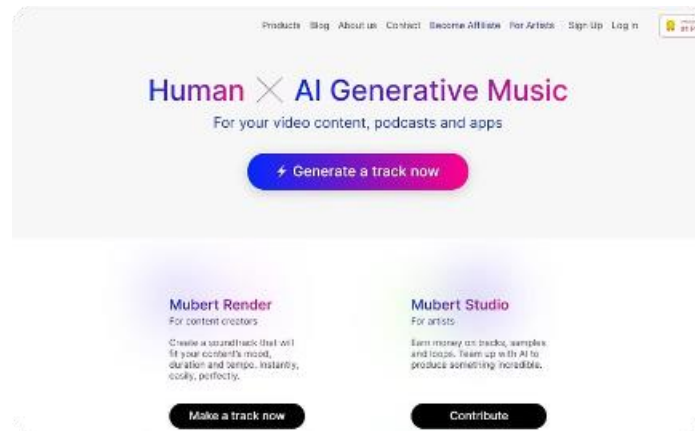


LLMs are evolving beyond text generation, creating and interacting with a broader range of content formats.



## Images

Create realistic images based on text prompts.



## Music

Generate unique musical compositions based on instructions and styles.






## Videos

Produce short video clips with diverse styles and effects.

# Enterprise Applications of LLMs



LLMs have found numerous applications in enterprise settings, automating tasks and enhancing user experiences.

	<b>Conversational Agents</b>	Provide personalized and efficient customer service through automated chat interfaces.
	<b>RAG Applications</b>	Extract insights from unstructured documents like PDFs to answer questions and provide summaries.
	<b>Task Automation</b>	Automate tasks based on email sentiment analysis, data extraction, and more.



# Thank you!

DATA FACT | Z

22260 Haggerty Rd, Northville, MI  
1 866 739 0820  
[hello@datafactz.com](mailto:hello@datafactz.com)