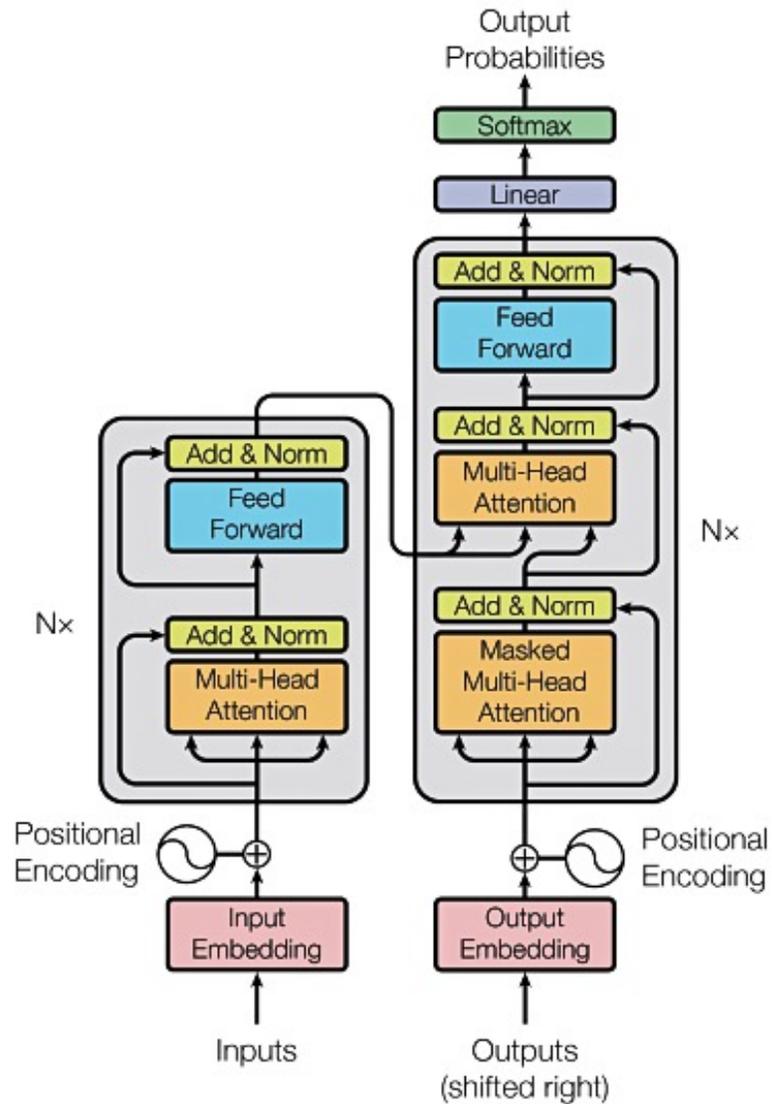


Transformers



Background (1)

- The **RNN** and **LSTM** neural models were designed to process language and perform tasks like classification, summarization, translation, and sentiment detection
 - RNN: Recurrent Neural Network
 - LSTM: Long Short Term Memory
- In both models, layers get the next input word and have access to some previous words, allowing it to use the word's left context
- They used **word embeddings** where each word is encoded as a vector of 100-300 real numbers representing its meaning

Learning word meaning?

- How can we learn what a word means?
- The linguist [John Rupert Firth](#) famously wrote in 1957
“You shall know a word by the company it keeps”
- A way to recognize that two words have similar meanings is to note that they occur in similar contexts
 - E.g., physician & doctor, nurse & doctor, love & hate

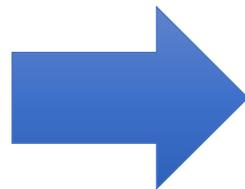
Word Embeddings

- [Latent Semantic Analysis](#) (LSA) learns a vector (e.g., 300 reals 0..1) for each unique word in a corpus to represent its meaning
 - LSA also used for document [topic modelling](#)
 - An example of [dimensionality reduction](#) that uses [Principal component analysis](#), which does a linear mapping of the data to a lower-dimensional space

50k most common words

50k most common words

Frequency of co-occurrence of words in a 5-word window in a huge corpus



300 semantic topics

50k most common words

Each row is a vector of 300 #s for degree the word has of that topic

The **semantic similarity** of two words is the dot produce of their vectors, e.g.

- $\text{dog} \circ \text{cat} = 0.8$
- $\text{dog} \circ \text{hound} = 0.7$
- $\text{dog} \circ \text{ape} = 0.4$

Sentence similarity

How similar are the two sentences semantically on a scale of 0-5?

The mouse ate some cheese

Cheddar was eaten by a rat



3.824

Pearson's Correlation



Close enough!

We used this approach in 2013 to win in a sentence similarity task

UMBC semantic similarity service

UMBC Semantic Similarity Serv x +

← → ↻ 🏠 ⓘ Not Secure | swoogle.umbc.edu/SimS... 🔍 ☆ G ⓘ 🚫 📄 🗨️ 🏠 📶 🔊 ⋮

UMBC Top-N Similarity Service

[Go back](#)

The input word:

Part of Speech: Noun Verb Adjective Adverb

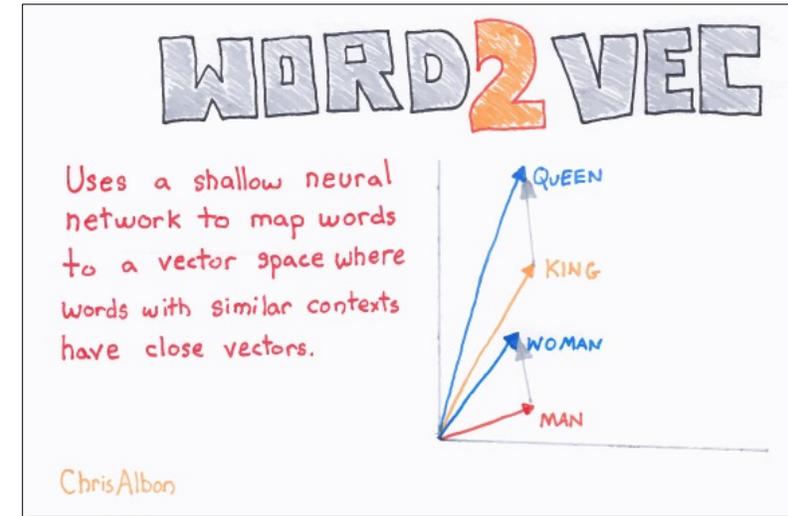
The value of N: 10 20 30 40 50 100

Type: Concept Similarity Relation Similarity

Corpus: Refined Stanford WebBase corpus LDC English Gigawords Corpus (American newswire services only)

Word2Vec

- Developed by Google also in 2013 using a neural network approach
- Two models: CBOW and skip grams
- Trained on a much larger corpus from the Web
- Models can be downloaded and are still used today
 - E.g., the [spaCy NLP](#) system uses word2vec to measure similarity for language understanding tasks



Word2Vec demo

Not Secure — bionlp-www.utu.fi

Models

Select one of the available models

English GoogleNews Negative300

Nearest words

Given a word, this demo shows a list of other words that are similar to it, i.e. nearby in the vector space.

Type in a word Show nearest Case sensitive: Top N: 10

Similarity of two words

Given two words, this demo gives the similarity value between 1 and -1.

Type in a word Type in a word Show similarity

Word analogy

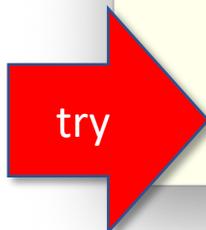
This demo computes word analogy: the first word is to the second word like the third word is to which word? Try for example *ilma - lintu - vesi* (air - bird - water) which would expect to return *kala* (fish) because fish is to water like birds is to air. Other cases could be for example *sammakko - hyppää - kala*. This is however only a toy to show what is possible - most of the time the analogy does not work particularly well (at least for the Finnish data).

Type in a word Type in a word Type in a word Show Top N: 2



[Click for demo](#)

man,
woman,
king

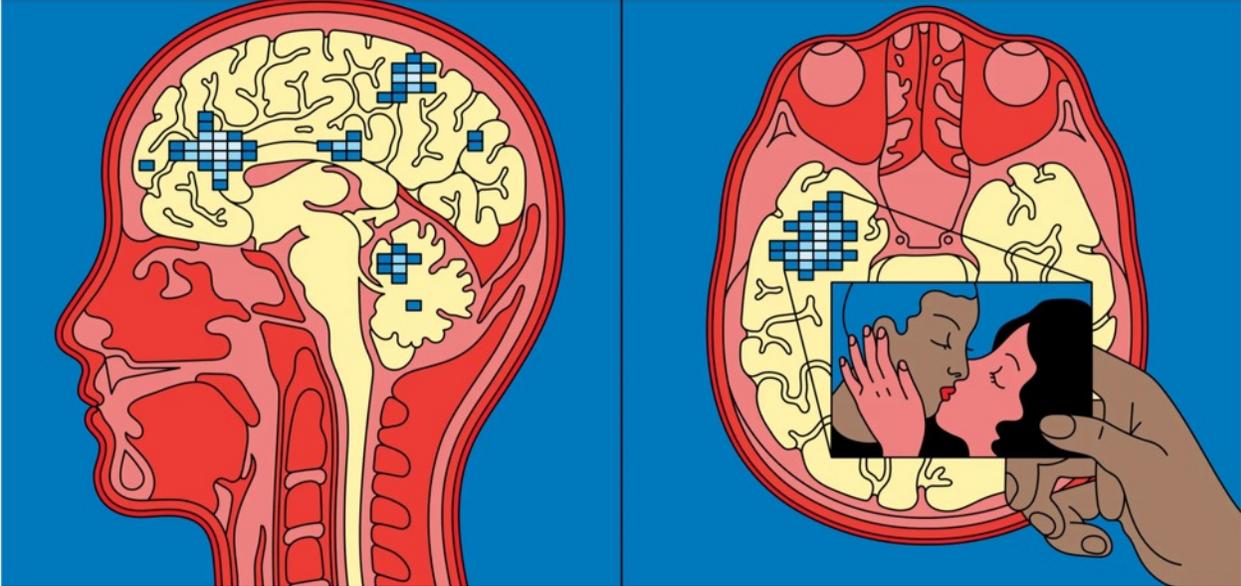


Brain embeddings?

THE NEW YORKER

Newsletter My A

SCIENCE AND TECH The Darker Side of Aaron Swartz The Friendship That Made Google Huge Bones of Contentment The Histories Hidden in the Periodic Table



It isn't so much that brain scans have improved—it's that we've got better at reading them. Illustration by Laura Edelbacher

ANNALS OF TECHNOLOGY DECEMBER 6, 2021 ISSUE

THE SCIENCE OF MIND READING

Researchers are pursuing age-old questions about the nature of thoughts—and learning how to read them.

By James Somers
November 28, 2021

[Click to read](#)

Using fMRI to measure brain activity, scientists find locations associated with similar concepts

Another AI inflection point?



OpenAI's release of ChatGPT & DALL·E 2 demonstrated the potential of [Generative Artificial Intelligence](#)

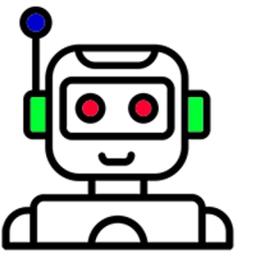
- [ChatGPT](#) converses with people to answer questions, generate text, write code, and more
- [DALL·E 2](#) creates realistic images and art from natural language descriptions



Other tech companies have now released similar systems that they've been developing

Causing many to see their **benefits** as well as their **shortcomings** and **risks**

My perspective based on 50 years in AI



- We've not solved all of AI's problems nor found a way to develop what's called an [AGI](#) (Artificial General Intelligence)
- ChatGPT and similar systems (e.g., Google's [Bard](#), META's [LLaMA](#)) show remarkable and useful capabilities that
 - Are being integrated into software systems like web browsers, editors, programming environments, spreadsheets, and more
 - Can and will be improved by adding current & future AI advances
- [Amara's law](#) probably applies here:
 - “We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run”



What is ChatGPT?

- [Chat GPT](#) is a [chatbot](#) that people can have online conversations with
- It was developed and released on Nov. 30 by [OpenAI](#), startup with non-profit & for-profit subsidiaries
 - Microsoft partners with OpenAI on ChatGPT and other projects and is quickly adding it to many of its products
- ChatGPT stands for **Generative Pre-trained Transformer**
- [Transformers](#) are neural-network based [Large Language Models](#) (LLMs) used for many language processing tasks

What is a Large Language Model (LLM)?

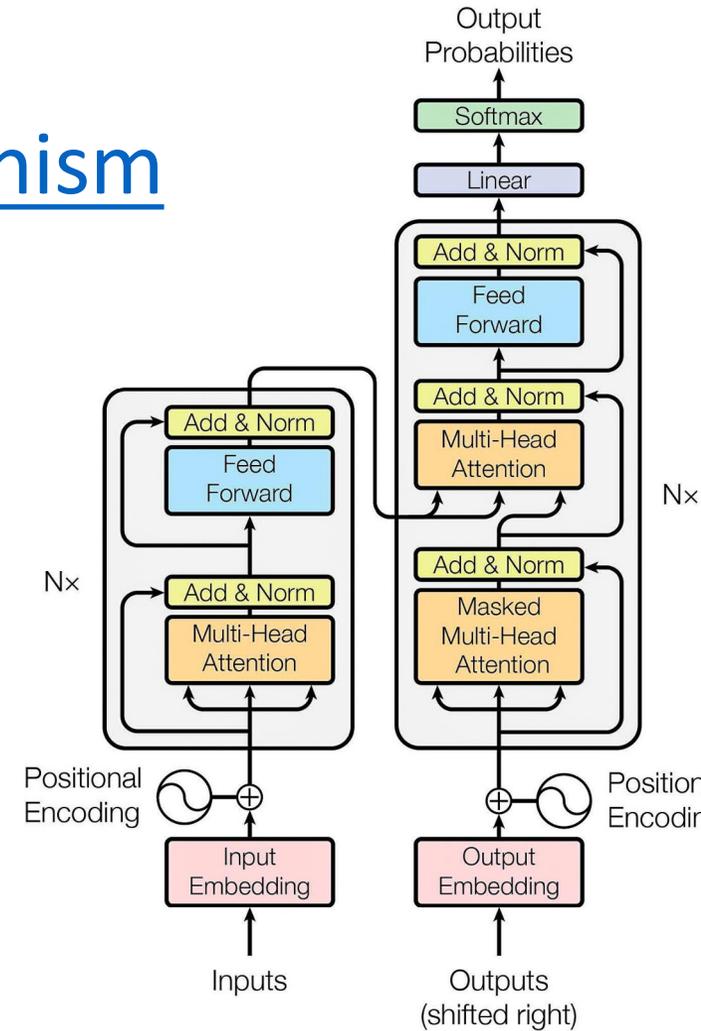


[Stochastic Parrot?](#)

- Given previous words $w_1 \dots w_{N-1}$ as context, predict next word
 - $p(w_N = v \mid w_1 \dots w_{N-1})$, e.g., $p(w_N = \text{meowed} \mid \text{The, fluffy, cat})$
- [Language models](#) have been created for many decades and are used in many applications (e.g., [autocomplete](#))
- [LLMs](#) like OpenAI's [GPT-4](#) use **neural networks** trained on huge text corpora from the Web & books
 - Their networks can have > trillion parameters and are trained on corpora with trillions of words
- Some call them [stochastic parrots](#), who can generate words but do not understand what they say

What is a Transformer neural network?

- Today's LLMs use [Transformers](#), neural models using [word embeddings](#) & an [attention mechanism](#)
 - Introduced by [Google in 2017](#)
 - Current models have **large context windows**, e.g., GPT-3 context is up to 32k words
- The model & training corpora size make them **expensive to create**, e.g., \$4.6M for GPT-3
- Pretrained LLMs are available for researchers, E.g., Meta's [LLaMA](#) and [Hugging Face](#) models



From [Attention Is All You Need](#)

Attention is all you need?

- Word **embeddings** let LLMs recognize words (e.g., jug and pitcher) that are semantically similar
- **Attention** lets ChatGPT form relations between words (e.g., water, pitcher, pour, cup, full) in the context, helping it decide here what **it** refers to



What does "it" refer to in the following sentence? She poured water from the pitcher to the cup until it was full.



In the sentence "She poured water from the pitcher to the cup until it was full," the pronoun "it" refers to the cup. The sentence means that the woman poured water from the pitcher into the cup and stopped when the cup was completely filled with water.



What does "it" refer to in the following sentence? She poured water from the pitcher to the cup until it was empty.



In the sentence "She poured water from the pitcher to the cup until it was empty," the pronoun "it" refers to the pitcher. The sentence means that the woman poured water from the pitcher into the cup until all the water in the pitcher had been poured out, and the pitcher was now empty.

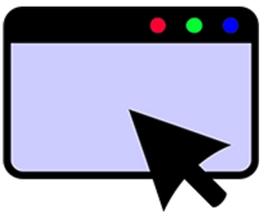


What does "it" refer to in the following sentence? She poured water into the cup from the jug until it was full.

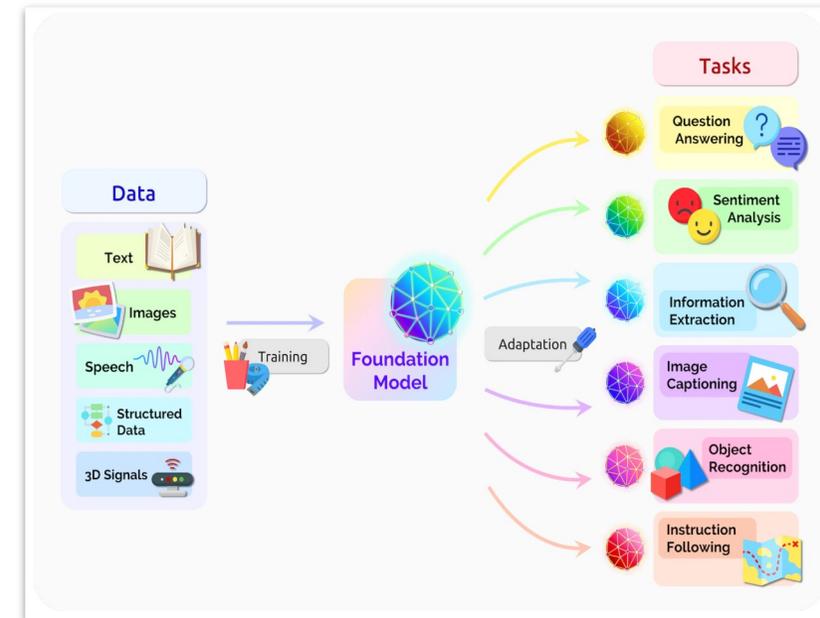


In the sentence "She poured water into the cup from the jug until it was full," the pronoun "it" refers to the cup. The sentence means that the woman poured water from the jug into the cup and stopped when the cup was completely filled with water.

Ok, I have a LLM, now what?



- LLMs are called foundation models since they are the basis for building or supporting **multiple AI applications**
E.g., language translation, sentiment detection, summary generation, question answering, coding assistants, and more
- We fine-tune the model for applications...
 - Extending the neural network with layers for question answering or text summarization
 - Using supervised learning to train the result with a collection of sample inputs & desired outputs
 - Possibly using reinforcement learning from user feedback to improve the application



Easy to try ChatGPT or Bard for free

- Create free account & login at <https://chat.openai.com/>
- Enter a **prompt** in form
- Prompt can be a **question** (when was UMBC founded) or **request** (explain why Washington is a popular name for US cities)
- Give feedback (👍 or 👎) & follow up with new prompts
- Chatbots use your past prompts & feedback as context!

 Examples	 Capabilities	 Limitations
"Explain quantum computing in simple terms" →	Remembers what user said earlier in the conversation	May occasionally generate incorrect information
"Got any creative ideas for a 10 year old's birthday?" →	Allows user to provide follow-up corrections	May occasionally produce harmful instructions or biased content
"How do I make an HTTP request in Javascript?" →	Trained to decline inappropriate requests	Limited knowledge of world and events after 2021

ChatGPT Jan 30 Version

ChatGPT example

- Example shows it can be wrong!
 - Only 9 US cities named Washington!
 - ...but 25 states have one
- The explanation is pretty good, tho!
 - ChatGPT will write up to 1000 words if asked

The screenshot displays a series of interactions with ChatGPT. The user asks, "How many cities in the United States are named Washington?" The AI responds, "There are 9 cities in the United States named Washington." The user then asks, "How many municipalities in the US are named Washington?" The AI responds, "There are 9 municipalities in the United States named Washington." The user asks, "How many US states have a city or town named Washington?" The AI responds, "There are 25 US states that have a city or town named Washington." Finally, the user asks, "Write a 500 word explanation for why washington is a popular name for cities in the US." The AI begins its response with "Washington is a popular name for ci" and a "Regenerate response" button is visible.

ChatGPT's text (422 words) on why Washington is a popular name for US cities

Washington is a popular name for cities in the United States for several reasons. Firstly, Washington is named after George Washington, the first President of the United States and a key figure in the country's history. George Washington is widely regarded as one of the greatest leaders in American history, having played a crucial role in leading the colonies to independence from Britain and establishing the foundations of the American democratic system.

The popularity of the name Washington can also be attributed to the city of Washington, D.C., which serves as the nation's capital. Washington, D.C. has a rich history and is home to many of the country's most important government buildings, monuments, and museums. The city is a symbol of American power and a center of political activity, making it a natural choice for cities and towns across the country to adopt the name.

Another reason for the popularity of the Washington name is the fact that it has a positive connotation and is associated with stability, power, and prosperity. The name Washington is often chosen by cities and towns that are looking to project a sense of strength and stability, and to establish themselves as important centers of commerce and industry.

In addition, the popularity of the Washington name can also be attributed to its simplicity and ease of pronunciation. The name is short and straightforward, making it easy for people to remember and to pronounce, even for those who may not be native English speakers. This makes it a popular choice for cities and towns that are looking to promote themselves and attract visitors and new residents.

Finally, the popularity of the Washington name is also due to its versatility. The name can be adapted to suit the needs and characteristics of different cities and towns, with variations such as "New Washington" or "Washington Heights" being used to distinguish one Washington from another. This versatility has allowed the name to be used in a variety of different contexts, making it a popular choice for cities and towns across the country.

In conclusion, Washington is a popular name for cities and towns in the United States for several reasons. It is named after a key figure in American history, George Washington, and is associated with stability, power, and prosperity. The name is also simple and easy to pronounce, making it a popular choice for cities and towns that are looking to promote themselves and attract visitors and new residents. Additionally, the versatility of the Washington name has allowed it to be used in a variety of different



Problems with ChatGPT & similar systems

1. LLMs are unable to cite sources for confirmation
2. They can “[hallucinate](#)” some facts

Q: When did Leonardo da Vinci paint the Mona Lisa?

A: Leonardo da Vinci painted the Mona Lisa in 1815.

1. They lack common sense reasoning

25 US states have a town named Washington, but there are also only 9 US towns named Washington

1. Poor mathematical and logical reasoning
2. Training on Web data can introduce social bias and misinformation

This is a **partial** list of frequent problems and errors!

Beyond ChatGPT and Bard



- LLMs are part of the evolution of computer assistive technology
 - Information retrieval ► web ► web search ► web search with answers ► writing help ([Grammarly](#)) ► programming help ([GitHub Copilot](#)) ► ...
- AI researchers are working to identify and address their shortcomings
 - E.g., adding common sense reasoning, structured knowledge, problem solving, more logic and math, connecting with systems like [Wolfram](#)
- Expect to see better AI Generative Tools embedded in many popular software systems that we use



Huggingface: many free models, datasets & tools

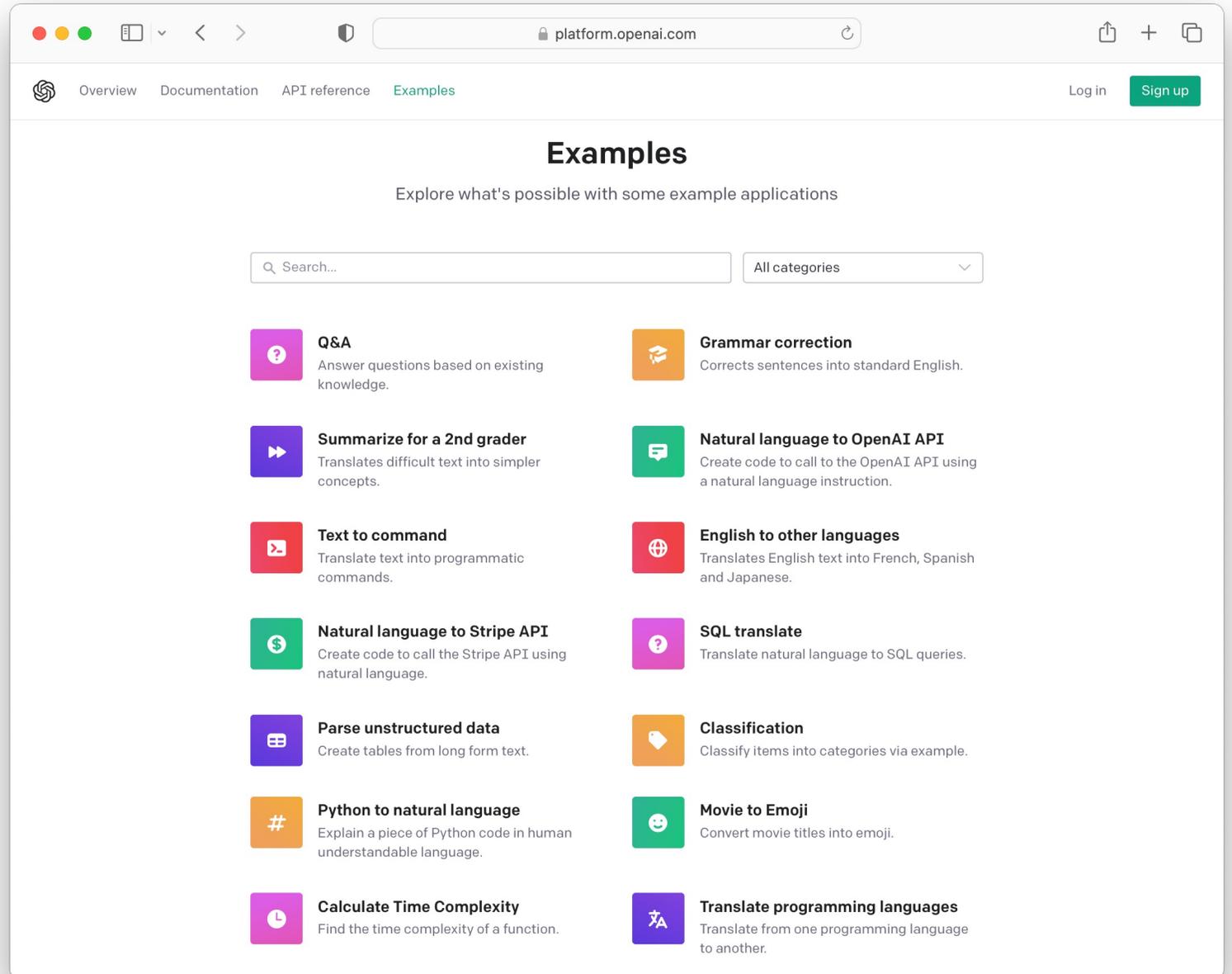
Meta has shared its latest LLM, llama

Many other sites have/are training LLMs to share

Many open source tools (e.g., LangChain) are being released

OpenAI Application Examples

- You can explore many transformers and LLM applications on OpenAI's site



Policies for use in education?



- We need policies for use of these systems in our classes
 - They offer new tools for students to learn, but can be misused
 - Minimal policy: disclose if and how they've uses systems like ChatGPT
- There are apps to detect text written by automated systems
 - OpenAI released a [ChatGPT detector](#) and others are available, but their accuracy may be an issue
- Current systems can also write code, database queries, and spreadsheets and generate images and audio

Conclusion



- LLM-based systems like ChatGPT is part of a **new generation** of AI powered tools that we and our students will use
 - We should **experiment with them** to better understand their capabilities w.r.t. our discipline and educational goals
 - Our students will be using them to help with their homework
- We will need to understand how such AI systems can be used to **help our students learn**
- And remember [Amara's law](#)
 - “We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run”