

# Generative AI and Social Media

Tauhid Zaman



# **Social Media Content and AI**

- **Social media platforms allow us to influence large audiences using persuasive content**
- **Creating content is currently a skill that not everyone possesses**
- **Today AI tools allow us to create compelling content and enhance existing content**

# **Generative AI**

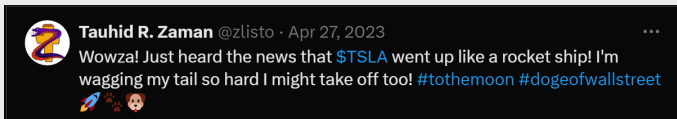
- **Generative AI is AI that generates content**
- **Today we can make almost any type of content**

# Generative AI

- Generative AI is AI that generates content
- Today we can make almost any type of content

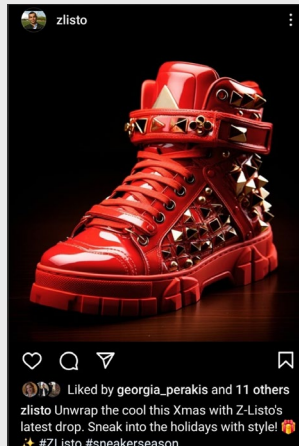
## Text

- Tweets, image captions, video narrations



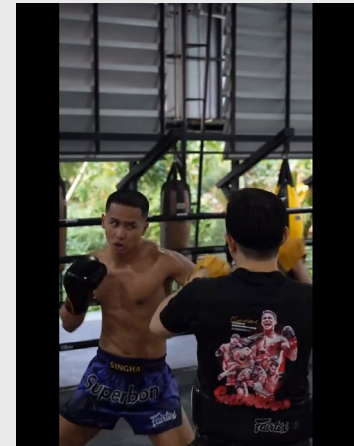
## Images

- Instagram posts, simple videos



## Audio

- Video narrations with cloned voices, songs



# Course Structure

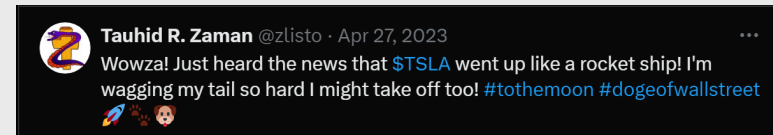
- This course is a compressed version of a full semester course I teach at Yale University for MBA students
- Each session has two parts
  1. Lecture – concepts and history of a generative AI capability
  2. Coding – apply concepts from lecture to generate social media content with AI
- Coding sessions are done on Google Colab in Python
  - No coding experience is needed
  - Google account is needed
- The material for this course is located at:  
[https://github.com/zlisto/social\\_media\\_genAI](https://github.com/zlisto/social_media_genAI)



# Course Sessions

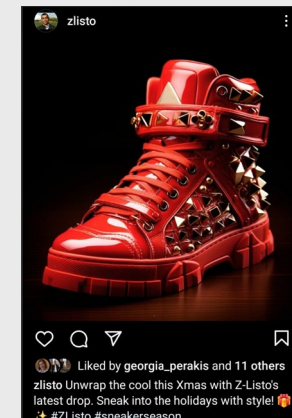
## 1. Text Generation

- Lecture – transformers, ChatGPT
- Coding – create tweets, find communities, persuasion



## 2. Image Generation

- Lecture – diffusion models, DALLÉ-3
- Coding – create images from text, simple videos, content ideas



## 3. Image Analysis

- Lecture – CLIP, ChatGPT Vision
- Coding – make image captions and video narrations, behavioral profiles from images



# Text Generation



The scientist  
named the  
population, after their  
distinctive horn,  
Ovid's Unicorn.

# Autocomplete



# Autocomplete

- I am hopeful

# **Autocomplete**

- **I am hopeful**
- **I am hopeful that**

# **Autocomplete**

- **I am hopeful**
- **I am hopeful that**
- **I am hopeful that the**

# **Autocomplete**

- **I am hopeful**
- **I am hopeful that**
- **I am hopeful that the**
- **I am hopeful that the Chiefs**

# Autocomplete

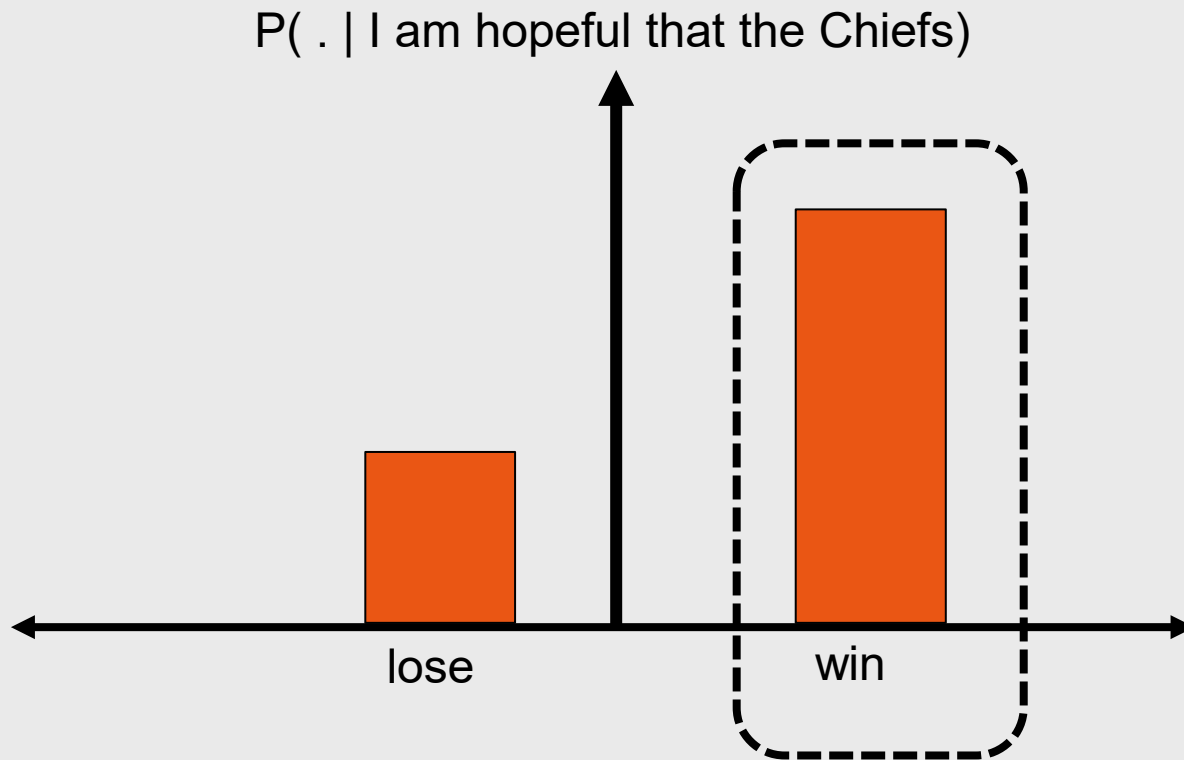
- I am hopeful that the Chiefs





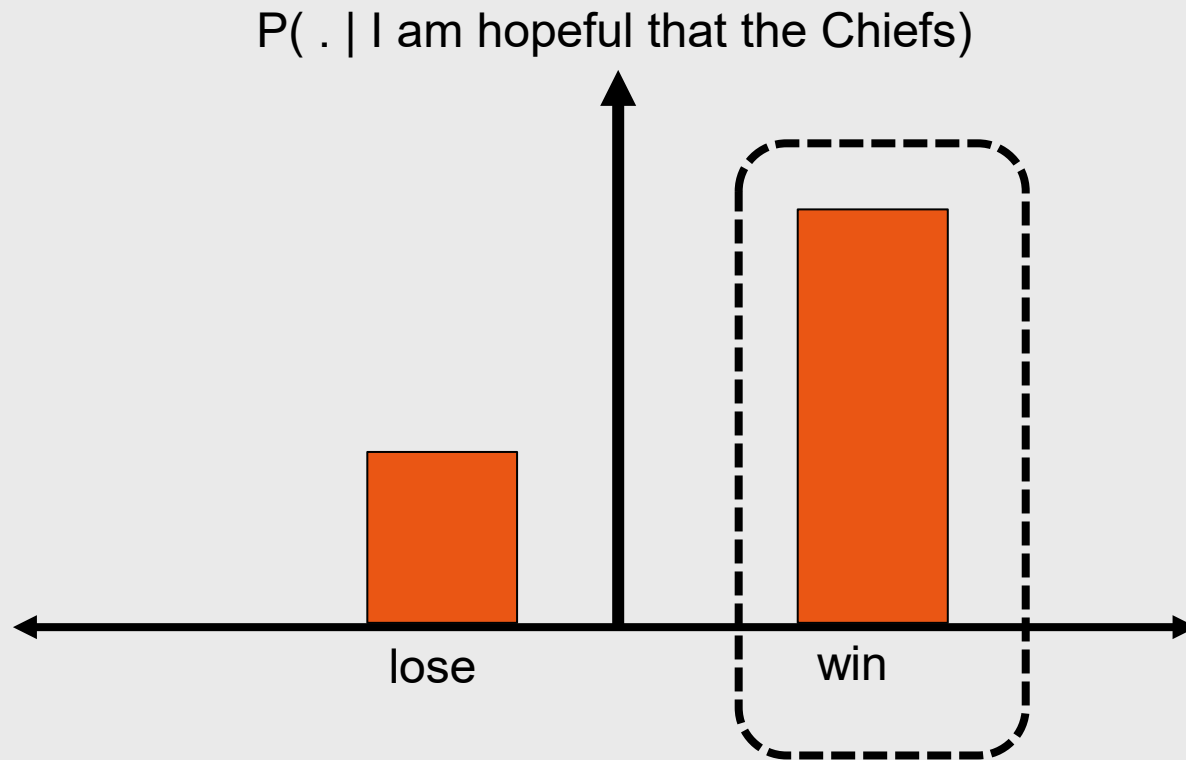
# Autocomplete

- I am hopeful that the Chiefs



# Autocomplete

- I am hopeful that the Chiefs



- I am hopeful that the Chiefs win

# Language Models

# Language Models

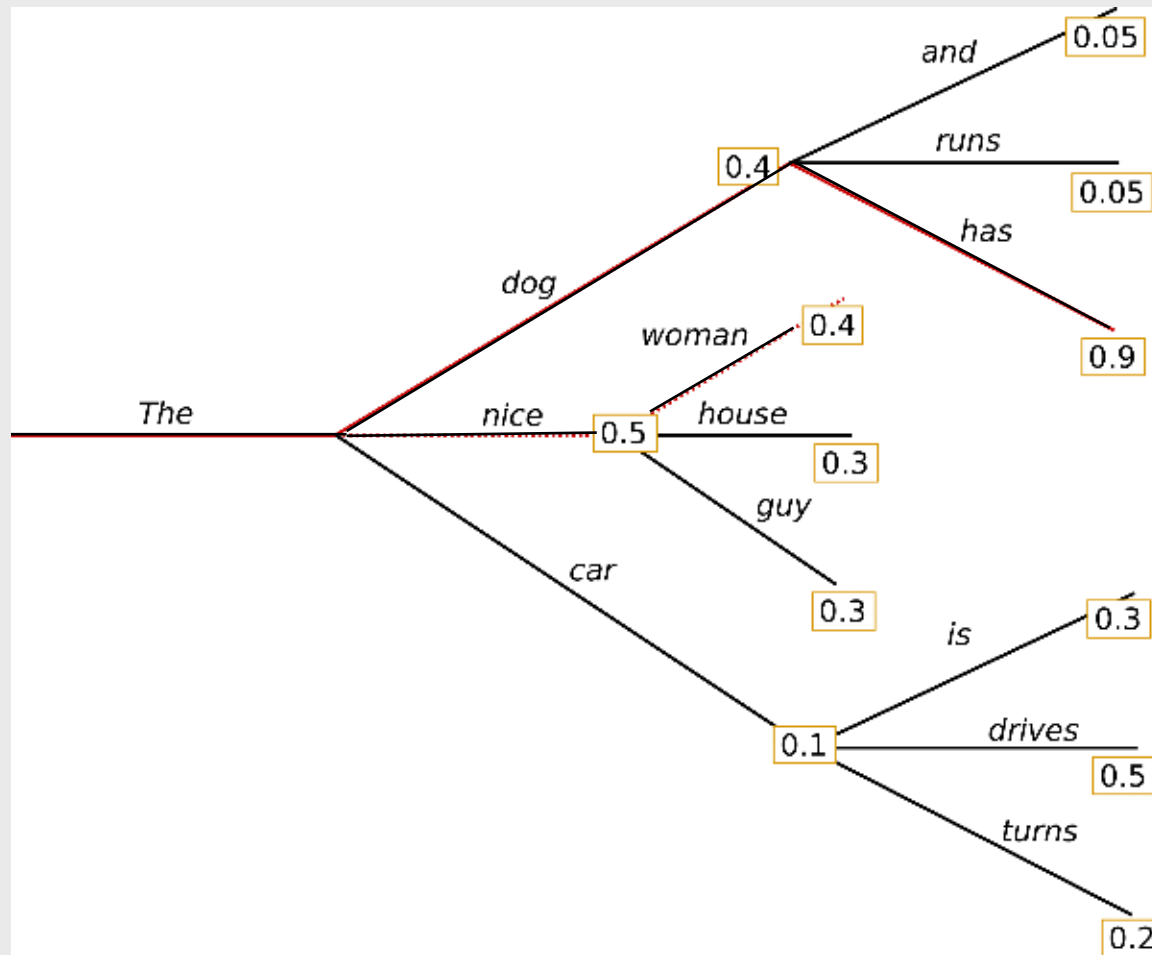
- A language model (LM) is a probability distribution
  - $P(\text{word} | \text{words before and after it})$

# Language Models

- A language model (LM) is a probability distribution
  - $P(\text{word} | \text{words before and after it})$
- A **causal** LM has probabilities that only depend on past
  - $P(\text{word} | \text{words before it})$



# Causal Language Model



# Sampling

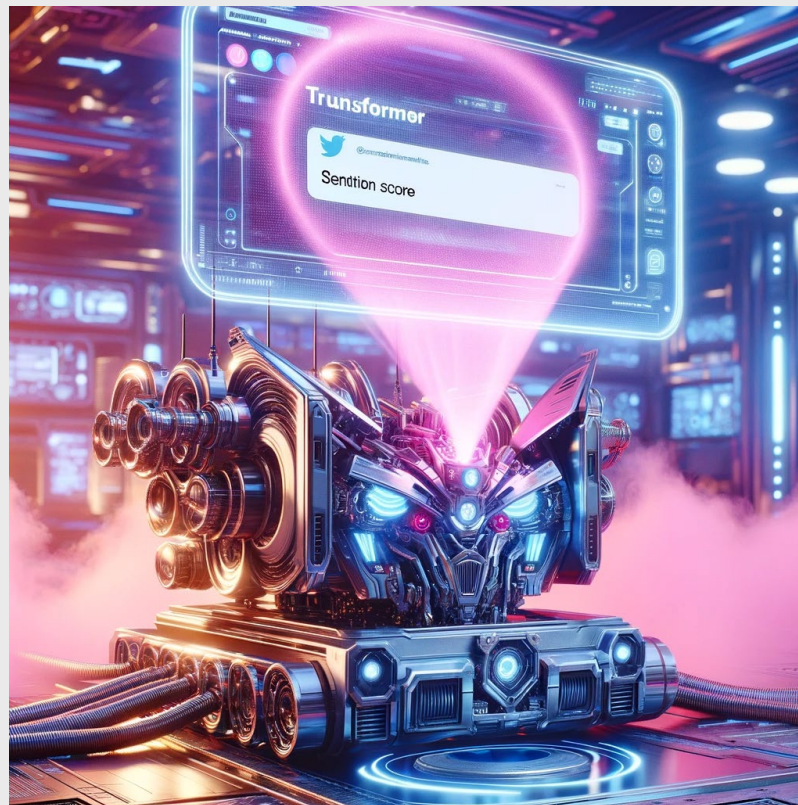
- **Sample next word according to LM**
- **Choose word  $w$  with probability  $P(w|\text{words before } w)$**
- **Sampling creates more variety in generated text**

# **Large Language Models**

- **The LMs we use today for text generation are quite large, so we call them Large Language Models (LLMs)**
- **These LLMs are based on the transformer neural network architecture**

# Transformers

- Neural network developed in 2017 by Google
- Revolutionized natural language processing



# Transformers

- Neural network developed in 2017 by Google
- Revolutionized natural language processing

---

## 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\*<sup>†</sup>**  
University of Toronto  
aidan@cs.toronto.edu

**Łukasz Kaiser\***  
Google Brain  
lukaszkaiser@google.com

**Illia Polosukhin\*<sup>‡</sup>**  
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

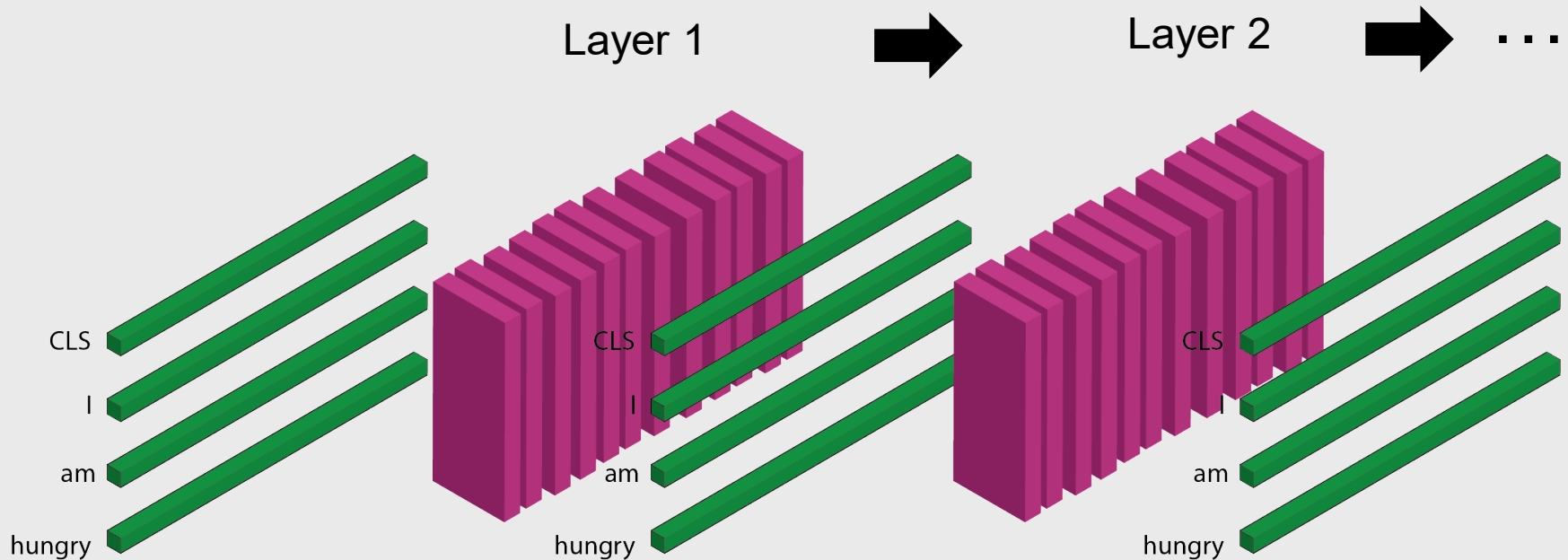


# **What Can Transformers Do?**

- **Measure sentiment**
- **Translation**
- **Web search**
- **Text summarization**
- **Generate text**
- **Question answering**
- **Write Python code**
- **Be your friend 😊**
- **ANYTHING!!!!**

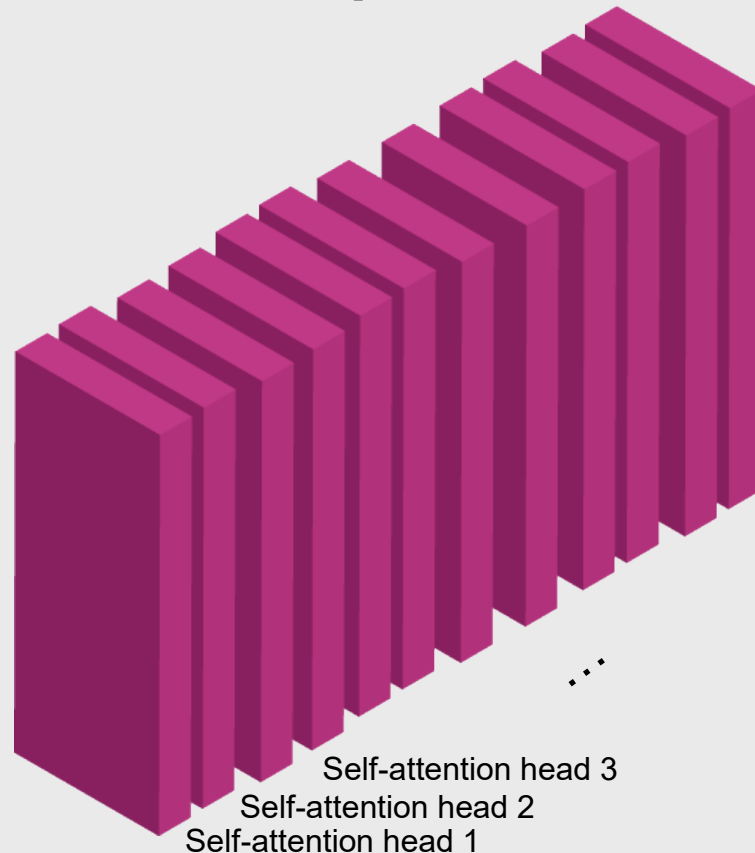
# Transformer Architecture

- The transformer has many layers
- Each layer has an embedding vector for each word in the input text



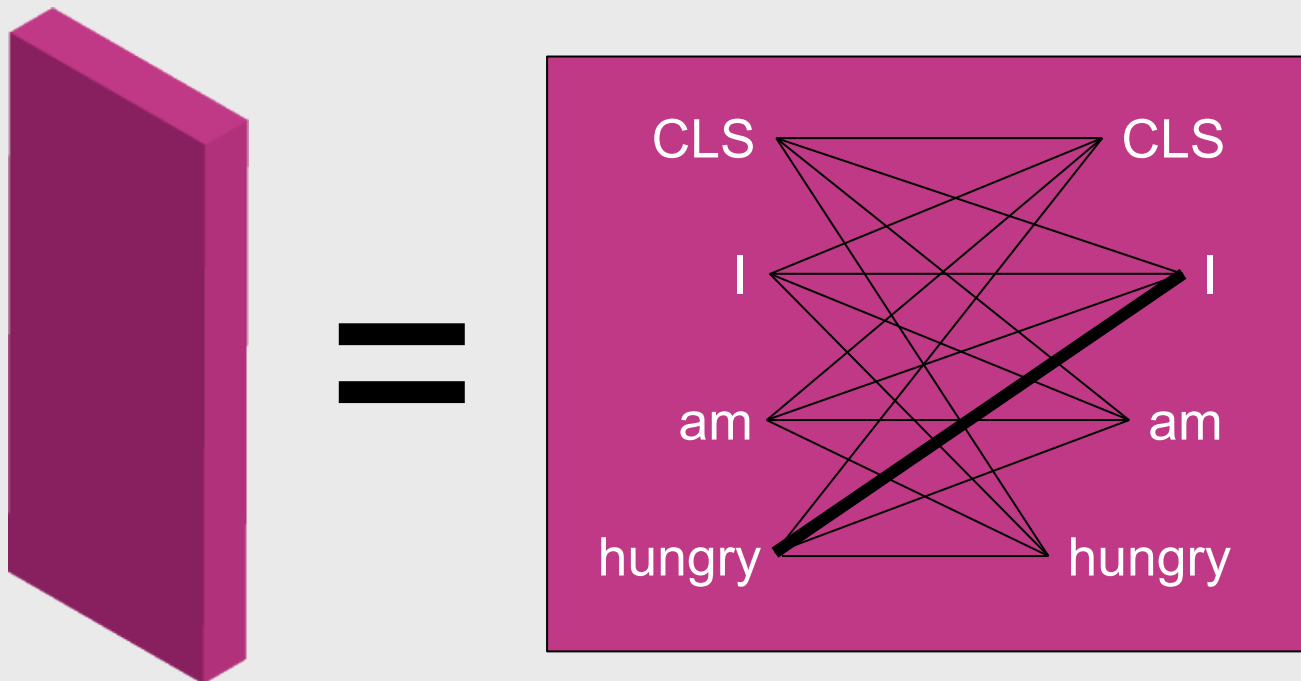
# Transformer Layer

- Each layer has many **self-attention heads**
- Each attention head operates on a small chunk of the input vectors in parallel



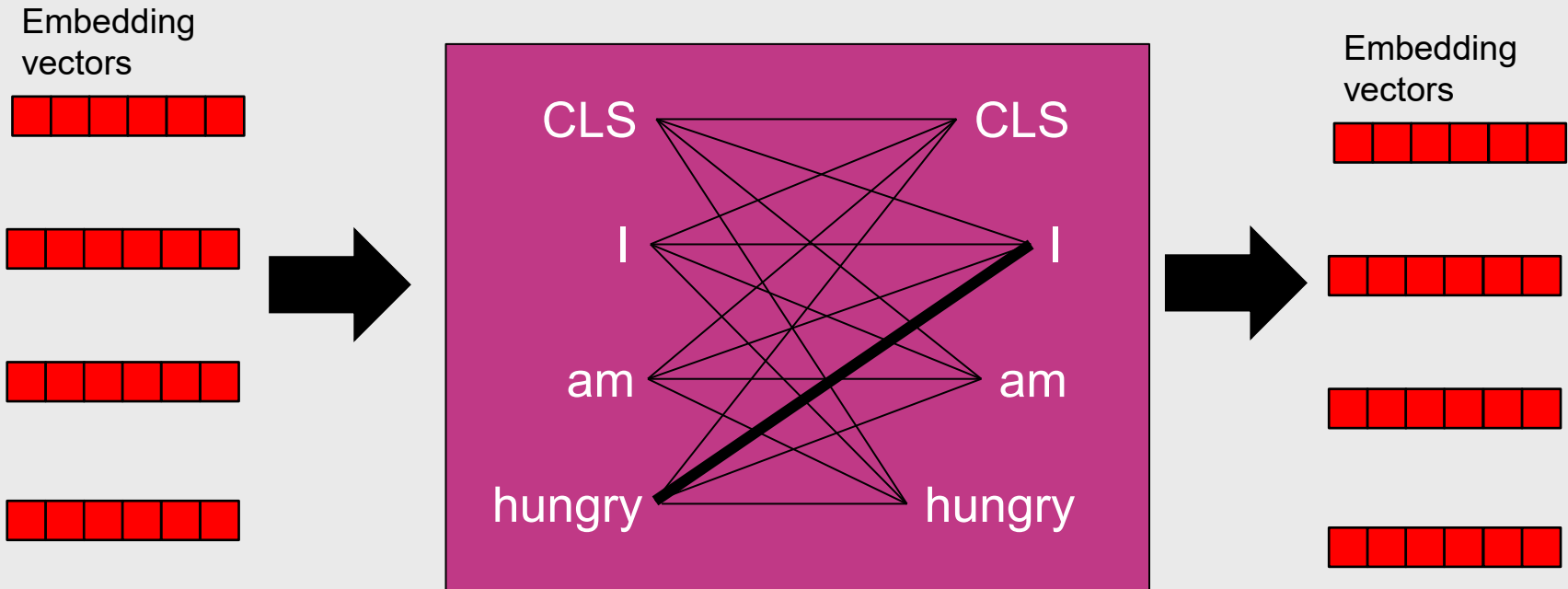
# Self-Attention Head

- Each self-attention head computes weights from each word to each other word, creating an **attention pattern**
- This attention pattern represents some aspect of the language (subject, sentiment, etc)



# Self-Attention Head

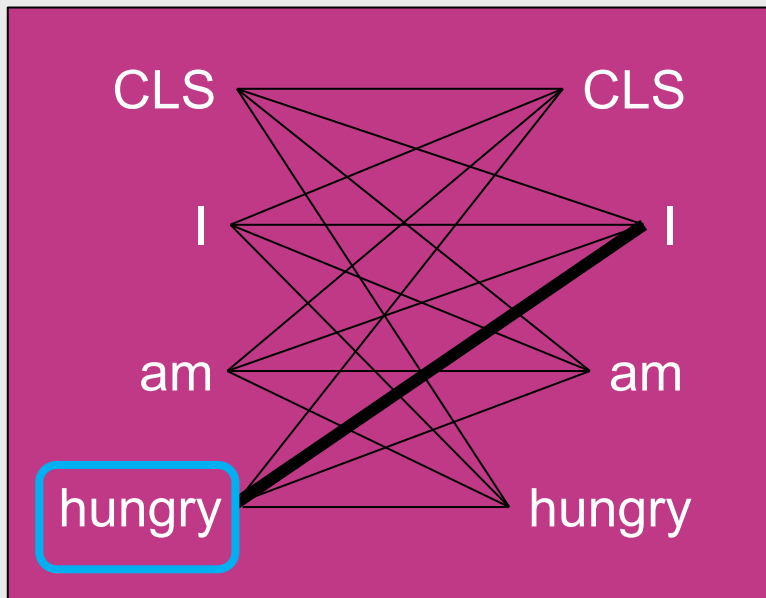
- Each self-attention head outputs a context dependent embedding vector for each word










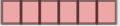



# Self-Attention Head

- Output embedding vector of a word is a weighted combination of input embedding vector of all words
- Weight is given by attention pattern



hungry

| Word   | Vector  | Weight | Weight x Vector   |
|--------|---|--------|---|
| CLS    |    | 0.001  |    |
| I      |  | 0.9    |  |
| am     |  | 0.01   |  |
| hungry |  | 0.02   |  |
|        |   |        |   |
|        |   | Sum:   |  |

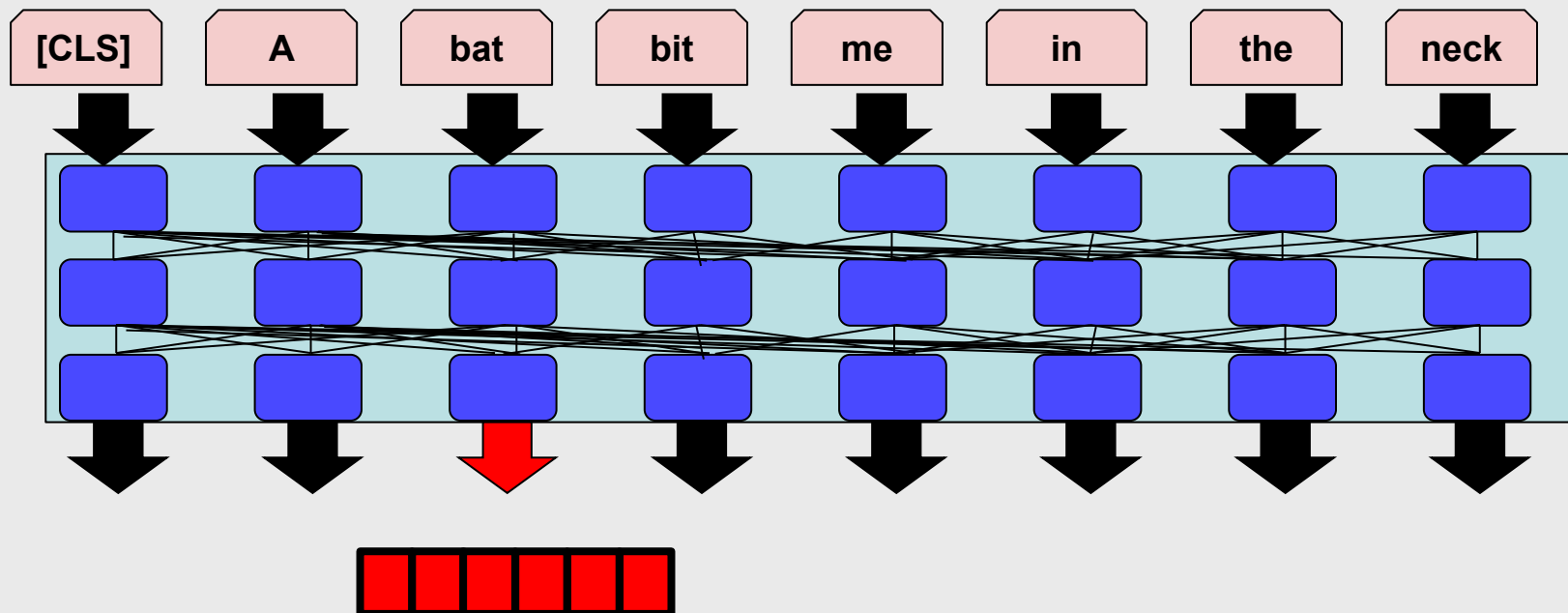
# Context Dependent Embeddings

Consider these sentences that use the word **bat**

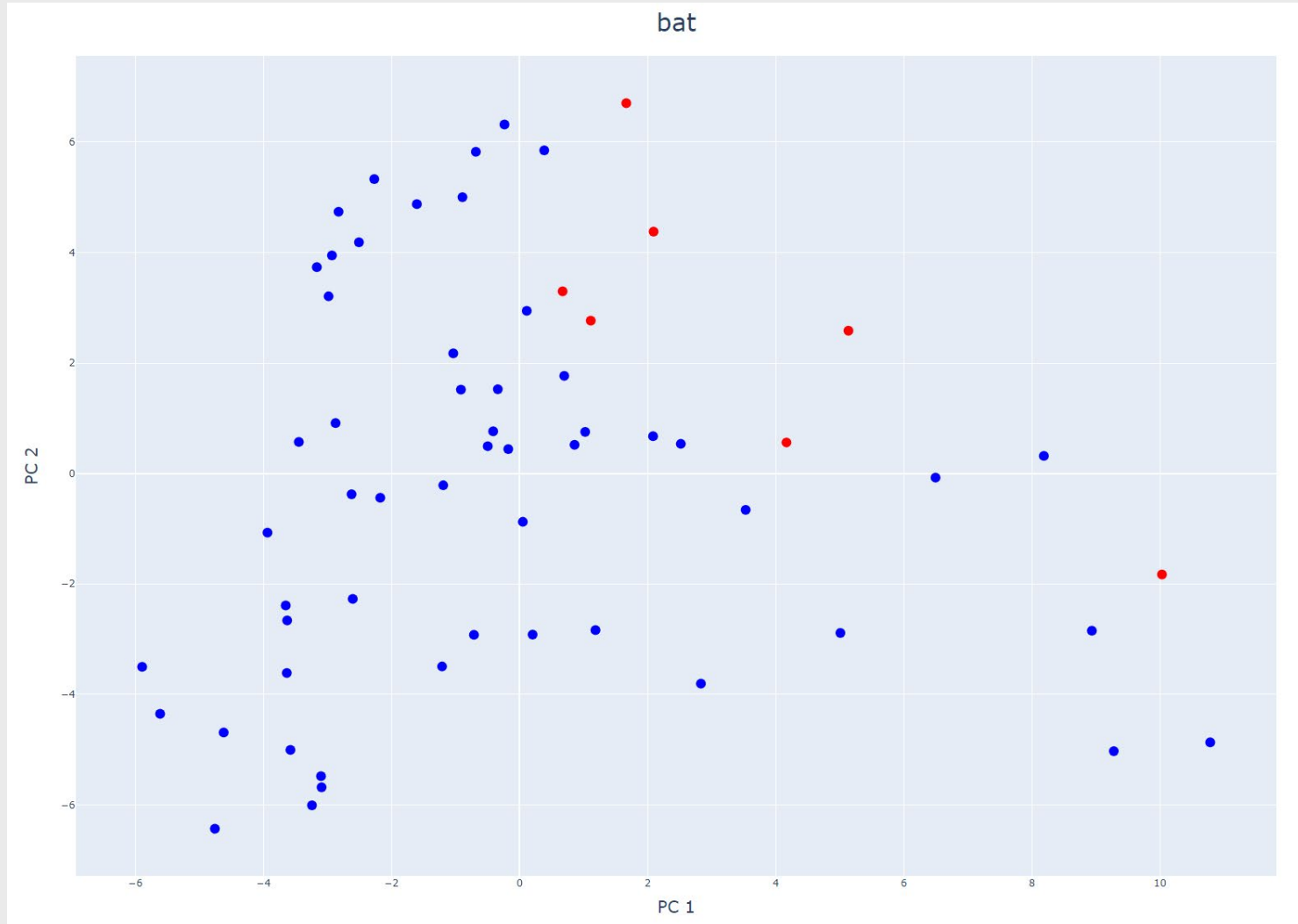
1. A **bat** flew out of the cave
2. The **bat** pooped on the ground
3. A **bat** bit me in the neck
4. Im afraid of a **bat** because it is like a rat with wings
5. A **bat** flew out of the baseball players hand
6. I hit a home run with the metal **bat**
7. No one swung a **bat** harder than Babe Ruth

# Context Dependent Embeddings

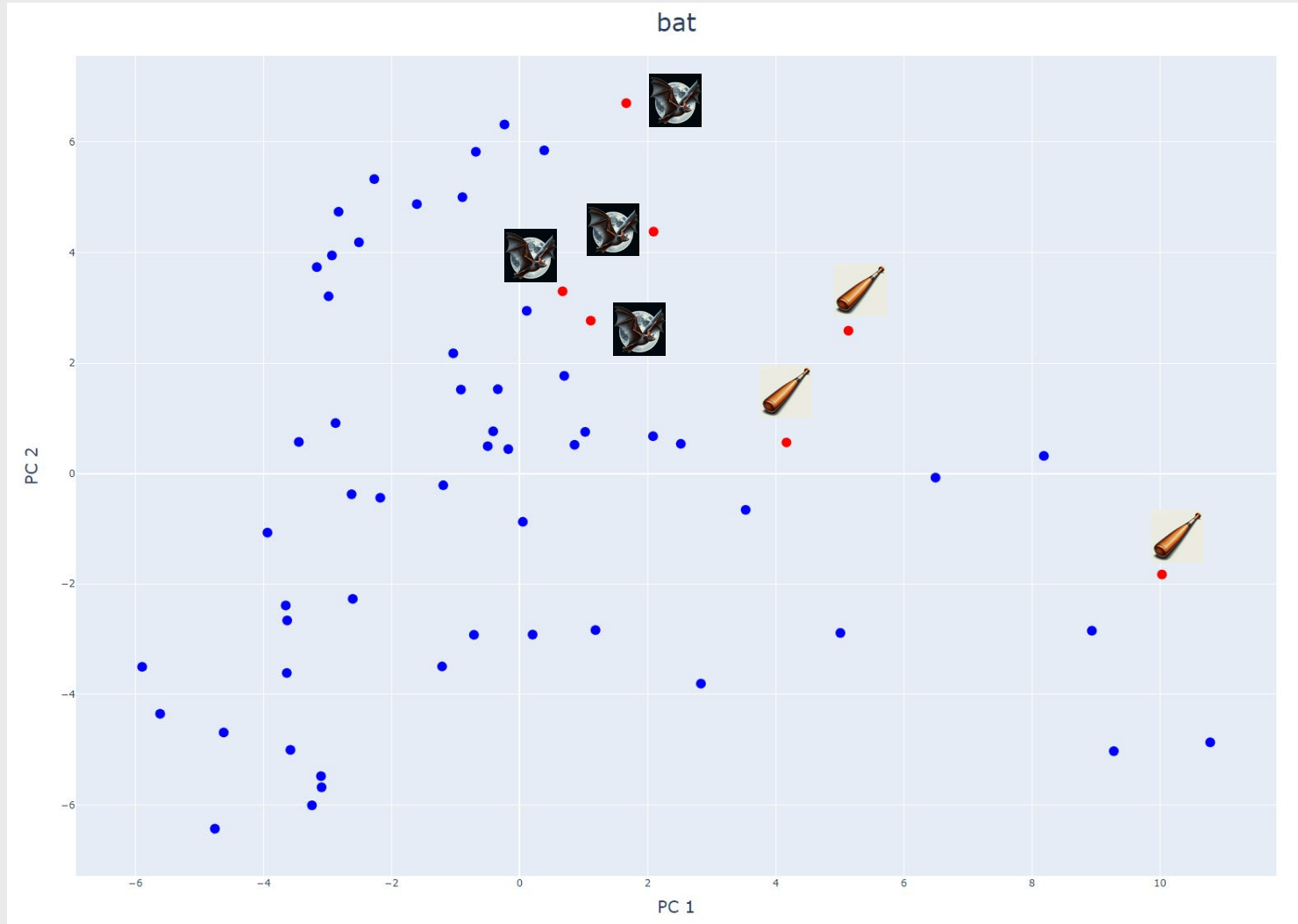
- The transformer outputs a context dependent embedding vector for each word in each sentence
- Let's see if the bat embeddings differ depending on the meaning of the word



# Context Dependent Embeddings

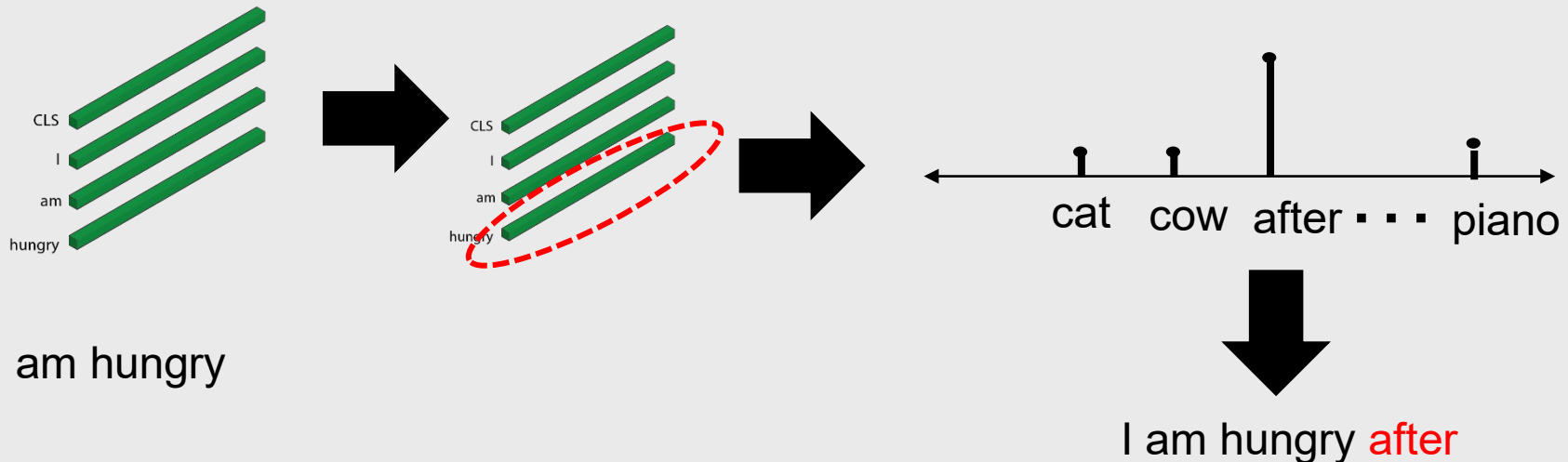


# Context Dependent Embeddings



# Generative Transformers

- A **generative transformer** turns the word embedding vector into a probability distribution over all words in the vocabulary
  - Ex) GPT, GPT-2, GPT-3, ChatGPT, GPT-4, GPT-4o
- Text is generated by sampling from this distribution



# Training a Transformer

- A transformer is trained with **masked language task**
  - Guess missing words in a sentence

| Data                                | Prediction       |
|-------------------------------------|------------------|
| I went to buy milk at the [MASK].   | [MASK] = store   |
| I graduated from [MASK]             | [MASK] = college |
| I had a burger and it tasted [MASK] | [MASK] = amazing |

# ChatGPT

- GPT = **Generative** Pre-trained Transformer
- ChatGPT Released in 2023 by OpenAI
- Newest version of ChatGPT (GPT-4o) has over 175 billion parameters
  - More than 96 layers
  - More than 96 attention heads per layer
  - At least 12,288 dimensional word embedding
- Trained on all text data in the world + human labeled data

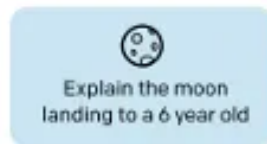


# Training ChatGPT: Fine Tuning

Step 1

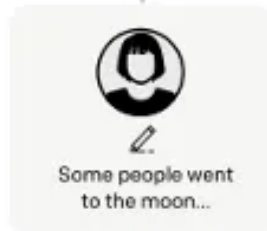
**Collect demonstration data,  
and train a supervised policy.**

A prompt is  
sampled from our  
prompt dataset.



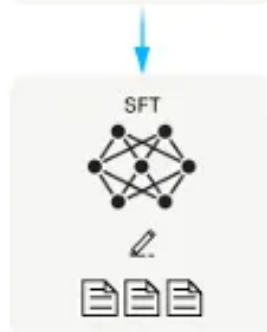
Prompt dataset is a series of  
prompts previously submitted to  
the Open API

A labeler  
demonstrates the  
desired output  
behavior.



40 contractors  
hired to write  
responses to  
prompts

This data is used  
to fine-tune GPT-3  
with supervised  
learning.



Input / output pairs are used to  
train a supervised model on  
appropriate responses to  
instructions.

# Training ChatGPT: Reward Model

Step 2

**Collect comparison data,  
and train a reward model.**

A prompt and  
several model  
outputs are  
sampled.

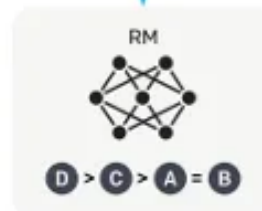


Responses are generated by  
the SFT model

A labeler ranks  
the outputs from  
best to worst.



This data is used  
to train our  
reward model.



$\binom{k}{2}$  combinations of  
rankings served to the  
model as a batch datapoint

# Training ChatGPT: RLHF

Step 3

**Optimize a policy against the reward model using reinforcement learning.**

Leverages Proximal Policy Optimization (PPO)

A new prompt is sampled from the dataset.

The policy generates an output.

The reward model calculates a reward for the output.

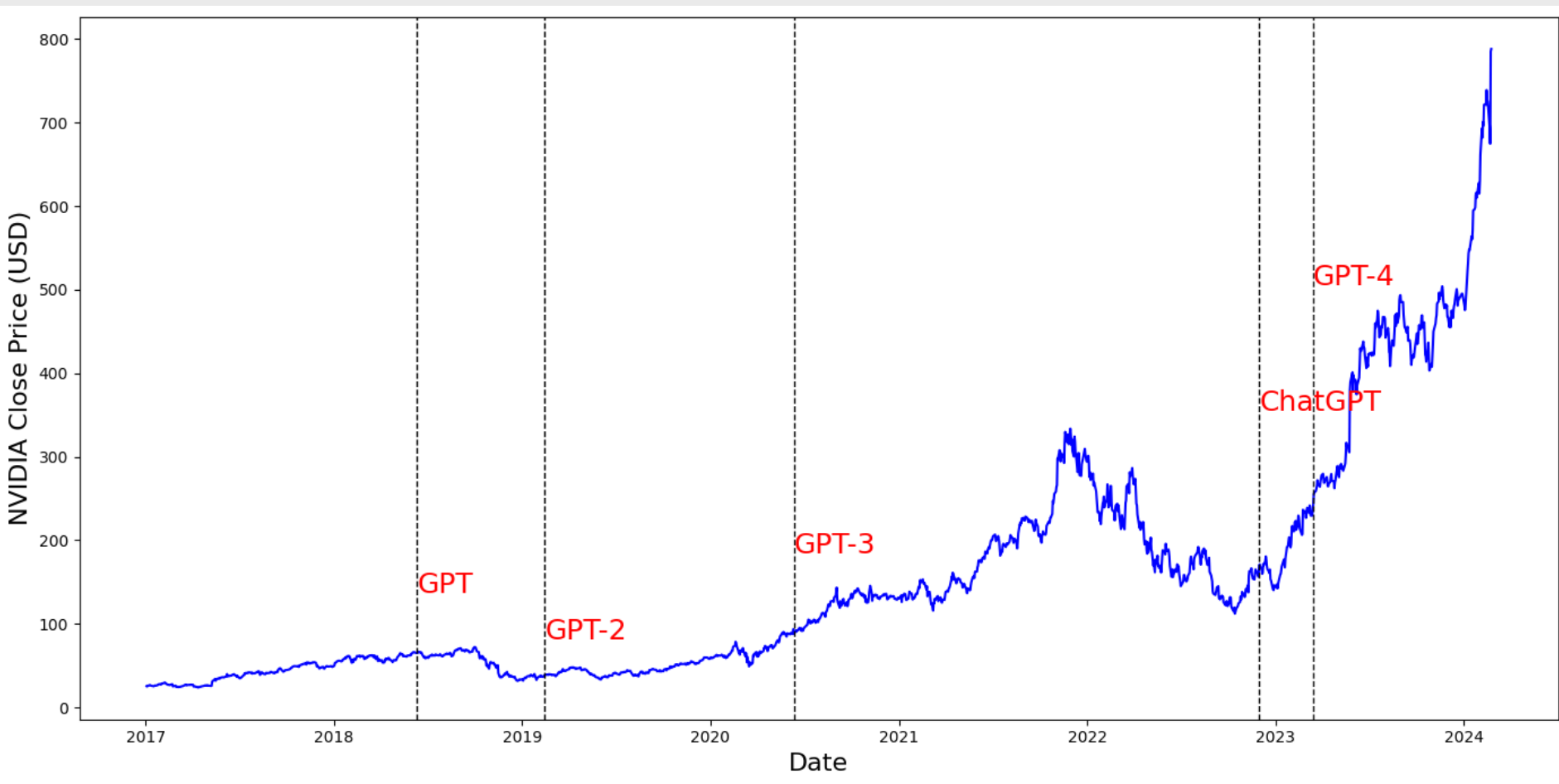
The reward is used to update the policy using PPO.



A policy is, a strategy that an agent uses in pursuit of goals

Kullback-Leibler penalty for SFT model to avoid overfitting

# GPTs Need GPUs



# Few-Shot Learning

- How do we make the language model generate text for a specific task?

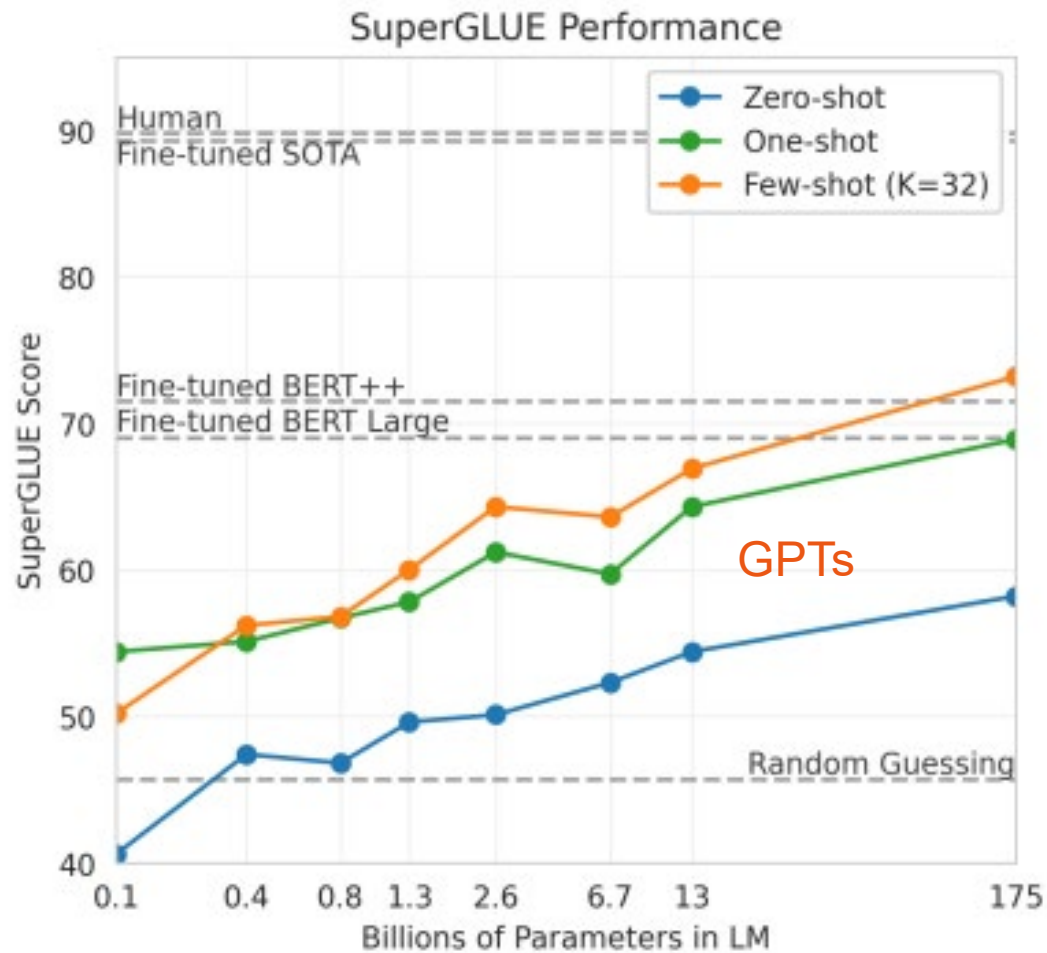
# Few-Shot Learning

- **How do we make the language model generate text for a specific task?**
- **Old way – fine tune on new set of data**
  - Requires a lot of compute power
  - Requires a lot of time

# Few-Shot Learning

- How do we make the language model generate text for a specific task?
- Old way – fine tune on new set of data
  - Requires a lot of compute power
  - Requires a lot of time
- New way – few-shot learning
  - Put a few example texts in the input
  - No training needed
  - Works surprisingly well if the model is large enough

# Few-Shot Learning





# Emergent Behavior

- **ChatGPT was trained to complete text**
  - Unsupervised training on huge dataset
  - Fine-tuning on smaller dataset
  - RLHF on even smaller dataset
- **ChatGPT exhibits emergent behavior – it can do things it was not trained to do**
  - Problem solving
  - Writing code
  - Creative content creation
  - Emulating writing styles
  - Many, many more

# **Social Media Text Generation**

- **Few-shot learning lets us do many useful social media tasks with small datasets**
  - **Measure sentiment**
  - **Learn community structure**
  - **Generate tweets with controlled style, topic, and sentiment**
  - **Generate persuasive content targeted for a person with a given set of interests to support a topic**

# Measuring Sentiment with ChatGPT

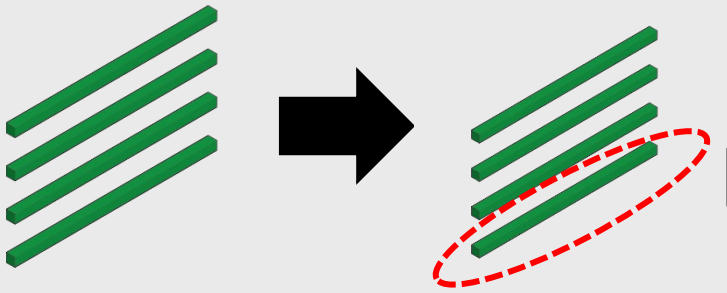
- ChatGPT can measure sentiment based on a prompt

# Measuring Sentiment with ChatGPT

- ChatGPT can measure sentiment based on a prompt
- Ex) “You will be given a sentence and must grade its sentiment from 0 to 100, 0 meaning very negative and 100 meaning very positive. Return only the numerical score: I hate my job

# Measuring Sentiment with ChatGPT

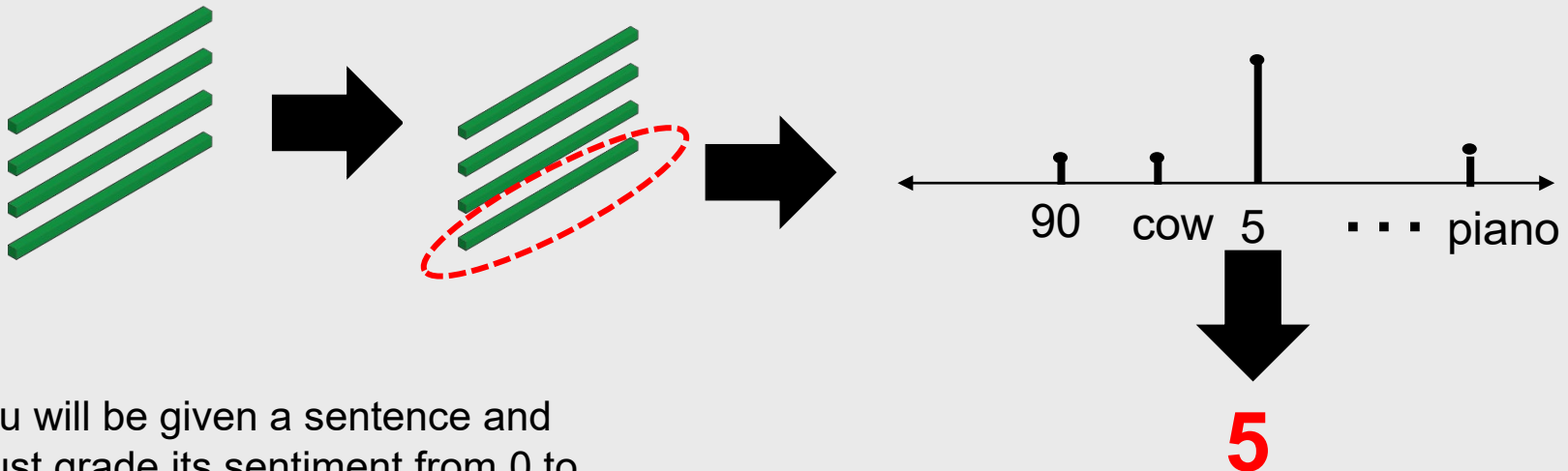
- ChatGPT can measure sentiment based on a prompt
- Ex) “You will be given a sentence and must grade its sentiment from 0 to 100, 0 meaning very negative and 100 meaning very positive. Return only the numerical score: I hate my job



You will be given a sentence and must grade its sentiment from 0 to 100, 0 meaning very negative and 100 meaning very positive. Return only the numerical score: I hate my job

# Measuring Sentiment with ChatGPT

- ChatGPT can measure sentiment based on a prompt
- Ex) “You will be given a sentence and must grade its sentiment from 0 to 100, 0 meaning very negative and 100 meaning very positive. Return only the numerical score: I hate my job



You will be given a sentence and must grade its sentiment from 0 to 100, 0 meaning very negative and 100 meaning very positive. Return only the numerical score: I hate my job

# Generating Tweets With Controlled Style, Topic, and Sentiment

You are going to help create content for a social media account. You will be asked to write clever tweets that will get engagement in the style of provided tweets. Learn from these tweets how to write your tweets. Here are the tweets:

Tweet: Starship will fundamentally change how we access space, with entire space stations like Starlab launched on a single mission

Tweet: One day, the chickens may turn on us ... <https://t.co/gP7epUz45y>

Tweet: Bill Maher is absolutely right

Tweet: A few other things you probably don't know: illegals in America can get bank loans, mortgages, insurance, driver's licenses, free healthcare (California & New York) and in-state college tuition. What's the point of being a citizen if an illegal gets all the benefits, but doesn't pay taxes or do jury duty?

Tweet: \$HOOD - Robinhood - SARAL Trade idea...!SARAL system is suggesting a swing long trade. Will create a trade using options. take 30 day free trial of SARAL Swing alerts.

Tweet: 😊

Tweet: It was always the plan

Tweet: Outrageous!

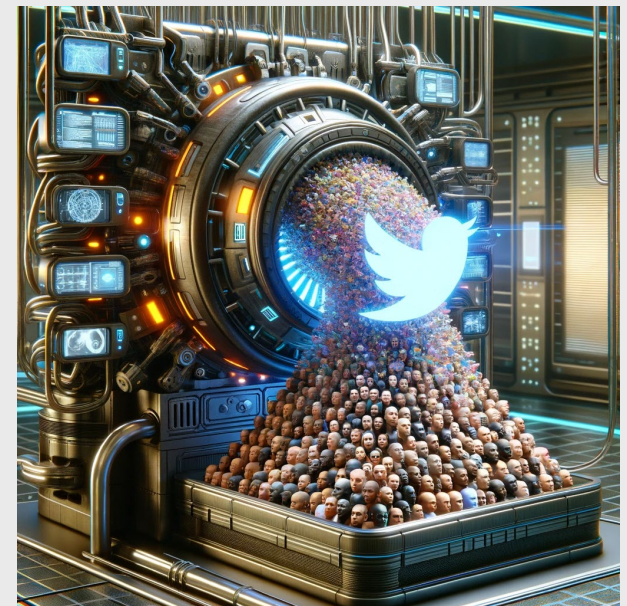
Tweet: #Earnings season continues this week, with results due from another 13% of the S&P 500. Consumer discretionary and healthcare lead. View our calendar for more. \$PYPL \$F \$CVS \$WYNN <https://t.co/CJoGLZDRTb>

Tweet: "Colon Cologne, the fragrance that takes you beyond Uranus!" – Grok.

**Write a tweet about healthy eating in the style of these tweets with a sentiment of 90/100, 100 being most positive, 0 being most negative.**

# Community Detection With ChatGPT

- User profile has a name, image, location, and bio
- We can feed all this raw data to ChatGPT and ask it to give us communities





# **Persuasive Text**

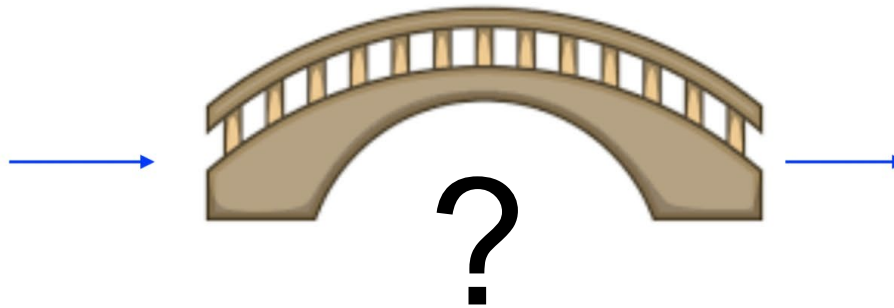
- **ChatGPT can generate persuasive text to shift a person's opinion on a topic**
- **It does this by bridging the person and the topic with a piece of text content**

# Persuasive Text

- ChatGPT can generate persuasive text to shift a person's opinion on a topic
- It does this by bridging the person and the topic with a piece of text content



Taxes are bad



Chocolate is good

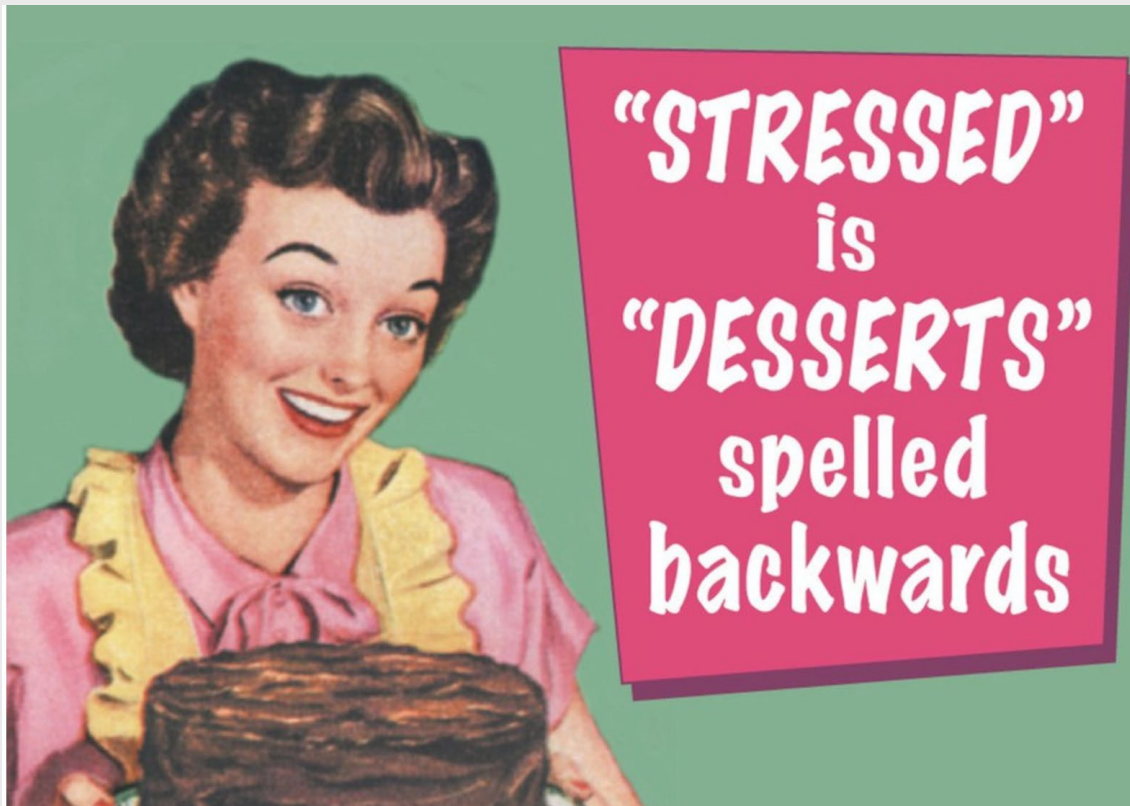
# Taxes and Chocolate

- Filing taxes is stressful



# Taxes and Chocolate

- Chocolate relieves stress

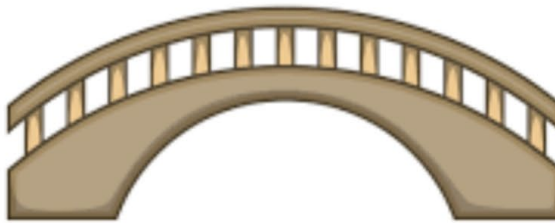


# Taxes and Chocolate

- Bridge: Taxes -> **Stress** -> Chocolate



Taxes are bad



**Stress**



Chocolate is good

# ChatGPT v4 Prompts

- **Prompt: write 3 tweets that convince someone who doesn't like paying taxes to eat chocolate**

# ChatGPT v4 Prompts

- Prompt: write 3 tweets that convince someone who doesn't like paying taxes to eat chocolate
- Response: 1. "Hey there! Did you know that eating chocolate can actually make you feel better about paying taxes? Studies show that chocolate releases endorphins, which can boost your mood and reduce **stress**. #ChocolateForTaxHaters"
- 2. "Taxes may be a bitter pill to swallow, but chocolate is sweet **relief**! Treat yourself to a delicious bar of chocolate and take the edge off tax day. You deserve it! #TaxDayTreat #ChocolateLover"
- 3. "Why **stress** about taxes when you can indulge in the sweet pleasure of chocolate? Trust us, it's a match made in heaven. Try it for yourself and thank us later! #ChocolatelsTheAnswer #TaxDayBliss"

# Persuasion Prompt Structure

- Given a person who is described by <A>
  - A sentence, tweets, Twitter bios, Instagram photos
- Convince them to support topic <B>
- Use content of type <C>
  - A tweet, song, picture, narration of TikTok video



# Persuasion Prompt Structure

- Given a person who is described by <A>
  - A sentence, tweets, Twitter bios, Instagram photos
- Convince them to support topic <B>
- Use content of type <C>
  - A tweet, song, picture, narration of TikTok video
- Prompt: “Here is the description and content posted by someone: \n<A>\n  
Convince them to support <B> with a <C>

# **Coding Session**

- **Creating social media text content with ChatGPT**
  - **Controlled tweet generation**
  - **Community detection**
  - **Persuasion**
- **You will need an OpenAI API key**
  - **One has been provided in an email**