

"Dreaming of Home and Mother", Song by John P. Ordway

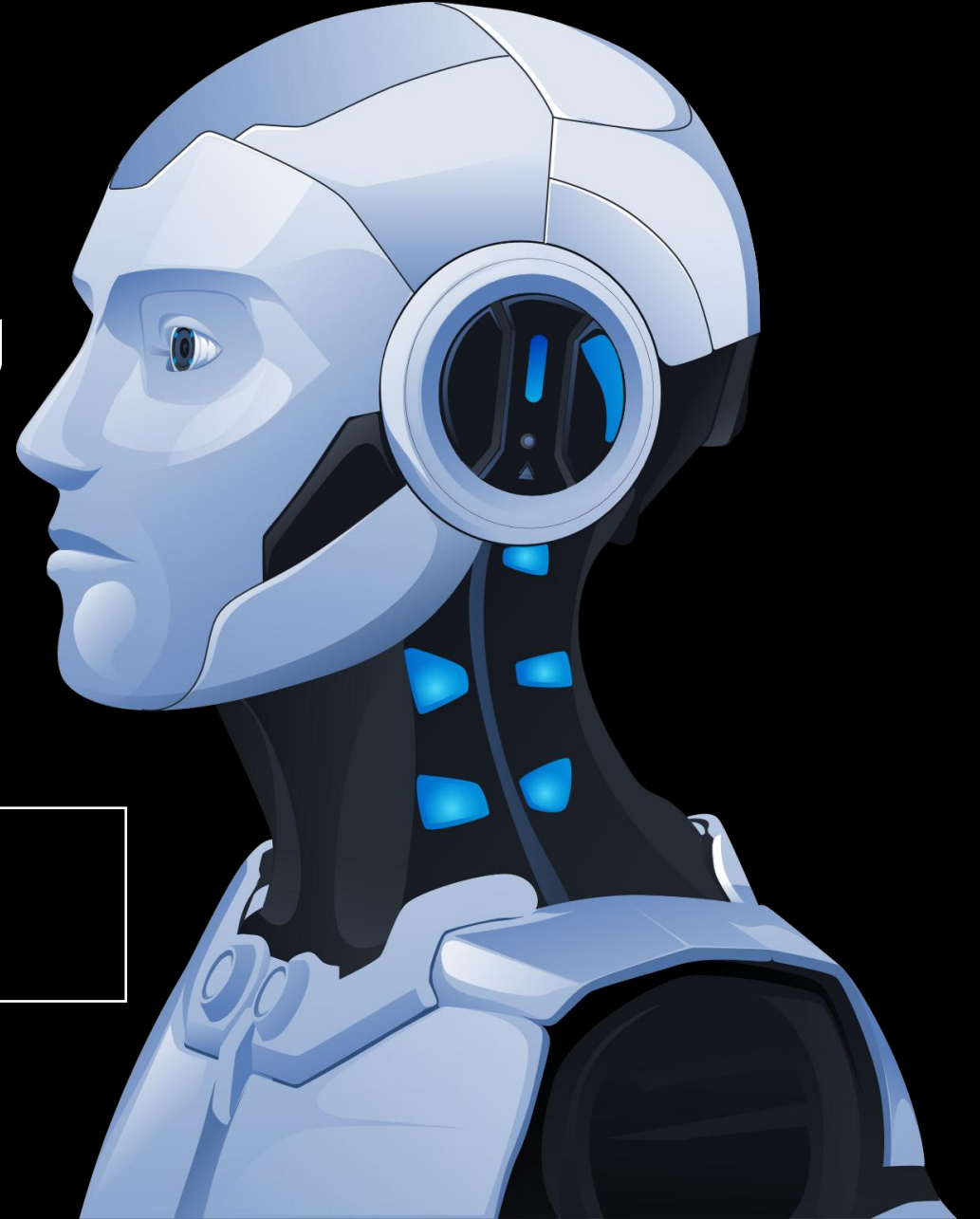


# Natural Language Processing CMPE297

1<sup>st</sup> Lecture, Sep. 18, 2025

Instructor: Michael Hu

Generated by Human  
Intelligence



# Self Introduction

---

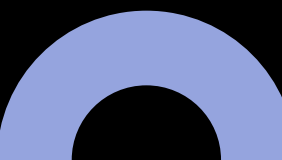
About Your Name, Academic  
Background, Course Motivations,  
Personal Interests, and Unique  
Attributes





# Class Contents

---

- **Understanding NLP:** Exploring its role and status in the industry
  - **Course Structure:** Understanding logistics and layout
  - **Word Representation Basics:** Exploring distributional semantics
  - **Tool Introduction:** Hands-on with GPT-5
- 

# Early Human Language Writing and Q&A System

- The Shang Dynasty in ancient China (c. 1600-1046 BCE) is famous for its use of **oracle** bones, which are inscribed with early Chinese characters.
- The oracle bone script primarily consists of divinatory inscriptions on turtle shells and animal bones.
- These inscriptions were used in divination practices where questions were posed to the spirits or deities, and the cracks that appeared on the heated bones were interpreted as answers.
- These **questions** often included **information about events, names of individuals, and other aspects of daily life** during the Shang Dynasty.



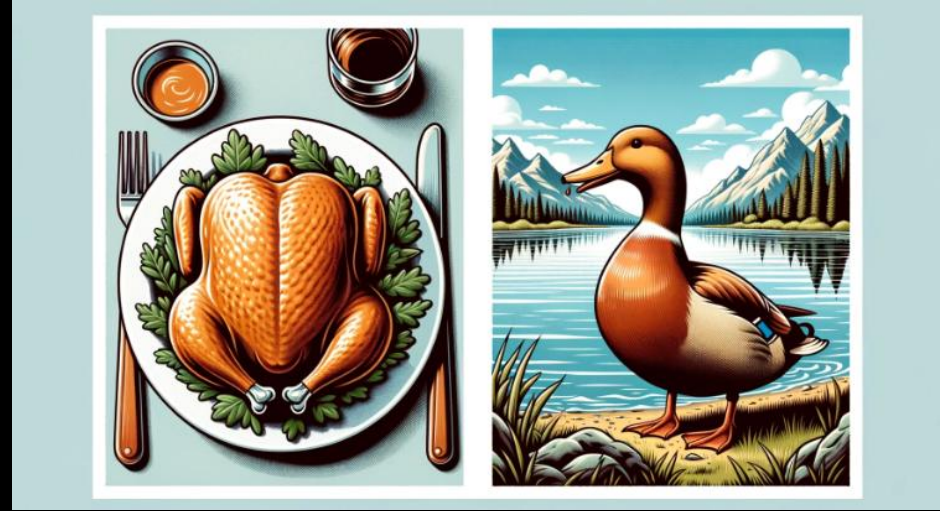
# Early Philosopher Talking about Language and Words

---

- 荃者所以在鱼，得鱼而忘荃
  - Nets are for fish; Once you get the fish, you can forget the net.
  - 言者所以在意，得意而忘言
  - Words are for meaning; Once you get the meaning, you can forget the words
- 
- 庄子(Zhuangzi, 4th century BCE ), Chapter 26

# Classic Examples for Language Ambiguity

- The duck is ready to eat.



- I saw the man with a telescope.



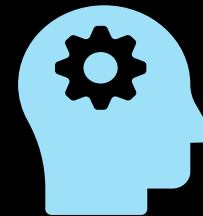
# Natural Language Processing (NLP)

---

Interaction between **computers and humans** via natural language with the aim to enable computers to **understand, interpret, and generate language**.

## ■ Applications:

- ✓ Machine Translation (e.g., Google Translate)
- ✓ Sentiment Analysis
- ✓ Human-machine communication (e.g. Question Answering, Dialog)
- ✓ Speech Recognition (e.g., Siri, Alexa)
- ✓ Text Summarization
- ✓ .....



# The Advancement of AI in the Past Decade

## ■ Foundation: Data explosion and computing advancement

- ✓ Enormous amount of available data
- ✓ Cloud computing and GPU (e.g. Nvidia)

## ■ Deep learning revolution

- ✓ Deep learning, a subset of AI, has revolutionized the field
- ✓ Significant advancements in areas such as computer vision, natural language processing, and speech recognition.
- ✓ AI Leaders: : **Geoffrey Hinton, Yoshua Bengio and Yann LeCun** (2018 Turing Award winners). **Geoffrey Hinton** won the 2024 Nobel Prize in Physics, alongside John J. Hopfield, for pioneering foundational work in machine learning via artificial neural networks.

## ■ Impact to different industries

- ✓ AI technologies are automating and optimizing various tasks and processes across industries, ranging from manufacturing, healthcare, logistics, banking etc.

## ■ Personalized experiences and pervasive presence.

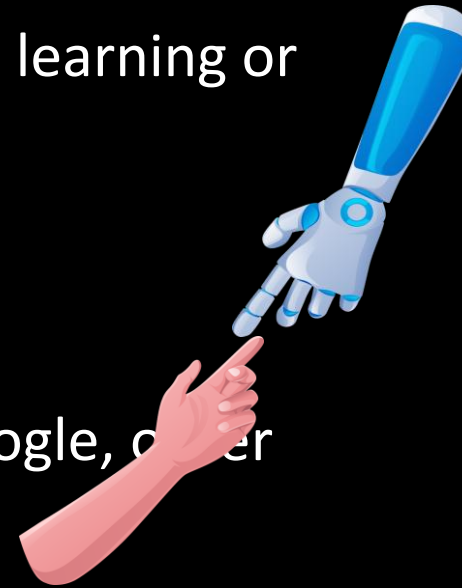
- ✓ Applications found in smartphones, virtual assistants, smarter homes, autonomous vehicles.
- ✓ AI has become an integral part of our daily lives, especially with AI Agents.



# Generative AI: A “Once-in-a-generation Shift”

## ■ Large Language Model (LLM)

- ✓ A large language model (LLM) is a language model consisting of a neural network with many parameters (tens of millions, billions (175B for GPT-3), trillions (x T for GPT-4).
- ✓ It is trained on large quantities of unlabeled text using self-supervised learning or semi-supervised learning.
- ✓ LLM is closely linked to Generative AI.
- ✓ Helps developers for **faster application development** – “Vibe coding”.
- ✓ **Multimodality.**
- ✓ Examples: ChatGPT from OpenAI, LLaMA from Meta, Gemini from Google, other open models include DeepSeek, Qwen etc.



■ **Generative AI seems to be on the lips of every venture capitalist, entrepreneur.**

# Diverse Perspectives on LLMs

- LLM **could understand natural languages**
  - ✓ A revolution
  - ✓ “Pause giant AI experiments” (Elon Musk, Steve Wozniak, **Yoshua Bengio**, Andrew Yang)
  - ✓ Needs regulation
- LLM ChatGPT is **‘not particularly innovate,’ and ‘nothing revolutionary’**
  - ✓ Bad idea to pause the research
  - ✓ Joint Embedding Predictive Architecture (JEPA)



Yoshua Bengio, Geoffrey Hinton and Yann LeCun  
Source: Yann Lecun's linkedin site

# The Paper behind Today's LLMs

## 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\* †  
University of Toronto  
aidan@cs.toronto.edu

Lukasz Kaiser\*  
Google Brain  
lukaszkaier@google.com

Illia Polosukhin\* †  
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 mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

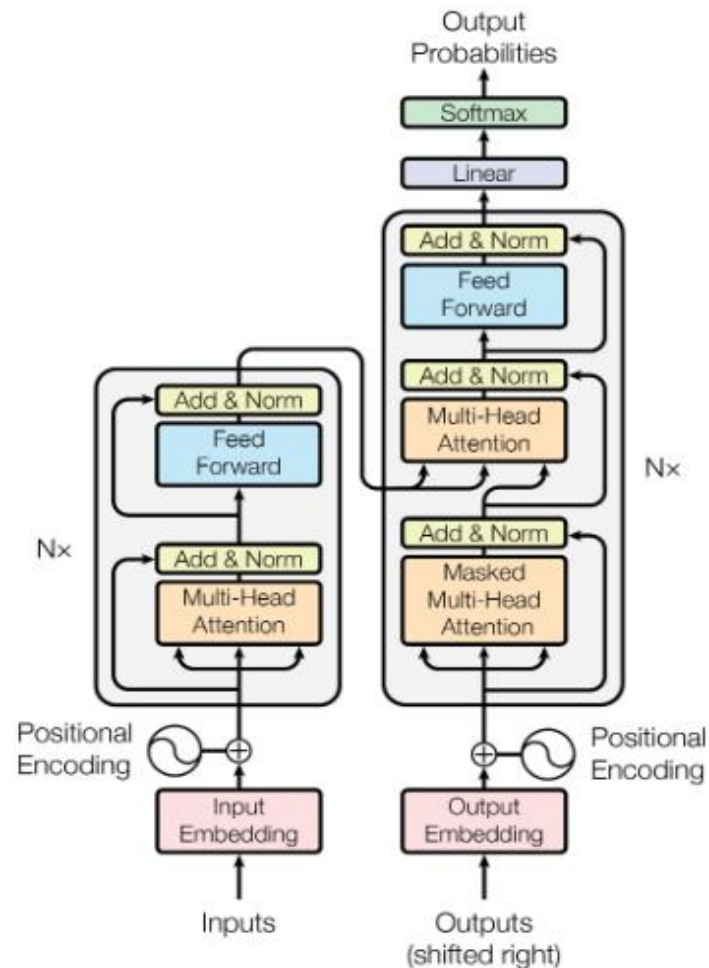
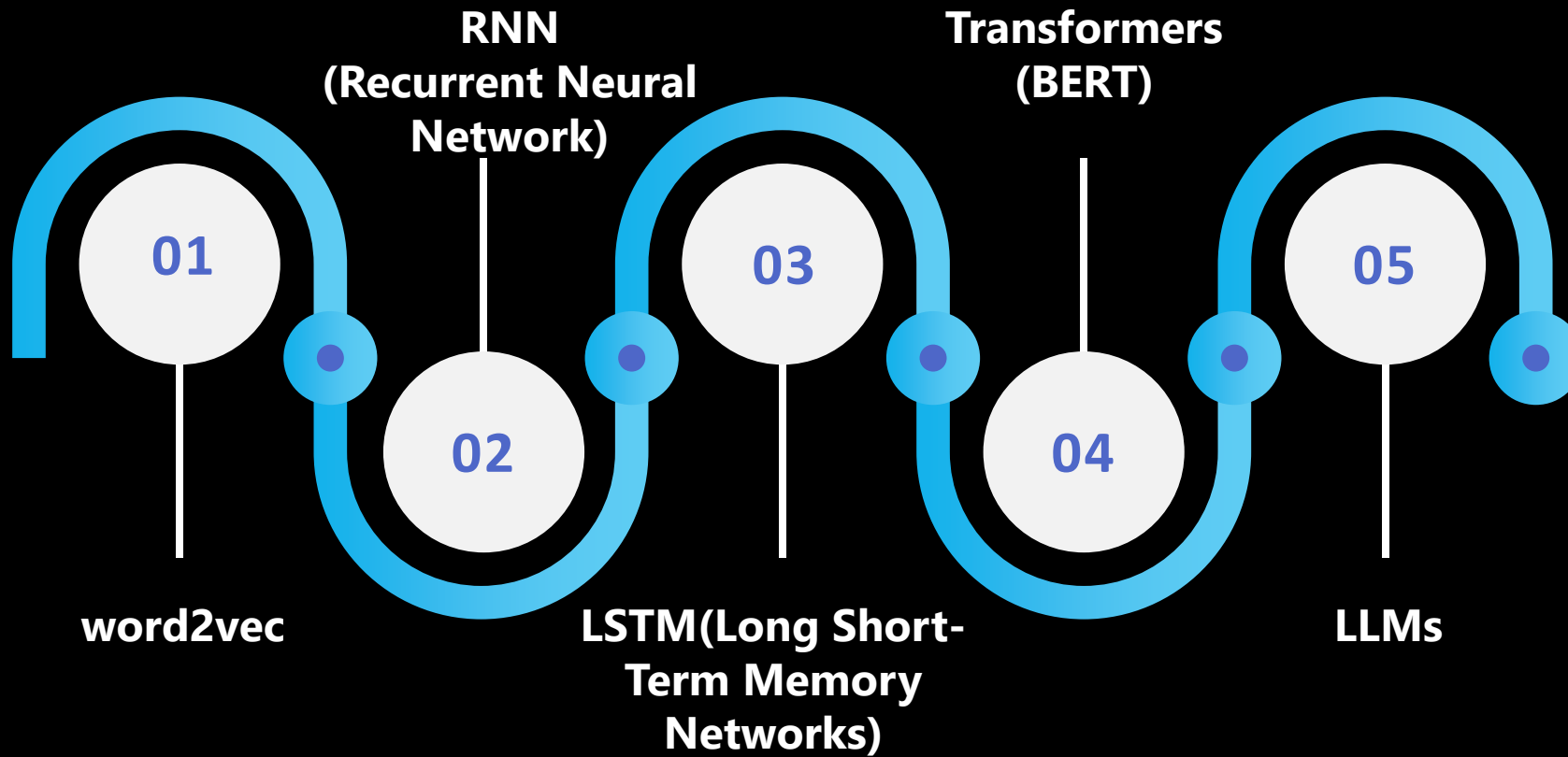


Figure 1: The Transformer - model architecture.

# The Evolution



# Understanding How LLMs Work



Once upon a time, in a land far far away

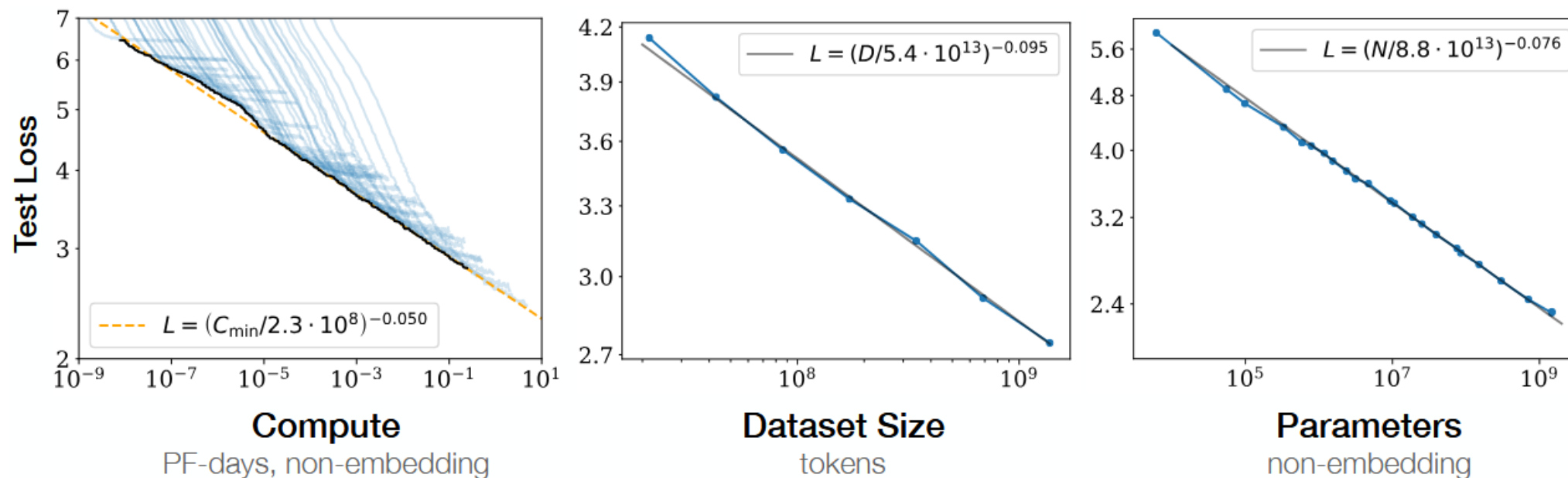


Generated training examples, using  
attention & positional encoding

Examples			Inputs				Correct Output (Label)		
Once	Upon	a	time	in	a				land
Once	Upon	a	time	In	a	land			far
One	Upon	a	time	in	a	land	far		far

Reinforcement Learning from Human Feedback (RLHF)

# Scaling Laws for Neural Language Models



Source: Kaplan et al. (OpenAI, 2020) Scaling Laws for Neural Language Models

# The GPT Use-cases

Table 1: Distribution of use case categories from our API prompt dataset.

Use-case	(%)
Generation	45.6%
Open QA	12.4%
Brainstorming	11.2%
Chat	8.4%
Rewrite	6.6%
Summarization	4.2%
Classification	3.5%
Other	3.5%
Closed QA	2.6%
Extract	1.9%

Table 2: Illustrative prompts from our API prompt dataset. These are fictional examples inspired by real usage—see more examples in Appendix A.2.1.

Use-case	Prompt
Brainstorming	List five ideas for how to regain enthusiasm for my career
Generation	Write a short story where a bear goes to the beach, makes friends with a seal, and then returns home.
Rewrite	This is the summary of a Broadway play: "" { summary } "" This is the outline of the commercial for that play: ""

Source: Training language models to follow instructions with human feedback, OpenAI, 2022



# The Future is Coming

AI is affecting people's lives deeply.  
We may find a way to reach the moon

—  
GENERATED BY DALL-E



It was the best of times, it was  
the worst of times, it was the  
age of wisdom, it was the age  
of foolishness.

—  
Charles Dickens,  
"A Tale of Two  
Cities"

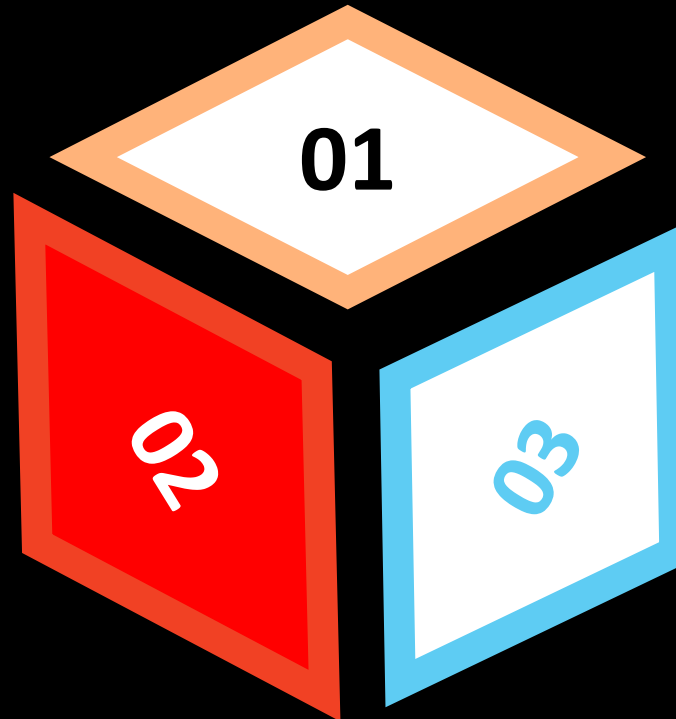


# The Course Structure

---

## Foundations knowledge to NLP:

Tokenization, Word embedding, FF Network, recurrent network, attention, transformers, pre-trained language model...



## Practical Application Tools/Best Practices:

LLM Tokenizer, GPT, LLM APIs, BERT, Chain of Thoughts, function calling, semantic database...

## Projects with Milestones:

ChatBot, function calling, MCP, linking with Enterprise assets, QA systems...



# The Basic Software Environment

---

- Basis: Python, Web, AI
  - Development
    - ✓ Jupyter Notebook
    - ✓ Ubuntu, Windows, MacOS
    - ✓ Visual Studio Code
    - ✓ Web Framework
    - ✓ JavaScript
    - ✓ TensorFlow/PyTorch
    - ✓ GPT Models, BERT
  - Deployment
    - ✓ Cloud
    - ✓ Semantic Database
    - ✓ GPT Model, BERT/Llama...
- 



# Reference Textbook

---

Speech and Language Processing (3rd ed. draft)

Dan Jurafsky and James H. Martin

## Speech and Language Processing

An Introduction to Natural Language Processing,  
Computational Linguistics, and Speech Recognition  
with Language Models

Third Edition draft

Daniel Jurafsky  
*Stanford University*

James H. Martin  
*University of Colorado at Boulder*

*Copyright ©2025. All rights reserved.*

Draft of August 24, 2025. Comments and typos welcome!

# Evaluation Criteria & Grading

Percent	Grade	Percent	Grade
96%-100%	A+	76%-79%	C+
93%-95%	A	73%-75%	C
90%-92%	A-	70%-72%	C-
86%-89%	B+	66%-69%	D+
83%-85%	B	63%-65%	D
80%-82%	B-	60%-62%	D-
		Below 60%	F

**HW (Individual): 45%**

**Interim Project Review (Team) : 30%**

6 Teams, weekly progress review from week 4

**Final Project Review(Team and Individual): 25%**

Total: 100%

Notes:

**Project Attendance:** Your attendance is tracked and contributes to your individual grade, recognizing your active participation.

**Grace Periods:** You may request up to **2** homework submission grace periods, each lasting up to **4** days, to accommodate unexpected situations.

**Approval:** Grace period requests require approval from the Instructional Student Assistant (ISA) or the instructor — reasonable requests are typically approved.

# Other Course Logistics

---

- Michael Hu **email:** [xiaozhuan.michael@sjsu.edu](mailto:xiaozhuan.michael@sjsu.edu)
- **Office Hours:**
  - ✓ Primary Slot: Tuesdays, 7pm-8pm (by appointment via email)
  - ✓ Alternative Arrangements: If this time is inconvenient, kindly email to propose alternative appointment times for consideration.
- Please complete your homework on time
- Instructional Student Assistant (ISA) :
  - Charmi Doshi **email:** [charmi.doshi@sjsu.edu](mailto:charmi.doshi@sjsu.edu)

# Human Languages

## ■ Language :

Is essential for human beings because it facilitates communication, expression, and the organization of complex thoughts.

## ■ Written language :

Is important because it allows for the preservation, dissemination, and structured representation of knowledge across time and space.

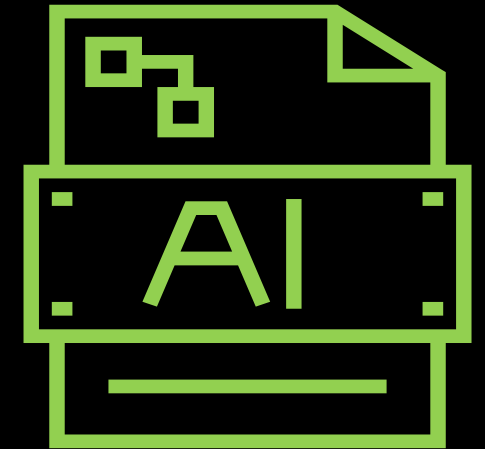
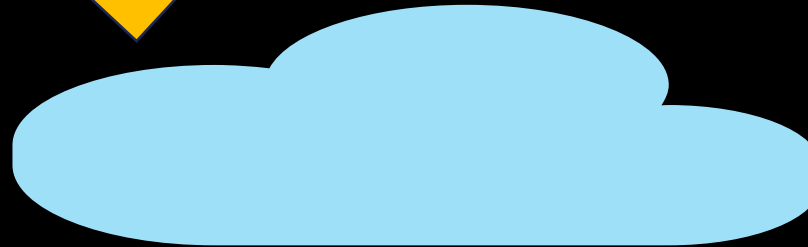
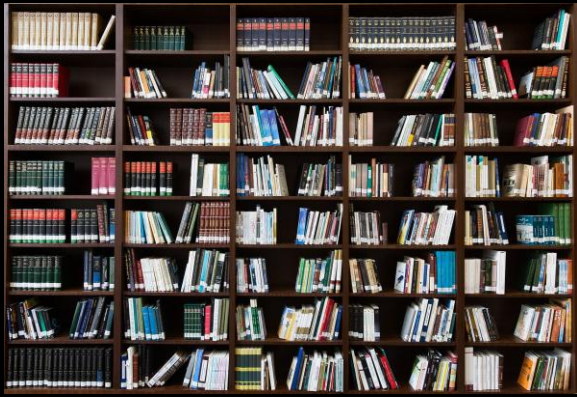
## ■ Human knowledge:

Is encapsulated in language, facilitating the sharing and understanding of collective experiences and insights.



# Knowledge of the World

"Knowledge is of no value unless you put it into practice." - Anton Chekhov



# Evolution of Natural Language Processing (NLP)

- **1950s-1960s: Rule-Based Systems & Turing Test.**

Example: ELIZA – An early computer program that simulated a psychotherapist.

- **1970s: Symbolic & Transformational Grammar.**

Example: SHRDLU – A system for natural language understanding, operating in a blocks world.

- **1980s-1990s: Rise of Statistical Models & Probabilistic Grammars.**

Example: IBM's statistical translation models for machine translation.

- **Late 1990s-2000s: Machine Learning & Pattern Recognition.**

Example: SPAM filters to distinguish spam emails from legitimate ones.

- **2010s: Deep Learning, Word Embeddings, and Transformers.**

Example: Google's Word2Vec embeddings, and IBM Watson winning in the Jeopardy! competition in 2011.

- **2020s: Efficient Models, LLM and Contextual Understanding.**

Example: Efficient versions of the BERT model for deployment in real-world applications, the emerging of LLM such as OpenAI GPT.



Source: IBM Research





# Word Meaning

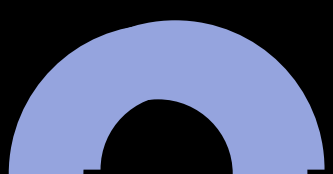
---

What is the meaning of word **Zyzzzyva**?

## Clues:

**Sentence 1:** The zyzzzyva is known for its unique ability to thrive in tropical environments, particularly among decaying plant matter.

**Sentence 2:** Many entomologists study the zyzzzyva to understand more about the diversity of beetles in rainforest ecosystems.

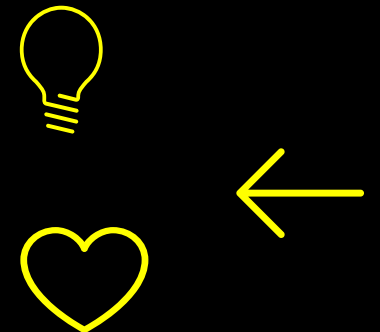


# The Meaning of Words

"For a large class of cases—though not for all—in which we employ the word 'meaning' it can be defined thus: the meaning of a word is its use in the language." - *Philosophical Investigations*, Ludwig Wittgenstein

- Denotation: The literal, dictionary definition of a word.
  - Connotation: The emotional and cultural associations tied to a word. (Positive/Negative, e.g. "slim")
  - Context: How the meaning of a word can change depending on its use in a sentence or situation.
  - Polysemy: Words that have multiple related meanings. ("star")
  - Homonymy: Words that sound alike but have different meanings. ("bat")
- A single sentence showing a word with different meanings based on context:  
"He faced the **bank** of cameras with a smile, but his mind was on the river **bank** where he spent his childhood."

Words have no meaning  
without context!





# Understanding Word Representation through Distributional Semantics

---

## ■ The Foundation of Meaning

- ✓ "You shall know a word by the company it keeps." - J.R. Firth, 1957
- ✓ The essence of a word's meaning is captured by the words that surround it.

## ■ Contextual Clues

- ✓ A word's context is defined by its neighbors within the text.
- ✓ Words gain significance from the patterns of words that frequently occur around them.

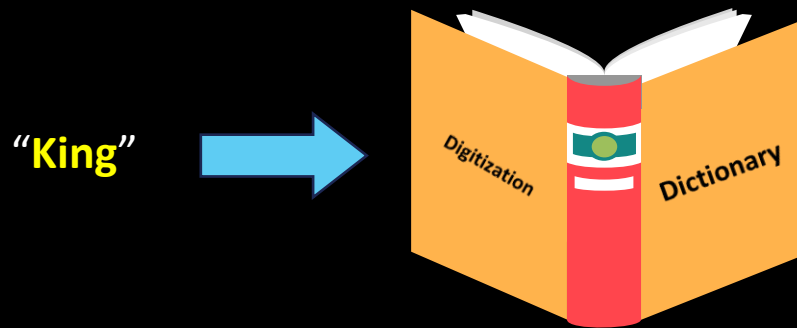
## ■ Building Representations

- ✓ Analyze multiple contexts: Gathering a word's "company" from large text corpora.
- ✓ Develop a semantic profile: Using statistical/neural network methods to create a multi-dimensional space.

## ■ Vector Space Model

- ✓ Each word is a point in space.
  - ✓ The proximity of words in this space reflects the similarity of their meanings.
- 

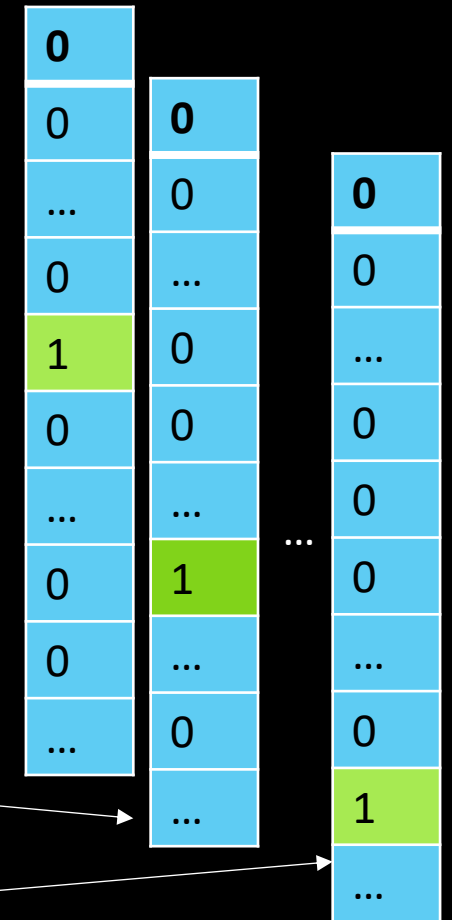
# Word Representation



#4914  
 $O_{4914}$

Words	Word ID
King	4914
Man	5391
Queen	7157
Woman	9853

motel = [0 0 0 0...0 0 0 0 0 0 0 0 0 0... 0 **1** 0...0 0 0 0 0]  
hotel = [0 0 0 0...0 0 0 0 **1** 0 0 0 0 0 ...0 0 0...0 0 0 0 0]

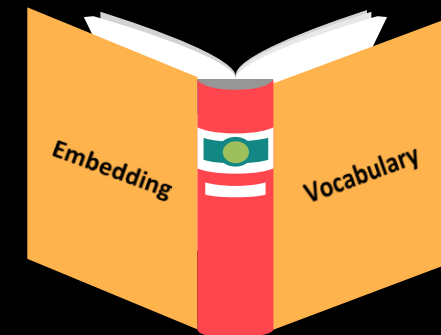


## One-hot vector

"orthogonal": two vectors are at right angles to each other, are independent or unrelated to each other.

# Feature Representation: Word Embedding

	Man(5391)	Woman(9853)	King(4914)	Queen(7157)	Apple(456)	Orange(6257)
Gender	-1.00	1.00	-0.95	0.97	0.00	0.01
Royal	0.01	0.02	0.93	0.97	-0.01	0.00
Age	0.03	0.02	0.70	0.69	0.03	-0.02
Food	0.09	0.01	0.02	0.01	0.95	0.97
...(up to 1536)	.....					
Size			.....			
Cost		.....			.....	
Alive				.....		.....
Verb						



**Notes:** The features are not the results of machine learning; they are for illustration purposes only.

$$\begin{pmatrix} -0.95 \\ 0.93 \\ 0.70 \\ 0.02 \\ \dots \end{pmatrix} - \begin{pmatrix} -1.00 \\ 0.01 \\ 0.03 \\ 0.09 \\ \dots \end{pmatrix} = \begin{pmatrix} 0.05 \\ 0.92 \\ 0.67 \\ -0.07 \\ \dots \end{pmatrix} + \begin{pmatrix} 1.00 \\ 0.02 \\ 0.02 \\ 0.01 \\ \dots \end{pmatrix} = \begin{pmatrix} 1.05 \\ 0.94 \\ 0.69 \\ -0.06 \\ \dots \end{pmatrix}$$

**King - Man = Queen - Woman ?**

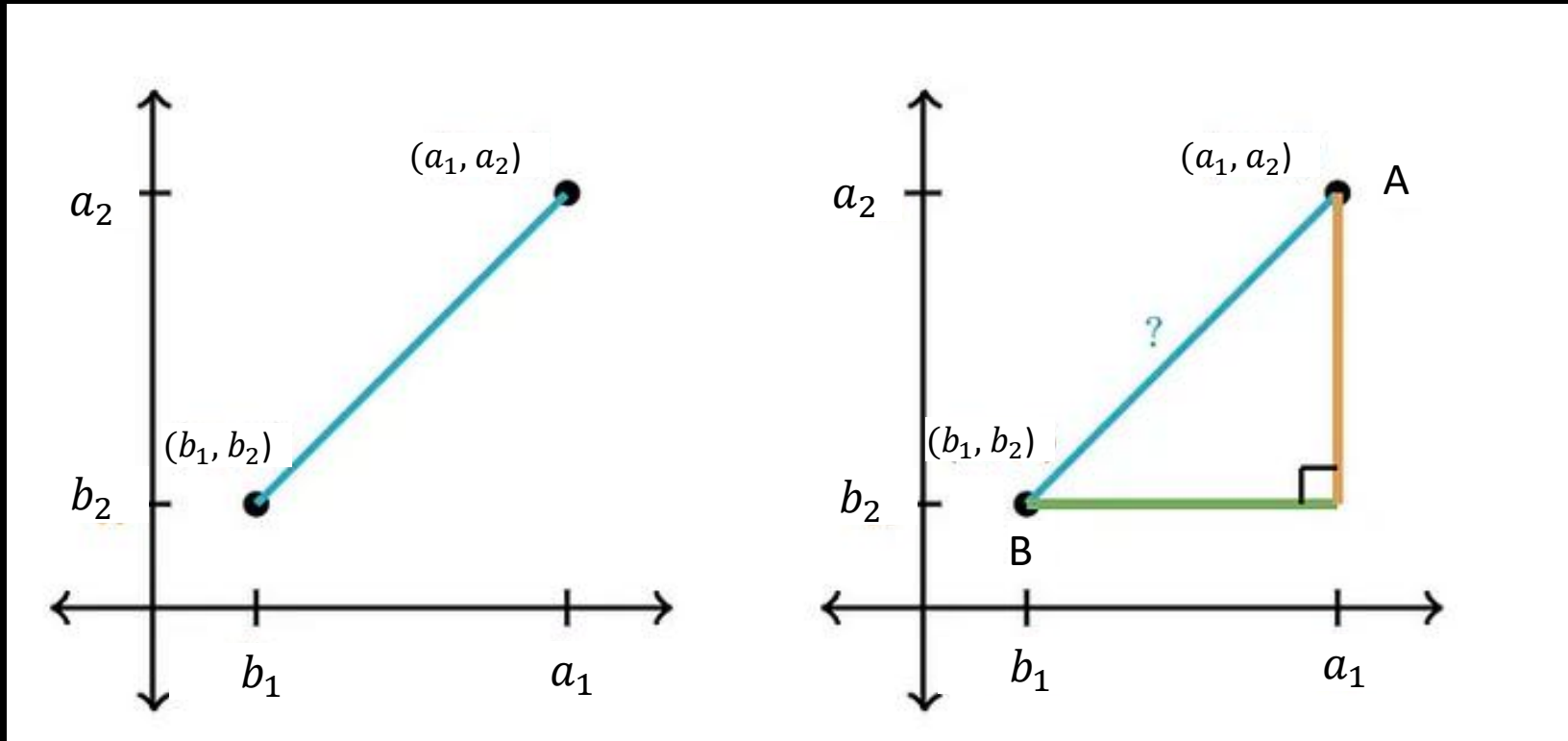
**King - Man + Woman = Queen ?**

biggest - big + small = smallest?  
Paris - France + Poland = Warsaw?

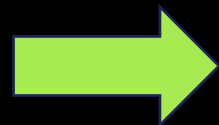


**Let's take a moment to review some math**

# Euclidean Distance Function



$$D(a, b) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$$



$$D(a, b) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2}$$

# Dot Product

The dot product between two vectors is a **scalar**:

$$\vec{v} \cdot \vec{w} = \sum_{i=1}^n v_i w_i = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n$$

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = 1 * 4 + 2 * 5 + 3 * 6 = 32$$

$v \cdot w$

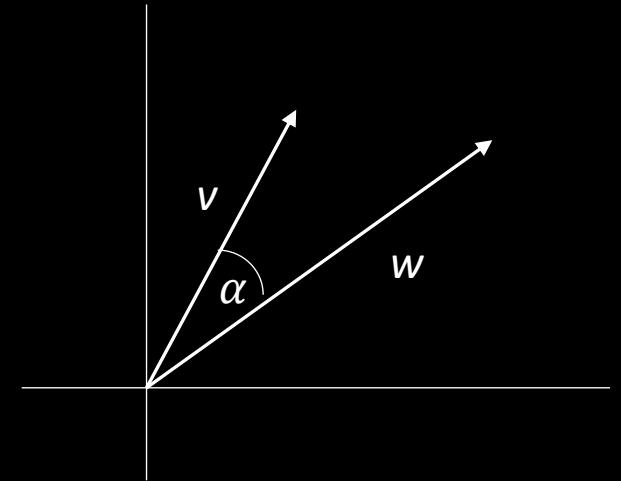
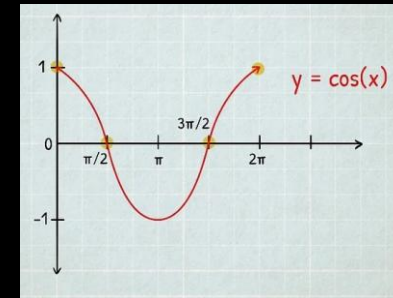
$$\vec{v} \cdot \vec{w} = \|\vec{v}\| \|\vec{w}\| \cos(\theta)$$



# Cosine Similarity

Cosine similarity is a measure of the angle between two vectors. It is computed by taking the dot product of the vectors and dividing it by the product of their magnitudes.

$$\text{cosine}(\vec{v}, \vec{w}) = \frac{\vec{v} \cdot \vec{w}}{|\vec{v}| |\vec{w}|} = \frac{\sum_{i=1}^N v_i w_i}{\sqrt{\sum_{i=1}^N v_i^2} \sqrt{\sum_{i=1}^N w_i^2}}$$



where  $\vec{v}$  and  $\vec{w}$  are the vectors being compared,  
“•” stands for the dot product

**Cosine Similarity measurement in two dimensions:**

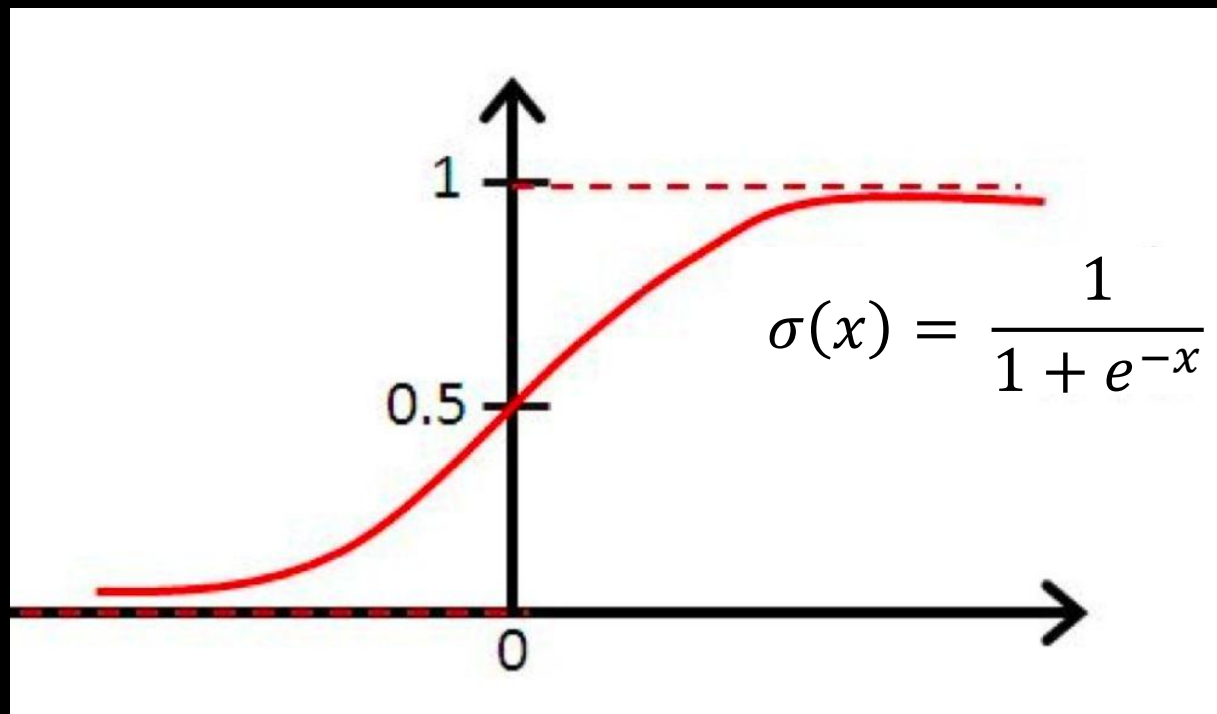
- +1: Vectors point in same directions
- 1: Vectors point in opposite directions
- 0: Vectors are orthogonal

Notes: since raw frequency values are non-negative, the cosine for these vectors ranges from 0–1.

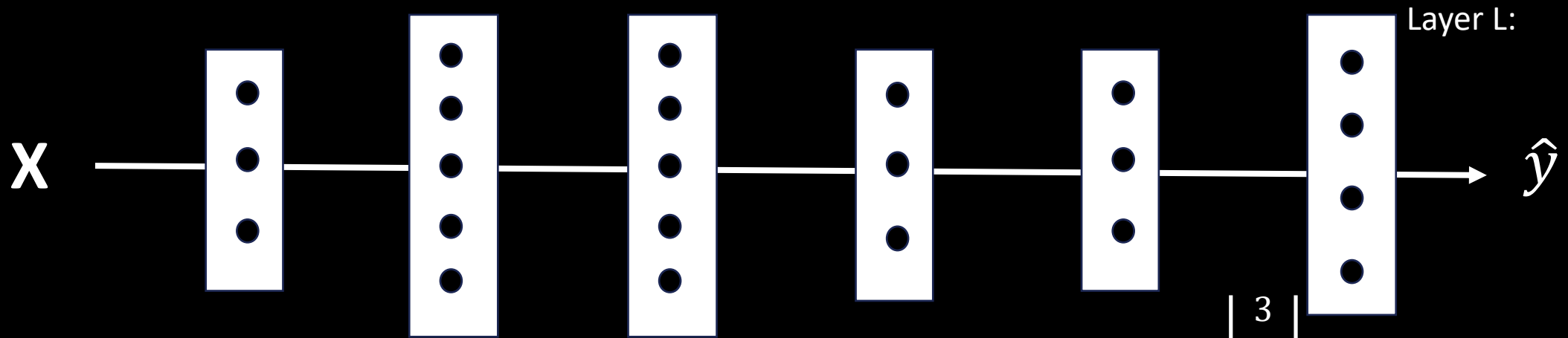
# Sigmoid Function

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

$$\sigma(-x) = 1 - \sigma(x)$$



# Softmax Activation



An example:

Element wise e to  $Z[L]$  with dimension (4,1):

$$a^{[L]} = \frac{e^{(z^{[L]})}}{\sum_{j=1}^4 t_j}$$

$$z^{[L]}$$

$$\begin{bmatrix} 3 \\ 4 \\ -1 \\ 1 \end{bmatrix}$$

$$t = e^{(z^{[L]})} \begin{bmatrix} 20.1 \\ 54.6 \\ 0.4 \\ 2.7 \end{bmatrix}$$

$$\sum_{j=1}^4 t_j = 77.8$$

$$a^{[L]} = \frac{e^{(z^{[L]})}}{\sum_{j=1}^4 t_j} \begin{bmatrix} 0.26 \\ 0.70 \\ 0.01 \\ 0.03 \end{bmatrix}$$

# Word2Vec: A Transformative Approach to Word Embeddings

---

## ■ Introduction to Word2Vec

- ✓ A neural network-based technique to learn word associations from a large corpus (“body”) of text.
- ✓ Developed by a team led by Tomas Mikolov at Google (2013).

## ■ Key Concepts

- ✓ **Word Embeddings:** Converts words into numerical form, allowing for the capture of semantic relationships.
- ✓ Continuous Bag-of-Words (CBOW): Predicts a target word from a set of context words.
- ✓ Skip-Gram: Predicts surrounding context words from a target word.



# Using GPT as the Learning Tool

---



# GPT Models

Language Models	Description	Context Window Size	Remarks
GPT-5 GPT-5-mini	High Intelligence Model  Affordable small model for focused task	400K context	Knowledge cutoff date: Sept 30, 2024 (GPT-5)
O3-pro	More compute for better responses, think before thy answer, perform reasoning. Affordable small model for reasoning.	200K context	Knowledge cutoff date: Jun 01, 2024
GPT Image 1	State-of-the-art image generation model		Support both text and image inputs
GPT-4o mini TTS	Use it to convert text to natural sounding spoken audio.		
GPT-4o mini Transcribe	Transcribe is a speech-to-text model, use it for more accurate transcripts.	16K context	For speech recognition
Embeddings (text-embedding-3-small)	Numerical representation of text.	8K context	Output dimension 1,536 for small



# To Access GPT

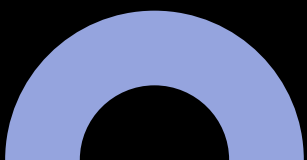
---

- ChatGPT: <https://chatgpt.com/>
- The OpenAI status: <https://status.openai.com/> (99.69% Uptime)
- Using API through API key



# APIs

---

- Chat Completions API
  - Responses API (Superset of Chat Completions API)
  - Using API through API calls
- 

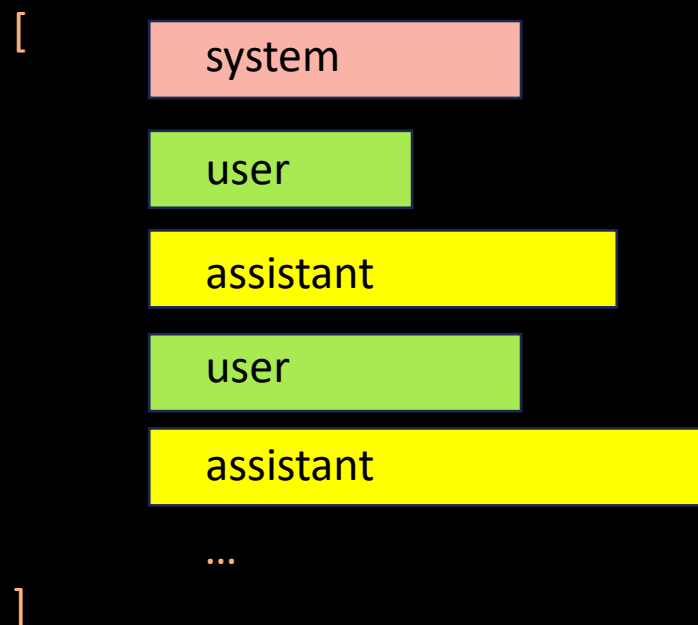


# Chat and Responses Completions

- Model, e.g. “gpt-5-mini”
  - ✓ Input: A list (array) of **messages** or **input**
  - ✓ Output: Generated **message** or **response**

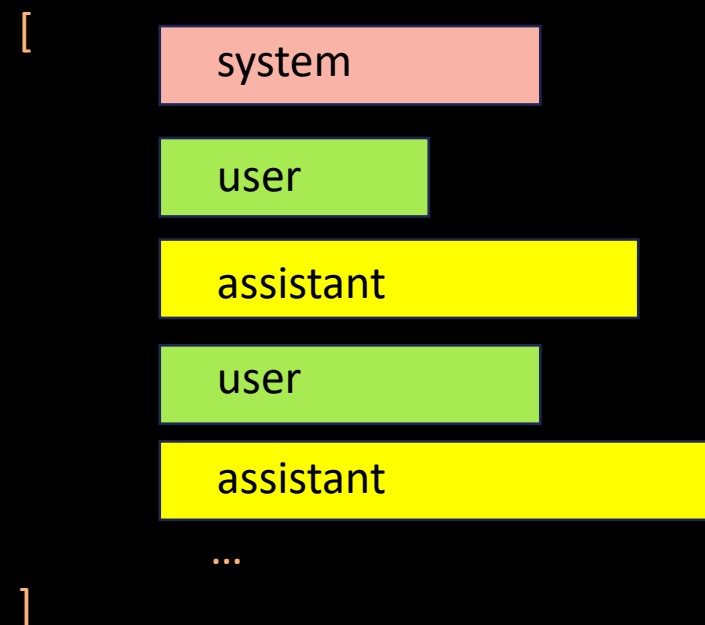
## Chat Completions API

messages =



## Responses API

input =



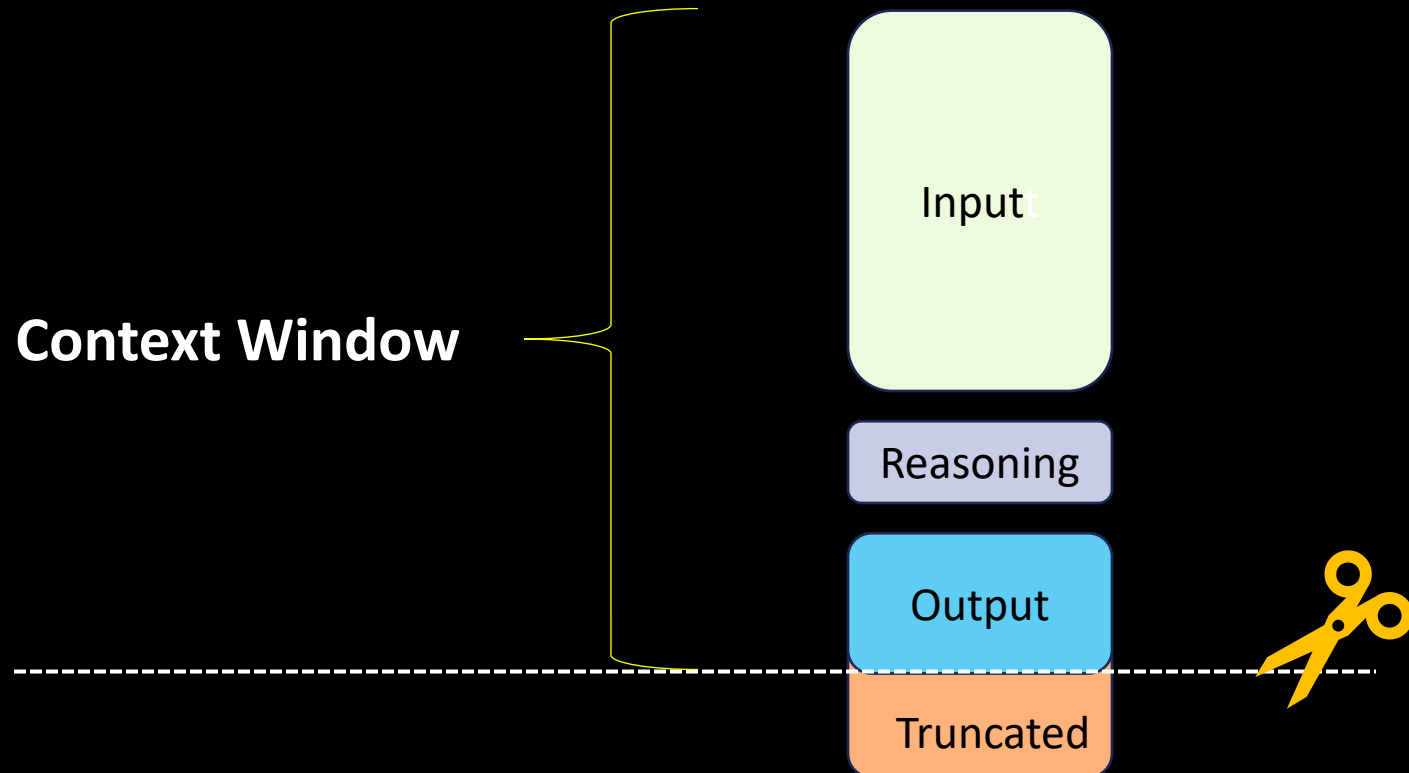
# Context Window

## ■ Context Window

- Represents the total tokens for input, output, and sometimes reasoning tokens..

## ■ Output Tokens

- Tokens generated by a model in response to a prompt. Example: GPT-4o-2024-08-06 can generate up to 16,384 output tokens.





# HW1

---

- Class Demo Notebooks code: The class demo notebook files can be found under the “Course Materials” section under each module in the Canvas.
- Create an account in OpenAI if you have not yet done that.
- Please set up prepaid billing for API usage. Please note this is **NOT** monthly subscription but **prepaid billing**. To the best of knowledge from our past experience, \$5 will be good enough for course purpose.
- Recommended Technical Papers: While it is optional, you are encouraged to browse through these papers.