# Artificial Intelligence

## CSC 665

tyler dae devlin

# Course Overview

*2.1.2024*

# Demo
# Eliza vs. ChatGPT

**ELIZA**

- Developed in the mid 1960s
- Powered by pattern matching and substitution rules
- A beginning programmer can implement a toy version from scratch

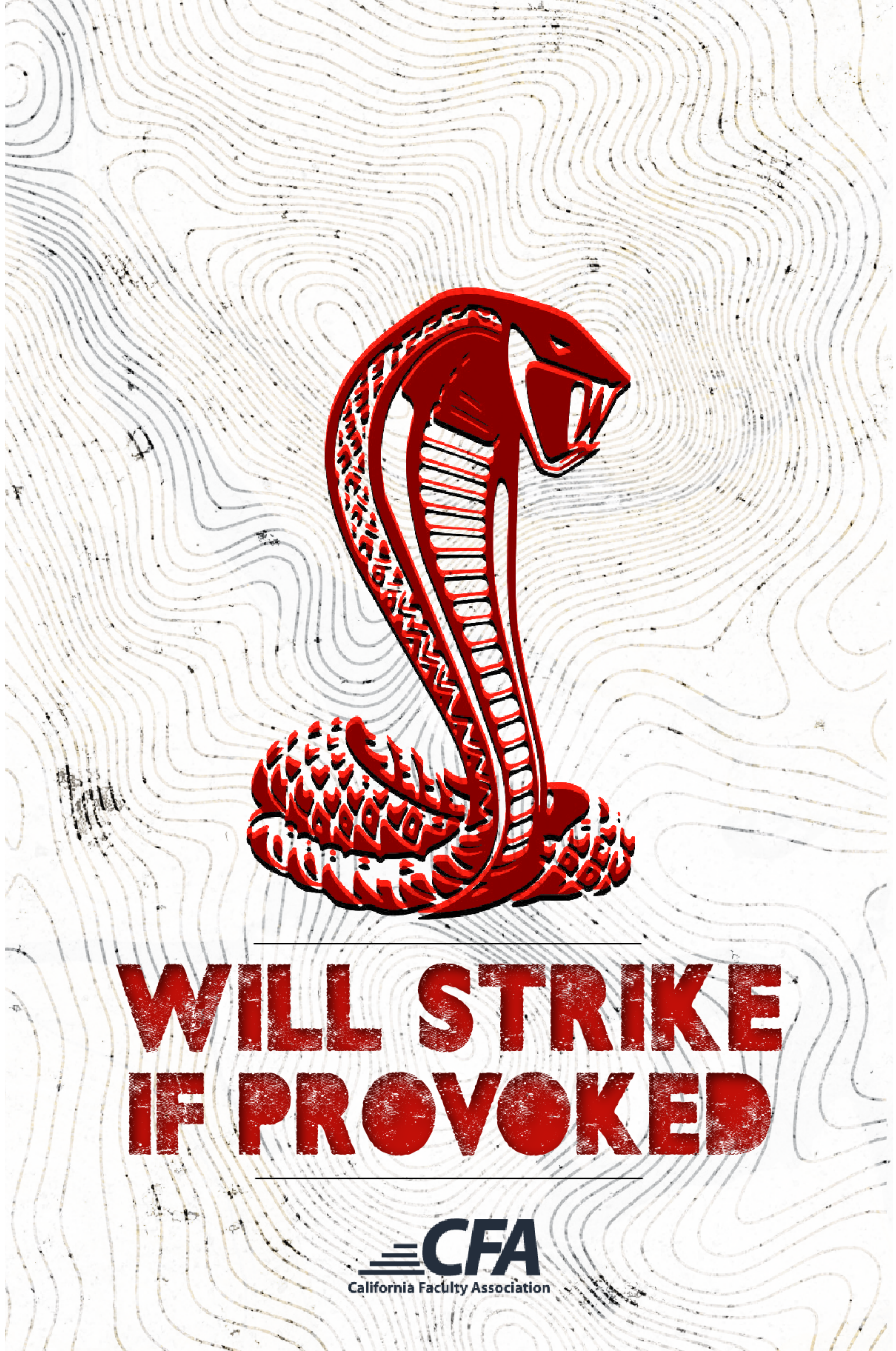
**ChatGPT**

- Released November 2022
- Powered by deep learning
- It is impossible for any individual to recreate this model from scratch
- But we will try to understand its core mechanisms by the end of the course

# Introductions

# Who am I?

- Tyler
- Machine learning engineering at Yelp and LinkedIn
- Applied mathematics and biology in college
- Been playing bass for 2 years
- Part-time EMT




# WHERE DOES THE MONEY GO?

CSU's mission to serve California's working families is undermined by management diversion of student tuition and tax-payer dollars from classroom instruction to **out-of-control executive enrichment** and **cash investments and reserves larger than the GDP of 60 nations.**

## Hoarded funds

In 2018, the California State Auditor criticized CSU management for **concealing a 1.5 billion dollar slush fund** from legislators and students as it proposed tuition increases. By diverting massive sums from instructional budgets, CSU cash investments, cash accounts, balances and reserves have now grown to an unprecedented

**$12.3 billion**

## Executive ghost campus

While rejecting CFA proposals to keep faculty pay ahead of inflation and lift the salaries of the lowest paid faculty, the **CSU funnels more resources to the Chancellor's Office than to six CSU campuses.** The budget for this "ghost campus" where no student graduates has now risen to almost

**$200 million**

## The Golden Chancellor

In an August 10 letter, interim Chancellor García called CFA's salary proposal “unreasonable because it would grossly undermine the CSU’s fiscal stability.” But the CSU **Chancelor's salary increased 66% since 2020**. On top of a $795,000 base salary, the new Chancellor gets a $96,000 housing allowance, $80,000 in deferred compensation, and a $12,000 car allowance, bringing the grand total to

**$983,000**

## Lavish executive pay

Growth in management pay outstrips that of all other campus employees. While working families struggle to pay college costs, **CSU campus presidents earn up to $533000**, plus perks like **housing allowances up to $60,000**. Before raising tuition 34%, CSU Trustees awarded campus presidents yet another raise; some got

**29% raises**

**The obstacle to a fair faculty contract isn't a budget shortfall but the misuse of public funds by an irresponsible managerial elite.**

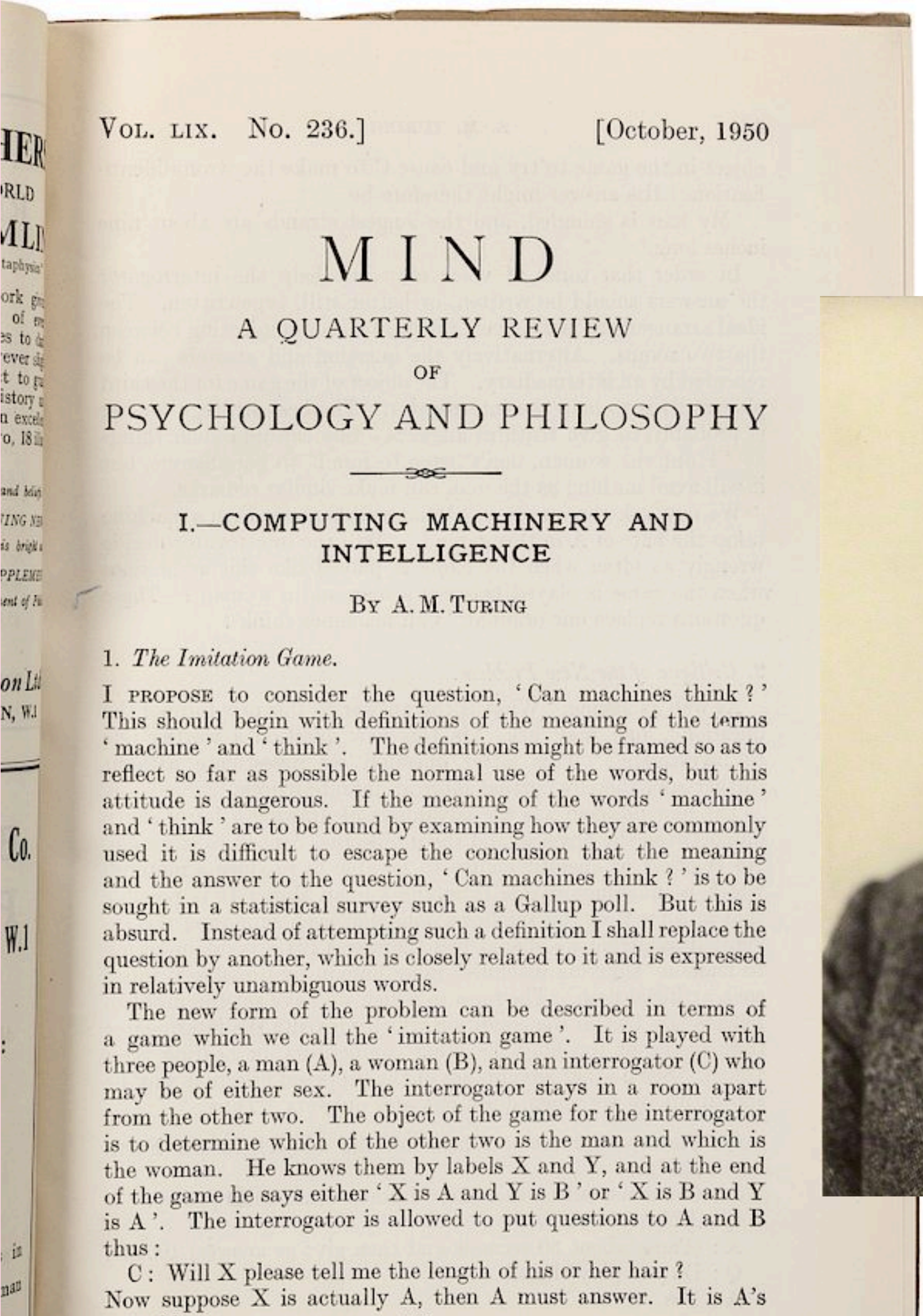
Sources:
https://csu.opengov.com
https://jacobin.com/2023/08/california-state-university-union-wages-inequality-administrators-public-education
https://www.auditor.ca.gov/reports/2016-122/index.html
https://www.calstate.edu/csu-system/about-the-csu/facts-about-the-csu/csu-funding
https://www.auditor.ca.gov/reports/2018-127/sections.html#section1

CFA
California Faculty Association

# Who are you?

# What this class is about

# Where it all began



VOL. LIX. No. 236.] [October, 1950

# MIND

## A QUARTERLY REVIEW OF PSYCHOLOGY AND PHILOSOPHY

### I.—COMPUTING MACHINERY AND INTELLIGENCE

BY A. M. TURING

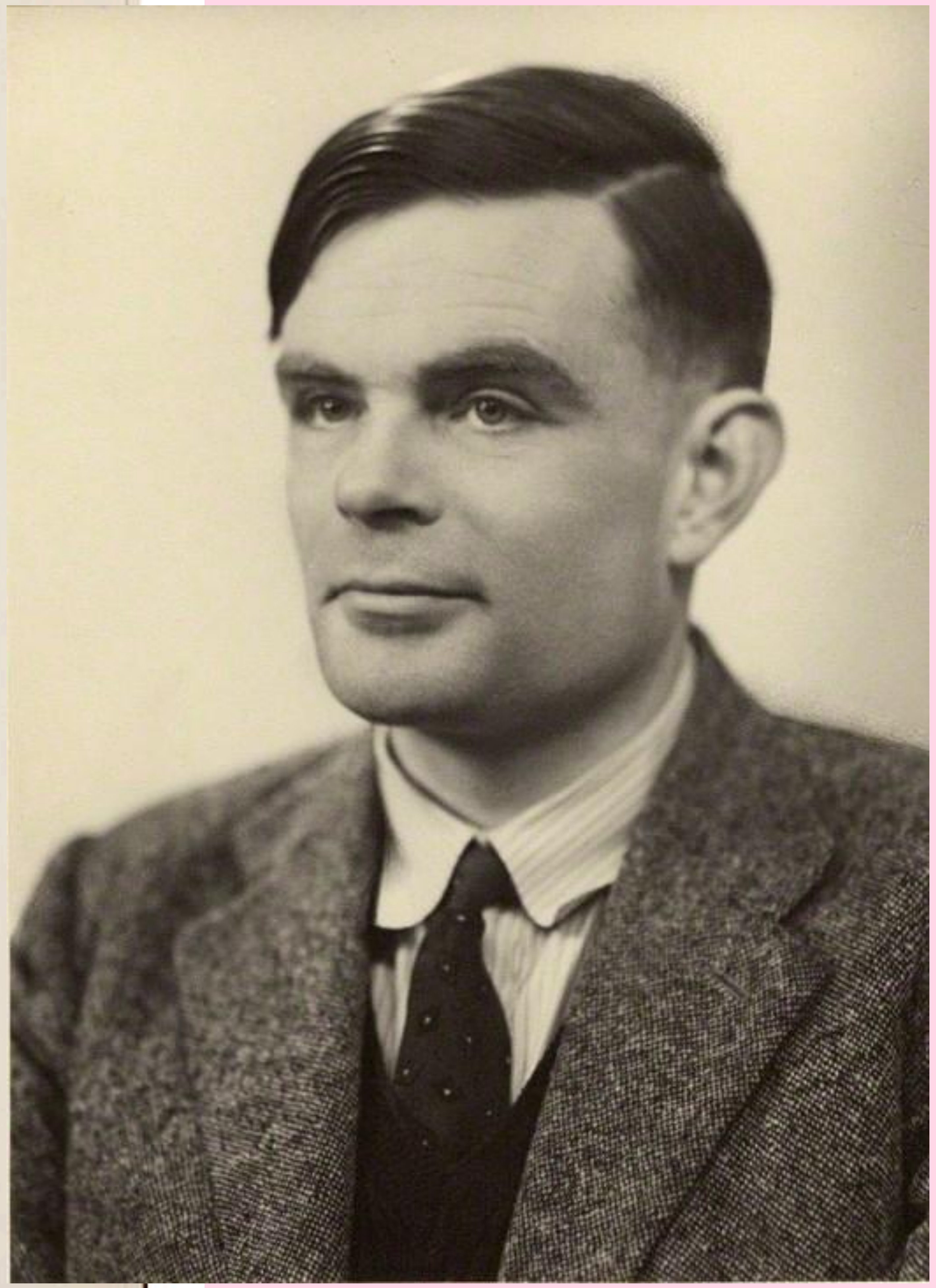
1. *The Imitation Game.*

I PROPOSE to consider the question, 'Can machines think ?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words 'machine' and 'think' are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, 'Can machines think ?' is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

The new form of the problem can be described in terms of a game which we call the 'imitation game'. It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either 'X is A and Y is B' or 'X is B and Y is A'. The interrogator is allowed to put questions to A and B thus :

C : Will X please tell me the length of his or her hair ?

Now suppose X is actually A, then A must answer. It is A's



Alan Turing (1912–1954)

The Turing test

Turing machines

Turing completeness

*People have been thinking of AI for as long as there have been computers.*

# What is special about AI?

AI versus

- programming languages
- operating systems
- databases
- networking

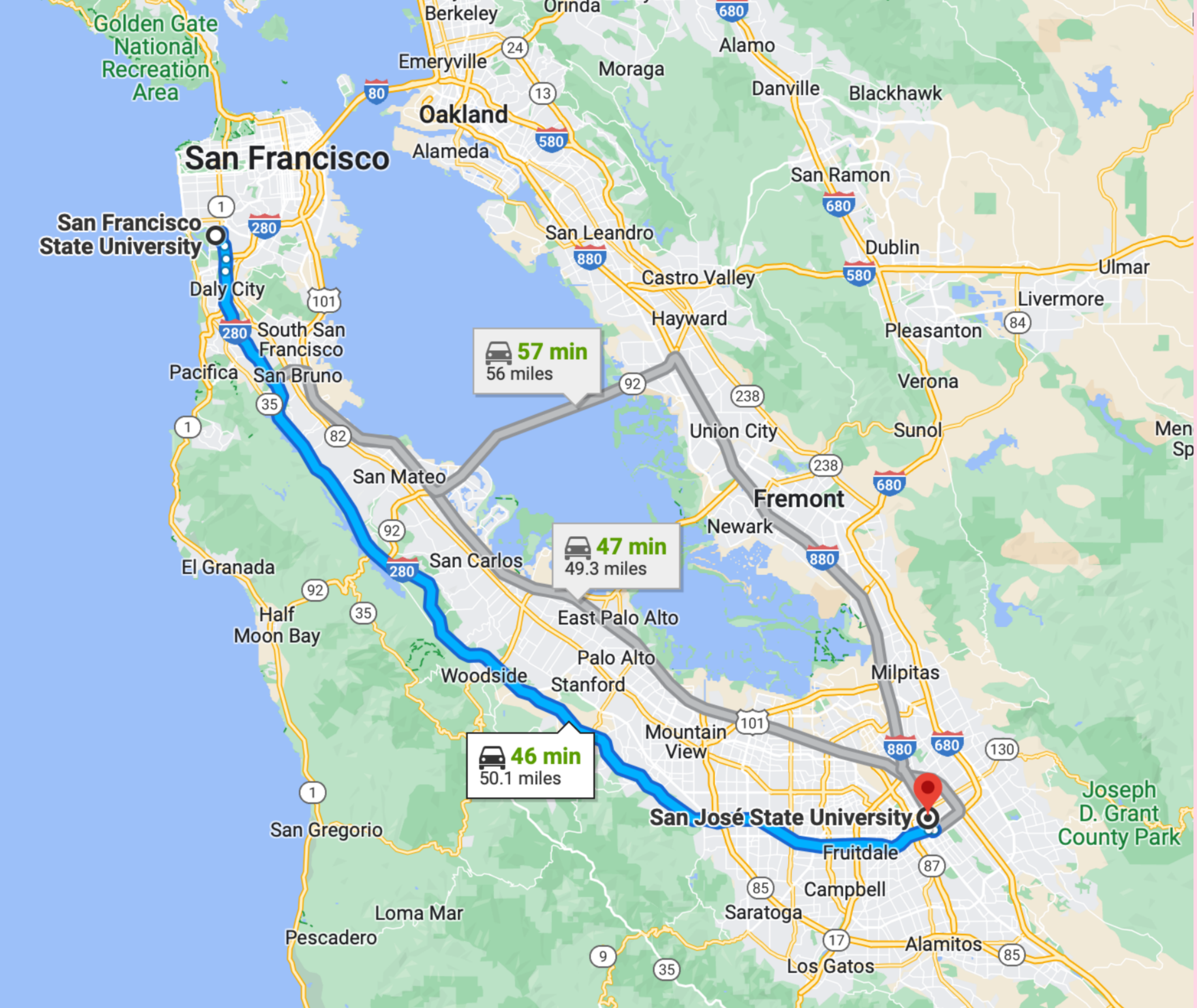
**This course:** a survey of some things in CS that were called AI at one point or another

# What is AI?

- **Two dimensions**
  - thought vs. behavior
  - human vs. rational
- AI research has mainly focused on producing agents capable of **rational behavior**
- Agents should "do the right thing"
- Agents should be capable of behavior beyond their explicit programming
- **Example:** Is a Roomba AI?

# Framework

- **Modeling and representation:** How do we build an abstract representation to model the real world? (Always a lossy process.)
- **Inference and prediction:** How do we answer a specific question given a model?
- **Learning:** How can we use data to produce better models and make better inferences?
- **Politics:** How should we allocate harms and benefits in society, and who decides?




# Example: Driving Directions

(Is this AI?)

# Framework

- **Model:** Use a graph to model the world — intersections are nodes, roads are edges.
- **Inference:** Use graph search algorithms to find the shortest path from start to destination.
- **Learning:** Use data to annotate edges with driving times.
- **Politics:** Road congestion, mode of transportation, car dominance, landmark prominence. Mapmaking is always political.

# Entire course on one slide

- **Search:** make decisions by looking ahead
- **Logic:** deduce new facts from existing facts
- **Constraints:** find a way to satisfy a given specification
- **Probability:** reason quantitatively about uncertainty
- **Learning:** make future predictions from past observations

# Course mechanics

[***tddevlin.com/csc665-spring24/***](https://tddevlin.com/csc665-spring24/)

# Prerequisites

- Officially, CSC 413 (software development)
- Transitively, CSC 220 (data structures)
- Ideally, CSC 230 (discrete math), CSC 510 (algorithms), MATH 324 (probability)
- If you don't have the official prereq, either:
  - submit evidence of background knowledge to me via email
  - talk to me after class
- *Students who don't have the prereqs will be dropped*

# Prerequisites

- **Prerequisites elsewhere**
  - SJSU: DS & algorithms, OOP
  - SCU: DS, discrete math
  - Cal: DS, discrete math & probability
  - Stanford: intro CS, discrete math, probability, linear algebra
- We will use Python for all homework assignments
- In AI, the more math you know the better
- Problems involving significant math will be optional extra credit
- Use the first homework to calibrate

Attention is All You Need, Vaswani et al.

85K citations

The mechanism at the core of all modern LLMs

Math content

- summation notation
- matrix operations
- big-oh notation

#### 3.2.2 Multi-Head Attention

Instead of performing a single attention function with $d_{\text{model}}$-dimensional keys, values and queries, we found it beneficial to linearly project the queries, keys and values $h$ times with different, learned linear projections to $d_k$, $d_k$ and $d_v$ dimensions, respectively. On each of these projected versions of queries, keys and values we then perform the attention function in parallel, yielding $d_v$-dimensional output values. These are concatenated and once again projected, resulting in the final values, as depicted in Figure 2.

Multi-head attention allows the model to jointly attend to information from different representation subspaces at different positions. With a single attention head, averaging inhibits this.

[4] To illustrate why the dot products get large, assume that the components of $q$ and $k$ are independent random variables with mean 0 and variance 1. Then their dot product, $q \cdot k = \sum_{i=1}^{d_k} q_i k_i$, has mean 0 and variance $d_k$.



$$\begin{aligned}\mathrm{MultiHead}(Q, K, V) &= \mathrm{Concat}(\mathrm{head_1}, ..., \mathrm{head_h})W^O \\ \text{where}\ \mathrm{head_i} &= \mathrm{Attention}(QW_i^Q, KW_i^K, VW_i^V)\end{aligned}$$

Where the projections are parameter matrices $W_i^Q \in \mathbb{R}^{d_{\text{model}} \times d_k}$, $W_i^K \in \mathbb{R}^{d_{\text{model}} \times d_k}$, $W_i^V \in \mathbb{R}^{d_{\text{model}} \times d_v}$ and $W^O \in \mathbb{R}^{hd_v \times d_{\text{model}}}$.

In this work we employ $h = 8$ parallel attention layers, or heads. For each of these we use $d_k = d_v = d_{\text{model}}/h = 64$. Due to the reduced dimension of each head, the total computational cost is similar to that of single-head attention with full dimensionality.

Table 1: Maximum path lengths, per-layer complexity and minimum number of sequential operations for different layer types. $n$ is the sequence length, $d$ is the representation dimension, $k$ is the kernel size of convolutions and $r$ the size of the neighborhood in restricted self-attention.

| Layer Type | Complexity per Layer | Sequential Operations | Maximum Path Length |
|---|---|---|---|
| Self-Attention | $O(n^2 \cdot d)$ | $O(1)$ | $O(1)$ |
| Recurrent | $O(n \cdot d^2)$ | $O(n)$ | $O(n)$ |
| Convolutional | $O(k \cdot n \cdot d^2)$ | $O(1)$ | $O(log_k(n))$ |
| Self-Attention (restricted) | $O(r \cdot n \cdot d)$ | $O(1)$ | $O(n/r)$ |

Compare with a "real" math paper

**Lemma 7.4.** *Let* $\mathrm{OG}^{\mathrm{stat,red}}_{(g,m,r,s,a,d)} \subset \mathrm{OG}^{\mathrm{stat,red}}_{g}$ *be as in Lemma 6.6. Then*

$$\int_{\mathrm{OG}^{\mathrm{stat,red}}_{(g,m,r,s,a,d)}} \alpha = \frac{-1}{(a_1!)\dots(a_s)!}\frac{(r+n-2)!}{r!} \cdot B_r,$$

*where* $B_r$ *is the Bernoulli number and* $n = |a| = a_1 + \cdots + a_s$.

*Proof.* First dispense with a special case that $(r,n) = (0,2)$. Then either $s = 1$, $a_1 = 2$, and both sides are $-1/2$, or $s = 2$, $a_1 = a_2 = 1$, and both sides are $-1$.

Otherwise, an object of $\mathrm{OG}^{\mathrm{stat,red}}_{(g,m,r,s,a,d)}$ has $n$ many tails, each of which consists of an edge ending in the tail point, which is a valence-1 vertex. If we pick a total ordering of the set of the $a_1$ many tails with value $d_1$, a total ordering of the set of the $a_2$ many tails with value $d_2$, etc, we arrive at an object of the groupoid $\Gamma^{\mathrm{p}}_{r,n}$ defined exactly as $\Gamma^{\circ}_{r,n}$, except that we do not require the marking function $m : \{1,\dots,n\} \to V$ to be injective. The superscript $p$ stands for "pure," as in [CGP22]. Accounting for the choices of orderings of marked points, and the $n$ many extra edges where the tails are attached, we therefore get

$$\int_{\mathrm{OG}^{\mathrm{stat,red}}_{(g,m,r,s,a,d)}} \alpha = \frac{(-1)^n}{(a_1!)\dots(a_s)!}\int_{\Gamma^{\mathrm{p}}_{r,n}} \alpha = \frac{(-1)^n}{(a_1!)\dots(a_s)!}\sum_{(G,m)\in\Gamma^{\mathrm{p}}_{r,|a|}} \frac{(-1)^{|E(G)|}}{|\mathrm{Aut}(G)|}.$$

The lemma now follows from the formula

$$\sum_{(G,m)\in\Gamma^{\mathrm{p}}_{r,n}} \frac{(-1)^{|E(G)|}}{|\mathrm{Aut}(G)|} = (-1)^{n+1}\frac{(r+n-2)!}{r!}\cdot B_r, \tag{18}$$

valid when $2r - 2 + n > 0$. The case $n = 0$, $r \geq 2$ of this formula was given by Kontsevich [Kon93]; see [Ger04, §7.1] for a proof. The general case is easily deduced by induction on $n$; see Appendix 10. ☐

## 8. Conclusion

In light of Lemma 7.1, we may apply (8) to the functor $\mathcal{R} : \mathrm{OG}^{\mathrm{stat}}_g \to \mathrm{OG}^{\mathrm{stat,red}}_g$ to rewrite the formula in Corollary 5.9 as

$$\begin{aligned} z_g &= \int_{\mathrm{OG}^{\mathrm{stat}}_g} \alpha\cdot\beta\cdot\gamma \\ &= \int_{\mathrm{OG}^{\mathrm{stat,red}}_g} (\mathcal{R}_*\alpha)\cdot\beta\cdot\gamma. \end{aligned}$$

We may now replace $\mathrm{OG}^{\mathrm{stat,red}}_g$ by the coproduct in Lemma 6.6, and observe by Proposition 7.2 that $\alpha$ and $\beta$ are constant functions on each $\pi_0(\mathrm{OG}^{\mathrm{stat,red}}_{(g,m,r,s,a,d)})$.

$$z_g = \sum_{(k,m,r,s,a,d)} \left(\prod_{i=1}^{s}(-\mu(m/d_i))^{a_i}\int_{\mathrm{OG}^{\mathrm{stat,red}}_{(g,m,r,s,a,d)}} \alpha\right)\left(P_m^{1-r}\prod_{i=1}^{s}\left(\frac{P_{d_i}}{P_m}\right)^{a_i}\right)\left(m^{r-1}\prod_{p|D}(1-p^{-r})\right).$$

Combining with the formula in 7.4 then finishes the proof of Theorem 1.1. ☐

# How hard is the class?

- **Last semester**
  - median grade: 54
  - high: 93
- **Letter grade distribution:** 11 A's 10 B's 11 C's 1 F
- "This is the hardest class I've ever taken. It was way harder than 415."
- "This professor teaches a very hard class but has his own way of making it simplified."

# Do students like the class?

- “This was the best course I've taken so far.”
- “The instructor does not know how to teach.”
- “Tyler is an exceptional instructor, arguably one of the finest I've encountered during my time at SFSU.”
- “The instructor was great, had everything organized, and provided materials that stimulated my learning.

# Attendance

- Lecture attendance is optional, but *this is not a remote-friendly course*
- Exam attendance is *mandatory*
- Participation grade: up to 2% extra credit (quality over quantity)
- Forms of participation:
  - Attend lecture and ask questions
  - Attend lecture and answer questions
  - Post on the discussion forum
- Please *do not* come to class if you are sick

# Academic integrity

- You are encouraged to study together, but any work you submit must be your own
- Talk to me if you feel yourself slipping
- Past students have received zeros because of plagiarism
- *You should not expect to pass the class if you violate the honor code*

# Academic integrity

- **Okay**
  - Discussing assignments verbally
  - Taking brief notes during discussions with classmates
  - Posting high-level pseudocode on the discussion board
- **Not okay**
  - Copy-pasting code
  - Taking pictures/screenshots of homework solutions
  - Typing on someone else's laptop
  - Typing on your own laptop while looking at someone else's work

# Other section

- CSC 665 is also offered on Tuesdays and Thursdays, 9:30 – 10:45 am
- **Instructor:** Akila de Silva
- Different approach, different style, different emphasis
- Consider checking it out

# Review of course resources

- **Lectures:** learn material, ask questions
- **Office hours:** ask more questions
- **Canvas:** submit assignments, view grades, get announcements, message me
- **Website:** see calendar, get homeworks, view lecture slides
- **Campuswire:** post questions, help your classmates
- **Textbook:** read for details

# Things to do

- Homework 0 is out; due next Wednesday 2/7 at midnight.
- Make a campuswire account and join the class.
- If you don't have the official prereq, either:
  - submit evidence of background knowledge to me via email
  - talk to me after class
- Watch for Canvas announcement on Monday about whether we're back in the classroom next Tuesday


POWER
GAME BOY
GAME BOY COLOR
Nintendo
B
SELECT START
POKéMON
DMG-AAXE-USA
THIS SIDE OUT
EVERYONE
SILVER VERSION

# Welcome and good luck!