



# Parrots and Pedagogues

Two Questions on AI

---

```
graph LR; A((What is AI and how does it work?)) --> B((What are the opportunities/risks?))
```

What is AI and how does it work?

What are the opportunities/risks?

Me.

10+ years as a researcher

Center for AI and Digital Ethics,  
CIS, at Melbourne Uni (2+ years)

Worked on technology policy at:

- Princeton University
- U.S. Senate
- Placement at Federal Trade Commission

**“I’ve got fantastic new tech  
for our unit!”**

---



**What questions  
would you ask  
this person?**

How much does it cost?

How do our stakeholders feel about it?

What problem does it solve?

How usable is the technology?



How compliant is it?

How safe is the technology?


What risks does it introduce?
















**“See, it works!”**





 Clear chat 

Chat History is off for this browser.  
When history is turned off, new chats on this browser won't appear in your history on any of your devices, be used to train our models, or stored for longer than 30 days. This setting does not sync across browsers or devices. [Learn more](#)

 Enable chat history

-  Add Beamer Page Numbers
-  Two-column frame with image
-  New chat
-  Define darkred in LaTeX
-  PCDR for Telecommunications
-  Helping with Baby Delivery
-  AI's Turning Point
-  AI's Turning Point
-  Reliability of ChatGPT
-  Algo Life Parody
-  Estimating  $\pi$  with Monte Carlo
-  Variable Sequence Determination
-  Grant Details Stored.
-  Masking Operations Program
-  Color-coded 32-cell Plot

 GPT-3.5  **GPT-4**

# ChatGPT PLUS

**Explain airplane turbulence**  
to someone who has never flown before

**Compare storytelling techniques**  
in novels and in films

**Come up with concepts**  
for a retro-style arcade game

**Plan a trip**  
to explore the rock formations in Cappadocia, Turkey

Send a message 



CO

I need to replace a broken motor on my Joyal B500 belt conveyor. What model motor is appropriate?





Choosing a replacement motor requires a thorough understanding of the specific application and the operating conditions the motor will face. The power, speed, efficiency, torque, voltage, insulation class, and other motor parameters must be suitable for the application.



However, since your previous example was for a conveyor system in a mining operation, an industrial-grade motor rated for heavy-duty applications would be appropriate.

Here is an example of a suitable replacement motor:

**Brand: Siemens**

**Model: 1LA1 256-4AA90**

This model is a Siemens squirrel-cage three-phase induction motor. It has an output power of 500 kW and is designed for a voltage of 690V. It's built for heavy-duty applications and is rated for continuous operation, making it suitable for use in a conveyor system.



siemens 1la1 256-4aa90 motor



Videos

Images

Maps

Shopping

News

Books

Flights

Finance

About 0 results (0.26 seconds)

Your search - **siemens 1la1 256-4aa90 motor** - did not match any documents.

Suggestions:

- Make sure that all words are spelled correctly.
- Try different keywords.
- Try more general keywords.





r/ChatGPT · Posted by u/151N 1 day ago 🏆 2 🗨️ 🔄 🌟



7.4k



## Chatgpt Helped me pass an exam with 94% despite never attending or watching a class.

Educational Purpose Only

Hello, This is just my review and innovation on utilizing Ai to assist with education

The Problem:

I deal with problems, so most of my semester was spent inside my room instead of school, my exam was coming in three days, and I knew none of the lectures.

How would I get through 12 weeks of 3-2 hours of lecture per week in three days?

The Solution: I recognized that this is a majorly studied topic and that it can be something other than course specific to be right; the questions were going to be multiple choice and based on the information in the lecture.

I went to Echo360 and realized that every lecture was transcribed, so I pasted it into Chat gpt and asked it to:

"Analyze this lecture and use your algorithms to decide which information would be relevant as an exam, Make a list."

The first time I sent it in, the text was too long, so I utilized <https://www.paraphraser.io/text-summarizer> to summarize almost 7-8k words on average to 900-1000 words, which chat gpt could analyze.

Now that I had the format prepared, I asked Chat Gpt to analyze the summarized transcript and highlight the essential discussions of the lecture.

It did that exactly; I spent the first day Listing the purpose of each discussion and the major points of every lecturer in the manner of 4-5 hours despite all of the content adding up to 24-30 hours.

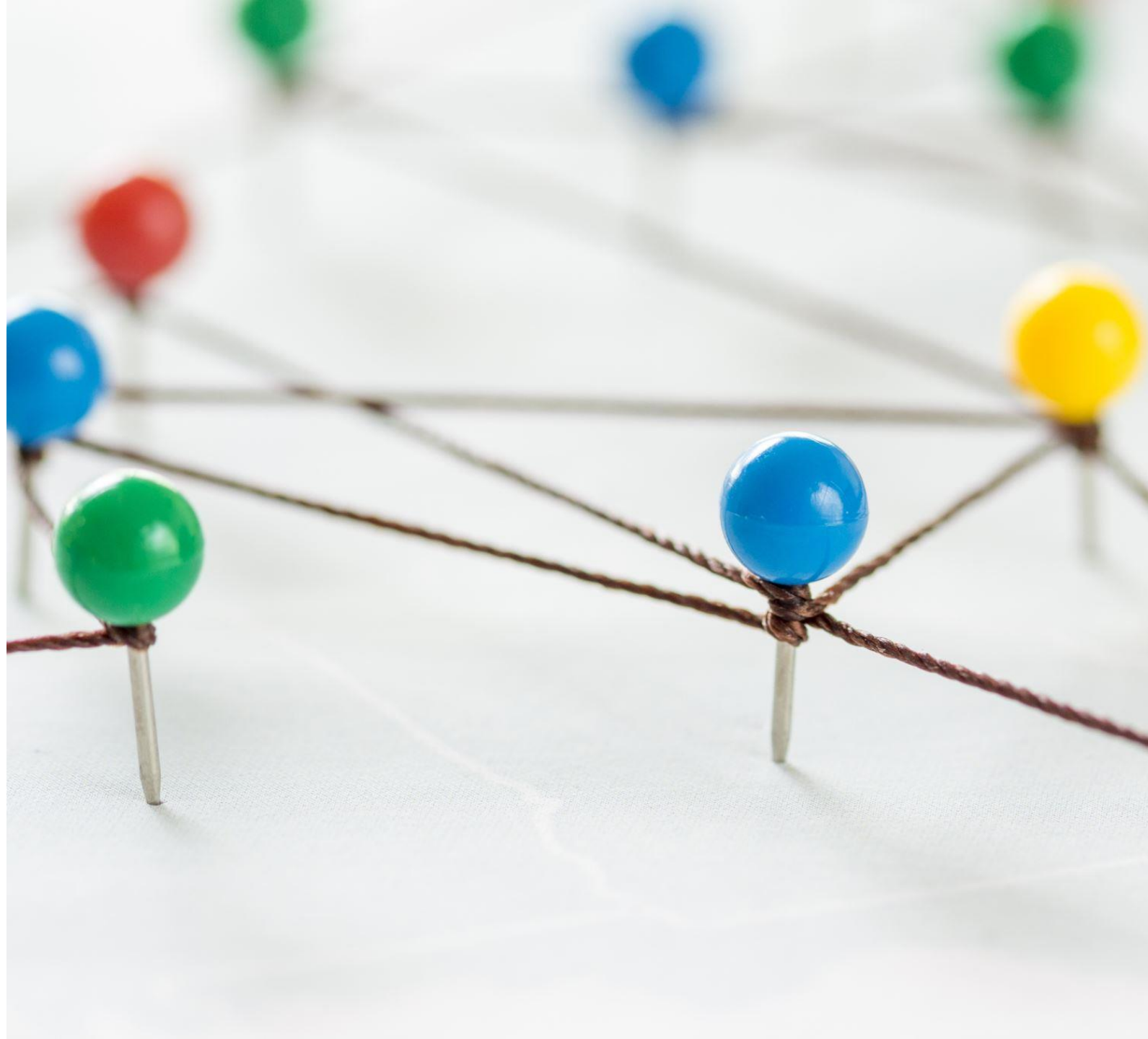
The next day, I asked Chat gpt to define every term listed as the significant "point" in every lecture **only** using the course textbook and the transcript that had been summarized; this took me 4-5 hours to make sure the information was accurate.

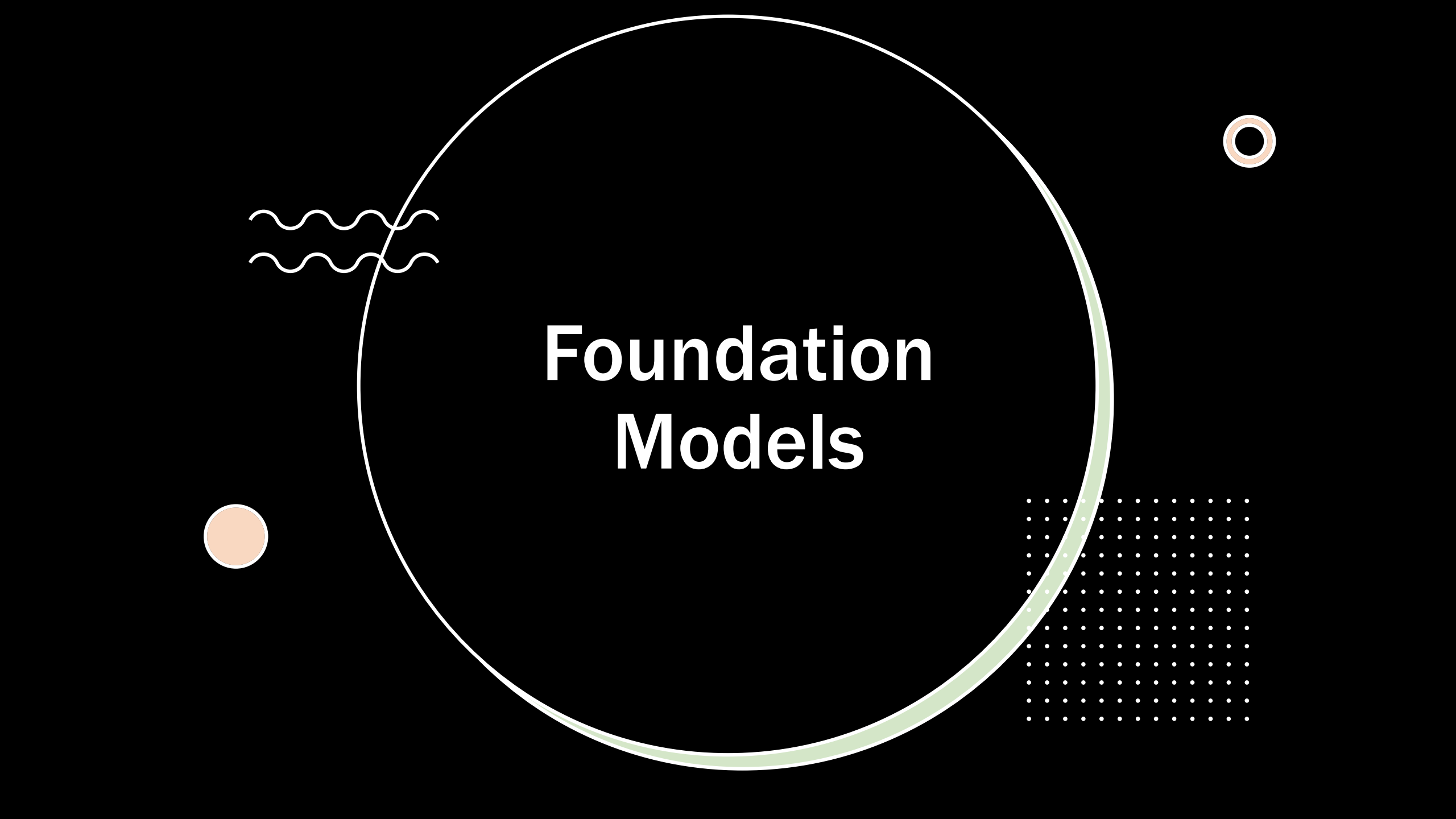
I spent the last day completely summarizing the information that chat gpt presented, and it was almost like the exam was an exact copy of what I studied,

The result: I got a 94 on the exam, despite me studying only for three days without watching a single lecture

# Machine Learning

- **Training:** build a map of the world (a 'statistical model')
- **Inference:** use the map to find what is at given coordinates
- Goal is to build a model that is both *accurate* and *useful*





# Foundation Models





Multi-functional statistical models

Can be *multi-modal* (input/output more than just text)



# Sequence Prediction

- ChatGPT (Microsoft)
- Bard (Google)
- LLaMA (Facebook)

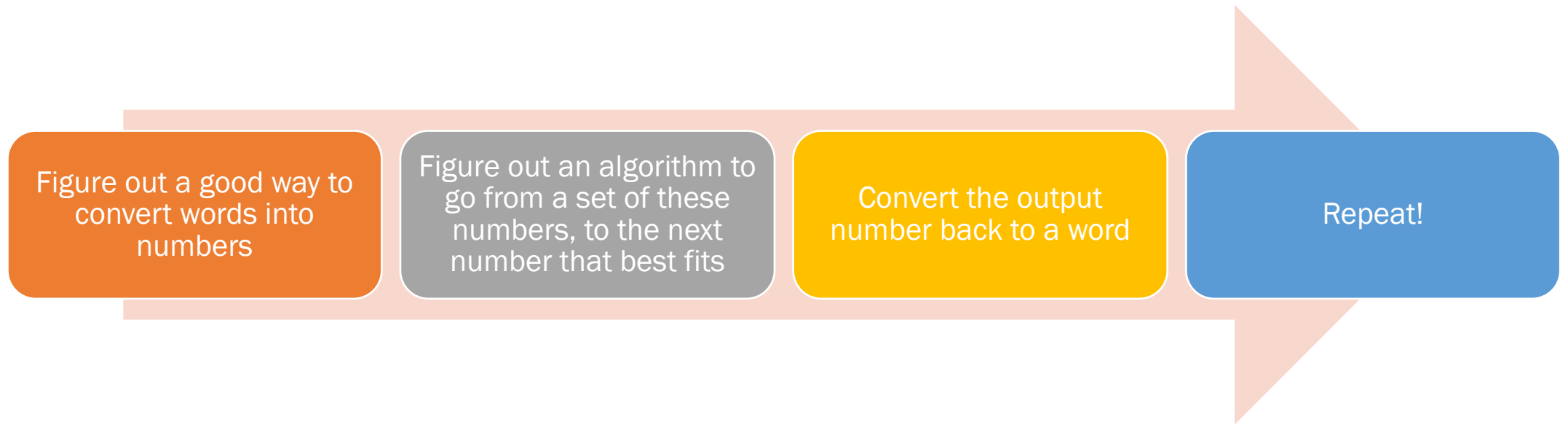
Given the start of a sequence:

”the” “student” “read” “from” “the” ???

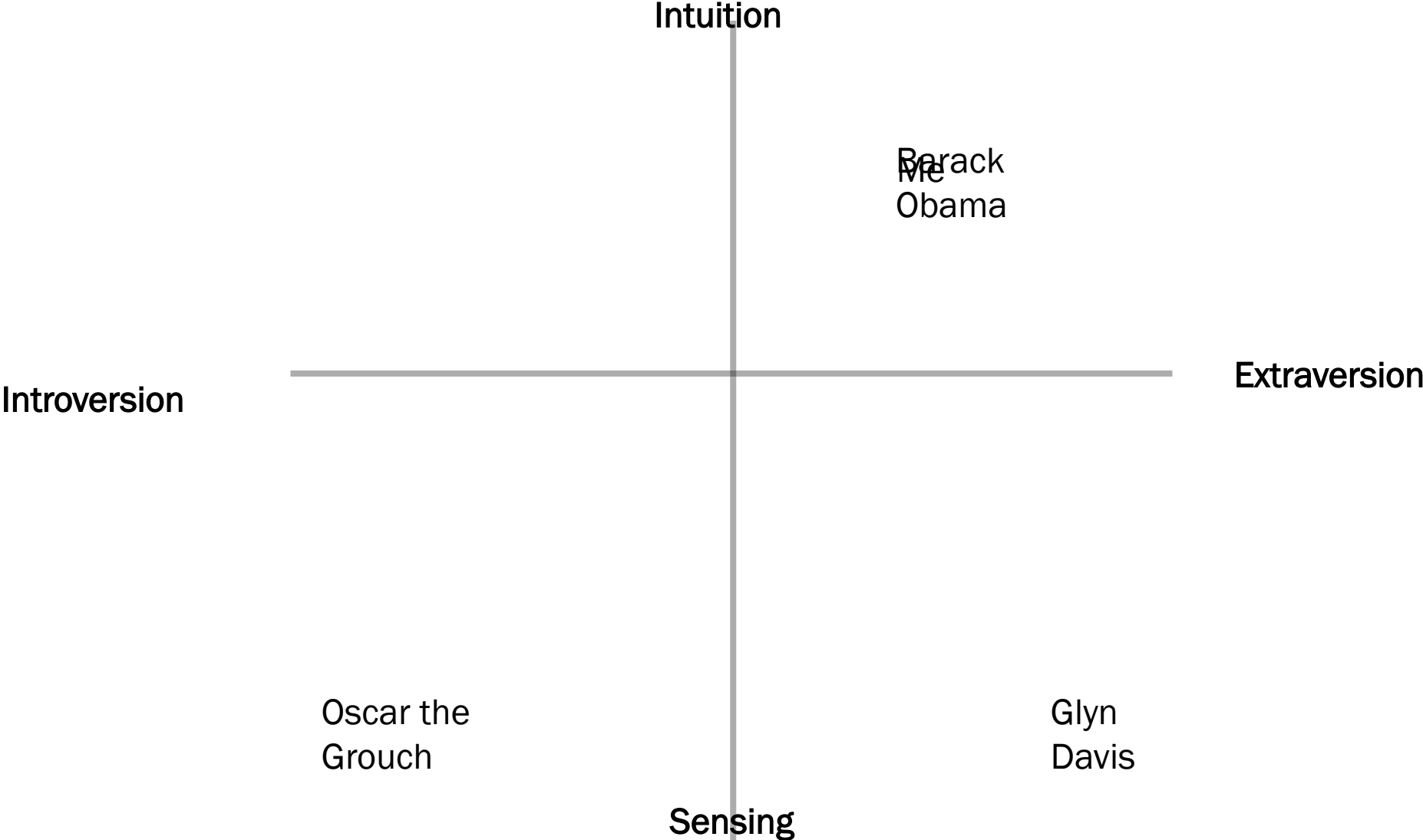




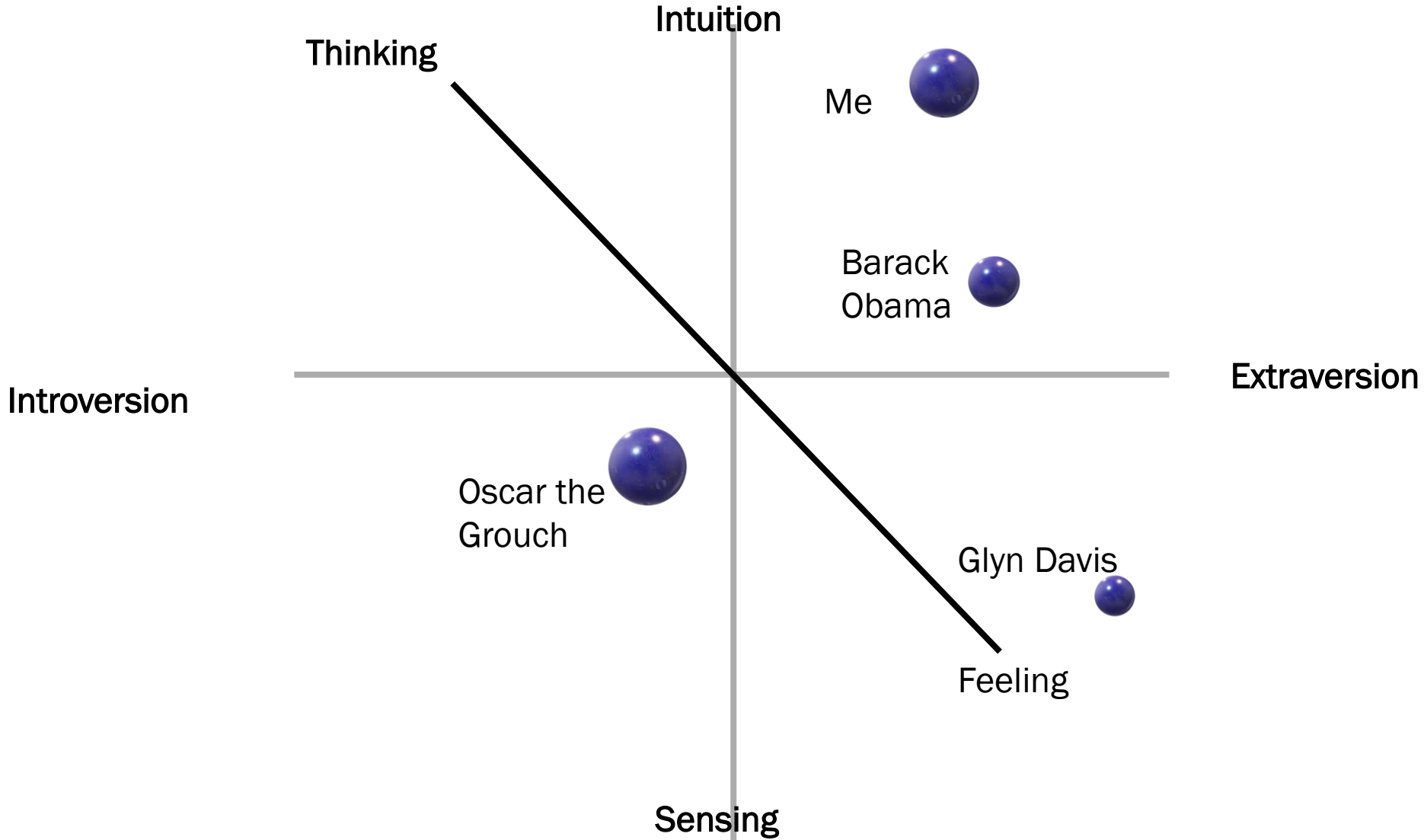
# To build a Large Language Model (LLM):



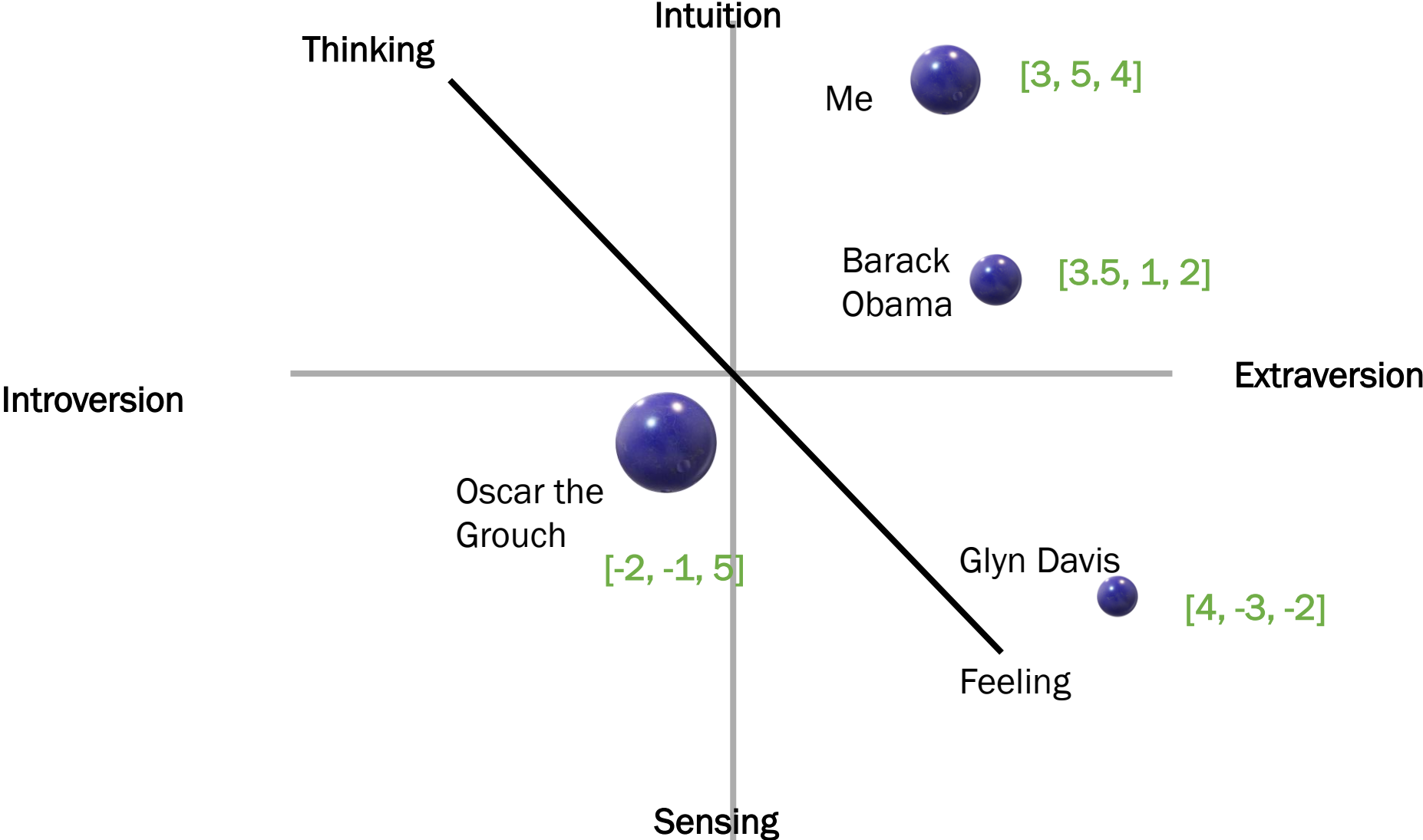
# Analogy: Myer-Briggs Type Indicators



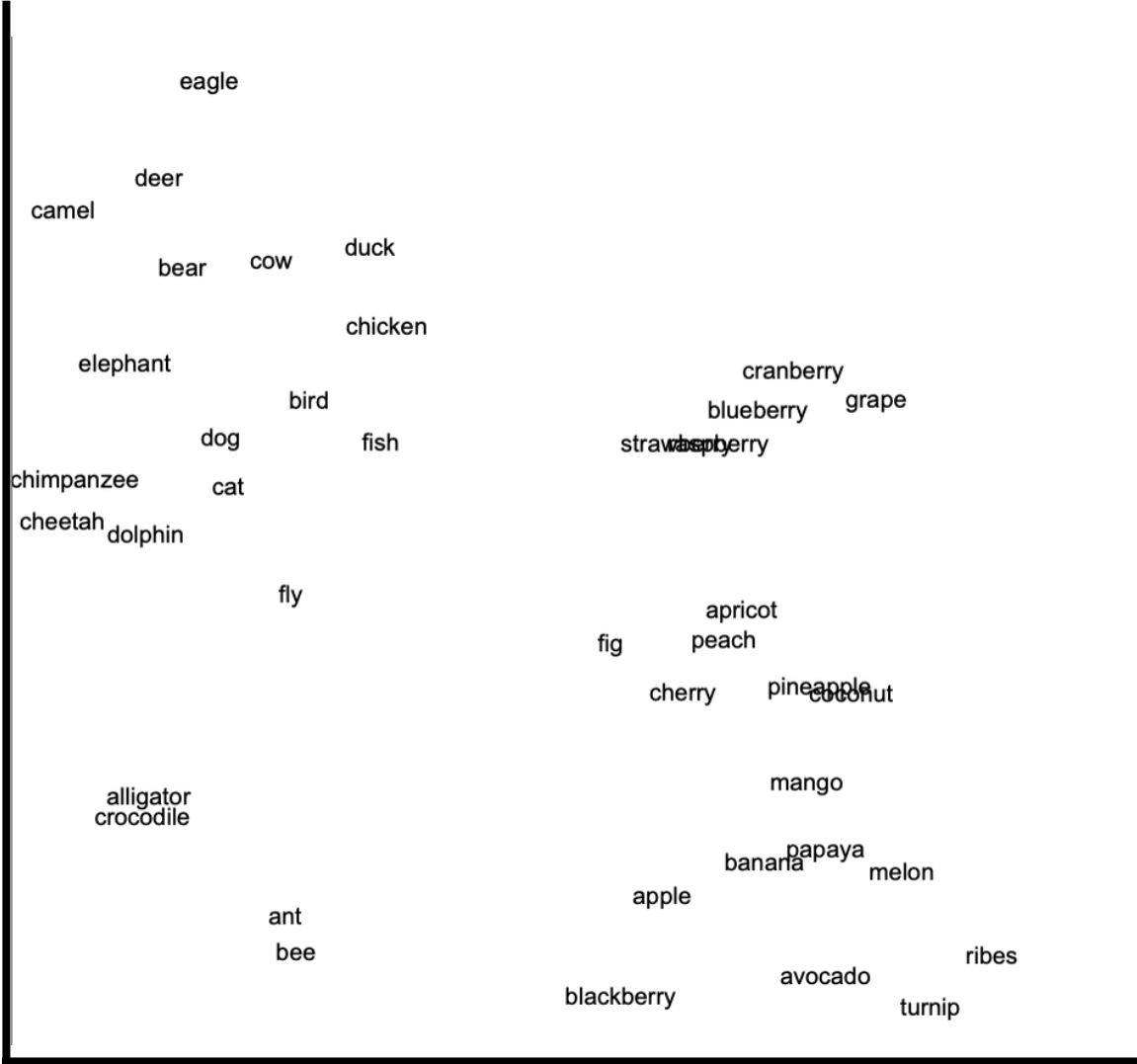
# Analogy: Myer-Briggs Type Indicators

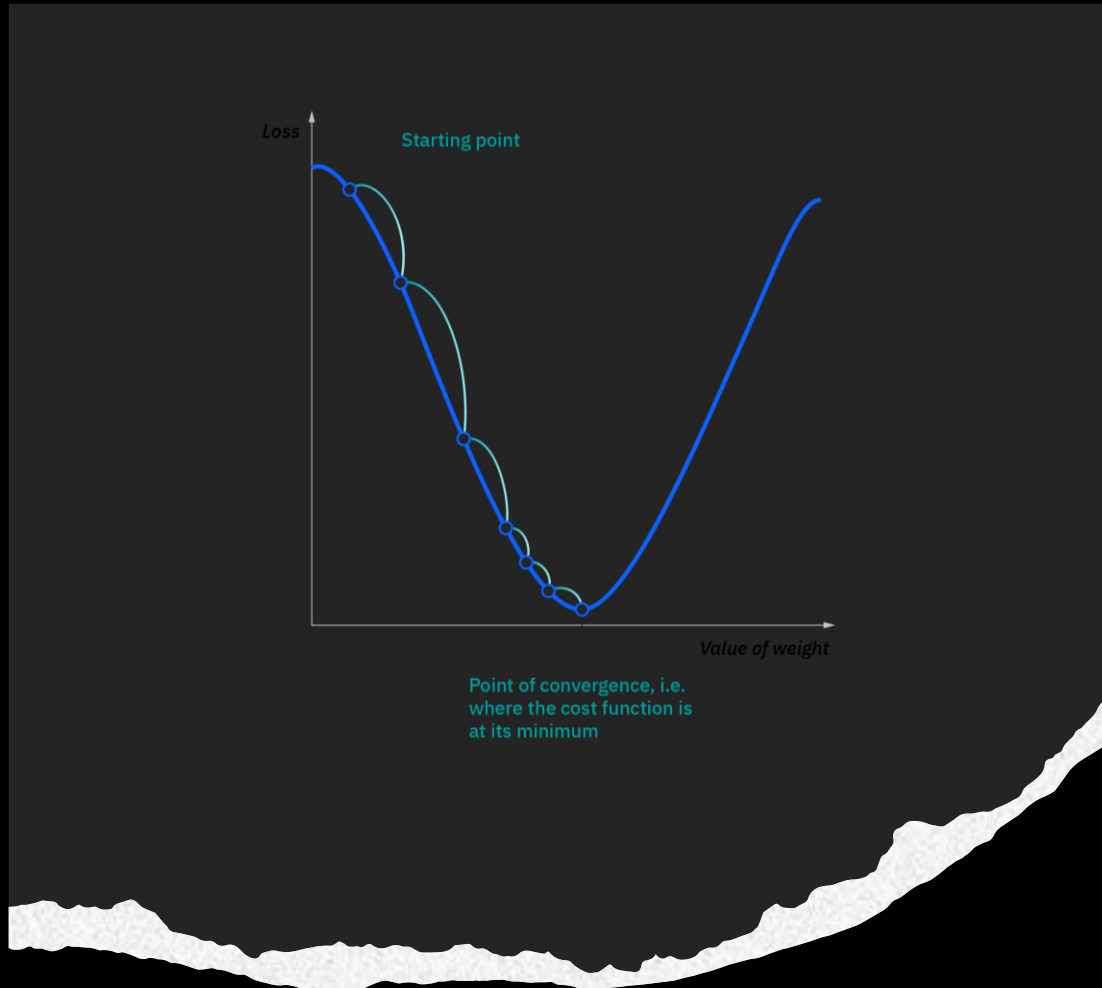


# Analogy: Myer-Briggs Type Indicators



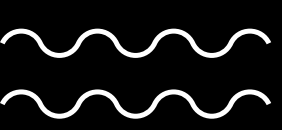
# Step 1: Convert Words to Numbers



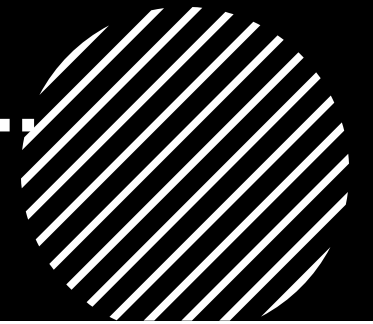


# Training

- Read text from the internet!
- Compare close-by words on our 'map' to close-by words in text
- Move things *a tiny amount* so the 'map' better matches text



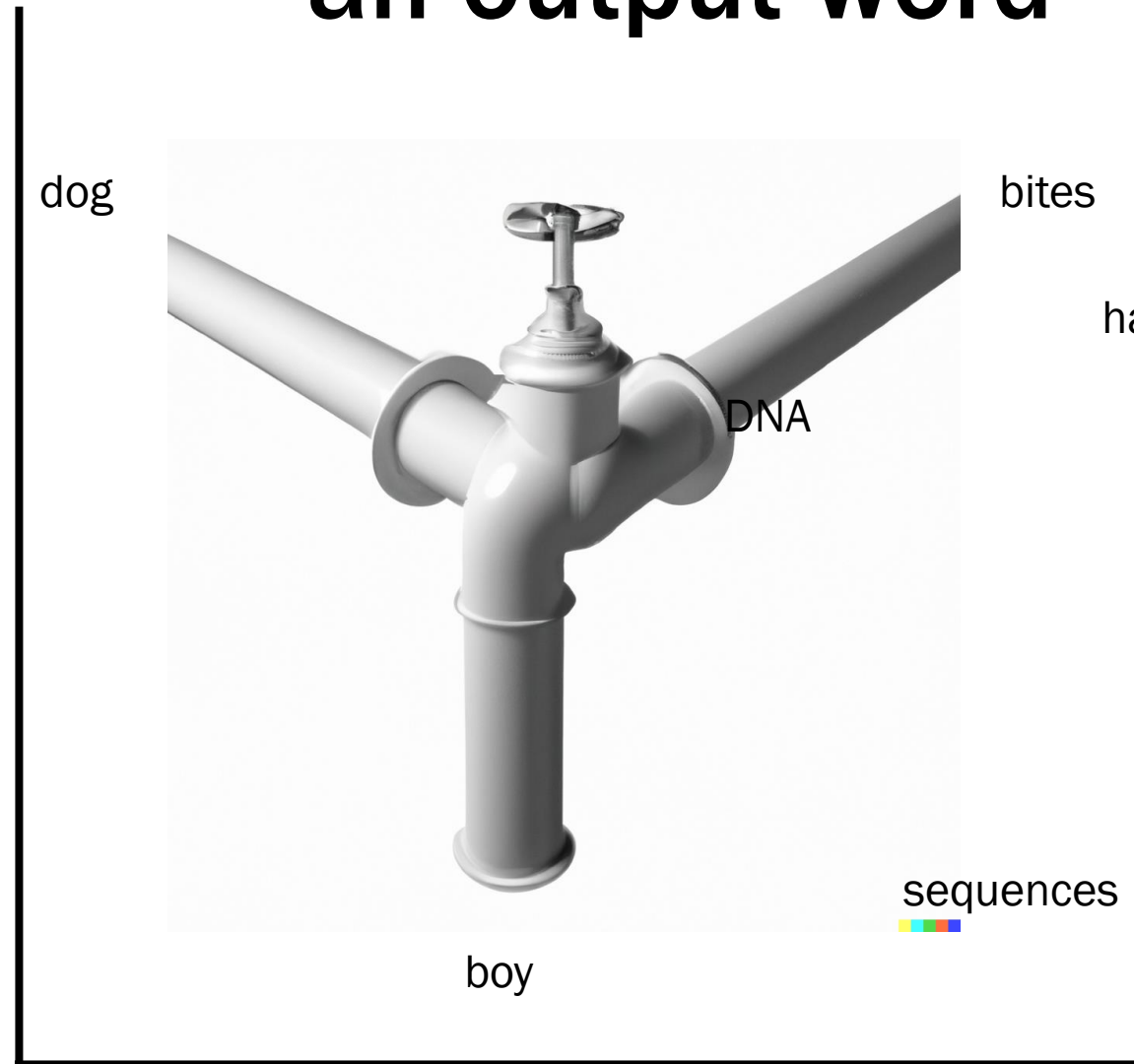
## Step 2: Go from a few words to the next word...



Build  
another map  
for this task

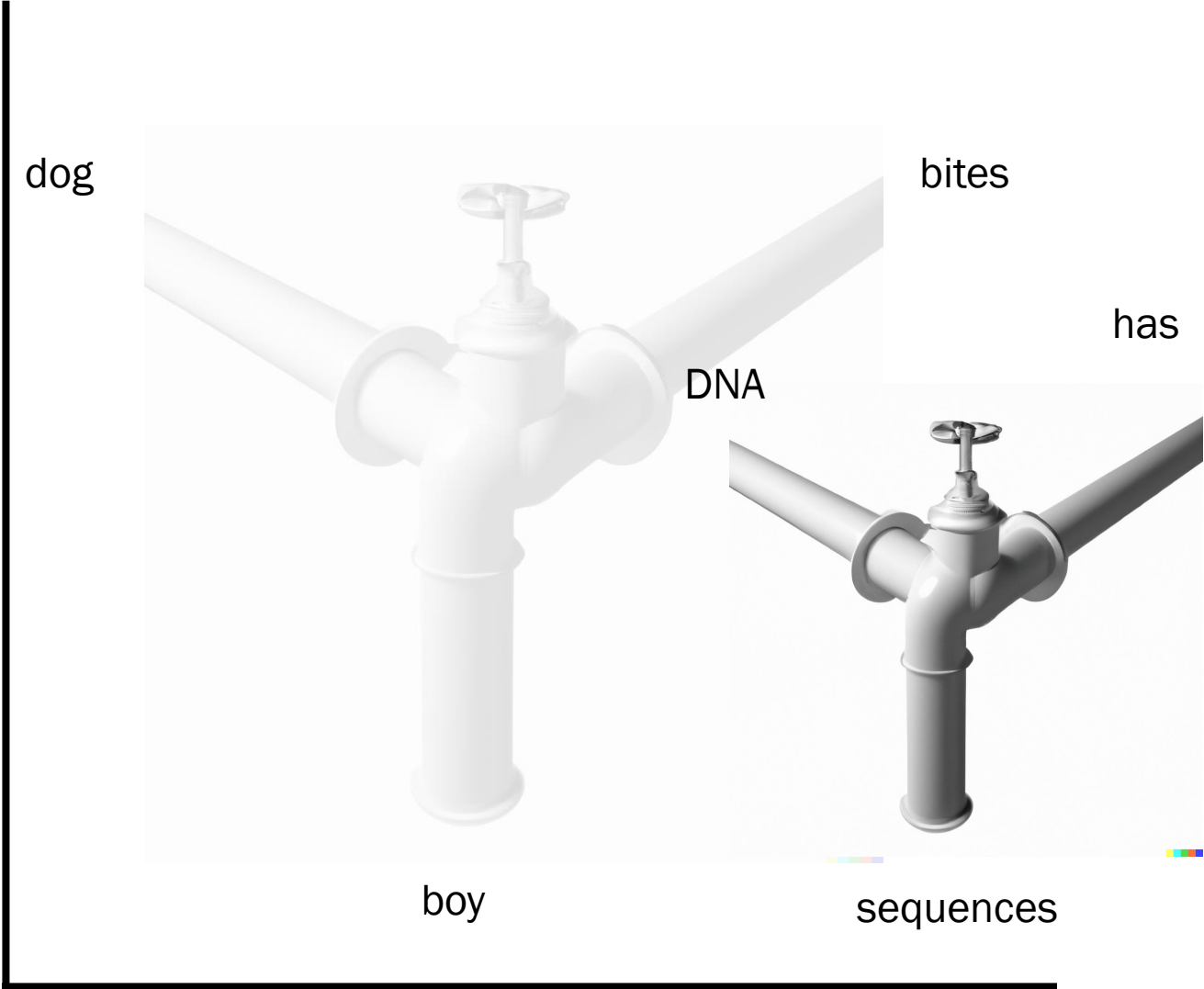
Use the  
map!


# Make rules that takes input words to give an output word





# Step 2: Our pipes are the map!





**The wettest  
word is the  
one we pick  
next**





**But we need too many pipes....**

**Each pipe is an individual rule.**

**Instead of lots of rules, we make a complicated equation...**

$$\begin{aligned} &w_{511}f(w_{311}f(b_{11} + xw_{111} + yw_{112}) + w_{312}f(b_{12} + xw_{121} + yw_{122}) + \\ &\quad w_{313}f(b_{13} + xw_{131} + yw_{132}) + w_{314}f(b_{14} + xw_{141} + yw_{142}) + b_{31}) + \\ &w_{512}f(w_{321}f(b_{11} + xw_{111} + yw_{112}) + w_{322}f(b_{12} + xw_{121} + yw_{122}) + \\ &\quad w_{323}f(b_{13} + xw_{131} + yw_{132}) + w_{324}f(b_{14} + xw_{141} + yw_{142}) + b_{32}) + \\ &w_{513}f(w_{331}f(b_{11} + xw_{111} + yw_{112}) + w_{332}f(b_{12} + xw_{121} + yw_{122}) + \\ &\quad w_{333}f(b_{13} + xw_{131} + yw_{132}) + w_{334}f(b_{14} + xw_{141} + yw_{142}) + b_{33}) + b_{51} \end{aligned}$$

**Owch...  
that was a  
lot of CS...**





**So what can LLMs  
actually do?**



# Strengths of Current Generation

---

1. **Stylism and Form**
  - Can easily match request *style* (eg: write an email to staff requesting comments on the proposed Process Framework)
2. **Explanations and Reading Comprehension**
3. **Data Munging**
  - Format conversions (eg text to table)
  - Extract all the X from the text
4. **Idea Generation**
  - eg: What are five ways Unimelb can improve its brand strategy?
5. **Basic coding!**



# Strengths Synthesis



Tasks that are easy to check/ground  
(What are 5 popular brands of motors)



Tasks where style > substance



Tasks where there is no ground truth  
*Write an inspiring ode to the PSC*





# Weaknesses of Current Generation

1. **Grounding**
  - Still large gaps on zero-shot responses to complex factual questions–improving rapidly
  - Cannot accurately cite sources
2. **Computation**
  - Numerical or computation results are frequently incorrect
3. **High-level organisation & curation**
  - Not yet good at strategic level organisation of information
4. **Current events (training is time limited)**

The background features a repeating pattern of light gray speech bubbles, each containing a dark teal question mark. The speech bubbles are arranged in a staggered, overlapping grid. The overall color palette is a mix of teal and gray.

**Not the only form of AI.**

And probably not the best for most of your imagined tasks.



# Biggest Opportunity: Integration into Existing Systems

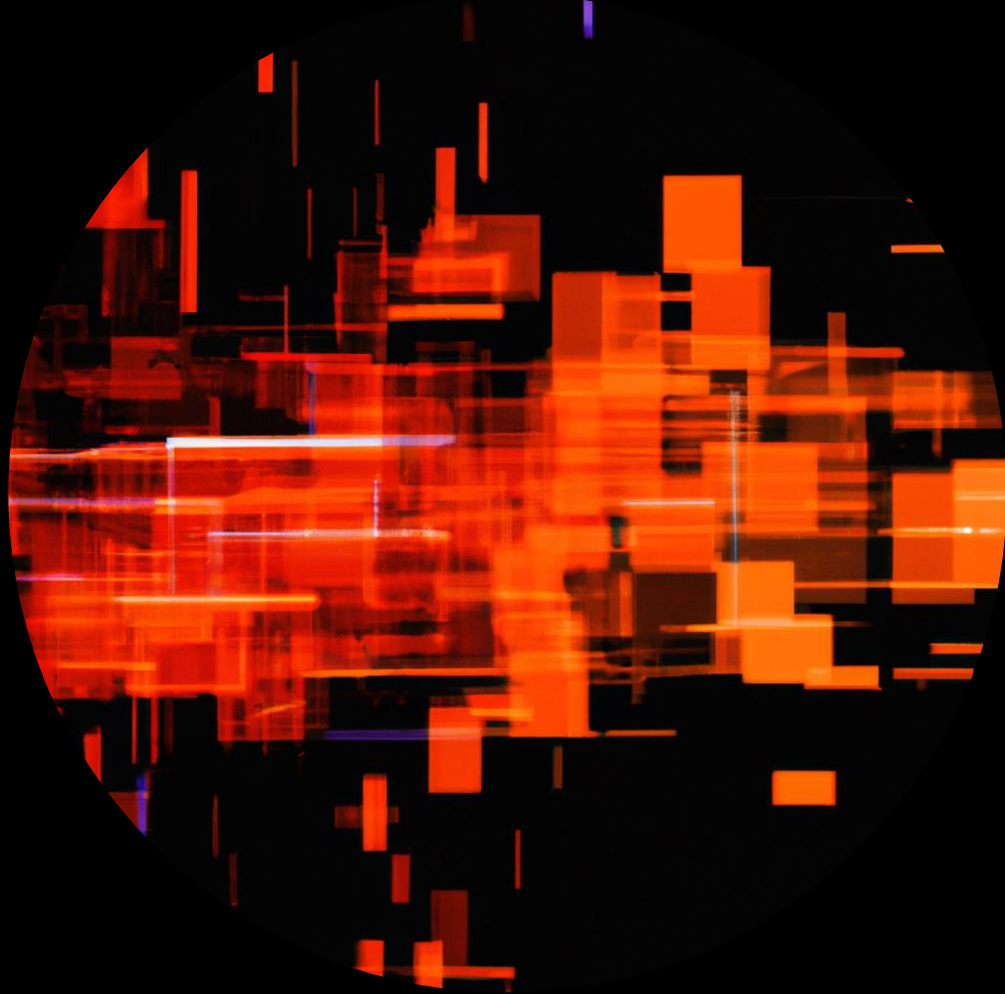


Example well-formed question for machine learning

What is the optimal calendar timing and keywords for advertising to potential students?

But... just adding AI+data  
does not solve **most** business  
problems.





- Collection of the *right data*
- Collection of *enough data*
- Using the *right* machine-learning model: simple before complex
- Evaluating if a model is **really necessary** for decision making



# AI Ethics



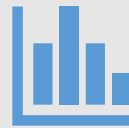
Compliance



Privacy, and  
Security



Explainability and  
Accountability



Data Provenance  
and Transparency



Reliability and  
Safety



Hype

# Australia's AI Ethics Principle's (Voluntary)

Human, societal and environmental wellbeing

Human-centred values

Fairness

Privacy protection and security

Reliability and safety

Transparency and explainability

Contestability

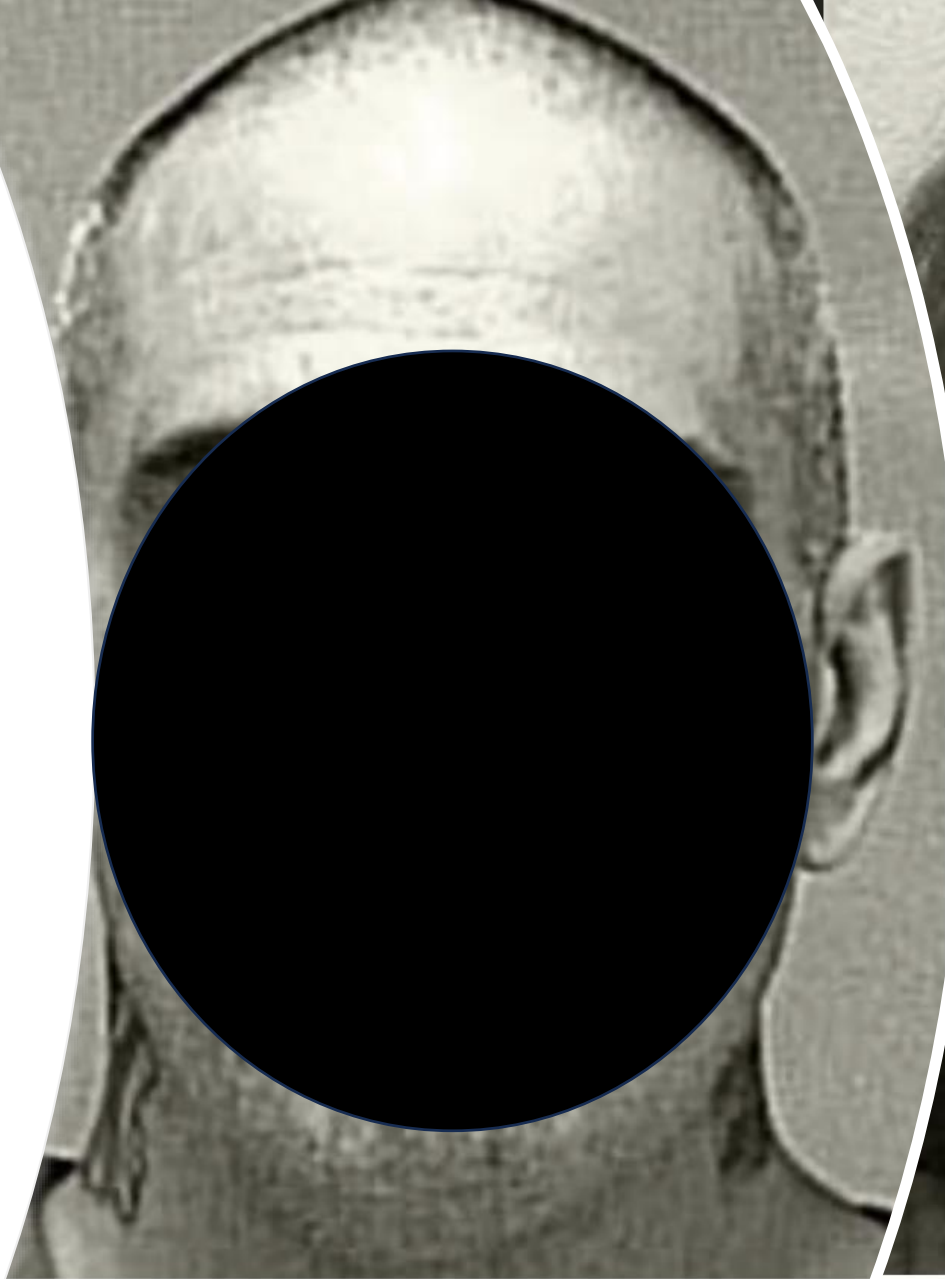
Accountability



# Bias

- Vernon Prater
  - Prior: 2 robberies, 1 attempt
  - Subsequent: 1 grant theft
- Brisha Borden
  - Prior 4 juvenile misdemeanors
  - Subsequent: None

Source: Machine Bias, Angwin et al.



VERNON PRATER



BRISHA BORDEN

# Bias

- Vernon Prater
  - Prior: 2 robberies, 1 attempt
  - Subsequent: 1 grant theft
- Brisha Borden
  - Prior 4 juvenile misdemeanors
  - Subsequent: None



LOW RISK

3

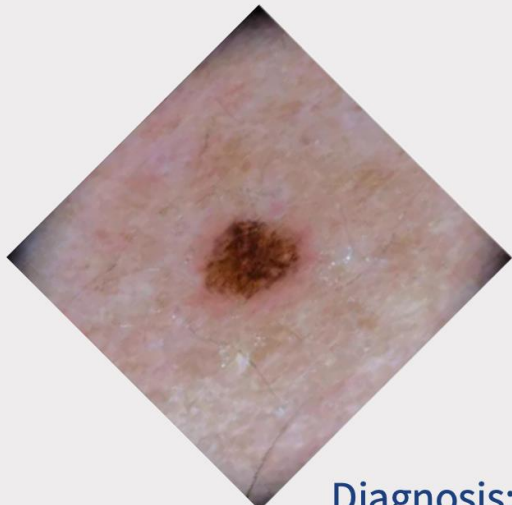
HIGH RISK

8

# RELIABILITY AND SAFETY



Diagnosis: Benign



Diagnosis: Malignant

277.7 Metabolic syndrome  
429.9 Heart disease, unspecified  
278.0 Obesity, unspecified

Reimbursement: Denied



401.0 Benign essential hypertension  
272.0 Hypercholesterolemia  
272.2 Hyperglyceridima  
429.9 Heart disease, unspecified  
278.0 Obesity, unspecified

Reimbursement: Approved

Source: It's disturbingly easy to trick AI into doing something deadly, Samuel. [www.vox.com/future-perfect/2019/4/8/18297410/ai-tesla-self-driving-cars-adversarial-machine-learning](http://www.vox.com/future-perfect/2019/4/8/18297410/ai-tesla-self-driving-cars-adversarial-machine-learning)

**"We tend to overestimate  
the effect of a technology  
in the short run and  
underestimate the effect in  
the long run." – Roy Amara  
(futurist)**