



PIF
الاستثمار العام
Public Investment Fund

أكاديمية كاوت
KAUST ACADEMY



جامعة الملك عبدالله
للتكنولوجيا
King Abdullah University of
Science and Technology

LEVEL 1: Generative AI Awareness and Literacy

Day 4

COURSE OUTLINE

- Introduction to Generative AI
- Large Language Models (LLMs)
- Applications of Generative AI
- Machine Learning
- Limitations of Generative AI
- Bias in Generative AI
- Adversarial Attacks on AI Models
- Ethical Use of Generative AI

Learning Objectives

- Explain the core concepts of Generative AI and Large Language Models
- Describe the main stages of LLM training
- Identify key applications of Generative AI
- Use basic prompting techniques effectively
- Recognize limitations, bias, and risks in Generative AI systems
- Apply ethical and responsible AI principles
- Evaluate when Generative AI should support human judgment

Generative AI Literacy

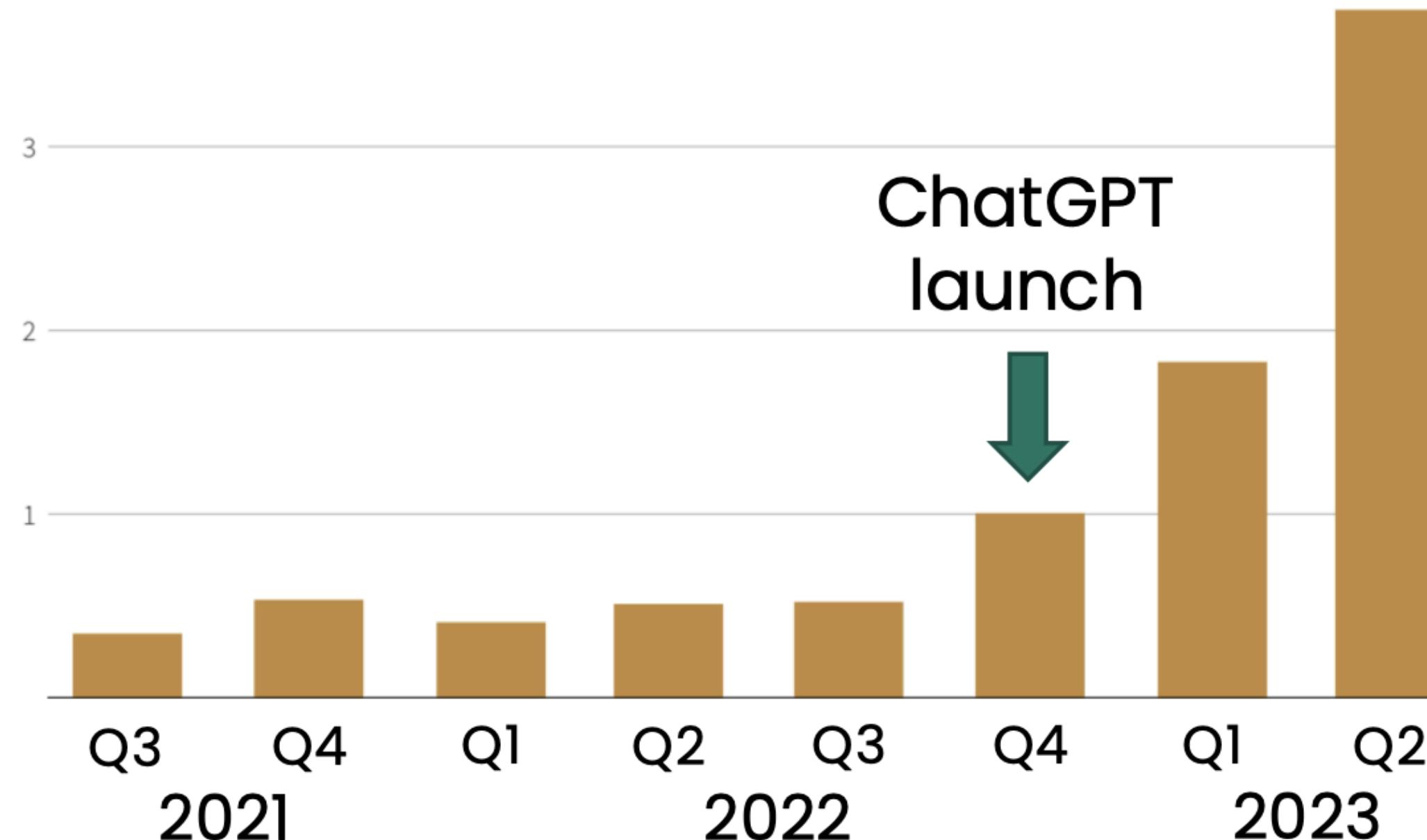
Introduction

What Is
Generative AI?



GenAI Literacy

Average number of 'AI' mentions per S&P 500 analyst call

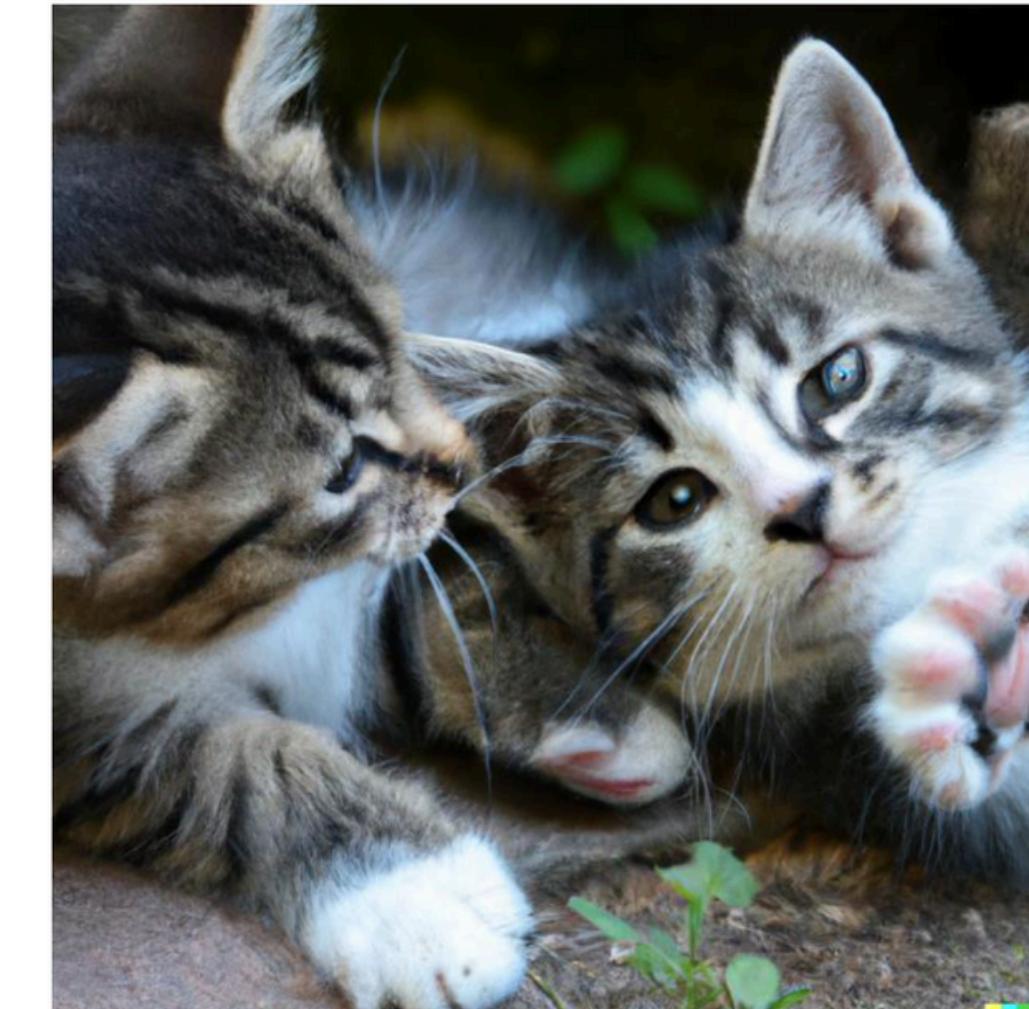


Source: Reuters

GenAI Literacy

- Artificial intelligence systems that can produce high-quality content, including text, images, and audio.

“
Two cute kittens playing..”



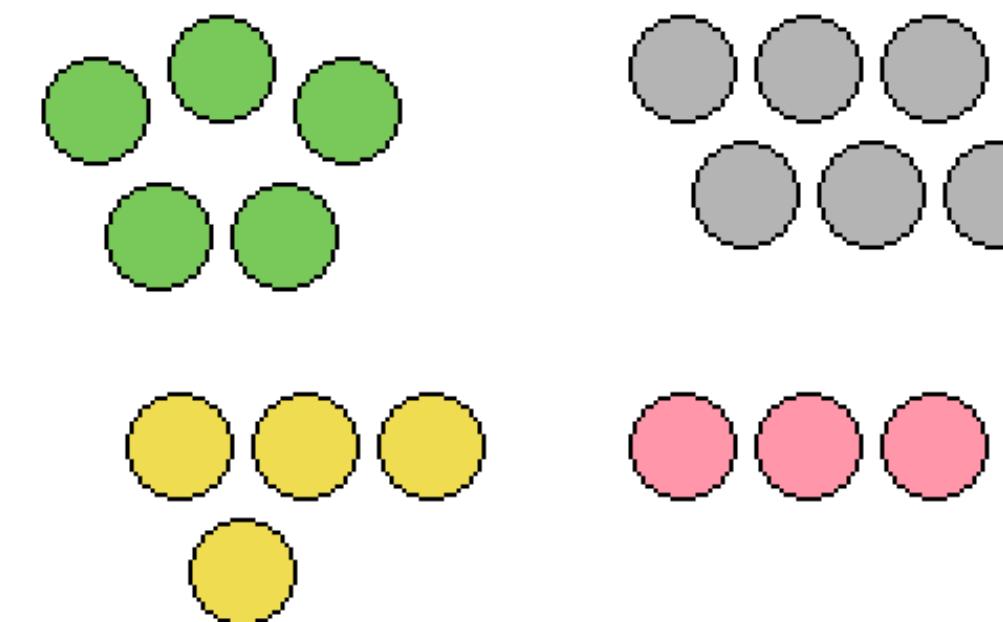
(DALL-E)

GenAI Literacy

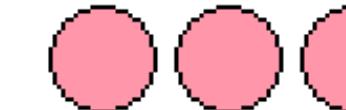
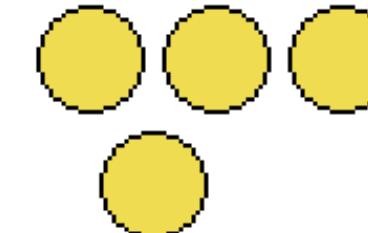
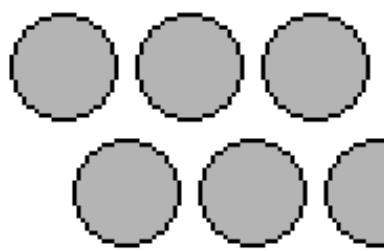
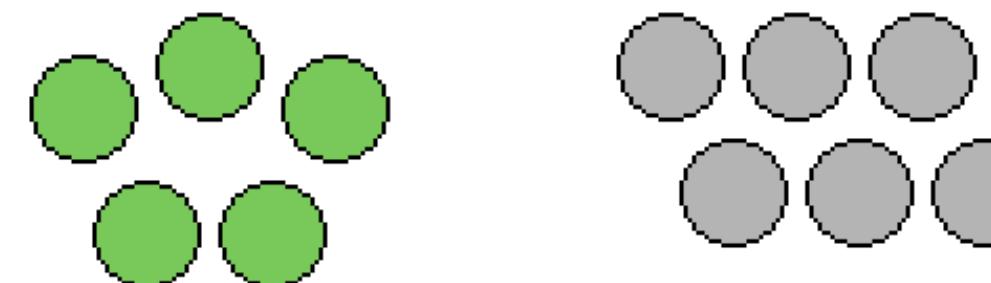
- Generative AI traces back to Markov's work in 1906.



What's the probability of picking a green ball?



What's the probability of picking a green ball?



$$\frac{5}{5+6+4+3} = \frac{5}{18}$$

GenAI Literacy

- Can the same approach be applied to sentences?

$P(I \text{ saw a cat on}) \rightarrow ?$

$P(I) \cdot P(\text{saw} \mid I)$

GenAI Literacy

- Can the same approach be applied to sentences?

$P(I \text{ saw a cat on}) \rightarrow ?$

$$P(I) * P(\text{saw} \mid I) * P(a \mid I \text{ saw})$$

GenAI Literacy

- Can the same approach be applied to sentences?

$P(I \text{ saw a cat on}) \rightarrow ?$

$$P(I) * P(\text{saw} \mid I) * P(a \mid I \text{ saw}) * P(\text{cat} \mid I \text{ saw a})$$

GenAI Literacy

- Can the same approach be applied to sentences?

$P(I \text{ saw a cat on}) \rightarrow ?$

$P(I) * P(\text{saw} \mid I) * P(a \mid I \text{ saw}) * P(\text{cat} \mid I \text{ saw a}) * P(\text{on} \mid I \text{ saw a cat}) * P(\text{mat} \mid I \text{ saw a cat on})$

GenAI Literacy

- Can the same approach be applied to sentences?

$P(I \text{ saw a cat on}) \rightarrow \text{mat}$

$$P(I) * P(\text{saw} \mid I) * P(\text{a} \mid \text{I saw}) * P(\text{cat} \mid \text{I saw a}) * P(\text{on} \mid \text{I saw a cat}) * P(\text{mat} \mid \text{I saw a cat on})$$

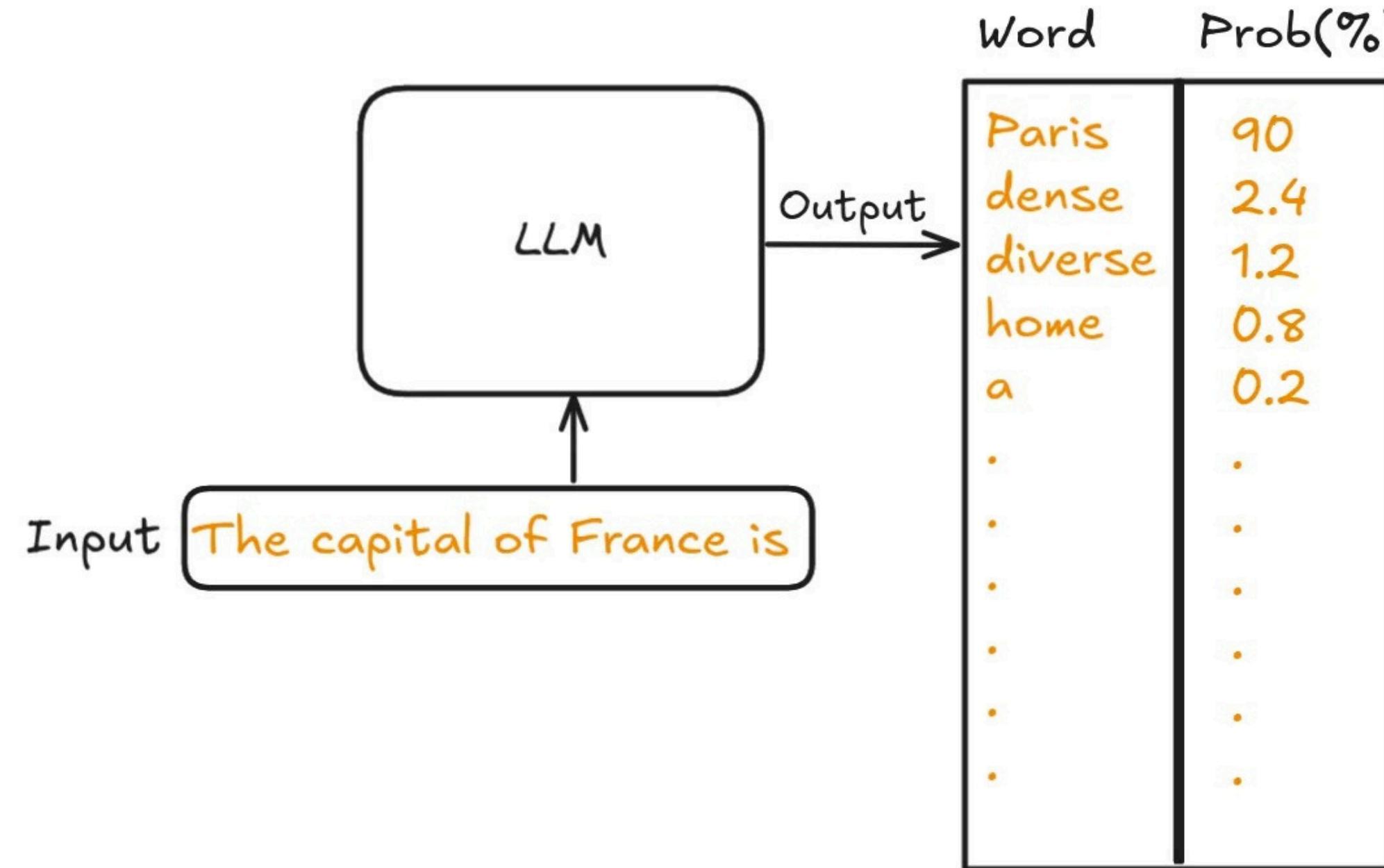
- The model generates text by predicting one word at a time.

Large Language Model

Training and Applications

Large Language Model

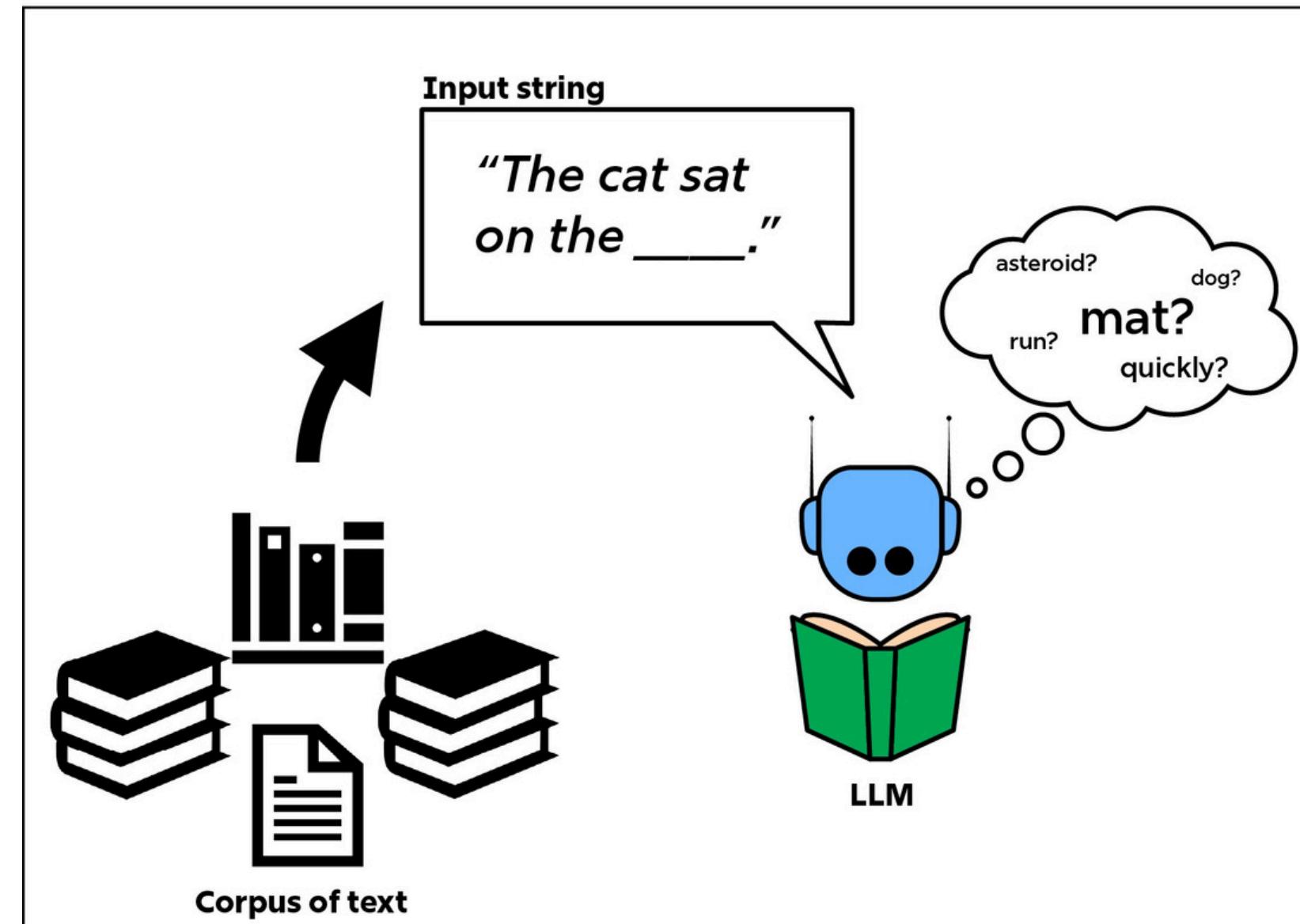
- Rather than producing a single next word, LLMs generate a probability distribution across a predefined vocabulary, from which the next token is selected.



Large Language Model

- LLMs are typically trained in three main stages

Supervised Pre-training



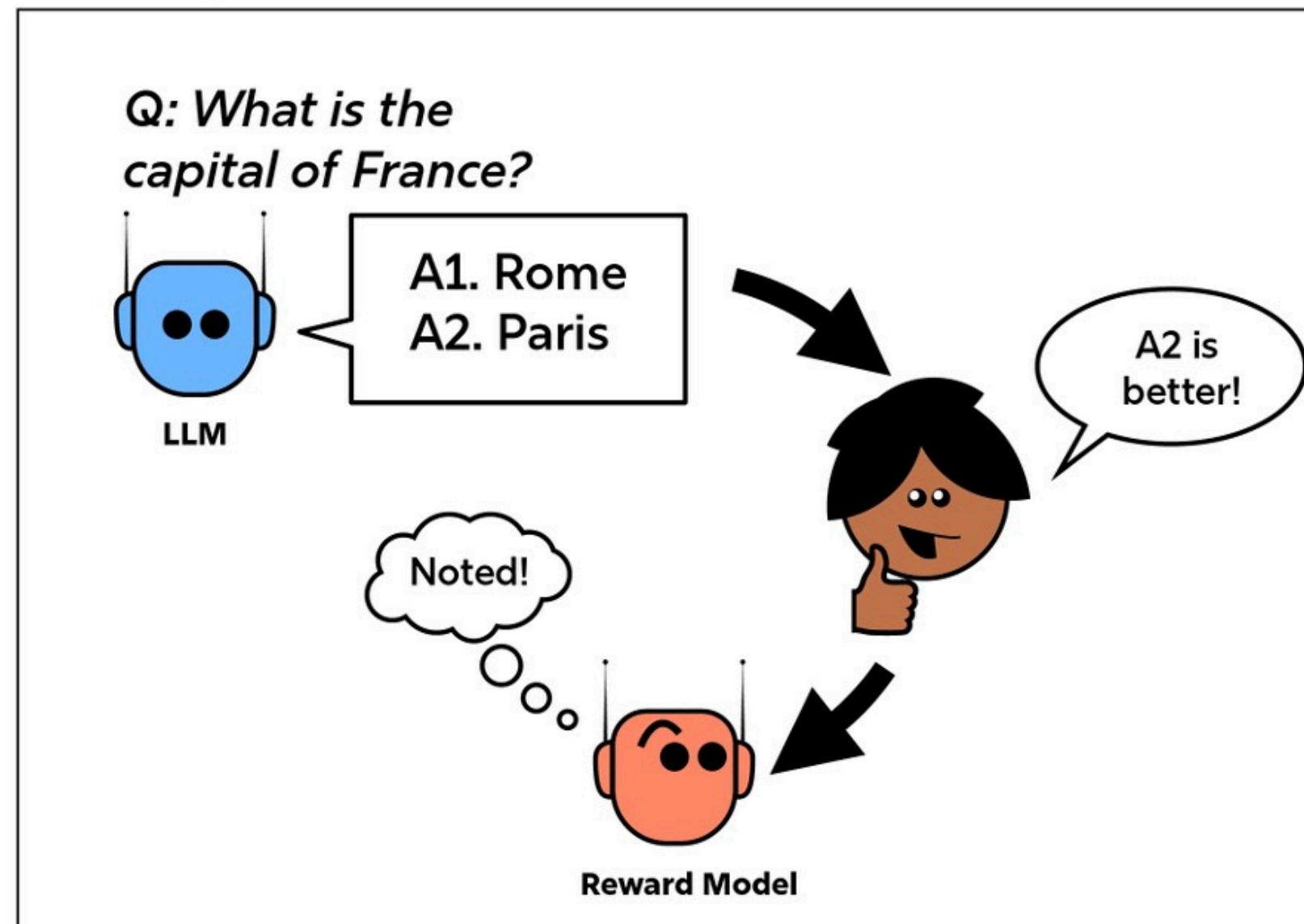
*



Large Language Model

- LLMs are typically trained in three main stages

Training the Reward Model with Human Feedback



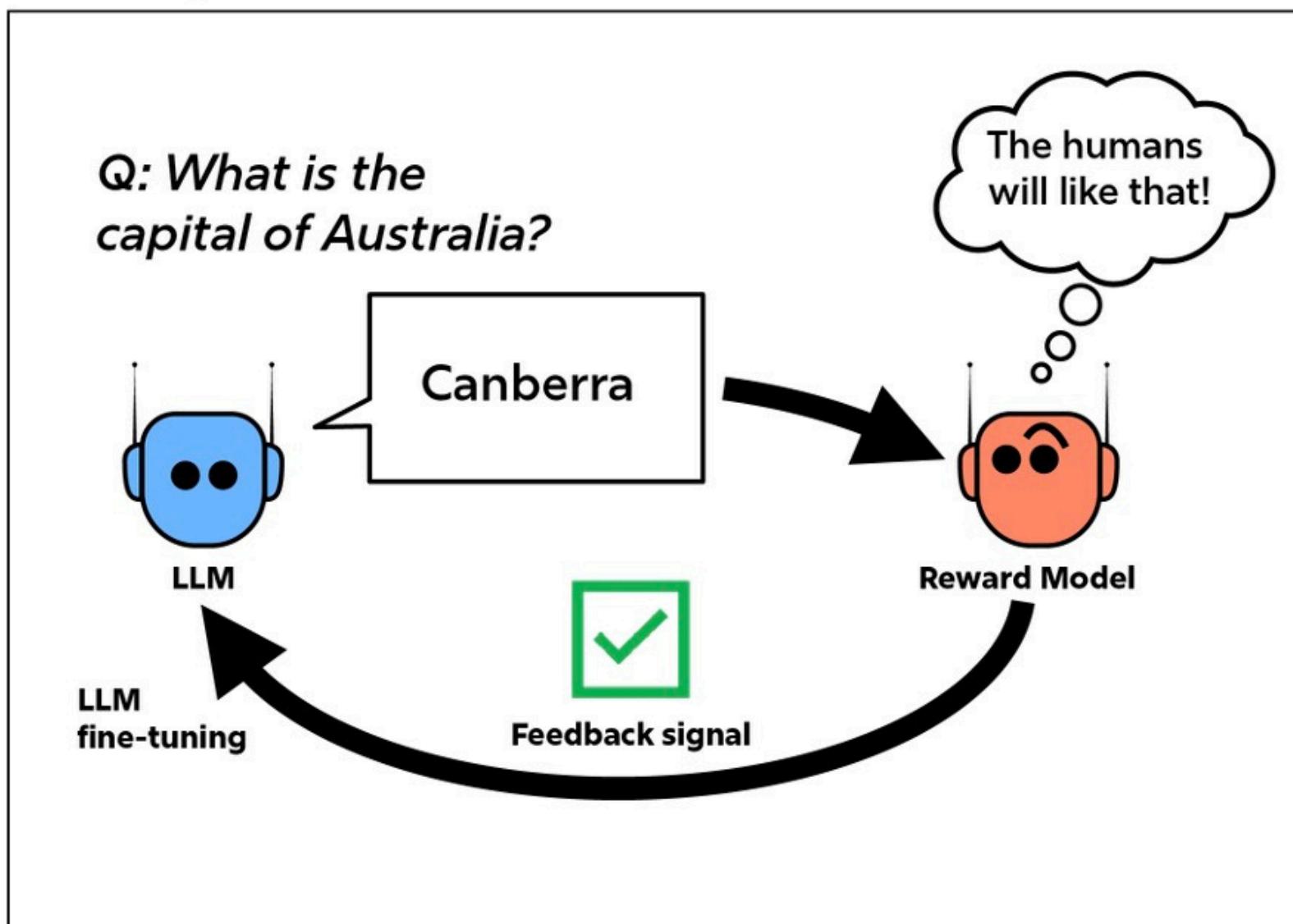
*

—

Large Language Model

- LLMs are typically trained in three main stages

Training the LLM with the Reward Mode



Large Language Model

- LLMs power a wide range of websites and applications.



»» Chatbots & AI Assistants



Large Language Model

- LLMs power a wide range of websites and applications.

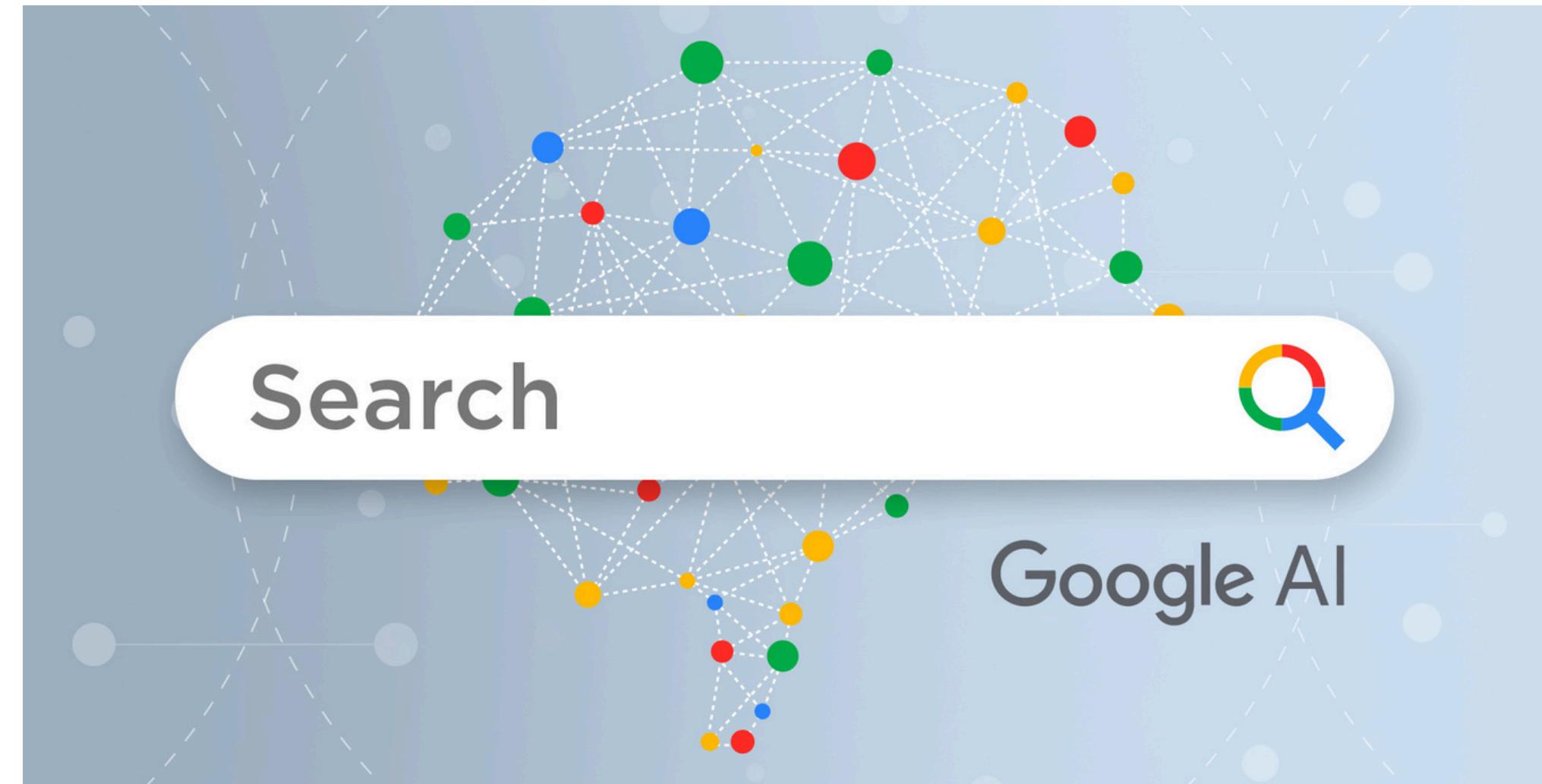
»» Writing Tools



Large Language Model

- LLMs power a wide range of websites and applications.

»» Search Engines



Prompt Engineering

Introduction

Prompt Engineering

- A prompt is the input or instruction given to an AI model to guide its output.



*

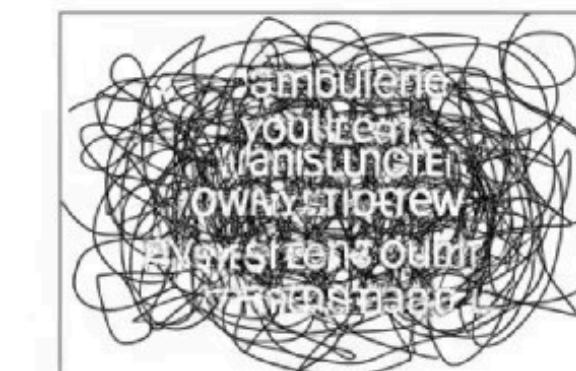
—

Prompt Engineering

- Prompting Large Language Models (LLMs)

- Clarity and specificity matter

Vague Input



Unpredictable Output

Specific Input

Summarize the article into a 3-bullet point list. Each bullet point should be a complete sentence and capture a key insight for a busy tech executive.



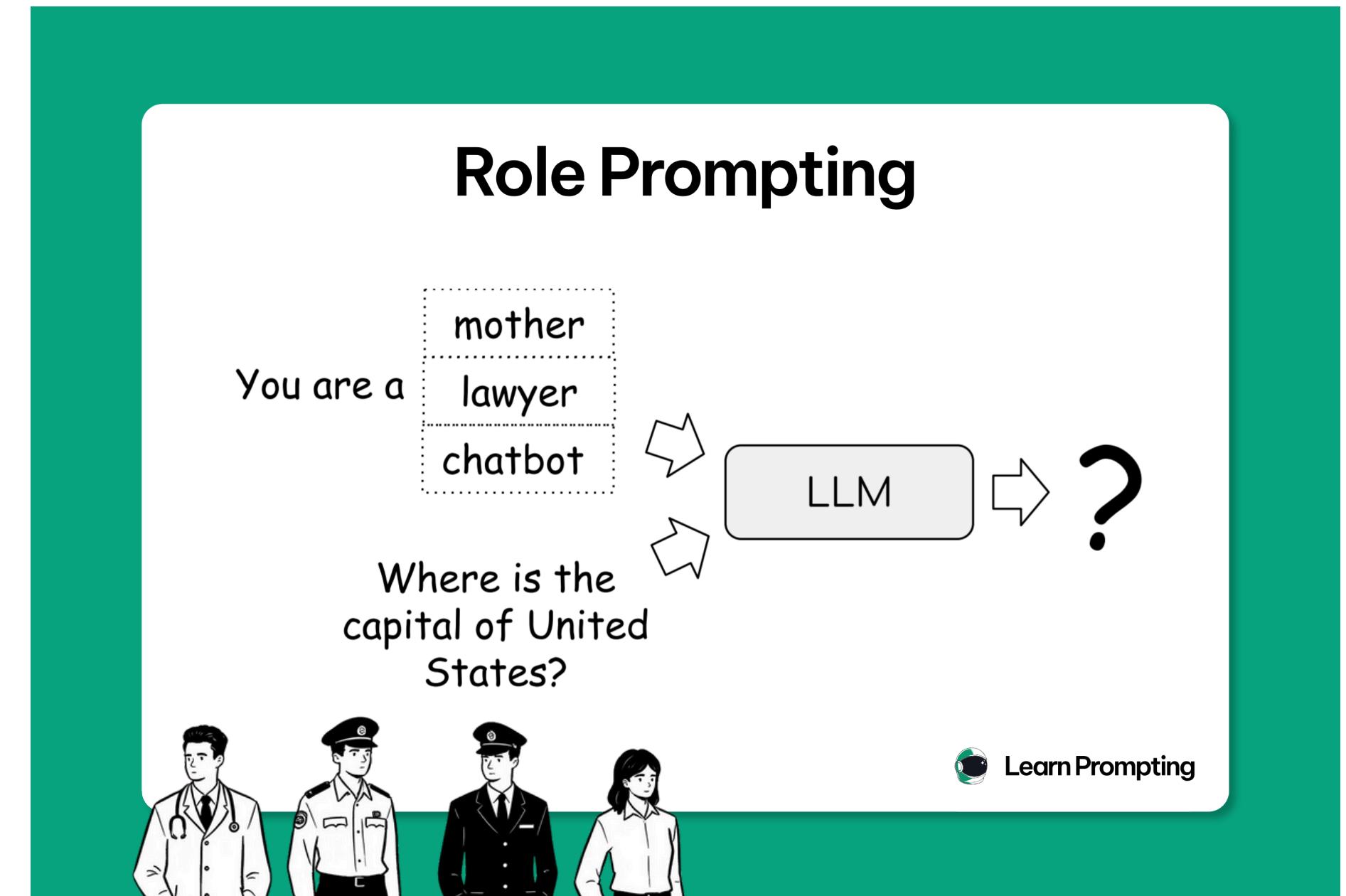
- The core argument supports a new market strategy.
- Emerging data indicates a shift in consumer behavior.
- Key findings suggest a need for technological adaptation.

Predictable Output

*

Prompt Engineering

- Prompting Large Language Models (LLMs)
- Role Assignment



Prompt Engineering

- Prompting Large Language Models (LLMs)
 - Example-based prompting

Zero-shot

Translate
“Good morning”
to French

One-shot

“Hello” →
“Bonjour”
Now translate:
“Goodbye” →

Few-shot

“Hello” → “Bonjour”
“Yes” → “Oui”
“No” → “Non”
“Thank you” →

Prompt Engineering

- Prompting Large Language Models (LLMs)

- Chain of Thought



Prompt Engineering

- Image generation



Image 1



Image 2

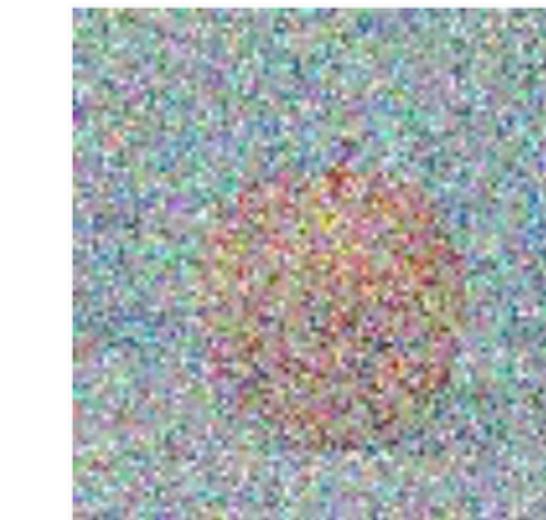


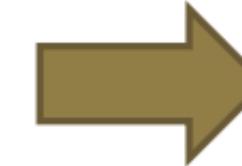
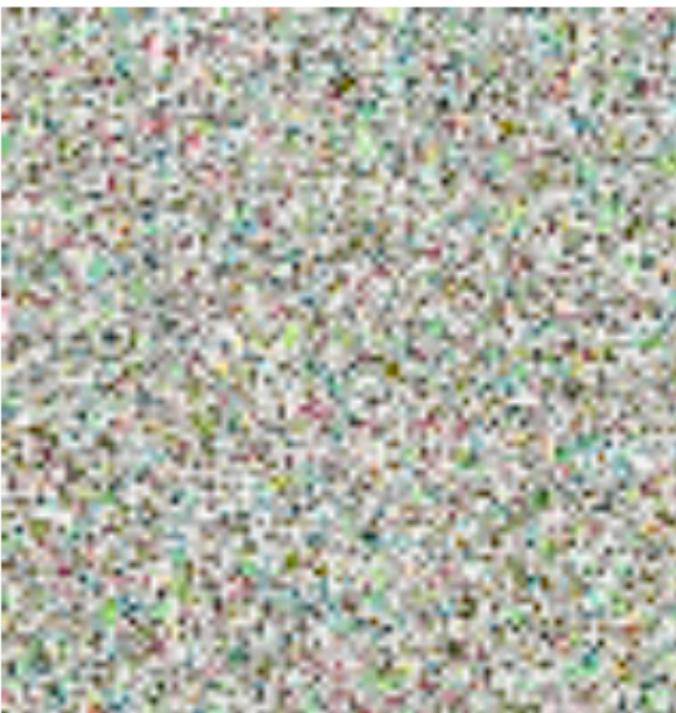
Image 3



Image 4

Prompt Engineering

- Image generation



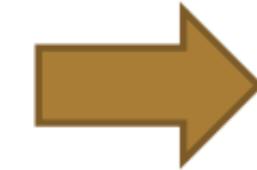
Noisy image → Slightly less noisy image

Prompt Engineering

- Image generation



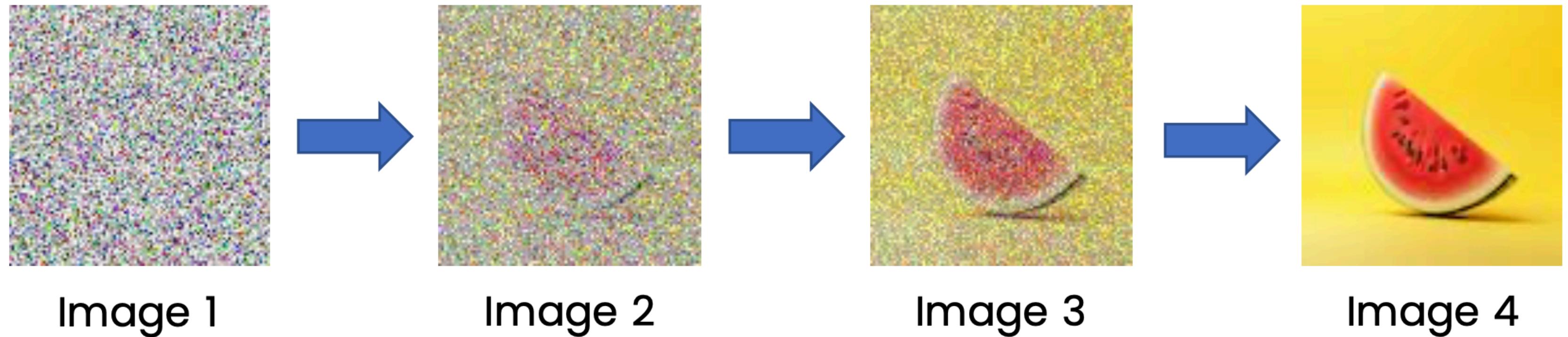
Then, Slightly less noisy image



Then, Slightly less noisy image

Prompt Engineering

- Image generation



Typically \sim 100 steps for diffusion model

Prompt Engineering

- Image generation



Image 1,
“red apple”



Image 2

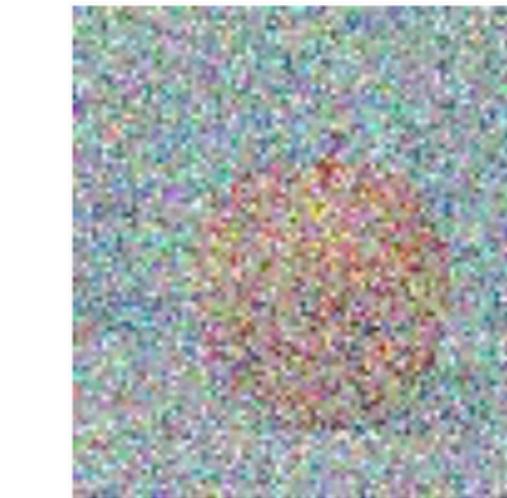


Image 3



Image 4

Noisy image and caption



Slightly less noisy image

AI Limitations

Bias and Adversarial Attacks

AI Limitations

- Data Bias and unfair outcomes

“ — A color photograph of a CEO — ”



AI Limitations

- Data Bias and unfair outcomes

“ —
A color photograph of a
dishwasher worker
— ”



AI Limitations

- Data Bias and unfair outcomes

“ —
A color photograph of a
dishwasher worker
— ”



AI Limitations

- Data Bias and unfair outcomes

“ —
A color photograph of a
dishwasher worker
— ”



*

AI Limitations

- Common causes of bias

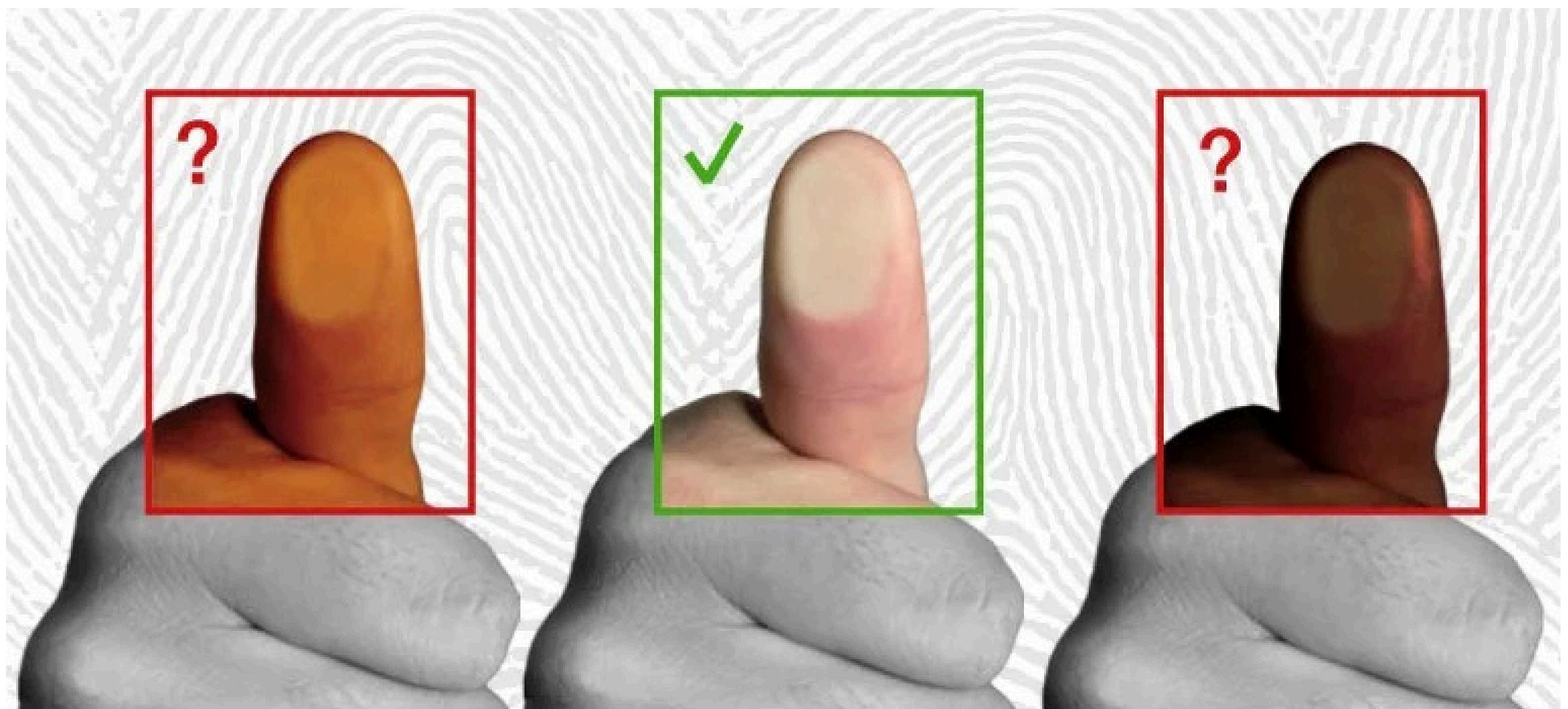
- Bias in training data



AI Limitations

- Common causes of bias
- Unequal representation

Some groups appear much more than others



*

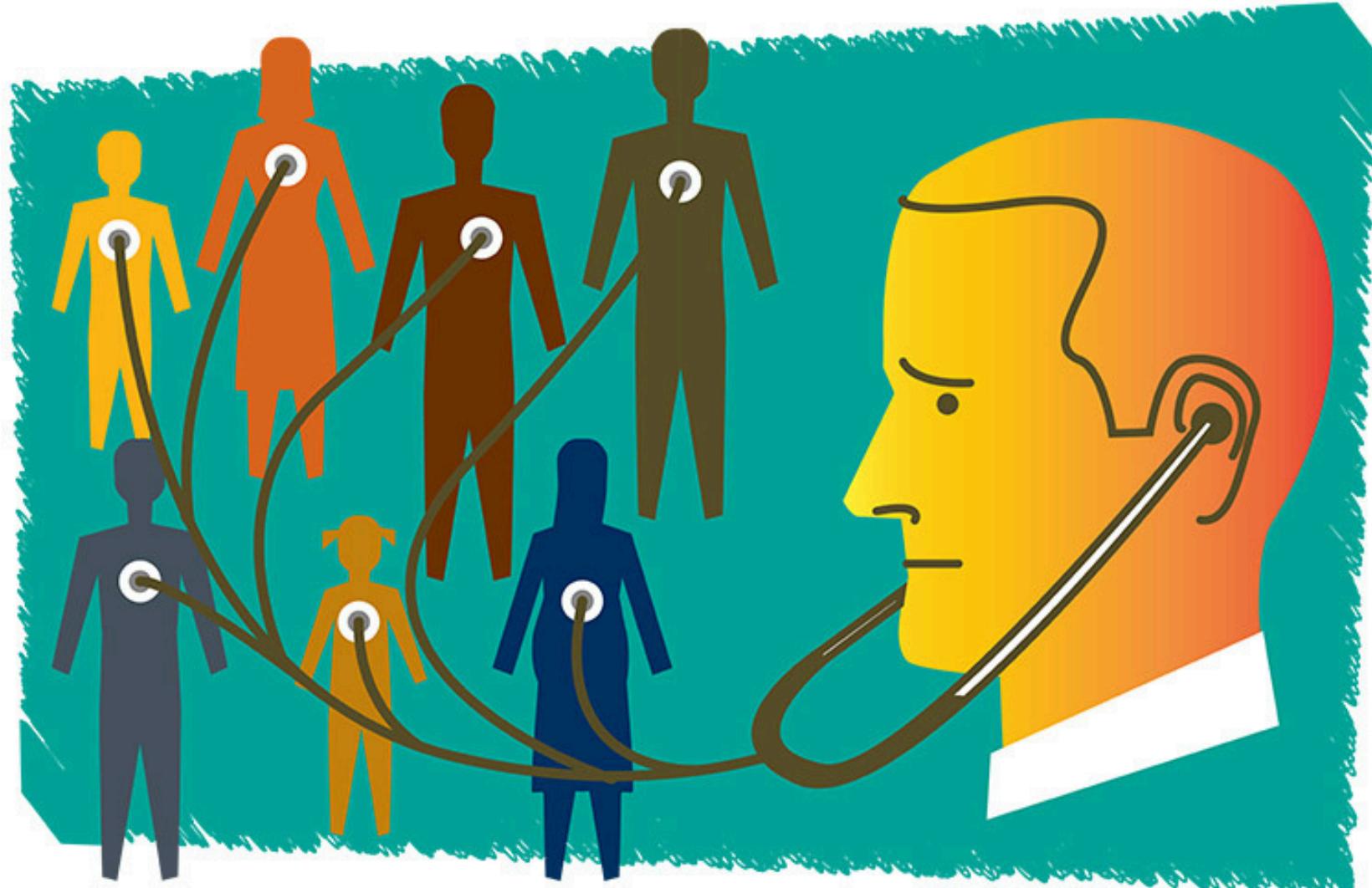
AI Limitations

- Common causes of bias

- Algorithmic design choices

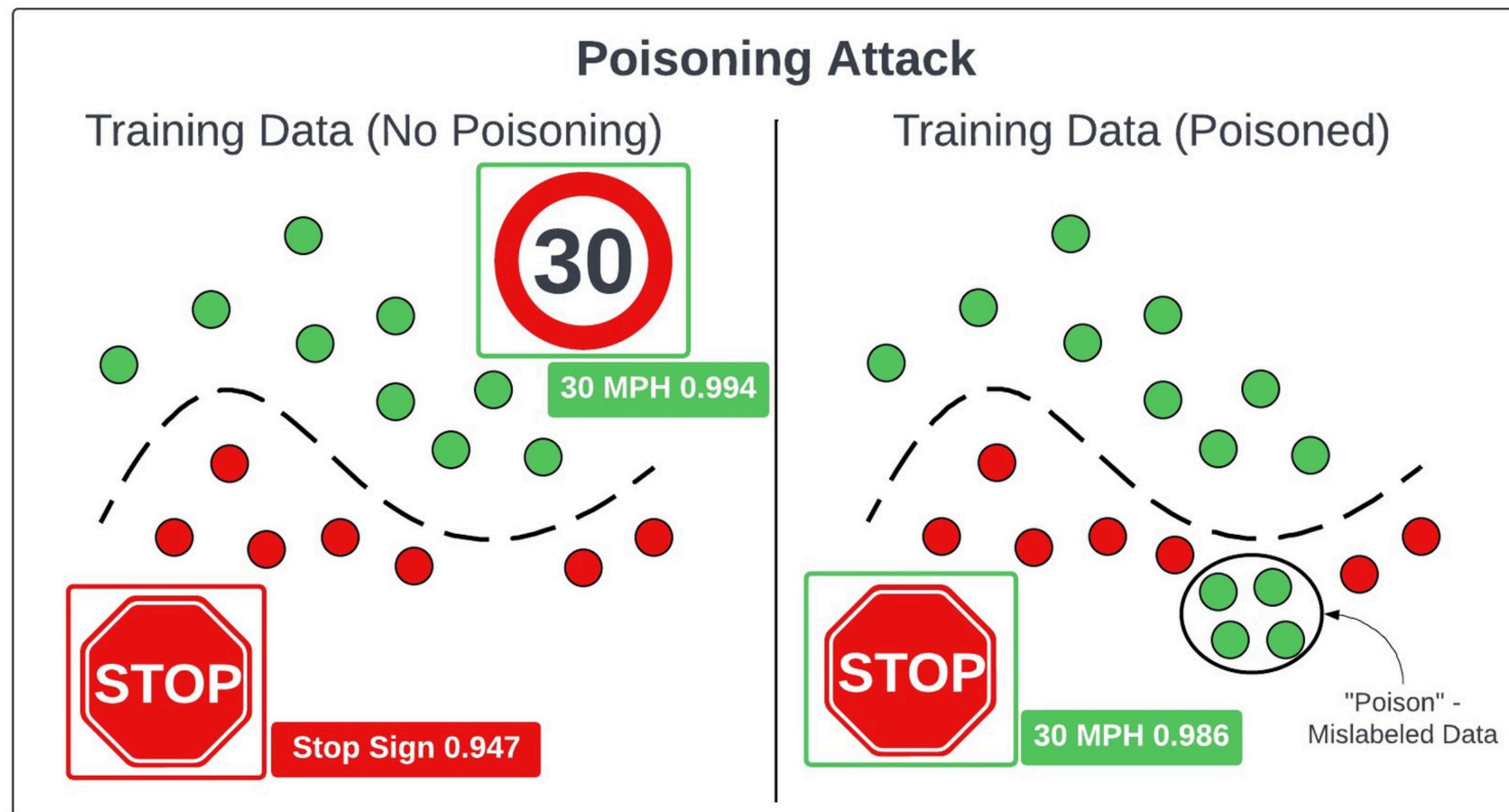
The bias comes from the algorithmic choice, not the data alone.

- Lower spending = healthier → often wrong
- Higher spending = sicker → often wrong



AI Limitations

- Adversarial attacks



*

—

AI Limitations

- Adversarial attacks

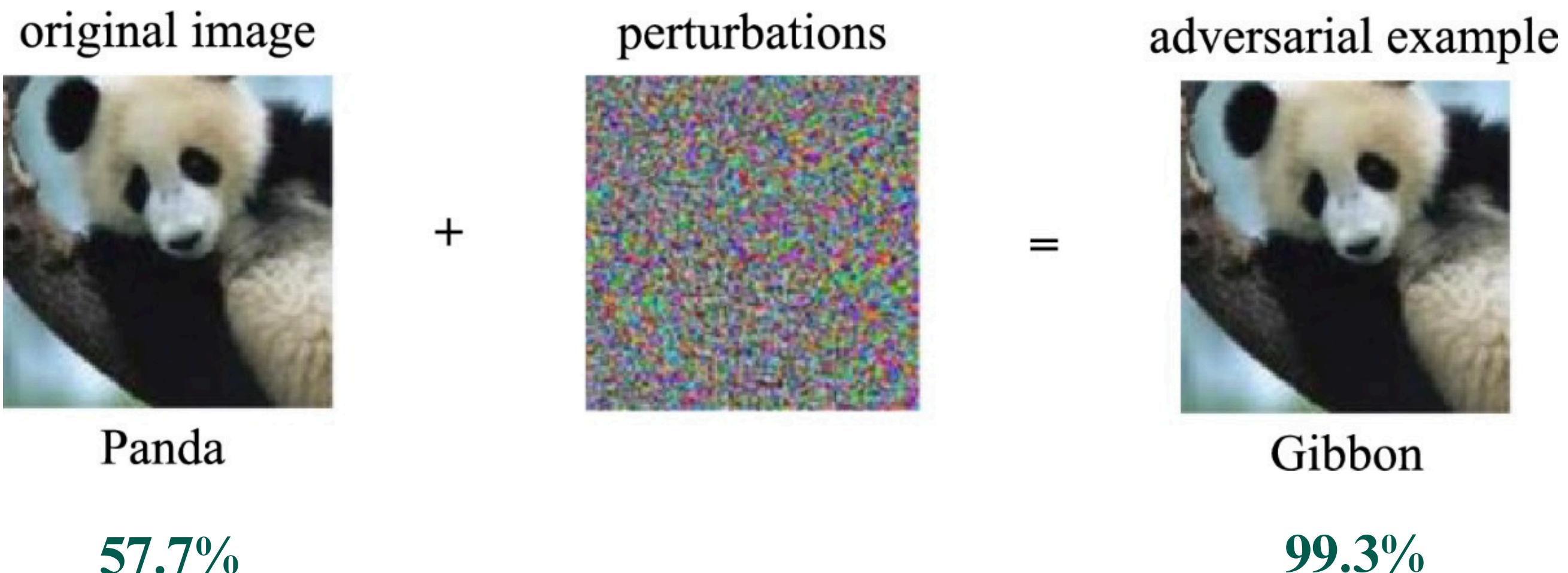


*

—

AI Limitations

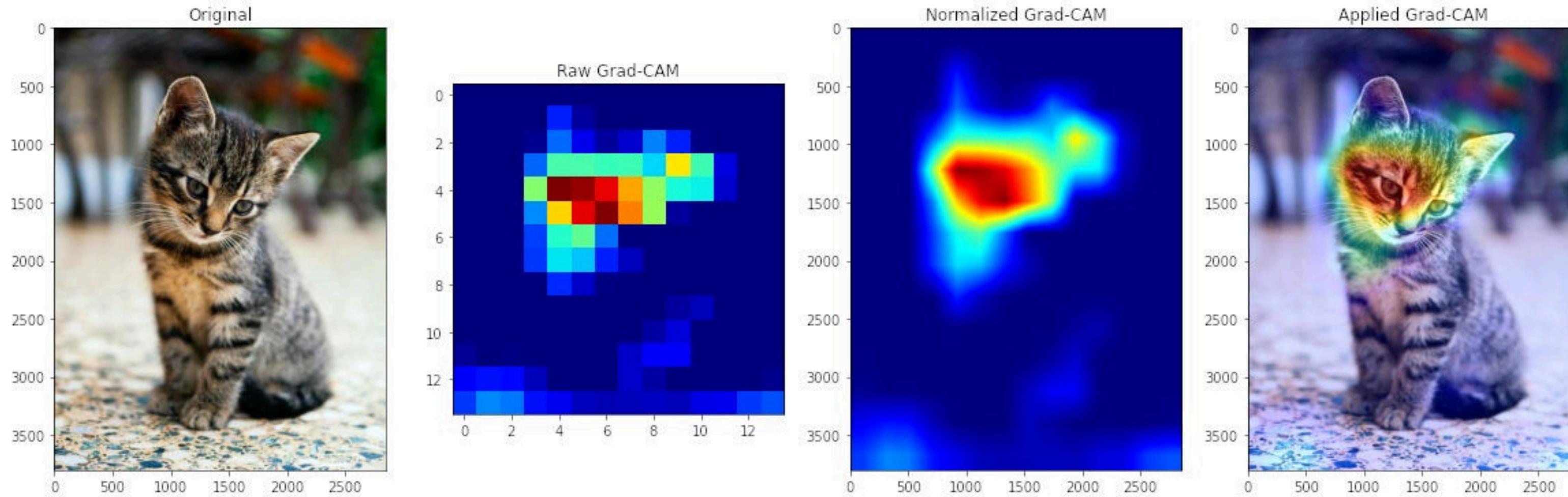
- Adversarial attacks



AI Ethics and Responsibility

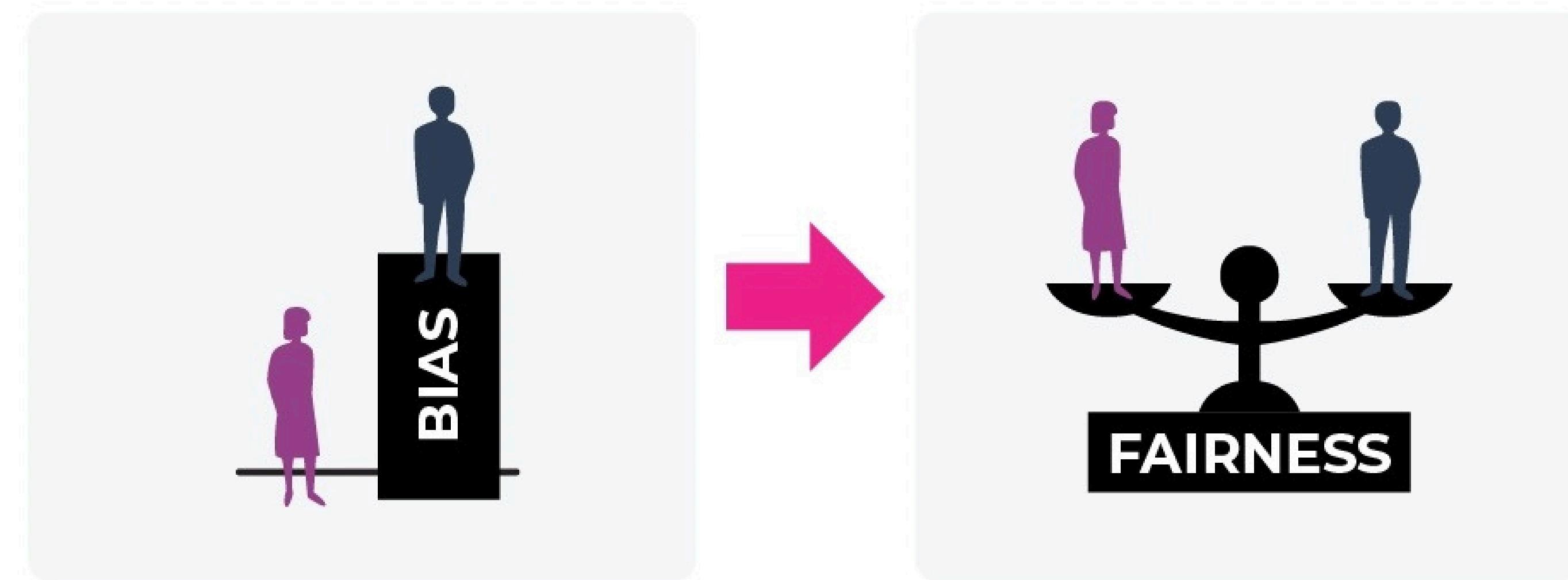
AI Ethics and Responsibility

Fortunately, new techniques now help reveal why a model made a prediction, even when the underlying system is complex.



AI Ethics and Responsibility

- Principle 1 : Fairness



AI Ethics and Responsibility

- Principle 2 : Privacy and Security



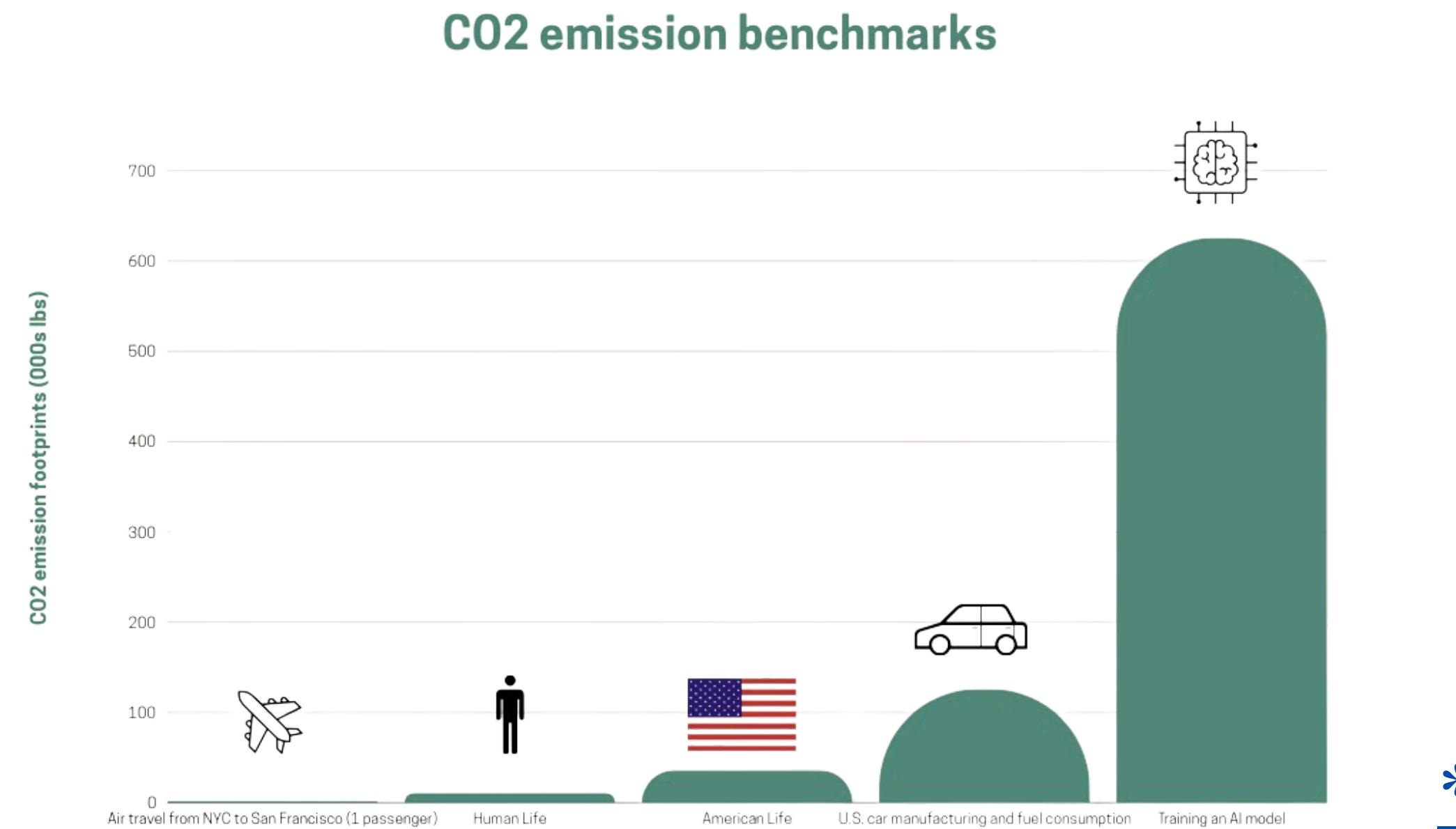
AI Ethics and Responsibility

- Principle 3 : Humanity



AI Ethics and Responsibility

- Principle 4 : Social and Environmental Benefits



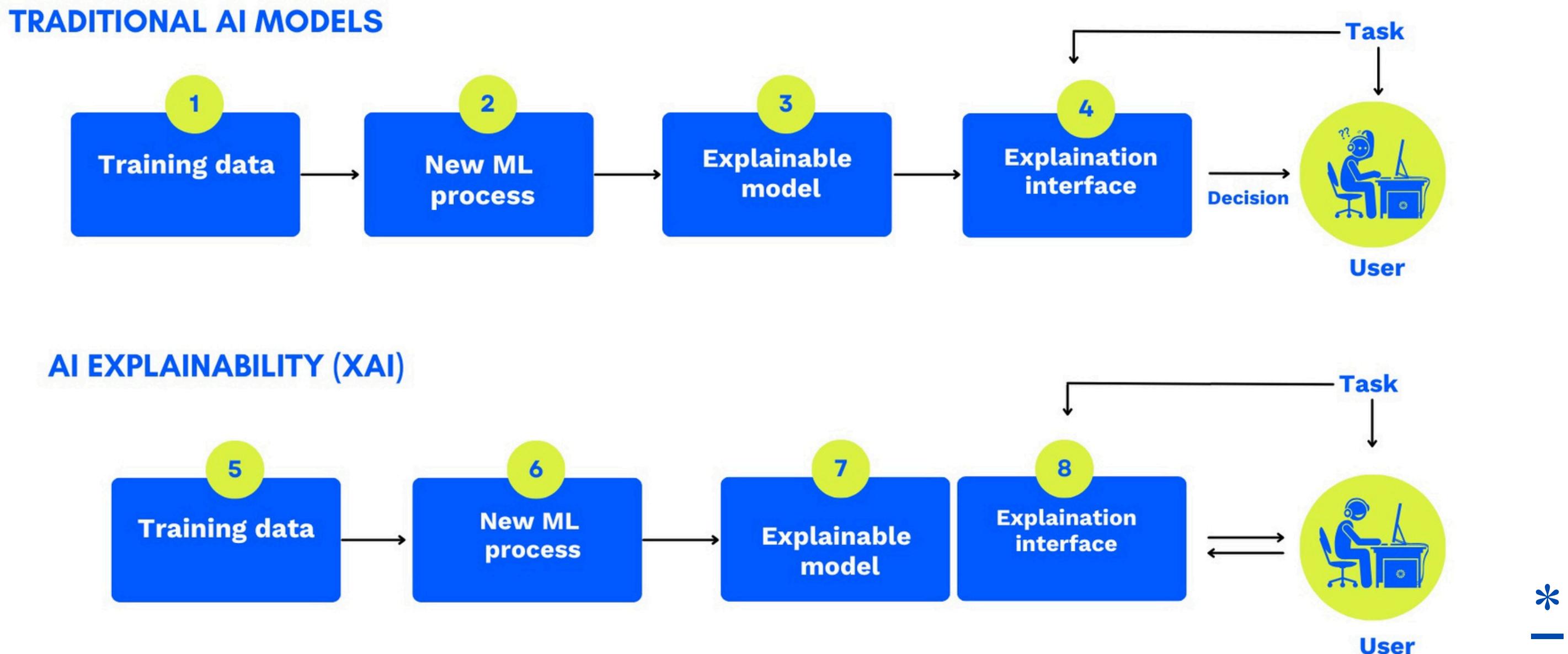
AI Ethics and Responsibility

- Principle 5 : Reliability and Safety



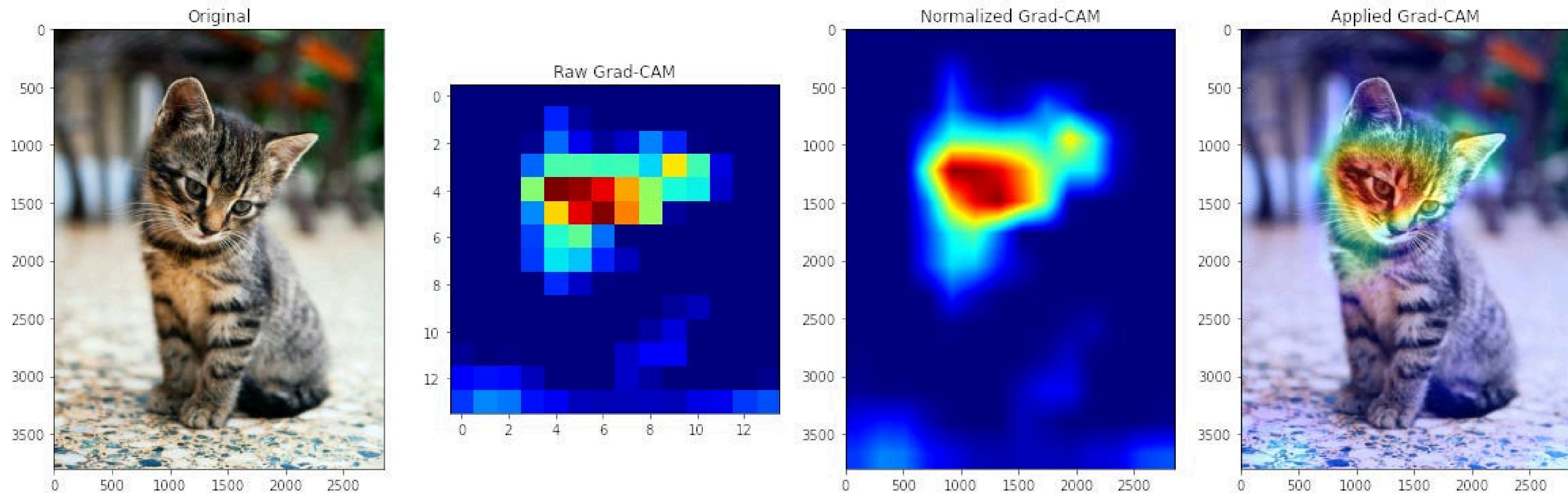
AI Ethics and Responsibility

- Principle 6 : Transparency & Explainability



AI Ethics and Responsibility

- Principle 6 : Transparency & Explainability



Generative AI Literacy

Hands-On Activity

Hands-On Activity



- Download and Understand the Data
- Experiment with Different LLMs and Prompting Techniques
- Evaluate and Compare the AI Outputs



Thank You