

Model design and data collection

GENERATIVE AI CONCEPTS



Daniel Tedesco
Data Lead, Google

Know how to fill the tank



¹ GM Fairfax Assembly Plant

Developing a model

Model Development Steps

1. Research and design
2. Training data collection
3. Model training
4. Model evaluation

Stable Diffusion's research and development

Example output from Stable Diffusion



Stable Diffusion's R&D

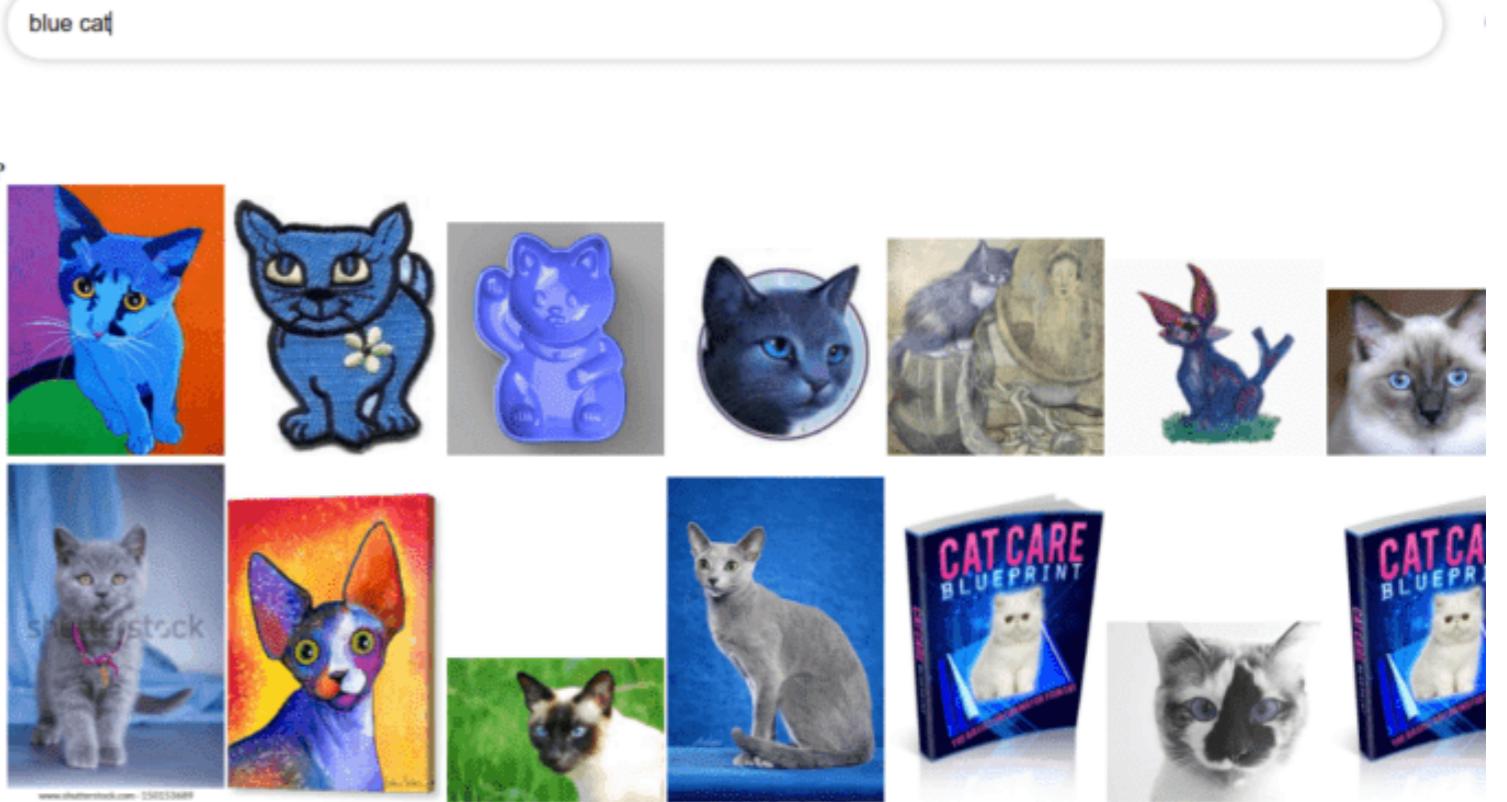
- Purpose: Decide on image generation
- Architecture: Settle on diffusion model
- Resources: 256 GPUs, 150k hours, \$600k

¹ Stability AI, Emad Mostaque Twitter post

Data collection: not your typical ML model

Training data preparation

- Massive amounts required
- Diverse, context-rich data
- Requires preprocessing



¹ Laion blog

Data collection: privacy and security are critical

Training data preparation

Personally Identifiable Information (PII)

- Anonymize or aggregate
- Store in secure location with controlled access



Let's practice!

GENERATIVE AI CONCEPTS

Model training

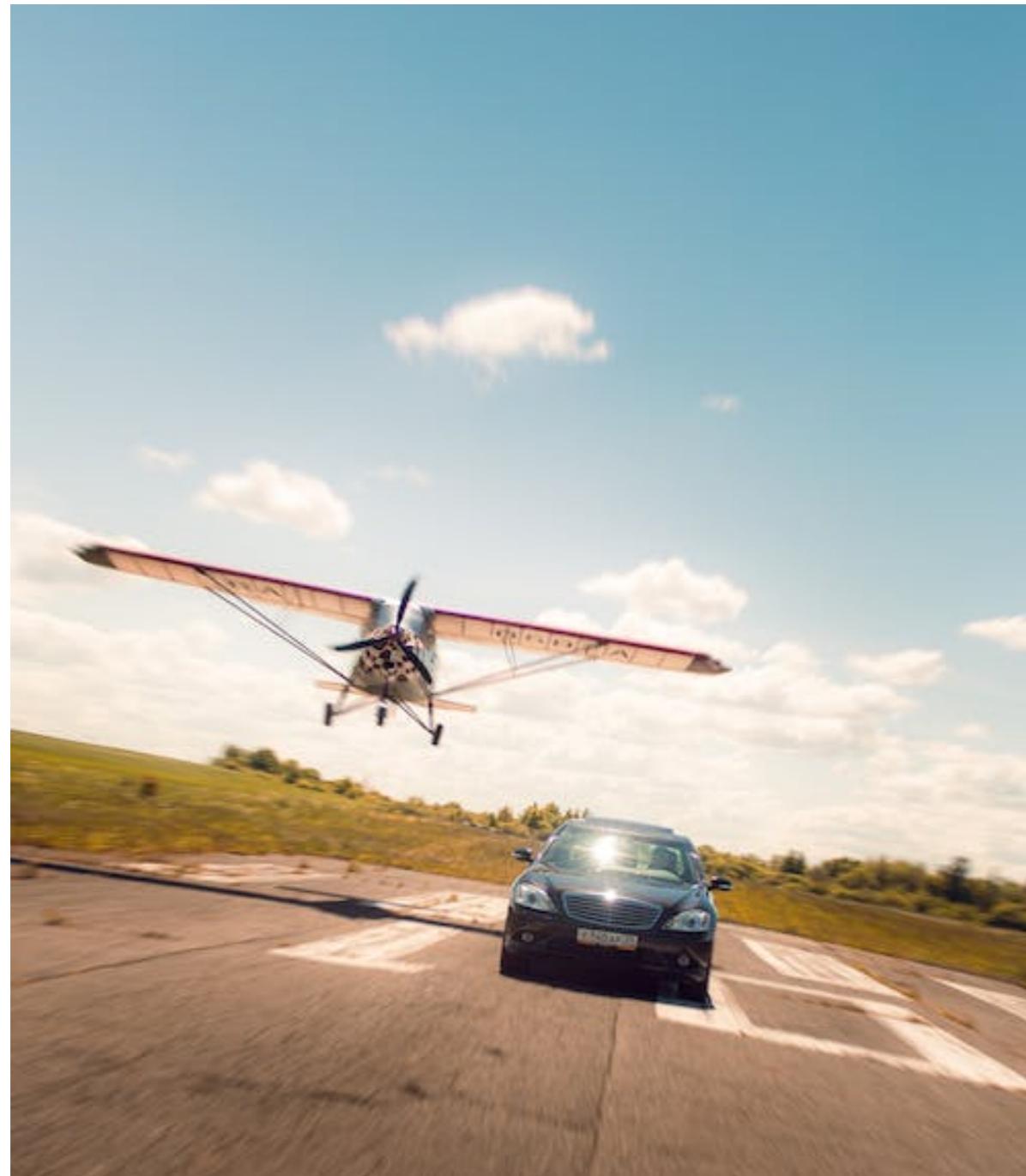
GENERATIVE AI CONCEPTS



Daniel Tedesco

Data Lead, Google

Pick your mode of train-sportation



1. Hardware
2. Time
 - Dataset size
 - Model complexity
 - Rounds of training
3. Cost

Graduate to advanced techniques

- Foundation generative AI model is a first step
- Advanced techniques specialize them for specific contexts

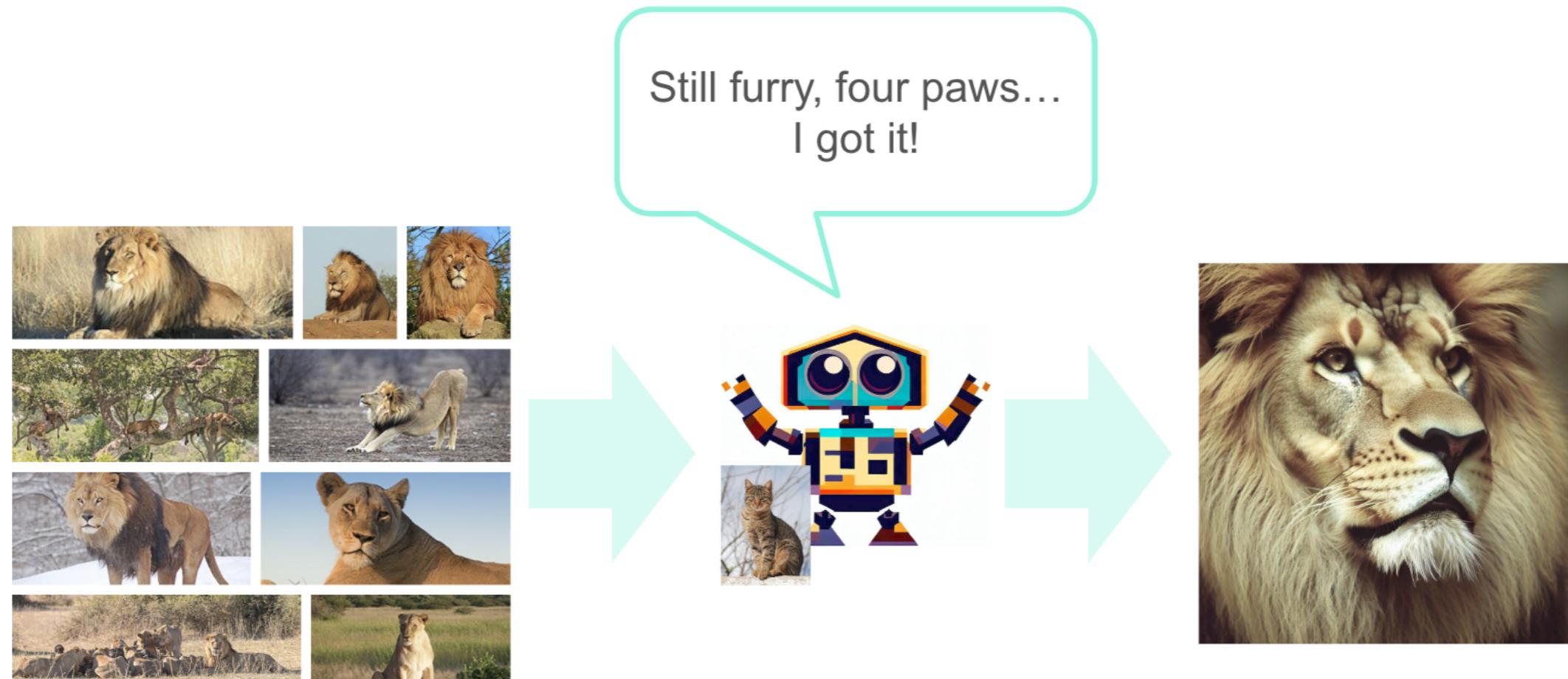
Key advanced techniques:

1. Transfer learning and fine-tuning
2. RLHF
3. Custom embeddings



From cats to lions

- Transfer learning transfers knowledge from one task to another
- Fine-tuning is a type of transfer learning for a small dataset



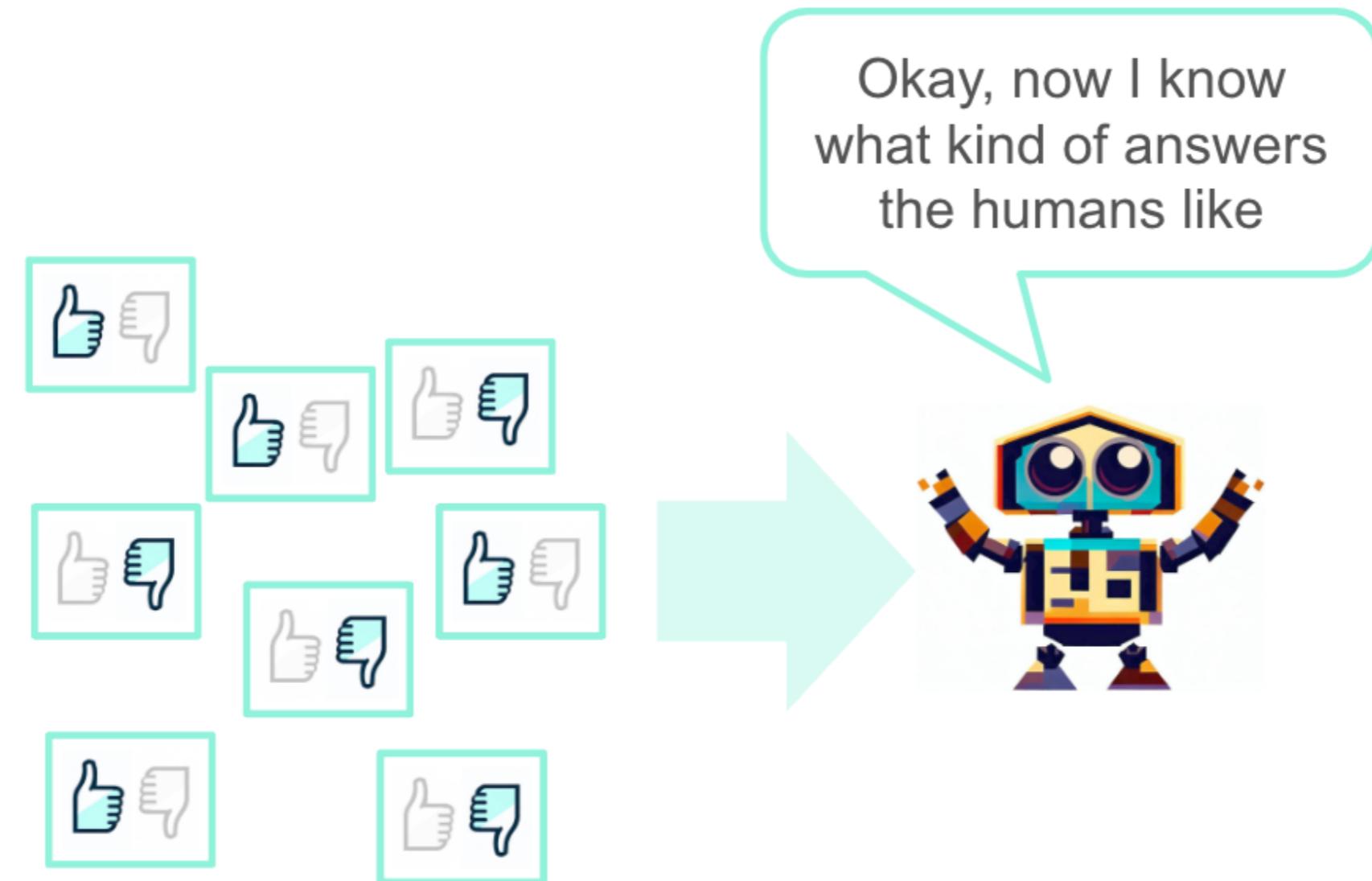
¹ Creative Commons Attribution-Share Alike 4.0, Bing image generator

Where does your feedback go?

The screenshot shows a conversation in the Google Bard interface. A user asks, "Can you give me a simple example of RLHF?" Bard responds with a simple example of RLHF, followed by a bulleted list of steps: "We want to train a robot to walk.", "We start by giving the robot a pre-trained model of how to walk.", "We then ask a human to watch the robot walk and provide feedback on how well it is doing.", "The robot uses this feedback to improve its walking gait.", "We continue this process until the robot is able to walk well." Below this, a note states, "In this example, the human feedback is used to train a reward model. The reward model is then used to train the robot to walk better." Another note says, "RLHF can be used to train a variety of agents, including robots, software agents, and even animals. It is a powerful technique that can be used to improve the performance of agents in a variety of tasks." A section titled "Here are some other examples of RLHF:" lists: "Training a chatbot to have better conversations with humans.", "Training a self-driving car to avoid obstacles.", "Training a robot to play chess.", "Training a software agent to make better financial decisions." A final note says, "RLHF is a rapidly growing field of research with many potential applications. It is an exciting time to be involved in this field!" At the bottom, there are social sharing icons (Facebook, Twitter, Google) and a "Google it" button, along with a prompt input field and a "Send" button.

¹ Google Bard, 2023

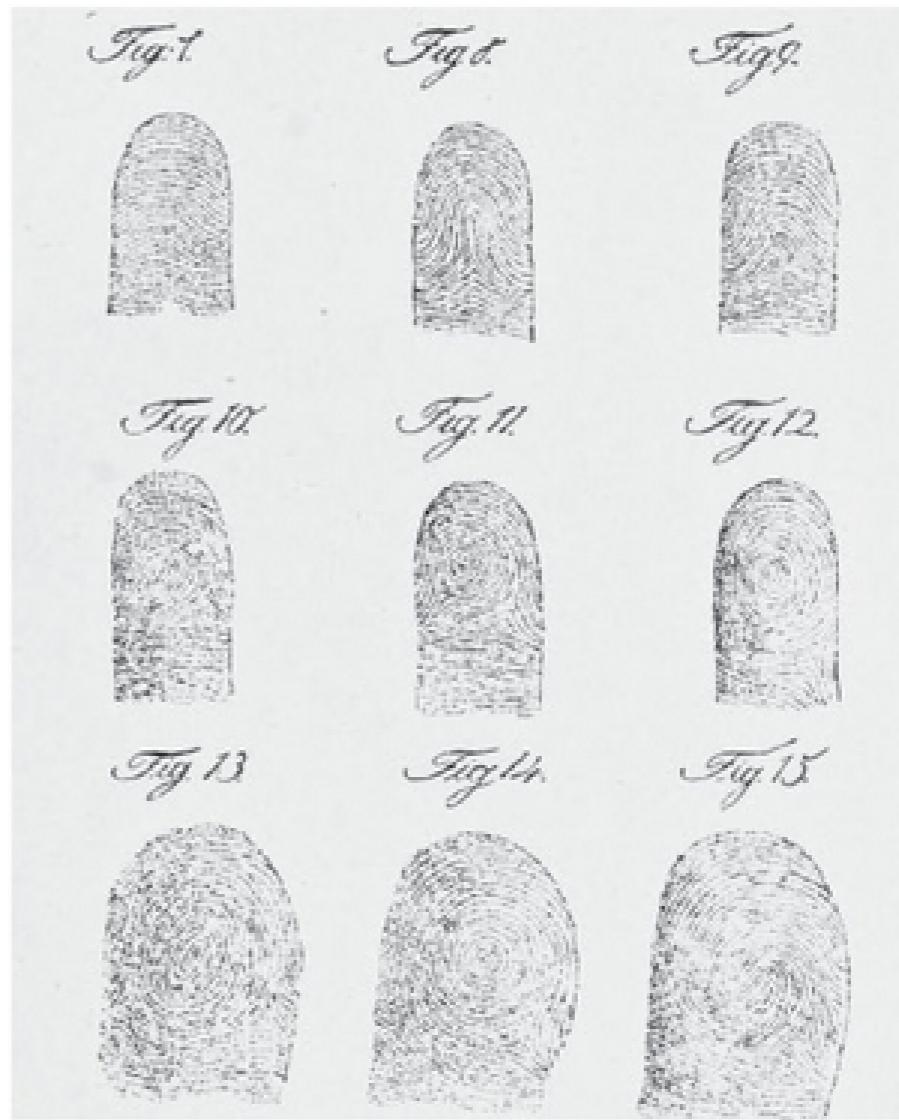
Thumbs up for better responses



¹ Google Bard, 2023

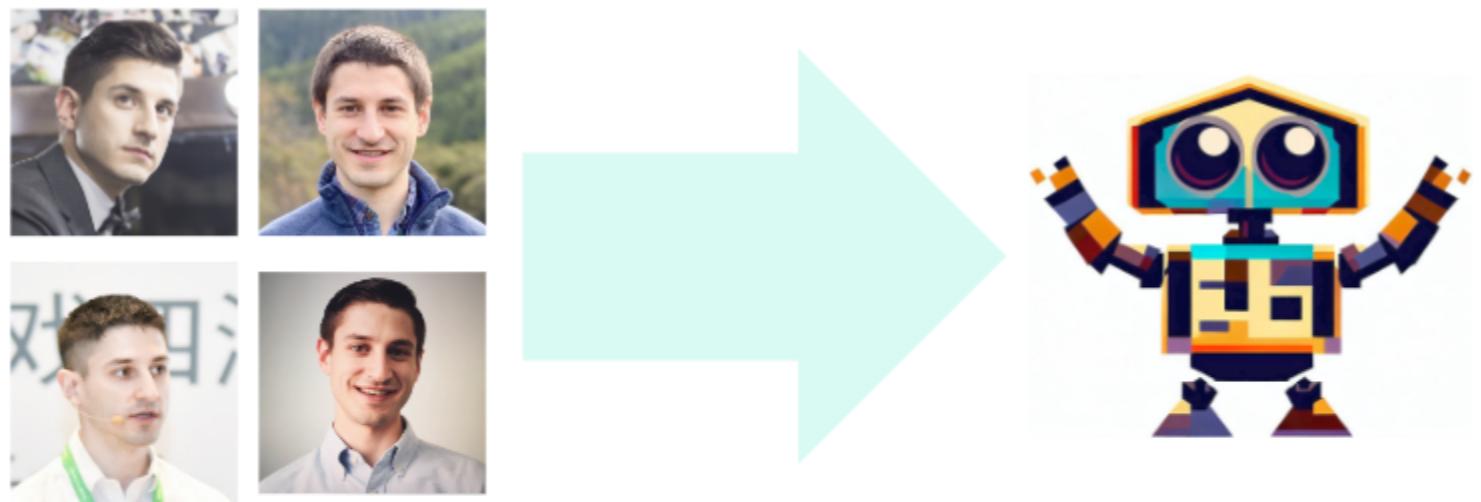
Embeddings as fingerprints

- Similar to recognizing a fingerprint
- Unique representations of data entities
- Capture meaning, context, and relationships in compact form



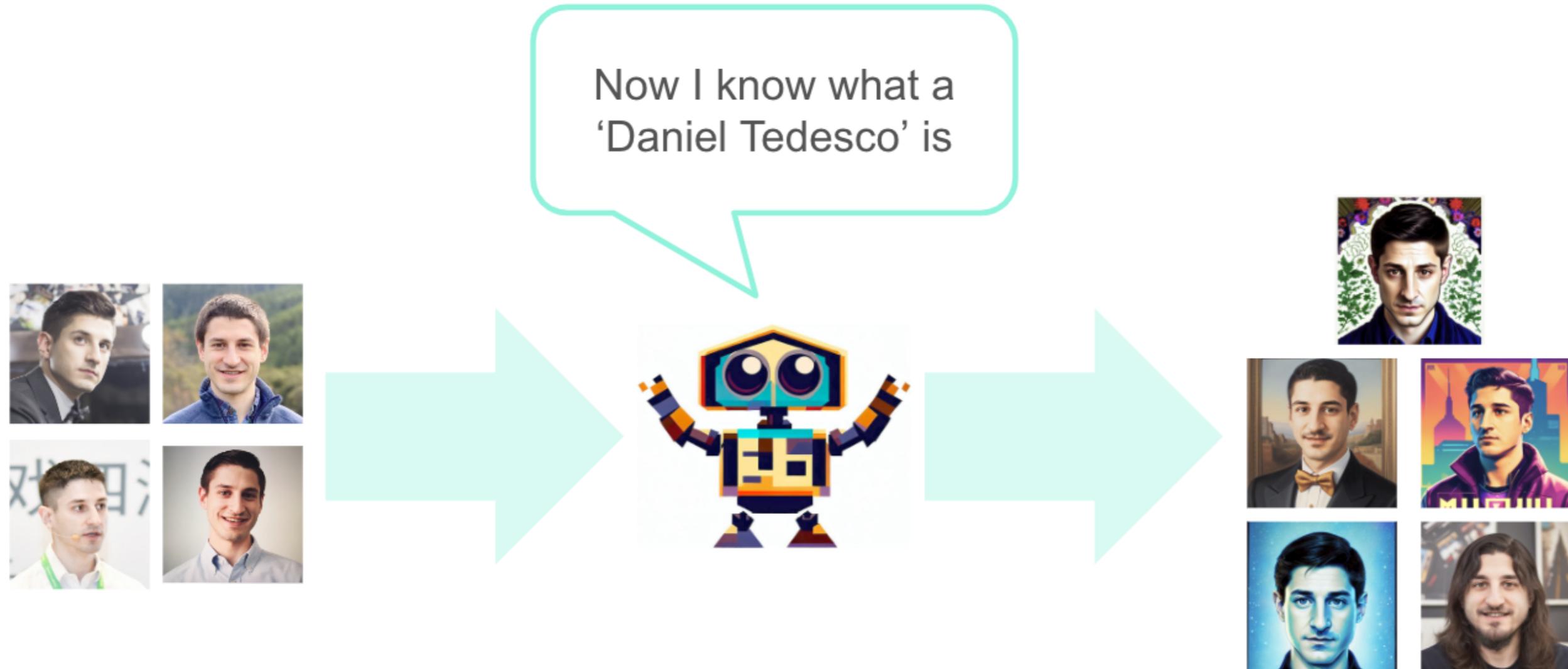
¹ Wikimedia commons

Embedding Dan



¹ Daniel Tedesco, 2023

Embedded Dan



¹ Daniel Tedesco portraits from vana.com

Let's practice!

GENERATIVE AI CONCEPTS

Model evaluation

GENERATIVE AI CONCEPTS



Daniel Tedesco
Data Lead, Google

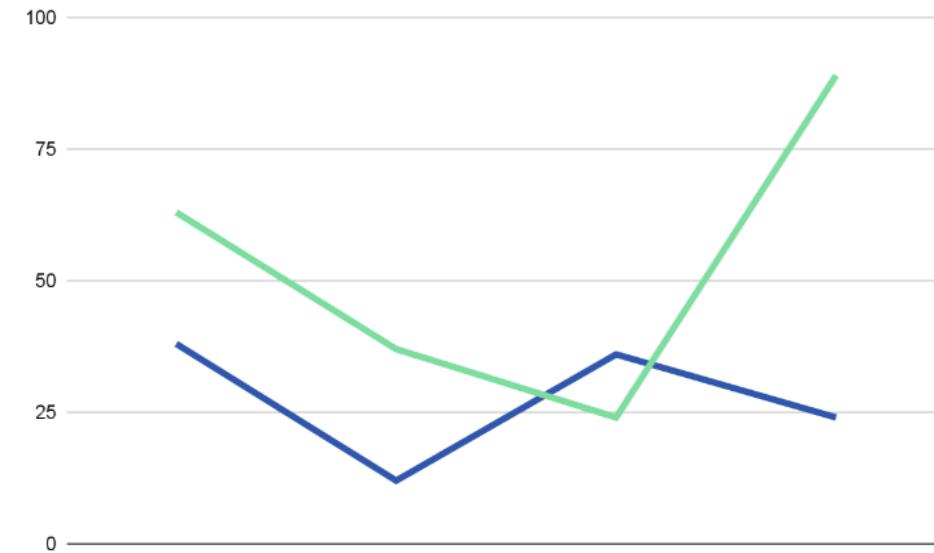
Why evaluate anyway?

Assess performance and effectiveness of a model:

- Measure progress
- Rigorous model comparison
- Benchmark human performance

Evaluating generative AIs

Quantitative Metrics



- Discriminative model evaluation metrics
- Generative model-specific metrics

Human-centric Metrics



- Human performance comparison
- Intelligent evaluation

Discriminative model evaluation techniques

Measure performance on well-defined tasks

Pros:

- Widely accepted and understood
- Easy to calculate and compare

Cons:

- Do not capture subjective nature of generated content



Generative model-specific metrics

Customized for particular generative tasks

Pros:

- Nuanced criteria, like realism, diversity, and novelty
- Many well-known metrics

Cons:

- Cannot capture many subjective elements
- Often do not generalize



Human performance comparison

Pros:

- Benchmarks against human abilities
- Demonstrates practical applicability

Cons:

- Unfair comparison



VS.



Award-winning Als

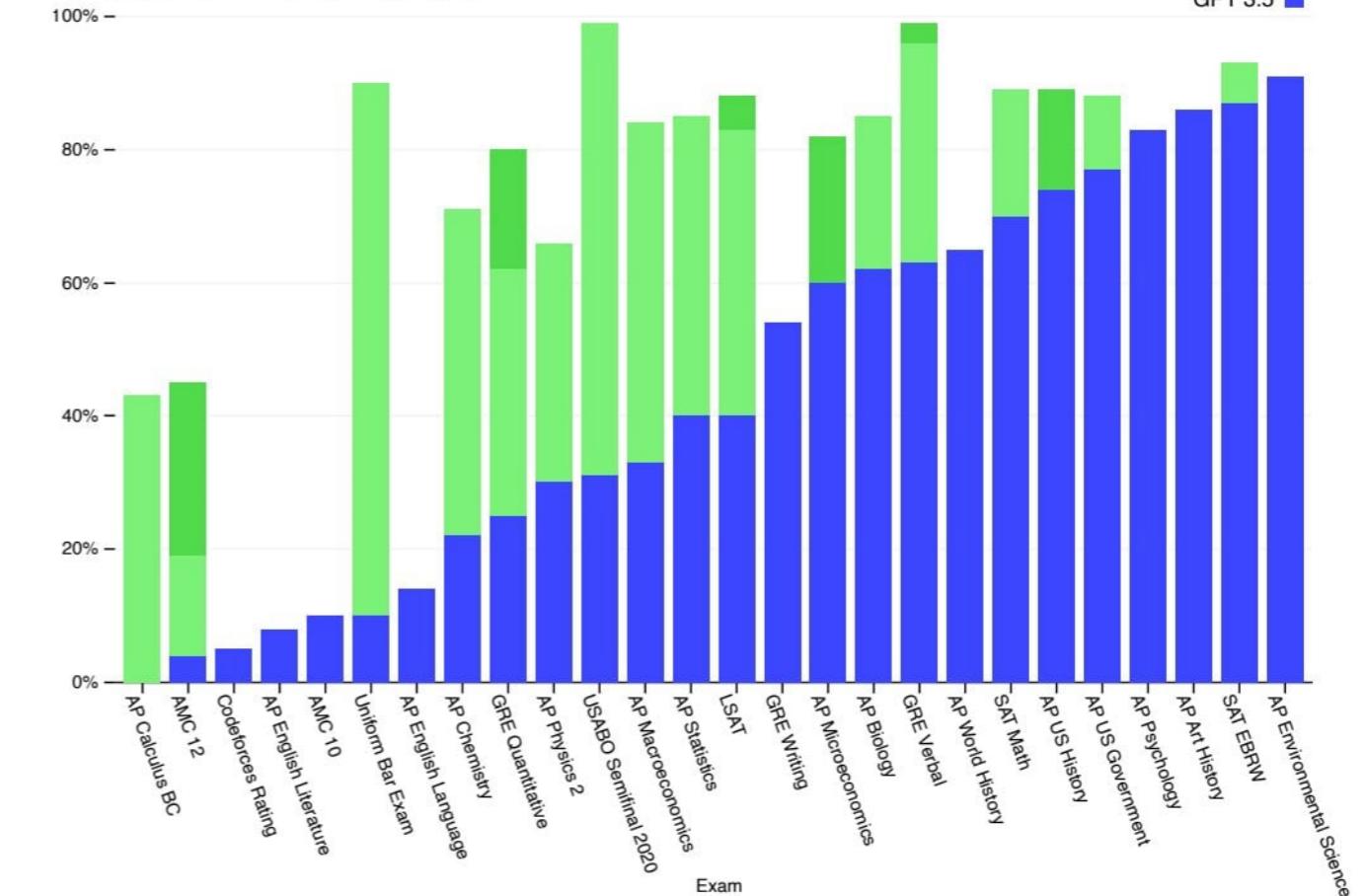
Human Competitions



Human Standardized Tests

Exam results (ordered by GPT 3.5 performance)

Estimated percentile lower bound (among test takers)



¹ <https://twitter.com/colostatefair/status/1565486317839863809>, OpenAI

The gold standard

Intelligent evaluation by humans or other AIs

Pros:

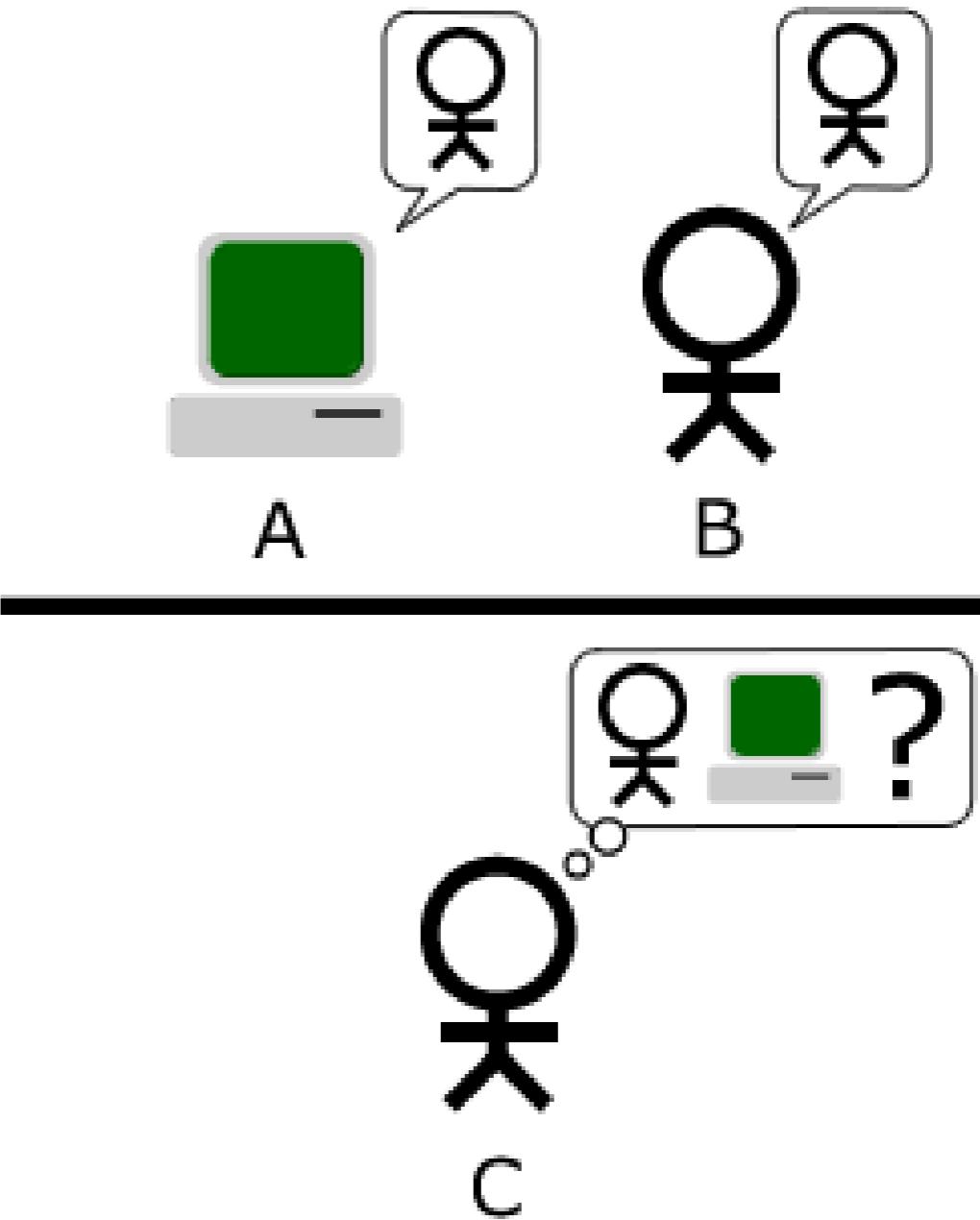
- Captures subjective aspects

Cons:

- Slow, costly, and difficult to standardize
- Subject to human biases and irregularity

Turing's classic test

- Proposed by computer scientist Alan Turing
- Human evaluator judges AI-generated content
- Passes if evaluator cannot distinguish AI from human
- But human behavior is not always the right standard



Let's practice!

GENERATIVE AI CONCEPTS