UC Berkeley
Teaching Professor
Dan Garcia

Berkeley
UNIVERSITY OF CALIFORNIA

Generative AI

Announcements

- Register iclicker
- Postterm this weekend
- Check GradeView
- No handouts today

“…democracy and social order could collapse, resulting in wars…”

In an AI manifesto published April 8, Japan's Nippon Telegraph and Telephone (NTT) and Yomiuri Shimbun Group Holdings called for legislation to rein in generative AI. Despite acknowledging the productivity benefits afforded by generative AI, the manifesto said that if AI remains unchecked, "in the worst-case scenario, democracy and social order could collapse, resulting in wars." The companies called for laws to safeguard elections and national security from generative AI abuse.

https://www.wsj.com/tech/ai/social-order-could-collapse-in-ai-era-two-top-japan-companies-say-1a71cc1d

Today's lecture

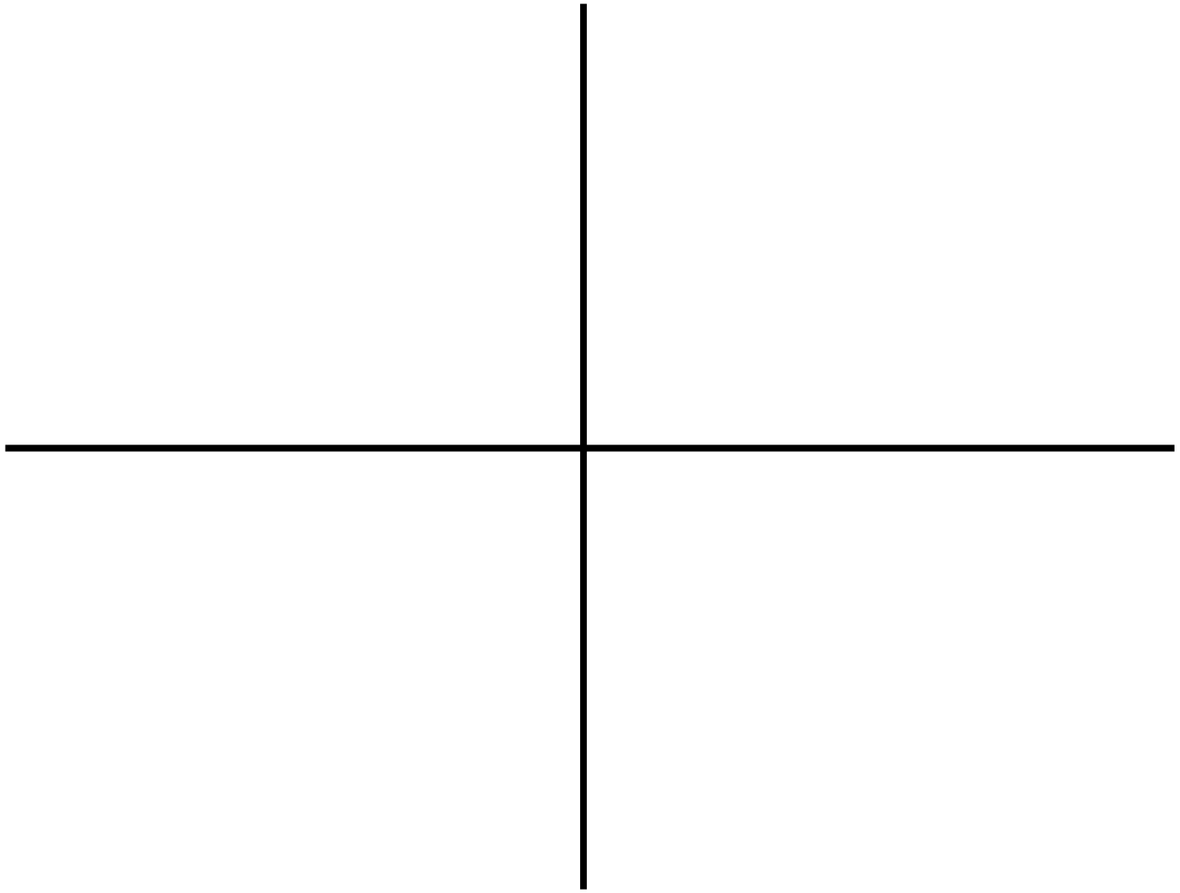
- Brief History of AI
- Machine Learning review, demo
- Generative AI

(Image generated by DALL-E)

Brief History of AI

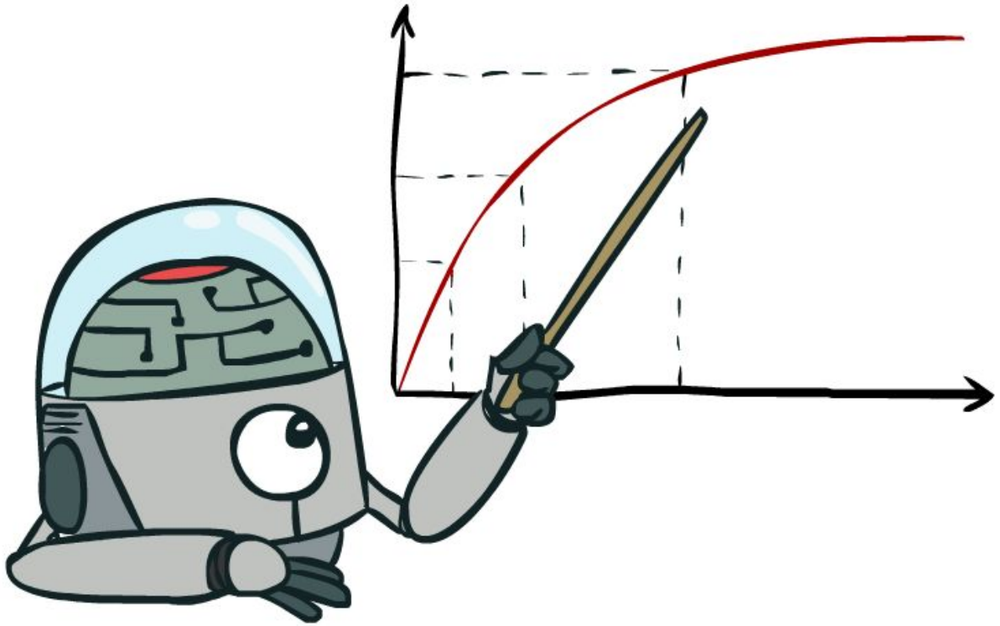
What should we build?

Should we make machines that…



Rational Decision Making

- We'll use the term **rational** in a very specific, technical way:
  - Rational: *maximally achieving pre-defined goals*
  - Goals are expressed in terms of the **utility** of outcomes
  - World is uncertain, so we'll use **expected** utility
  - Being rational means acting to **maximize your expected utility**

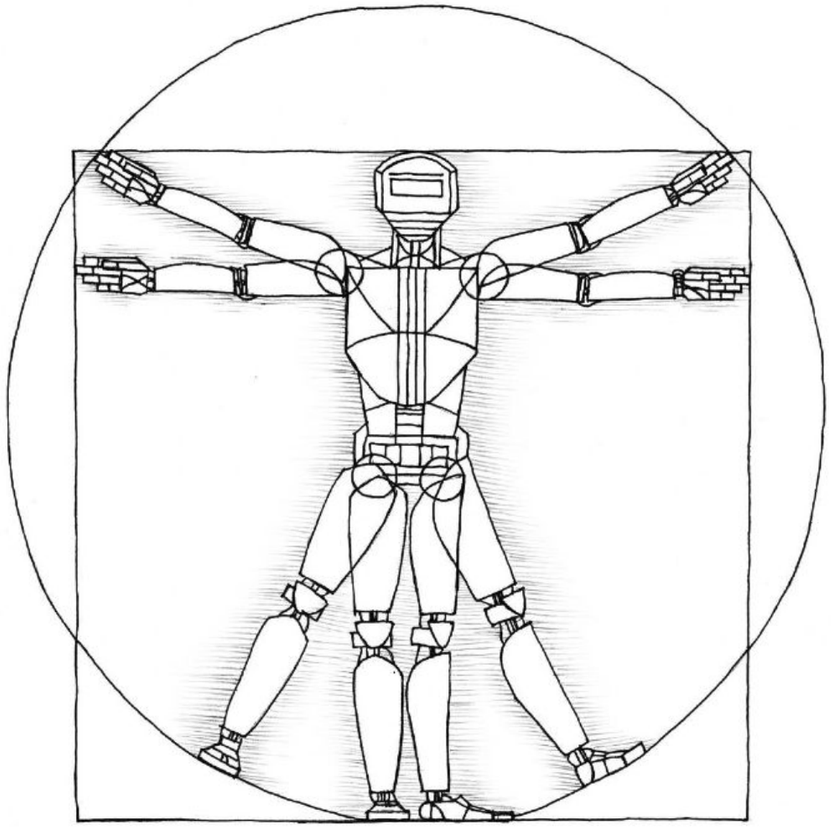


Other Perspectives at Intelligence

- Skills-based perspective
- “A system is only intelligent if it can do [X]”
  - Play chess?
  - Learn from experience?
  - Use words properly?
  - Make mistakes?
  - Make no mistakes?

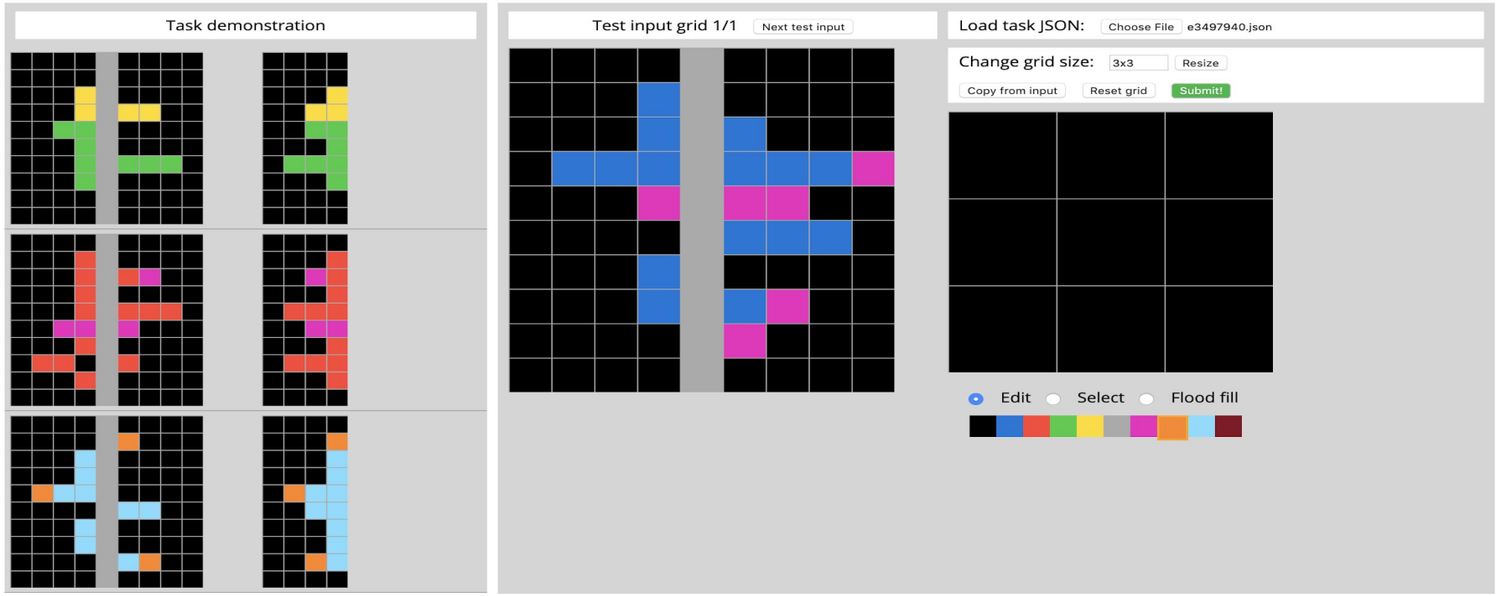
Perspectives at Intelligence

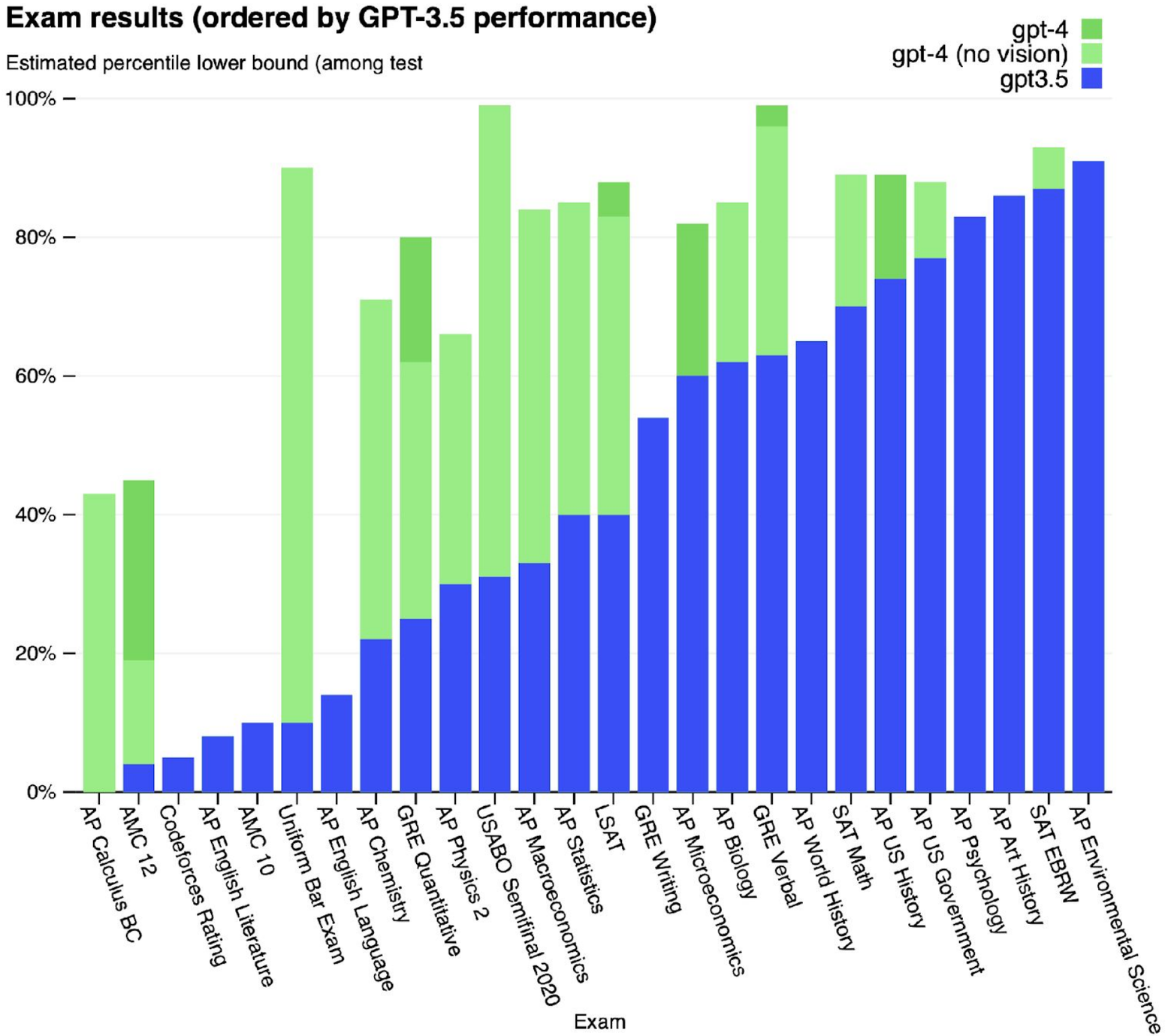
- Embodiment perspective (Rodney Brooks)



Perspectives at Intelligence

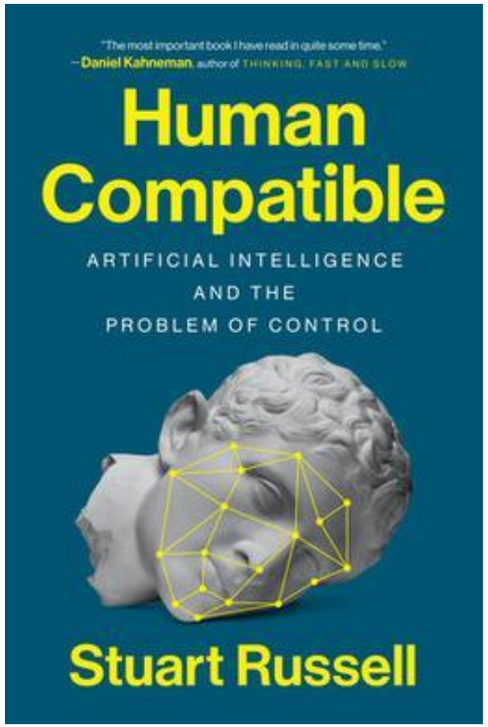
- Psychometrics perspective (François Chollet)
- “Measuring abilities, not skills […] across a broad range of tasks, including tasks that were previously unknown to the ability-enabled system and its developers.”




Exam results (ordered by GPT-3.5 performance)
Estimated percentile lower bound (among test
gpt-4
gpt-4 (no vision)
gpt3.5
100%
80%
60%
40%
20%
0%
AP Calculus BC
AMC 12
Codeforces Rating
AP English Literature
AMC 10
Uniform Bar Exam
AP English Language
AP Chemistry
GRE Quantitative
AP Physics 2
USABO Semifinal 2020
AP Macroeconomics
AP Statistics
LSAT
GRE Writing
AP Microeconomics
AP Biology
GRE Verbal
AP World History
SAT Math
AP US History
AP US Government
AP Psychology
AP Art History
SAT EBRW
AP Environmental Science
Exam

Perspectives at Intelligence

- Human-compatible perspective (Stuart Russell)



1. Machine's objective is to maximize human utility.
2. Initially uncertain about human preferences.
3. Must learn about preferences from human behavior.

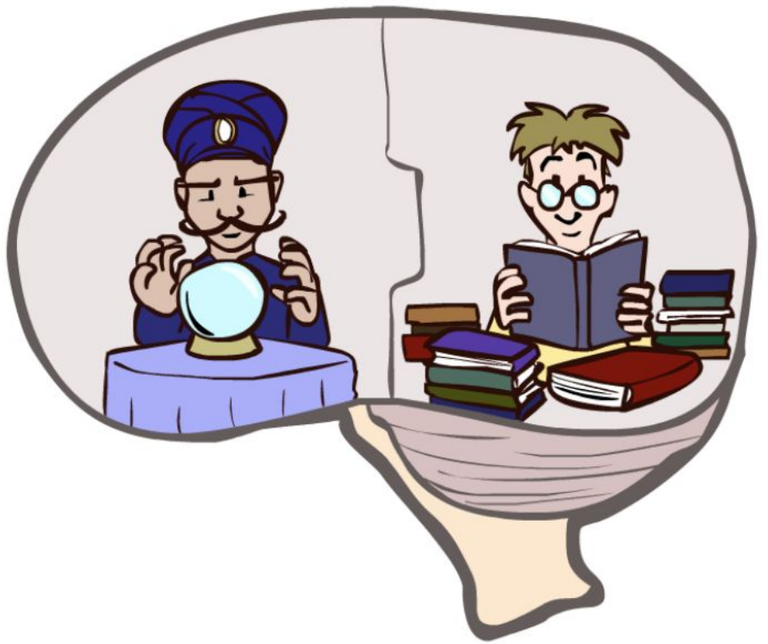
Perspectives at Intelligence

A human being should be able to change a diaper, plan an invasion, butcher a hog, conn a ship, design a building, write a sonnet, balance accounts, build a wall, set a bone, comfort the dying, take orders, give orders, cooperate, act alone, solve equations, analyze a new problem, pitch manure, program a computer, cook a tasty meal, fight efficiently, die gallantly. Specialization is for insects.

—Robert A. Heinlein

What About the Brain?

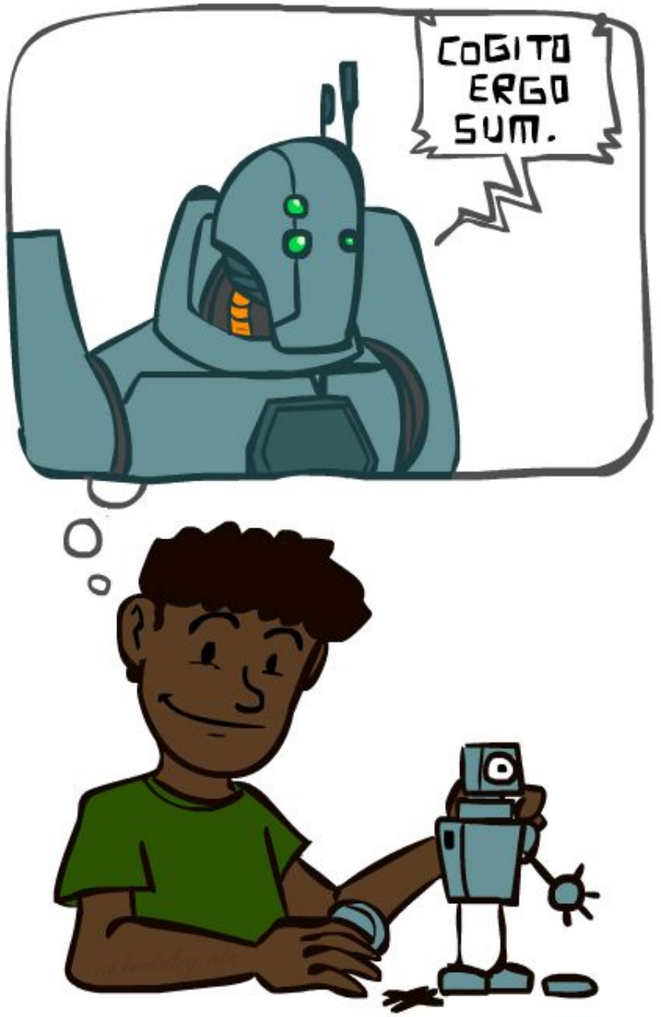
- Brains (human minds) are very good at making rational decisions, but not perfect
- Brains aren't as modular as software, so hard to reverse engineer!
- AI may be better than brains at some tasks
- *"Brains are to intelligence as wings are to flight"*
- We can't yet build AI on the scale of the brain
  - ~100T synapses in the human brain vs
    ~1.8T weights in GPT4
- Still, the brain can be a great inspiration for AI!



A (short) history on AI

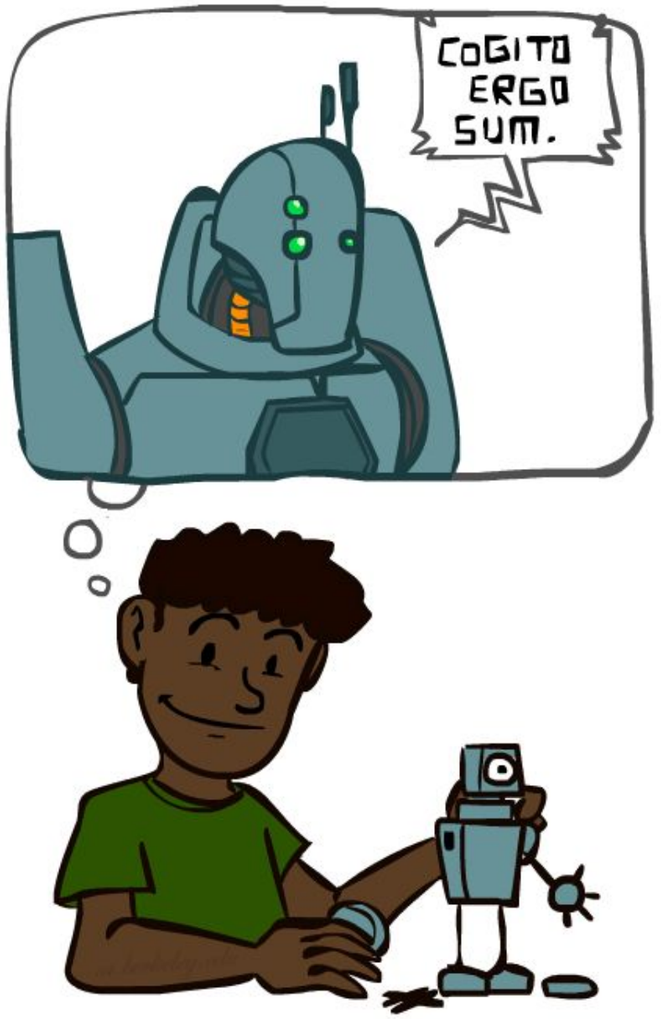
- **1940-1950:** Early days: neural and computer science meet
  - 1943: McCulloch & Pitts: Perceptron–boolean circuit model of brain
  - 1950: Turing's “Computing Machinery and Intelligence” … Turing Test!
- **1950—70:** Excitement! Logic-driven
  - 1950s: Early AI programs, including Samuel's checkers program, Newell & Simon's Logic Theorist, Gelernter's Geometry Engine
  - 1956: Dartmouth meeting: “Artificial Intelligence” adopted

*“We propose that a **2-month, 10-man study of artificial intelligence** be carried out **during the summer of 1956** at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that **every aspect of learning** or **any other feature of intelligence** can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. **We think that a significant advance can be made** in one or more of these problems if a carefully selected group of scientists work on it together for a summer.”*



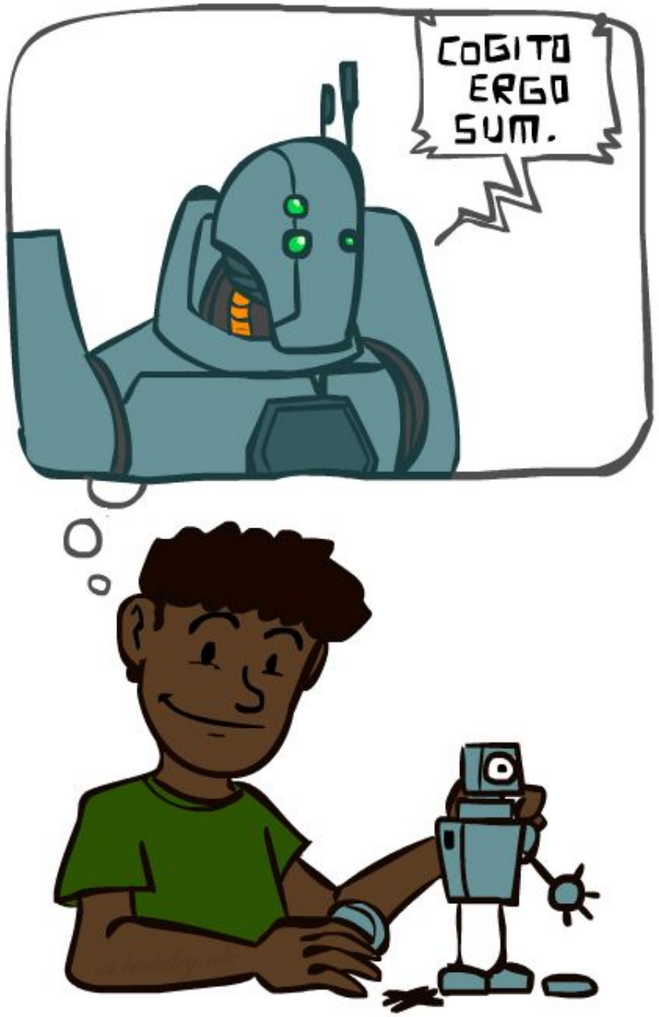
A (short) history on AI

- **1940-1950: Early days: neural and computer science meet**
  - 1943: McCulloch & Pitts: Perceptron–boolean circuit model of brain
  - 1950: Turing's “Computing Machinery and Intelligence”
- **1950—70: Excitement! Logic-driven**
  - 1950s: Early AI programs, including Samuel's checkers program, Newell & Simon's Logic Theorist, Gelernter's Geometry Engine
  - 1956: Dartmouth meeting: “Artificial Intelligence” adopted
  - 1969: Minsky & Papert: Perceptrons can’t learn XOR / parity!
- **1970—90: Knowledge-based approaches**
  - 1969—79: Early development of knowledge-based systems
  - 1980—88: Expert systems industry booms; backpropagation makes it feasible to train multi-layer neural networks
  - 1988—93: Expert systems industry busts: “AI Winter”
- **1990—2010: Statistical approaches, agents**
  - Resurgence of probability, focus on uncertainty
  - Agents and learning systems… “AI Spring”?
  - 1992: TD-Gammon achieves human-level play at backgammon
  - 1997: Deep Blue defeats Gary Kasparov at chess
  - 2002: Embodied AI; Roomba vacuum invented



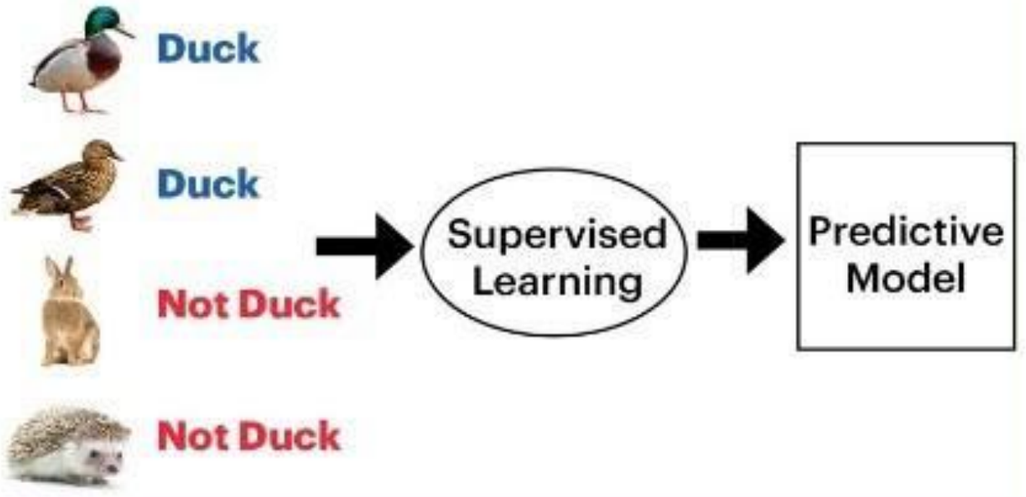
A (short) history on AI

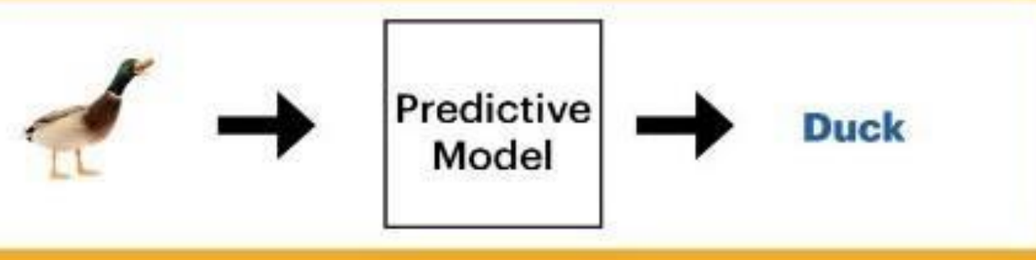
- **2010—2017: Big Data, GPUs, Deep Learning**
  - 2011: Apple releases Siri
  - 2012: AlexNet wins ImageNet competition
  - 2015: DeepMind achieves “human-level” control in Atari games
  - 2016: DeepMind’s AlphaGo defeats Lee Sedol at Go
  - 2016: Google Translate migrates to neural networks
- **2017—: Scaling Up, Large Language Models**
  - 2017: Google invents Transformer architecture
  - 2017: DeepStack/Libratus defeat humans at poker
  - 2018-2020: AlphaFold predicts protein structure from amino acids
  - 2021-2022: Modern text-to-image generation
  - 2022: OpenAI releases ChatGPT
  - 2023: Every other company also releases a chatbot



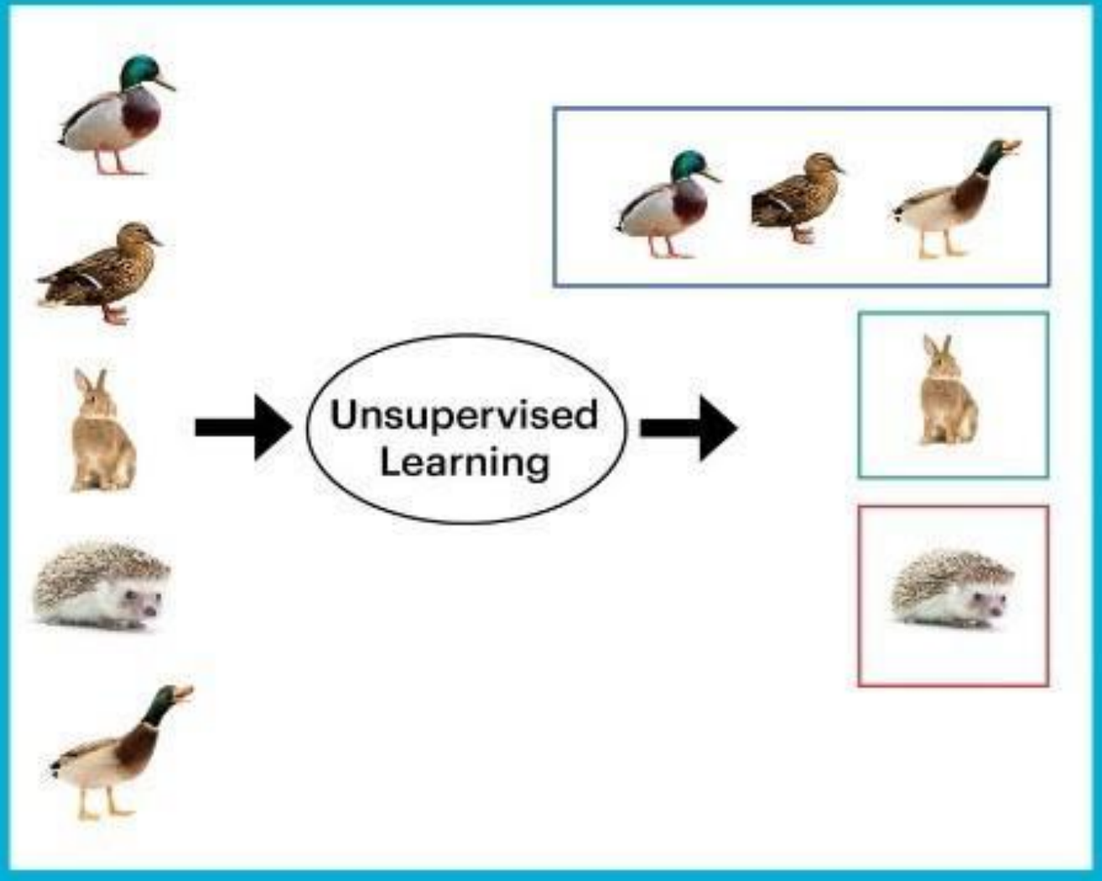
Machine Learning Review

Supervised Learning (Classification Algorithm)





Unsupervised Learning (Clustering Algorithm)



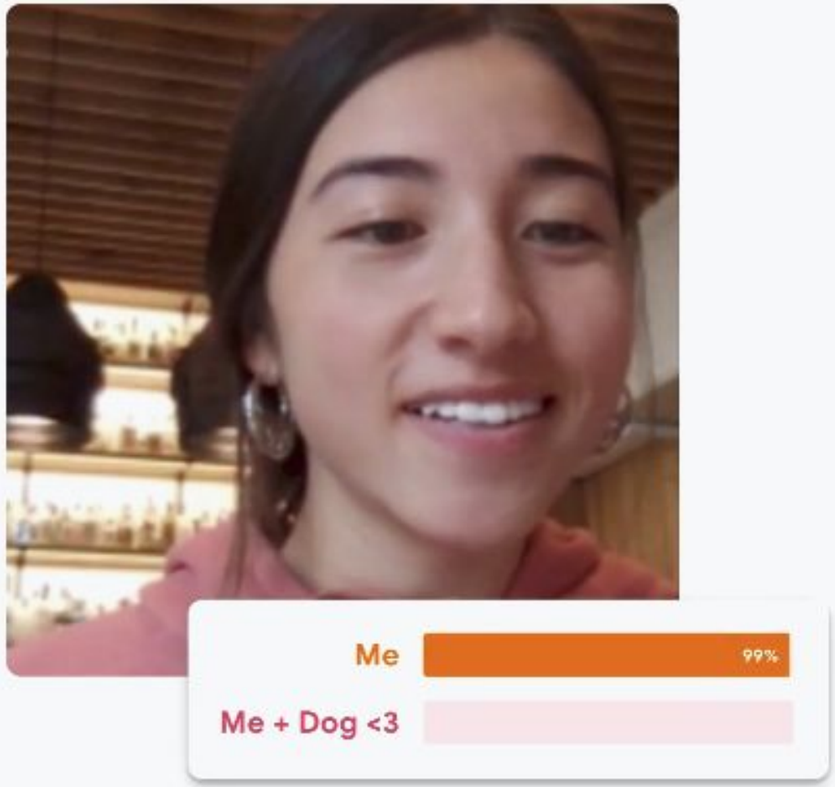
Western Digital.

Teachable Machine

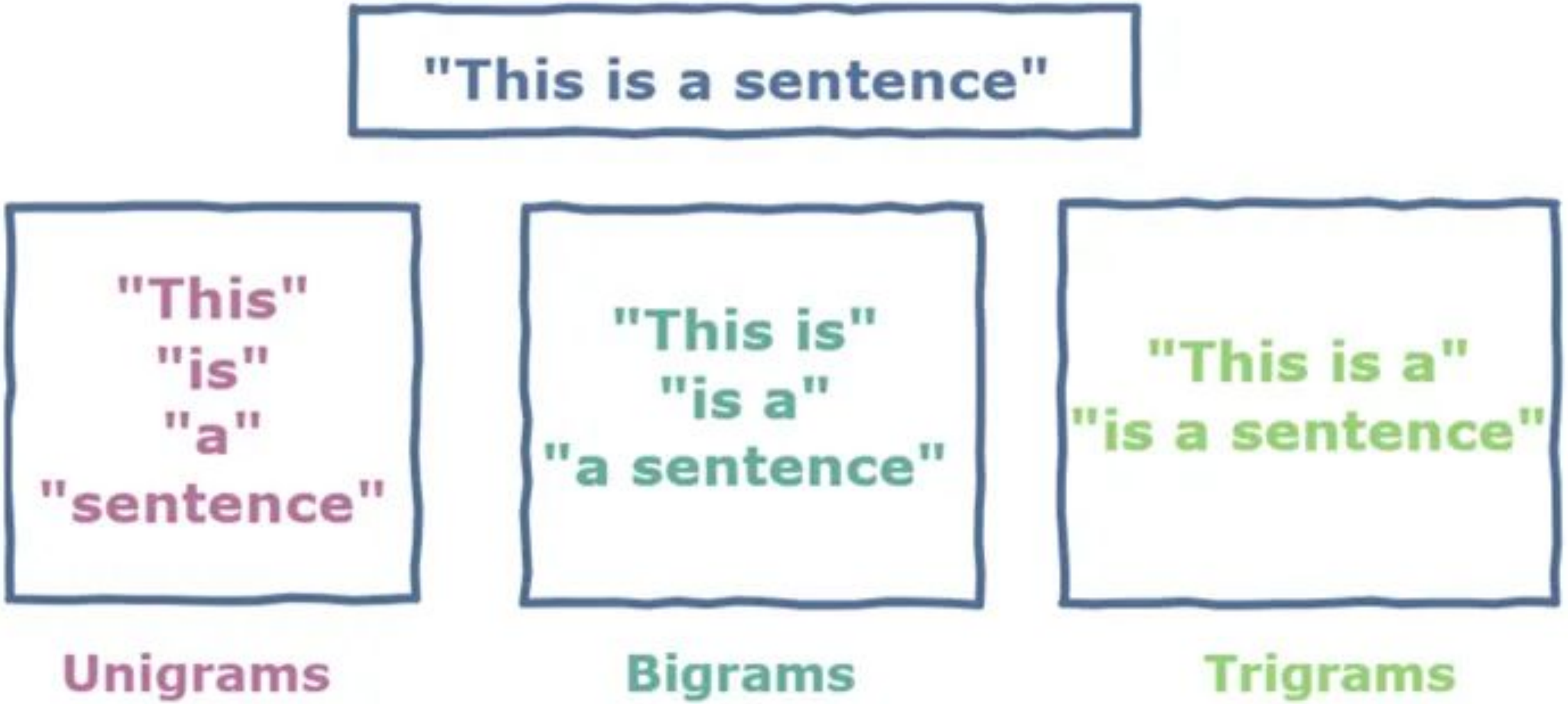
Train a computer to recognize your own images, sounds, & poses.

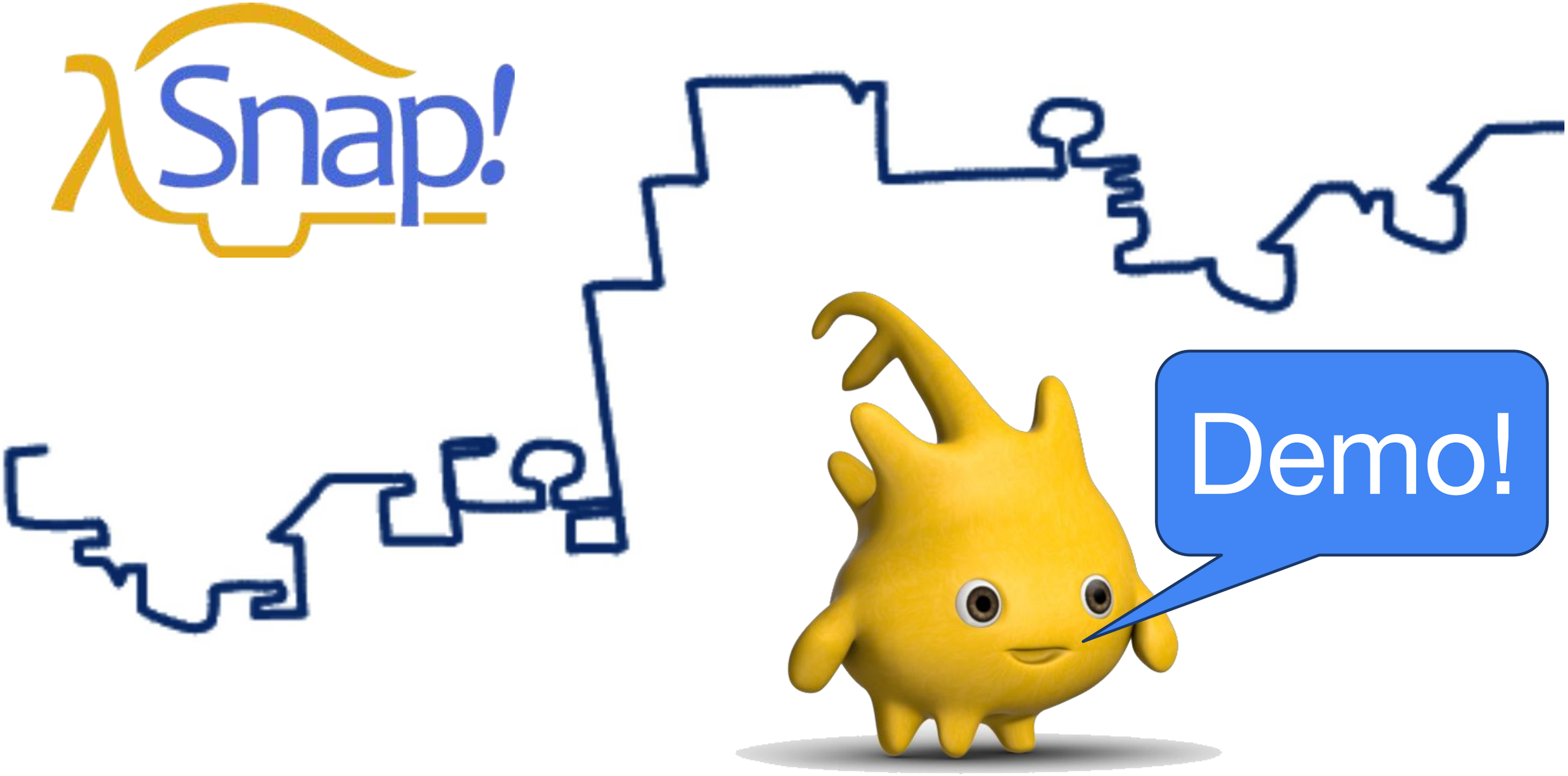
A fast, easy way to create machine learning models for your sites, apps, and more – no expertise or coding required.

Get Started



Generative AI


"This is a sentence"
"This"
"is"
"a"
"sentence"
"This is"
"is a"
"a sentence"
"This is a"
"is a sentence"
Unigrams
Bigrams
Trigrams


λSnap!
Demo!

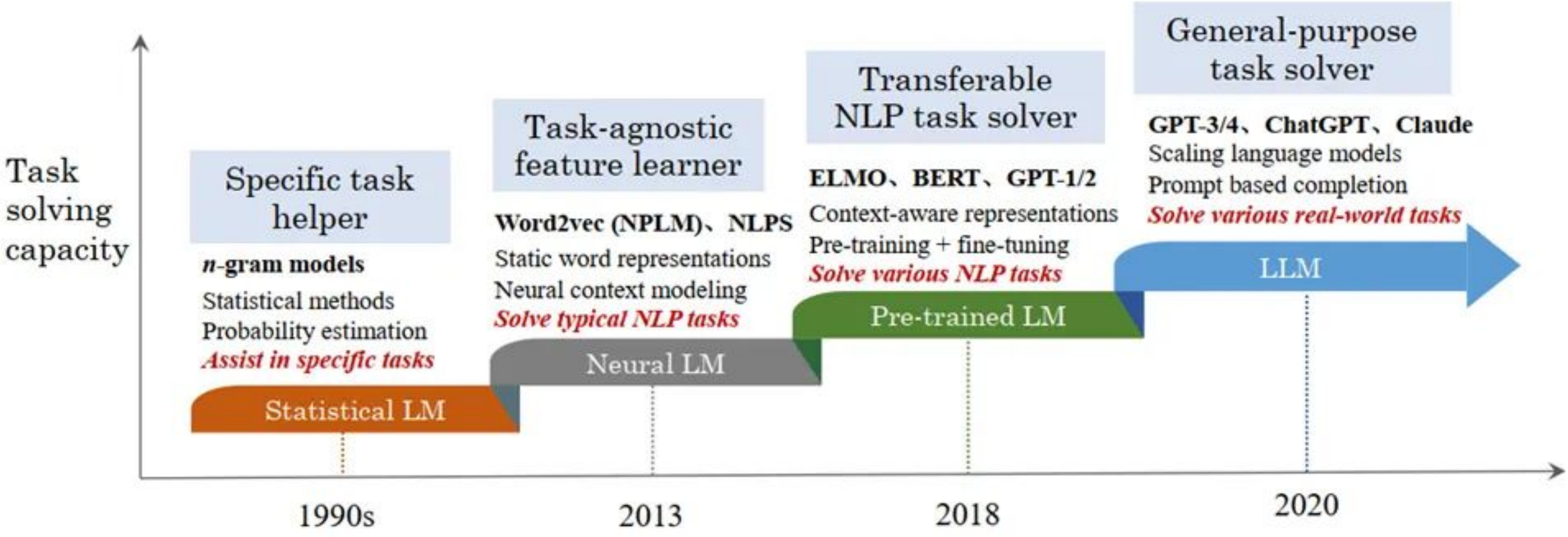
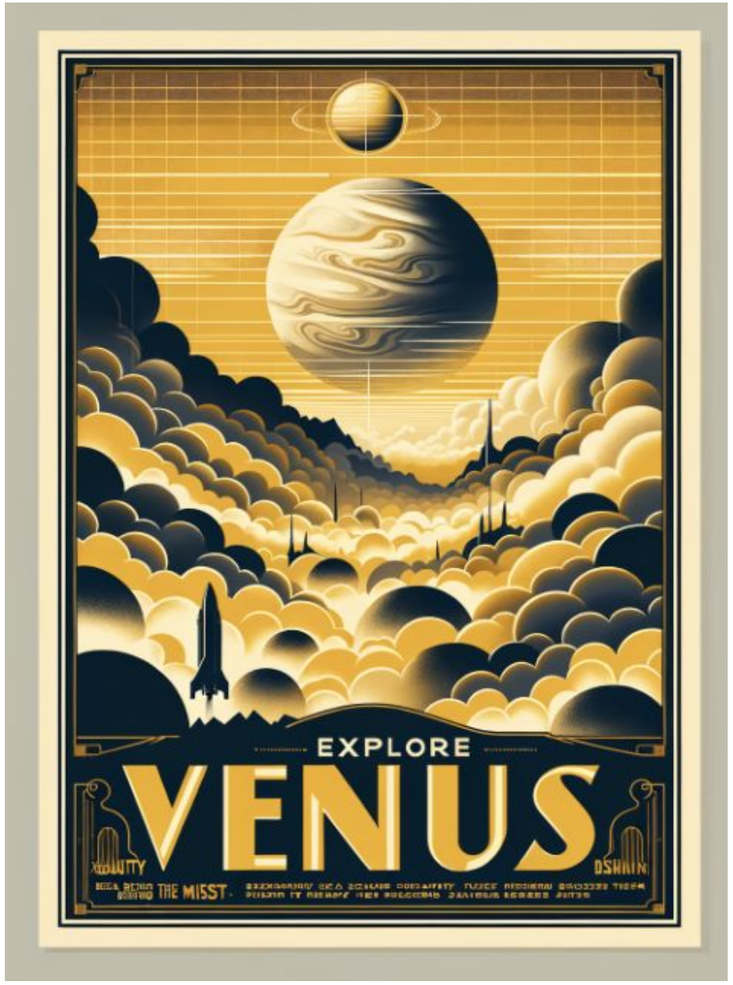

Task solving capacity
Specific task helper
n-gram models
Statistical methods
Probability estimation
Assist in specific tasks
Statistical LM
Task-agnostic feature learner
Word2vec (NPLM)、NLPS
Static word representations
Neural context modeling
Solve typical NLP tasks
Neural LM
Transferable NLP task solver
ELMO、BERT、GPT-1/2
Context-aware representations
Pre-training + fine-tuning
Solve various NLP tasks
Pre-trained LM
General-purpose task solver
GPT-3/4、ChatGPT、Claude
Scaling language models
Prompt based completion
Solve various real-world tasks
LLM
1990s
2013
2018
2020

Image Generation



Progress in image generation

“A cup of coffee”

“A cat”

Image credit: cs280-berkeley.github.io

Progress in image generation

“A cup of coffee”

“A cat”

“A cup of cat”

Image credit: cs280-berkeley.github.io

Progress in image generation

“A cup of coffee”

“A cat”

“A cup of cat”

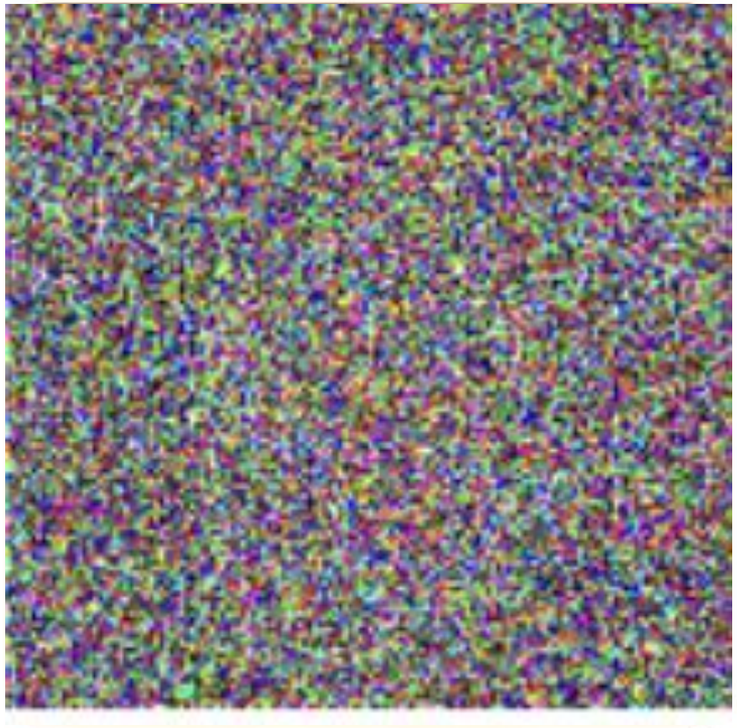


Image credit: cs280-berkeley.github.io

DALL-E and Diffusion Models

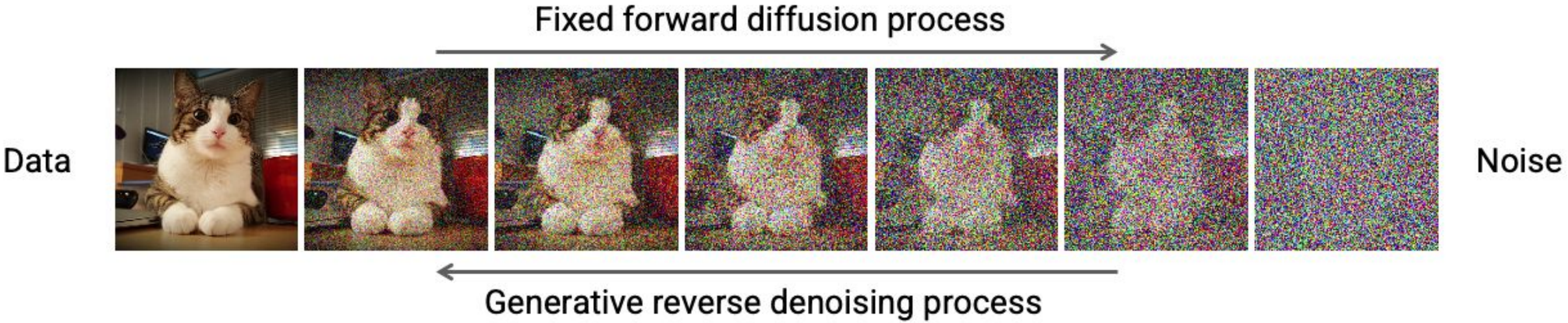
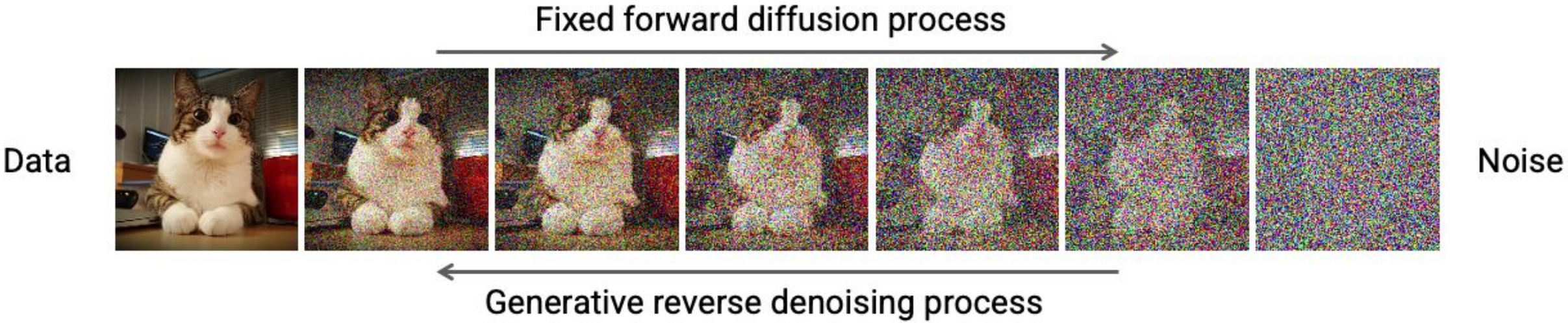


Image credit: (Amatriain, 2023)

DALL-E and Diffusion Models



Model learns to reverse the process of adding noise

Image credit: (Amatriain, 2023)