Generative Artificial Intelligence

Siddhartha Singh, 07.09.2023

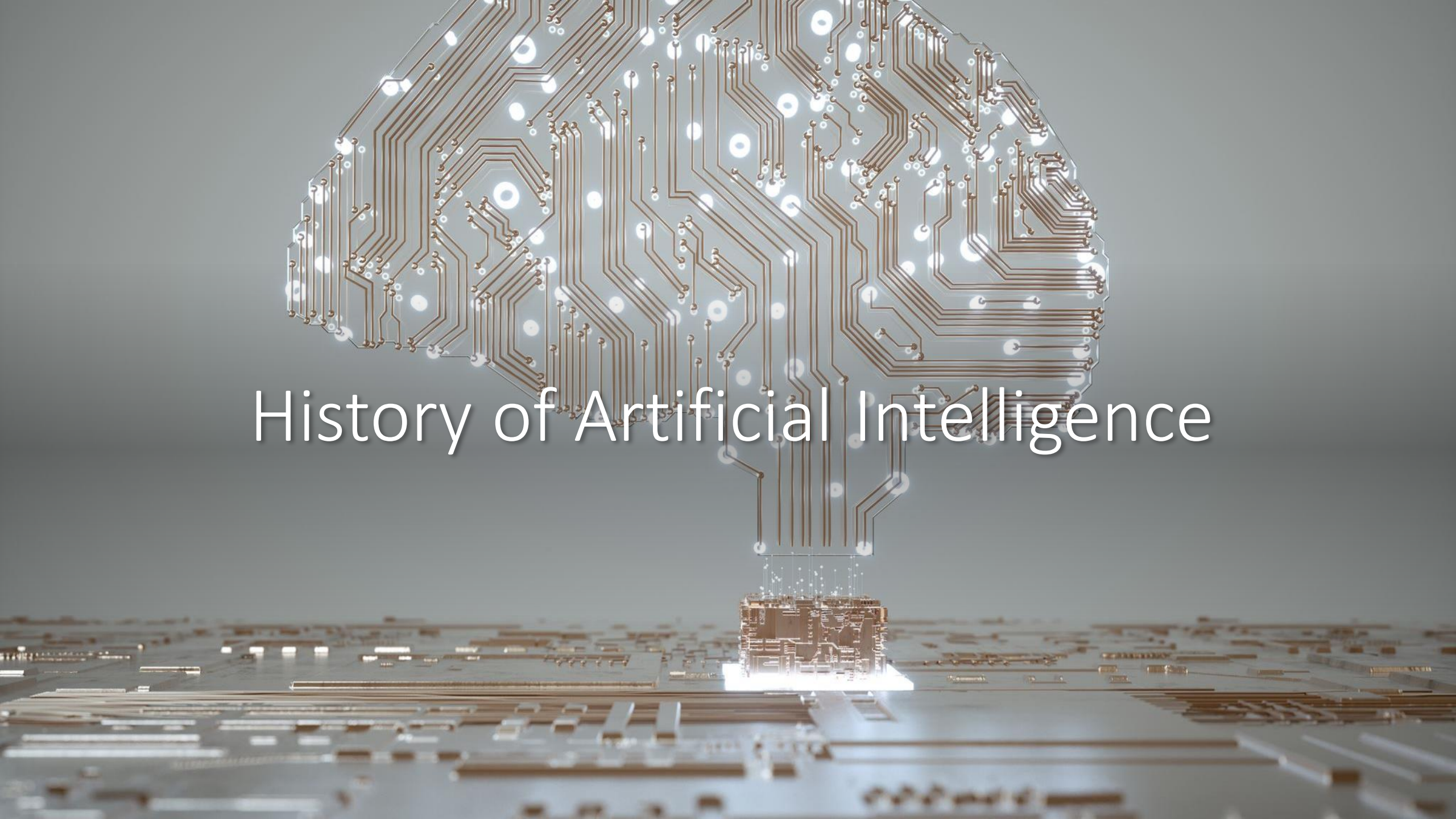
Goals

- History of AI
- Go through all the elements that make ChatGPT
  - You must do this hands-on throughout the day
  - You won't make your own ChatGPT but rather understand each element
  - You can make your own ChatGPT at home after the workshop
- Debate on effects of AI

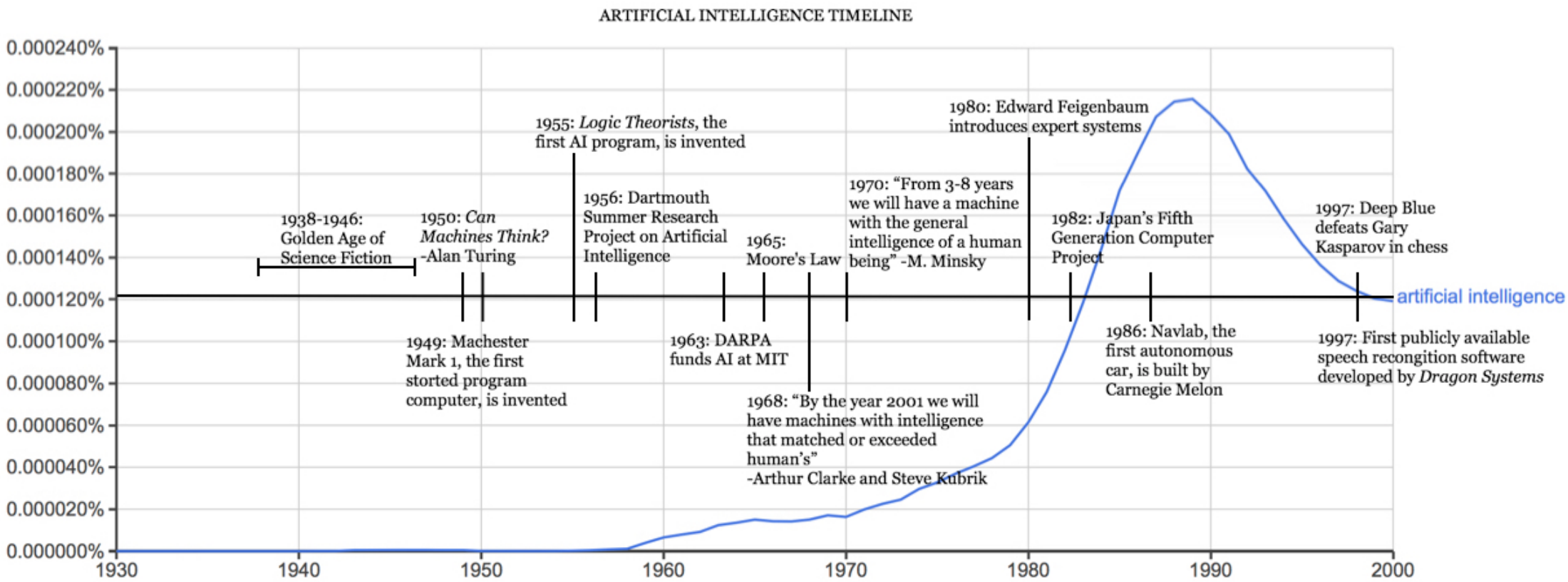
Out-of-Scope

- Basic introduction to machine learning
- Parameter tuning to create perfect results
  - You can do this on your own at home
- Understanding why the methods work
  - Most theoretical principals are out-of-scope for this workshop

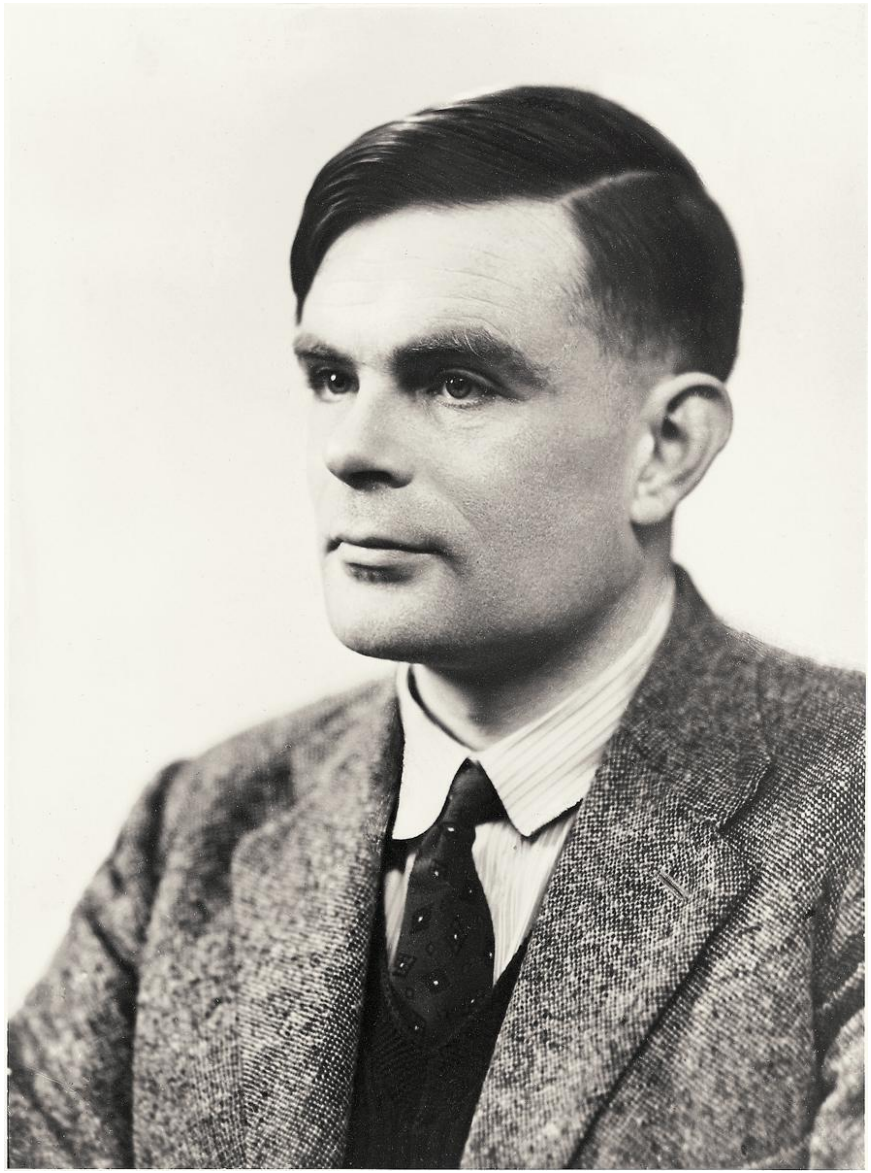
How is Generative AI different from other types of AI?

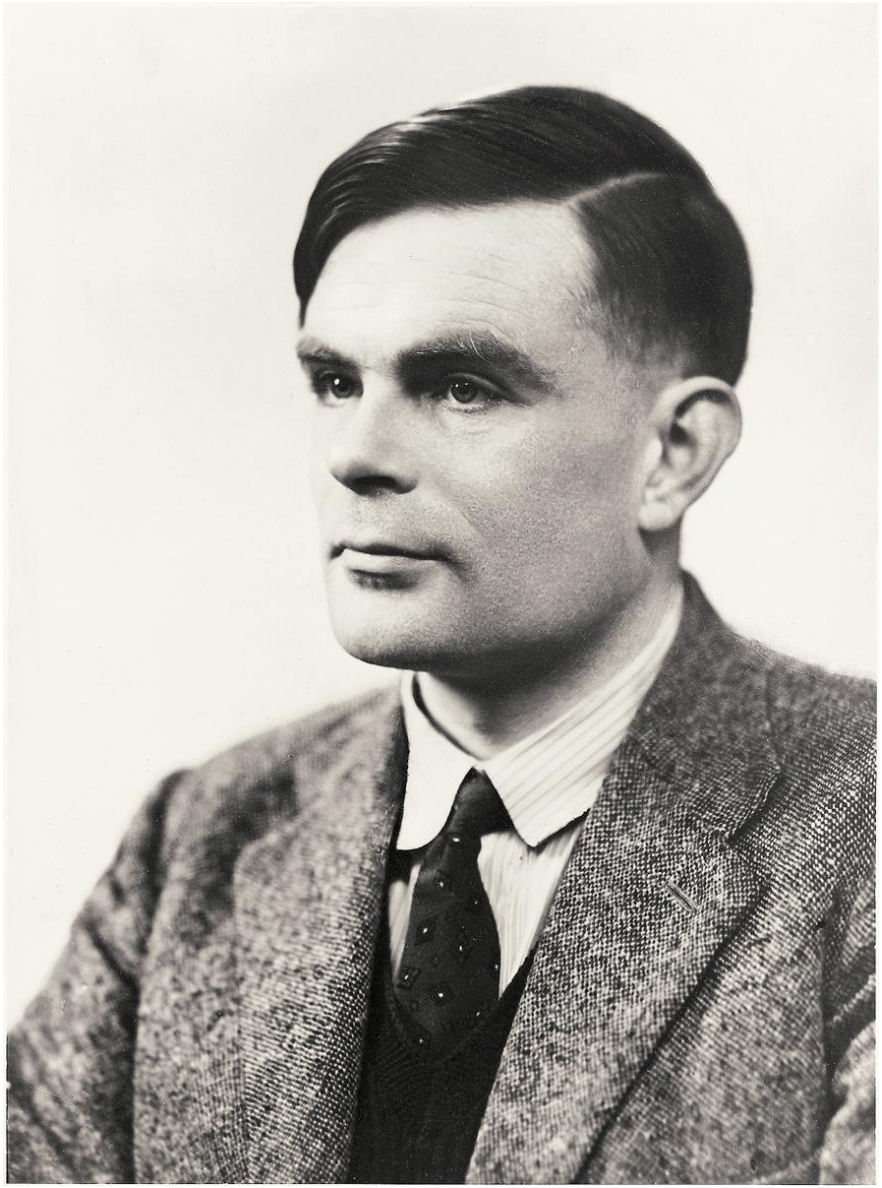


History of Artificial Intelligence


ARTIFICIAL INTELLIGENCE TIMELINE
0.000240%
0.000220%
0.000200%
0.000180%
0.000160%
0.000140%
0.000120%
0.000100%
0.000080%
0.000060%
0.000040%
0.000020%
0.000000%
1930
1940
1950
1960
1970
1980
1990
2000
1938-1946: Golden Age of Science Fiction
1950: *Can Machines Think?* -Alan Turing
1955: *Logic Theorists*, the first AI program, is invented
1956: Dartmouth Summer Research Project on Artificial Intelligence
1965: Moore's Law
1970: "From 3-8 years we will have a machine with the general intelligence of a human being" -M. Minsky
1980: Edward Feigenbaum introduces expert systems
1982: Japan's Fifth Generation Computer Project
1997: Deep Blue defeats Gary Kasparov in chess
artificial intelligence
1949: Machester Mark 1, the first storted program computer, is invented
1963: DARPA funds AI at MIT
1968: "By the year 2001 we will have machines with intelligence that matched or exceeded human's" -Arthur Clarke and Steve Kubrik
1986: Navlab, the first autonomous car, is built by Carnegie Melon
1997: First publicly available speech recongition software developed by *Dragon Systems*


1950-1974





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

MIND

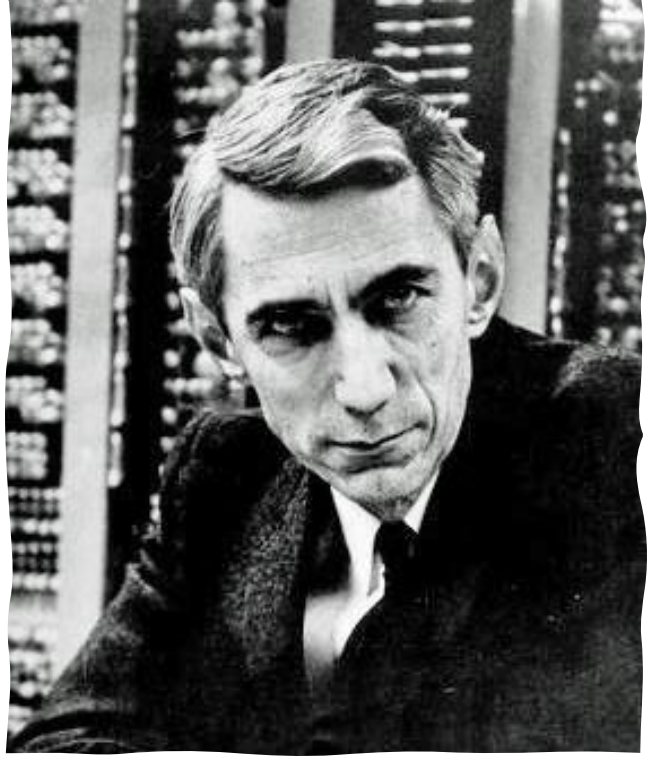
A QUARTERLY REVIEW

OF

PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND INTELLIGENCE

BY A. M. TURING



- The name "Artificial Intelligence" originated from the 1956 Dartmouth Worshop organised by **Marvin Misky**, **John McCarthy**, **Claude Shannon** and **Nathan Rochester**.

Developments

- Work on reasoning as *search problems.*
- Chatbots like ELIZA could carry out conversations, sometimes fooling users.
- Micro-worlds creation
  - SHRDLU: NLP system developed at MIT
  - Could move objects around in the "blocks world" using english language
- Japan initiated work into humanoid robots in this era (and they are still the leaders today).

Person: Pick up a big red block.

Computer: OK.

Person: Grasp the pyramid.

Computer: I DON'T UNDERSTAND WHICH PYRAMID YOU MEAN.

Person (changing their mind): Find a block which is taller than the one you are holding and put it into the box.

Computer: BY "IT", I ASSUME YOU MEAN THE BLOCK WHICH IS TALLER THAN THE ONE I AM HOLDING.

Computer: OK.

Person: What does the box contain?

Computer: THE BLUE PYRAMID AND THE BLUE BLOCK.

Person: What is the pyramid supported by?

Computer: THE BOX.

Person: How many blocks are not in the box?

Computer: FOUR OF THEM.

Person: Is at least one of them narrower than the one which i told you to pick up?

Computer: YES, THE RED CUBE.

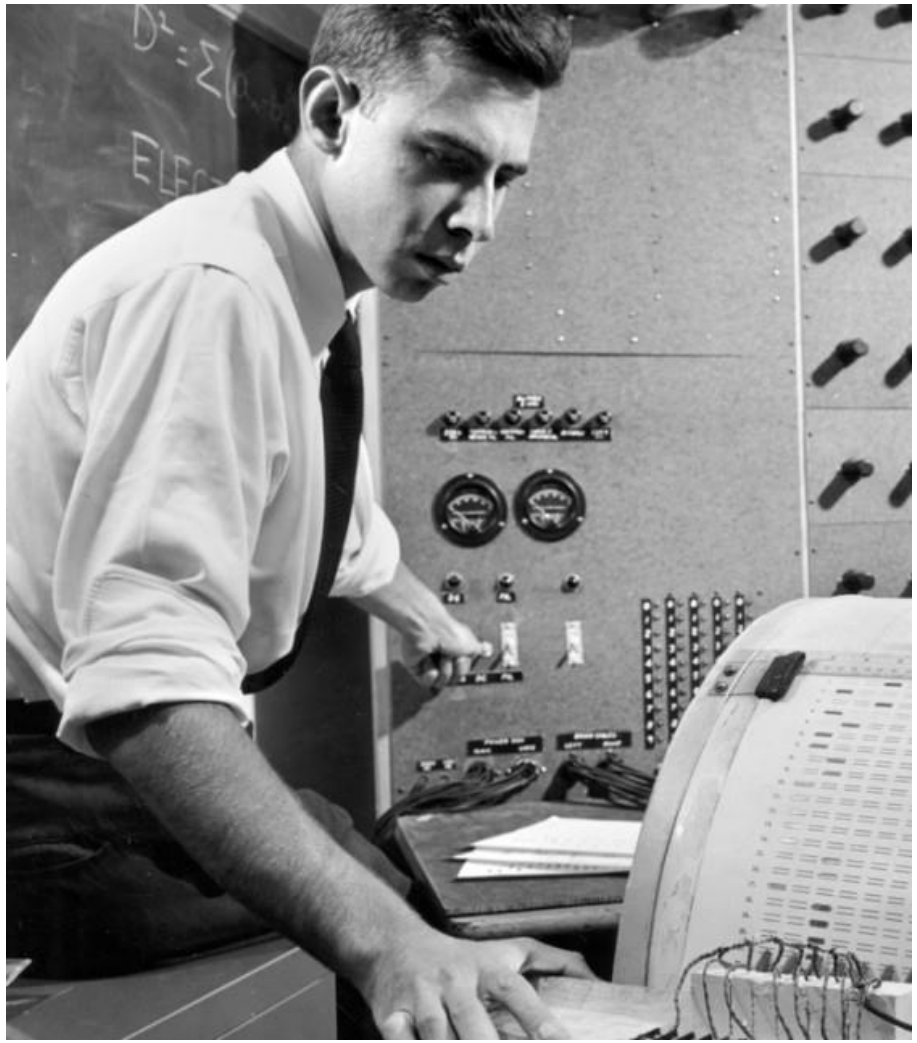
Person: Is it supported?

Computer: YES, BY THE TABLE.

Person: Can the table pick up blocks?

Computer: NO.

Foundations in Deep Learning, 1958



1969 book by Misky and Papert, called *Perceptrons* killed the entire funding in the field.

Schools of AI

- The AI Community already started segregating into different schools:
  - Symbolism
  - Connectionism
  - Evolutionaries
  - Bayesians
  - Analogizers

Great Optimism and Hype

"within 10 years a digital computer will be the world's chess champion", "within 10 years a digital computer will discover and prove an important mathematical theorm"
- H.A. Simon, Allen Newell 1958

"machines will be capable, within 20 years, of doing any work man can do"
- H.A. Simon 1965

"within a generation ... the problem of creating 'artificial intelligence' will substantially be solved"
- Marvin Minsky 1967

"In from 3 to 8 years we will have a machine with the general intelligence of an average human being"
- Marvin Minsky, Life Magazine 1970

Complexity Research Progress

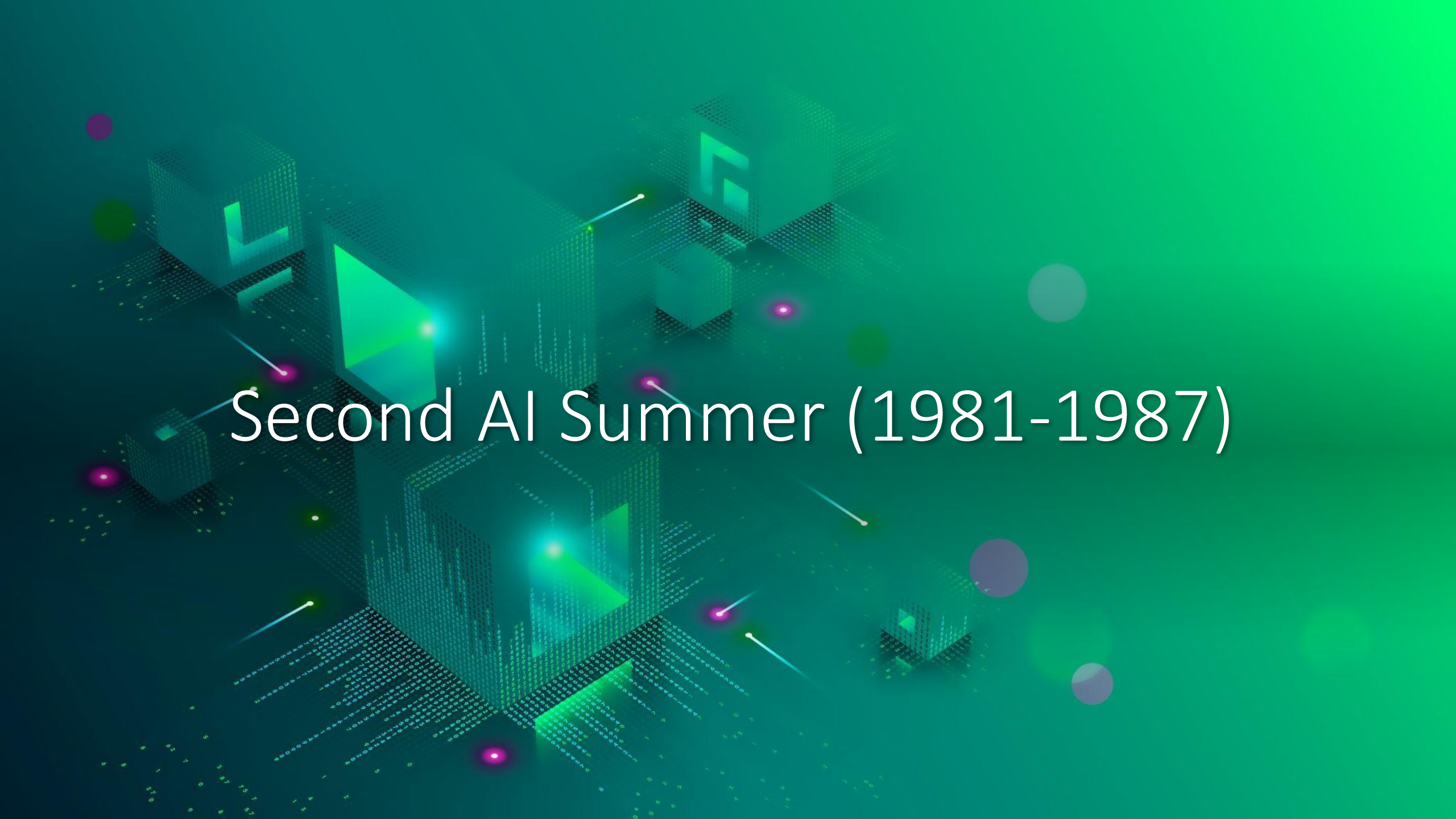
- Combinatorial Explosion was realised to be the single biggest issue in AI.
- Research into the limits of compute made great strides in this time.
- AI crashed into the complexity barrier
  - NP-Completness: to this date there are no solution to these problems

The Lighthill Report (1973)


1974-1980
First AI Winter

First AI Winter

- The 1973 Lighthill Report criticised the utter failure of AI to achive all the grandiose objectives it promised.
- Sir James Lighthill specifically mentioned the *combinatorial explosion* and *intractability* problems.
- Around same time, DARPA cuts back on funding on AI research.


Second AI Summer (1981-1987)

Boom Time Again!

- Rise of the Expert Systems
  - Based on rules
  - Explicit knowledge representations
  - Can also "explain" the answers
  - **MYCIN** was developed to diagnose infectious blood diseases
- Knowledge Based Systems
  - **Cyc** was the first attempt to commonsense knowledge problem by creating a huge database, created by Douglas Lenat.
- The Japanese set aside $850million in 1981 for Fifth Generation Computer.
- UK began their £350million Alvey project.
- Neural Network were revived.
- Judea Pearl introduces Bayesian Networks for causal analysis.

1989 Demo: Neural Networks


1987-2000
Second AI Winter

Second AI Winter (1987-2000)

- LISP machine market collapses
- Strategic Computing Initiative cancels new spending on AI
- By 1990s, many *Expert Systems* are abandoned
- In the same timeline, the Japanese abandon the Fifth Generation computer's goals.
- Many companies shut-down.

Another AI Summer (1993 - Present)

New AI Summer

- 1997: Deep Blue defeats Gary Kasparov
- 1997: LSTM proposed by Hochreiter and Schmidhuber
- Rise of AI Agents that can understand and react to their environment
- Progress in probabilistic reasoning, greater rigour
- Many problems where tackled such as speech recognition, search (giving rise to companies like Google), data mining etc.
- 2006: Fei-Fei Li starts working on ImageNet
- 2009: Raina et.al. introduce using GPU with Neural Networks
- 2012: Geoffery Hinton, Ilya Sutskever and Alex Krizhevsky introduces CNN winning the ImageNet challenge, triggering the explosion of interest shown in AI today!!

2013: Google Deepmind introduces Atari Agent

2014: Generative Adversarial Networks are introduced by Goodfellow et. al.

2016: Google Deepmind introduces AlphaGo

2017: Vaswani et. al. at Google introduce Attention Models

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

2018: OpenAI introduces GPT-1 Model

Improving Language Understanding by Generative Pre-Training

Alec Radford
OpenAI
alec@openai.com

Karthik Narasimhan
OpenAI
karthikn@openai.com

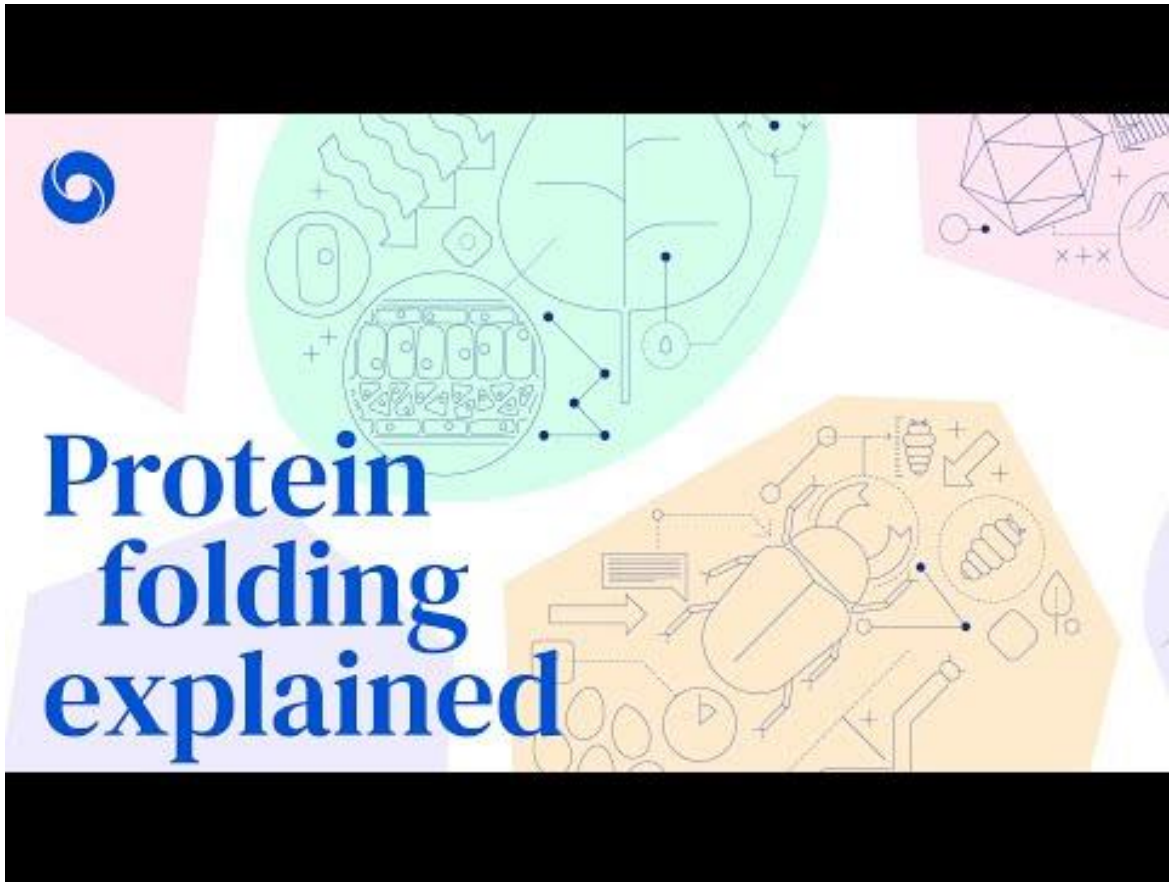
Tim Salimans
OpenAI
tim@openai.com

Ilya Sutskever
OpenAI
ilyasu@openai.com

Abstract

Natural language understanding comprises a wide range of diverse tasks such as textual entailment, question answering, semantic similarity assessment, and document classification. Although large unlabeled text corpora are abundant, labeled data for learning these specific tasks is scarce, making it challenging for discriminatively trained models to perform adequately. We demonstrate that large gains on these tasks can be realized by *generative pre-training* of a language model on a diverse corpus of unlabeled text, followed by *discriminative fine-tuning* on each specific task. In contrast to previous approaches, we make use of task-aware input transformations during fine-tuning to achieve effective transfer while requiring minimal changes to the model architecture. We demonstrate the effectiveness of our approach on a wide range of benchmarks for natural language understanding. Our general task-agnostic model outperforms discriminatively trained models that use architectures specifically crafted for each task, significantly improving upon the state of the art in 9 out of the 12 tasks studied. For instance, we achieve absolute improvements of 8.9% on commonsense reasoning (Stories Cloze Test), 5.7% on question answering (RACE), and 1.5% on textual entailment (MultiNLI).

2021: Google Deepmind introduces AlphaFold



2022: OpenAI opens ChatGPT

ChatGPT — Release Notes

The latest update for ChatGPT

Written by Natalie
Updated over a week ago

Introducing ChatGPT Enterprise (August 28, 2023)

Today we're launching ChatGPT Enterprise, which offers enterprise-grade security and privacy, unlimited higher-speed GPT-4 access, longer context windows for processing longer inputs, advanced data analysis capabilities, customization options, and much more.

ChatGPT Enterprise also provides unlimited access to Advanced Data Analysis, previously known as Code Interpreter.

Learn more on our website and connect with our sales team to get started.

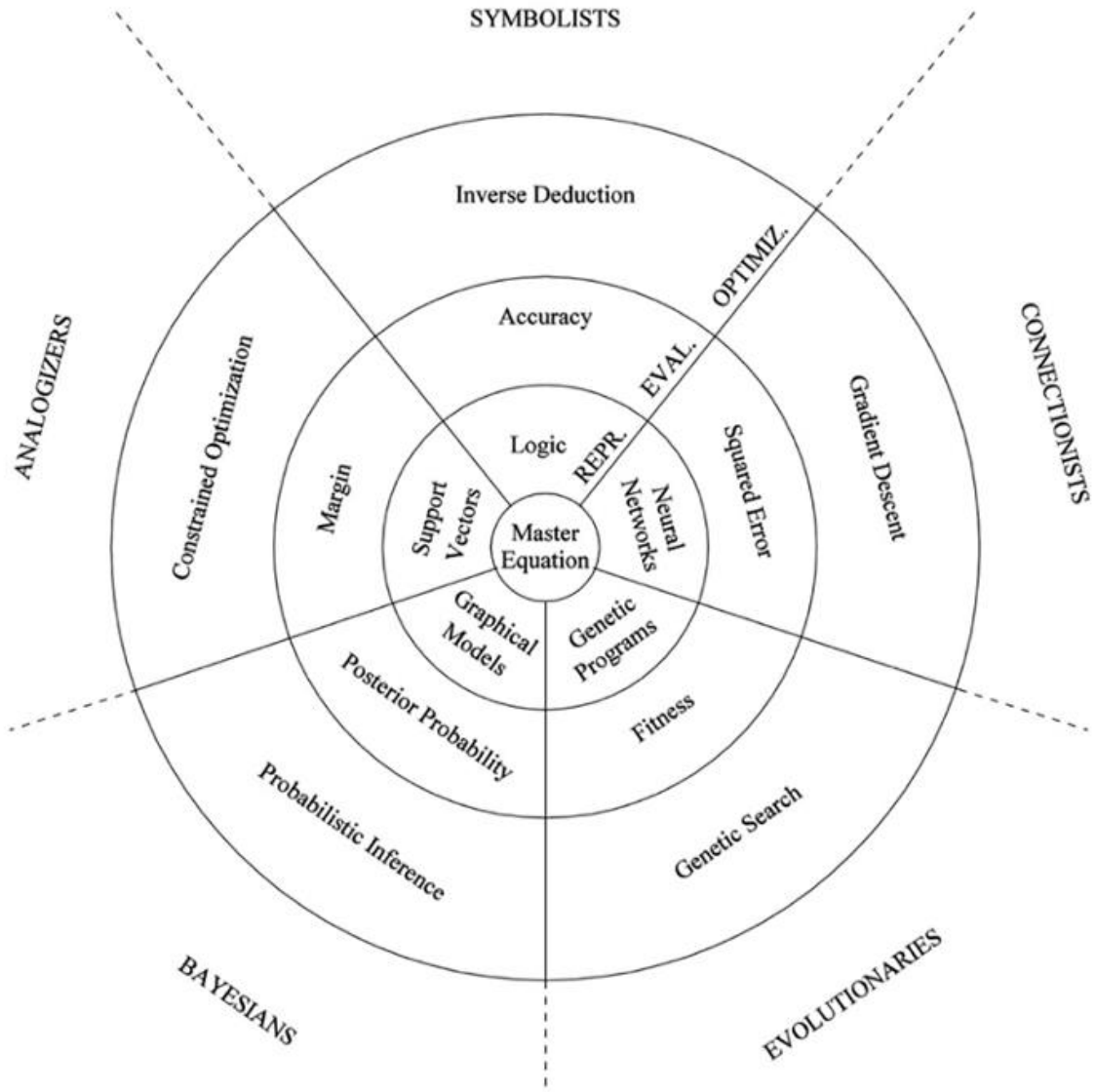
Custom instructions are now available to users in the EU & UK (August 21, 2023)

Custom instructions are now available to users in the European Union & United Kingdom.

To add your instructions:

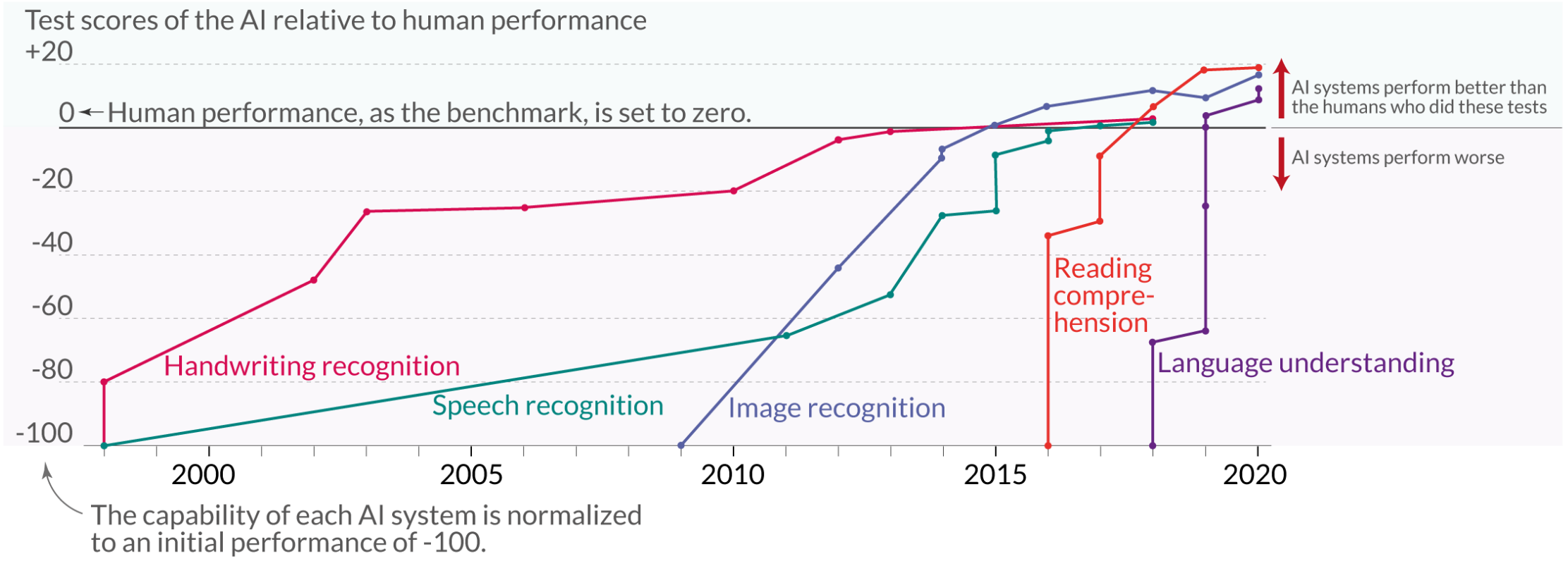
- Click on your name
- Select 'Custom instructions'

Schools of AI



Language and image recognition capabilities of AI systems have improved rapidly

Our World in Data



Data source: Kiela et al. (2021) – Dynabench: Rethinking Benchmarking in NLP

OurWorldinData.org – Research and data to make progress against the world's largest problems.

from ELIZA to Bard

Early chatbots struggled to understand human language. Over time, machine learning (ML) advancements have ushered in chatbots that can process natural language, learn from experience and generate full-length articles.

1966 — ELIZA
A computer scientist at MIT develops ELIZA, the first chatbot.

1995 — A.L.I.C.E.
Inspired by ELIZA, Richard Wallace creates a similar yet more complex chatbot.

2001 — SmarterChild
ActiveBuddy develops a chatbot to interact with AOL Instant Messenger users.

Basic chatbots

The first generation of chatbots used decision trees and simple keyword-recognition capabilities to generate scripted responses. While these chatbots can only process a strict set of inquiries, they continue to help modern contact centers answer customers' frequently asked questions.

Conversational agents

These tools use advanced natural language processing and ML to understand complex human language, process voice commands and learn from past interactions. This kind of chatbot can remind contact center agents of upcoming appointments and answer complex customer questions.

2010 — IBM Watson
IBM develops Watson, a question-answering computer system, to compete on the TV show, *Jeopardy!*

2011 — Siri
Apple incorporates virtual assistant, Siri, into its iPhone 4S.

2014 — Alexa
Amazon releases a virtual assistant, Alexa, alongside the company's Echo speaker.

2021 — Jasper AI
Jasper AI launches a generative copywriting platform for business users.

2022 — ChatGPT
OpenAI releases a free, general-purpose generative AI chatbot to the public.

2023 — Bard
Google releases its own generative AI chatbot to the public.

Generative AI chatbots

Advancements in ML, such as transformers, have let developers train ML models on massive data sets to create generative AI chatbots. In contact centers, these chatbots can help agents compose emails, summarize past customer conversations and draft social media responses.

Types of chatbot technology

As chatbots have evolved, they've broken off into smaller subsets: conversational agents, virtual assistants, generative AI chatbots and generative AI assistants.

Chatbots
Programs that simulate human conversation

Conversational agents
Advanced chatbots that can complete tasks and learn from past interactions

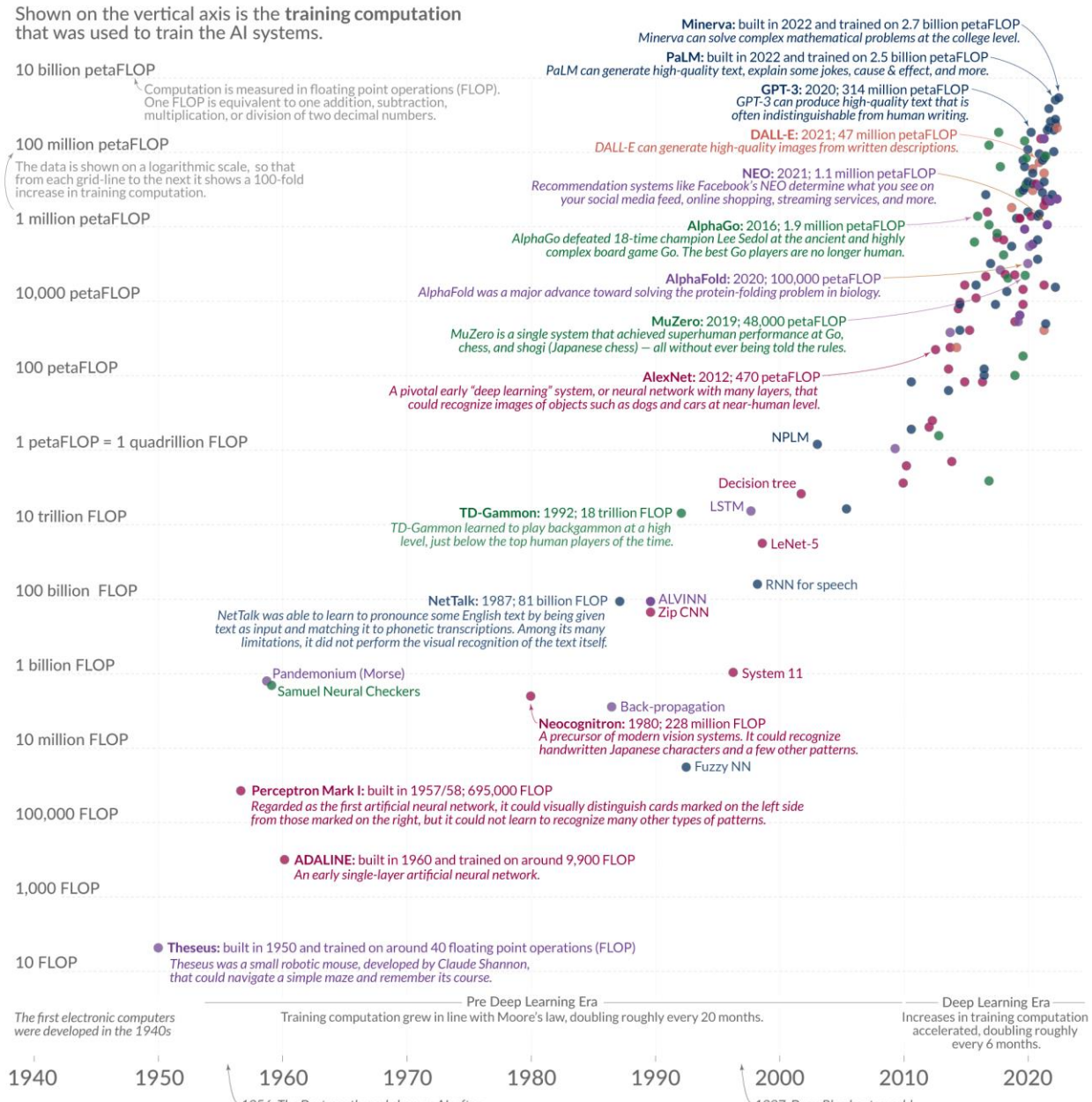
- **Virtual assistants** — Conversational agents personalized for individuals
- **Generative AI chatbots** — Conversational agents that use transformers to generate complex content
- **Generative AI assistants** — Generative AI chatbots trained on an individual's data

DESIGN: LINDA KOURY; ICON: YLIVDESIGN/ADOBE STOCK

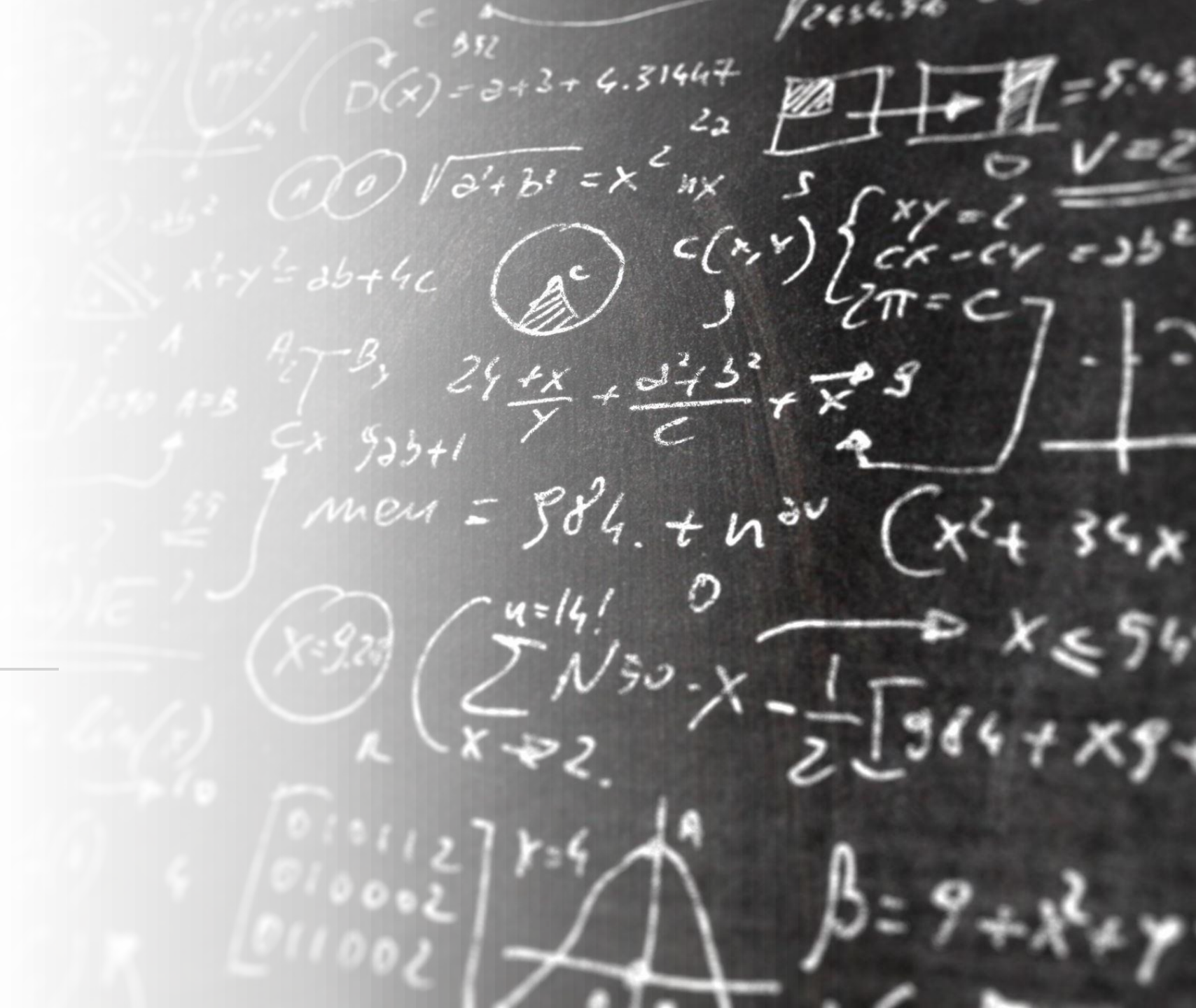
The rise of artificial intelligence over the last 8 decades: As training computation has increased, AI systems have become more powerful

Our World in Data

The color indicates the domain of the AI system: Vision · Games · Drawing · Language · Other

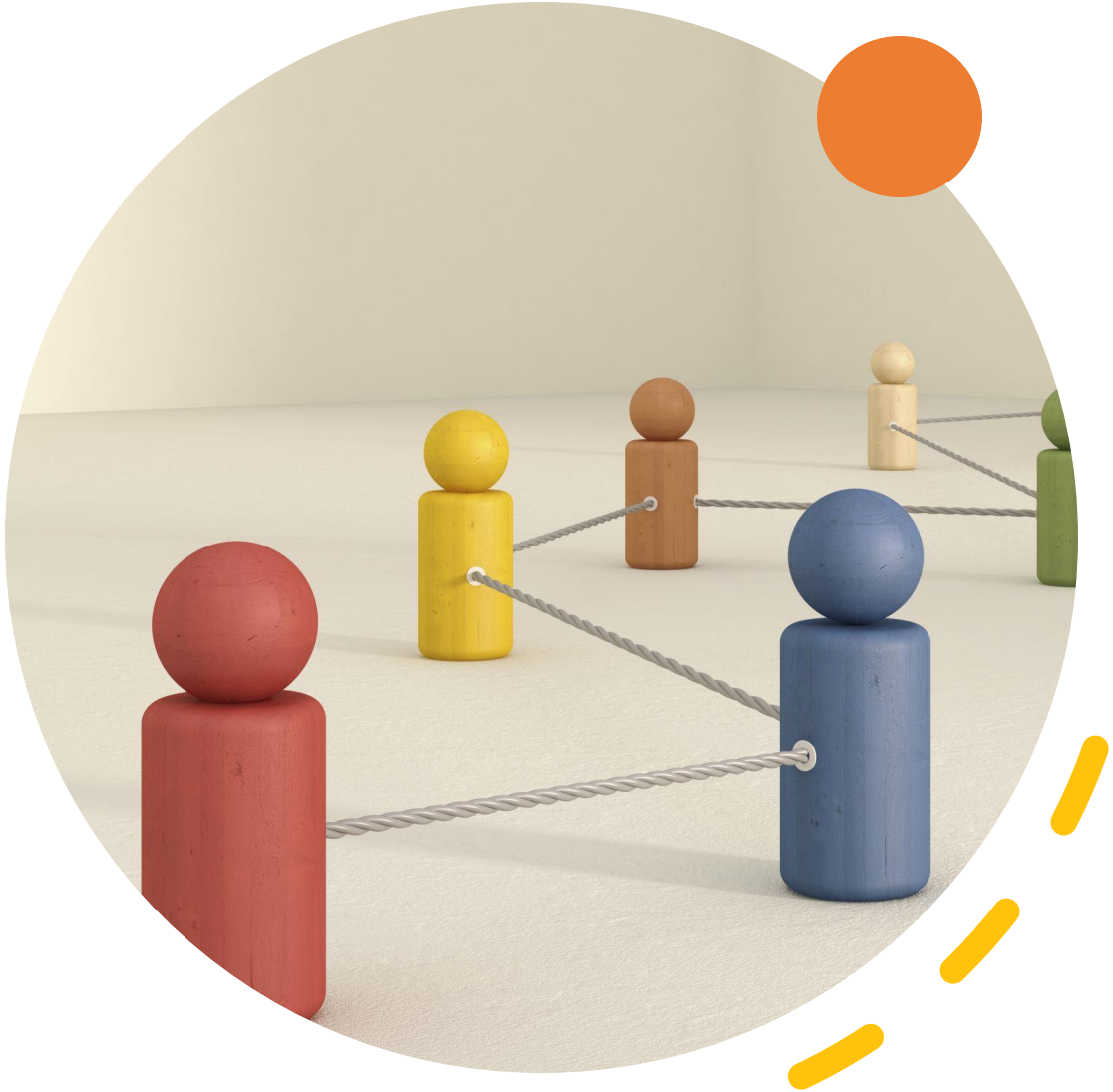


Elements of Large Language Models



3 Phases of training an LLM

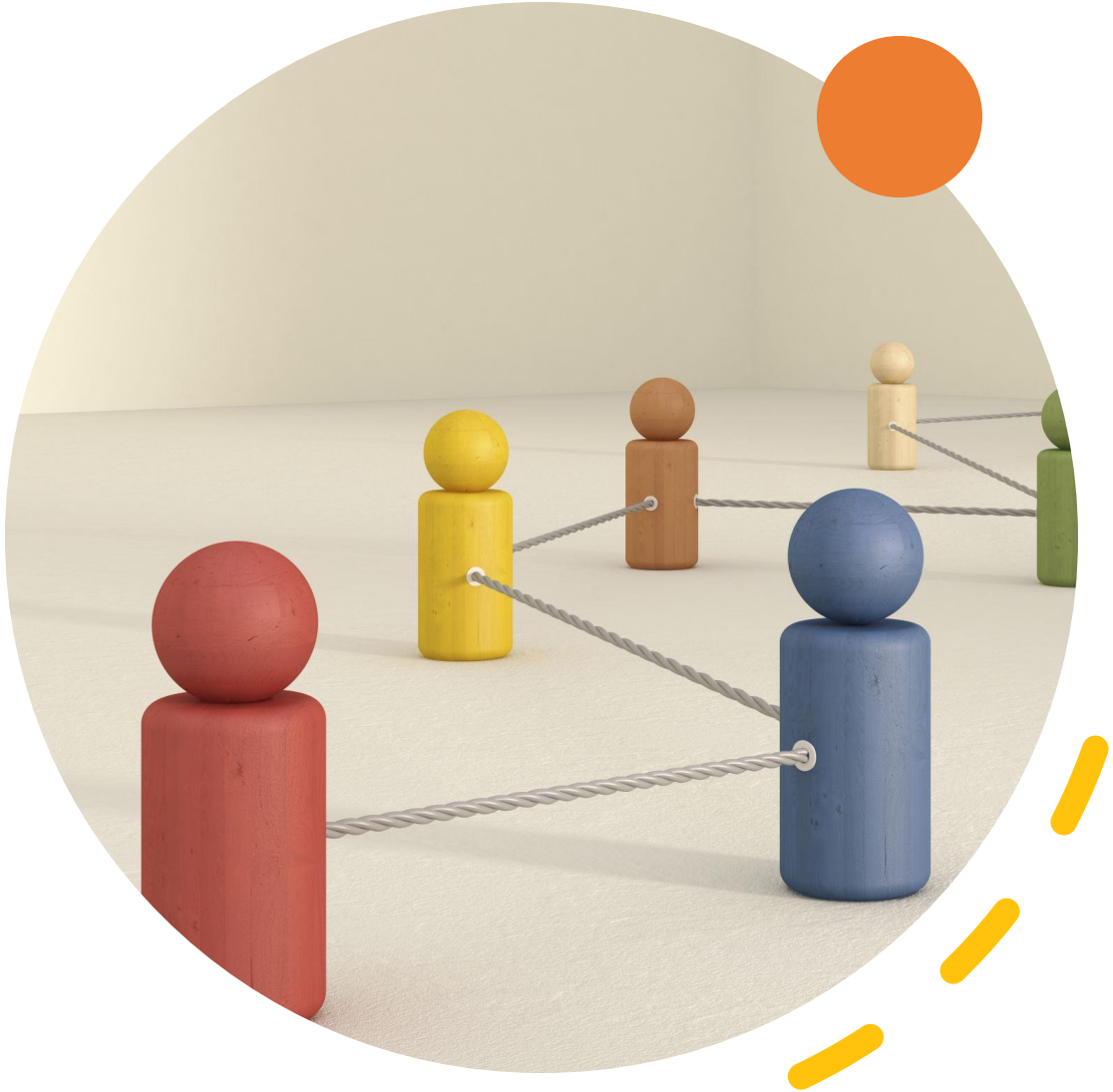
- Pre-training
- Supervised Fine-tuning (SFT)
- Reinforcement Learning from Human Feedback (RLHF)

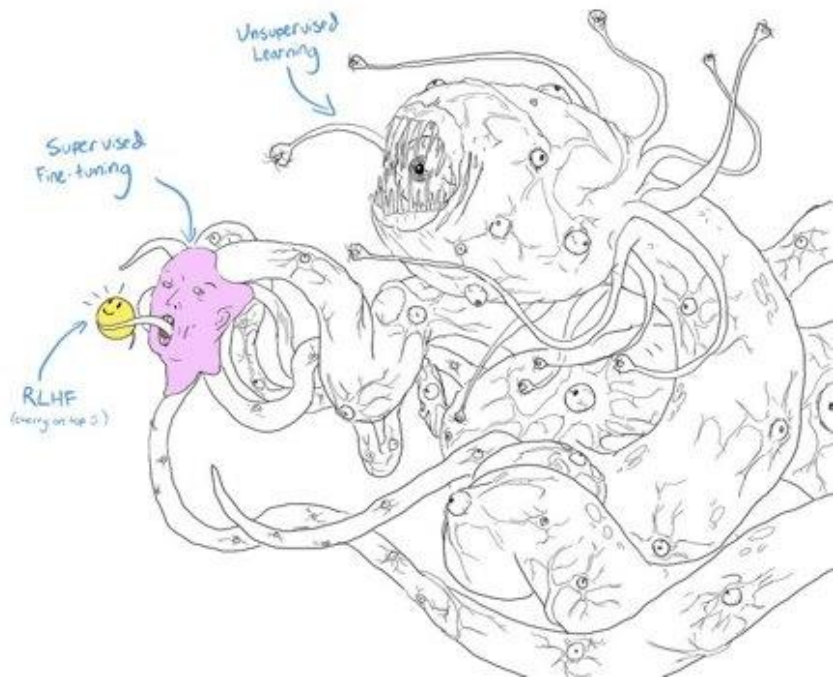


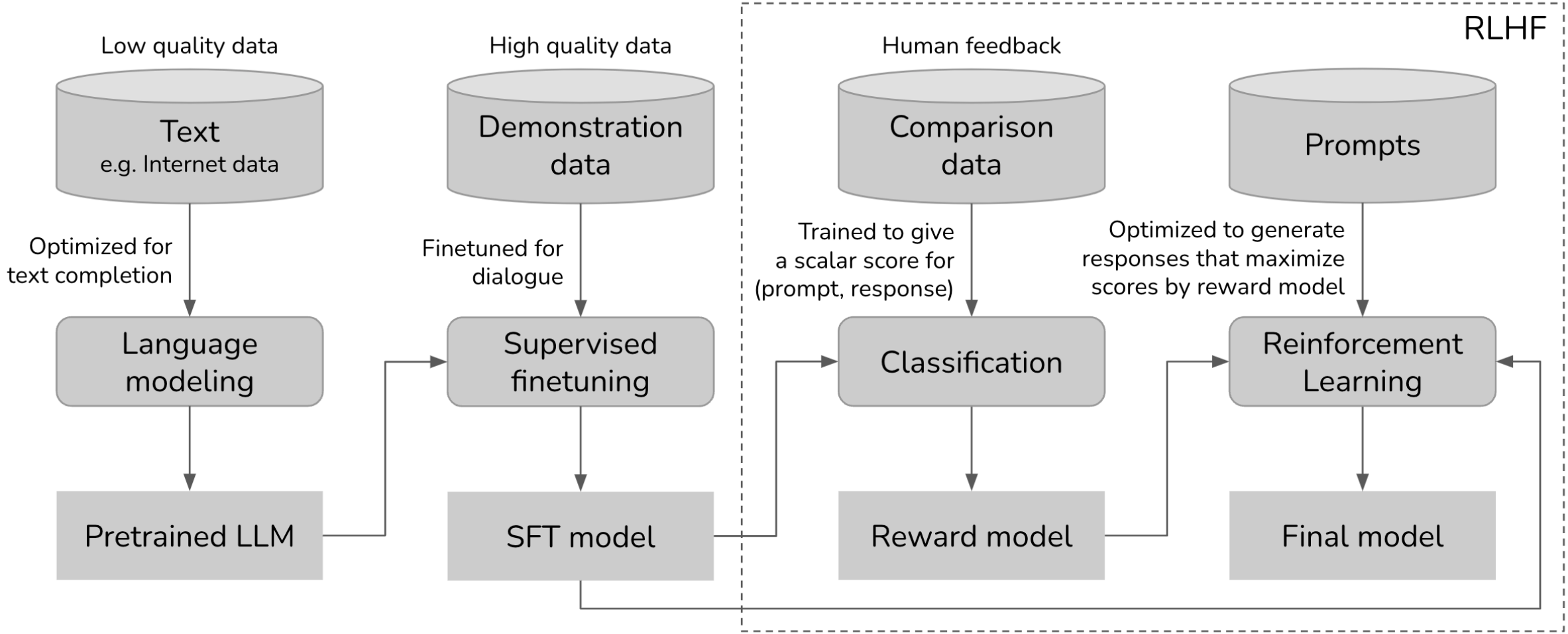
3 Phases of training an LLM

- Pre-training
- Supervised Fine-tuning (SFT)
- Reinforcement Learning from Human Feedback (RLHF)

The goal in the next few hours is for you to understand the steps in practice.




Unsupervised
Learning
Supervised
Fine-tuning
RLHF



Scale May '23	**>1 trillion** tokens	**10K - 100K** (prompt, response)	**100K - 1M** comparisons (prompt, winning_response, losing_response)	**10K - 100K** prompts
Examples **Bolded**: open sourced	GPT-x, Gopher, **Falcon**, LLaMa, **Pythia**, **Bloom**, **StableLM**	**Dolly-v2, Falcon-Instruct**		InstructGPT, ChatGPT, Claude, **StableVicuna**

Phase 1: Pretraining for completion

Task: language modeling

Data Quality: Low

Data Scale: trillions of tokens for large models

Output: Large Language Model (LLM)

Phase 2: Supervised Finetuning (SFT) for dialogue

Task: language modeling

Data Quality: high, formatted as (prompt, response)

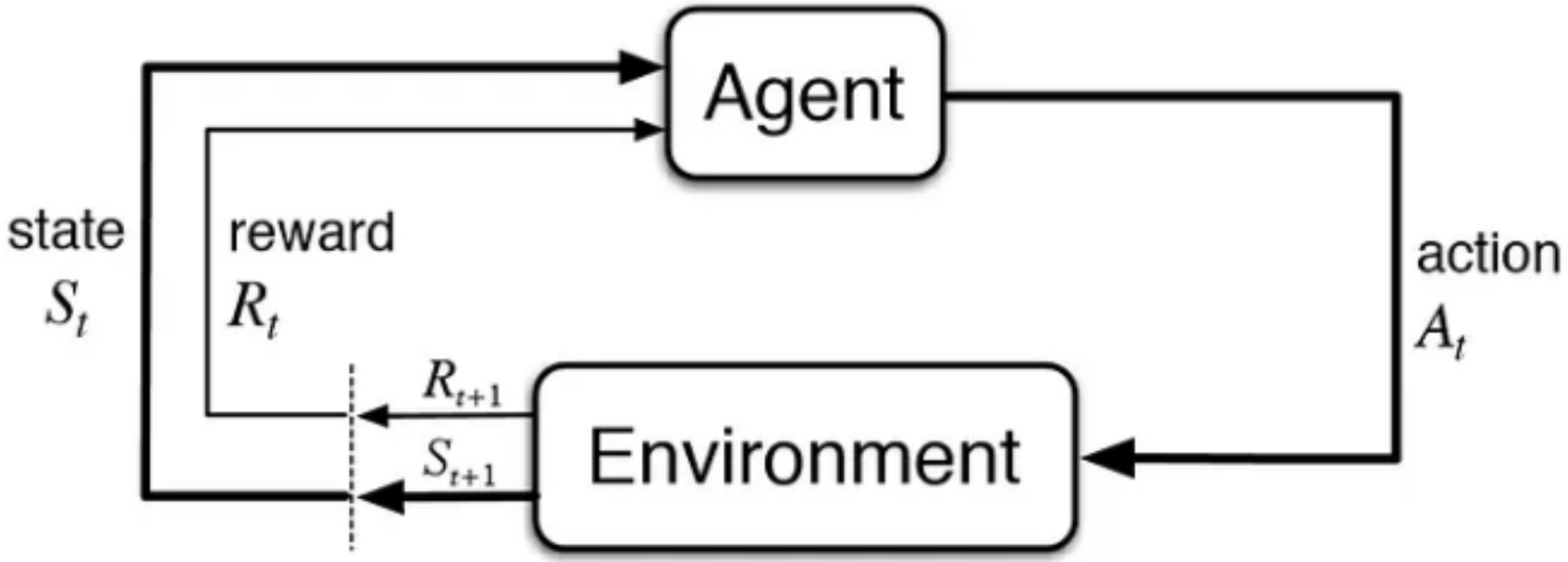
Data Scale: 10,000 to 100,000 (prompt, response) pairs

Input: prompt
Output: response
Loss function: minimize the cross entropy

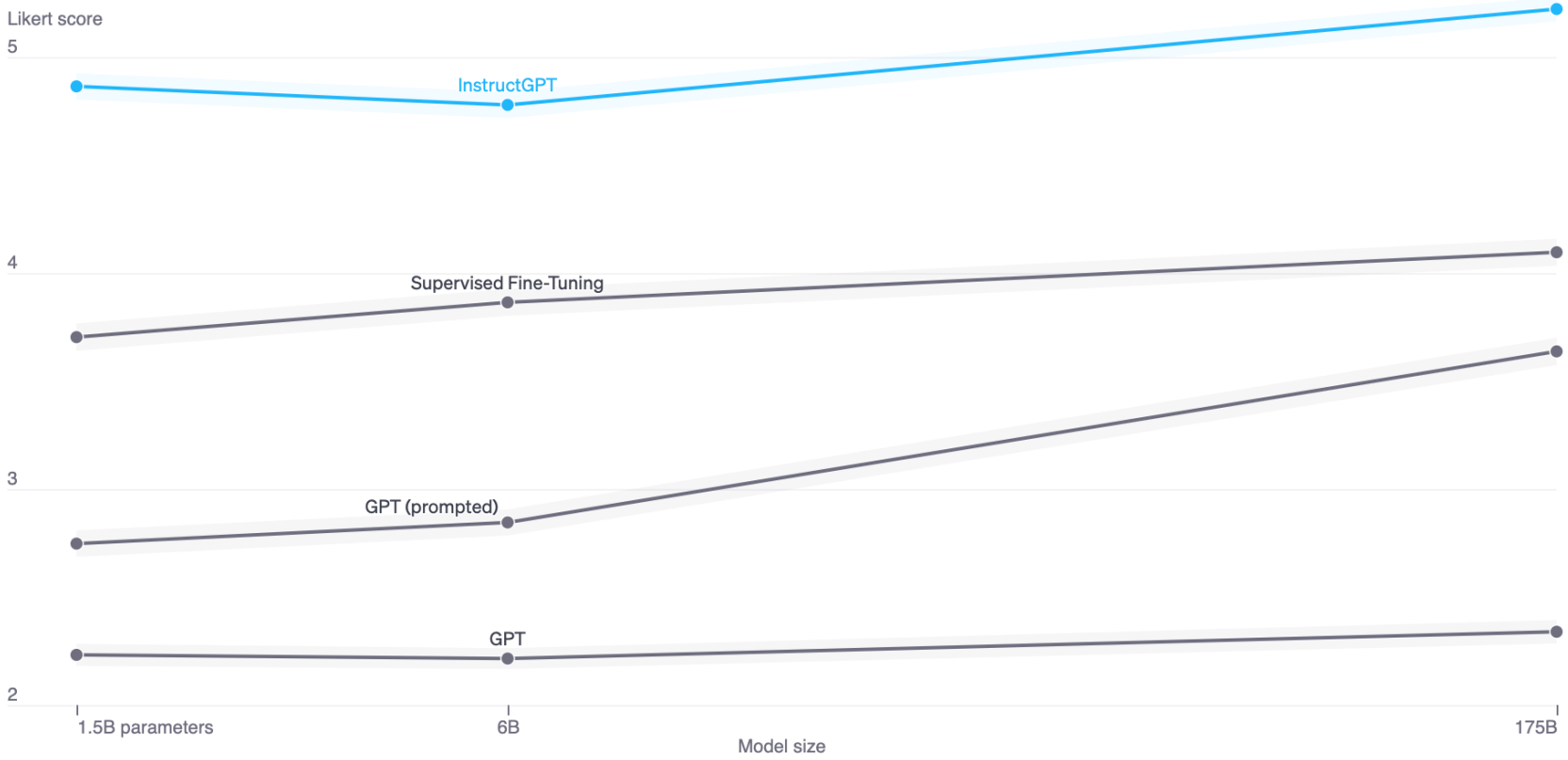
Phase 2: SFT

Prompt	Response
Serendipity means the occurrence and development of events by chance in a happy or beneficial way. Use the word in a sentence.	Running into Margaret and being introduced to Tom was a fortunate stroke of serendipity.
ELI5: What's the cause of the "anxiety lump" in our chest during stressful or disheartening experiences?	The anxiety lump in your throat is caused by muscular tension keeping your glottis dilated to maximize airflow. The clenched chest or heartache feeling is caused by the vagus nerve which tells the organs to pump blood faster, stop digesting, and produce adrenaline and cortisol.
Create a shopping list from this recipe: Trim the ends off zucchini. Cut zucchini in half lengthwise; scoop out pulp, leaving 1/2-in. shells. Finely chop pulp. In a skillet, cook beef, zucchini pulp, onion, mushrooms and peppers over medium heat until meat is no longer pink; drain. Remove from the heat. Add 1/2 cup cheese, ketchup, salt and pepper; mix well. Spoon into the zucchini shells. Place in a greased 13x9-in. baking dish. Sprinkle with remaining cheese.	Zucchini, beef, onion, mushroom, peppers, cheese, ketchup, salt, pepper

Reinforcement Learning Primer



Phase 3: RLHF



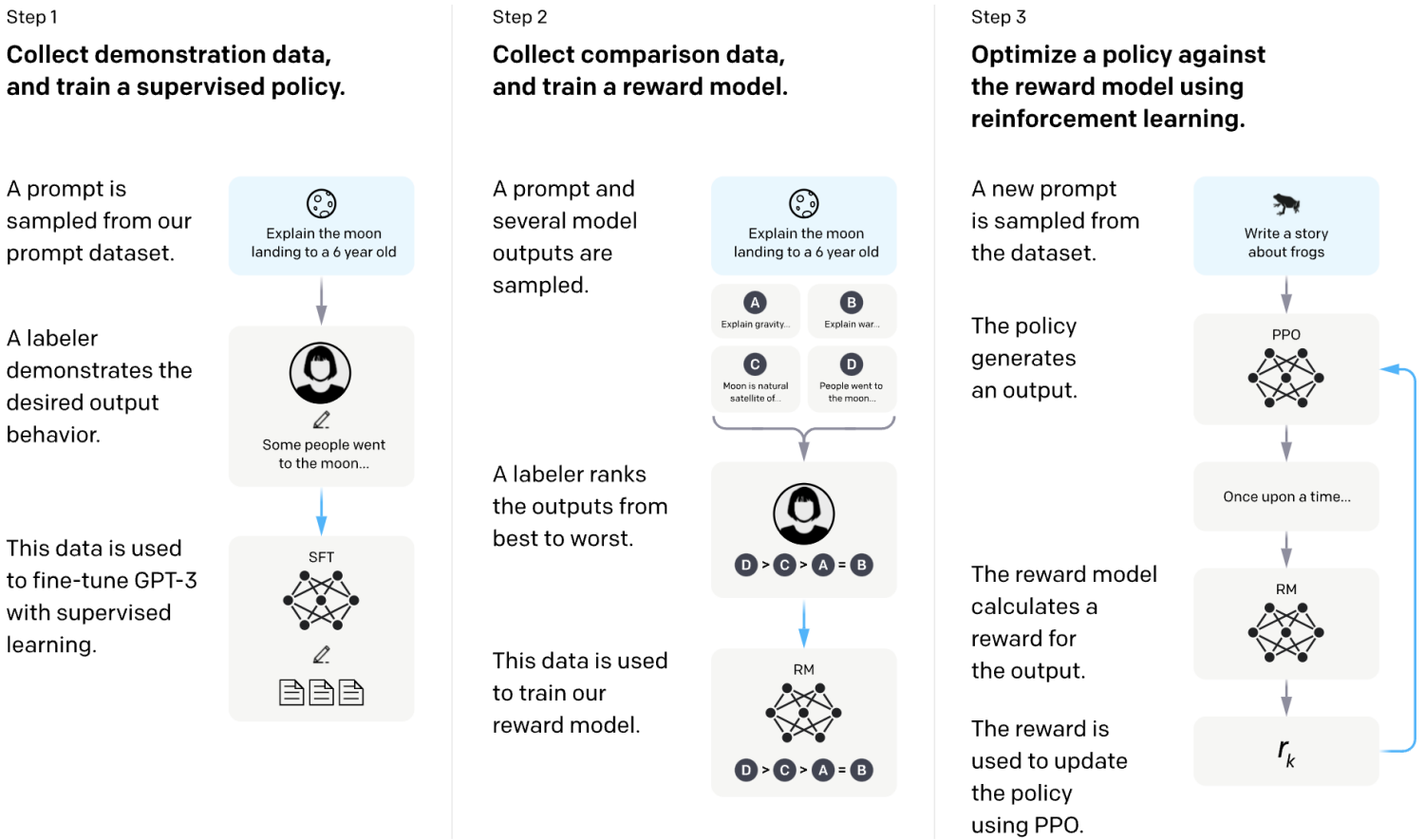
Quality ratings of model outputs on a 1–7 scale (y-axis), for various model sizes (x-axis), on prompts submitted to InstructGPT models on our API. InstructGPT outputs are given much higher scores by our labelers than outputs from GPT-3 with a few-shot prompt and without, as well as models fine-tuned with supervised learning. We find similar results for prompts submitted to GPT-3 models on the API.

Phase 3.1: Reward Model

Training Data: **high, formatted as (prompt,** winning_response, loosing_response)

Data Scale: 100,000 **to** 1,000,000 examples

Phase 3.2: Finetuning using Reward Model



Phase 3.2: Finetuning using Reward Model

Task: Reinforement Learning

Action Space: Vocabulary of all the tokens

Observation space: the distribution over all possible prompts

Reward function: reward model

Training Data: randomly selected prompts

Data Scale: 10,000 to 100,000 prompts

The Debate

2014

Elon Musk: artificial intelligence is our biggest existential threat

The AI investor says that humanity risks 'summoning a demon' and calls for more regulatory oversight

Stephen Hawking warns artificial intelligence could end mankind

2 December 2014 · Comments

Open Letters (2015)

https://futureoflife.org/data/documents/research_priorities.pdf

As capabilities in these areas and others cross the threshold from laboratory research to economically valuable technologies, a virtuous cycle takes hold whereby even small improvements in performance are worth large sums of money, prompting greater investments in research. There is now a broad consensus that AI research is progressing steadily, and that its impact on society is likely to increase. The potential benefits are huge, since everything that civilization has to offer is a product of human intelligence; we cannot predict what we might achieve when this intelligence is magnified by the tools AI may provide, but the eradication of disease and poverty are not unfathomable. Because of the great potential of AI, it is important to research how to reap its benefits while avoiding potential pitfalls.

Open Letters (2023)

Signatories

Yoshua Bengio, Founder and Scientific Director at Mila, Turing Prize winner and professor at University of Montreal

Stuart Russell, Berkeley, Professor of Computer Science, director of the Center for Intelligent Systems, and co-author of the standard textbook "Artificial Intelligence: a Modern Approach"

Elon Musk, CEO of SpaceX, Tesla & Twitter

Steve Wozniak, Co-founder, Apple

Yuval Noah Harari, Author and Professor, Hebrew University of Jerusalem.

Emad Mostaque, CEO, Stability AI

Andrew Yang, Forward Party, Co-Chair, Presidential Candidate 2020, NYT Bestselling Author, Presidential Ambassador of Global Entrepreneurship

John J Hopfield, Princeton University, Professor Emeritus, inventor of associative neural networks

Valerie Pisano, President & CEO, MILA

Connor Leahy, CEO, Conjecture

Jaan Tallinn, Co-Founder of Skype, Centre for the Study of Existential Risk, Future of Life

Pause Giant AI Experiments: An Open Letter

We call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4.

Signatures
33712

Add your signature

Published
March 22, 2023

IDEAS • TECHNOLOGY

Pausing AI Developments Isn't Enough. We Need to Shut it All Down

Eliezer Yudkowsky, TIME Magazine, 2023

Open Letters (2023)

Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war.

Signatories:

☑ AI Scientists ☑ Other Notable Figures

Geoffrey Hinton
Emeritus Professor of Computer Science, University of Toronto

Yoshua Bengio
Professor of Computer Science, U. Montreal / Mila

Demis Hassabis
CEO, Google DeepMind

Sam Altman
CEO, OpenAI

Dario Amodei
CEO, Anthropic

Dawn Song
Professor of Computer Science, UC Berkeley

Ted Lieu
Congressman, US House of Representatives

Bill Gates
Gates Ventures

Ya-Qin Zhang
Professor and Dean, AIR, Tsinghua University

Ilya Sutskever
Co-Founder and Chief Scientist, OpenAI

Igor Babuschkin
Co-Founder, xAI

Shane Legg
Chief AGI Scientist and Co-Founder, Google DeepMind

Martin Hellman

Geoffery Hinton Quits (2023)

Some Books

Human Compatible, Stuart Russell

A Brief History of Artificial Intelligence, Michael Wooldridge

Englignment Now, Steven Pinker

The Alignment Problem, Brian Christian

The Master Algorithm, Pedro Domingos

Super Intelligence: Paths, Dangers, Strategies, Nick Bostrom

Algorithms to Live By: The Computer Science of Human Decisions, Brian Christian

Rebooting AI: Building Artificial Intelligence We Can Trust, Gary Marcus and Ernest Davis

Life 3.0: Being Human in the Age of Artificial Intelligence, Max Tegmark

AI Snake Oil, Arvind Narayanan and Sayash Kapoor

The End