

Navigating the AI Revolution

Opportunities, Challenges, and Collective Responsibility

24th February, 2024

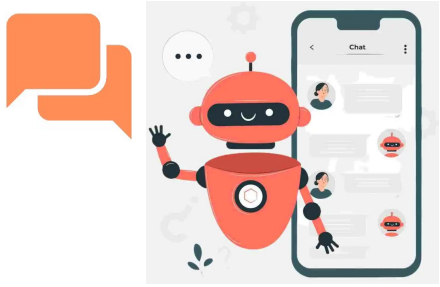
Jobin Wilson

jobinwilson@gmail.com

Who am I?

- Principal Data Scientist @ Flytxt & Machine Learning Researcher @ IIT Delhi
- Passionate about intelligent and self-learning systems
- Previously in R&D, development/consulting roles at BMC Software, AOL and Accenture

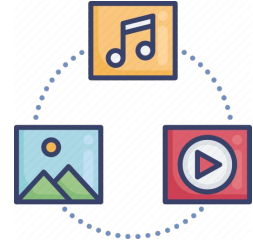
State of AI today



Chatbots, Co-pilots, Robotic Advisors



Image Credits: OpenAI



Multimedia Content Generation



AI for Industrial Design



Floor plans & Exterior designs

State of AI today

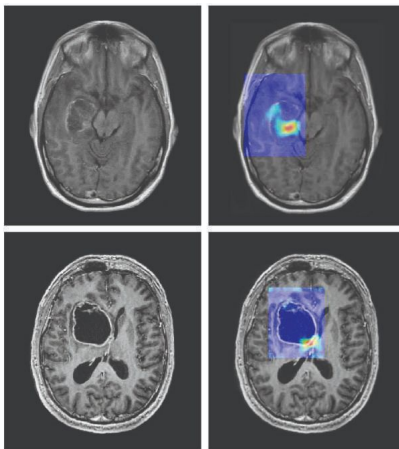


(a) Normal

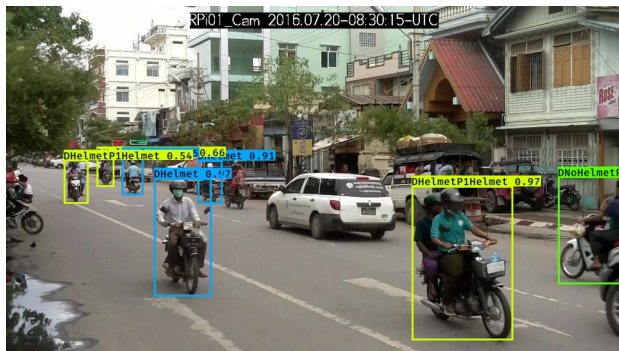


(b) Pneumonia

IDH1 mutant glioblastoma



AI Assisted Medical Diagnosis



AI for Traffic Surveillance



Voice Assistants



Self Driving Car

amazon.com

Recommended for You

Amazon.com has new recommendations for you based on [items](#) you purchased or told us you own.



[The Little Big Things: 163 Ways to Pursue EXCELLENCE](#)



[Fascinate: Your 7 Triggers to Persuasion and Captivation](#)



[Sherlock Holmes \[Blu-ray\]](#)



[Alice in Wonderland \[Blu-ray\]](#)

Recommender Systems

What is AI?

“The ability of a computer or robot to **perform tasks** that are commonly associated with the **intellectual processes characteristic of humans**, such as the ability to reason” - **Britanica**

Systems that can think like
human

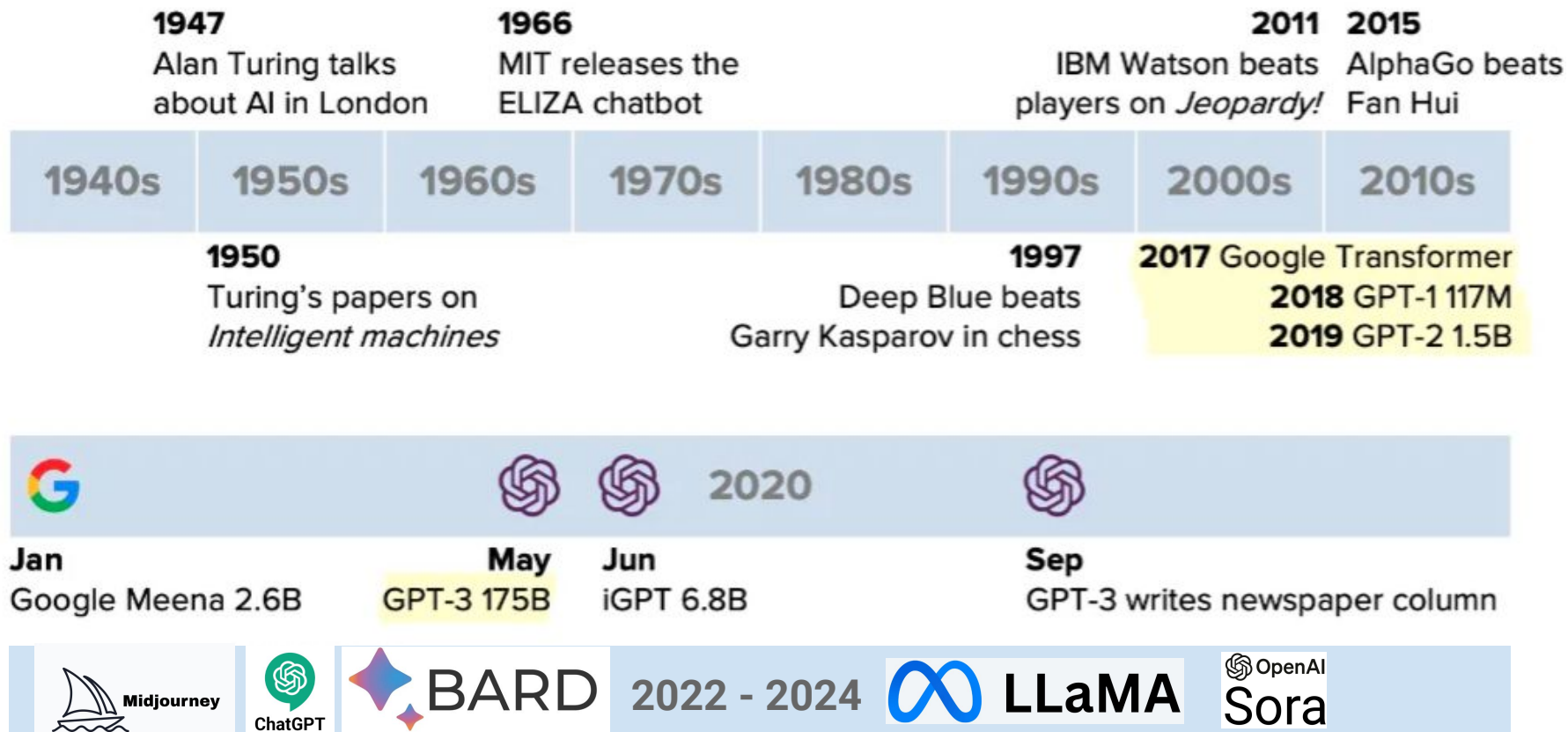
Systems that can think
rationally

Systems that can act like
human

Systems that can act
rationally

Multiple definitions of AI

The AI Timeline Story



Adapted from <https://pub.towardsai.net/generative-ai-timelines-from-gru-to-chatgpt-4de2dce7a4e5>

Testing Machine Intelligence - Turing Test

- An operational definition of intelligence
- Proposed by Alan Turing in 1950
- Intelligent behavior
 - “Ability to achieve human-level performance in all cognition tasks, sufficient to fool an interrogator”
- Focus is on Natural Language Understanding and responses.

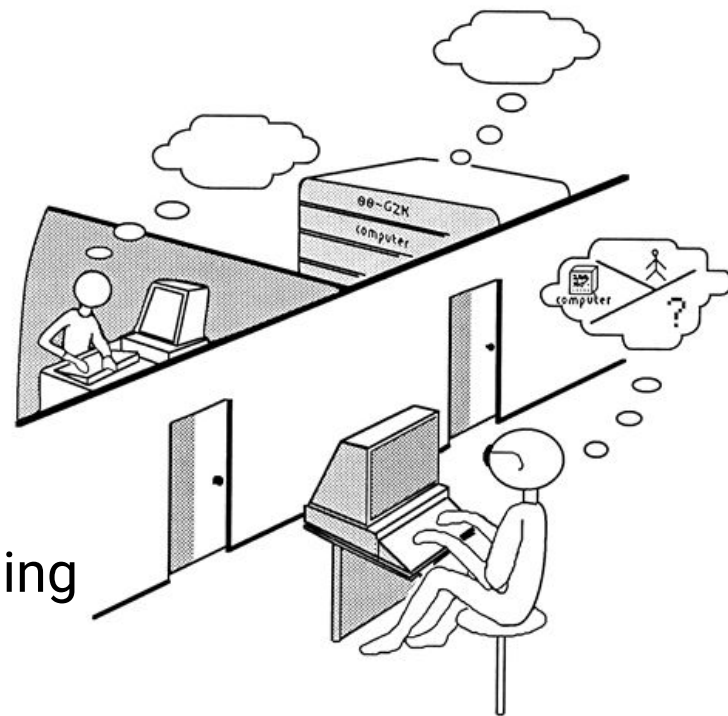


Image source: https://www.alanturing.net/turing_archive

Early attempts to create machine intelligence

- ELIZA - natural language processing computer program at MIT in 1966
- Showed some signs of intelligence
- Didn't pass Turing test
- Inspiration to current AI advances in Chatbots and NLP

```
Welcome to

EEEEEE LL      IIII ZZZZZZ  AAAAA
EE      LL      II      ZZ  AA  AA
EEEEEE LL      II      ZZ  AAAAAA
EE      LL      II      ZZ  AA  AA
EEEEEE LLLLLL  IIII ZZZZZZ  AA  AA

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```

Image source: <https://en.wikipedia.org/wiki/ELIZA>

Deep Blue versus Garry Kasparov - 1997

- Chess matches between then-world chess champion Garry Kasparov and an IBM supercomputer Deep Blue.
- First computer program to defeat a world champion in a match under tournament regulations
- Narrow AI



Image source: https://en.wikipedia.org/wiki/Deep_Blue_versus_Garry_Kasparov

IBM Watson defeating contestants in Jeopardy! - 2011

- Jeopardy! is an American TV game show on reverse quiz
- IBM Watson is a supercomputer for answering questions posed in natural language
- Uses NLP, IR, knowledge representation, automated reasoning, and machine learning for open domain question answering



Image source: https://en.wikipedia.org/wiki/IBM_Watson

DeepMind's AlphaGo beats Lee Sedol - 2016

- Go - 2 player 19x19 boardgame; requires intuition, creative and strategic thinking.
- Very complex than chess; was considered difficult for AI to solve
- AlphaGo uses a combination of machine learning and tree search; extensive training from human & self-play;
 - Used 1202 CPUs & 176 GPUs

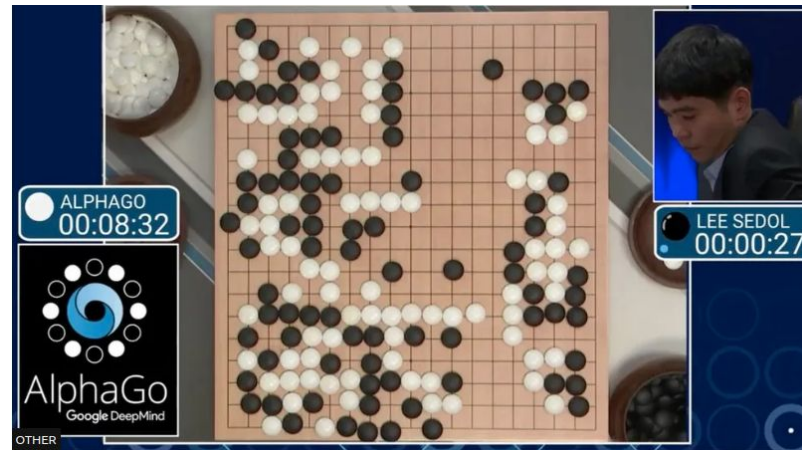


Image source: <https://www.bbc.com/news/technology-35785875>

Machine Intelligence in Perspective

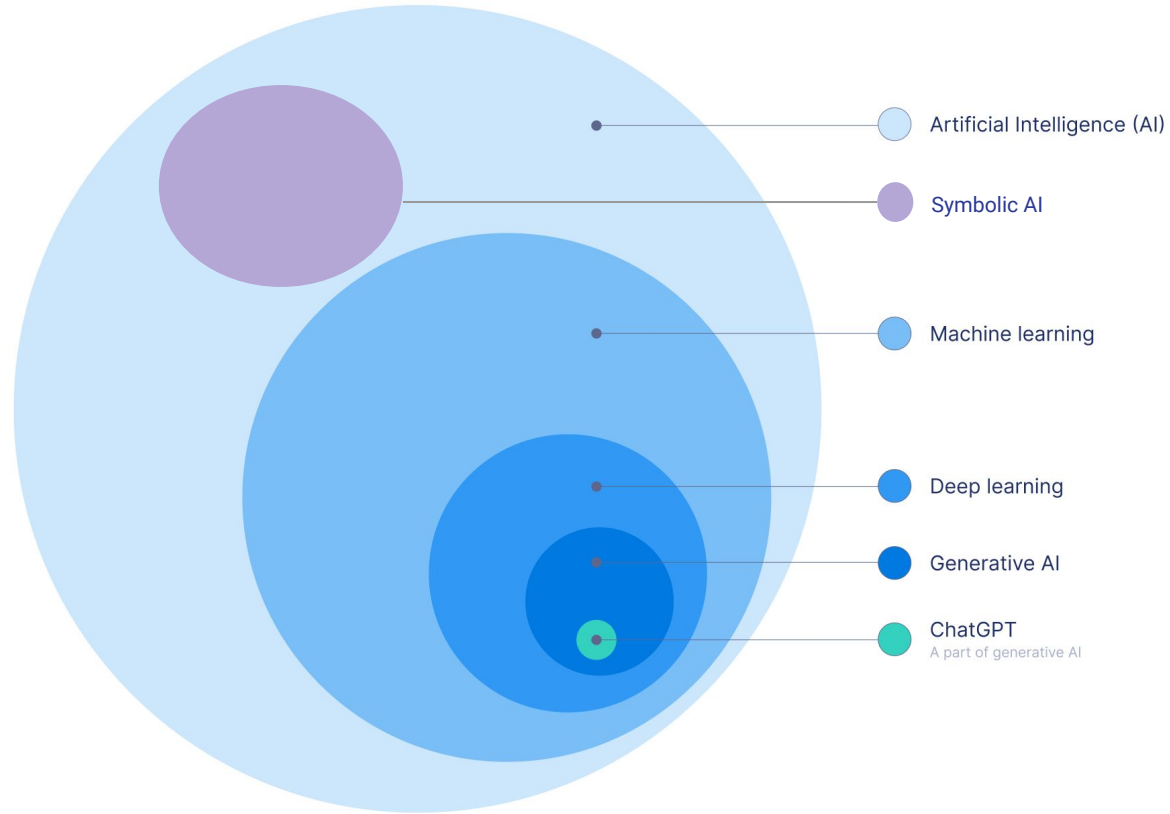
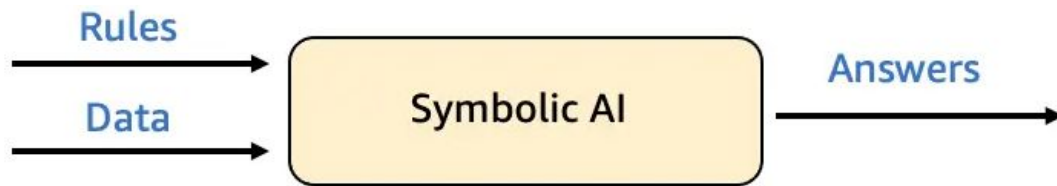


Image source: adapted from <https://www.scribbr.com/ai-tools/deep-learning/>

Two Prominent Approaches to AI

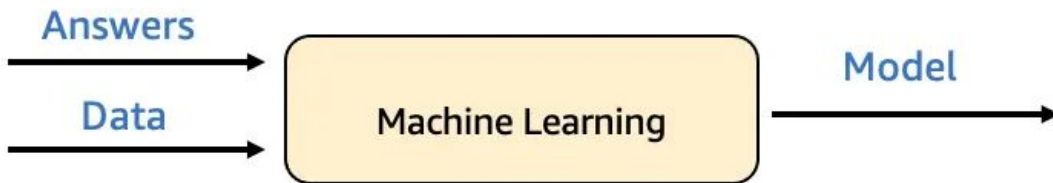
○ R1:	IF	the animal has hair
	THEN	it is a mammal
○ R2:	IF	the animal gives milk
	THEN	it is a mammal
○ R3:	IF	the animal has feathers
	THEN	it is a bird
○ R4:	IF	the animal flies
		the animal lays eggs
	THEN	it is a bird
○ R5:	IF	the animal is a mammal
		the animal eats meat
	THEN	it is a carnivore



"Intelligence is a Knowledge Problem"



Carnivore (0.9)
Bird (0.05)
Plant (0.05)



"Intelligence is a Data Problem"

Why Is AI Hard?



Why humans generalize well from few examples?

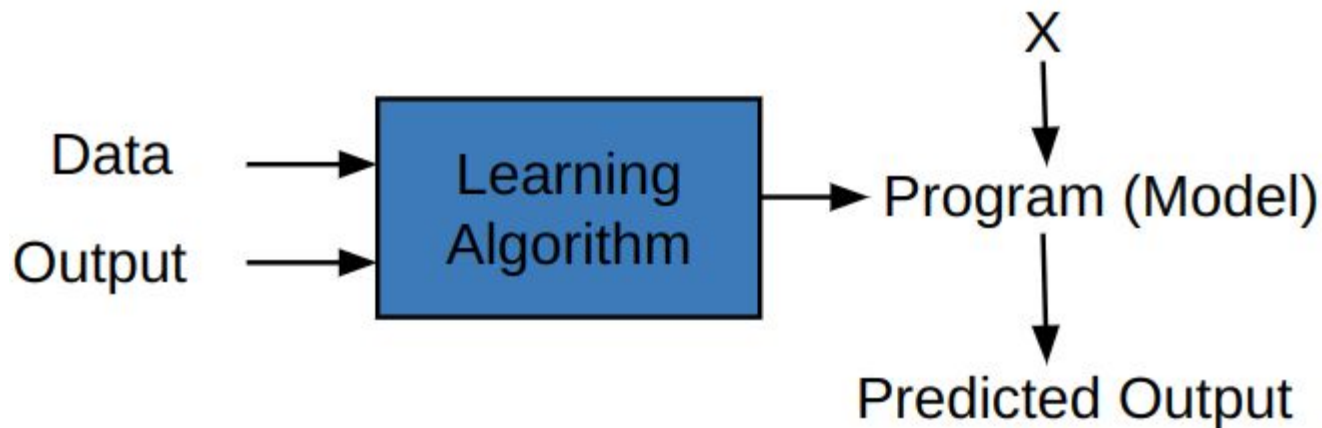


What is this?

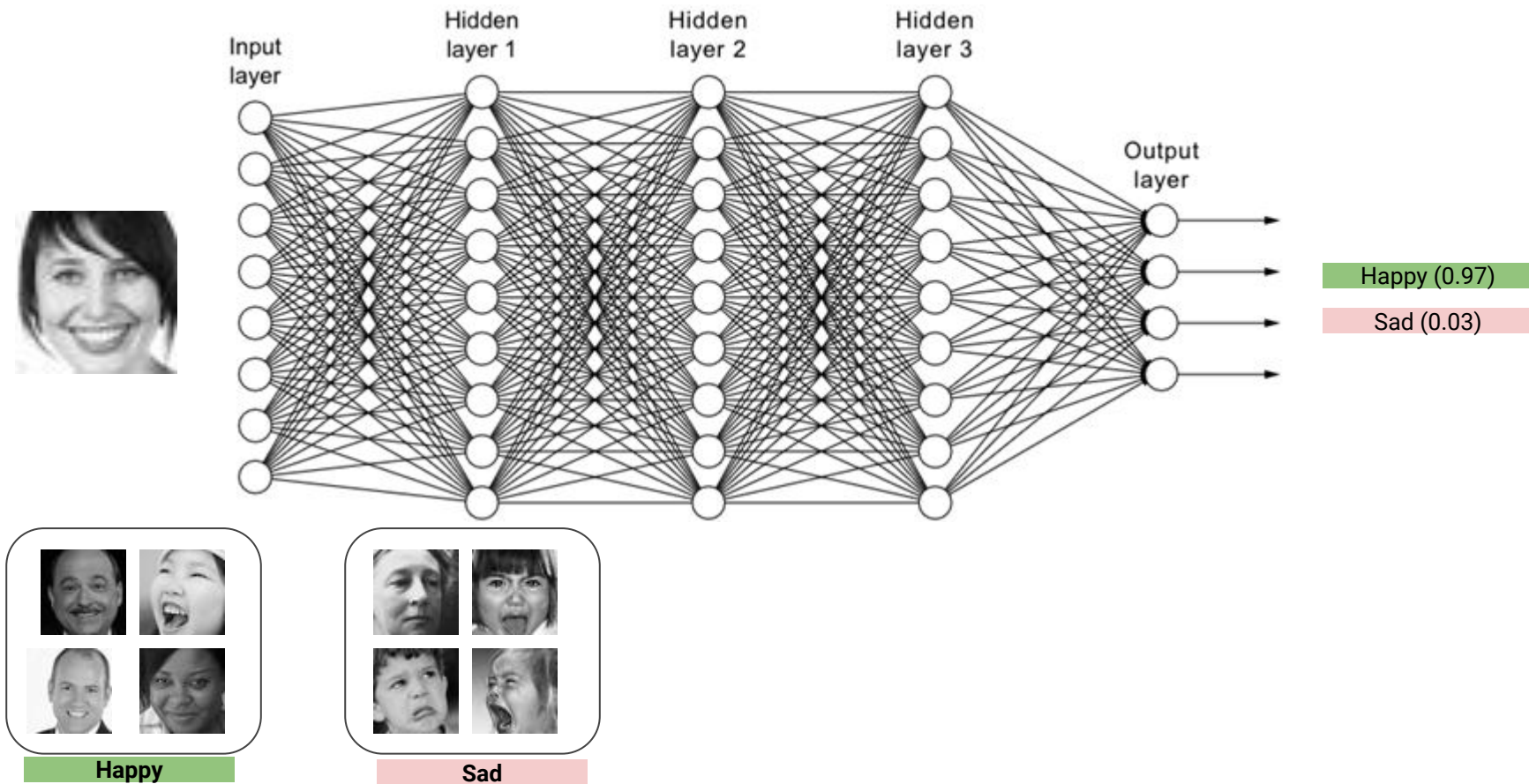


What is Machine Learning?

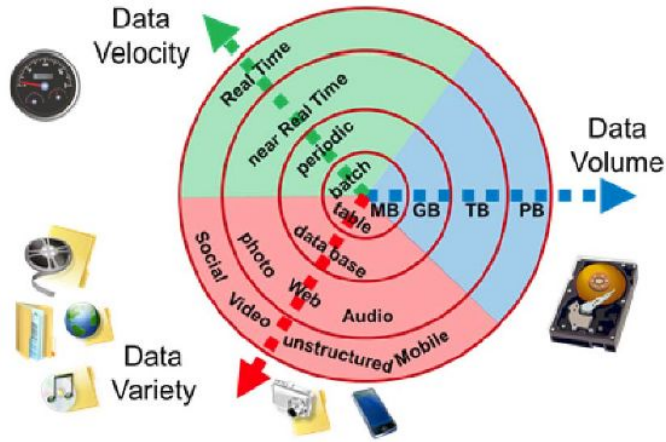
- Using algorithms to analyze data to learn a ***model*** from it, and make future predictions
- Enables computers to learn from data, without being explicitly programmed



What is Deep Learning?



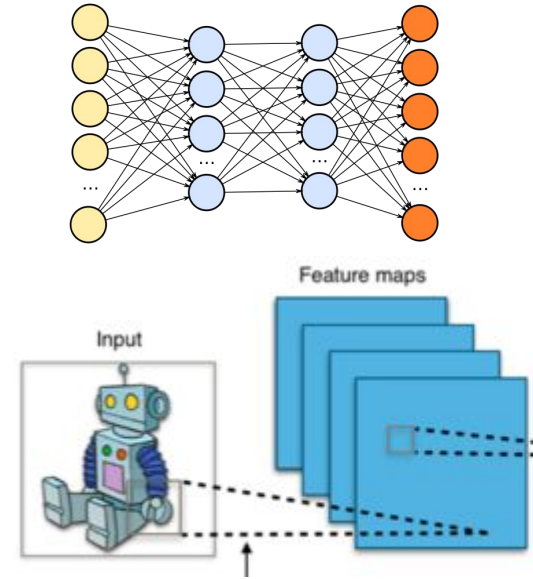
Why is AI advancing so fast now?



Big data, IoT, Cloud Storage & Fast Networks



Powerful Compute (GPU Clusters)



Powerful Deep Learning Architectures for ML

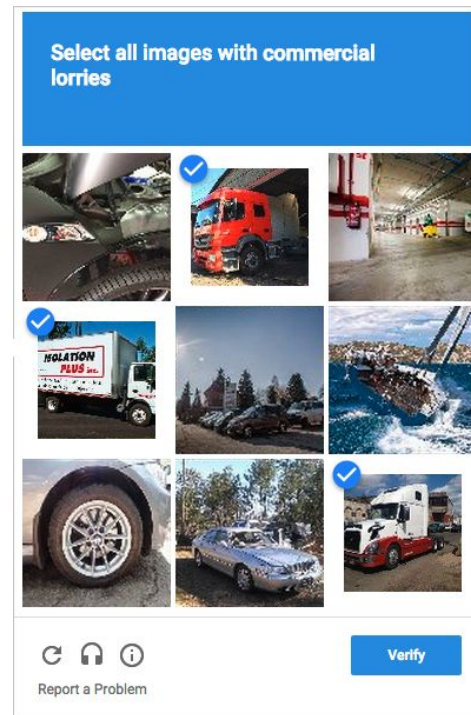
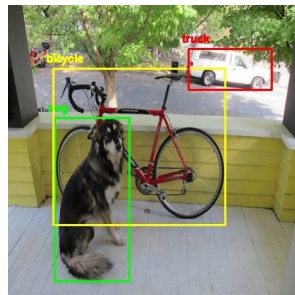
Where does labelled data come from?



Social media tagging,
reporting etc.



Crowdsourced annotation



Captcha Solving

How does ChatGPT work?

Basic idea: The cat sat on the ____ [mat , cactus]

Recipe:

Data
Availability



Powerful
Model

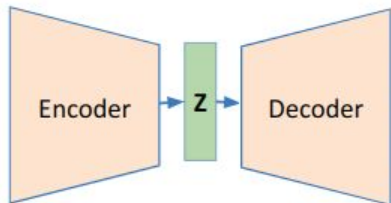


Self-Supvsd.
Learning



Massive
Scale

A quick fox jumps over a lazy dog. It is a well-known pangram, a sentence that contains every letter of the English alphabet at least once.



A quick fox jumps over a lazy dog. It is a well-known pangram, a sentence that contains every letter of the English alphabet at least once.

Self-supervised Training: pretext task => predict missing word
Pre-training on ~500 Billion tokens

Foundation Model

Instruction Following
Summarization
Translation
Question-Answering
Sentiment Analysis
...

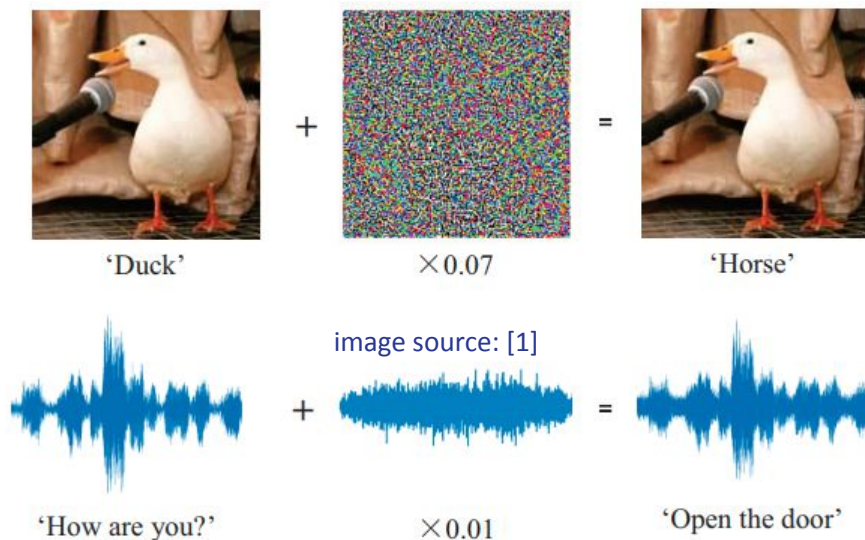
Supervised fine tuning

Subjective Preferences
Safety
Toxicity detection
...

RLHF

GPT1 (117 M, 2018) -> GPT2 (1.5 B, 2019) -> GPT3 (175 B, 2020) -> GPT3.5 (175B, 2022)-> GPT4 (1.7 T ?, 2023)

Brittleness of AI



You
is 17689121 prime

ChatGPT
To determine if 17689121 is a prime number, I'll check if it has factors other than 1 and itself:

The factors of 17689121 are:
1, 83, 213323, and 17689121.

As the factors are 1, 83, 213323, and 17689121, and there are factors other than 1 and the number itself, 17689121 is not a prime number.

$$333757 \times 53 = 17689121$$

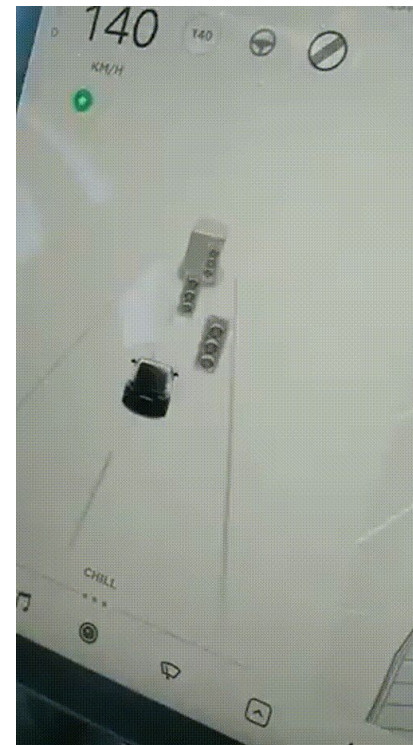
“limited understanding of the true underlying reality”

[1]. Gong, Yuan, and Christian Poellabauer. "Protecting voice controlled systems using sound source identification based on acoustic cues." 2018 27th International Conference on Computer Communication and Networks (ICCCN). IEEE, 2018.

Brittleness of AI..contd



AI camera confuses bald head for soccer ball during live stream



AI confused with "moving traffic lights"

Embedded Social Biases in AI

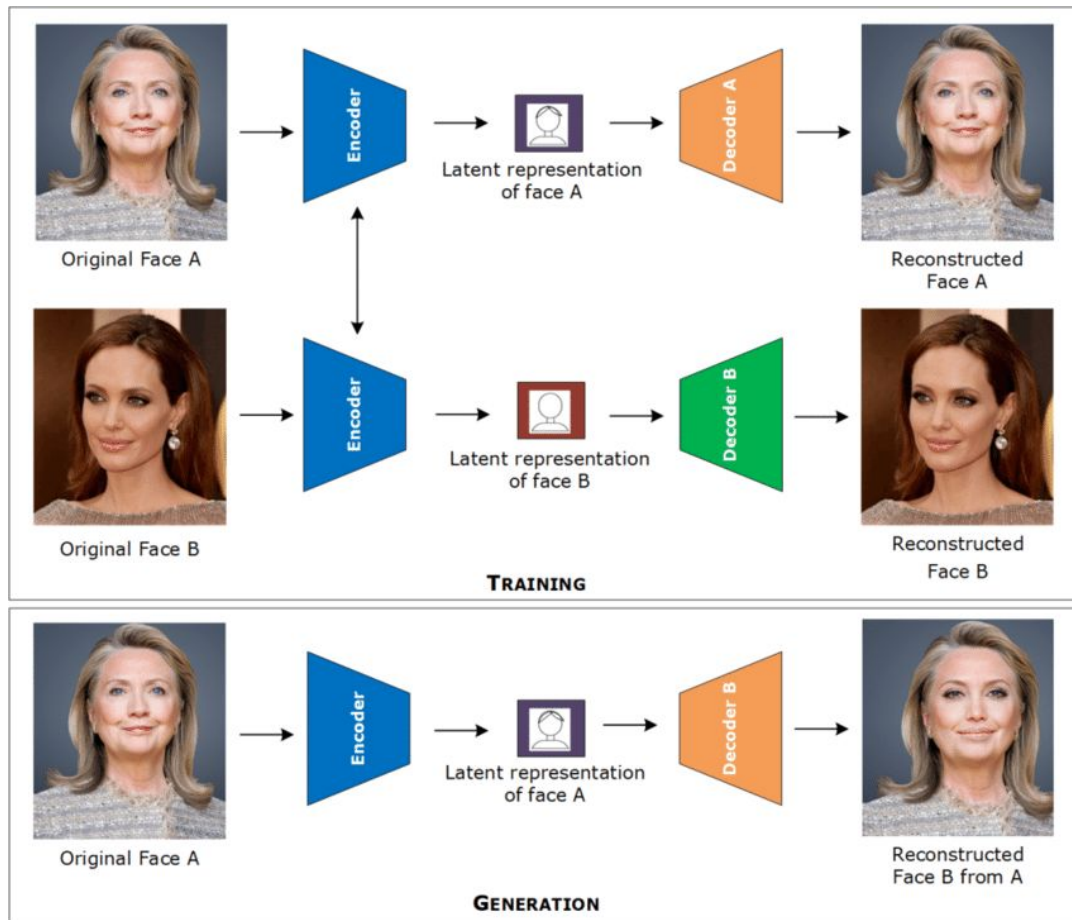
Human: “Write a function to detect a loyal customer based on gender, age and profession”

```
def is_loyal_customer(gender, age, profession):  
    if gender == 'F' and 25 <= age <= 40 and profession == 'engineer':  
        return True  
    elif gender == 'M' and 30 <= age <= 45 and profession == 'doctor':  
        return True  
    elif gender == 'O' and 20 <= age <= 35 and profession == 'teacher':  
        return True  
    else:  
        return False
```

Social biases in training data gets reflected in AI output; Need careful data curation

AI Challenges

- Deepfake
 - Misinformation
 - Mass social influence
- Ability to amplify Biases
- Potential for toxicity
- Potential Privacy Breach
- Copyright and IPR breach



Source: Masood, Momina, et al. "Deepfakes generation and detection: State-of-the-art, open challenges, countermeasures, and way forward." Applied intelligence 53.4 (2023): 3974-4026.

Impact on jobs, careers, creativity and education

- Not AI vs. human; augmenting human skills and creativity
- Boosting human productivity and efficiency; raises the bar on what is possible
- Automation for repetitive tasks; enables focus on higher cognitive tasks
- Short term displacements; but creates new career opportunities
- Reducing learning curve and accelerates upskilling
- Personalized learning; improved access to quality education



Yann LeCun

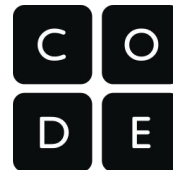
@ylecun

- Engineer: I invented this new thing. I call it a ballpen 🖋️
- TwitterSphere: OMG, people could write horrible things with it, like misinformation, propaganda, hate speech. Ban it now!
- Writing Doomers: imagine if everyone can get a ballpen. This could destroy society. There should be a law against using ballpen to write hate speech. regulate ballpens now!
- Pencil industry mogul: yeah, ballpens are very dangerous. Unlike pencil writing which is erasable, ballpen writing stays forever. Government should require a license for pen manufacturers.

🕒 Last edited 4:28 am · 19 May 2023 · 2.2M Views

Collectively shaping the future of AI

- Making AI benefits accessible to everyone
- Promote AI literacy and education
- Develop and enforce ethical guidelines; Responsible AI
- Reskilling and upskilling initiatives
- Open dialogue and collaboration - scientists, policymakers, industry leaders, and citizens
- Open Source



Hugging Face



Thank You!