Intro to AI

04/05/21

Agenda

- So far :
  - we have learned a bit about Machine Learning
- Today:
  - What is AI?
  - How it is related to machine Learning

What is AI?

Some early definitions:

- " The exciting new effort to make computers **think** … machines with minds, in the full and literal sense." (Haugeland, 1985)
- "The art of creating machines that perform functions that require intelligence when performed by **people**" (Kurzweil, 1990)

What is AI?

Some early definitions:

- "The study of **mental** faculties through the use of computational models" (Charniak and McDermott, 1985)
- Computational Intelligence is the study of the design of **intelligent** agents." (Poole et al. 1992)

What is AI?

A brief definition could be :

Agents that can **think** and **act humanly** and **rationally**.

Turing Test

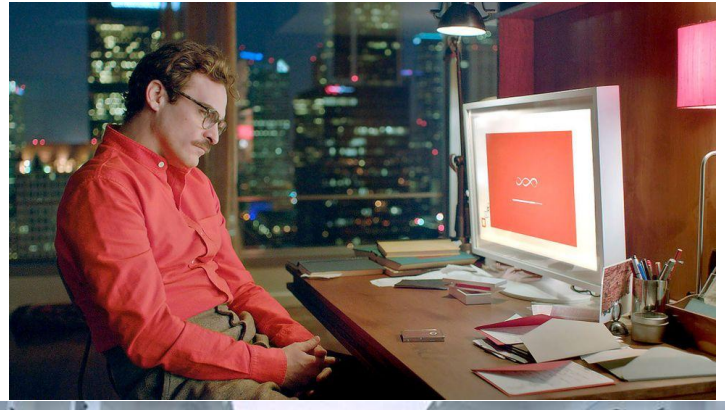
Proposed by Alan Turing (1950), was designed to provide a satisfactory operational definition of AI:

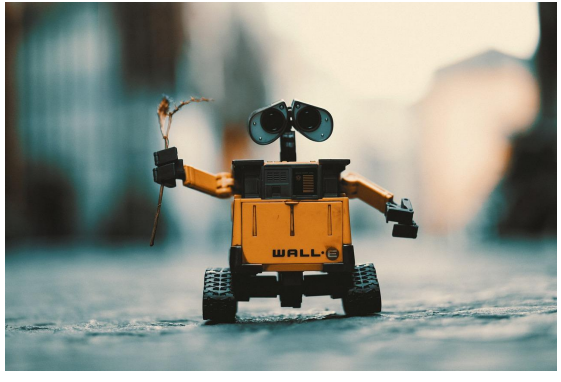
A computer passes the test if a human interrogator, after posing some written questions, can not tell whether the written responses came from a person or from a computer.

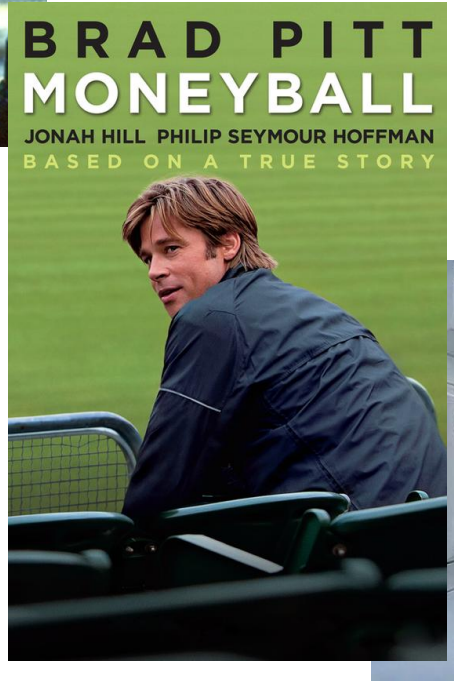
AI Birthday: 1956

"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 if we work on it together for a **summer.**" (John McCarthy and Claude Shannon ,Dartmouth Workshop Proposal)

AI in Movies




WALL·E


BRAD PITT
MONEYBALL
JONAH HILL PHILIP SEYMOUR HOFFMAN
BASED ON A TRUE STORY

AI in the News

Can artificial intelligence predict whether someone will die from COVID?

An artificial intelligence algorithm developed by the University of Copenhagen managed to predict, with up to 90% accuracy whether someone undiagnosed is at risk of dying from COVID-19.

AI in your daily life

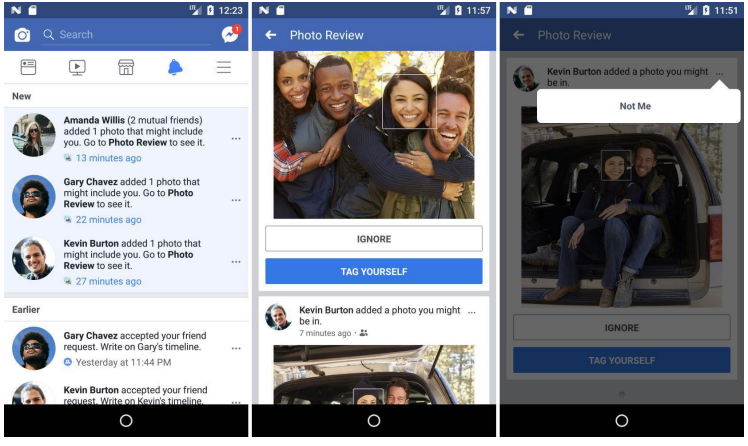
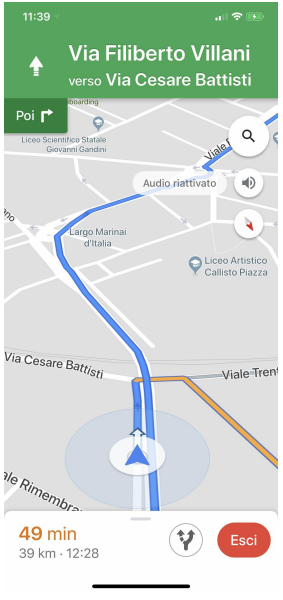


Image recognition

Dialog systems



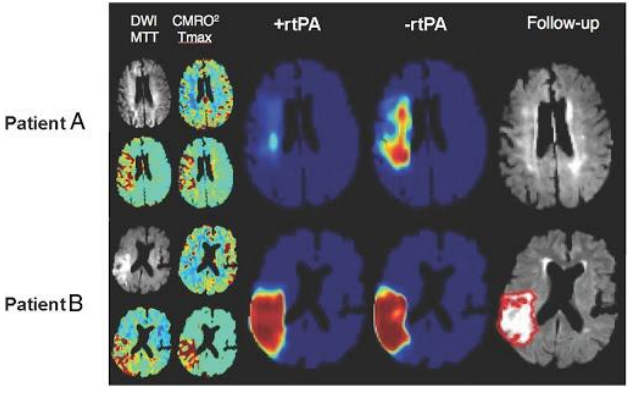
Path planning

What other applications of AI you can think of?

Recommender systems

Translation

Autonomous Driving



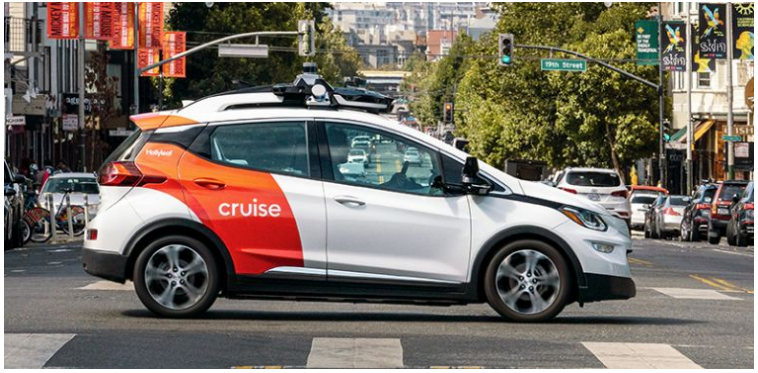
Disease detection

Robotics (delivery)

Stanley (2005)

Tartan (2007)

Nuro



Cruise

Waymo

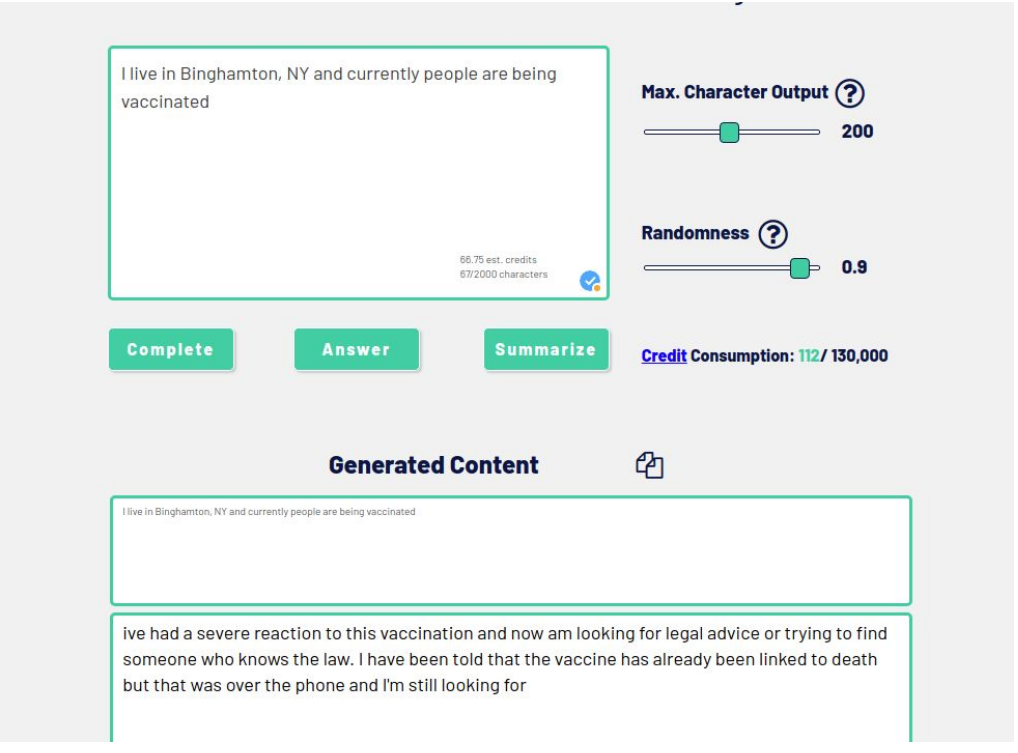
Breakthroughs: DQN (2014)

Breakthroughs: AlphaGo (2016)

Breakthroughs: GPT-3 (2020)

GPT-3 (Link)

Transformer (Link)



Breakthroughs: Dall-E (2020)

https://openai.com/blog/dall-e/

Major Fields of AI

Decide: Planning

Given a task, how an agent can take multiple actions (decisions) to achieve the goal (accomplish the task)

Think: Also known as knowledge representation and reasoning

Learn: Given some data or some interaction with the environment, how can the agent learn a pattern or a behaviour?

Fields of AI: Planning

Classical vs. Probabilistic

- Classical Planning: Generating a sequence of actions prior to executing them, assuming that action execution does not fail
- Probabilistic Planning: Finding a policy, that tells the agent what to do at each state. Assumes uncertainty in action execution.

Fields of AI: KRR

Using logical rules to define the relationship between objects, entity, etc and query conclusions.

Could be probabilistic

Example: Finding the meeting time that works for everyone

Fields of AI: Learning

Learn: Given some data or some interaction with the environment, how can the agent learn a pattern or a behaviour?

Machine learning types

In Machine Learning, we want to find some pattern in data, without explicitly coding those patterns. The goal is make predictions or decisions.

Three main types:

d Learning

Requires: annotated dataset
Used for: Mainly prediction

Unsupervised Learning

Requires: Non-annotated data
Used for: Finding similarity in data. Also, data generation

Reinforcement Learning

Requires: No initial data. Instead, interacting with the environment and learn from feedback
Used for: Decision making

How projects make progress? (Ideally)

1-Literature review

2- Problem statement

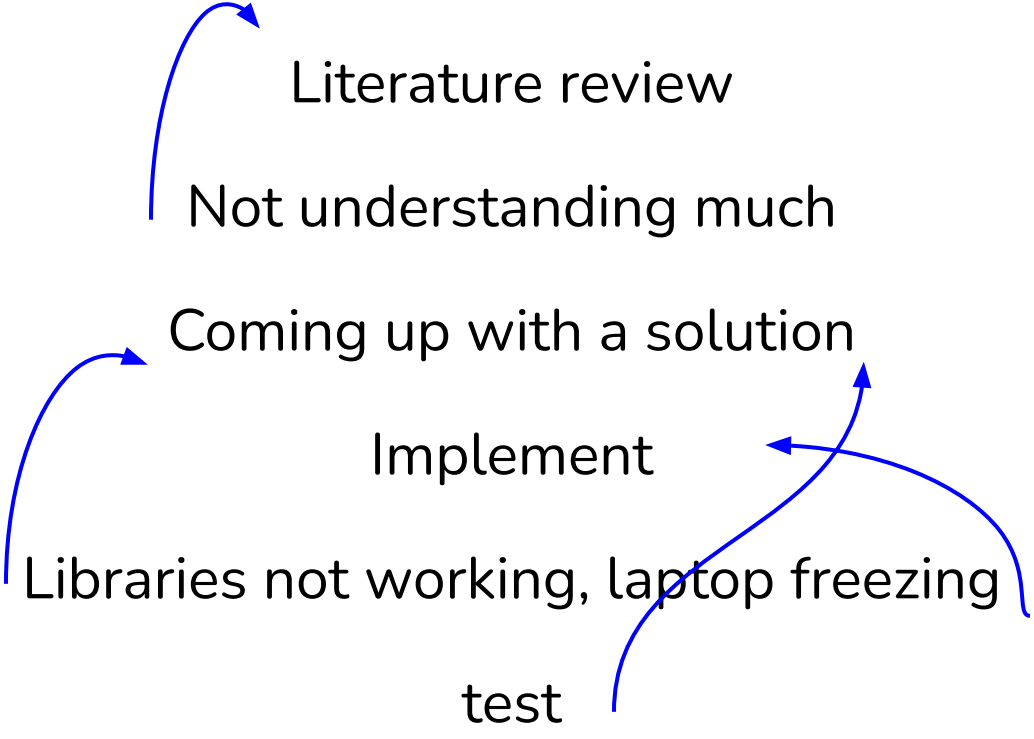
3- Solution

4- Implement

5- Debug

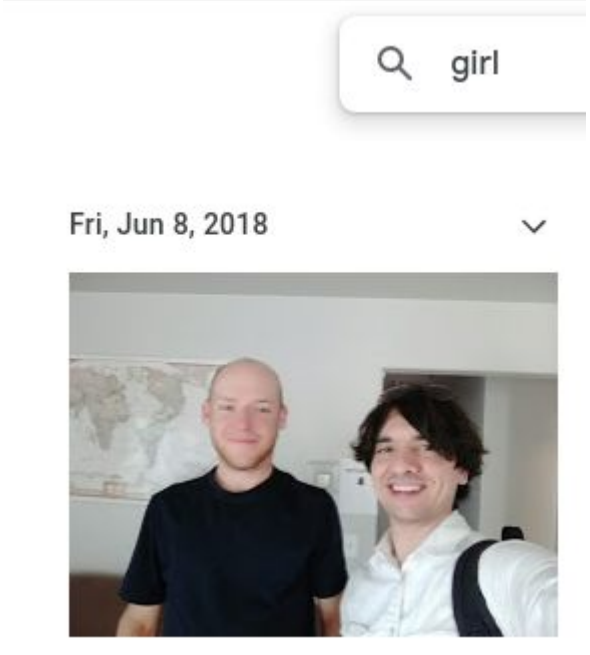
6- Finish

How projects make progress? (Reality)



New emerging Fields of AI

Explainability: If the neural networks makes a wrong prediction/decision, can it be explained?



New emerging Fields of AI

Bias: Is the prediction system biased in their prediction based on gender, race, color?

New emerging Fields of AI

Bias:

How diverse this face generation tool is?

New emerging Fields of AI

Bias: Google Translate not translating pronouns correctly

New emerging Fields of AI

Fairness/Ethics: Is using autonomous weapons ethical?

Steps taken

nature

Explore content · Journal information · Publish with us · Subscribe

nature › news › article

NEWS · 23 DECEMBER 2020 · CORRECTION 23 DECEMBER 2020

Prestigious AI meeting ta
improve ethics of researc

For the first time, the organizers of NeurIPS requ
societal impact of their work.

Davide Castelvecchi

View

Broader Impact

Our work applies differential privacy and generalization bounds to make streaming algorithms robust to adversarial attacks and feedback loops (in which the value reported by the algorithm affects future updates). The idea of using differential privacy as a tool to protect against adversarial attacks on the randomness of the algorithm may be applicable more generally, when a randomized ML model that reports continuously is exposed to a dangerous feedback loop or malicious users. We believe that the connection we establish in this work is only the beginning, and that, following our work, ideas from the literature of differential privacy will continue to find new applications in the field of robust streaming and other related areas.

Acknowledgments and Disclosure of Funding

The authors are grateful to Amos Beimel and Edith Cohen for many helpful discussions.

Haim Kaplan is partially supported by Israel Science Foundation (grant 1595/19), German-Israeli Foundation (grant 1367/2017), and the Blavatnik Family Foundation. Yishay Mansour has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 882396), and by the Israel Science Foundation (grant number 993/17). Uri Stemmer is partially supported by the Israel Science Foundation (grant 1871/19), and by the Cyber Security Research Center at Ben-Gurion University of the Negev.

References

[1] K. J. Ahn, S. Guha, and A. McGregor. Analyzing graph structure via linear measurements. In *SODA 2012*.

Real-world problems are tough

Real-world problems are tough

References

Russell, Stuart, and Peter Norvig. "Artificial intelligence: a modern approach." (2002).