MONASH University

INTRO TO ARTIFICIAL INTELLIGENCE

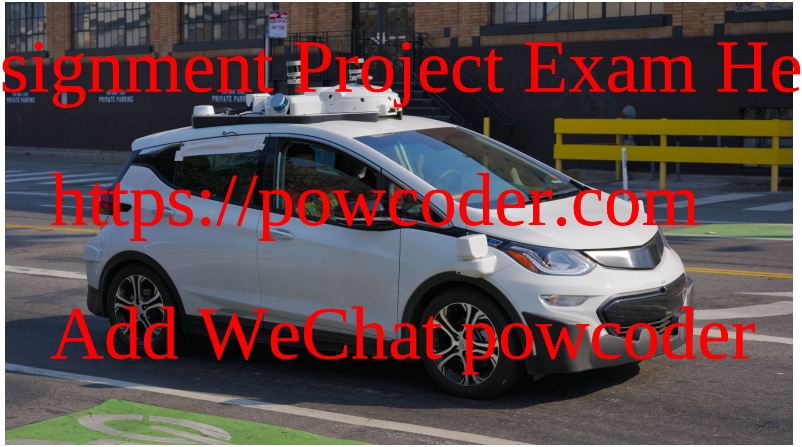
FIT3080

Hamid Rezatofighi

AlphaGo vs. Lee Sedol (2016)

Scene from AlphaGo documentary,
https://www.youtube.com/watch?v=WXuK6gekU1Y

Progress in self-driving cars



Testing of Cruise autonomous vehicle, by Dllu

https://en.wikipedia.org/wiki/Cruise_(autonomous_vehicle)#/media/File:Cruise_Automation_Bolt_EV_third_generation_in_San_Francisco.jpg

Amazon Go: A “just walk out” store

https://www.youtube.com/watch?v=zdbumR6Bhd8

Artificial Intelligence becoming mainstream

- Industrial Revolution → major increase in manufacturing productivity
- AI Revolution → promises major increase in knowledge work productivity

Currently AI research is receiving significant attention and investment,

Estimated 133 million new jobs globally by 2022 (Dwivedi et al., 2019)

Artificial Intelligence becoming mainstream

- Industrial Revolution → major increase in manufacturing productivity
- AI Revolution → promises major increase in knowledge work productivity

Currently AI research is receiving significant attention and investment,

Estimated 133 million new jobs globally by 2022 (Dwivedi et al., 2019)

However, we have seen 'AI winters' before...

Goals of the field of AI research

- Understand what intelligence is, and
- Build intelligent systems

What is intelligence?

Ingredients of an intelligent entity:

- Communication
- Internal knowledge, memory
- World knowledge, model of the world
- Intentions, and plans to achieve them
- Creativity

Definitions of AI

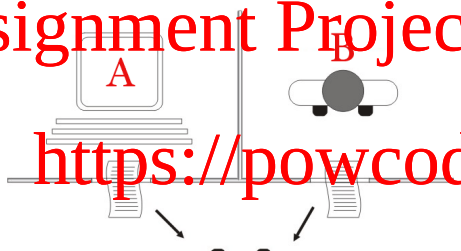
- “[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning ..." (Bellman, 1978)
- “The study of computations that make it possible to perceive, reason, and act." (Winston, 1992)
- “The study of how to make computers do things at which, at the moment, people are better." (Rich and Knight, 1991)

Nuances in definition of AI

- Human ideal:
  - Thinking humanly
  - Acting humanly
- Rational ideal:
  - Thinking rationally
  - Acting rationally

Acting humanly: Turing test (1950)

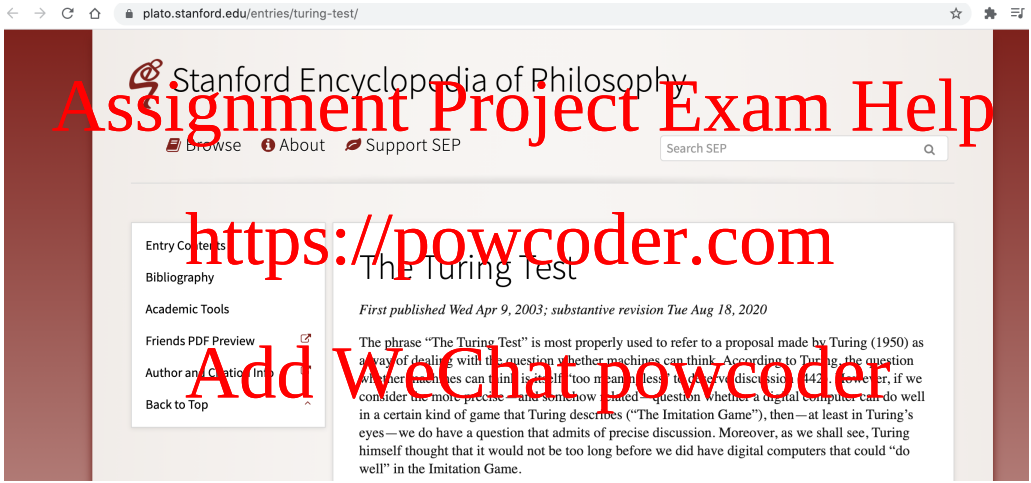
Judge C uses written responses to questions to decide whether A or B is the computer



Requires knowledge, reasoning, language, learning

By definition an empirical test to judge success

Turing test, A read



https://plato.stanford.edu/entries/turing-test/

Acting rationally

Rational behaviour: always doing the *best* thing

Acting rationally

Rational behaviour: always doing the *best* thing

Where *best* means: that which maximizes the expected performance on the goal, given all available information.

Acting rationally

Rational behaviour: always doing the *best* thing

Where *best* means: that which maximizes the expected performance on the goal, given all available information.

Limits on rational behavior: CPU time, memory capacity

Bounded rationality: do the best possible under given resources

Acting rationally

Rational behaviour: always doing the *best* thing

Where *best* means: that which maximizes the expected performance on the goal, given all available information.

Limits on rational behavior: CPU time, memory capacity

Bounded rationality: do the best possible under given resources

This unit covers algorithms that target rational behaviour

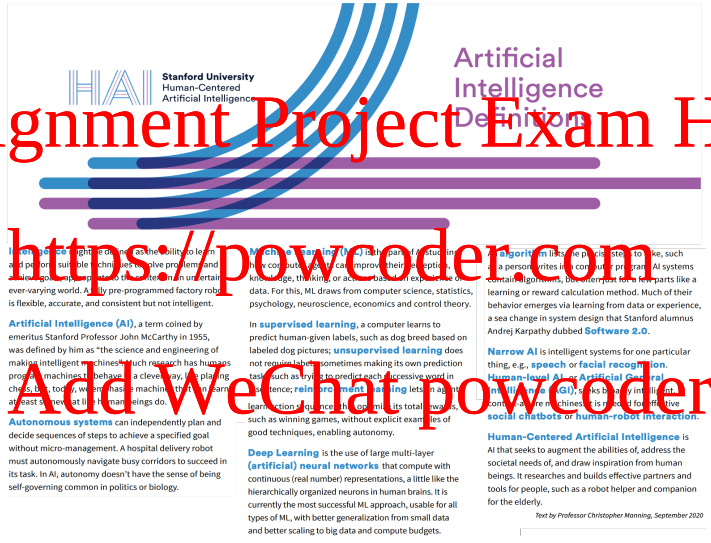
Nuances in the path to AI

- '*Strong*' AI
  - Build a machine that is actually thinking
  - Top-down philosophical perspective: does my program think?
  - Create general strategy to intelligence
    - One program that can do any task.
- '*Weak*' AI
  - Build a machine that acts *as if* it were intelligent
  - Like a submarine acts *as if* it can swim
  - Bottom-up, goal-driven perspective: does my program work?
  - Create domain-specific strategies:
    - a chess program,
    - a train scheduling program,
    - a language translation program,
    - an image classification program

Nuances in the path to AI

- '*Strong*' AI
  - Build a machine that is actually thinking
  - Top-down, philosophical perspective: does my program think?
  - Create general strategy to intelligence
    - One program that can do any task.
- '*Weak*' AI
  - Build a machine that acts *as if* it were intelligent
  - Like a submarine acts *as if* it can swim
  - Bottom-up, goal-driven perspective: does my program work?
  - Create domain-specific strategies:
    - a chess program,
    - a train scheduling program,
    - a language translation program,
    - an image classification program

More AI Definitions



Stanford University
Human-Centered
Artificial Intelligence

Artificial
Intelligence
Definitions

Intelligence highlights the general ability to learn and perform suitable techniques to solve problems and create models that can deal with new situations in an ever-varying world. A fully pre-programmed factory robot is flexible, accurate, and consistent but not intelligent.

Artificial Intelligence (AI), a term coined by emeritus Stanford Professor John McCarthy in 1955, was defined by him as "the science and engineering of making intelligent machines." Much research has humans program machines to behave in a clever way, like playing chess, but, today, we emphasize machines that can learn, at least as well as humans and animals do.

Autonomous systems can independently plan and decide sequences of steps to achieve a specified goal without micro-management. A hospital delivery robot must autonomously navigate busy corridors to succeed in its task. In AI, autonomy doesn't have the sense of being self-governing common in politics or biology.

Machine learning (ML) is a branch of statistics and computer science that can improve their performance, knowledge, thinking, or actions based on experience or data. For this, ML draws from computer science, statistics, psychology, neuroscience, economics and control theory.

In **supervised learning**, a computer learns to predict human-given labels, such as dog breed based on labeled dog pictures; **unsupervised learning** does not require labels, sometimes making its own prediction task such as trying to predict each successive word in a sentence; **reinforcement learning** lets an agent learn action sequences through trial and error to maximize total rewards, such as winning games, without explicit examples of good techniques, enabling autonomy.

Deep Learning is the use of large multi-layer **(artificial) neural networks** that compute with continuous (real number) representations, a little like the hierarchically organized neurons in human brains. It is currently the most successful ML approach, usable for all types of ML, with better generalization from small data and better scaling to big data and compute budgets.

Algorithms like some problem solving techniques, such as a person writes in a computer program. AI systems contain algorithms, but interrupt for a few parts like a learning or reward calculation method. Much of their behavior emerges via learning from data or experience, a sea change in system design that Stanford alumnus Andrej Karpathy dubbed **Software 2.0**.

Narrow AI is intelligent systems for one particular thing, e.g., **speech** or **facial recognition**.

Human-level AI, or **Artificial General Intelligence (AGI)**, seeks broadly intelligent computers. Here machines' AI is rated for effective **social chatbots** or **human-robot interaction**.

Human-Centered Artificial Intelligence is AI that seeks to augment the abilities of, address the societal needs of, and draw inspiration from human beings. It researches and builds effective partners and tools for people, such as a robot helper and companion for the elderly.

Text by Professor Christopher Manning, September 2020

AI Definitions by Prof. Christopher Manning from Stanford University Human-Centered AI. `shorturl.at/cqrJZ`

Scope of FIT3080

Problems and approaches in **designing intelligent software**

Examples:

- Abstracting problems to allow solving them
  - Representing states of the world, properties of environments (PEAS)

Scope of FIT3080

Problems and approaches in **designing intelligent software**

Examples:

- Abstracting problems to allow solving them
- Finding the best sequence of decisions to take
  - Search algorithms (A*), planning algorithms (value iteration)

Scope of FIT3080

Problems and approaches in **designing intelligent software**

Examples:

- Abstracting problems to allow solving them
- Finding the best sequence of decisions to take
- Learning how to perform well in an unknown environment
  - Reinforcement learning algorithms (Q-learning, Deep Q-Networks)

Scope of FIT3080

Problems and approaches in **designing intelligent software**

Examples:

- Abstracting problems to allow solving them
- Finding the best sequence of decisions to take
- Learning how to perform well in an unknown environment
- Reasoning about information to derive new knowledge
  - Logic and probabilistic (Bayesian) reasoning

Scope of FIT3080

Problems and approaches in **designing intelligent software**

Examples:

- Abstracting problems to allow solving them
- Finding the best sequence of decisions to take
- Learning how to perform well in an unknown environment
- Reasoning about information to derive new knowledge
- Generalizing to unseen data from historic examples
  - Principles of machine learning