Most materials in this course are based on materials from http://aima.cs.berkeley.edu/ and http://ai.berkeley.edu

0. Introduction

Dirk Schnieders

File generated on 19 January 2024 at 1:09 PM

Intelligence

- We call ourselves Homo Sapiens
  - Latin: Man the wise
  - Intelligence is important to us
- For thousands of years, we have tried to understand how we humans think
  - How can our brain perceive, understand, predict, and manipulate a world far larger than itself ?
- Intelligence is most widely studied in humans, but has also been observed in animals and in plants

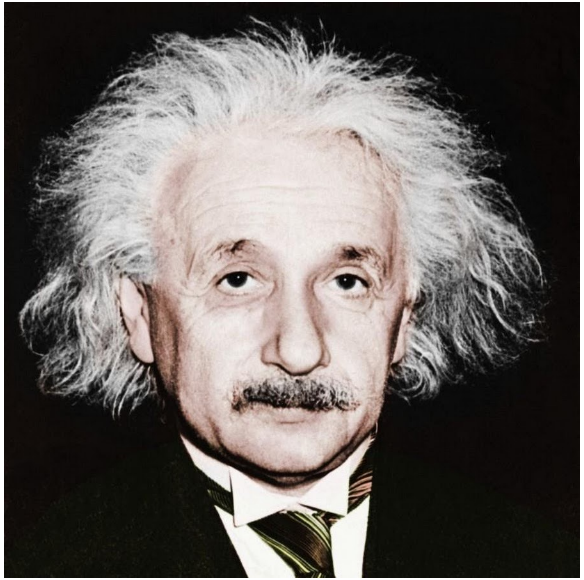
Artificial Intelligence

- AI goes further than just understanding intelligence
    - It attempts to build intelligent entities
    - Computing to act effectively and safely in a wide variety of novel situations
- AI: science, or engineering?

Why study AI?

AI’s impact will be “more than anything in the history of mankind.”

Why study AI?



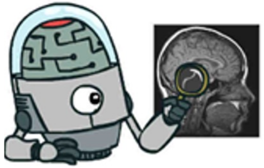
Physics

AI

Historically,
researchers have pursued several different versions of AI

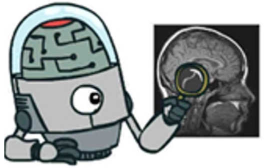
- Let’s build machines that …
  - Think Humanly
  - Act Humanly
  - Think Rationally
  - Act Rationally

A. Think Humanly

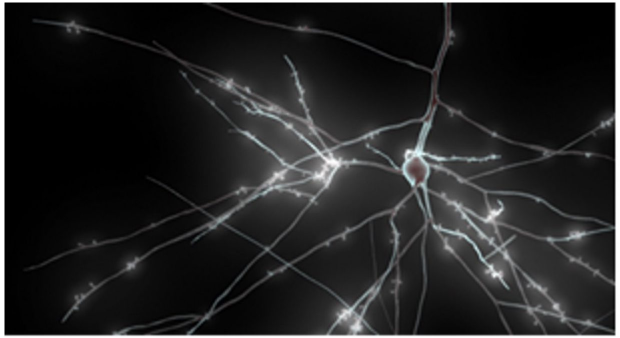


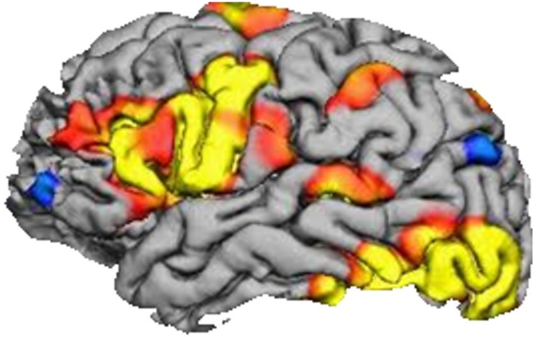
- Cognitive Modelling Approach: How do we think?
  - Introspection
  - Psychological experiments
  - Brain imaging
- Cognitive science constructs precise and testable theories of the human mind
  - E.g., express a theory as a computer program and compare input-output behaviors to a human
  - If there is a match, some of the programs mechanism could also be operating in humans

A. Think Humanly



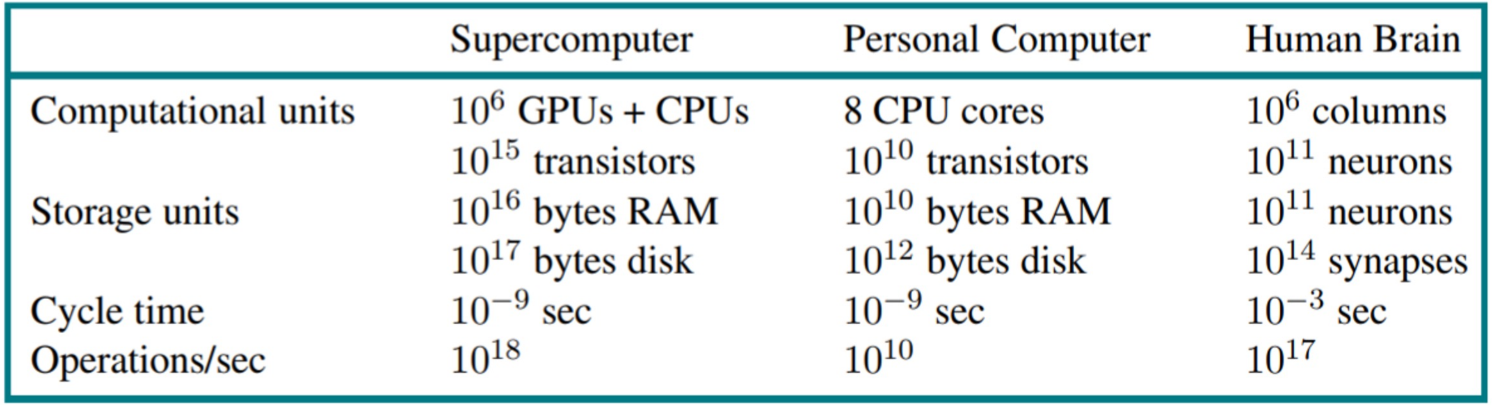
- The human brain is one of the great mysteries of science
  - How does our brain process information?
- The brain consists of nerve cells (aka neurons) and the collection of these simple cells leads to thought, action and consciousness
- The recent development of functional magnetic resonance imaging (fMRI) provides neuroscientists with details of brain activities





A. Think Humanly

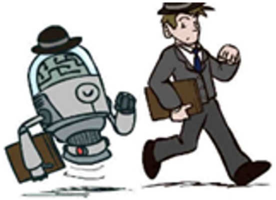
- ➢ Brains and digital computers have somewhat different properties
- ➢ A crude comparison of the raw computational resources



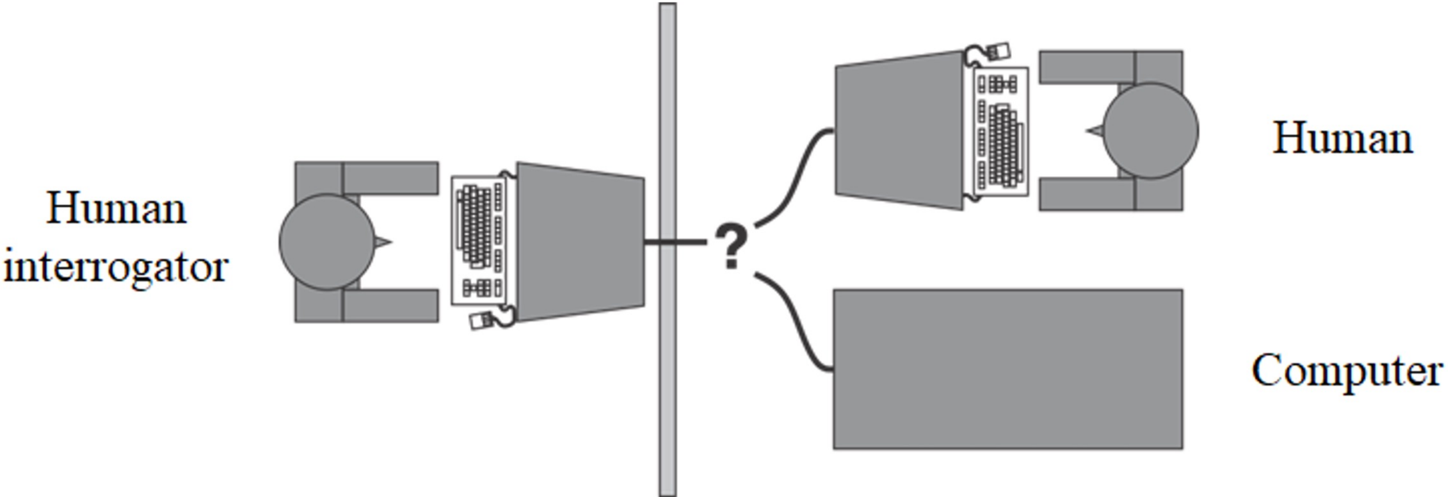
	Supercomputer	Personal Computer	Human Brain
Computational units	10^6 GPUs + CPUs	8 CPU cores	10^6 columns
	10^{15} transistors	10^{10} transistors	10^{11} neurons
Storage units	10^{16} bytes RAM	10^{10} bytes RAM	10^{11} neurons
	10^{17} bytes disk	10^{12} bytes disk	10^{14} synapses
Cycle time	10^{-9} sec	10^{-9} sec	10^{-3} sec
Operations/sec	10^{18}	10^{10}	10^{17}

- ➢ Would we be able to achieve the brain’s level of intelligence with a computer of unlimited capacity?

B. Act Humanly



- ➢ Turing Test Approach: Imitation Game was designed to provide a definition of intelligence
- ➢ A computer passes the test if a human interrogator, after posing some questions, cannot tell whether the response come from a human or a computer

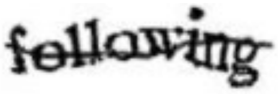


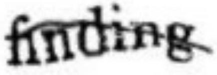
B. Act Humanly

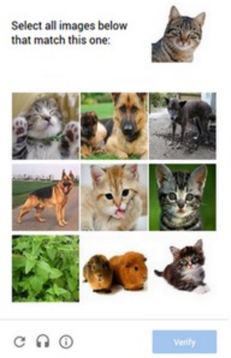
- The underlying principles of intelligence are more important than to duplicate an exemplar
- Consider another field: Artificial Flight
    - The Wright brothers succeeded because they stopped imitating birds and started using wind tunnels and learn about aerodynamics
    - It was not their goal to make “machines that fly so exactly like pigeons that they can fool even other pigeons”

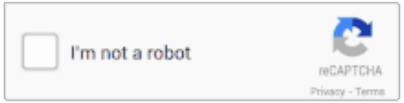
B. Act Humanly

- Reverse Turing Test
    - Turing test in which the objective / roles have been reversed
    - Interrogator is a computer. Interrogatee is human or computer.
- CAPTCHA
    - Completely Automated Public Turing test to tell Computers and Humans Apart






Select all images below that match this one:


□ I'm not a robot

C. Think Rationally

- The “laws of thought” approach
- What are the laws that guide and underlie our thinking?
- Greek schools developed various forms of logic
  - Notation and rules of derivation for thoughts
  - Example: Socrates is a man; all men are mortal; therefore, Socrates is mortal
- By 1965, programs existed that could (in principle) solve any solvable problem described in logic notation
- Problems with this approach
  - How to take informal knowledge and state it in formal terms? How about uncertainty?

D. Act Rationally

- The rational agent approach
    - Act to achieve the best outcome
        - With uncertainty: the best expected outcome
- Advantages over the other approaches
    - More general
        - Correct inference is just one of several possible mechanisms for achieving rationality
    - Rationality is well defined
        - Human behaviour is well adapted only for one specific environment
- Most researchers in AI focus on the general principles of rational agents and how to build them

Rational Agent

- Agent
  - agent comes from the Latin *agere*, to do
  - Something that perceives and acts
  - E.g., robot or softbot
- Rational Agent
  - Acts to achieve the best outcome or, when there is uncertainty, the best expected outcome
- For any given class of environments and tasks, we seek the agent (or class of agents) with the best performance

Rational Agent

- AI focuses on the study and construction of rational agents
  - Agents that do the right thing
  - What counts as the right thing is defined by the objective that we provide to the agent
- Like other areas of research
  - Control Theory
  - Operations Research
  - Statistics
  - Economics
- This is called the standard model

Perfect Rationality

- Always taking the exactly optimal action is not feasible in complex environments

Issues with the Standard Model

- The standard model assumes that we will supply a fully specified objective function
  - Difficult in practice
- Value alignment problem
  - Example
    - Domain: Self-driving car
    - Objective: Reach destination safely
    - Problem
      - Strict goal of safety requires staying in the garage
      - There is a tradeoff between making progress towards the destination and incurring a risk of injury
      - How should this tradeoff be made?

Value Alignment Problem

- Example: Chess

Machine Learning

- An agent is learning if it improves its performance on future tasks
- Why would we want an agent to learn?
    - If the design of an agent can be improved, why not design the agent with that improvement to begin with?

The Thinking Machine - MIT 1961

http://www.youtube.com/watch?v=aygSMgK3BEM

History of AI - Turing Award Winners

- Marvin Minsky (1969)
- John McCarthy (1971)
- Edward Feigenbaum and Raj Reddy (1994)
- Judea Pearl (2011)
- Yoshua Bengio, Geoffrey Hinton, and Yann LeCun (2018)

History of AI - Milestones

- Inception (1943 - 1956)
- Early Enthusiasm (1952 - 1969)
- A dose of reality (1966 - 1973)
- Expert systems (1969 - 1986)
- Return of NN (1986 - present)
- Probabilistic reasoning (1987 - present)
- Big data (2001 - present)
- Deep Learning (2011 - present)

The State of the Art

- Publications
  - AI papers increased 20 fold between 2010 to 2019 to 20,000 a year
- Conferences
  - Attendance of NeurIPS increased 800% since 2012 to 13,500
- Industry
  - AI start-ups in the US increased 20 fold from 2010 to 2019
- Internationalization (in 2019)
  - China publishes more AI papers per year then US and about as many as Europe
  - In citation weighted impact, US is ahead by 50% vs. China

The State of the Art

- Vision
  - Error rates for object detection improved from 28% to less than 2%
- Speed
  - Training time for image recognition dropped by a factor of 100 in last 2 years
  - Amount of computing power used in top AI applications is doubling every few month
- Humans vs. AI (in 2019)
  - AI is better in chess, go, poker, pac-man, jeopardy!, object detection, speech recognition in limited domain, chinese-to-english in restricted domain, Quake III, Dota 2, StarCraft II, many Atari games, Skin cancer detection, prostate cancer detection, protein folding, ...

Benefits of AI

- First solve AI, then use AI to solve everything else.

Demis Hassabis, Google DeepMind

Risks of AI

- Lethal autonomous weapons
- Surveillance
- Biased decision making
- Impact on employment
- Safety-critical applications
- Cybersecurity

Risks of AI - Superhuman AI

- Most experts agree that we will eventually be able to create a superhuman AI
  - An intelligence that far surpasses human ability

Risks of AI - The Gorilla Problem

- About seven million years ago, a now-extinct primate evolved
  - one branch led to gorillas
  - another to humans
- Today the gorillas are probably not too happy about the human branch
  - They have no control over their future

Risks of AI - The Gorilla Problem

- If the gorilla problem is the result of developing AI then we should stop working on it
- If superhuman AI (aka AGI) were a black box from outer space, we should be careful in opening the box
    - But it is not, we design the AI systems
    - If AI does end up taking control, it would be a design failure
- We need to understand the source of potential failure
    - Philosophical foundations of AI
    - Maybe the most important area of AI research

AI Experts on the AI Apocalypse

- Worried AI experts signed on open letter in March 2023 asking all AI labs to immediately pause “giant AI experiments”
- Where do AI experts stand regarding the probability of AGI and the probability of disaster by AGI?
  - Find out here
    - https://dirk.hk/ai/IEEESpectrumAug2023_1.jpg
    - https://dirk.hk/ai/IEEESpectrumAug2023_2.jpg

Reading

- AIMA: Chapter 1
  - 66 pages available here: https://dirk.hk/ai/AIMA_Chapter_1.pdf
    - Password: Schnieders
- AIMA: Chapter 2
  - 49 pages available here: https://dirk.hk/ai/AIMA_Chapter_2.pdf
    - Password: Dirk
- The Thinking Machine – MIT 1961
  - 53 min video available here: https://dirk.hk/ai/TheThinkingMachine.html