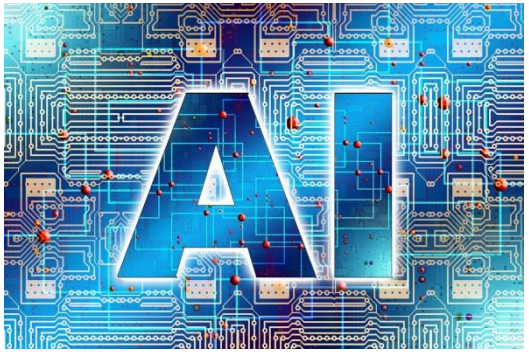
Australian National University

COMP3620/6320



Course Organisation and Introduction

Felipe Trevizan (Convenor)

https://cs.anu.edu.au/courses/comp3620/

Course Organization

- All the information about the course, assignments, lab, tutorials, policies are in the website:

https://cs.anu.edu.au/courses/comp3620/

- You must the **policies** and **outline** sections at least once
  - Not knowing the policies in place is not a valid excuse

5 Minutes Summary of the Course

- **3 Topics** (6 lectures each):
  - Search
  - Knowledge Representation and Reasoning (KRR)
  - Planning
- **4 Assignments**
  - 100% penalty if late
  - Plagiarism detection software, some automated testing and manually checked
- **6 Tutorials**
  - 2 per topic
  - 1 quiz per tutorial (0 marks if you miss the tutorial)
- **7 Labs**
- **1 Final exam**
- **1 Hurdle:** grade on the final exam >= 40 (out of 100)

Tutorials, Labs, Assignments and Quizzes

- **Tutorials**:
  - Goal is to help understand the material and prepare exam
  - Will discuss a list of questions, try answering them before the tutorial
- **Quizzes**:
  - Goal is to provide a reality check
  - Are keeping up with the content or should you spend more time studying?
- **Assignments**:
  - Goal is to put the course into practice by building AI programs
  - Essential to build a deep understanding of the course
- **Labs**:
  - Goal is to get help from the tutors with the assignments
  - **Unstructured and self-guided** (that is, you need to bring questions)
  - Get started well in advance to make the most of the opportunity

Contact & Information

- The course page is the main source of information
  - **It has priority in case of conflicting information**
- Use **Piazza** for all communications
  - See communication policy https://cs.anu.edu.au/courses/comp3620/policies/#communication for more details of whom to contact in different situations

Course Representatives

Please nominate yourself via the **CECS Course Representative EOI form** **by midday 1st March 2021**. *You are free to nominate yourself whether you are currently on-campus or overseas.*

You will be contacted by CECS Student Services, Employability and Experience by 5th March with the outcome of your self-nomination.

Meetings will be held via Zoom and there will be **three meetings this semester**, meeting details will be provided to course representatives shortly.

More information about roles and responsibilities can contact:

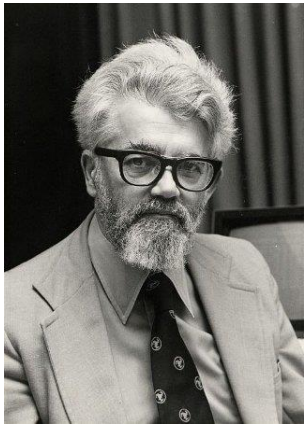
- ANUSA CECS representatives: Sandy Ma and Swatantra Roy: sa.cecs@anu.edu.au
- ANUSA President: Madhumitha Janagaraja: sa.president@anu.edu.au

Introduction

Topics

- What is AI?
- Foundational and Current Disciplines
- Brief History
- Ethics

Artificial Intelligence



John McCarthy
1927-2011

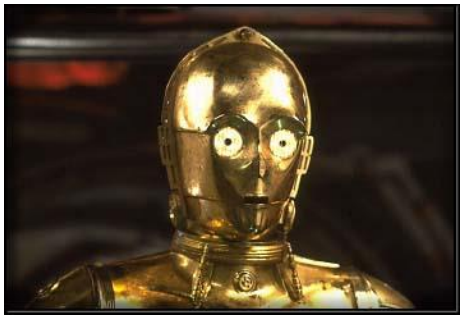
"The science and engineering of making intelligent machines"

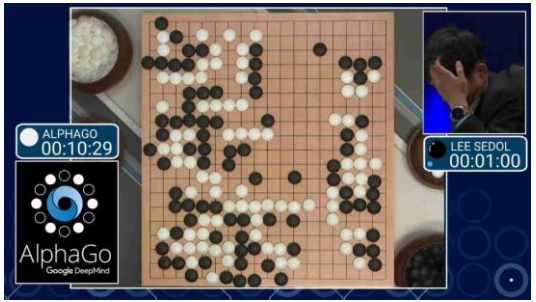
- **Official birth:** Dartmouth College Meeting, 1956

- **Ambitious goals:**
  1. Understand "intelligence"
  2. Build "intelligent" machines

But … what counts as an intelligent machine?

What is an Intelligent Machine?





What is an Intelligent Machine?

- 4 different perspectives corresponding to two dimensions
  1. **thinking vs acting** (thought/reasoning vs actions/behavior)
  2. **human vs rational** (close to a human vs close to optimal)

Systems that think like humans	Systems that think rationally
Systems that act like humans	Systems that act rationally

- **Rationale:**
  1. The ability to think is what distinguishes humans from animals. But a machine could behave intelligently without thinking.
  2. Human exhibit intelligence. But why not aim at super-human, optimal behavior, especially for specific tasks?

What is an Intelligent Machine?

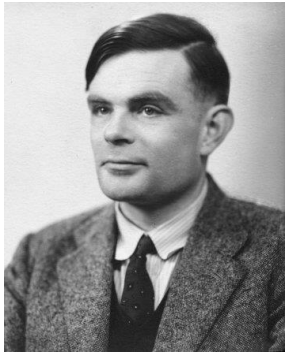
Systems that think like humans	Systems that think rationally
Systems that act like humans	Systems that act rationally

- Making machines with minds in the full and literal sense [Haugeland 85]
- Making computers do things at which, for now, people are better [Kurzweil 90]
- Studying the computations that enable perception, reasoning, and actions [Winston 92]
- Concerned with the automation of intelligent behavior [Luger & Subblefield 93]

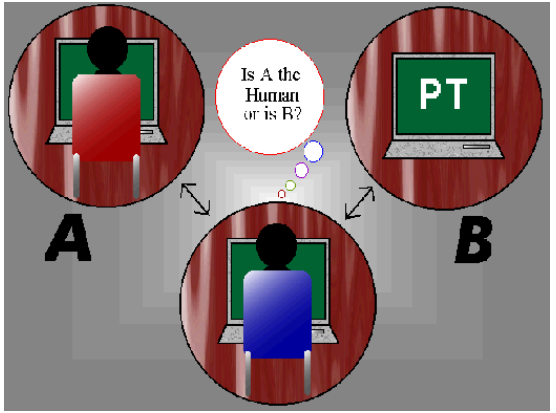
Whether machines can think is as relevant as whether submarines can swim [Dijkstra 84]

Acting Humanly: The Turing Test

- "Computing Machinery and Intelligence" [Turing 50]
- Operational test for intelligent behavior
- Imitation game: computer passes the test if interrogator cannot tell which is the machine (no physical presence)
- Focused on major AI components:
  - Natural language processing
  - Knowledge representation & reasoning
  - Learning
- Prediction: by 2000, 30% chance to fool a human in a 5-minutes conversation



Alan Turing
1912-1954



Acting Humanly: The Turing Test

- Eliza (Weizenbaum 66) rule-based program replicating a psychotherapist.
  Primitive natural language processing.
- Loebner Prize ($100,000) since 1991
  - A.L.I.C.E 3 bronze medals
  - Mitsuku 3 bronze medal in 2013, 2016-2019 www.mitsuku.com

Patient: I feel sad.
Eliza: Do you often feel sad?
Patient: Not very often.
Eliza: Please, go on.

Human: Does god exist?
Mitsuku: The bible is the most significant piece of evidence that God exists, but it is not a scientific proof.
Human: Okay, but do you believe in god?
Mitsuku: Yeah, okay, I believe there is probably something out there but I doubt it is a man with a long beard sat on a throne.

Thinking Humanly: Modelling Cognition

- Objective: develop scientific theories of the activities of the brain
- Two approaches:

1. **Cognitive Science** (top down):
   uses computer models and experimental psychology techniques to predict and test behavior of human subjects
2. **Cognitive Neuroscience** (bottom up):
   uses computer imaging & other neurological data to observe the brain in action
   - Project to simulate the brain www.humanbrainproject.eu
   - Related to the AI field of neural networks (see deep learning)

- These days, both disciplines are distinct from AI

Thinking Rationally: Laws of Thought

- Objective: formalise and mechanise valid reasoning
- Direct line through maths and philosophy to modern AI

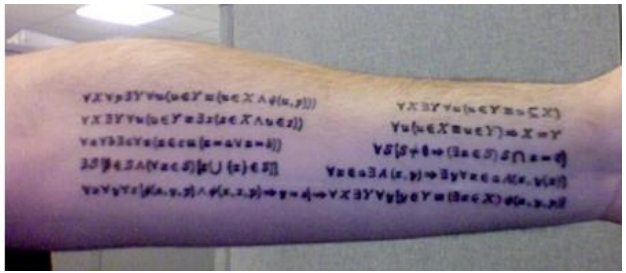
- **Logic**: notation and rules to derive valid conclusions
  - Aristotle's syllogism
  - Mathematical development of classical logic
    - Propositional & first-order logic (Boole, Frege, 1850s)
    - Most of mathematics can be derived from axioms of set theory
  - Non-classical logic to formalise common-sense reasoning
    - Default logic (by default, birds fly)

Tweety is a bird
Birds fly

Tweety flies

P(a)
$\forall x\ P(x) \rightarrow Q(x)$

Q(a)



Thinking Rationally: Laws of Thought

- **Limit 1: Undecidability**
  - Goedl's Theorem: every axiomatisable consistent theory extending arithmetic has formulas that are true but not provable within the theory.
- **Limit 2: Complexity**
  - Non-trivial to formalise a real-world problem in logic
  - Most problems are NP-complete or harder
- **Limit 3: Scope**
  - Not all intelligent behavior requires reasoning (much doesn't)
- **Limit 4: Purpose**
  - Reasoning to prove what? Notion of "goal" is missing

Acting Rationally: Rational Agents

- An **agent** is an entity that perceives and acts in its environment (driverless car, electronic trading system, energy management system)

- **Rationality** is about doing the right thing:
- Decision which achieves the best (expected) outcome, given the information available and time available (limited rationality)

- **This course (and much of today's AI) is about designing rational agents**: for any given class of environment and task, we seek the agent with the best performance.

Artificial Intelligence

"The science and engineering of making intelligent machines"

- **Ambitious goals:**

1. Understand "intelligence"
   - ➢ Accurate models of cognition are now the focus of cognitive science, neuroscience and psychology

2. Build "intelligent" machines
   - ➢ Focus on developing methods that match or exceed human performance in certain domains, possibly by different means.

Australian National University

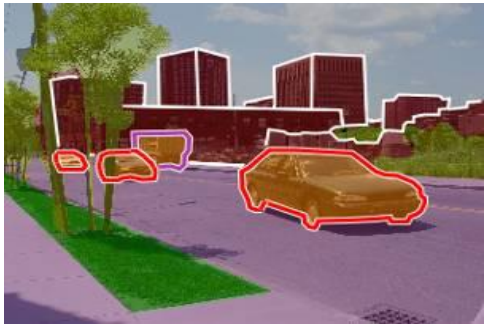
Foundational and Current Disciplines

Foundational Disciplines

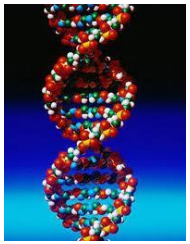
- **Philosophy:** logic, reasoning methods, foundations of learning, language, and rationality
- **Mathematics:** proofs, decidability, complexity, probability
- **Economics:** theory of rational decisions, game theory
- **Computer Sc. & Engineering:** algorithms, efficient computer design, control theory concepts (e.g. stability)
- **Cognitive Science:** behaviorism, adaptation, perception, experimental methods
- **Neuroscience:** information processing by the brain
- **Linguistics:** language representation, language & thought

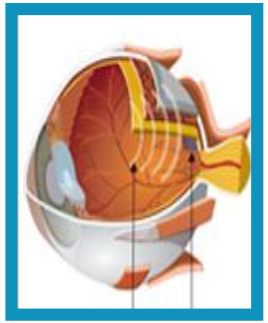
Current Disciplines

- **Knowledge representation & reasoning:** how to efficiently represent information and use this to answer questions and draw conclusions
- **Problem solving, planning, and search:** how to constructively solve problems and make decisions.
- **Machine learning:** inference from data to extrapolate patterns and adapt to new situations.
- **Natural language processing:** verbal communication with humans.
- **Computer vision:** processing and making sense of visual information about the environment.



Applications





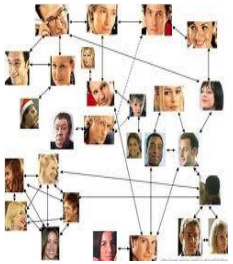
Health

Financial markets

Space

Energy

Defence



Web mining and applications

Transport

Australian National University

Brief History

Brief History

1950: Turing test
1950s: Early programs including checkers, theorist, neural nets
1956: Dartmouth meeting, “Artificial Intelligence” adopted
1965: Robison’s complete algorithm for logical reasoning
1966-74: AI discovers complexity, neural nets research disappears
1969-79: Early knowledge-based systems
1980-88: Expert systems industry booms
1988-93: Expert systems industry “busts”, AI Winter
1988-00: Greater technical depth, resurgence of probabilities
1985-95: Neural nets return, lead to, and replaced by modern SML
2003-: Human-level AI back on the agenda
2010-: Deep learning: neural nets research is favour again
2013-: Ethical issues make the headlines

NNs

Birth

Optimism

Realism

Expert Systems

Winter

Foundations

NN returns

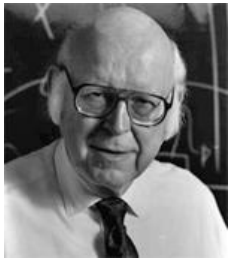
Data, multicore

NN again

Maturity?

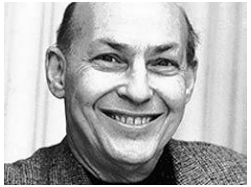
AI Achievements – Predictions

- **1958**: “within ten years a digital computer will be the world's chess champion” [Allen Newell]
- **1965:** “machines will be capable, within twenty years, of doing any work a man can do.” [Herb Simon]
- **1970:** “In from three to eight years we will have a machine with the general intelligence of an average human being.” [Marvin Minsky]



Allen Newell
1927-1992

Herb Simon
1916-2001



Marvin Minsky
1927-2016

AI Achievements – The Reality

- **1991:** Proverb solves crosswords better than human
- **1991:** AI solves Gulf-war logistics planning problems
- **1997**: IBM Deep Blue beats chess champion Kasparov
- **1999:** AI agent controls NASA deep space 1 probe
- **2001:** autonomous military drones unveiled
- **2005:** Driverless vehicles complete the 212km DARPA Grand Challenge through the Mojave desert
- **2007:** Checkers game completely solved
- **2009:** Google autonomous car drives in traffic
- **2011:** IBM Watson wins Jeopardy!
- **2016:** Google alphago beats go champion Lee Sedol
- **Today:** AI is everywhere, injects billions into economy

AI Achievements – New Predictions

- **2030**: “an AI system with an ongoing existence at the level of a mouse" [Rodney Brooks]
- **Not in his lifetime:** “a robot that has any real idea about his own existence, or the existence of humans in a way a 6 years old child would” [Rodney Brooks]
- **2050:** “Germany will loose to a robot soccer team.” [Toby Walsh]

Rod Brooks
1954-

Toby Walsh
1964-

Humanoid Robot Soccer - 1998

Humanoid Robot Soccer - 2018

Robocup 2018 SPL Finals - Nao-Team HTWK vs. B-Human

AI Ethics

AI Ethics and Risks

- People might lose their jobs
+ AI creates wealth and does dangerous and boring jobs for us

- Accountability loss: who is responsible, AI, owner, creator?
+ Similar issues elsewhere (medicine, software, plane crash)

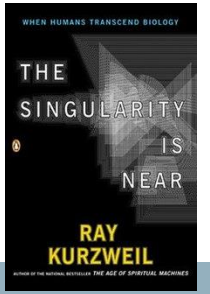
- AI reproducing our negative biases and attitudes (e.g., racism)
+ AI should share our *positive* values

- Use of AI as weapon (e.g., drones)
+ Can also save lives? Every beneficial invention can be misused

AI Ethics and Risks

- AI Success might end of the human era
  - Kurtzweil, Musk, Hawking!
  - Once machine surpasses human intelligence it can design smarter machines.
  - Intelligence explosion and **singularity** at which human era ends

- Many counter arguments
  - limits to intelligence
  - nothing special about human intelligence
  - computational complexity
  - “intelligence to do a task” ≠ “ability to improve intelligence to do a task”



Robotics Laws

The Three Laws of Robotics [Azimov 1942]

1. A robot may not injure a human being, or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law

- A robot may not injure humanity, or, through inaction, allow humanity to come to harm

UK Principles of Robotics [EPSRC 2011]

1. Robots are multi-use tools. Robots should not be designed solely or primarily to kill or harm humans, except in the interests of national security.
2. Humans, not robots, are responsible agents. Robots should be designed & operated as far as is practicable to comply with existing laws & fundamental rights freedoms, including privacy.
3. Robots are products. They should be designed using processes which assure their safety and security.
4. Robots are manufactured artefacts. They should not be designed in a deceptive way to exploit vulnerable users; instead their machine nature should be transparent.
5. The person with legal responsibility for a robot should be attributed.

Summary

- How to think or how to behave? Being like humans or being rational?
- This course about acting rationally

- AI related to many fields including philosophy, mathematics, economics, neuroscience, psychology, computer sci. and control theory
- 50+ years of progress along many different paradigms: logic, expert systems, neural nets, learning, probabilities
- Increasingly scientific: focus on experimental comparisons and theoretical foundations

- AI is a high-risk high gain area with major ethical implications