CS 5/7320 Artificial Intelligence

Introduction
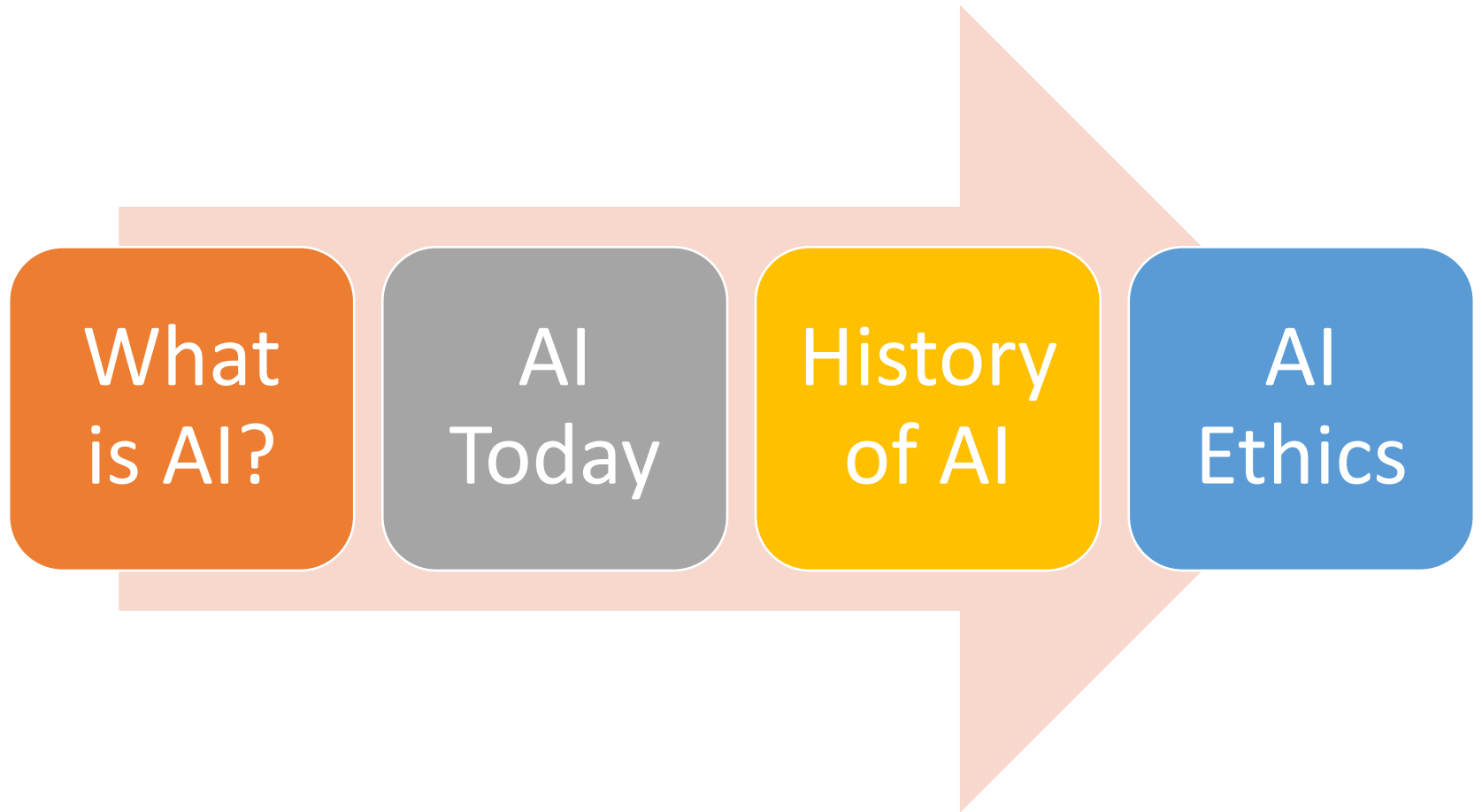
AIMA Chapter 1

Slides by Michael Hahsler
based on slides by Svetlana Lazepnik with figures from the AIMA textbook.

MAURICE ASHLEY: After Ngf6 Deep
has responded instantly with Bf1-d3,
developing the bishop, putting it on a
solid square. Potentially Kasparov mi
castle king-side, so the

Topics

What is AI?

ASIMO (**Advanced Step in Innovative Mobility**) is a humanoid robot created by Honda in 2000

What is it the goal of AI?

Create an agent that

| thinks like a human? | acts like a human? | thinks rationally? | acts rationally? |

Thinking Like a Human

| The brain as an information processing machine.	How to understand cognition as a computational process?
• Requires scientific theories of how the brain works	• Introspection: try to think about how we think. • Predict the behavior of human subjects. • Image the brain, examine neurological data

= Cognitive Sciences

Acting Like a Human

- Alan Turing (1950) "Computing machinery and intelligence"
- The Turing Test tries to define what acting like a human means
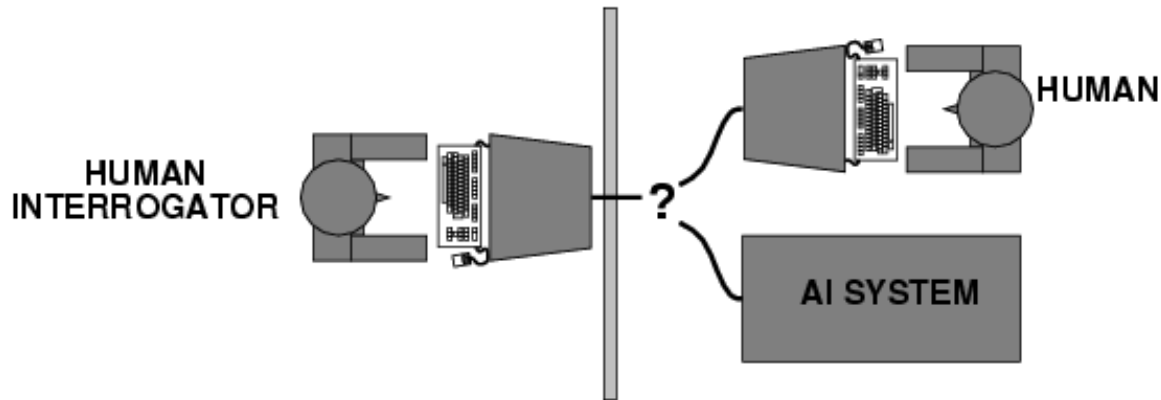
- What capabilities would a computer need to have to pass the Turing Test?
  - Natural language processing
  - Knowledge representation
  - Automated reasoning
  - Probably also machine learning

- Turing predicted that by the year 2000, machines would be able to fool 30% of human judges for five minutes.

Turing Test: Criticism

What are some potential problems with the Turing Test?

- Some human behavior is not intelligent.
- Some intelligent behavior may not be human.
- Human observers may be easy to fool
  - A lot depends on expectations
  - *Anthropomorphic fallacy* (humans tend to humanize things)
  - Chatbots, e.g., ELIZA imitate without intelligence.
  - Chinese room argument: one may simulate intelligence without having true intelligence (more of a philosophical objection)

Is passing the Turing test a good scientific goal?

- Engineering perspective: Not a good way to solve practical problems.
- We can create intelligent agents without trying to imitate humans.

Thinking rationally

- Idealized or "right" way of thinking.
- **Logic:** patterns of argument that always yield correct conclusions when supplied with correct premises
  - "Socrates is a man; all men are mortal; therefore Socrates is mortal."
  - Beginning with Aristotle, philosophers and mathematicians have attempted to formalize the rules of logical thought.
- *Logic*-**based approach to AI:** describe problem in formal logical notation and apply general deduction procedures to solve it
- Problems with the logic-based approach to AI
  - Describing real-world problems and knowledge in logical notation.
  - Computational complexity of finding the solution.
  - A lot of intelligent or "rational" behavior in an uncertain world has nothing to do with logic.

What about the following logical implication:

$$study\ hard \Rightarrow A\ in\ AI$$

Acting rationally: Rational agent

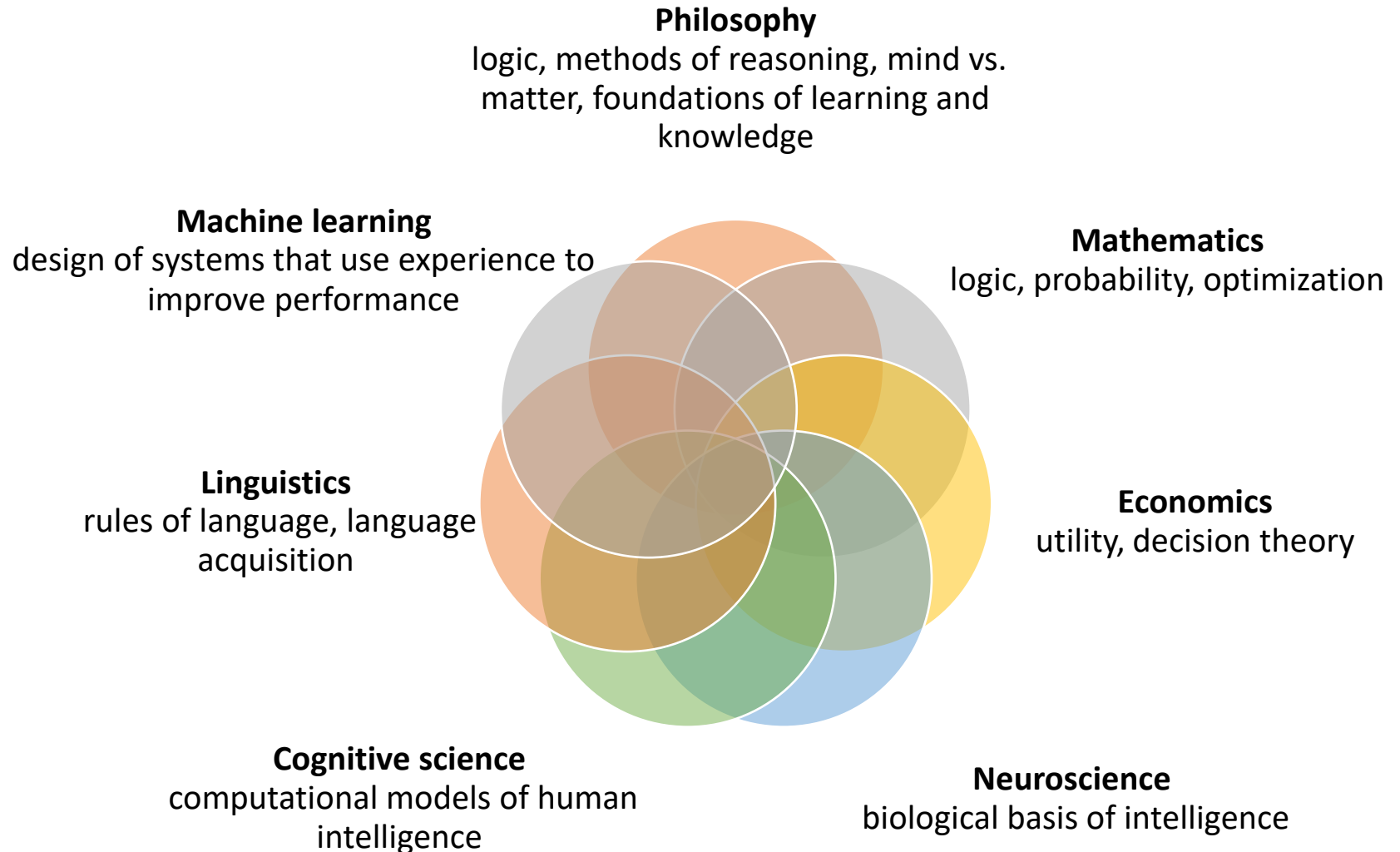A rational agent acts to achieve the best expected outcome:

- Goals are application-dependent and are expressed in terms of the **utility of outcomes.**
- Being rational means acting to **maximizing your expected utility.** Expectation means that different outcomes are possible (probabilities).
- In practice, utility optimization is subject to the agent's computational constraints (**bounded rationality** or bounded optimality).
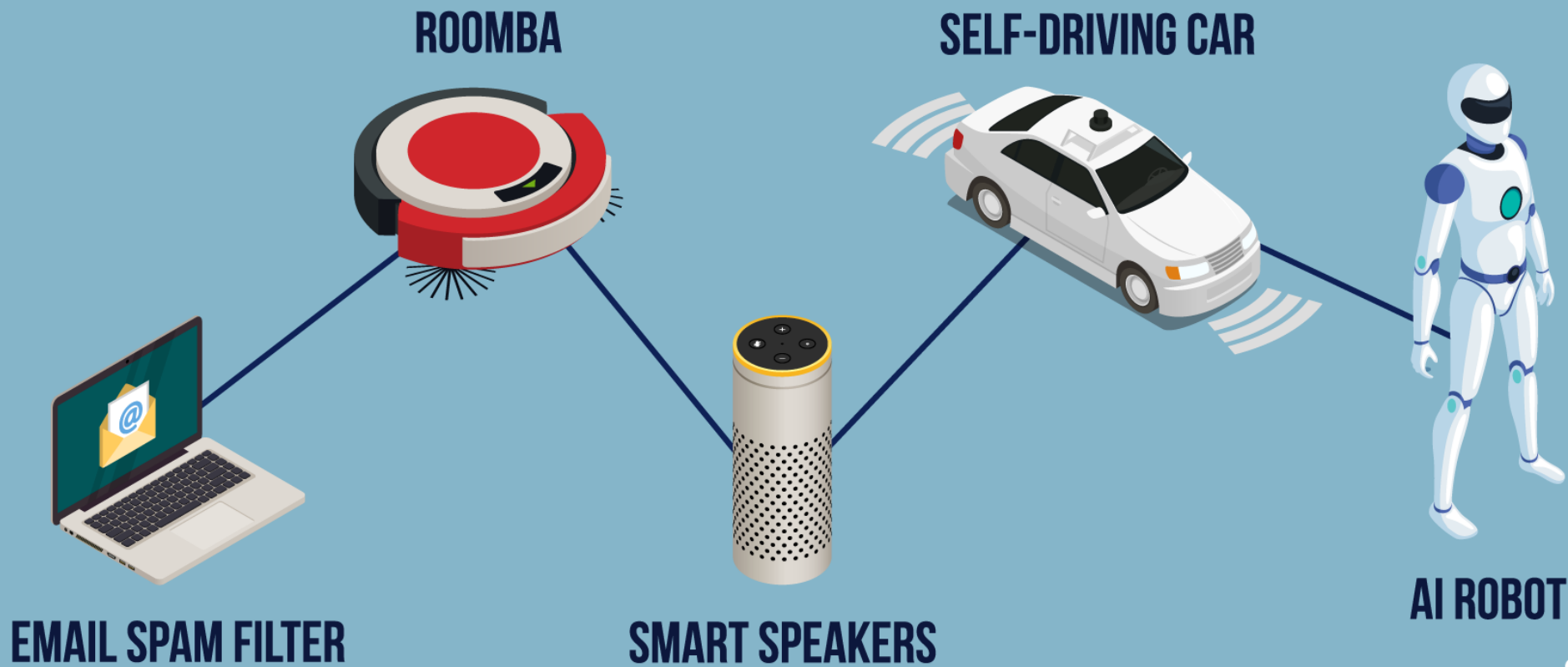
Acting rationally: Rational agent

Advantages of the "expected utility maximization" formulation

- **Generality**: an optimization that goes beyond explicit reasoning with rules.
- **Practicality**: can be adapted to many real-world problems.
- Amenable to good scientific and engineering methodology including simulation and experimentation.
- Only concerns the decisions/actions that are made, not the cognitive process behind them. Avoids philosophy and psychology (at least mostly) in favor of a clearly defined objective.

AI Connections

Philosophy
logic, methods of reasoning, mind vs. matter, foundations of learning and knowledge

Machine learning
design of systems that use experience to improve performance

Mathematics
logic, probability, optimization

Linguistics
rules of language, language acquisition

Economics
utility, decision theory

Cognitive science
computational models of human intelligence

Neuroscience
biological basis of intelligence

ROOMBA

SELF-DRIVING CAR

EMAIL SPAM FILTER

SMART SPEAKERS

AI ROBOT

AI Today

IBM Watson

- http://www.research.ibm.com/deepqa/
- NY Times article
- Trivia demo
- YouTube video
- IBM Watson wins on Jeopardy (February 2011)
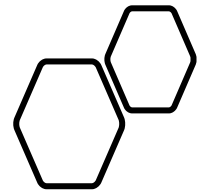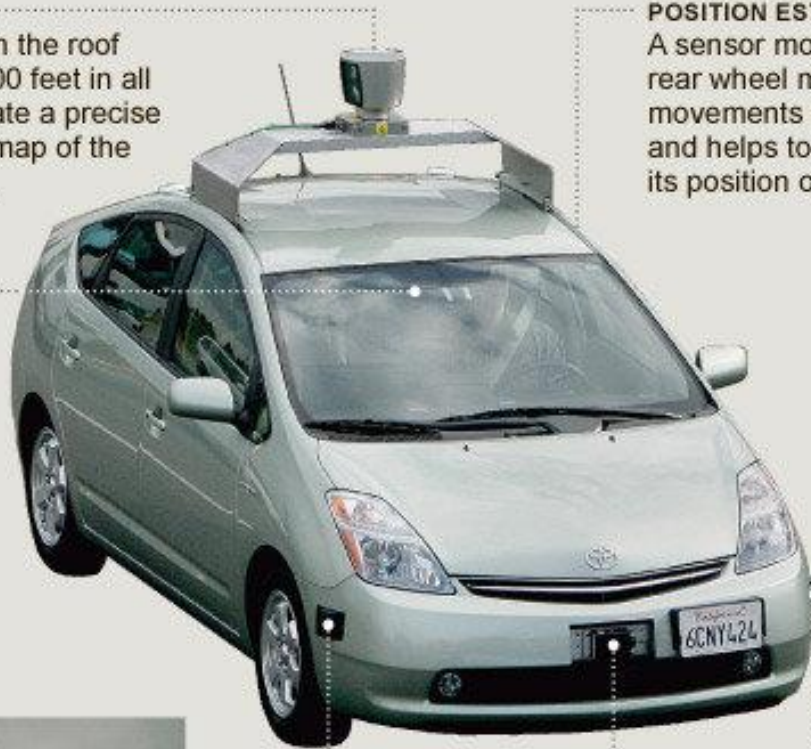
Autonomous Driving

Google's modified Toyota Prius uses an array of sensors to navigate public roads without a human driver. Other components, not shown, include a GPS receiver and an inertial motion sensor.

LIDAR
A rotating sensor on the roof scans more than 200 feet in all directions to generate a precise three-dimensional map of the car's surroundings.

POSITION ESTIMATOR
A sensor mounted on the left rear wheel measures small movements made by the car and helps to accurately locate its position on the map.

VIDEO CAMERA
A camera mounted near the rear-view mirror detects traffic lights and helps the car's onboard computers recognize moving obstacles like pedestrians and bicyclists.
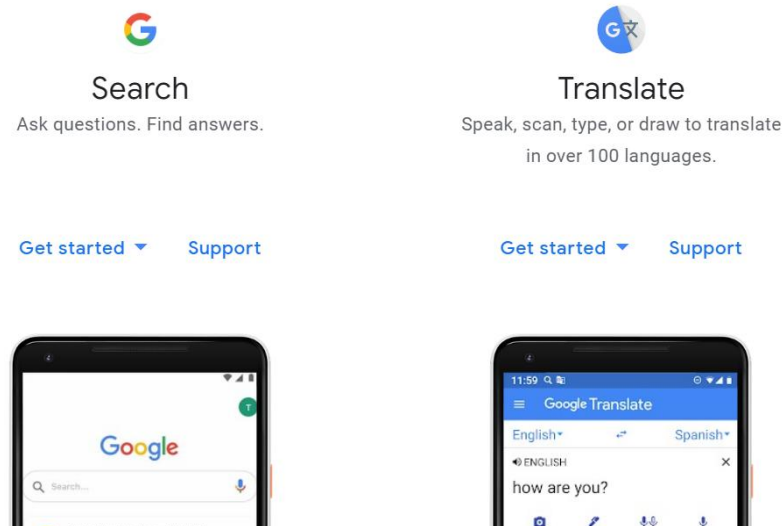
RADAR
Four standard automotive radar sensors, three in front and one in the rear, help determine the positions of distant objects.

Source: Google

Natural Language

- Speech technologies
  - Text-to-speech synthesis
  - Automatic speech recognition
    - Google voice search

- Machine translation
  - translate.google.com
  - Comparison of several translation systems

Search
Ask questions. Find answers.

Get started   Support

Translate
Speak, scan, type, or draw to translate in over 100 languages.

Get started   Support

Source	The Original Text	Human Translation	Google Translate
French "Le Petit Prince" ("The Little Prince") By Antoine de Saint-Exupéry	Le premier soir je me suis donc endormi sur le sable à mille milles de toute terre habitée. J'étais bien plus isolé qu'un naufragé sur un radeau au milieu de l'océan. Alors vous imaginez ma surprise, au lever du jour, quand une drôle de petite voix m'a réveillé. Elle disait: -S'il vous plaît... dessine-moi un mouton!	On the first night, I fell asleep on the sand, a thousand miles from any human habitation. I was far more isolated than a shipwrecked sailor on a raft in the middle of the ocean. So you can imagine my surprise at sunrise when an odd little voice woke me up. It said: "Please ... draw me a sheep." - Wordsworth Children's Classics, 1995	The first night I went to sleep on the sand a thousand miles from any human habitation. I was more isolated than a shipwrecked sailor on a raft in the middle of the ocean. So imagine my surprise at daybreak, when a funny little voice woke me. She said: "If it pleases you ... draw me a sheep!"

Vision

- OCR, read license plates, handwriting recognition (e.g., mail sorting)

- Face detection/recognition: now standard for smart phone cameras

- Visual search: Google Google Lense
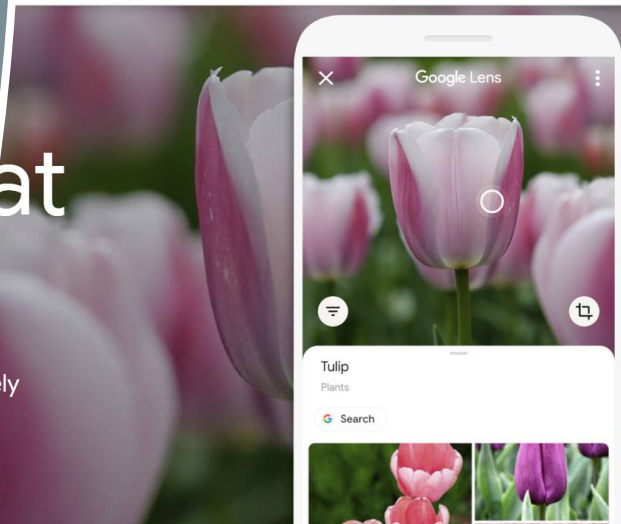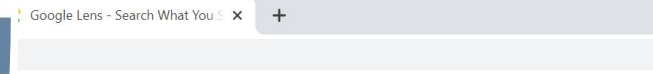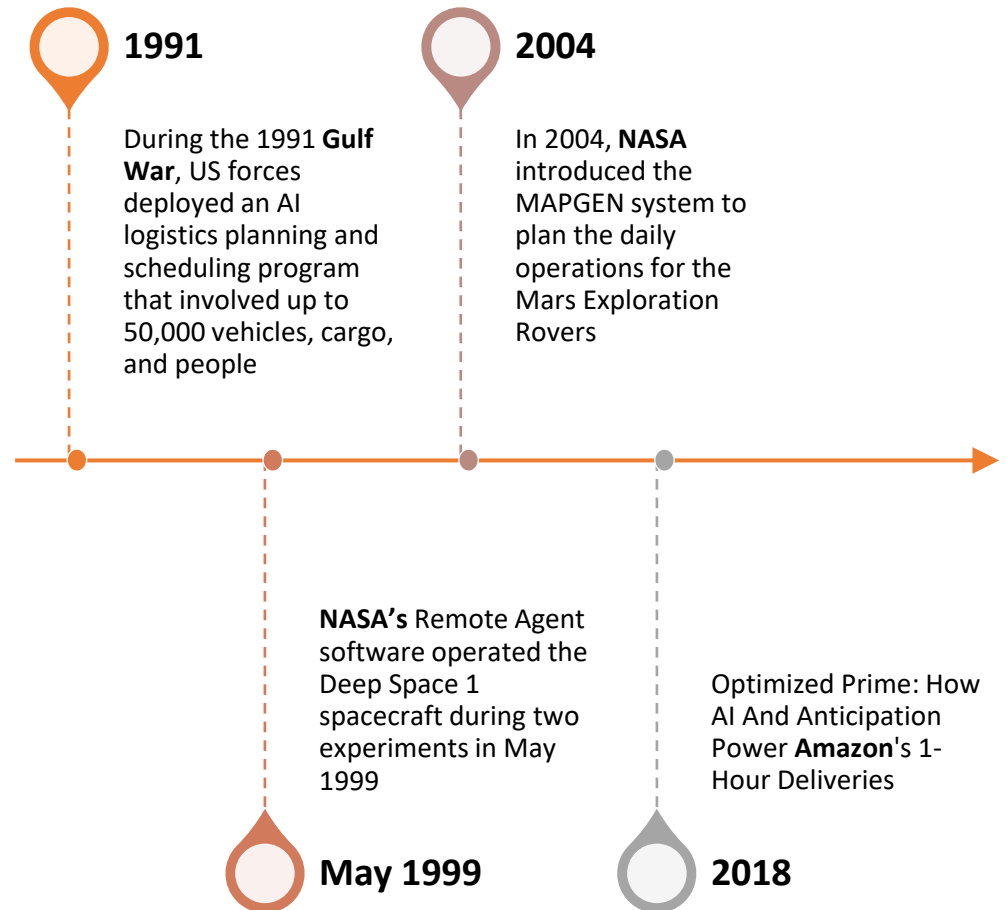
- Vehicle safety systems: Mobileye

Math, games, puzzles

- 1996: A computer program written by researchers at Argonne National Laboratory proved a mathematical conjecture (Robbins conjecture) unsolved for decades
    - NY Times story: "[The proof] would have been called creative if a human had thought of it"

- 1996/97: IBM's Deep Blue defeated the reigning world chess champion Garry Kasparov in 1997
    - **1996: Kasparov Beats Deep Blue**
      "I could feel --- I could smell --- a new kind
      of intelligence across the table."
    - **1997: Deep Blue Beats Kasparov**
      "Deep Blue hasn't proven anything."

- 2007: Checkers was "solved" --- a computer system that never loses was developed. Science article

- 2017+: AlphaZero learns chess, shogi and go by playing itself.
  Science article

Logistics Scheduling Planning

1991

During the 1991 **Gulf War**, US forces deployed an AI logistics planning and scheduling program that involved up to 50,000 vehicles, cargo, and people

2004

In 2004, **NASA** introduced the MAPGEN system to plan the daily operations for the Mars Exploration Rovers

NASA's Remote Agent software operated the Deep Space 1 spacecraft during two experiments in May 1999

May 1999

Optimized Prime: How AI And Anticipation Power **Amazon**'s 1-Hour Deliveries

2018

Information agents

- Search engines
- Recommendation systems
- Spam filtering
- Automated helpdesks
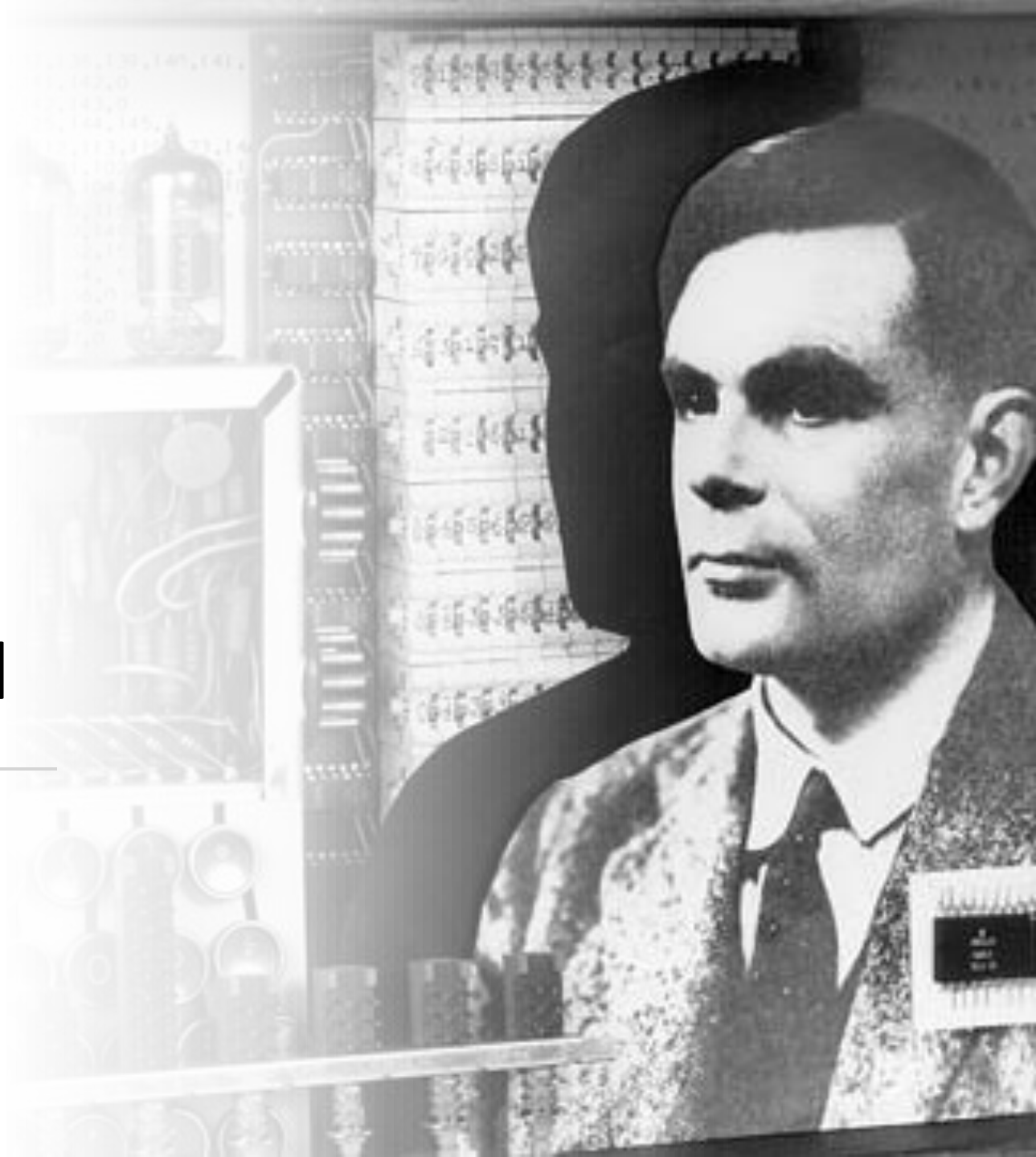- Medical diagnosis systems
- Fraud detection
- Automated trading

Robotics

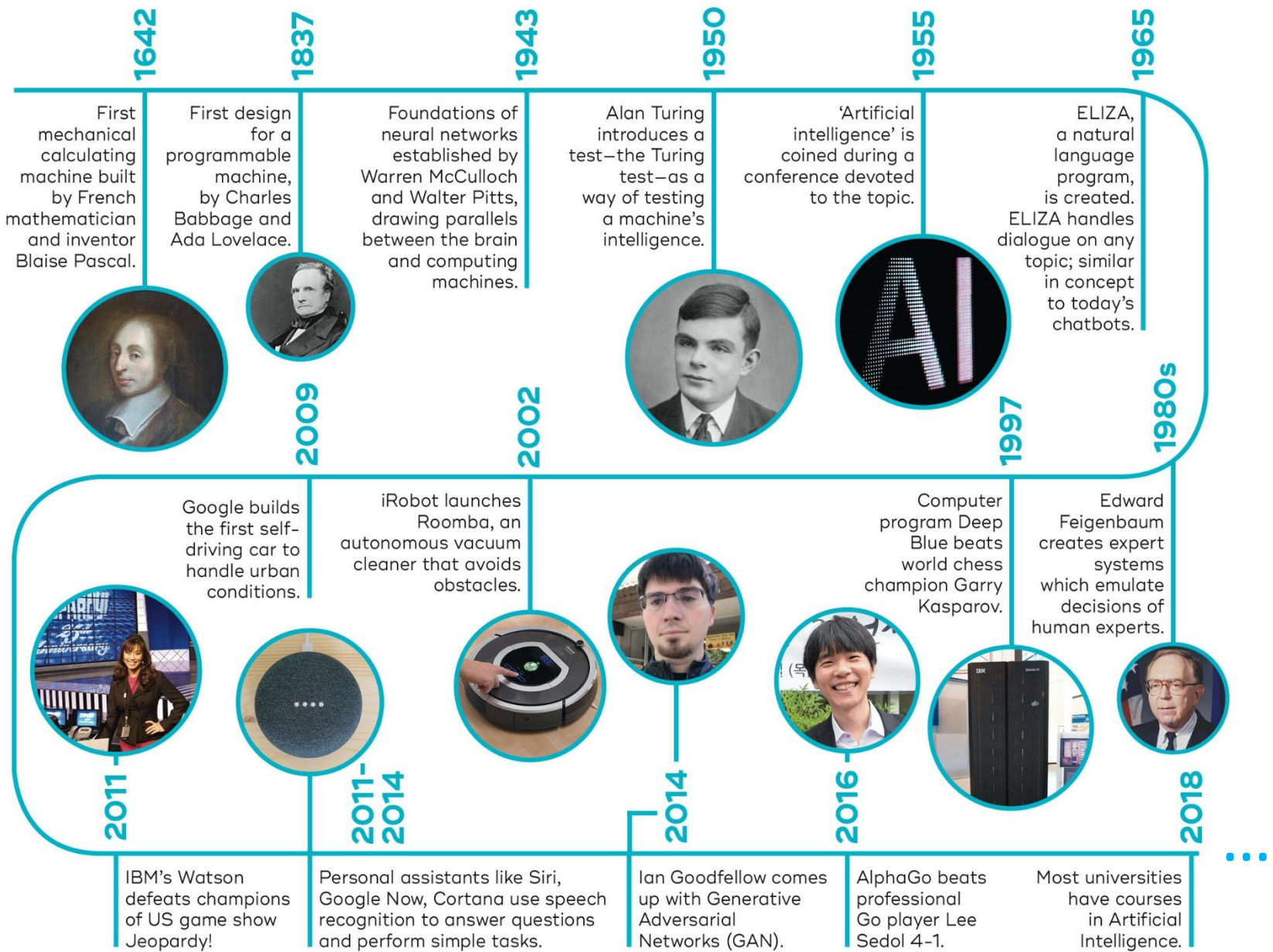- Mars rovers
- Autonomous vehicles
  - DARPA Grand Challenge
  - Google self-driving cars
- Autonomous helicopters and drones
- Robot soccer
  - RoboCup
- Personal robotics
  - Humanoid robots
  - Robotic pets
  - Personal assistants?

History of AI

1642 — First mechanical calculating machine built by French mathematician and inventor Blaise Pascal.

1837 — First design for a programmable machine, by Charles Babbage and Ada Lovelace.

1943 — Foundations of neural networks established by Warren McCulloch and Walter Pitts, drawing parallels between the brain and computing machines.

1950 — Alan Turing introduces a test—the Turing test—as a way of testing a machine's intelligence.

1955 — 'Artificial intelligence' is coined during a conference devoted to the topic.

1965 — ELIZA, a natural language program, is created. ELIZA handles dialogue on any topic; similar in concept to today's chatbots.

1980s — Edward Feigenbaum creates expert systems which emulate decisions of human experts.

1997 — Computer program Deep Blue beats world chess champion Garry Kasparov.

2002 — iRobot launches Roomba, an autonomous vacuum cleaner that avoids obstacles.

2009 — Google builds the first self-driving car to handle urban conditions.

2011 — IBM's Watson defeats champions of US game show Jeopardy!

2011–2014 — Personal assistants like Siri, Google Now, Cortana use speech recognition to answer questions and perform simple tasks.

2014 — Ian Goodfellow comes up with Generative Adversarial Networks (GAN).

2016 — AlphaGo beats professional Go player Lee Sedol 4–1.

2018 — Most universities have courses in Artificial Intelligence.

•••

Source: https://qbi.uq.edu.au/brain/intelligent-machines/history-artificial-intelligence

AI is harder than originally thought

Herbert Simon, 1957

"It is not my aim to surprise or shock you--- but … there are now in the world machines that think, that learn and that create. Moreover, their ability to do these things is going to increase rapidly until---in a visible future---the range of problems they can handle will be coextensive with the range to which human mind has been applied. **More precisely: within 10 years a computer would be chess champion, and an important new mathematical theorem would be proved by a computer***."*
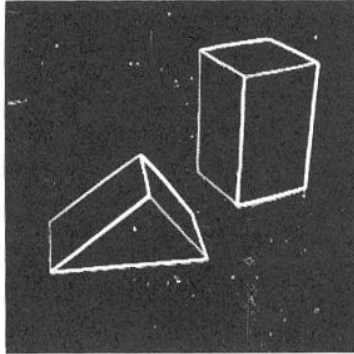
Simon's prediction came true --- but 40 years later instead of 10
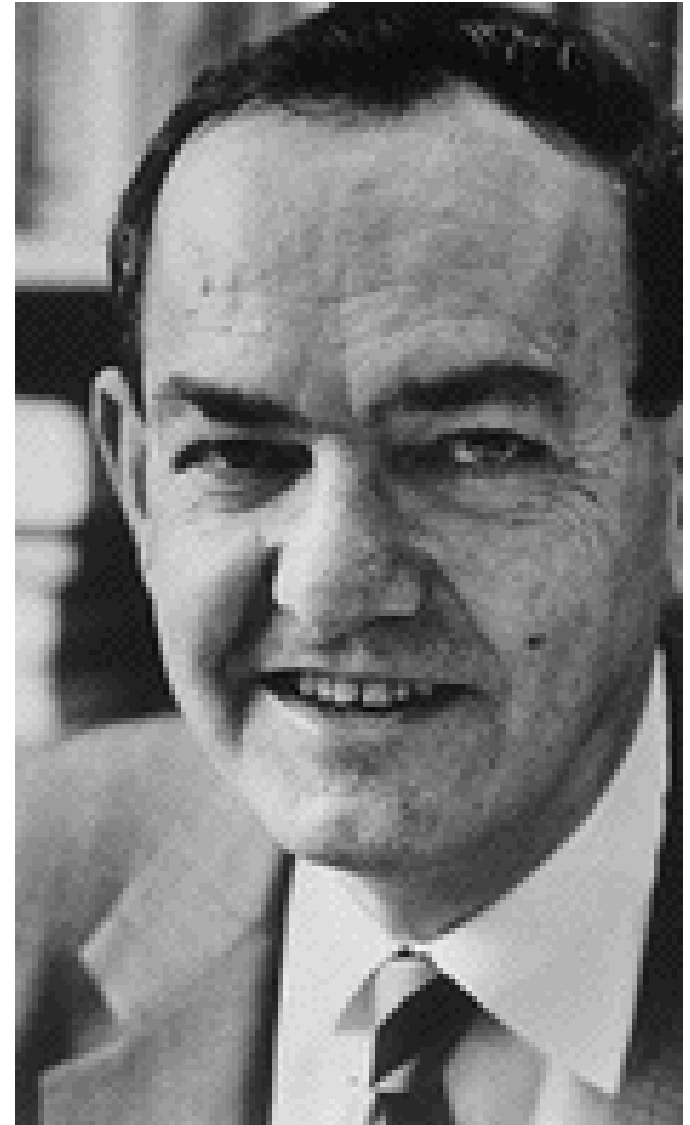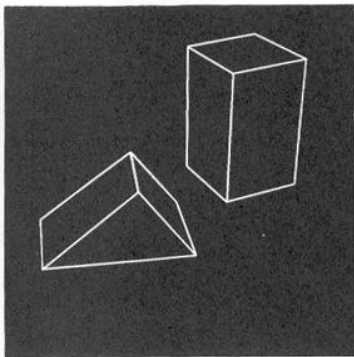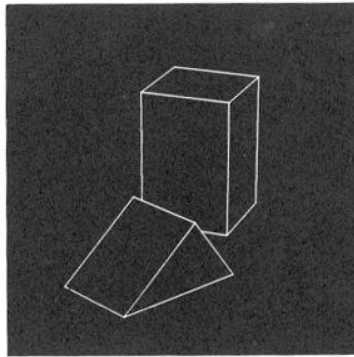
Blocks world (1960s – 1970s)
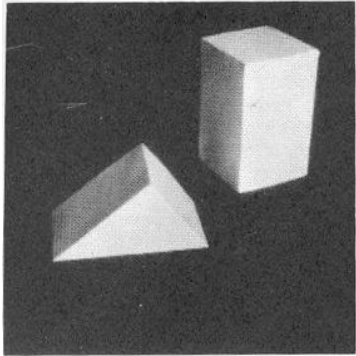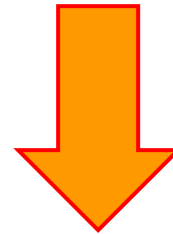

(a) Original picture.

(b) Differentiated picture.

(c) Line drawing.

(d) Rotated view.

Roberts, 1963

???

"Moravec's Paradox"

Hans Moravec (1988): *"It is comparatively easy to make computers exhibit adult level performance on intelligence tests or playing checkers, and **difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility**."*

Why is this?

- Early AI researchers concentrated on the tasks that "white male scientists" found the most challenging, abilities of animals and two-year-olds were overlooked.
- We are least conscious of what our brain does best.
- Sensorimotor skills took millions of years to evolve.
- Our brains were not designed for abstract thinking.

https://www.newsweek.com/googles-new-two-legged-robot-future-warfare-429831

What accounts for recent successes in AI?

- Faster computers
  - The IBM 704 vacuum tube machine that played chess in 1958 could do about **50,000 calculations per second**
  - Deep Blue could do **50 billion calculations per second** – a million times faster!
- Lots of storage (Cloud)
- Lots of data (Sensors, IoT)
- Dominance of statistical approaches, machine learning, deep learning.

AI gets no respect?

- [AI effect](): As soon as a machine gets good at performing some task, the task is no longer considered to require much intelligence
- Calculating ability used to be prized – not anymore
- Chess was thought to require high intelligence
  - Now, massively parallel computers essentially use brute force search to beat grand masters
- Learning once thought uniquely human
  - Ada Lovelace (1842): "The Analytical Engine has no pretensions to *originate* anything. It can do *whatever we know how to order it* to perform."
  - Now machine learning is a well-developed discipline
- Similar picture with animal intelligence
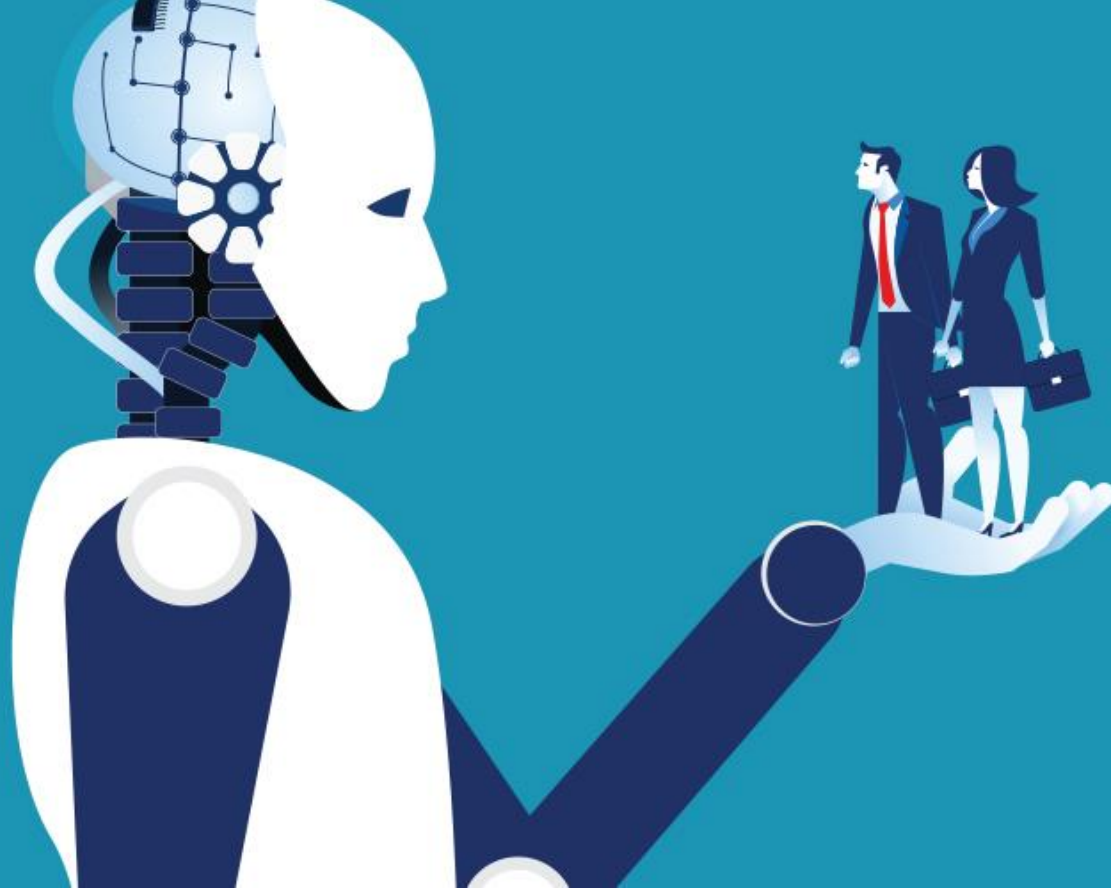
AI Ethics
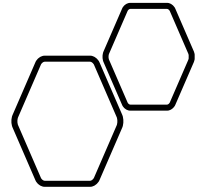
A new Frontier for Fairness and Freedom

Australia: AI Ethics Framework for Industry
A set of **voluntary AI Ethics Principles** (2019)

Core principles for AI

1. Generates net-benefits. The AI system must generate benefits for people that are greater than the costs.

2. Do no harm. Civilian AI systems must not be designed to harm or deceive people and should be implemented in ways that minimise any negative outcomes.

3. Regulatory and legal compliance. The AI system must comply with all relevant international, Australian Local, State/Territory and Federal government obligations, regulations and laws.

4. Privacy protection. Any system, including AI systems, must ensure people's private data is protected and kept confidential plus prevent data breaches which could cause reputational, psychological, financial, professional or other types of harm.

5. Fairness. The development or use of the AI system must not result in unfair discrimination against individuals, communities or groups. This requires particular attention to ensure the "training data" is free from bias or characteristics which may cause the algorithm to behave unfairly.

6. Transparency & Explainability. People must be informed when an algorithm is being used that impacts them and they should be provided with information about what information the algorithm uses to make decisions.

7. Contestability. When an algorithm impacts a person there must be an efficient process to allow that person to challenge the use or output of the algorithm.

8. Accountability. People and organisations responsible for the creation and implementation of AI algorithms should be identifiable and accountable for the impacts of that algorithm, even if the impacts are unintended.

https://consult.industry.gov.au/strategic-policy/artificial-intelligence-ethics-framework/supporting_documents/ArtificialIntelligenceethicsframeworkdiscussionpaper.pdf

European Union

Has regulations since 2016 included in the General Data Protection Regulation (GDPR)

Art. 22 GDPR – Automated individual decision-making, including

Art. 22 GDPR
Automated individual decision-making, including profiling

1. The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.

2. Paragraph 1 shall not apply if the decision:

   (a) is necessary for entering into, or performance of, a contract between the data subject and a data controller;

   (b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or

   (c) is based on the data subject's explicit consent.

3. In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.

4. Decisions referred to in paragraph 2 shall not be based on special categories of personal data referred to in Article 9(1), unless point (a) or (g) of Article 9(2) applies and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place.

The US

116TH CONGRESS
1ST SESSION **H. R. 2231**

To direct the Federal Trade Commission to require entities that use, store, or share personal information to conduct automated decision system impact assessments and data protection impact assessments.

(2) AUTOMATED DECISION SYSTEM IMPACT ASSESSMENT.—The term "automated decision system impact assessment" means a study evaluating an automated decision system and the automated decision system's development process, including the design and training data of the automated decision system, for impacts on accuracy, fairness, bias, discrimination, privacy, and security that includes, at a minimum—

(A) a detailed description of the automated decision system, its design, its training, data, and its purpose;

(B) an assessment of the relative benefits and costs of the automated decision system in light of its purpose, taking into account relevant factors, including—

(i) data minimization practices;

(ii) the duration for which personal information and the results of the automated decision system are stored;

(iii) what information about the automated decision system is available to consumers;

(iv) the extent to which consumers have access to the results of the automated decision system and may correct or object to its results; and

(v) the recipients of the results of the automated decision system;

(C) an assessment of the risks posed by the automated decision system to the privacy or security of personal information of consumers and the risks that the automated decision system may result in or contribute to inaccurate, unfair, biased, or discriminatory decisions impacting consumers; and

(D) the measures the covered entity will employ to minimize the risks described in subparagraph (C), including technological and physical safeguards.

European Union Study (2019)

European Parliament

This study develops policy options for the governance of algorithmic transparency and accountability, based on an analysis of the social, technical and regulatory challenges posed by algorithmic systems. Based on a review and analysis of existing proposals for governance of algorithmic systems, a set of four policy options are proposed, each of which addresses a different aspect of algorithmic transparency and accountability: 1. awareness raising: education, watchdogs and whistleblowers; 2. accountability in public-sector use of algorithmic decision-making; 3. regulatory oversight and legal liability; and 4. global coordination for algorithmic governance.

https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_STU(2019)624262

Algorithmic Bias and Fairness

"**Algorithmic bias** describes systematic and repeatable errors in a computer system that create unfair outcomes, such as privileging one arbitrary group of users over others" Wikipedia

Pre-existing bias

- Social and institutional norms influence design and training data choices.
- For example: Evaluate job applicants for a job which is historically almost exclusively held by males.

Technical bias

- Limitations of a program or computational power.
- For example: instead of a random sample, the program uses the first n data points.

Emergent bias

- Use of algorithms for new data without checking for bias (e.g., existing correlations in the data).
- Use of an algorithm for an unanticipated application.

What type of AI do we cover in this course?

Create an agent that

~~thinks like a human?~~   ~~acts like a human?~~   ~~thinks rationally?~~   **acts rationally.**
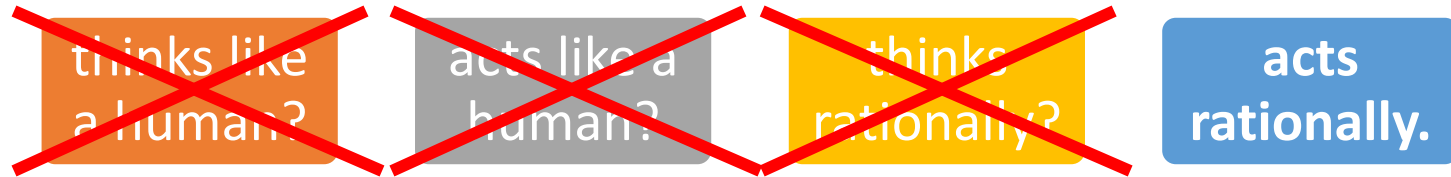
That is, use machines to solve hard problems that traditionally would have been thought to require human intelligence.