Artificial Intelligence

Chapter 1, Sections 1–3

What is Intelligence?

The dream of AI has been to build...

"...machines that can think, that learn and that create."

"The question of whether Machines Can Think...
...is about as relevant as the question whether Submarines Can Swim."

Dijkstra (1984)

Strong and Weak AI

One may dream about...

...that computers can be made to think on a level at least equal to humans, that they can be conscious and experience emotions.

Strong AI

This course is about...

...adding "thinking-like" features to computers to make them more useful tools. That is, "not obviously machine like".

Weak AI

Weak AI

Weak AI is a category that is flexible, as soon as we understand how an AI-program works, it appears less "intelligent".

And as soon as a part of AI is successful, it becomes an own research area! E.g. large parts of advanced search, parts of language understanding, parts of machine learning and probabilistic learning etc.

And AI is left with the remaining hard-to-solve problems!

Contributing research fields

- philosophy
- mathematics
- economics
- neuroscience
- psychology
- computer engineering
- control theory and cybernetics
- linguistics

What is AI?

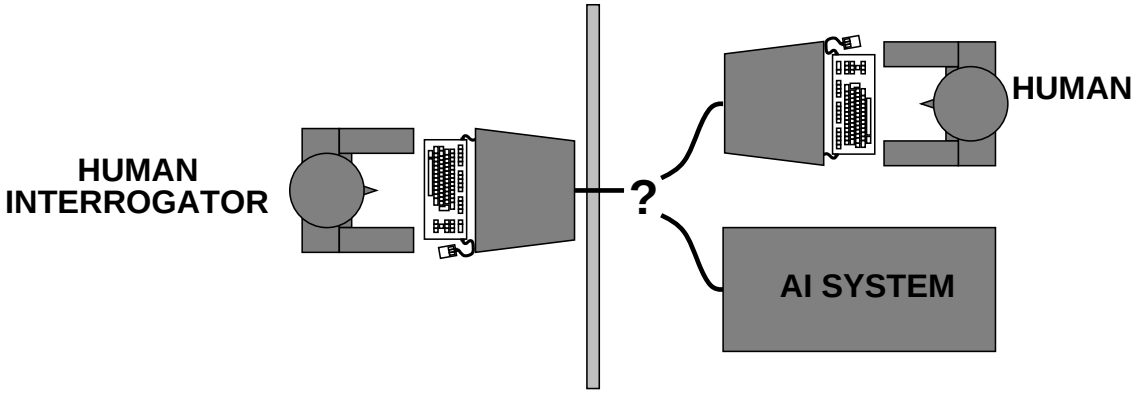
Systems that...

...think like humans?	...think rationally?
...act like humans?	...act rationally?

Acting humanly: The Turing test

Turing (1950) "Computing machinery and intelligence":

- "Can machines think?" $\longrightarrow$ "Can machines behave intelligently?"
- Operational test for intelligent behavior: the Imitation Game



- Predicted that by the year 2000, a machine might have a 30% chance of fooling a lay person for 5 minutes
- Anticipated all major arguments against AI in the following 50 years
- Suggested major components of AI: knowledge, reasoning, language understanding, learning

Problem: Turing test is not **reproducible**, **constructive**, or amenable to **mathematical analysis**

Thinking humanly: Cognitive Science

1960s: "Cognitive revolution"

Requires scientific theories of internal activities of the brain

- What level of abstraction? "Knowledge" or "circuits"?
- How to validate? Requires
  1) Predicting and testing behavior of human subjects (top-down)
  or 2) Direct identification from neurological data (bottom-up)

Both approaches (roughly, Cognitive Science and Cognitive Neuroscience) are now distinct from AI

Both share with AI the following characteristic:

the available theories do not explain (or engender) anything resembling human-level general intelligence

Hence, all three fields share one principal direction!

Thinking rationally: Laws of Thought

Normative (or prescriptive) rather than descriptive

Aristotle: what are correct arguments/thought processes?

Several Greek schools developed various forms of logic:
notation and **rules of derivation** for thoughts;
may or may not have proceeded to the idea of mechanization

Direct line through mathematics and philosophy to modern AI

Problems:
1) Not all intelligent behavior is mediated by logical deliberation
2) What is the purpose of thinking? What thoughts **should** I have
out of all the thoughts (logical or otherwise) that I **could** have?

Acting rationally: Rational agents

Rational behavior: "doing the right thing", i.e., that which is expected to maximize goal achievement, given the available information

— doesn't necessarily involve thinking (e.g., blinking reflex), but thinking should be in the service of rational action

An agent is an entity that perceives and acts
This course (and the course book) is about designing rational agents

Abstractly, an agent is a function from percept histories to actions:

$$f : \mathcal{P}^* \rightarrow \mathcal{A}$$

For any given class of environments and tasks, we seek the agent (or class of agents) with the best performance

Caveat: **computational limitations make perfect rationality unachievable**
→ design best program for given machine resources

AI prehistory

Philosophy	logic, methods of reasoning mind as physical system foundations of learning, language, rationality
Mathematics	formal representation and proof algorithms, computation, (un)decidability, (in)tractability probability
Economics	formal theory of rational decisions
Neuroscience	plastic physical substrate for mental activity
Psychology	adaptation phenomena of perception and motor control experimental techniques (psychophysics, etc.)
Computer science	algorithms, data structures, hardware
Control theory	homeostatic systems, stability simple optimal agent designs
Linguistics	knowledge representation grammar, interaction

Potted history of AI

1943	McCulloch & Pitts: Boolean circuit model of brain
1950	Turing's "Computing Machinery and Intelligence"
1952–69	Look, Ma, no hands!
1950s	Early AI programs: e.g., Samuel's checkers program, Gelernter's Geometry Engine, Newell & Simon's Logic Theorist and General Problem Solver
1956	Dartmouth meeting: "Artificial Intelligence" adopted
1965	Robinson's complete algorithm for logical reasoning
1966–74	AI discovers computational complexity Neural network research almost disappears
1969–79	Early development of knowledge-based systems
1971	Terry Winograd's Shrdlu dialogue system
1980–88	Expert systems industry booms
1988–93	Expert systems industry busts: "AI Winter"
1985–95	Neural networks return to popularity
1988–	Resurgence of probability; general increase in technical depth "Nouvelle AI": ALife, GAs, soft computing
1995–	Agents, agents, everywhere . . .
1997	IBM Deep Blue beats the World Chess Champion
2001–	Very large datasets: Google gigaword corpus, Wikipedia
2003–	Human-level AI back on the agenda
2011	IBM Watson wins Jeopardy
2012	US state of Nevada permits driverless cars