Introduction

Jihoon Yang

Machine Learning Research Laboratory
Department of Computer Science & Engineering
Sogang University

What is Intelligence?

- A wish-list of general characteristics of intelligence
  - **Perception**: manipulation, interpretation of data provided by sensors
  - **Action**: control, and use of effectors to accomplish a variety of tasks
  - **Reasoning**: adapting behavior to better cope with changing environments, discovery of patterns, learning to reason, plan, and act
  - **Communication**: with other intelligent agents including humans using signals, signs, icons, $\cdots$
  - **Planning**: formulation of plans – sequences or agenda of actions to accomplish externally or internally determined goals
  - $\ldots$

What is Artificial Intelligence (AI)?

- The exciting new effort to make computers think.. *machines with minds*
- AI is the art of creating machines that perform functions that require intelligence when performed by humans
- AI is the study of the computations that make it possible to perceive, reason, and act
- AI is the enterprise of design and analysis of intelligent agents

- **Weak** AI: machines could act *as if* they were intelligent
- **Strong** AI: machines that do so are *actually* consciously thinking (not just *simulating* thinking); shifted to "human-level" or "general" AI that can solve an arbitrarily wide variety of tasks, and do so as well as a human

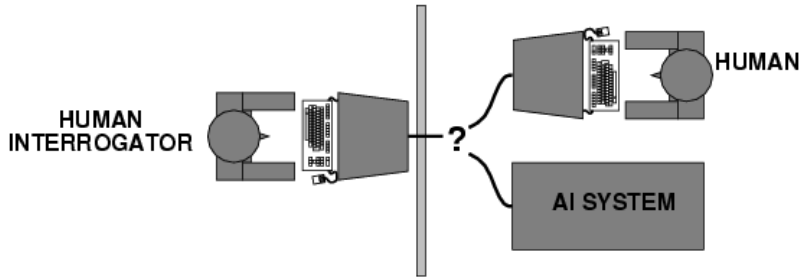
What is AI?

Thinking humanly	Thinking rationally
Acting humanly	Acting rationally

- Are you concerned with *thought process/reasoning* or *behavior*?
- Do you want to model *humans* or measure against an *ideal* concept of intelligence, **rationality**?

Acting humanly: Turing Test

- Alan Turing (1950) "Computing machinery and intelligence":
- "Can machines think?" $\rightarrow$ " Can machines behave intelligently?"
- Operational test for intelligent behavior: the Imitation Game



- Predicted that by 2000, a machine might have a 30% chance of fooling a lay person for 5 minutes
- Annual Loebner prize competition (since 1990): the first prize of $100,000 to be awarded to the first program that passes the "unrestricted" Turing test
- Suggested major components of AI: knowledge, reasoning, language understanding, learning, etc.

Thinking humanly: Cognitive Science

- 1960s "cognitive revolution": information-processing psychology replaced prevailing orthodoxy of behaviorism
- Cognitive science brings together computer models from AI and experimental techniques from psychology to construct precise and testable theories of human mind
- AI and cognitive science fertilize each other
  - Incorporation of neurophysiological evidence into computational models in computer vision
  - Combination of neuroimaging methods with machine learning techniques led to the beginnings of a capacity to "read minds" (i.e. to ascertain the semantic content of a person's inner thoughts), shed further light on how human cognition works

Thinking rationally: Laws of Thought

- Aristotle: what are correct arguments/thought processes?
- Several Greek schools developed various forms of logic: notation and rules of derivation for thoughts
- Direct line through philosophy, mathematics and logic to AI, so-called logicist
- Problems:
  - Not all intelligent *behavior* is mediated by logical deliberation
  - 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 agent

- This course is about designing rational agents
- An agent is an entity that perceives and acts
- Rational behavior: **doing the right thing**
- The right thing: that which is expected to maximize goal achievement, given the available information
- A rational agent is one that acts so as to achieve the best outcome
- Caveat: computational limitations make perfect rationality unachievable $\rightarrow$ design best program for given machine resources

Foundations

- 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: physical substrate for mental activity
- Psychology: experimental techniques (psychophysics, etc.), behaviorism (percept/stimulus and action/response), cognitive psychology/science (views brain as info-processing device)
- Computer engineering: building efficient computers
- Control theory and cybernetics: homeostatic systems, stability, simple optimal agent designs
- Linguistics: knowledge representation, grammar

Brief history of AI

- 1943: McCulloch & Pitts: model of artificial neurons
- 1950: Turing's "Computing Machinery and Intelligence" introduced Turing test, ML, GA, RL
- 1956: McCarthy, Minsky, Rochester, Shannon, et al., Dartmouth workshop: "Artificial Intelligence" adopted
- 1952-69: Early enthusiasm, great expectations, optimism fueled by early success on some problems thought to be hard (e.g. theorem proving); NN flourished – Hebb's learning and its enhancement in Widrow's adalines, Rosenblatt's perceptrons with convergence theorem
- 1965: Robinson's complete algorithm for logical reasoning (resolution)

Brief history of AI

- 1966-73: Collapse in AI research: Progress was slower than expected; Unrealistic predictions, Herbert Simon (1957) chess champion in 10 years; AI discovers computational complexity; NN research almost disappears
- 1969-79: Early development of knowledge-based systems
- 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

Brief history of AI

- Mid 1990s–present: The emergence of intelligent agents in various applications
  - information retrieval
  - data mining and knowledge discovery
  - customized software systems
  - smart devices (e.g. homes, automobiles)
  - agile manufacturing systems
  - autonomous vehicles
  - bioinformatics
  - internet tools: search engines, recommender systems
  - ...
- 2001–: Big data
- 2011–: Deep learning

State of the art

AI100 st Stanford produces AI Index (`aiindex.org`) tracking progress

- Robotic vehicles: DARPA Grand Challenge and Urban Challenge (2005, 2007); Waymo passed landmark of 10 million miles on public roads (2018), followed by commercial robotic taxi service
- Legged locomotion: Boston Dynamics' BigDog (2008) resembles an amimal; Atlas walks, jumps, and backflips (2016)
- Autonomous planning and scheduling: EUROPA planning toolkit for daily operation of NASA's Mars rovers; SAXTANT system for autonomous navigation in deep space; DARPA's DART (Dynamic Analysis and Replanning Tool) for automated logistic planning and scheduling for transportation, deployed during the Gulf war (1991); Dynamic driving directions provided by Uber and Google Maps
- Machine translation: online MT systems produce adequate results
- Speech recognition: human-level, Alexa, Siri, Cortana, Google
- Recommendation: Amazon, Facebook, Netflix, Spotify, YouTube, Walmart, Coupang, ...

State of the art

- Game playing: Deep Blue defeated the world chess champion Garry Kasparov (1997); Chinook defeated human checkers champion (1994), can't lose at checkers (2007); The IBM supercomputer Watson beat human champions on 'Jeopardy!' (2011); AlphaGo/AlphaGoZero/AlphaZero (for Go, chess, shogi)/AlphaStar beat human champions
- Image understanding: ImageNet object recognition task, image captioning, etc.
- Medicine: disease diagnosis with multimodal data
- Climate science, ...