Artificial Intelligence

Lecture1: Introduction

Credit: Ansaf Salleb-Aouissi, and “Artificial Intelligence: A Modern Approach”, Stuart Russell and Peter Norvig, and “The Elements of Statistical Learning”, Trevor Hastie, Robert Tibshirani, and Jerome Friedman, and “Machine Learning”, Tom Mitchell.

Outline

- What is AI
- Application of AI
- Foundation of AI
- History of AI

Definitions of AI

“Intelligence: The ability to learn and solve problems”

Webster’s Dictionary.

“Artificial intelligence (AI) is the intelligence exhibited by machines or software”

Wikipedia.

“The science and engineering of making intelligent machines”

McCarthy.

“The study and design of intelligent agents, where an intelligent agent is a system that perceives its environment and takes actions that maximize its chances of success.”

Russell and Norvig AI book.

Why AI?

“Just as the Industrial Revolution freed up a lot of humanity from physical drudgery, I think AI has the potential to free up humanity from a lot of the mental drudgery.”

Andrew Ng.

What is AI?

Thinking Humanly	Thinking Rationally
"The exciting new effort to make computers think . . . *machines with minds*, in the full and literal sense." (Haugeland, 1985) "[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning . . ." (Bellman, 1978)	"The study of mental faculties through the use of computational models." (Charniak and McDermott, 1985) "The study of the computations that make it possible to perceive, reason, and act." (Winston, 1992)
Acting Humanly	**Acting Rationally**
"The art of creating machines that perform functions that require intelligence when performed by people." (Kurzweil, 1990) "The study of how to make computers do things at which, at the moment, people are better." (Rich and Knight, 1991)	"Computational Intelligence is the study of the design of intelligent agents." (Poole *et al.*, 1998) "AI . . . is concerned with intelligent behavior in artifacts." (Nilsson, 1998)

Four schools of thoughts

Thinking humanly: cognitive approach

- Requires to determine how humans think!
  1960's "cognitive revolution".
  Requires scientific theories of internal activities of the brain
  - What level of abstraction? "Knowledge" or "circuits"?
  - How to validate?
- **Today, Cognitive Science and Artificial Intelligence are distinct disciplines.**

Acting humanly: Turing test

- Turing test (Alan Turing 1950): A computer passes the test of intelligence, if it can fool a human interrogator.
- Major components of AI: knowledge, reasoning, language understanding, learning.

Alan Turing, *Computing Machinery and Intelligence*, 1950.

Thinking rationally: “laws of thought”

- Codify “right thinking” with **logic**.
- Several Greek schools that developed various forms of logic: *notations* and *rules of inference/ derivation* for thoughts.
- Problems:
  1. Not all knowledge can be expressed with logical notations.
  2. Computational blow up.

Acting rationally: rational agent

- Agents do the right thing: which is expected to maximize the goal achievement, given the available information (environment, background knowledge, etc.)
- A **rational agent** is one that acts so as to achieve the ***best*** outcome, or when there is uncertainty (stochastic environment), to achieve the ***best expected*** outcome.
- Aristotle (Nicomachean Ethics):
  "*Every art and every inquiry, and similarly every action and pursuit, is thought to aim at some good*."

What is AI? Our Approach

Thinking Humanly	**Thinking Rationally**
"The exciting new effort to make computers think . . . *machines with minds*, in the full and literal sense." (Haugeland, 1985) "[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning . . ." (Bellman, 1978)	"The study of mental faculties through the use of computational models." (Charniak and McDermott, 1985) "The study of the computations that make it possible to perceive, reason, and act." (Winston, 1992)
Acting Humanly	**Acting Rationally: Our approach**
"The art of creating machines that perform functions that require intelligence when performed by people." (Kurzweil, 1990) "The study of how to make computers do things at which, at the moment, people are better." (Rich and Knight, 1991)	"Computational Intelligence is the study of the design of intelligent agents." (Poole *et al.*, 1998) "AI . . . is concerned with intelligent behavior in artifacts." (Nilsson, 1998)

How to design *intelligent* systems that act rationally in order to achieve their goals.

State-of-the-art applications

- Speech recognition
- Recommendation systems
- Financial forecasting
- Game playing, video games
- Spam filtering
- Logistics planning
- Machine translation
- Autonomous car
- Web search engines
- Automatic assembly
- Sentiment analysis
- Medical diagnosis, imaging
- Computer animation
- Fraud detection
- Social network analysis
- Route finding
- Protein design (bioinformatics)
- Document summarization
- Transportation/scheduling
- Information extraction
- VLSI layout
- Energy optimization
- Question answering systems
- Traveling salesperson
- Autonomous planning and scheduling
- Robotics (household, surgery, navigation)
- etc.

Many more!

AI is not only compelling/interesting, it actually touches many aspects of our lives.

Speech recognition

- Virtual assistants: Siri (Apple), Echo (Amazon), Google Now, Cortana (Microsoft).
- Leverage **deep neural networks** to handle speech recognition and natural language understanding.

Handwriting recognition

- For the history, the USPS was very interested in automatically sorting the handwritten addresses and zip codes on the envelopes.
- LeCun, published a solution using convolutional neural networks, CNN, to recognize the handwritten digits on the envelopes.

Machine translation

- Historical motivation: translate Russian text to English.
- First systems using **mechanical translation (or one-to-one correspondence)** failed!
- "Out of sight, out of mind" → "Invisible, imbecile".

Machine translation

- **Machine translation** has gone through ups and downs.
- Today, **Statistical Machine Translation** leverages a vast amounts of available **translated corpuses**.
- While there is room for improvement, machine translation has made significant progress.

Google Translate: 100+ languages

Recommendation system

- Key technology: **collaborative filtering**

Search engines

- Key technology (near real time): **ranking (PageRank)**

Spam filtering

- The baseline technology: **Naive Bayes classifiers**

Face detection

Viola-Jones algorithm

Face recognition

Cancer detection

- Skin Cancer Detection & Tracking using **Deep Learning**

Chess (1997)

- Garry Kasparov vs. IBM Deep Blue

Powerful search algorithms!

Jeopardy! (2011)

- Ken Jennings vs. IBM Watson

Natural language understanding and information extraction!

Go (2016)

- Lee Sedol (1) vs. Google AlphaGo (4)

Deep Learning, reinforcement learning, and search algorithms!

Robotics

- Awesome robots today! ASIMO, SPOT/Atlas, Unitree, and more!
- Robotics is an interdisciplinary branch of engineering and science
  - **Power source**
  - **Actuation**
  - **Sensing**
  - **Manipulation**
  - **Locomotion**

Autonomous driving

The Foundation of AI

- AI is an **interdisciplinary field** and several other disciplines have contributed to the progress of AI.

Philosophy

- Can formal rules be used to draw valid conclusions?
- How does the mind arise from a physical brain?
- Where does knowledge come from?
- How does knowledge lead to action?

Aristotle
(384–322 B.C.)

✓Logic, methods of reasoning.

✓Mind as a physical system that operates as a set of logical rules.

✓Foundations of learning, language, rationality.

Mathematics

- What are the formal rules to draw valid conclusions?
- What can be computed?
- How do we reason with uncertain information?

✓Logic: Formal representation and proof.
✓Computation, algorithms.
✓Probability.

George Boole
(1815–1864)

Thomas Bayes
(1702–1761)

Economics

- How should we make decisions so as to maximize payoff?
- How should we do this when others may not go along?
- How should we do this when the payoff may be far in the future?

✓Formal theory of rational decisions.

✓Combined decision theory and probability theory for decision making under uncertainty.

✓Game theory.

✓Markov decision processes.

Adam Smith
(1723–1790)

Herbert Simon
(1916–2001)

Neuroscience

- How brains process information?
- How brains and computers are (dis)similar?

Psychology

- How do humans and animals think and act?

✓Cognitive psychology specifically perceives the brain as an information processing machine.

✓Led to the development of the field *cognitive science*: how could computer models be used to study *language*, *memory*, and *thinking* from a psychological perspective.

Computer engineering

- How can we build an efficient computer?

✓E.g., Self-driving cars are possible today thanks to advances in computer engineering.

Control theory and cybernetics

- How can artifacts operate under their own control?

✓Design simple optimal agents receiving feedback from the environment.

✓Modern control theory design systems that maximize an objective function over time.

Linguistics

- How does language relate to thought?

✓Modern linguistics + AI = Computational linguistics (Natural language processing).

The History of AI

- Gestation of AI (1940–1950)
- Early enthusiasm, great expectations (1950–1970)
- Knowledge-based AI (1970–1990)
- AI becomes “scientific” (1990–present)

AI is a broad field with a long history, it went through ups and downs, successes and failures, optimism and disappointment, big enthusiasm with large funding and then cutting funding and so on and so forth.

Gestation of AI (1940–1950)

- McCulloch and Pitts, model of artificial neurons, 1943
- Alan Turing, *Computing Machinery and Intelligence*, 1950.

Alan Turing (1912-1954)
(Turing test, machine learning, genetic algorithms, and reinforcement learning)

Early enthusiasm (1950–1970)

- Early AI programs, Samuel's checkers program
- Birth of AI @ Dartmouth meeting 1956.

Dartmouth Workshop

John McCarthy
(Lisp Language)

Marvin Minsky
(SNARC)

Claude Shannon
(Information theory)

Ray Solomonoff
(Algorithmic probability)

Allen Newell
(General Problem Solver)

Herbert Simon
(Satisficing)

Arthur Samuel
(Computer checkers)

And three others: Oliver Selfridge (Pandemonium theory), Nathaniel Rochester (Designed IBM 701), and Trenchard More (Natural deduction)

Knowledge-based AI (1970–1990)

- Expert systems, AI becomes an industry
- AI winter

AI becomes “scientific” (1990–present)

- Neural Networks: le retour (reinvented)
- The emergence of intelligent agents
- AI becomes “scientific”, use of probability to model uncertainty
- AI Spring!
- The availability of very large datasets.

Data will drive future discoveries and alleviate the complexity in AI

Summary

- AI is a hard (computational complexity, language, vision, etc.), and a broad field with high impact on humanity and society.
- What can AI do for us is already amazing!
- AI systems do not have to model human/nature but can act like or be inspired by human/nature.
- How human think is beyond the scope of this course.
- **Rational (do the right thing) agents are central to our approach of AI.**
- Note that rationality is not always possible in complicated environment, but we will still aim to build rational agents.

To be continued

ISOMAP

- A global geometric framework for nonlinear dimensionality reduction, *Science* 2000.

LLE

- Nonlinear dimensionality reduction by locally linear embedding, *Science* 2000.

Deep Auto-encoder

- Reducing the dimensionality of data with neural networks, *Science* 2006.

CFDP

- Clustering by fast search and find of density peaks, *Science* 2014.

Bayesian Program Learning

- Human-level concept learning through probabilistic program induction, *Science* 2015.

Deep Q-network

- Human-level control through deep reinforcement learning, *Nature* 2015.

Review: Deep learning

- Deep learning
- Yann LeCun, Yoshua Bengio & Geoffrey Hinton
- *Nature* 2015.

AWARD WINNER

Yann LeCun

ACM A. M. Turing Award (2018)

2018 ACM A.M. Turing Award

AWARD WINNER

Yoshua Bengio

ACM A. M. Turing Award (2018)

2018 ACM A.M. Turing Award

AWARD WINNER

Geoffrey E Hinton

ACM A. M. Turing Award (2018)

2018 ACM A.M. Turing Award

AlphaGo

- Mastering the game of Go with deep neural networks and tree search, *Nature* 2016.

AlphaGo Zero

- Mastering the game of Go without human knowledge, *Nature* 2017.

Skin Cancer Detection

- Dermatologist-level classification of skin cancer with deep neural networks, *Nature* 2017.

AlphaFold

- Improved protein structure prediction using potentials from deep learning, *Nature* 2020.

AlphaFold2

- Highly accurate protein structure prediction with AlphaFold, *Nature* 2021.

Chip Placement

- A graph placement methodology for fast chip design, *Nature* 2021.

Eva

- Efficient and targeted COVID-19 border testing via reinforcement learning, *Nature* 2021.

AlphaTensor

- Discovering faster matrix multiplication algorithms with reinforcement learning, *Nature* 2022.

AlphaDev

- Faster sorting algorithms discovered using deep reinforcement learning, *Nature* 2023.

PanguWeather

- Accurate medium-range global weather forecasting with 3D neural networks, *Nature* 2023.

AlphaGeometry

- Solving olympiad geometry without human demonstrations, *Nature* 2024.

Tearing Avoidance Control

- Avoiding fusion plasma tearing instability with deep reinforcement learning, *Nature* 2024.

Continual Learning

- Loss of plasticity in deep continual learning, *Nature* 2024.

LLM Reliability

- Larger and more instructable language models become less reliable, *Nature* 2024.

GenCast

- Probabilistic weather forecasting with machine learning, *Nature* 2025.

Gameplay Ideation

- World and Human Action Models towards gameplay ideation, *Nature* 2025.

To be continued

ChatGPT - RLHF

DeepSeek – “Aha Moment”

Question: If $a > 1$, then the sum of the real solutions of $\sqrt{a - \sqrt{a + x}} = x$ is equal to

Response: <think>

To solve the equation $\sqrt{a - \sqrt{a + x}} = x$, let's start by squaring both $\cdots$

$\left(\sqrt{a - \sqrt{a + x}}\right)^2 = x^2 \implies a - \sqrt{a + x} = x^2.$

Rearrange to isolate the inner square root term:

$(a - x^2)^2 = a + x \implies a^2 - 2ax^2 + (x^2)^2 = a + x \implies x^4 - 2ax^2 - x + (a^2 - a) = 0$

$\cdots$

Wait, wait. Wait. That's an aha moment I can flag here.

Let's reevaluate this step-by-step to identify if the correct sum can be $\cdots$

We started with the equation:

$\sqrt{a - \sqrt{a + x}} = x$

First, let's square both sides:

$a - \sqrt{a + x} = x^2 \implies \sqrt{a + x} = a - x^2$

Next, I could square both sides again, treating the equation: $\cdots$

$\cdots$

Thank you !