Fondamenti di Intelligenza Artificiale

1. Introduction

What is AI, Anyway?

Prof Sara Bernardini
bernardini@diag.uniroma1.it
www.sara-bernardini.com

SAPIENZA
UNIVERSITÀ DI ROMA

Spring Term

Agenda

1. Introduction
2. AI Concepts
3. AI Foundations
4. AI History
5. AI Today
6. Conclusion

Our Agenda for This Chapter

- **AI Concepts:** What are we actually talking about?

  → Clarify what the (modern) research field of AI does (and does not try) to do.

- **AI Foundations:** Where AI come from?

  → Brief history of the disciplines that contributed ideas, viewpoints, and techniques to AI.

- **AI History:** How did AI come about?

  → Just a little background to illustrate how we came from "classical AI" to "modern AI".

- **AI Today:** What is the landscape of techniques and applications?

  → Rough overview and some examples.

What? Take 1

What is “intelligence”?

See what I mean? It’s impossible to agree on this . . .

- Ability to think . . . ?
- Simulating the brain . . . ?
- Creativity . . . ?
- Ability to learn . . . ?
- Being good at math . . . ?
- Being good at Chess or Go . . . ?
- Passing an IQ test with high marks?

→ There are entire dissertations written on this subject (e.g. in Neuroscience, Philosophy, etc.)

What? Take 2

What is "artificial intelligence"?

	Humanly	**Rationally**
Thinking	*"The exciting new effort to make computers think . . . machines with human-like minds."*	*"The formalization of mental faculties in terms of computational models. "*
Acting	*"The art of creating machines that perform actions requiring intelligence when performed by people."*	*"The branch of CS concerned with the automation of appropriate behavior in complex situations."*

→ "Rational": Performance-oriented as opposed to imitating humans.

What? Recap

What is "artificial intelligence"?

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

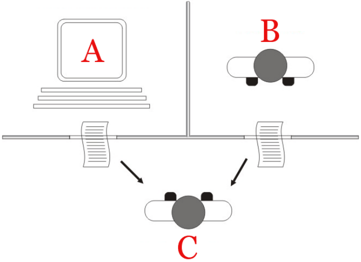
EU Vision

What is “artificial intelligence”?

“Artificial intelligence (AI) refers to systems that display intelligent behavior by analyzing their environment and taking actions – with some degree of autonomy – to achieve specific goals.

AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications).”

Acting Humanly: The Turing Test



- Turing (1950): *"Computing machinery and intelligence"*
- The classical benchmark for AI (for "Acting Humanly").
- Yearly competitions, e.g., Loebner Prize

  → Frequent winners: Richard Wallace with "A.L.I.C.E." and Steve Worswick with "Mitsuku"

A Conversation With Alice

https://www.pandorabots.com/pandora/talk?botid=b8d616e35e36e881

https://www.pandorabots.com/kuki/

So, do the following elements count as “intelligence”?

- Deception: The machine has to pretend to be a human
- Conversation: Avoid answering questions (e.g. using jokes)
- Ambiguous evaluation

Winograd Schema Challenge

- Evaluates NLP: common sense and basic deduction
- Two or more entities are mentioned in a sentence
- Multiple-choice question: identify to who/what a pronoun refers
- The answer changes if a special word is replaced by another

The trophy would not fit in the brown suitcase because it was too big (small). What was too big (small)?

1. the trophy
2. the suitcase

The town councilors refused to give the demonstrators a permit because they feared (advocated) violence. Who feared (advocated) violence?

1. the town councilors
2. the demonstrators

→In AAAI 2018, 25,000$ for getting more than 90% accuracy

So, What Does Modern AI Do?

There's a lot to say about the 4 AI "categories": (but we'll cut it short)

Acting Humanly: Turing Test. Not much pursued otherwise.

→ ≈ Aeronautics: "Machines that fly so similarly to pigeons that they can even fool other pigeons".

Not reproducible, not amenable to mathematical analysis.

Thinking Humanly: Cognitive Science \Neuro-science. How do humans think, how does the human brain work.

→ Neural networks are an extremely simple approximation.

Thinking Rationally: Logics. Formalization of knowledge and deduction.

→ We cover the basics. Fairly wide-spread in modern AI.

Acting Rationally: How to make good action choices? Doing the right thing.

→ Our main approach. Contains logics (one possible way to make intelligent decisions). Rationality can be formalized mathematically. The right thing is expected to maximize goal achievement, given available information.

Rational agents

- An agent is an entity that perceives and acts.
- This course is about designing rational agents. Another title for the course could be Computational Rationality.
- 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

Philosophy

- Can formal rules be used to draw valid conclusions?

  Aristotle (384–322 B.C.): system of syllogisms for proper reasoning
- Can computation be automated?

  Hobbes (1588–1679): "We add and subtract in our silent thoughts."
  Pascal (1642): Pascaline (mechanical calculator)
- How does the mind arise from a physical brain?

  Descartes (1596–1650): Distinction between the two
- Where does knowledge come from?

  Bacon's (1620): Empiricism Hume's (1739): Problem of induction
- How does knowledge lead to action?

  Aristotle: "We deliberate not about ends, but about means."

Mathematics

The leap from philosophy to AI as a formal science required a level of mathematical formalization in three fundamental areas:

- Logic: what are the formal rules to draw valid conclusions?

  Boolean logic (Boole, 1847); First-order logic (Frege, 1879)
- Computation: what can be computed?
  - Algorithms for logical deduction
  - Godel (1931): incompleteness theorem (decidability)
  - Church–Turing thesis (1936) (computability)
  - NP-completeness theory (Cook, 1971) (tractability)
- Probability: how do we reason with uncertain information?

  from Cardano (1564) to Bayes' theorem (1763)

Economics

- Smith (1776): science of agents maximizing their economic well-being
- Economics study how people make choices that lead to preferred outcomes. Utility: property of being useful/beneficial to a person
- Decision theory: decisions made under uncertainty (combines probability theory with utility theory)
- In small economies: Game Theory
- When payoff depends on sequence of actions: Operations Research
- AI (often) looks for sub-optimal (not optimal) solutions

Neuroscience

- How do brains process information?
- Study of the nervous system, the brain in particular
- Golgi (1873): First observations of neurons in the brain
- How does the brain give rise to cognitive processes and consciousness?
- “It is competence, not consciousness, that matters!” (Russell)

Psychology

- How do humans think and act?
- Wundt (1879): first laboratory of experimental psychology (human vision)
- Cognitive Psychology: brain as information-processing device
- Craink (1943): “The Nature of Explanation” laid the foundations of knowledge-base agents
- Cognitive Science: born at a workshop in 1956 at MIT where papers by Miller, Chomsky, Newell & Simon showed how computer models could be used to address the psychology of memory, language, and logical thinking

Control Theory

- How can artifacts operate under their own control?
- Ktesibios of Alexandria (c. 250 B.C.) built the first self-controlling machine: a water clock with a regulator maintaining a constant flow rate
- Wiener (1948): "Cybernetics"

  Purposive behavior arises from a regulatory mechanism trying to minimize error—the difference between current state and goal state
- Modern Control Theory: design systems that optimize an objective functions over time
- AI vs. Control Theory: different mathematical tools

  Calculus and matrix algebra vs. logical inference and computation

Linguistics

- How does language relate to thought?
- Computational Linguistics
- Natural Language Processing

  Understanding language requires an understanding of:

  - structure of sentences
  - subject matter
  - context
- Early work in knowledge representation was tied to language

The History of AI

Origins: The dream of an "artificial intelligence" (broadly interpreted) is age-old (philosophy mainly)

1956: Inception of AI at Dartmouth Workshop. John McCarthy proposes the name "Artificial Intelligence". Early enthusiasm, famous quote:

"It is not my aim to surprise or shock you – but the simplest way I can summarize is to say that 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 the visible future – the range of problems they can handle will be coextensive with the range to which the human mind has been applied."

60's: Early successes. "Intelligent Behavior" is shown in many demonstration systems for microworlds (Blocksworld, Checkers)

The History of AI, ctd.

70's: How to scale from microworlds to real applications?
→ Knowledge-based systems

80's: Commercial success of rule-based expert systems

End of 80's: Expert systems prove less promising than imagined (difficult to update/maintain, cannot learn, brittle) → "AI Winter"
Renaissance of neural networks

90's: Formalization of AI techniques and increased use of mathematics.
Support-vector machines & recurrent neural networks become popular
Quote from 1st edition of Russel & Norvig's text book [1995]:

"Gentle revolutions have occurred in robotics, computer vision, machine learning, and knowledge representation. A better understanding of the problems and their complexity properties, combined with increased mathematical sophistication, has led to workable research agendas and robust methods."

The History of AI, ctd.

'95 - present: Intelligent agents emerge
→ 1997: Deep Blue beats Garry Kasparov in a chess match

2000 - present: Very large data-sets become available
→ 2005: First DARPA Challenge on autonomous cars
→ 2011: IBM Watson defeats Jeopardy champions

2010s Deep learning becomes feasible and triggers significant advancements in vision and text processing
→ 2016: Deepmind's AlphaGo beats Go world champion
→ 2017: Autonomous taxis (MIT's spin off NuTonomy)
→ 2018-19: Super human AI for multi player poker

2020s The generative AI era
→ 2022: OpenAI creates Generative Pre-trained Transformer 3.5
→ Nov 2022: OpenAI launches ChatGPT based on GPT-3.5
→ Mar 2023: OpenAI releases GPT-4 (and ChatGPT Plus)

AI Today: Sub-Areas

Modern AI is a conglomerate of highly technical sub-areas:

Search: How to effectively find solutions in problems with large search spaces (**NP**-hard and far beyond)

→ **This course!**

CSP & SAT: General formulation and solution of search problems that involve satisfying a set of constraints

→ **This course!**

KR: Knowledge representation and reasoning (logic and deduction)

→ **This course!**

Planning: General formulation and solution of search problems that involve finding goal-leading action strategies

AI Today: Sub-Areas, ctd.

Modern AI is a conglomerate of highly technical sub-areas:

Uncertainty: Reasoning about uncertain knowledge.

Machine Learning: How to learn from experience?

Multi-Agent Systems: How to control/analyze systems of agents perceiving/acting individually?

Robotics: How to control/design robots?

Vision and Image Processing: How to interpret/analyze camera input?

Natural Language Processing: How to understand and produce language?

Tasks: Speech recognition, speech tagging , word disambiguation, etc.

Use cases: machine translation, chatbots, spam detection, text summarization, etc.

→ Intimate relations to many other areas of CS. Logic Programming, Databases, Verification, Game Theory, . . .

AI Today: Underlying Technologies

The core technology behind most of the most visible advances is machine learning, especially

- deep learning, including generative adversarial networks (GANs)
- reinforcement learning

They are powered by large-scale data and computing resources

AI Today: Language Processing

Major leap in the last five years thanks to neural network language models: GPT, ELMo, mT5, BERT

- These models learn about how words are used in context from sifting through the patterns in naturally occurring text
- They use billions of tunable parameters and are engineered to be able to process unprecedented quantities of data
  → over one trillion words for GPT-3!
- Applications: machine translation, text classification, speech recognition, writing aids, chatbots
- Challenges: smaller data sets, biases, deep text understanding

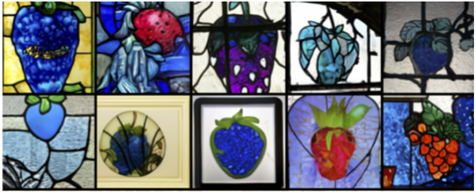
Tremendous growth in conversational interfaces: Google Assistant, Siri, Alexa

AI Today: Vision & Image Processing

Image-processing technology is now widespread and training time has been substantially reduced

- Many programs use ImageNet (massive standardized collection of over 14 million photographs) to train and test visual identification programs
- Real-time object-detection systems, e.g. YOLO (you look only once), are used in video surveillance and mobile robotics
- Face-recognition (e.g. to unlock your phone)
  → issues around bias and privacy
- Generation of photorealistic images and even videos using GANs
  → deepfakes: could used in illicit activity, e.g. revenge porn, identity theft

AI Today: Integration of Images and Text



- Images produced by OpenAI's DALL-E given prompt:
  a stained glass window with an image of a blue strawberry
- Similar query to a web-based image search produces blue strawberries, blue stained-glass windows, or stained-glass windows with red strawberries

  → the system is not merely retrieving relevant images but producing novel combinations of visual features
- Integration of GAN technology for generating images and the transformer technology for producing text

AI Today: Games

Developing algorithms for games has long been a fertile training ground and a showcase for the advancement of AI techniques.

- 1997: Deep Blue beats Garry Kasparov in a chess match
  → Kasparov said that he felt a *"new kind of intelligence"* across the board from him
- 2016: Deepmind'S AlphaGo beats Go world champion Lee Sedol
- AlphaGoZero: no use of direct human guidance (large collection of data from past Go matches)
- AlphaZero: a single network architecture that could learn to play expert-level Go, Shogi, or Chess
- AI agents has outperformed humans in many games, e.g. poker, StarCraft II, Quake III, Alpha Dogfight

AI Today: Robotics

Remarkable progress in intelligent robotics driven by:

- machine learning
- powerful computing and communication capabilities
- increased availability of sophisticated sensor systems

Progress due to a combination of learning techniques, classical control theory and painstaking engineering and design

- Combined with deep-learning vision systems, manipulator-type robots can effectively pick up randomly placed overlapping objects
- Legged robots much more agile, e.g. Boston Dynamics Atlas, Spot

BUT majority of types of robotics systems remain lab-bound!

AI Today: And Several Other Advances

- Mobility
- Health
- Finance
- Recommender Systems
- ...

See: https://ai100.stanford.edu/

AI Today: Open Challanges

- Generalizability
  → Capacity for generalizing or transferring learning from a training task to a novel one
- Causality
  → Today's ML models have only limited capacity to discover causal knowledge of the world
- Normativity
  → AI needs to have good normative models and to be capable of integrating its behavior into human normative institutions and processes
  → AI alignment, unbiased and fair algorithms, accountability

Philosophical Questions

- Can a machine *think*?
- Can a machine have an *intelligent behaviour*?
- Can a machine have *consciousness*?
- Can a machine have *self-consciousness*?
- Can a machine have *emotions*, ...?

Societal Issues

- Advantages and disadvantages of *automation*
- Impact on *human intelligence*
- *Ethics*
- *Privacy*
- *Legal Issues*

Ethical Principles by the EU Expert Group

- Human agency and oversight
- Technical robustness and safety
- Privacy and data governance
- Transparency
- Diversity, non-discrimination and fairness
- Societal and environmental well-being
- Accountability

See: https://ec.europa.eu/digital-single-market/en/artificial-intelligence

Isaac Asimov's Laws of Robotics

- A robot may not injure a human being or, through inaction, allow a human being to come to harm
- A robot must obey the orders given it by human beings, except when such orders would conflict with the previous law
- A robot must protect its own existence as long as such protection does not conflict with the previous two laws

Oren Etzioni's Laws of AI

- An AI system must be subject to the full gamut of laws that apply to its human operator
- An AI system must clearly disclose that it is not human
- An AI system cannot retain or disclose confidential information without explicit approval from the source of that information

State of the Art

- Play a decent game of table tennis → Yes
- Drive safely along a curving mountain road → Yes
- Drive safely along Telegraph Avenue → In progress
- Buy a week's worth of groceries on the web → Yes
- Buy a week's worth of groceries at Berkeley Bowl → No
- Play a decent game of bridge→ Yes
- Discover and prove a new mathematical theorem → In progress
- Design and execute a research program in molecular biology → In progress
- Write an intentionally funny story → Yes
- Give competent legal advice in a specialized area of law → Yes
- Translate spoken English into spoken Swedish in real time → Yes
- Converse successfully with another person for an hour → Yes
- Perform a complex surgical operation → In progress
- Unload any dishwasher and put everything away → Yes

Summary

- "Artificial intelligence" as an idea can be roughly classified along the dimensions thinking vs. acting and humanly vs. rationally.
- The research area of Artificial Intelligence (AI) today, as well as this course, are about "acting rationally".
- Early AI had ambitious dreams, and successes in simple problems, but then faced difficulties to scale up. Since the early 90s, AI has become more formal and systematic.
- Modern AI is a conglomerate of highly technical sub-areas, many of which have intimate relations to other areas of Computer Science.
- There are numerous AI applications in various forms of control, robotics, speech processing, vision, verification, security, . . .

  → I'll list some applications within each chapter.

Reading

- *Chapter 1: Introduction* [Russell and Norvig (2010)].

  Content: A much more detailed account of the issues I have overviewed here. (33 pages in the book . . .)

References I

Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach (Third Edition)*. Prentice-Hall, Englewood Cliffs, NJ, 2010.