Artificial Intelligence 1

Prof. Dr. Frank Hopfgartner
Dr. Matthias Horbach

Institute for Web Science and Technologies (WeST)
University of Koblenz

WeST
People and Knowledge Networks

Overview

What is “A

Definitions 1/2

Artificial intelligence (AI, also machine intelligence, MI) is intelligence demonstrated by machines, in contrast to the natural intelligence (NI) displayed by humans and other animals. In computer science AI research is defined as the study of "intelligent agents": any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals.

[. . .]

The scope of AI is disputed: as machines become increasingly capable, tasks considered as requiring "intelligence" are often removed from the definition, a phenomenon known as the AI effect, leading to the quip "AI is whatever hasn't been done yet." *[Wikipedia]*

Definitions 2/2

Artificial Intelligence, the science of making machine do things that would require intelligence if done by man.

[Minsky, 1968]

The study of how to make computers do things at which, at the moment, people are better.

[Rich and Knight, 1991]

Artificial Intelligence is not the study of computers, but of intelligence in thought and action. Computers are its tools, because its theories are expressed as computer programs [. . .]

[Boden, 1987]

Overview

Humans and rationality (Russel, Norvig) 1/5

Classification of AI approaches, what is being imitated?

	human	rational
Acting	human acting	rational acting
Thinking	human thinking	rational thinking

Humans and rationality (Russel, Norvig) 2/5

	human	rational
Acting	human acting	rational acting
Thinking	human thinking	rational thinking

- Goal: Development of systems that act like humans
- Develop methods, so that a computer can do things
  - that currently only humans can do
  - where humans can do it better
- Evaluation of methods: compare machine and humans (Turing test)

Humans and rationality (Russel, Norvig) 3/5

	human	rational
Acting	human acting	rational acting
Thinking	human thinking	rational thinking

- Goal: Machine thinks like a human
- Develop models of human cognition (e.g. neural networks)
- Models are tested by psychological experiments.
- More part of *cognitive sciences*, not really of concern in AI

Humans and rationality (Russel, Norvig) 4/5

	human	rational
Acting	human acting	rational acting
Thinking	human thinking	rational thinking

- Agent paradigm: autonomous, intelligent and acting entities
- Rational agent: optimise course of action to achieve goal
- Humans do not necessarily act rationally

Humans and rationality (Russel, Norvig) 5/5

	human	rational
Acting	human acting	rational acting
Thinking	human thinking	rational thinking

- Formalisation of thinking through logic and formal reasoning
- Mathematical formulations of "laws of thought"
- Identification of the foundational principles of thinking (not necessarily of its *implementation* in humans through neural nets)
- Has certain psychological and philosophical aspects

Strong and weak AI hypothesis

Weak AI hypothesis (weak AI):

> *Machines (computers, robots,. . .) can act, as if they were intelligent.*

Strong AI hypothis (strong AI):

> *Machines (computers, robots,. . .) can actually think and do no just simulate intelligence.*

- The weak AI hypothesis is usually accepted by AI researchers (even if intelligent acting is not yet achieved in many areas)
- The strong AI hypothesis implies a philosophical discussion (what is consciousness, what is intelligence?)

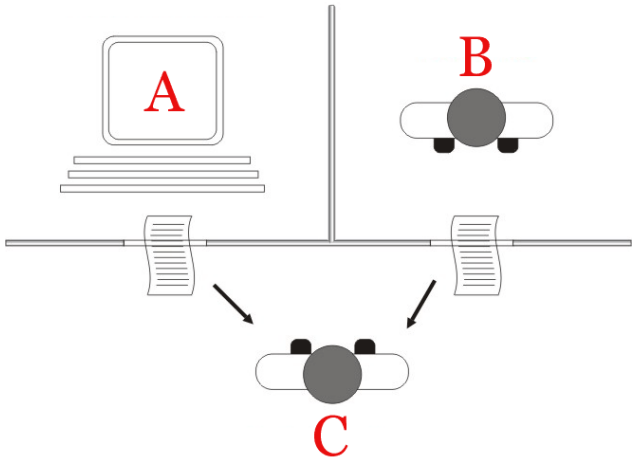
AI-hard and AI-easy

- Similar as problems can be categorised as NP-hard in computational complexity, problems can characterised by their difficulty in AI terms
- AI-easy (no problems of AI research anymore!)
  - Finding shortest paths with complete information
  - Sorting
  - Calculating
- AI-hard:
  - Natural language understanding
  - Image recognition
  - Poetry
- . . . and, of course, a lot of problems in between (these are the ones we will discuss in this course)

The Turing test

Alan Turing, 1912–1954

The Turing test is the *prototype* of an AI-hard problem



- Humans asks computer/other human any written question
- Test is passed, if human cannot tell whether his partner is human or computer

The Chinese room

- Criticism on the Turing test: is a system intelligent that has access to an all-encompassing data base with all ever needed answers?
- Example of the Chinese room (Searle)
  - A human who does not know Chinese is in a closed room
  - In the room there is
    - a stack with papers, each having some Chinese writing on it
    - an instruction book (in the humans native language) with rules how to transform some given paper with Chinese writing to another paper with Chinese writing
  - A paper with Chinese writing is entered through a mail slot (for example, a story with some questions about the story)
  - Human returns some paper from the stack as instructed according to the rules in the instruction book
- Does the human understand Chinese?
- Does the system (human+room) understand Chinese (and what does that mean)?

Winograd Schemes 1/2

- Alternative Turing test for natural language understanding
- Question for intelligent system: "Can a crocodile run a steeplechase?"
  - Question cannot be answered by information retrieval methods and availability of lots of data (hypothetical question, Google has also no answer)
  - Requires reasoning over physical capabilities (length of legs) and pyhysics (ability to jump)
- The question requires special background knowledge (about crocodiles and steeplechases), but it can be way simpler

Winograd Schemes 2/2

- Winograd schemes: simple "A or B" questions:

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

  A) The trophy

  B) The brown suitcase

- Compare with

  The trophy would not fit in the brown suitcase because it was too **big**.

  What was too big?

  A) The trophy

  B) The brown suitcase

- A system has to have deep language understanding in order to be able to distinguish these two variants.
- *Commonsense reasoning*: Subarea of AI aiming at modelling the human ability to draw meaningful conclusions in everyday situations.

Overview

The term “Artificial Intelligence”

The term AI is usually attributed to John McCarthy (Dartmouth, 1956):

> *I won’t swear that I hadn’t seen it before [. . .] but artificial intelligence wasn’t a prominent phrase particularly. Someone may have used in a paper or a conversation or something like that, but there were many other words that were current at the time. The Dartmouth conference made the phrase dominate the others.*

Some history 1/2

Antique	Hephaistos builds Talos (bronze robot)
16th cent.	Rabbi Loew creates the Golem
18th cent.	Mechanical turk
1837	C. Babbage constructs the "Analytical machine"
1854	George Boole publishes "The Laws of Thought"
1920	First appearance of the term "robot" (Capek, play "Rossum's Universal Robots")
1937	Turing and the halting problem (bounds of computability)
1940	Asimov and the laws of robotics
1950	Turing test
1950	Shannon: first chess algorithm
1955	Samuel writes first checkers program that can defeat himself

Some history 2/2

1957 Simon/Newell develop the General Problem Solver
1965 Zadeh develops fuzzy logic
1966 Weizenbaum develops Eliza (first chatbot)
1972 Prolog
1976 Shortliffe and Buchanan develop one of the first expert systems (MYCIN)
1992 Tesauro develops backgammon program, which plays better than every human
1993 Beginning of the RoboCups
1997 First victory of a chess computer against a world champion (Deep Blue vs. Kasparow)
2011 IBM Watson wins Jeopardy!
2016 First victory of a Go computer against a professional player
2017 First victory of a Poker bot against a professional player

AI Enthusiasm

Simon and Newell's (1958) predictions:

> 1. That within ten years, a digital computer will be the world's chess champion, unless the rules bar it from competition.
> 2. That within ten years a digital computer will discover and prove an important new mathematical theorem.
> 3. That within ten years a digital computer will write music that will be accepted by critics as possessing considerable aesthetic value.
> 4. That within ten years most theories in psychology will take the form of computer programs, or of qualitative statements about the characteristics of computer programs.
>
> . . .The simplest way I can summarize is to say that there are now in the world machines that can 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 the human mind has been applied.

History of AI = History of Computer Science

- In the beginning of Computer Science (early/mid 20th century), research on general Computer Science and "Artificial Intelligence" was not distinguishable
- AI is a research discipline at the frontier of science

Overview

AI - an overview 1/2

AI has many facets:

- Knowledge representation and reasoning
  - Formalisation of and reasoning with information
  - Approaches: logics, probabilistic systems,...
- Automatic planning
  - Problem solving, planning actions to achieve some goal
  - Approaches: search algorithms, optimisation,...
- Learning
  - Rule learning, pattern recognition in data
  - Approaches: statistics, neural networks,...
- Natural language understanding
  - Chatbots, query-answering systems
  - Approaches: information retrieval, text mining,...

AI - an overview 1/2

- Agents
- Digital image processing
- Language processing
- Robotics
- Computational creativity
- Games
- ...

In this course:

- Automatic planning
- Knowledge representation and reasoning
- Agents

Chapter 1: Introduction

Summary

Chapter 1: Summary

- What is "Artificial Intelligence?"
- History of AI = History of Computer Science
- Humans and rationality
- Strong and weak AI hypothesis
- Turing test, the Chinese room, Winograd schemes
- AI has many facets