

Foundations of Artificial Intelligence

A1. Organizational Matters

Malte Helmert

University of Basel

February 16, 2026

Foundations of Artificial Intelligence

February 16, 2026 — A1. Organizational Matters

A1.1 People

A1.2 Format

A1.3 Assessment

A1.4 About this Course

Introduction: Overview

Chapter overview: introduction

- ▶ **A1. Organizational Matters**
- ▶ A2. What is Artificial Intelligence?
- ▶ A3. AI Past and Present
- ▶ A4. Rational Agents
- ▶ A5. Environments and Problem Solving Methods

A1.1 People

Teaching Staff: Lecturer

Lecturer

Malte Helmert

- ▶ **email:** `malte.helmert@unibas.ch`
- ▶ **office:** room 06.004, Spiegelgasse 1



Teaching Staff: Assistant

Assistant

Clemens Büchner

- ▶ **email:** `clemens.buechner@unibas.ch`
- ▶ **office:** room 04.005, Spiegelgasse 1



Teaching Staff: Tutors

Tutors

Claudia Grundke

- ▶ **email:** `claudia.grundke@unibas.ch`
- ▶ **office:** room 04.001, Spiegelgasse 5



Aeneas Meier

- ▶ **email:** `aeneas.meier@unibas.ch`



Carina Schrenk

- ▶ **email:** `carina.schrenk@stud.unibas.ch`



Students

target audience:

- ▶ Bachelor Computer Science, ~3rd year
- ▶ Bachelor Computational Sciences, ~3rd year
- ▶ Master Data Science
- ▶ other students welcome

prerequisites:

- ▶ algorithms and data structures
- ▶ basic mathematical concepts
(formal proofs; sets, functions, relations, graphs)
- ▶ complexity theory
- ▶ programming skills (mainly for exercises)

A1.2 Format

Structure Overview

Foundations of AI *week structure*:

- ▶ *Monday*: release of exercise sheet
- ▶ *Monday* and *Wednesday*: lectures
- ▶ *Wednesday/Friday*: exercise session
- ▶ *Sunday*: exercise sheet due
- ▶ *exceptions* due to holidays

Time & Place

Lectures

- ▶ Mon 16:15–18:00 in Biozentrum, lecture hall U1.141
- ▶ Wed 14:15–16:00 in Biozentrum, lecture hall U1.141

Exercise Sessions

- ▶ Wed 16:15–18:00 in Biozentrum, SR U1.193
- ▶ Wed 16:15–18:00 in Spiegelgasse 1, room U1.001
- ▶ Fri 14:15–16:00 in Pharmazentrum, room 1067
(with some exceptions, see course directory/VV online)

first exercise session: February 18/20 (this week)

Exercises

exercise sheets (homework assignments):

- ▶ mostly theoretical exercises
- ▶ sometimes programming exercises

exercise sessions:

- ▶ initial part:
 - ▶ discuss **common mistakes** in previous exercise sheet
 - ▶ answer **questions** on previous exercise sheet
- ▶ main part:
 - ▶ we **support** you solving the current exercise sheet
 - ▶ we **answer** your questions
 - ▶ we **assist** you comprehend the course content

Theoretical Exercises

theoretical exercises:

- ▶ exercises on ADAM every Monday
- ▶ covers material of **that week** (Monday and Wednesday)
- ▶ due Sunday of **the same week** (23:59) via ADAM
- ▶ solved in **groups of at most two** ($2 = 2$)
- ▶ **support** in exercise session of current week
- ▶ discussed in exercise session of following week

Programming Exercises

programming exercises (project):

- ▶ project with ~ 4 parts over the duration of the semester
- ▶ additional one-off programming exercises (occasionally)
- ▶ integrated into the exercise sheets (no special treatment) (solved in the same groups)
- ▶ implemented in Java; need working Linux system for some
- ▶ solutions that obviously do not work: 0 marks

A1.3 Assessment

Course Material

course material that is relevant for the exam:

- ▶ slides
- ▶ content of lecture
- ▶ exercise sheets

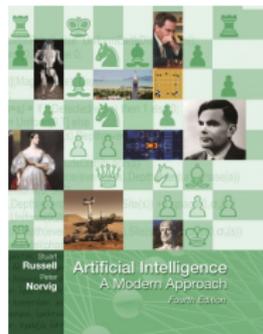
additional (optional) course material:

- ▶ textbook
- ▶ bonus material

Textbook

Artificial Intelligence: A Modern Approach
by Stuart Russell and Peter Norvig
(4th edition, Global edition)

- ▶ covers large parts of the course
(and much more), but not everything



Exam

- ▶ **written exam** on Wednesday, June 24
 - ▶ 14:00-16:00
 - ▶ 105 minutes for working on the exam
 - ▶ location: Organische Chemie, Grosser Hörsaal (Organic Chemistry, large lecture hall)
- ▶ 8 ECTS credits
- ▶ admission to exam: 50% of the exercise marks
- ▶ class participation **not required** but **highly recommended**
- ▶ **no repeat exam**

Course Homepage and Enrolment

Course Homepage

`https://dmi.unibas.ch/en/studies/computer-science/
course-offer-spring-2026/
13548-lecture-foundations-of-artificial-intelligence/`

- ▶ course information
- ▶ slides
- ▶ bonus material (not relevant for the exam)
- ▶ link to ADAM workspace

enrolment:

- ▶ `https://services.unibas.ch/`

Communication Channels

Communication Channels

- ▶ lectures and exercise sessions
- ▶ ADAM workspace (linked from course homepage)
 - ▶ link to Discord server
 - ▶ exercise sheets and submission
 - ▶ exercise FAQ
 - ▶ bonus material that we cannot share publicly
- ▶ Discord server (linked from ADAM workspace)
 - ▶ opportunity for Q&A and informal interactions
- ▶ contact us by email
- ▶ meet us in person (by arrangement)
- ▶ meet us on Zoom (by arrangement)

Plagiarism

Plagiarism

Plagiarism is presenting someone else's work, ideas, or words as your own, without proper attribution.

For example:

- ▶ Using someone's text without citation
- ▶ Paraphrasing too closely
- ▶ Using information from a source without attribution
- ▶ Passing off AI-generated content as your own original work

Long-term impact:

- ▶ You undermine your own learning.
- ▶ You start to lose confidence in your ability to think, write, and solve problems independently.
- ▶ Damage to academic reputation and professional consequences in future careers

Plagiarism in Exercises

- ▶ You may discuss material from the course, including the exercise assignments, with your peers.
- ▶ **But:** You have to independently write down your exercise solutions (in your team).
- ▶ Help from an LLM is acceptable to the same extent as it is acceptable from someone who is not a member of your team.
- ▶ The exercise submission includes a declaration on plagiarism and use of AI tools.

Immediate consequences of plagiarism:

- ▶ 0 marks for the exercise sheet (first time)
- ▶ exclusion from exam (second time)

If in doubt: check with us what is (and isn't) OK **before submitting**
Exercises too difficult? We are happy to help!

Special Needs?

- ▶ We (and the university) strive for equality of students with disabilities or chronic illnesses.
- ▶ Contact the lecturers for small adaptations.
- ▶ Contact the Students Without Barriers (StoB) service point for general adaptations and disadvantage compensation.

A1.4 About this Course

Classical AI Curriculum

“Classical” AI Curriculum

1. introduction
2. rational agents
3. uninformed search
4. informed search
5. constraint satisfaction
6. board games
7. propositional logic
8. predicate logic
9. modeling with logic
10. classical planning
11. probabilistic reasoning
12. decisions under uncertainty
13. acting under uncertainty
14. machine learning
15. deep learning
16. reinforcement learning

↔ wide coverage, but somewhat superficial

Our AI Curriculum

Our AI Curriculum

1. introduction
2. rational agents
3. uninformed search
4. informed search
5. constraint satisfaction
6. board games
7. propositional logic
8. predicate logic
9. modeling with logic
10. classical planning
11. probabilistic reasoning
12. decisions under uncertainty
13. acting under uncertainty
14. machine learning
15. deep learning
16. reinforcement learning

Topic Selection

guidelines for topic selection:

- ▶ fewer topics, more depth
- ▶ more emphasis on programming projects
- ▶ connections between topics
- ▶ avoiding overlap with other courses
 - ▶ Pattern Recognition (B.Sc.)
 - ▶ Machine Learning (M.Sc.)
- ▶ focus on **algorithmic core** of model-based AI

Under Construction...



- ▶ A course is never “done”.
- ▶ We are always happy about feedback, corrections and suggestions!

Foundations of Artificial Intelligence

A2. Introduction: What is Artificial Intelligence?

Malte Helmert

University of Basel

February 16, 2026

Foundations of Artificial Intelligence

February 16, 2026 — A2. Introduction: What is Artificial Intelligence?

A2.1 What is AI?

A2.2 Thinking Like Humans

A2.3 Acting Like Humans

A2.4 Thinking Rationally

A2.5 Acting Rationally

A2.6 Summary

Introduction: Overview

Chapter overview: introduction

- ▶ A1. Organizational Matters
- ▶ A2. What is Artificial Intelligence?
- ▶ A3. AI Past and Present
- ▶ A4. Rational Agents
- ▶ A5. Environments and Problem Solving Methods

A2.1 What is AI?

What is AI?

What do we mean by **artificial intelligence**?

↪ no generally accepted definition!

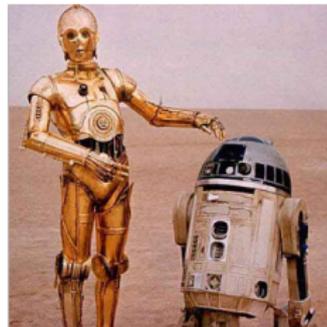
often pragmatic definitions:

- ▶ “AI is what AI researchers do.”
- ▶ “AI is the solution of hard problems.”

in this chapter: some common attempts at defining AI

What Do We Mean by Artificial Intelligence?

what **pop culture** tells us:



What is AI: Humanly vs. Rationally; Thinking vs. Acting

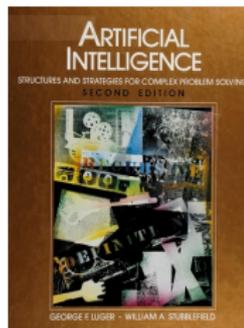
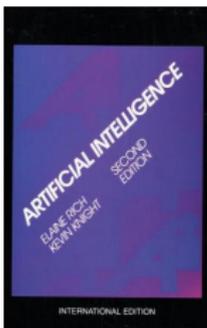
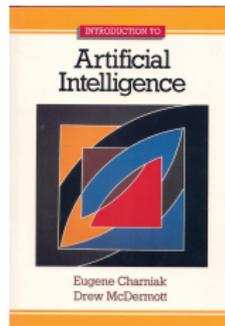
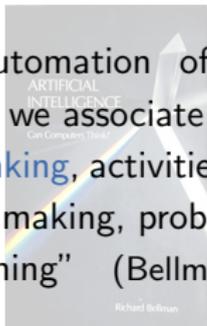
what **scientists** tell us:



What is AI: Humanly vs. Rationally; Thinking vs. Acting

what **scientists** tell us:

“[the automation of] activities that we associate with **human thinking**, activities such as decision-making, problem solving, learning” (Bellman, 1978)



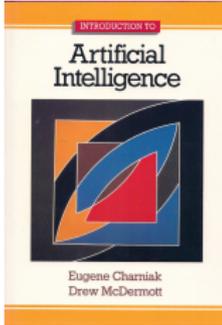
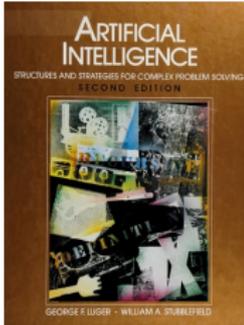
What is AI: Humanly vs. Rationally; Thinking vs. Acting

what **scientists** tell us:



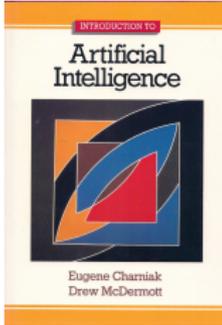
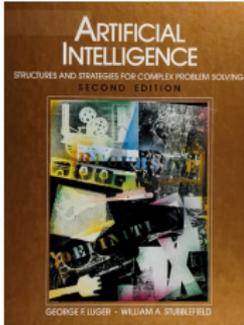
What is AI: Humanly vs. Rationally; Thinking vs. Acting

what **scientists** tell us:

 <p>thinking like humans</p>	
<p>“the study of how to make computers do things at which, at the moment, people are better”</p> <p>(Rich & Knight, 1991)</p>	

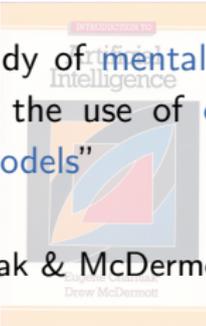
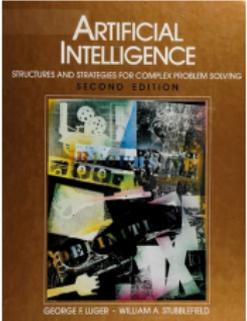
What is AI: Humanly vs. Rationally; Thinking vs. Acting

what **scientists** tell us:

 <p>thinking like humans</p>	
 <p>acting like humans</p>	

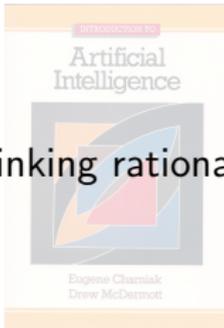
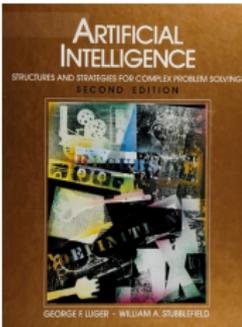
What is AI: Humanly vs. Rationally; Thinking vs. Acting

what **scientists** tell us:

 <p>thinking like humans</p>	 <p>“the study of mental faculties through the use of computational models” (Charniak & McDermott, 1985)</p>
 <p>acting like humans</p>	

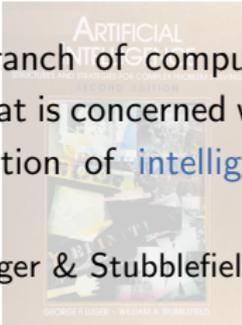
What is AI: Humanly vs. Rationally; Thinking vs. Acting

what **scientists** tell us:

 <p>thinking like humans</p>	 <p>thinking rationally</p>
 <p>acting like humans</p>	

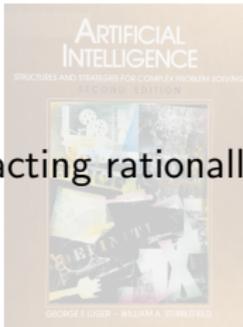
What is AI: Humanly vs. Rationally; Thinking vs. Acting

what **scientists** tell us:

 <p>thinking like humans</p>	 <p>thinking rationally</p>
 <p>acting like humans</p>	 <p>“the branch of computer science that is concerned with the automation of intelligent behavior” (Luger & Stubblefield, 1993)</p>

What is AI: Humanly vs. Rationally; Thinking vs. Acting

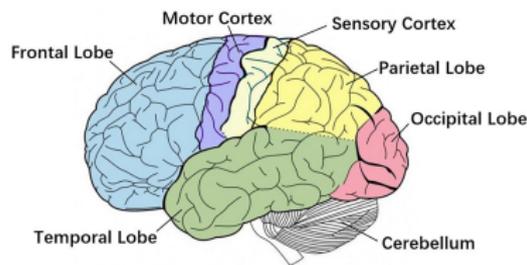
what **scientists** tell us:

 <p>thinking like humans</p>	 <p>thinking rationally</p>
 <p>acting like humans</p>	 <p>acting rationally</p>

A2.2 Thinking Like Humans

Cognitive (Neuro-) Science

- ▶ requires knowledge of **how humans think**
- ▶ two ways to a scientific **theory of brain activity**:
 - ▶ **psychological**: observation of human behavior
 - ▶ **neurological**: observation of brain activity
- ▶ roughly corresponds to **cognitive science** and **cognitive neuroscience**
- ▶ today separate research areas from AI



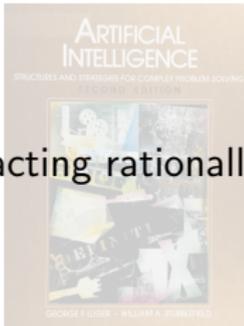
Machines that Think Like Humans



“brains are to intelligence as wings are to flight”



What Do We Mean by Artificial Intelligence?

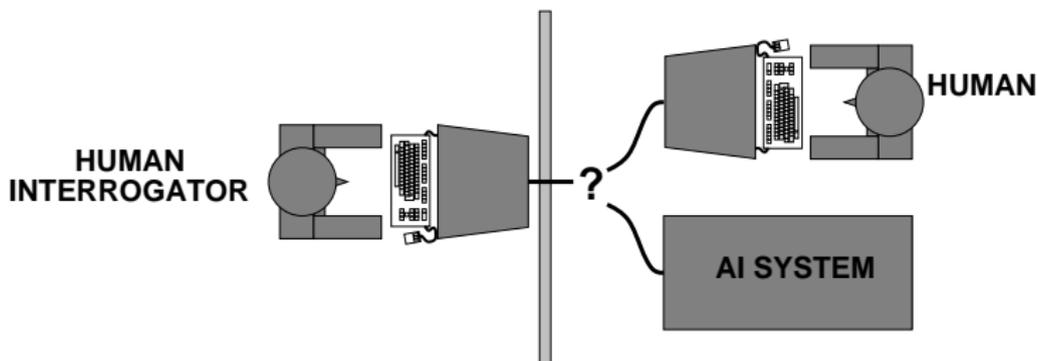
 <p>thinking like humans</p>	 <p>thinking rationally</p>
 <p>acting like humans</p>	 <p>acting rationally</p>

A2.3 Acting Like Humans

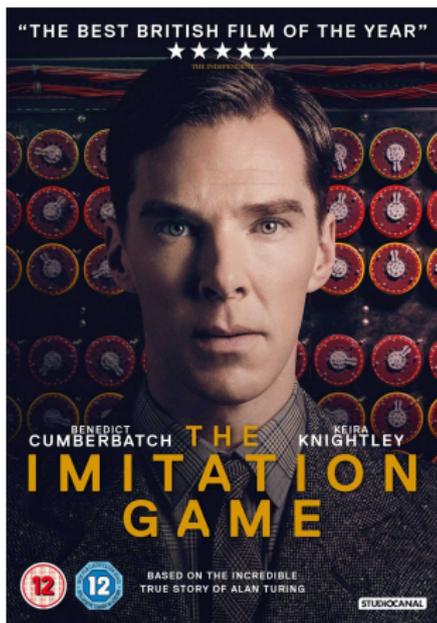
The Turing Test

Alan Turing, Computing Machinery and Intelligence (1950):

- ▶ central question: Can machines think?
- ▶ hypothesis: yes, if they can act like humans
- ▶ operationalization: the imitation game



Turing Test in Cinema



Turing Test: Brief History

- ▶ Eliza
- ▶ Loebner Prize
- ▶ Eugene Goostman
- ▶ Kuki (formerly Mitsuku)
- ▶ Google Duplex
- ▶ LaMDA & ChatGPT

```

Welcome to
EEEEEE LL      IIII ZZZZZZ  AAAA
EE  LL      II      ZZ  AA  AA
EEEE  LL      II      ZZZ  AAAAAA
EE  LL      II      ZZ   AA  AA
EEEEEE LLLLLL IIII ZZZZZZ  AA  AA

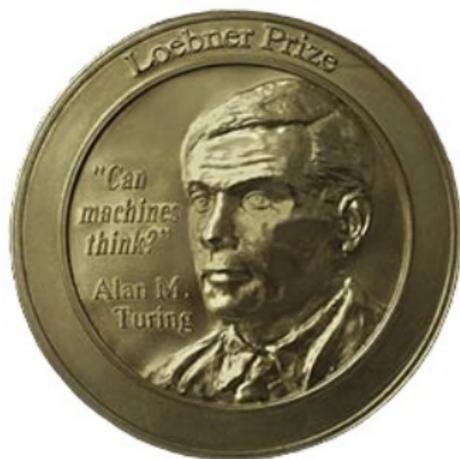
Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?
YOU:  Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:  They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:  Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:  He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:  It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:  █
  
```

- ▶ developed in 1966 by J. Weizenbaum
- ▶ uses combination of **pattern matching** and **scripted rules**
- ▶ most famous script mimics a **psychologist** ↔ many questions
- ▶ fooled early users

Turing Test: Brief History

- ▶ Eliza
- ▶ Loebner Prize
- ▶ Eugene Goostman
- ▶ Kuki (formerly Mitsuku)
- ▶ Google Duplex
- ▶ LaMDA & ChatGPT



- ▶ annual competition between 1991–2019
- ▶ most human-like AI is awarded
- ▶ highly controversial

Turing Test: Brief History

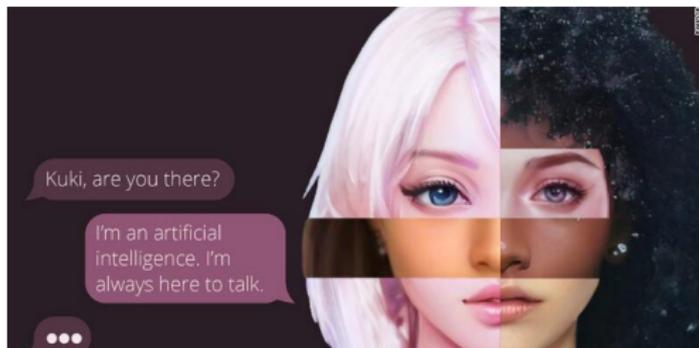
- ▶ Eliza
- ▶ Loebner Prize
- ▶ Eugene Goostman
- ▶ Kuki (formerly Mitsuku)
- ▶ Google Duplex
- ▶ LaMDA & ChatGPT



- ▶ mimics a 13-year-old boy from Odessa, Ukraine with a guinea pig
- ▶ “not too old to know everything and not too young to know nothing”
- ▶ 33% of judges were convinced it was human in 2014
↪ first system that passed the Turing test (?)

Turing Test: Brief History

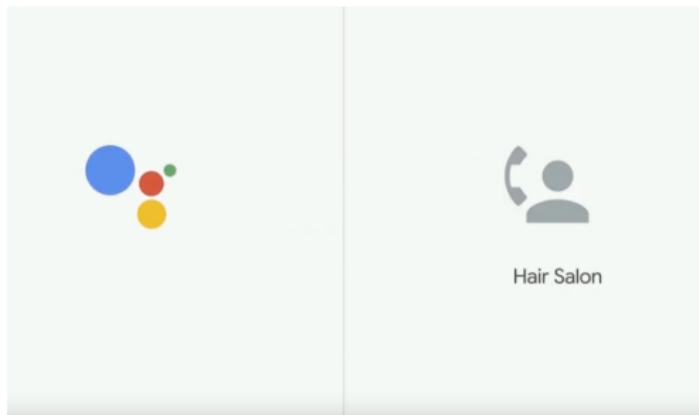
- ▶ Eliza
- ▶ Loebner Prize
- ▶ Eugene Goostman
- ▶ **Kuki** (formerly Mitsuku)
- ▶ Google Duplex
- ▶ LaMDA & ChatGPT



- ▶ **five times winner** of Loebner prize competitions (2015-2019)
- ▶ winner of "bot battle" versus Facebook's **Blenderbot**
↪ <https://youtu.be/RBK5j0yXDT8>

Turing Test: Brief History

- ▶ Eliza
- ▶ Loebner Prize
- ▶ Eugene Goostman
- ▶ Kuki (formerly Mitsuku)
- ▶ Google Duplex
- ▶ LaMDA & ChatGPT



- ▶ commercial product announced in 2018
- ▶ performs phone calls (making appointments) **fully autonomously**
- ▶ after criticism, it now starts conversation by **identifying as a robot**

Turing Test: Brief History

- ▶ Eliza
- ▶ Loebner Prize
- ▶ Eugene Goostman
- ▶ Kuki (formerly Mitsuku)
- ▶ Google Duplex
- ▶ LaMDA & ChatGPT

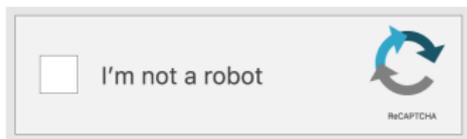


- ▶ systems like LaMDA and ChatGPT would likely pass the Turing test
- ▶ example conversation: <https://www.nytimes.com/2023/02/16/technology/bing-chatbot-transcript.html>
- ▶ ChatGPT even **passed some exams** (but failed on others)

Value of the Turing Test

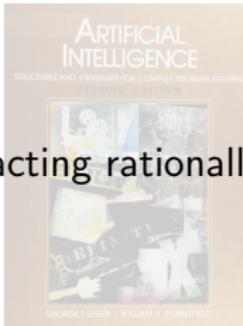
- ▶ human actions **not always intelligent**
- ▶ **scientific value** of Turing test questionable:
 - ▶ Test for AI or for interrogator?
 - ▶ results not reproducible
 - ▶ strategies to succeed \neq intelligence:
 - ▶ **deceive** interrogator
 - ▶ **mimic** human behavior

⇒ not important in AI “mainstream”



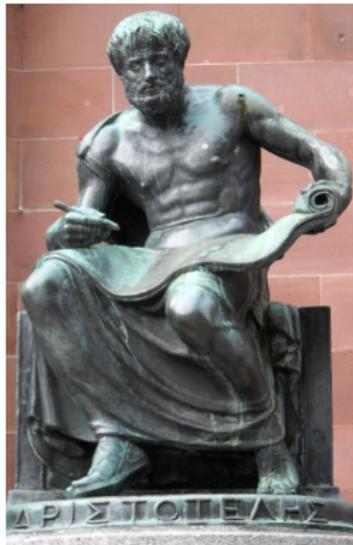
practical application: CAPTCHA
 (“**C**ompletely **A**utomated **P**ublic Turing
 test to tell **C**omputers and **H**umans **A**part”)

What Do We Mean by Artificial Intelligence?

 <p>thinking like humans</p>	 <p>thinking rationally</p>
 <p>acting like humans</p>	 <p>acting rationally</p>

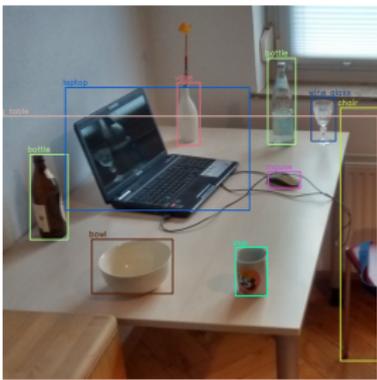
A2.4 Thinking Rationally

Thinking Rationally: Laws of Thought

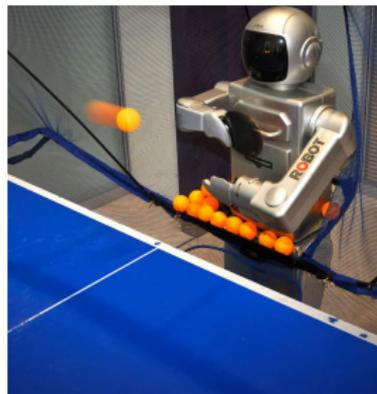
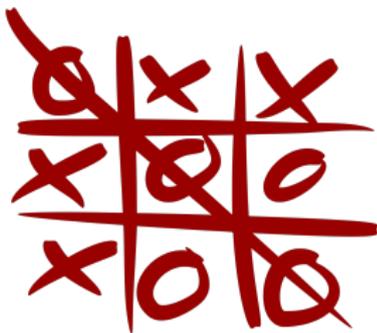


- ▶ **Aristotle:** What are correct arguments and modes of thought?
- ▶ **syllogisms:** structures for arguments that always yield correct conclusions given correct premises:
 - ▶ Socrates is a human.
 - ▶ All humans are mortal.
 - ▶ Therefore Socrates is mortal.
- ▶ direct connection to modern AI via mathematical **logic**

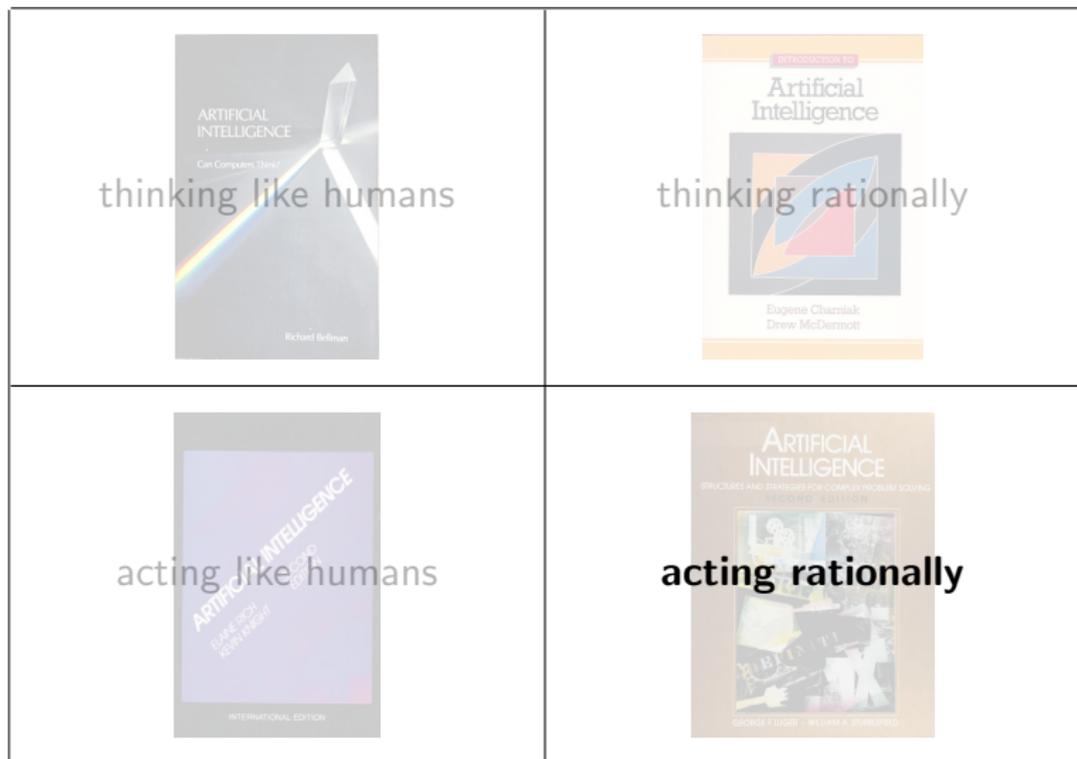
Problems of the Logical Approach



not all intelligent behavior stems from **logical thinking** and **formal reasoning**



What Do We Mean by Artificial Intelligence?



A2.5 Acting Rationally

Acting Rationally

acting rationally: “doing the right thing”

- ▶ the right thing: maximize utility given available information
- ▶ does not necessarily require “thought” (e.g., reflexes)

advantages of AI as development of rational agents:

- ▶ more general than thinking rationally (logical inference only one way to obtain rational behavior)
- ▶ better suited for scientific method than approaches based on human thinking and acting

↪ most common view of AI scientists today

↪ what we use in this course

A2.6 Summary

Summary

What is AI? \rightsquigarrow many possible definitions

- ▶ guided by **humans** vs. by utility (**rationality**)
- ▶ based on externally observable **actions** or inner **thoughts**?

\rightsquigarrow four combinations:

- ▶ acting like humans: e.g., Turing test
- ▶ thinking like humans: cf. cognitive (neuro-)science
- ▶ thinking rationally: logic
- ▶ **acting rationally**: most common view today
 - \rightsquigarrow amenable to scientific method
 - \rightsquigarrow used in this course

Foundations of Artificial Intelligence

A3. Introduction: AI Past and Present

Malte Helmert

University of Basel

February 18, 2026

Foundations of Artificial Intelligence

February 18, 2026 — A3. Introduction: AI Past and Present

A3.1 A Short History of AI

A3.2 Where are We Today?

A3.3 Summary

Introduction: Overview

Chapter overview: introduction

- ▶ A1. Organizational Matters
- ▶ A2. What is Artificial Intelligence?
- ▶ A3. AI Past and Present
- ▶ A4. Rational Agents
- ▶ A5. Environments and Problem Solving Methods

A3.1 A Short History of AI

Precursors (Until ca. 1943)

1950

1960

1970

1980

1990

2000

...

Philosophy and mathematics ask similar questions that influence AI.

- ▶ Aristotle (384–322 BC)
- ▶ Leibniz (1646–1716)
- ▶ Hilbert program (1920s)

Gestation (1943–1956)



1950

1960

1970

1980

1990

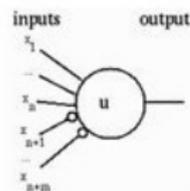
2000

...

Invention of electrical computers raised question:
Can computers mimic the human mind?

Gestation (1943–1956)

Artificial Neurons



1950

1960

1970

1980

1990

2000

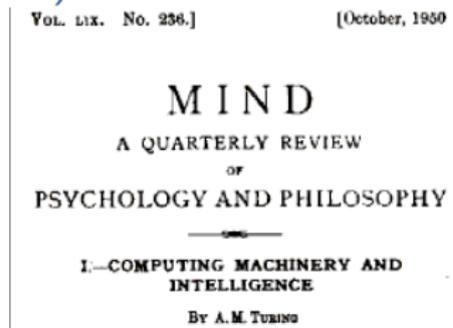
...

W. McCulloch & W. Pitts (1943)

- ▶ first computational model of **artificial neuron**
- ▶ **network of neurons** can compute any computable function
- ▶ basis of **deep learning**

Gestation (1943–1956)

Artificial
Neurons



1950

1960

1970

1980

1990

2000

...

Turing Test

Computing Machinery and Intelligence (A. Turing, 1950)

- ▶ famous for introducing **Turing test**
- ▶ (still) relevant discussion of **AI potential** and **requirements**
- ▶ suggests core AI aspects: **knowledge representation, reasoning, language understanding, learning**

Gestation (1943–1956)

Artificial Neurons



Dartmouth

1950

1960

1970

1980

1990

2000

...

Turing Test

Dartmouth workshop (1956)

- ▶ ambitious proposal: “An attempt will be made to find how to make machines use language, [...] solve kinds of problems now reserved for humans, and improve themselves.”
- ▶ J. McCarthy coins term **artificial intelligence**

Early Enthusiasm (1952–1969)

Artificial Neurons



Turing Test

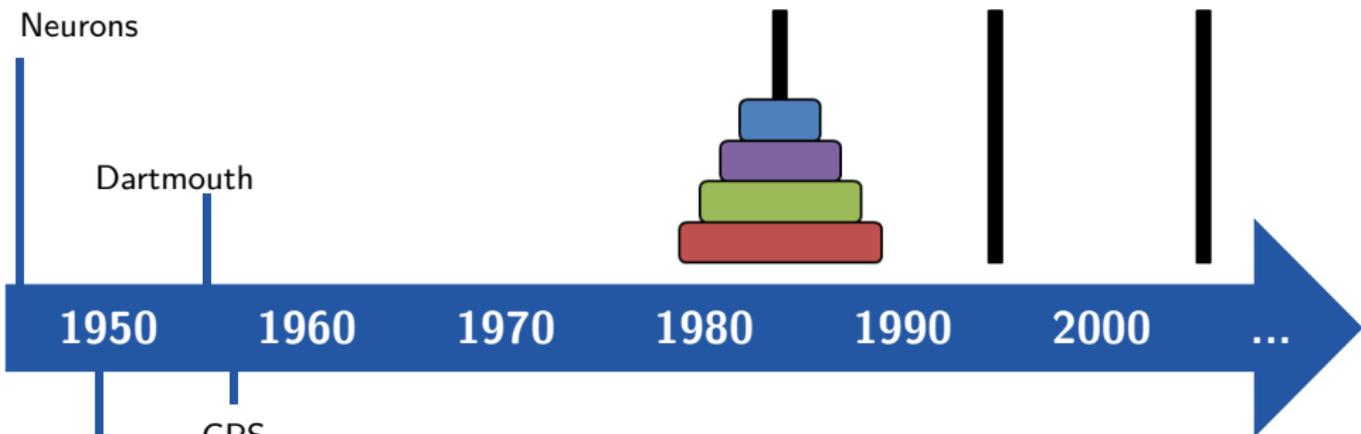
Dartmouth

early enthusiasm (H. Simon, 1957):

“[...] 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.”

Early Enthusiasm (1952–1969)

Artificial
Neurons



Turing Test

GPS

Dartmouth

General Problem Solver (H. Simon & A. Newell, 1957)

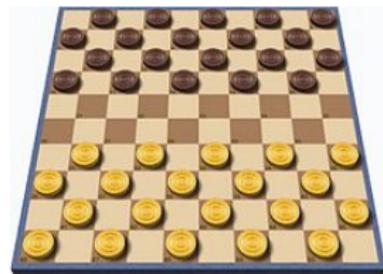
- ▶ universal problem solving machine
- ▶ imitates human problem solving strategies
- ▶ in principle able to solve every formalized symbolic problem
- ▶ in practice, GPS solves simple tasks like towers of Hanoi

Early Enthusiasm (1952–1969)

Artificial
Neurons

RL for
Checkers

Dartmouth



1950

1960

1970

1980

1990

2000

...

GPS

Turing Test

Checkers AI (A. Samuel, 1959)

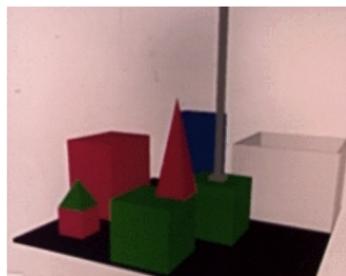
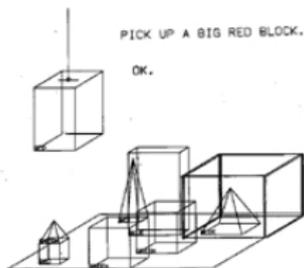
- ▶ popularized term [machine learning](#)
- ▶ learned to play at strong amateur level
- ▶ uses ideas of [reinforcement learning](#)

Early Enthusiasm (1952–1969)

Artificial
Neurons

RL for
Checkers

Dartmouth



1950

1960

1970

1980

1990

2000

...

GPS

Microworlds

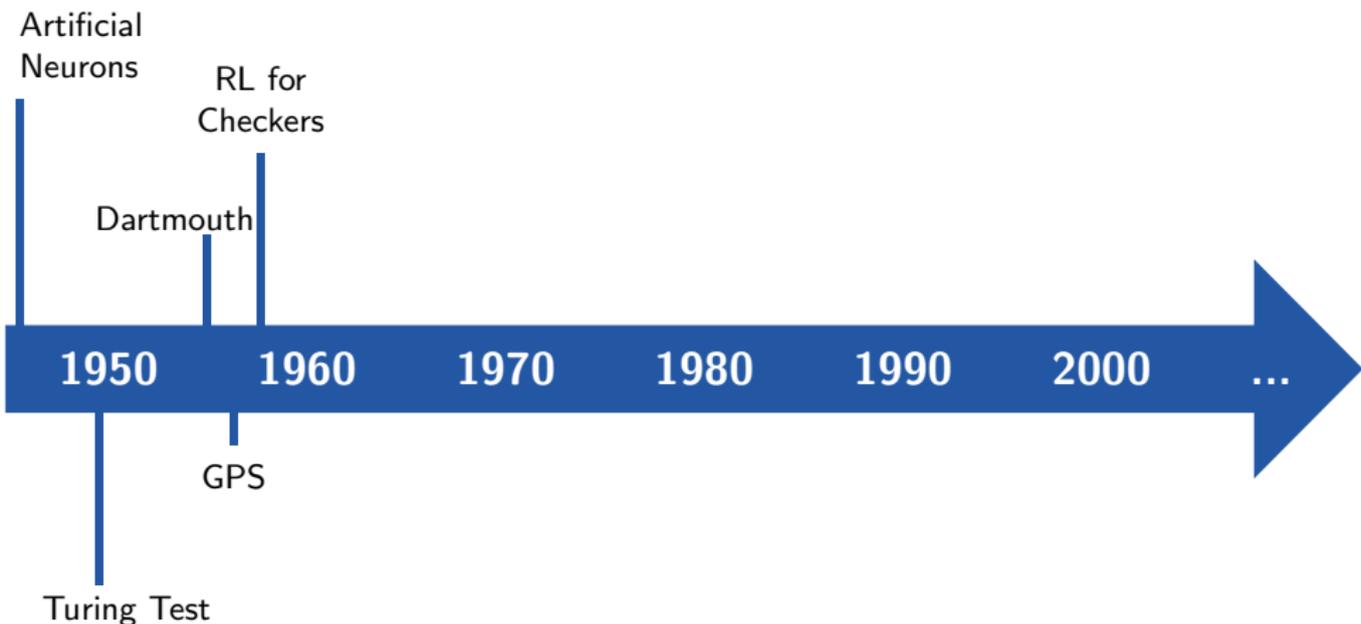
Turing Test

intelligence in **microworlds**, e.g. **SHRDLU** (T. Winograd, 1968)

- ▶ understands natural language
- ▶ communicates with user via teletype on **blocks world**
- ▶ graphical representation

↪ <https://hci.stanford.edu/winograd/shrdlu/>

Early Enthusiasm (1952–1969)



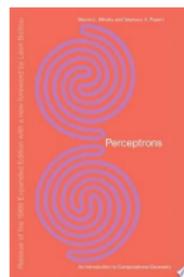
A Dose of Reality (1966–1973)

Artificial
Neurons

RL for
Checkers

Dartmouth

Limitations



1950

1960

1970

1980

1990

2000

...

GPS

Microworlds

Turing Test

- ▶ realization that unlimited computational power is illusion (birth of complexity theory, NP-completeness)
- ▶ AI systems (e.g., GPS, systems for micro worlds) *fail to scale*
- ▶ fundamental **limitations on basic structures** e.g., XOR problem of perceptrons

Expert Systems (1969–1986)

Artificial
Neurons

RL for
Checkers

Dartmouth

Limitations

DISTRIBUTE-MB-DEVICES-3

IF: the most current active context is distributing massbus devices
& there is a single port disk drive that has not been assigned to a massbus
& there are no unassigned dual port disk drives
& the number of devices that each massbus should support is known
& there is a massbus that has been assigned at least one disk drive and that should support additional disk drives
& the type of cable needed to connect the disk drive to the previous device on the disk drive is known
THEN: assign the disk drive to the massbus

1950

1960

1970

1980

1990

2000

...

GPS

Microworlds

Expert
Systems

Turing Test

1980s: AI gold rush

- ▶ rule-based expert systems commercially successful
- ▶ (human) expert knowledge as input
- ▶ allows automatic reasoning on larger problems in narrower applications
- ▶ also: second heyday of neural networks

Expert Systems (1969–1986)

Artificial
Neurons

RL for
Checkers

Dartmouth

Limitations

DISTRIBUTE-MB-DEVICES-3

```
IF:  the most current active context is distributing massbus devices
&   there is a single port disk drive that has not been assigned to a massbus
&   there are no unassigned dual port disk drives
&   the number of devices that each massbus should support is known
&   there is a massbus that has been assigned at least one disk drive and that should support additional
    disk drives
&   the type of cable needed to connect the disk drive to the previous device on the disk drive is known
THEN: assign the disk drive to the massbus
```

1950

1960

1970

1980

1990

2000

...

Turing Test

GPS

Microworlds

Expert
Systems

example: R1/XCON (J. McDermott, 1978)

- ▶ **input:** desired properties of a VAX computer system according to customer specifications
- ▶ **output:** specification of the computer system
- ▶ **inference engine:** simple forward chaining of rules

Expert Systems (1969–1986)

Artificial
Neurons

RL for
Checkers

Dartmouth

Limitations

DISTRIBUTE-MB-DEVICES-3

IF: the most current active context is distributing massbus devices
& there is a single port disk drive that has not been assigned to a massbus
& there are no unassigned dual port disk drives
& the number of devices that each massbus should support is known
& there is a massbus that has been assigned at least one disk drive and that should support additional disk drives
& the type of cable needed to connect the disk drive to the previous device on the disk drive is known
THEN: assign the disk drive to the massbus

1950

1960

1970

1980

1990

2000

...

GPS

Microworlds

Expert
Systems

AI Winter

Turing Test

end of 1980s: AI Winter

- ▶ companies failed to deliver promises
- ▶ expert systems difficult to maintain
- ▶ expert systems susceptible to uncertainty

Broad Visibility in Society (Since 2010s)

Artificial
Neurons

RL for
Checkers

Dartmouth

Limitations



1950

1960

1970

1980

1990

2000

...

Turing Test

GPS

Microworlds

Expert
Systems

AI Winter

well known systems and famous breakthroughs, e.g.,

- ▶ broadly used systems (e.g., virtual assistants)
- ▶ AI systems act in real-world (e.g., self-driving cars)
- ▶ systems outperform humans in hard tasks (e.g., AlphaGo)
- ▶ AI and human-written text hard to distinguish (ChatGPT)

A3.2 Where are We Today?

AI Approaching Maturity

Russell & Norvig (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.

Where are We Today?



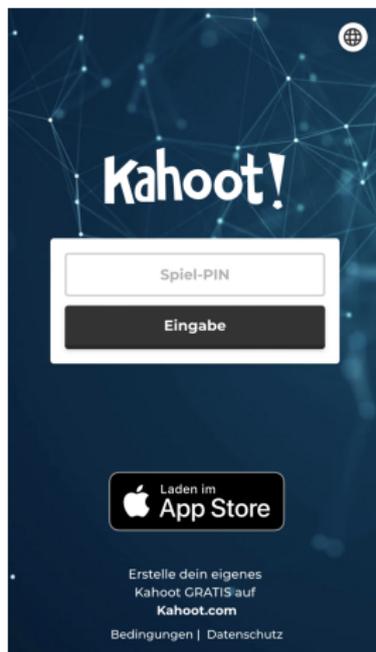
- ▶ many coexisting paradigms
 - ▶ reactive vs. deliberative
 - ▶ data-driven vs. model-driven
 - ▶ often hybrid approaches
- ▶ many methods, often borrowing from other research areas
 - ▶ logic, decision theory, statistics, ...
- ▶ different approaches
 - ▶ theoretical
 - ▶ algorithmic/experimental
 - ▶ application-oriented

Focus on Algorithms and Experiments

Many AI problems are inherently difficult (NP-hard), but strong search techniques and heuristics often solve large problem instances regardless:

- ▶ **satisfiability in propositional logic**
 - ▶ 10,000 propositional variables or more via **conflict-directed clause learning**
- ▶ **constraint solvers**
 - ▶ good scalability via **constraint propagation** and automatic exploitation of **problem structure**
- ▶ **action planning**
 - ▶ 10^{100} search states and more by search using **automatically inferred heuristics**

What Can AI Do Today?



<https://kahoot.it/>

A3.3 Summary

Summary

- ▶ 1950s/1960s: beginnings of AI; early enthusiasm
- ▶ 1970s: micro worlds and knowledge-based systems
- ▶ 1980s: gold rush of expert systems followed by “AI winter”
- ▶ 1990s/2000s: AI comes of age; research becomes more rigorous and mathematical; mature methods
- ▶ 2010s: AI systems enter mainstream

Foundations of Artificial Intelligence

A4. Introduction: Rational Agents

Malte Helmert

University of Basel

February 18, 2026

Foundations of Artificial Intelligence

February 18, 2026 — A4. Introduction: Rational Agents

A4.1 Systematic AI Framework

A4.2 Example

A4.3 Rationality

A4.4 Summary

Introduction: Overview

Chapter overview: introduction

- ▶ A1. Organizational Matters
- ▶ A2. What is Artificial Intelligence?
- ▶ A3. AI Past and Present
- ▶ A4. Rational Agents
- ▶ A5. Environments and Problem Solving Methods

A4.1 Systematic AI Framework

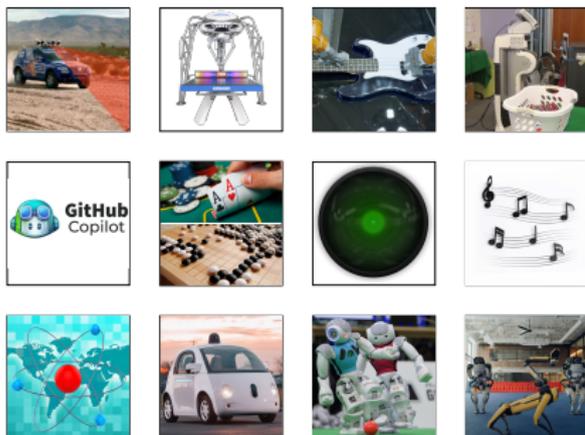
Systematic AI Framework

so far we have seen that:

- ▶ AI systems act rationally



- ▶ AI systems applied to wide variety of challenges



now: describe a **systematic framework** that

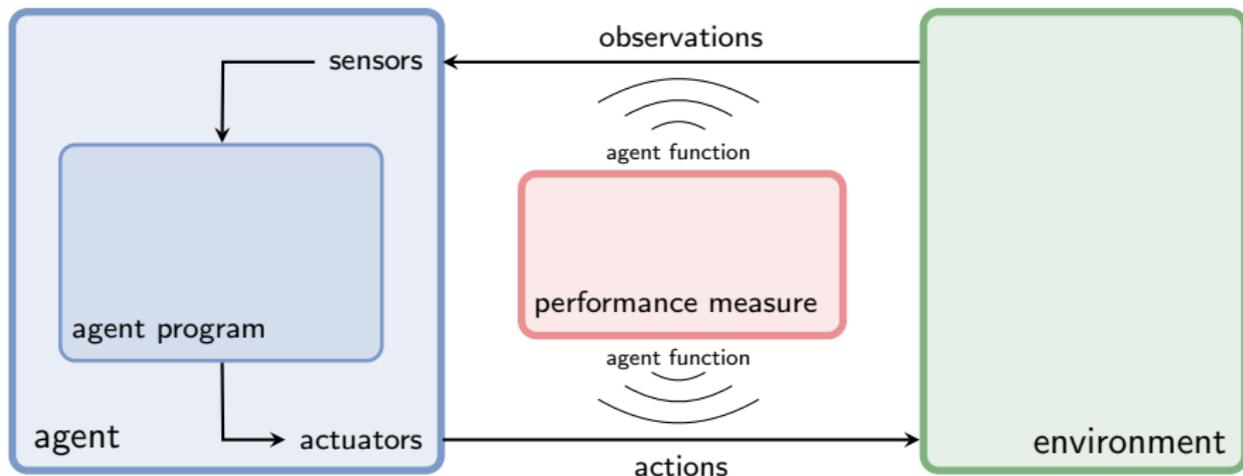
- ▶ captures this **diversity of challenges**
- ▶ includes an entity that **acts** in the environment
- ▶ determines if the agent acts **rationally** in the environment

Systematic AI Framework

so far we have seen that:

- ▶ AI systems act rationally

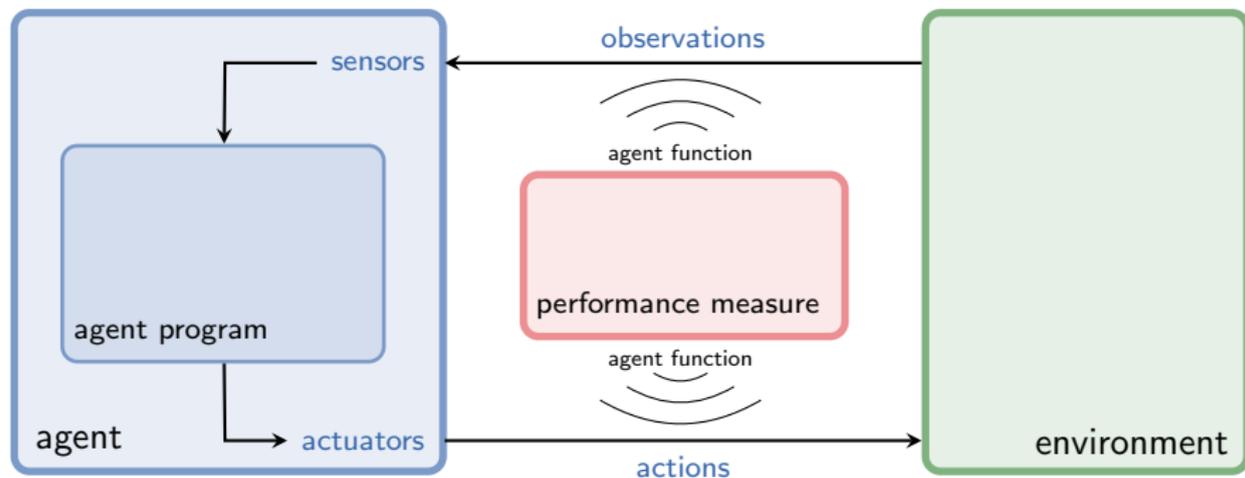
- ▶ AI systems applied to wide variety of challenges



now: describe a systematic framework that

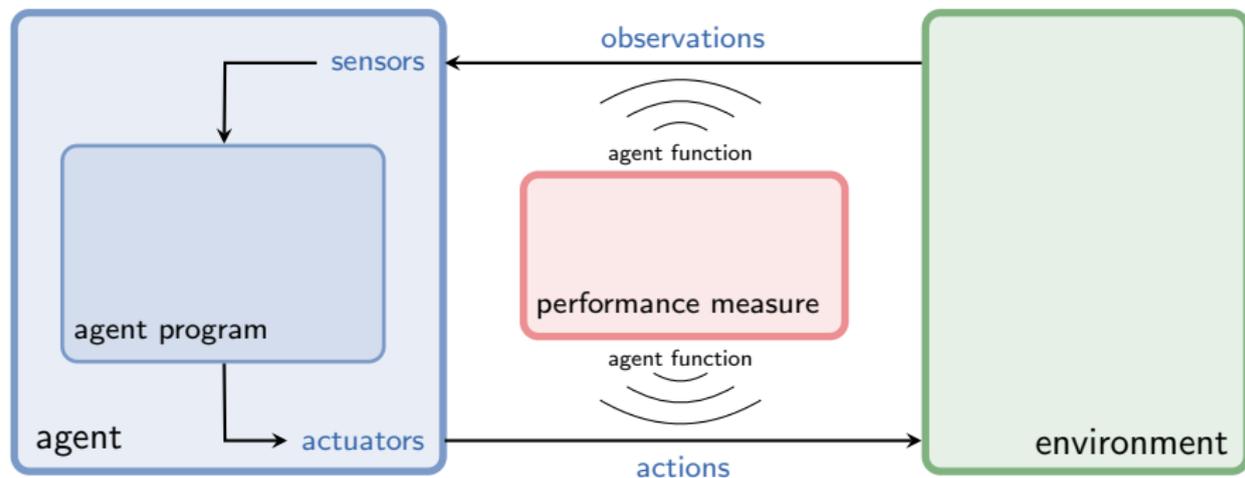
- ▶ captures this diversity of challenges
- ▶ includes an entity that acts in the environment
- ▶ determines if the agent acts rationally in the environment

Agent-Environment Interaction



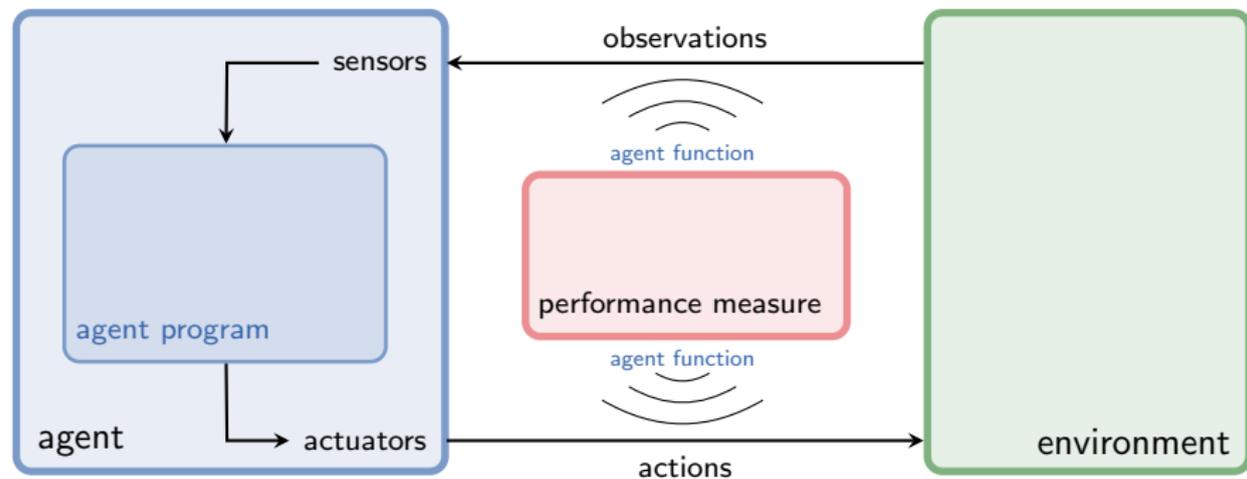
- ▶ **sensors**: physical entities that allow the agent to **observe**
- ▶ **observation**: data perceived by the agent's sensors
- ▶ **actuators**: physical entities that allow the agent to **act**
- ▶ **action**: abstract concept that affects the state of the environment

Agent-Environment Interaction



- ▶ **sensors** and **actuators** are not relevant for the course
(\rightsquigarrow typically covered in courses on **robotics**)
- ▶ **observations** and **actions** describe the agent's capabilities
(the **agent model**)

Formalizing an Agent's Behavior



① as agent program:

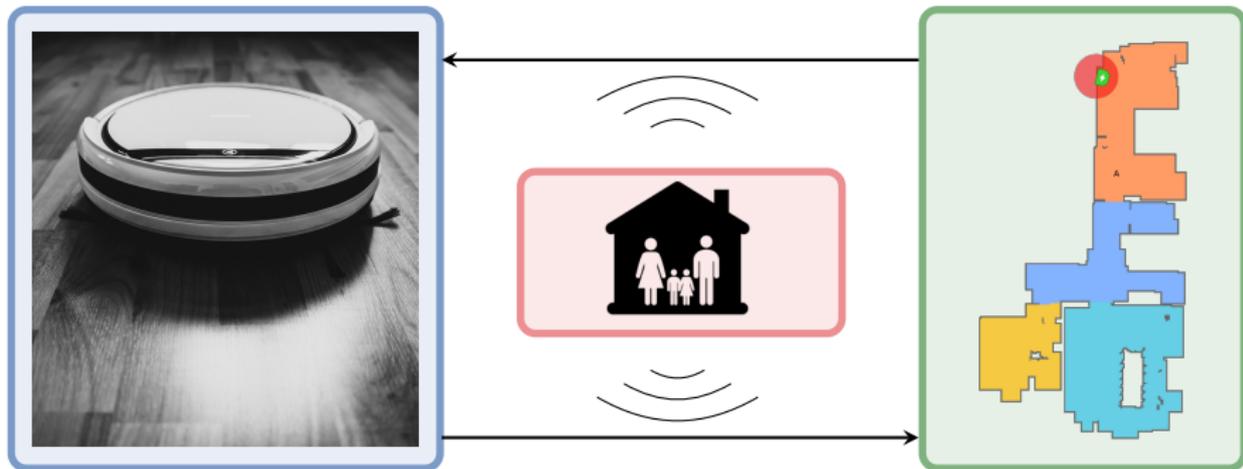
- ▶ internal representation
- ▶ specifics possibly *unknown* to outside
- ▶ takes *observation* as input
- ▶ outputs an *action*
- ▶ computed on physical machine (the *agent architecture*)

② as agent function:

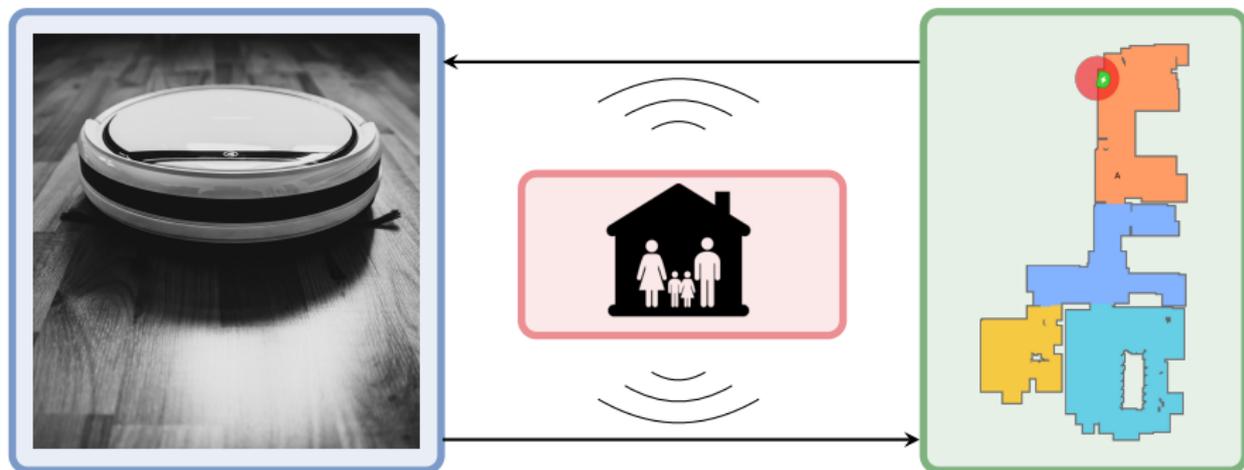
- ▶ external characterization
- ▶ maps *sequence of observations* to (probability distribution over) *actions*
- ▶ *abstract mathematical formalization*

A4.2 Example

Vacuum Domain

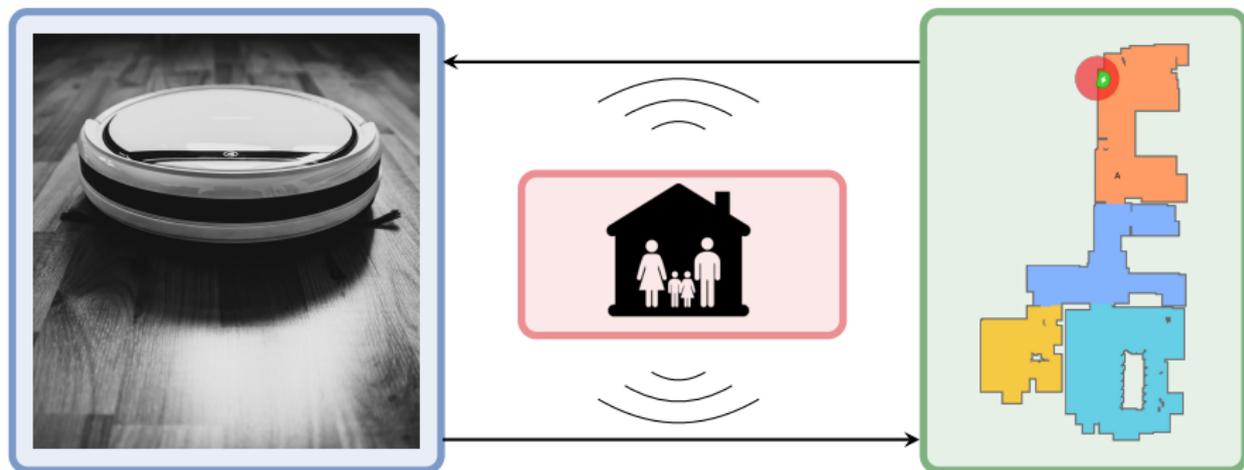


Vacuum Agent: Sensors and Actuators



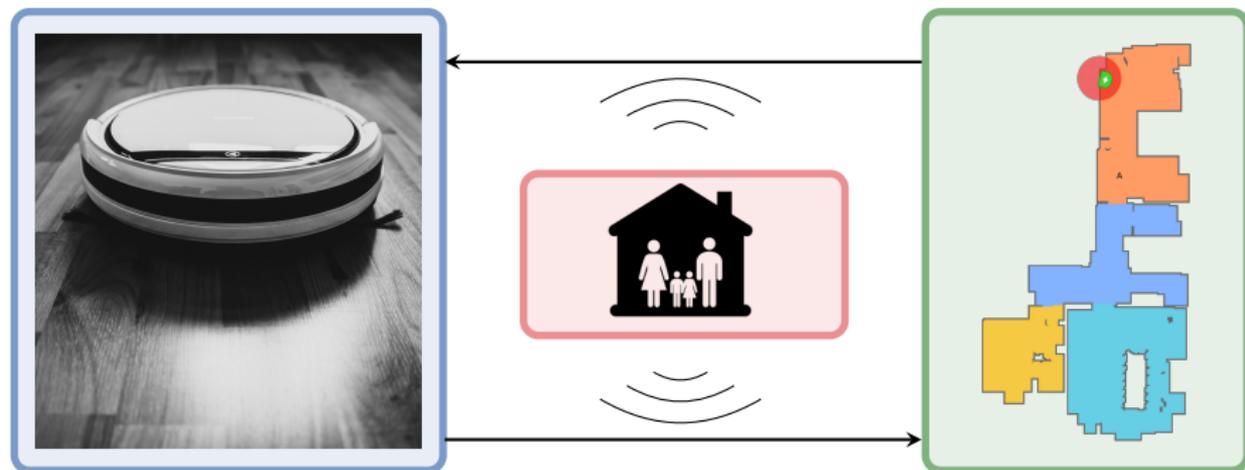
- ▶ **sensors:** cliff sensors, bump sensors, wall sensors, state of charge sensor, WiFi module
- ▶ **actuators:** wheels, cleaning system

Vacuum Agent: Observations and Actions



- ▶ **observations:** current location, dirt level of current room, presence of humans, battery charge
- ▶ **actions:** move-to-next-room, move-to-base, vacuum, wait

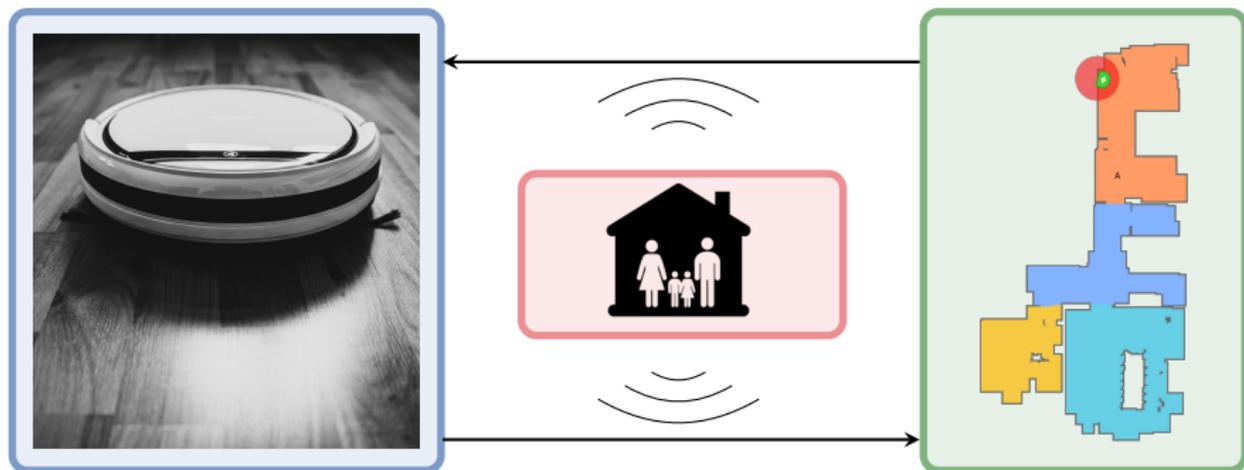
Vacuum Agent: Agent Program



```

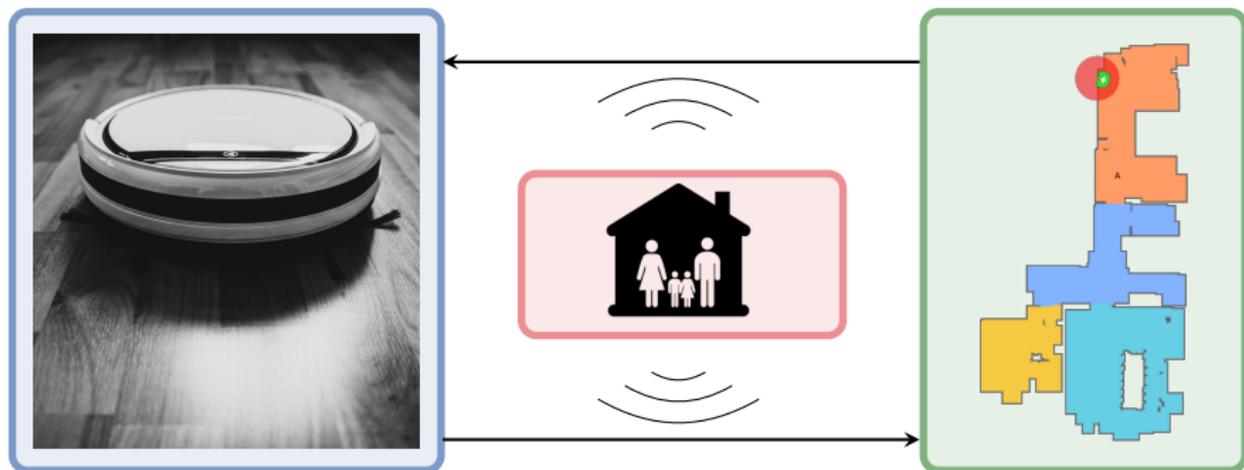
1 def vacuum-agent([location, dirt-level, owner-present, battery]):
2   if battery ≤ 10%: return move-to-base
3   else if owner-present = True: return move-to-next-room
4   else if dirt-level = dirty: return vacuum
5   else: return move-to-next-room
  
```

Vacuum Domain: Agent Function



observation sequence	action
$\langle [\text{blue, clean, False, 100\%}] \rangle$	<i>move-to-next-room</i>
$\langle [\text{blue, dirty, False, 100\%}] \rangle$	<i>vacuum</i>
$\langle [\text{blue, clean, True, 100\%}] \rangle$	<i>move-to-next-room</i>
...	...
$\langle [\text{blue, clean, False, 100\%}], [\text{blue, clean, False, 90\%}] \rangle$	<i>move-to-next-room</i>
$\langle [\text{blue, clean, False, 100\%}], [\text{blue, dirty, False, 90\%}] \rangle$	<i>vacuum</i>
...	...

Vacuum Domain: Performance Measure



potential influences on **performance measure**:

- ▶ dirt levels
- ▶ noise levels
- ▶ energy consumption
- ▶ safety

A4.3 Rationality

Evaluating Agent Functions



What is the **right** agent function?

Rationality

rationality of an **agent** depends on **performance measure** (often: **utility**, **reward**, **cost**) and **environment**

Perfect Rationality

- ▶ for each possible **observation sequence**
- ▶ select an action which **maximizes**
- ▶ **expected value** of future performance
- ▶ given **available information** on **observation history**
- ▶ and **environment**

Perfect Rationality of Our Vacuum Agent

Is our vacuum agent **perfectly rational**?



depends on performance measure and environment, e.g.:

- ▶ Do actions reliably have the desired effect?
- ▶ Do we know the initial situation?
- ▶ Can new dirt be produced while the agent is acting?

Performance Measure

- ▶ specified by designer
- ▶ sometimes clear, sometimes not so clear
- ▶ significant impact on
 - ▶ desired behavior
 - ▶ difficulty of problem



Performance Measure

- ▶ specified by designer
- ▶ sometimes clear, sometimes not so clear
- ▶ significant impact on
 - ▶ desired behavior
 - ▶ difficulty of problem



Perfect Rationality of Our Vacuum Agent

consider **performance measure**:

- ▶ +1 utility for cleaning a dirty room

consider **environment**:

- ▶ actions and observations reliable
- ▶ world only changes through actions of the agent

our vacuum agent is **perfectly rational**

Perfect Rationality of Our Vacuum Agent

consider **performance measure**:

- ▶ -1 utility for each dirty room in each step

consider **environment**:

- ▶ actions and observations reliable
- ▶ world only changes through actions of the agent

our vacuum agent is **not perfectly rational**

Perfect Rationality of Our Vacuum Agent

consider **performance measure**:

- ▶ -1 utility for each dirty room in each step

consider **environment**:

- ▶ actions and observations reliable
- ▶ yellow room may spontaneously become dirty

our vacuum agent is **not perfectly rational**

Rationality: Discussion

- ▶ perfect rationality \neq omniscience
 - ▶ incomplete information (due to limited observations) reduces achievable utility
- ▶ perfect rationality \neq perfect prediction of future
 - ▶ uncertain behavior of environment (e.g., stochastic action effects) reduces achievable utility
- ▶ perfect rationality is rarely achievable
 - ▶ limited computational power \rightsquigarrow bounded rationality

A4.4 Summary

Summary (1)

common metaphor for AI systems: **rational agents**

agent interacts with **environment**:

- ▶ sensors perceive **observations** about state of the environment
- ▶ actuators perform **actions** modifying the environment
- ▶ formally: **agent function** maps observation sequences to actions

Summary (2)

rational agents:

- ▶ try to maximize performance measure (utility)
- ▶ perfect rationality: achieve maximal utility in expectation given available information
- ▶ for “interesting” problems rarely achievable
 \rightsquigarrow bounded rationality