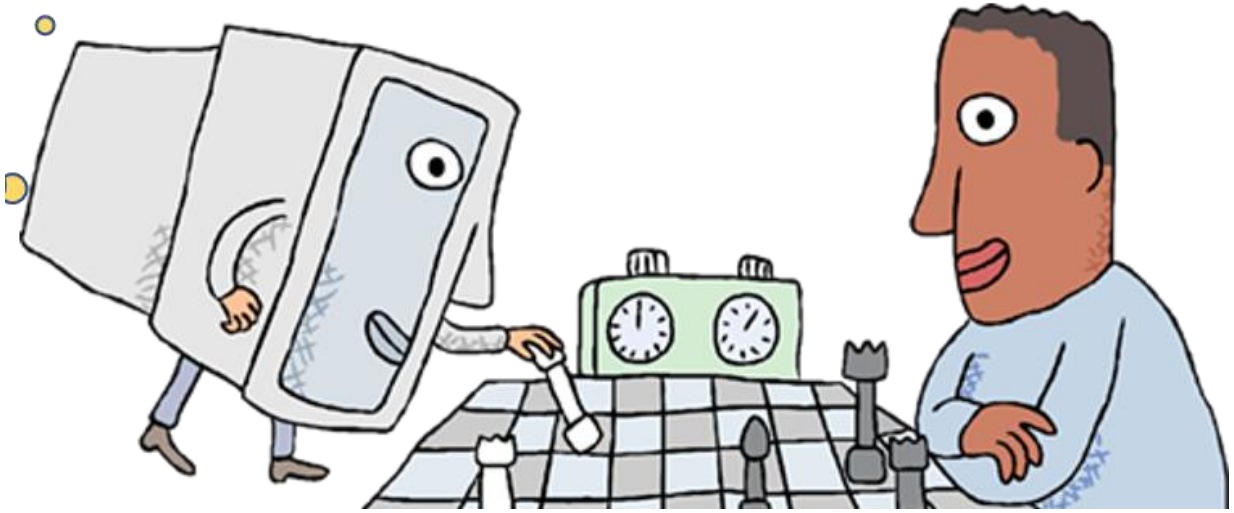
CS5491: Artificial Intelligence

Introduction



Instructor: Kai Wang

Acknowledgement

Developer of CS5491

To have fun!

Ying WEI (魏 颖)

ying.wei [at] zju [dot] edu [dot] cn

I am currently a ZJU-100 Young Professor with College of Computer Science and Technology, Zhejiang University.

I am generally interested in developing algorithms that equip machines with more general intelligence via knowledge transfer and compositionality. This includes allowing continuous transfer and adaptation of the knowledge previous learned (nowadays in LLMs) to quickly learn the current task with minimal human supervision, and autonomously evaluating the success of knowledge transfer. I am also passionate about applying these algorithms into real-world applications with small data, e.g., AI for Science.

Previously, I was a Nanyang Assistant Professor with College of Computing and Data Science, Nanyang Technological University, an Assistant Professor at Department of Computer Science, City University of Hong Kong and a senior researcher at Tencent AI Lab. I completed my Ph.D. in Computer Science and Engineering at Hong Kong University of Science and Technology under the supervision of Professor Qiang Yang, and my B.S. in Automation at Huazhong University of Science and Technology. I have also spent time interning at Microsoft Research Asia.

Course Information

Prerequisites

- CS 3334 Data Structures
- CS 4335 Design and Analysis of Algorithms

There will be some math/programming.

Canvas

- Syllabus
- Schedule and lecture slides
- Programming projects
- Other material

Check it often!

Course Material

Textbook

- Russell, S., & Norvig, P. *Artificial intelligence: a modern approach*. 3rd edition.
- Sutton, R. S., & Barto, A. G. *Reinforcement learning: An introduction*. MIT press. 2018

Highly recommend for students who want to read more.

Lecture Slides

- I will post a version of each lecture's slides.
- The version is possibly updated on Sunday with changes and annotations after lecture.

Download the slides!

Communication

Canvas

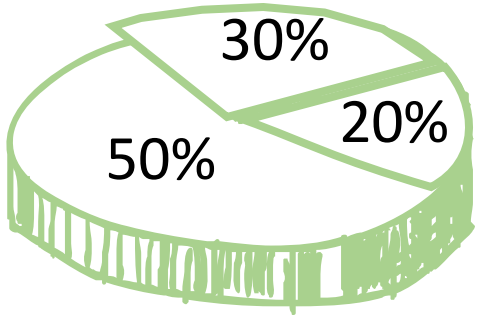
- Post questions on course material
- Post questions on assignments
- Expect 24 hours to respond by the teaching staff

We will not answer questions on course material via email.

Evaluation

Two Assignments

Final examination



Mid-term Exam

- We have up to +5 in your final mark upon question answering in class.
- Academic integrity policy

Late/Missing Policy

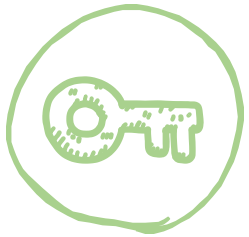
Late

- 2 late days for each project
- Additional late days will not be granted except under exceptional circumstances.
- If all late days are used up, you lose 20% each day.

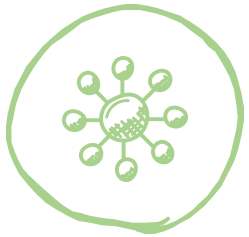
Missing

- Late days should cover all the reasons you will be late, unless something more seriours like an extended illness occur.
- A note from your doctor is required.
- If you miss a project, your score will be reweighted to exclude that project.
- If you miss the final, a make-up is required.

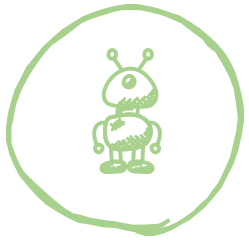
Today



What can AI do?



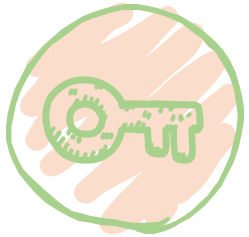
What is AI?



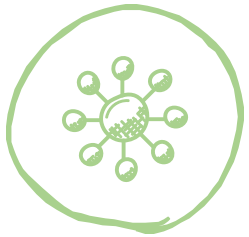
What is an intelligent agent?

What is this course?

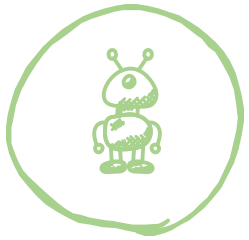
Today



What can AI do?



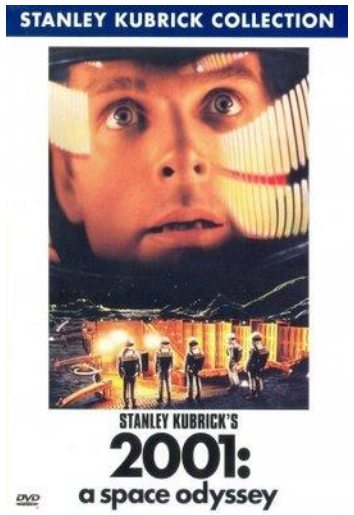
What is AI?

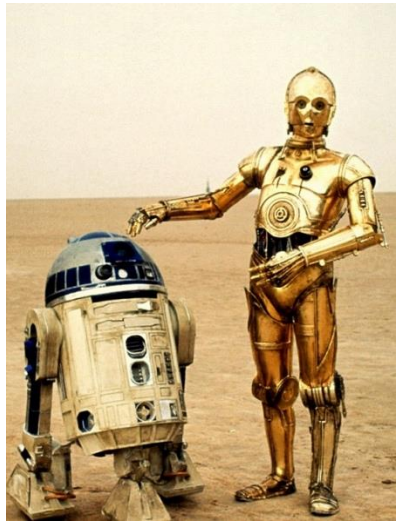


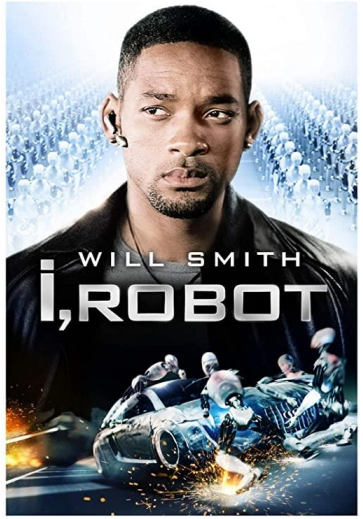
What is an intelligent agent?

What is this course?

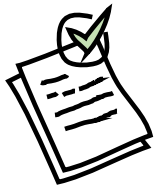
What Can AI Do?







What Can AI Do?



Clicker question: Which of the following can be done at present?

Play a game of badminton.
Play a game of chess/go.
Drive safely along a curving mountain road.
Drive safely along Songshan Lake Road.
Buy a week's worth of food on the web.
Buy a week's worth of food at Alibaba/Amazon offline Shop.
Discover and prove a new mathematical theorem.
Chat succesfully with another person for an hour.
Perform a surgical operation.
Translate spoken Chinese into spoken English in real time.
Write an intentially funning story.

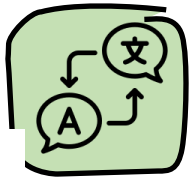
Nature Language Processing

Question answering

Context: California is a major economic center for US…

Question: What is a major importance of California?

Anaswer: major economic center



Machine translation

French: Cette rubrique comprend plusieurs tâches connexes.

English: This rubric includes several related tasks.

Sentiment analysis

Resturant review: The dessert is excellent.

Sentiment: Positive

Web search

Query: CityU CS Artificial Intelligence

Top-2 results:

www.cityu.edu.hk › current › course ▾ 翻译此页

CS4486 - Artificial Intelligence - City University of Hong Kong

CS4486 - Artificial Intelligence ... Department of Computer Science ... skills of problem solving

www.cs.cityu.edu.hk › research_areas › Artifi... ▾ 翻译此页

Artificial Intelligence And Knowledge Management - CityU CS

Artificial Intelligence And Knowledge Management - Research Areas - Department of Computer

Nature Language Processing

Speech/Audio

- InterSpeech
- ICASSP

Speech

Automatic speech recognition

Speech: 

Text: This rubric includes several related tasks.

Text-to-speech synthesis

Text: This rubric includes several related tasks.

Speech: 

Chatbot

User: Are you mad at me?

Siri: I can't answer that.

User: Do you love me?

Siri: Would you like me to search the web for "love"?

Speaker verification

Speech: 

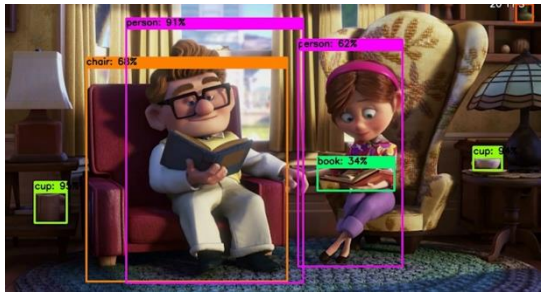
Question: Is this Bob's voice?

Answer: Yes

Speech

Perception (Computer Vision)

Object and face recognition



Scene segmentation

Image classification

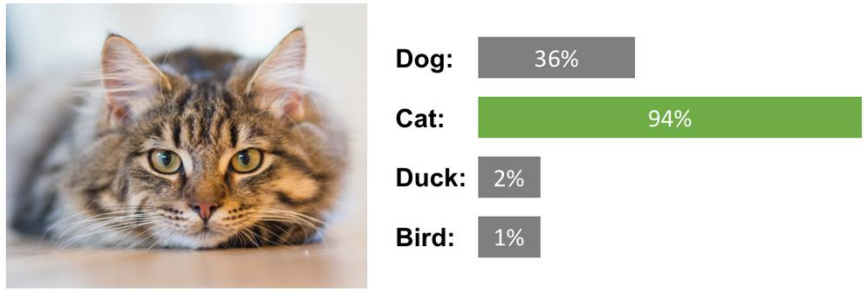
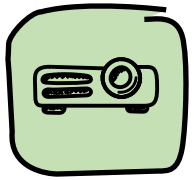


Image captioning

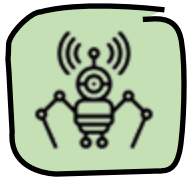
Perception

Robotics

Self-driving car



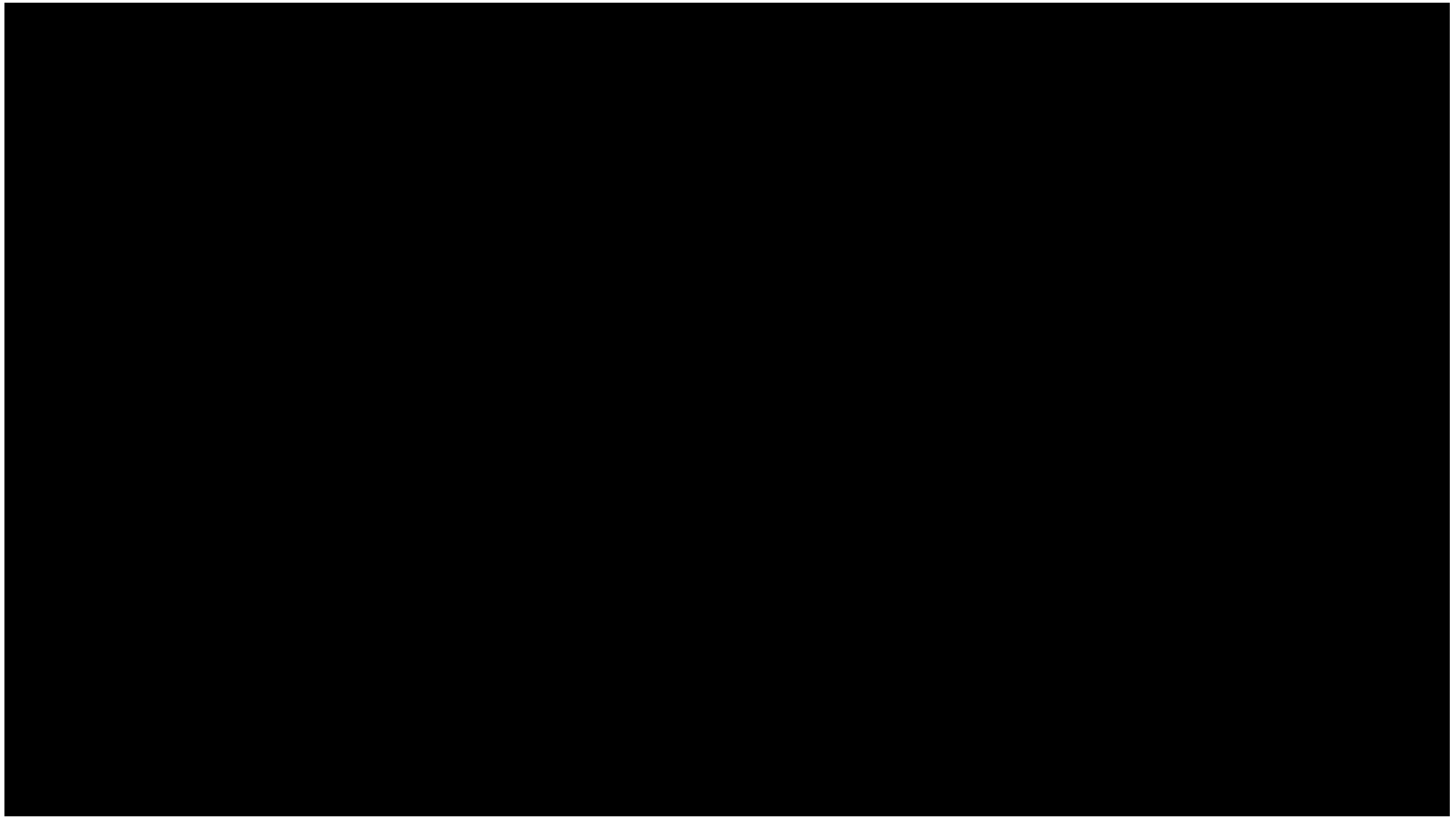
Warehousing robot



Rescue robot

Soccer cobots

Self-driving Car



Warehousing Robot

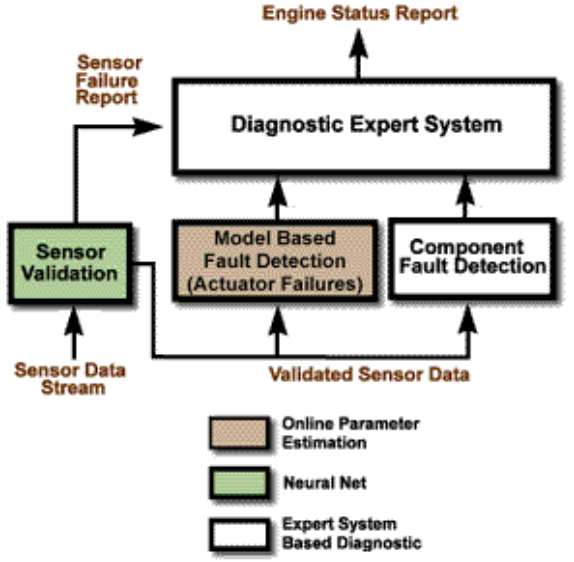
Logic

Theorem provers

Machine proof in LT

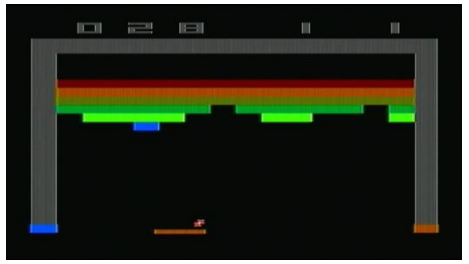
*2.85	$p \vee q . \supset . p \vee r : \supset : p . \vee . q \supset r$	
1.	$p \supset q . \supset : q \supset r . \supset . p \supset r$	*2.06
	$p/q, q/p \vee q, r/p \vee r$	
2.	$q \supset p \vee q . \supset : p \vee q \supset p \vee r . \supset . q \supset p \vee r$	Sub. 1
3.	$q \supset p \vee q$	*1.3
4.	$p \vee q \supset p \vee r . \supset . q \supset p \vee r$	*Modus ponens* 3, 2
5.	$p \vee q \supset p \vee r . \supset . \sim q \vee (p \vee r)$	Def. of $\supset$
6.	$p \vee (q \vee r) . \supset . q \vee (p \vee r)$	*1.5
	$p/\sim q, q/p$	
7.	$\sim q \vee (p \vee r) . \supset . p \vee (\sim q \vee r)$	Sub. 6
8.	$p \vee q \supset p \vee r . \supset . p \vee (\sim q \vee r)$	Chain with 5 and 7
9.	$p \vee q \supset p \vee r . \supset . p . \vee . q \supset r$	Def of $\supset$

NASA fault diagnosis

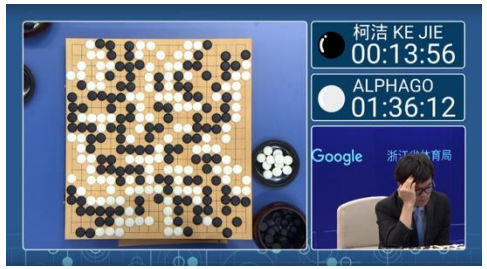


Game Playing

Chess



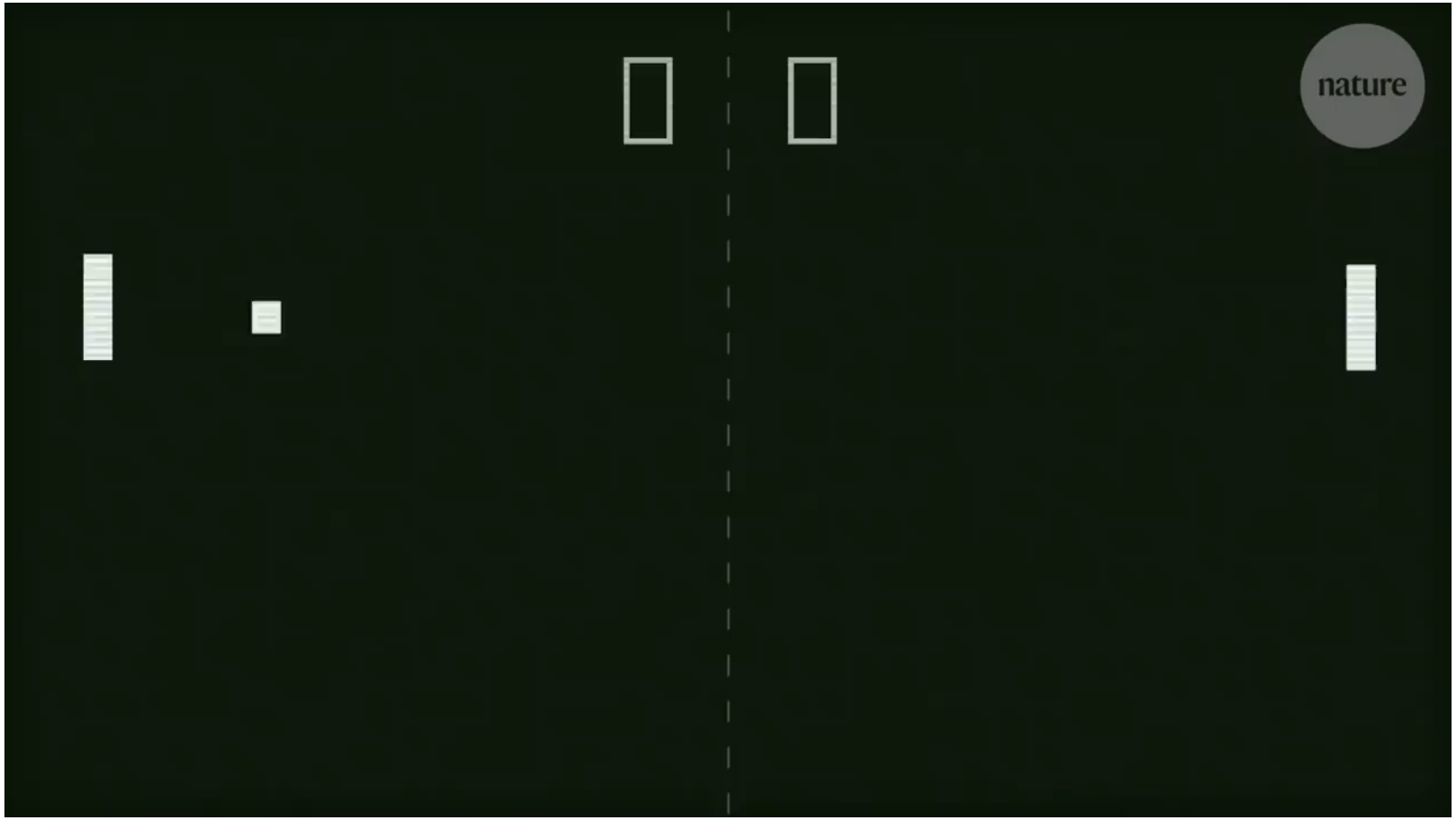
Atari games



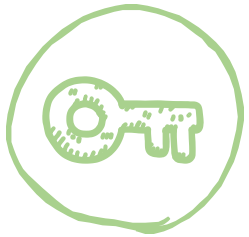
Go

Jeopardy

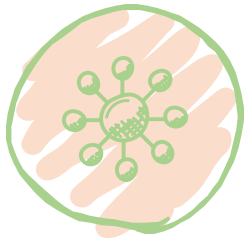
Game Playing



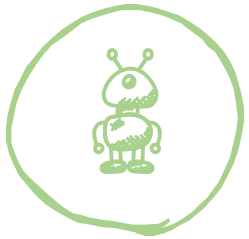
Today



What can AI do?



What is AI?

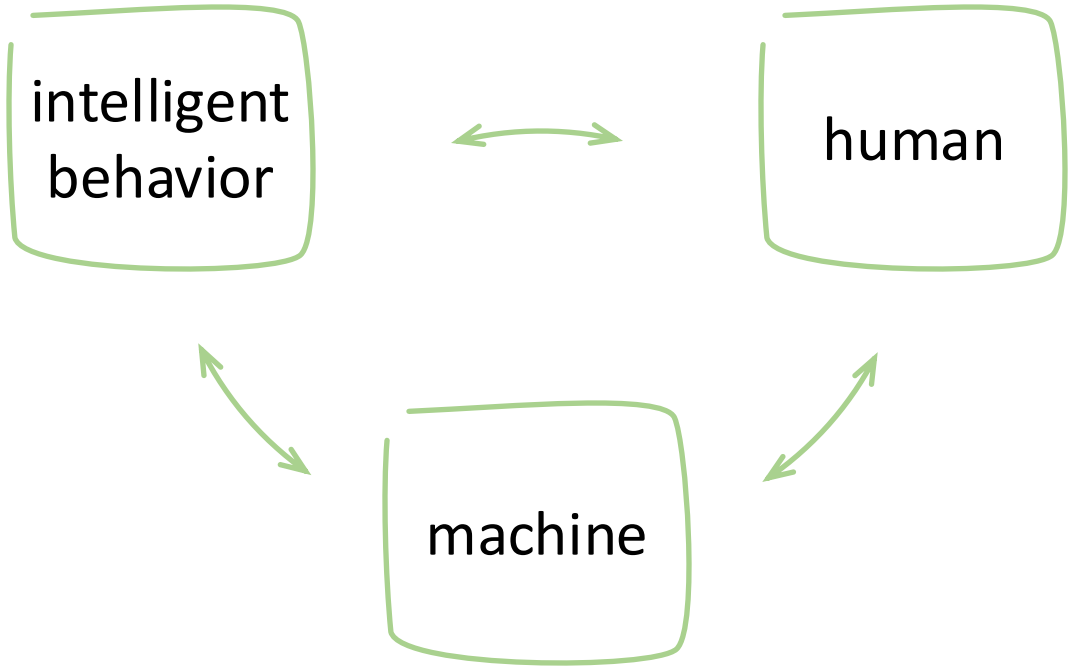


What is an intelligent agent?

What is this course?

What is AI?

AI is an attempt of reproduction of human reasoning and intelligent behavior by computational methods.



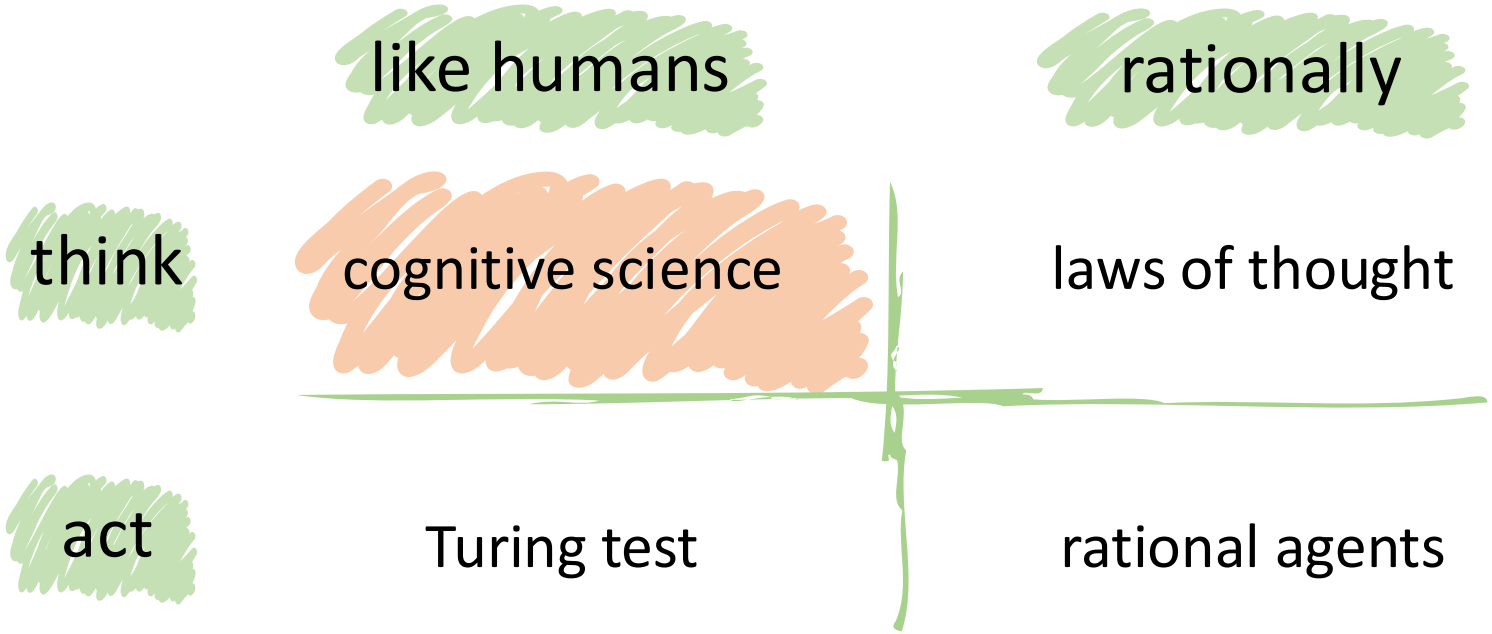
What is AI?

[Russell & Norvig] AI is a discipline that systematizes and automates reasoning processes to create machines that

	like humans	rationally
think	cognitive science	laws of thought
act	Turing test	rational agents

What is AI?

[Russell & Norvig] AI is a discipline that systematizes and automates reasoning processes to create machines that



Thinking Humanly

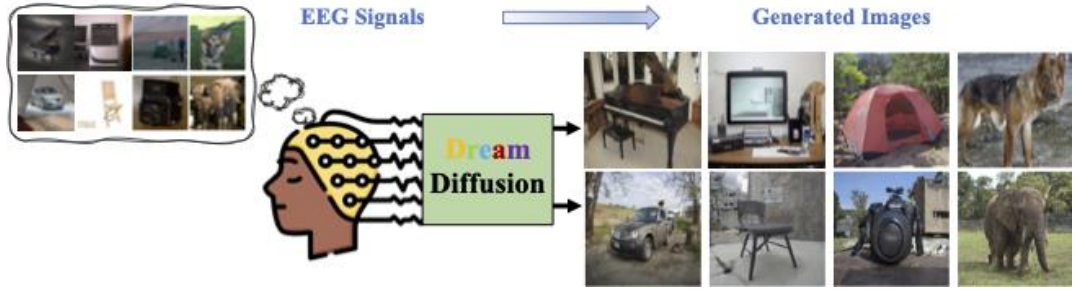
Modelling how humans think

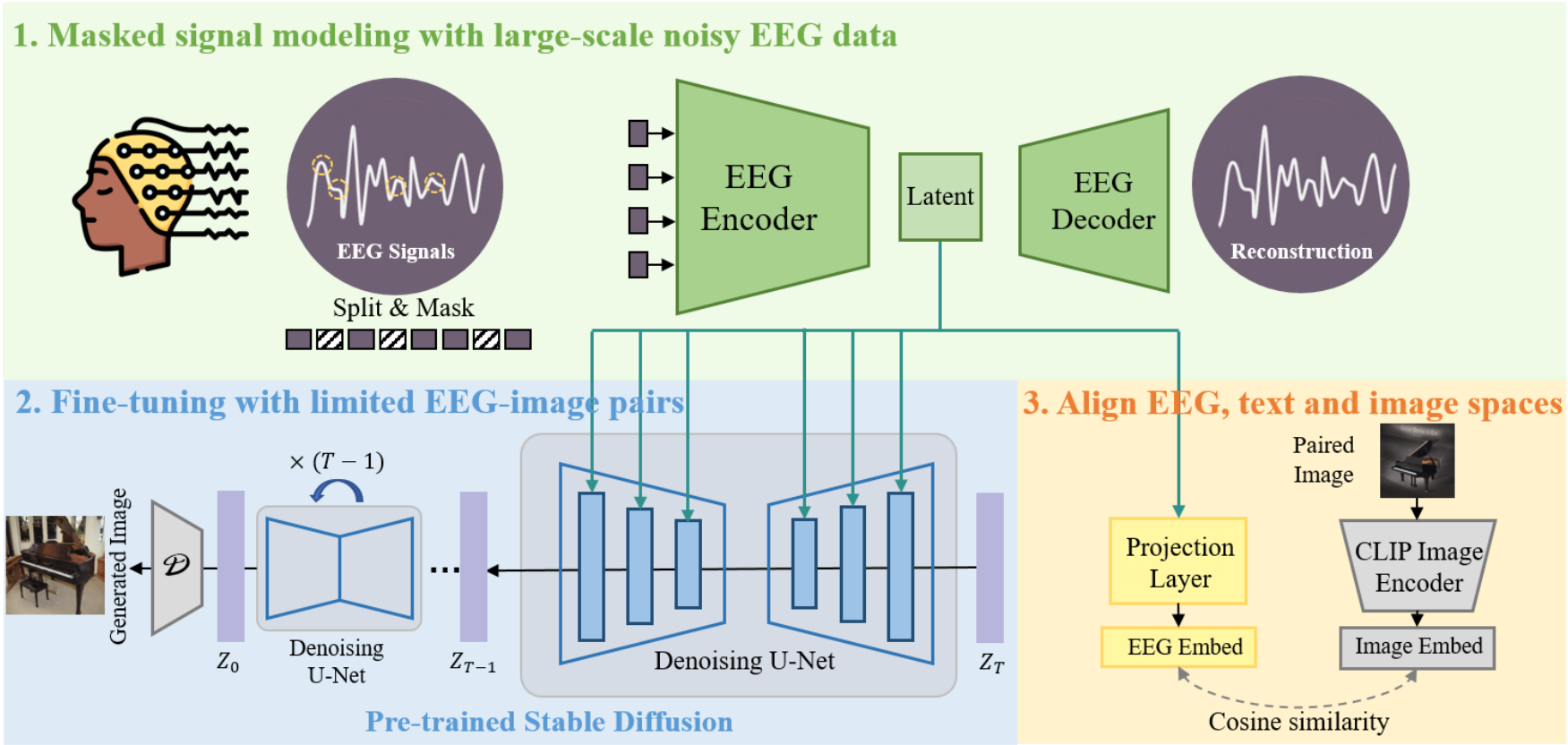
- Introspection (too subjective)
- Psychological experiments (predicting behaviors of humans based on cognitive theory and then testing, with matched predictions as supporting evidence)
- Brain imaging (observing the brain in action)

Cognitive science

- Top-down: Computer models from AI to simulate a theory and experimental techniques from psychology
- Bottom-up: Neuropsysiological evidence incorporated into computational models

DreamDiffusion





Thinking Humanly

- Cognitive science and AI are distinct disciplines, except the close connection of cognitive neuroscience to AI in neural networks.

- It is especially challenging to model how humans think, as brains are not as modular as software.

What is AI?

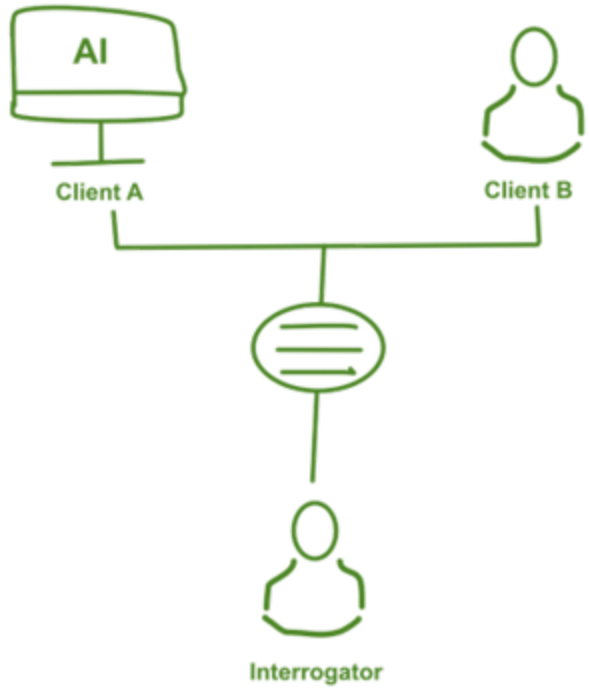
[Russell & Norvig] AI is a discipline that systematizes and automates reasoning processes to create machines that

	like humans	rationally
think	cognitive science	laws of thought
act	Turing test	rational agents

Acting Humanly

Turing test

- Proposed by Alan Turing in 1950
- "Can machines think?" → "Can machines behave intelligently?"
- Operational definition of intelligent behavior
- Can a human interrogator tell whether (written) responses to her (written) questions come from a human or an AI machine?
- Required capabilities: natural language processing, knowledge representation, automated reasoning, learning…



Acting Humanly

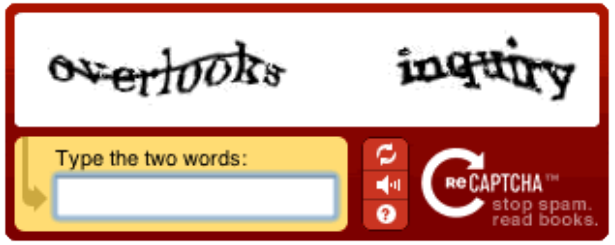
Attempts on passing the Turing test

- 1960s: ELIZA (by Joseph Weizenbaum)
- 1990s: ALICE
- Yearly: Loebner prize
  (http://www.loebner.net/Prizef/loebner-prize.html)

No machines has fully passed the test yet.

Applications of the Turing test

- CAPTCHA: Completely Automatic Public Turing tests to tell Computers and Humans Apart



Why do we want to replicate human behavior, including the imperfections?

Rationality

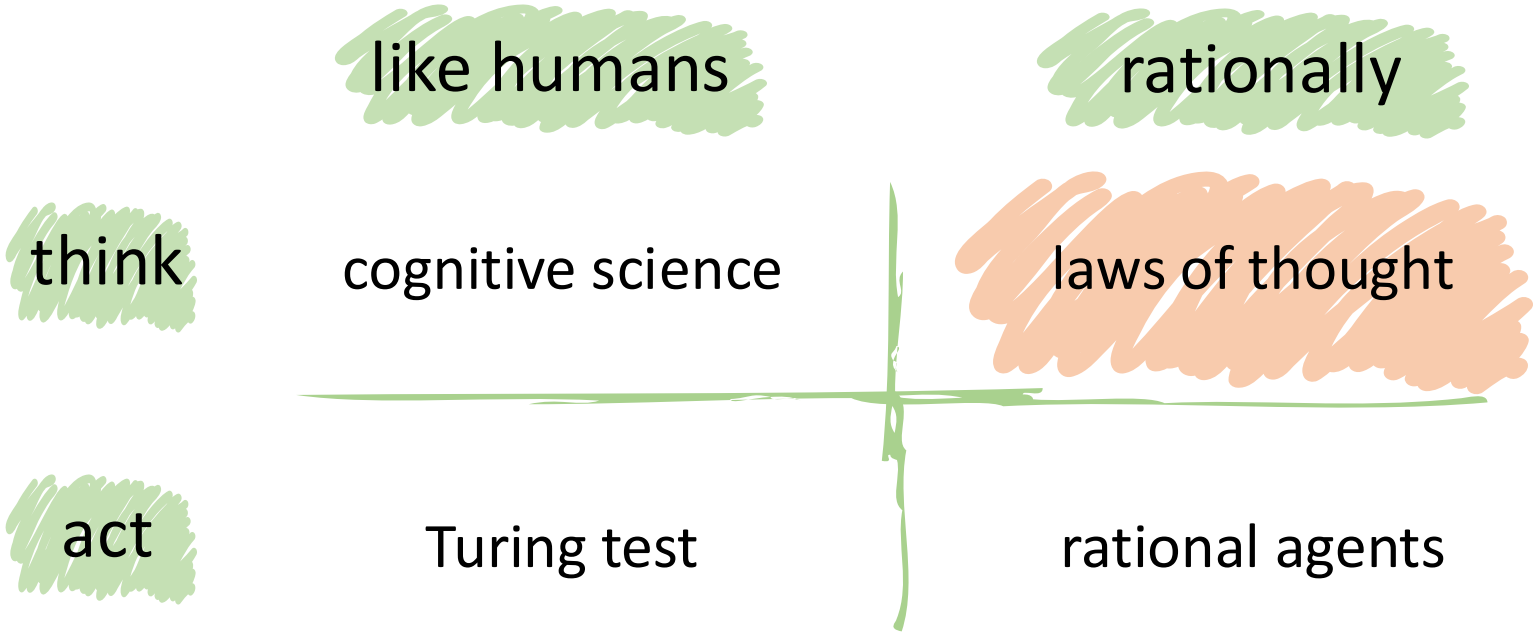
An abstract ideal of intelligence — do the right thing rather than “whatever humans think or act”. (are humans rational?)

Definitions of “right”

- Logic: Conclusions are provable from inputs.
- Economics: The utility of outcomes is maximized.

What is AI?

[Russell & Norvig] AI is a discipline that systematizes and automates reasoning processes to create machines that



Thinking Rationally

Logicist tradition

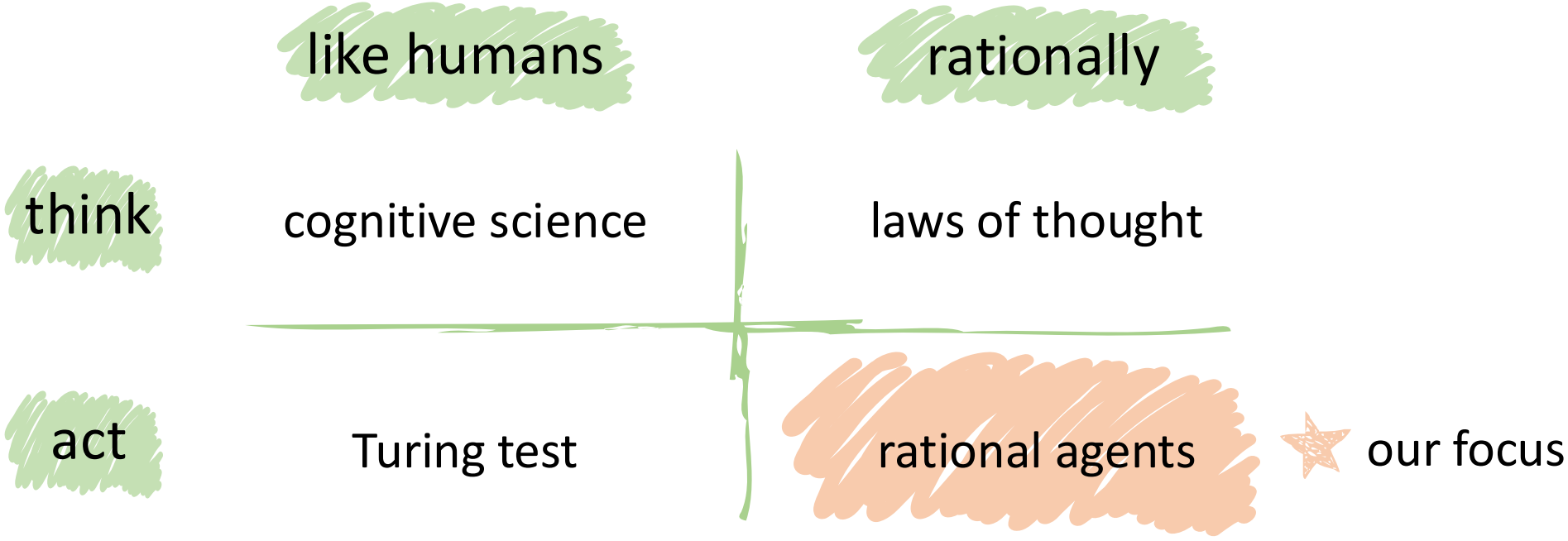
- Syllogisms by ancient Greeks: argument structures that always yield correct conclusions given correct premises
- Logic: notation and rules for derivation for thoughts

Limitations

- Not all intelligent behavior is mediated by logical deliberation.
- It is difficult to represent informal knowledge in logic.
- Logical systems tend to do the wrong thing in the presence of uncertainty.

What is AI?

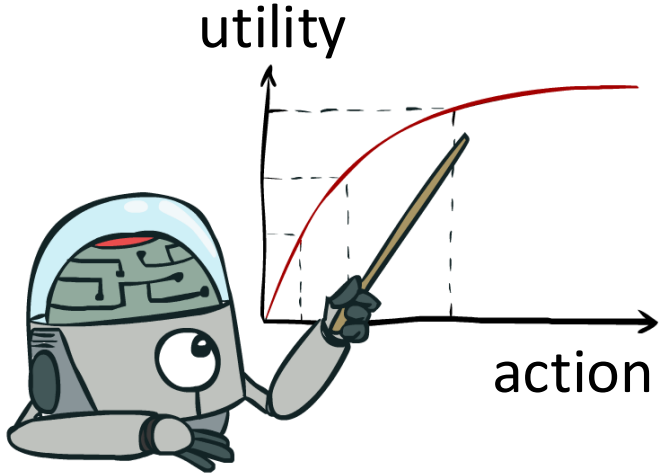
[Russell & Norvig] AI is a discipline that systematizes and automates reasoning processes to create machines that



Acting Rationally

- Rationality concerns what actions are made instead of the thought process behind them.

- Getting machines to be rational means maximizing the expected utility based on their circumstances and what they know.



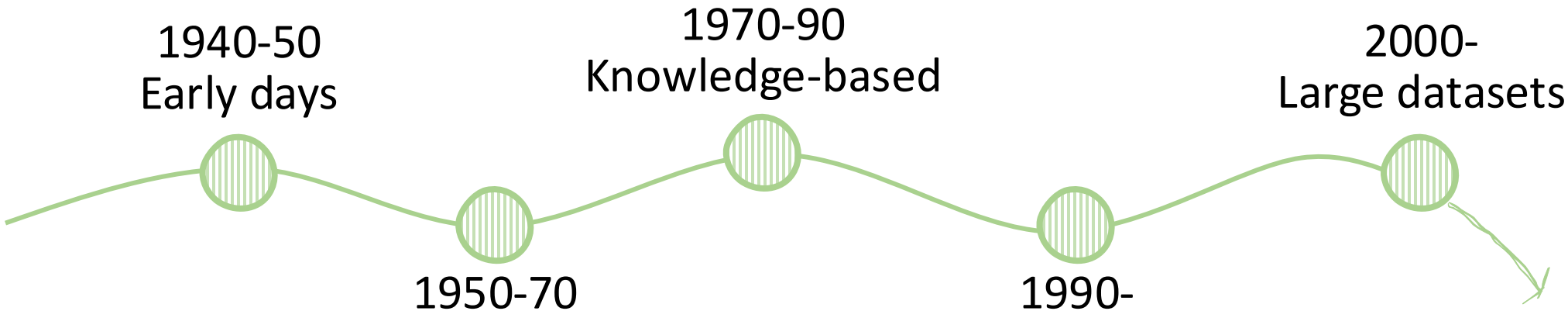
Irrationally ≠ insane ; an irrationally solution could be sub-optimal.

Rationally ≠ successful ; the most rational action may fail due to out-of-control circumstances or due to incomplete knowledge.

A Short History of AI

- 1943: McCulloch & Pitts: Boolean circuit model of brain
- 1950: Turing's "Computing Machinery and Intelligence"

- 1969—79: Early development of knowledge-based systems
- 1980—88: Expert systems industry booms
- 1988—93: Expert systems industry busts: "AI Winter"

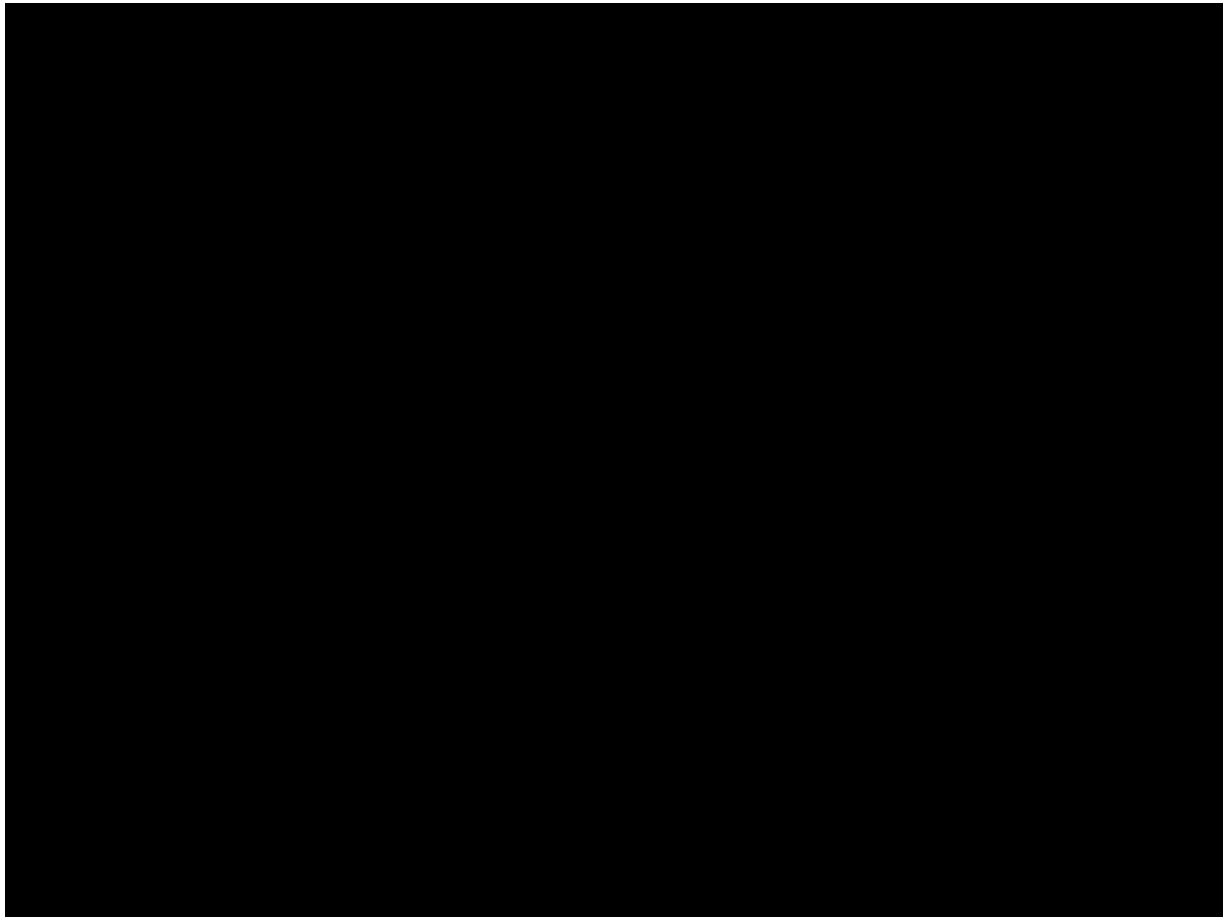


- 1950s: Early AI programs, including Samuel's checkers program, Newell & Simon's Logic Theorist, Gelernter's Geometry Engine
- 1956: Dartmouth meeting: "Artificial Intelligence" adopted
- 1965: Robinson's complete algorithm for logical reasoning

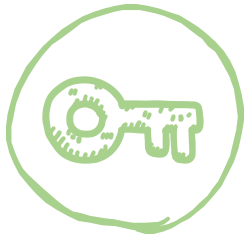
- Resurgence of probability, focus on uncertainty
- General increase in technical depth
- Agents and learning systems… "AI Spring"?

Statistic vs. Neural Network

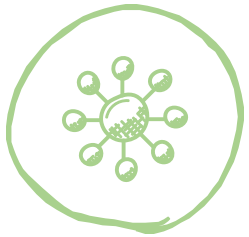
AI in the 1960s



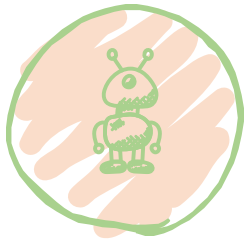
Today



What can AI do?



What is AI?



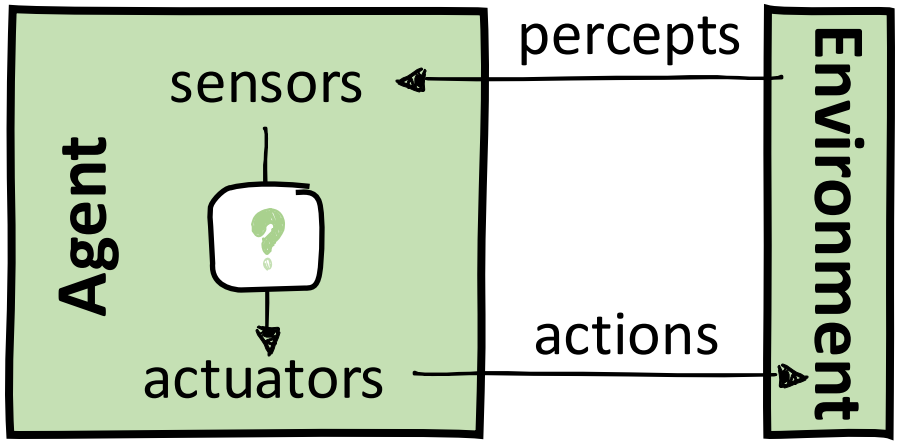
What is an intelligent agent?

What is this course?

Agent

An agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators.

Agent = architecture + program



Agent

Architecture

- Includes sensors and actuators.
- E.g., for a human agent
  - Sensors: eyes, ears, other organs
  - Actuators: hands, legs, mouth, other body parts
- E.g., for a robotic agent
  - Sensors: cameras, infrared range finders
  - Actuators: various motors

Program

- Runs internally on the physical architecture to produce the agent function $f: \mathcal{P}^* \rightarrow \mathcal{A}$
- $\mathcal{P}^*$: percept histories
- $\mathcal{A}$: actions.

Intelligent Agent

- A sufficiently complex rational agent can be viewed as an intelligent agent.

- Rational agent: For each possible percept sequence, a rational agent should select an action that is expected to maximize its performance measure (i.e., the utility), given the evidence provided by the percept sequence and whatever built-in knowledge the agent has.

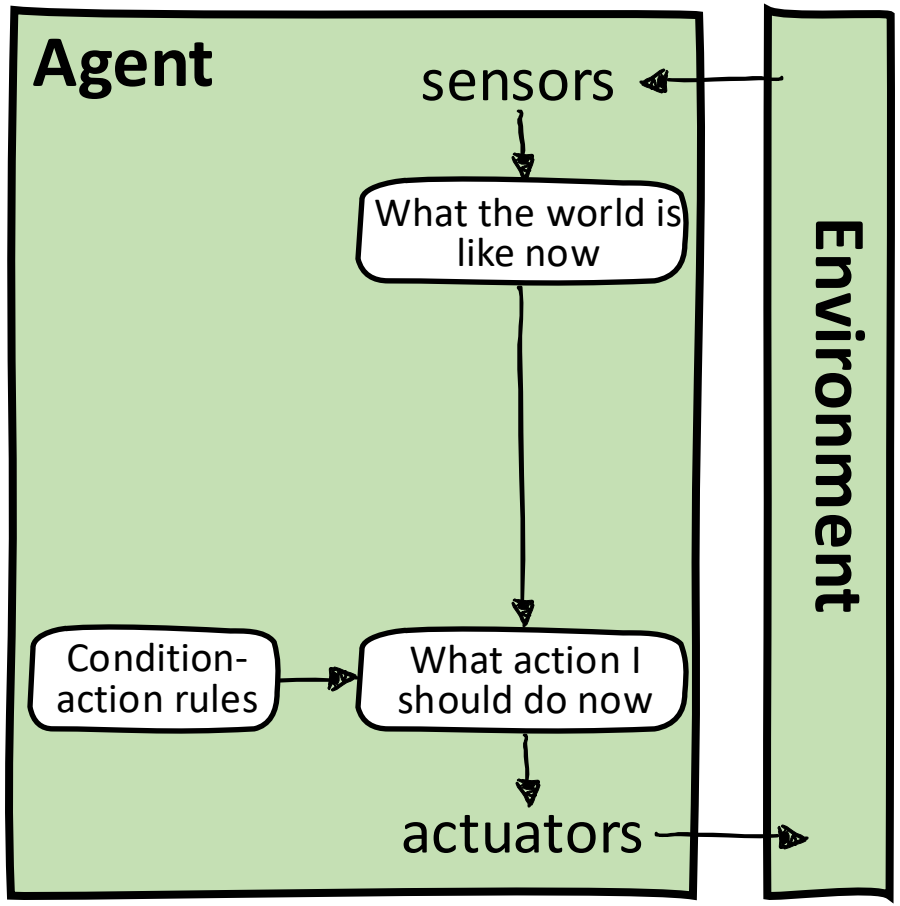
Exemples of Intelligent Agents

	performance measure	environment	actuators	sensors
 Self-driving car	safe, fast, legal, comfortable	roads, other traffic, pedestrians	steering wheel, brake, accelerator, signal, horn	cameras, GPS, LIDAR, engine sensors, speedometer, odometer
 AI microscopy	accurate diagnosis, fast, confident	patient, hospital	prediction, treatment suggestion	microscopy cameras

Types of Agents

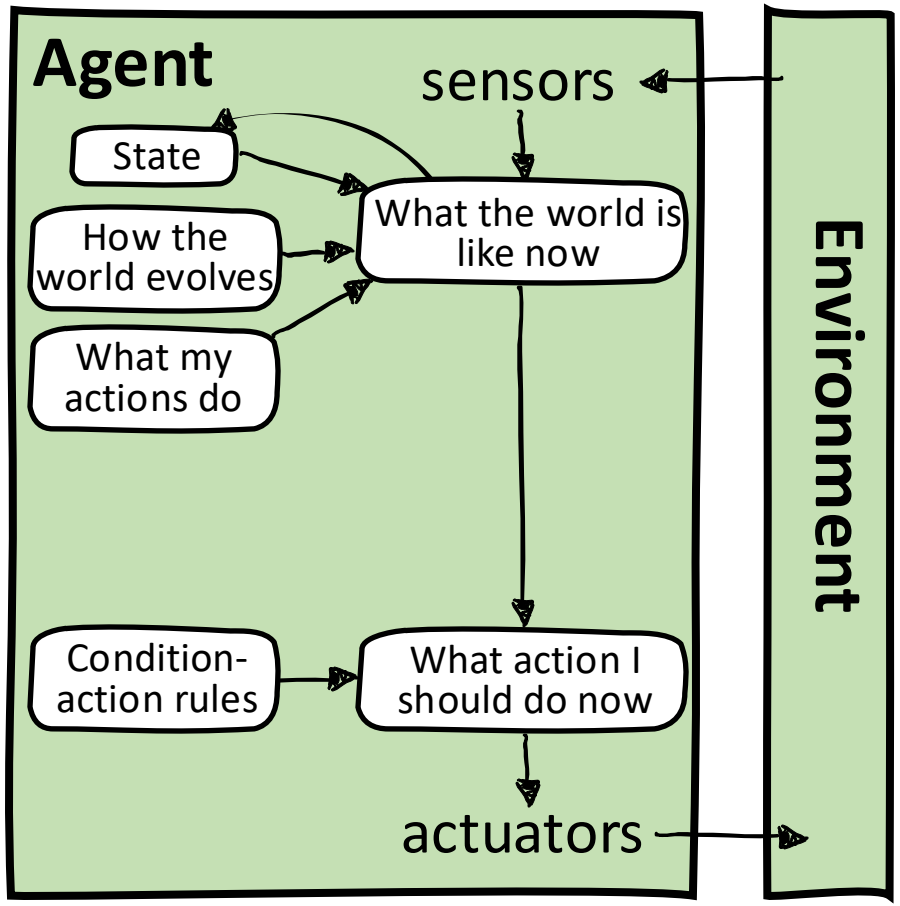
Simple Reflex Agents

- Use if-then rules to match certain percepts to an action.
  - E.g., if tail-light of car in front is red, then brake.
  - Use current percept only.
  - Generalize percepts via mapping them to the same action.
  - Adapt to changes in the environment by adding new rules.
- Does not track its percept histories.



Model-based Reflex Agents

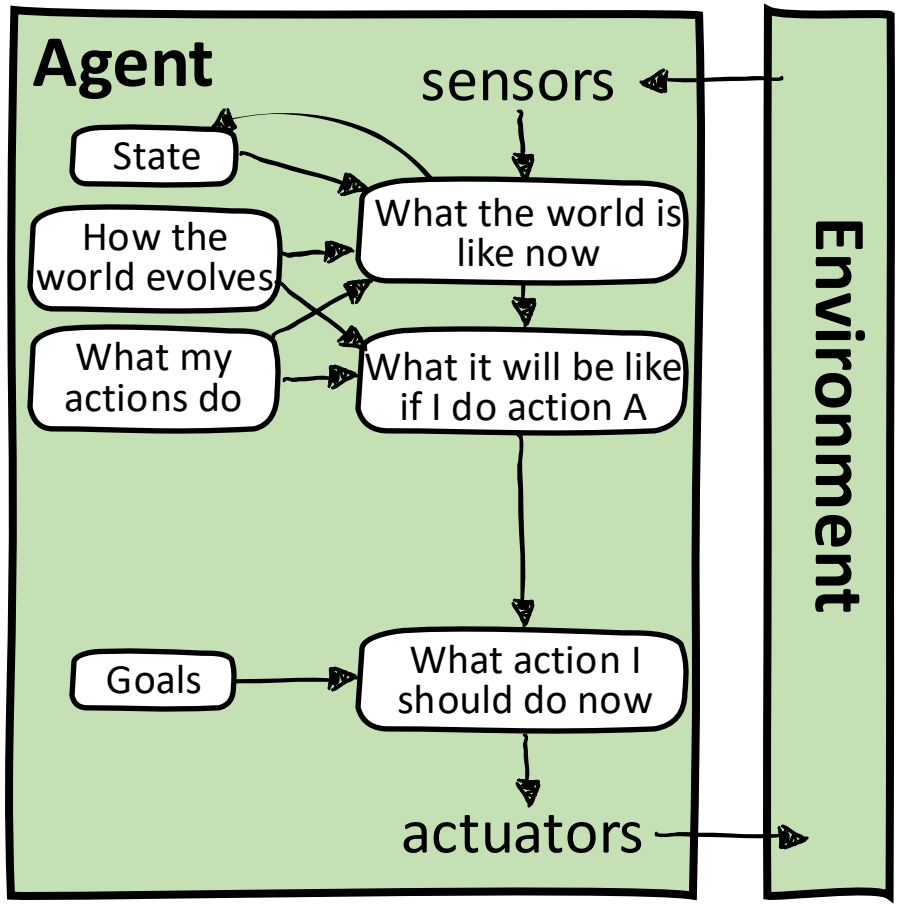
- Internally encode state of the world (via maintaining a model) based on past percepts and prior knowledge.
  - E.g., if dangerous driver in front, then keep distance (must infer potentially dangerous driver in front).
- Could just represent the current state.



Goal-based Agents

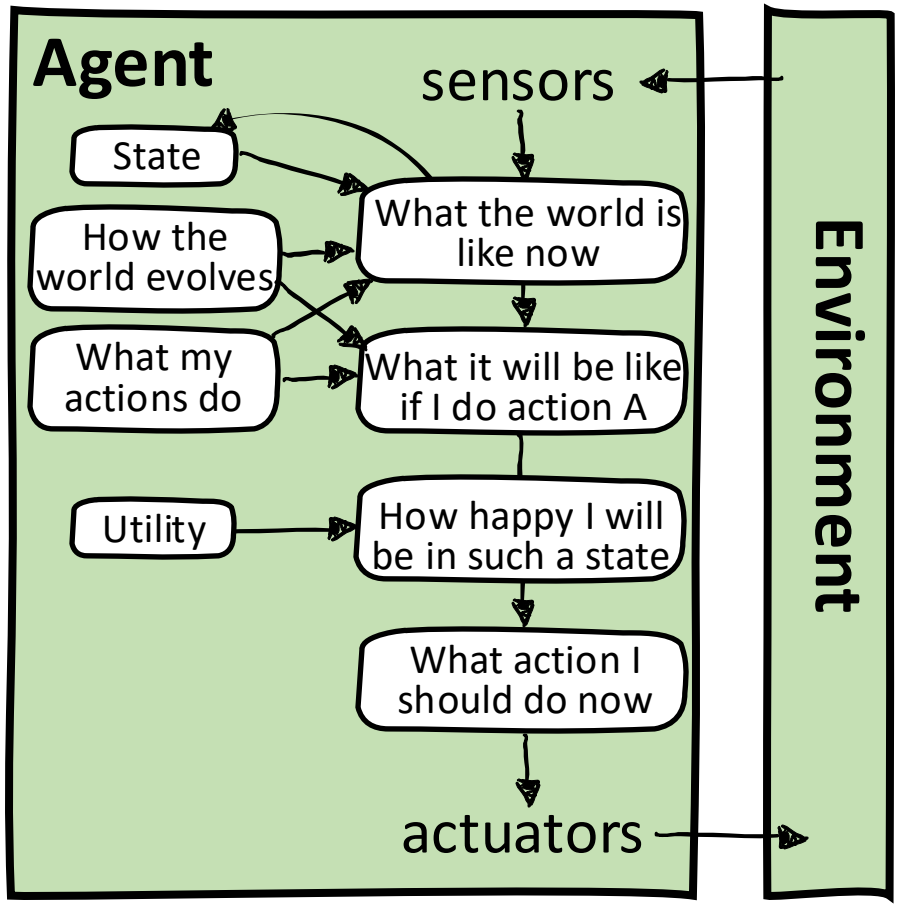
- Keep track of the state and a set of goals – choose actions to eventually reach given/computed goals.
  - E.g., go to a specific destination.
  - Goals help decide which actions are good.
  - Consider the future.

- Does not allow for degrees of goodness, especially when there are multiple possible alternatives.



Utility-based Agents

- Choose which alternative is best, according to the utility function.
  - E.g., go to the destination with the shortest path.
  - The utility function gives a quantitive measure of success or goodness at a give state.
  - Allow decisions comparing choice between conflicting goals, and choice between likelihood of success and importance of goal (if achievement is uncertain).



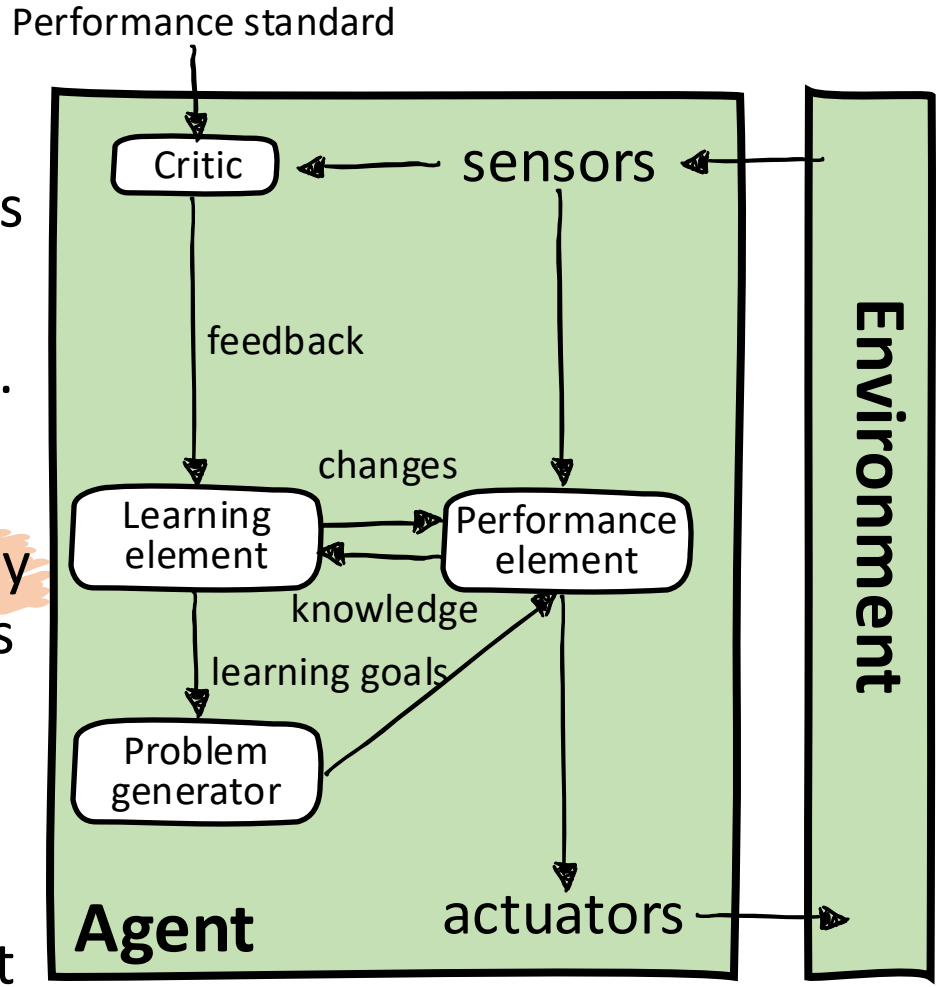
Types of Agents

How to program these agents?

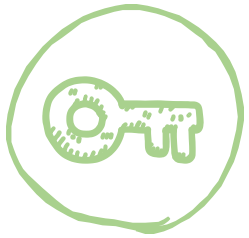
Manually coding the program is laborious and even impossible.

Learning Agents

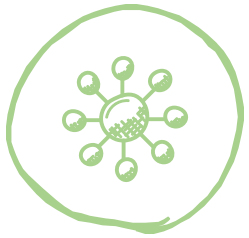
- Learn, adapt and improve over time
  - Feedback from the critic: how the agent is doing, e.g., quick turn is not safe.
  - Learning element: making improvements.
  - Changes: e.g., no quick turn.
  - Performance element: what we previously considered to be the entire agent; it takes in percepts and decides on actions.
  - Knowledge: e.g., road conditions etc.
  - Problem generator: suggesting actions that lead to new experiences, e.g., try out the brakes on different road surfaces.



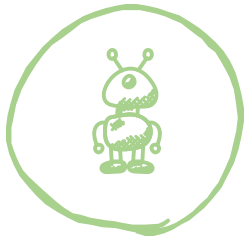
Today



What can AI do?



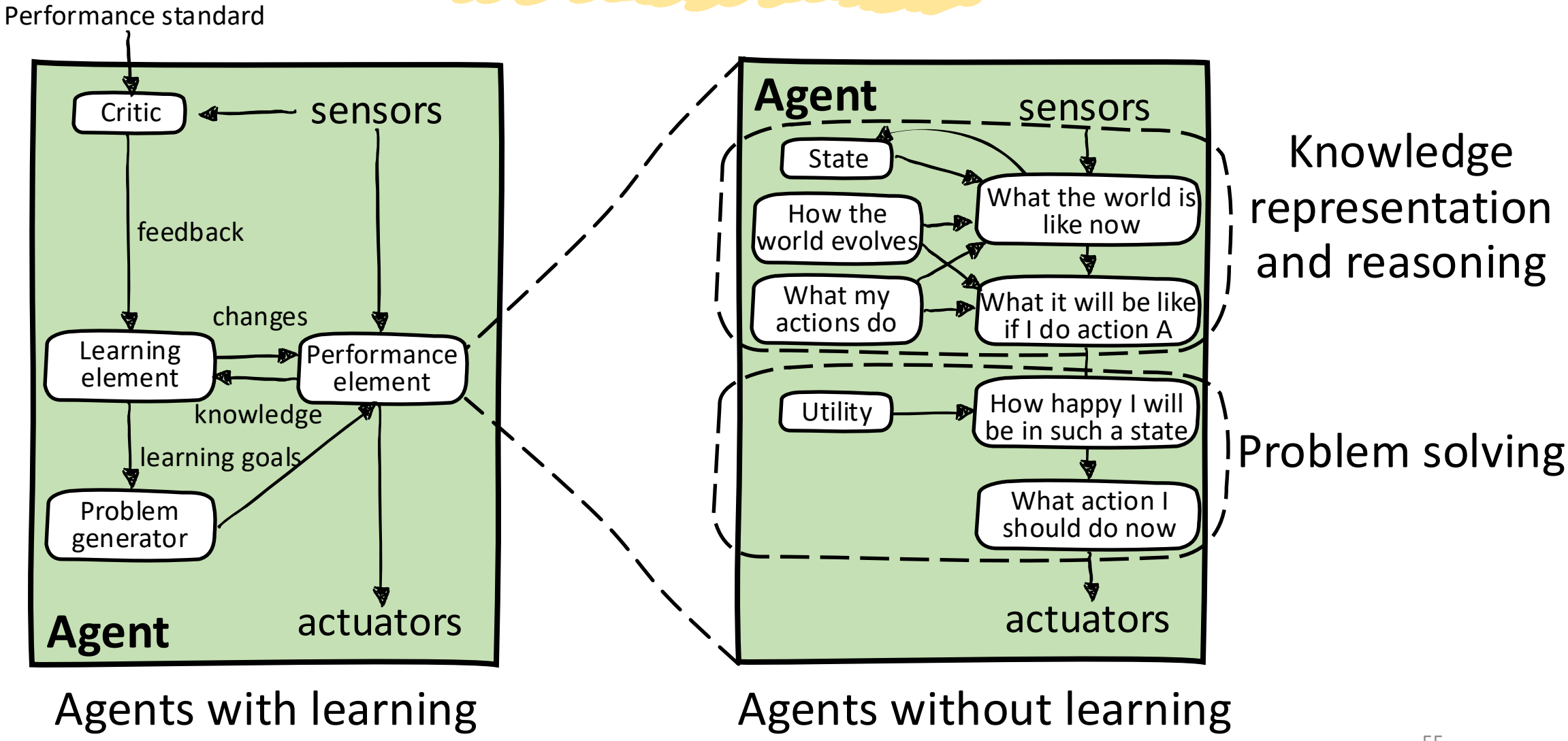
What is AI?



What is an intelligent agent?

What is this course?

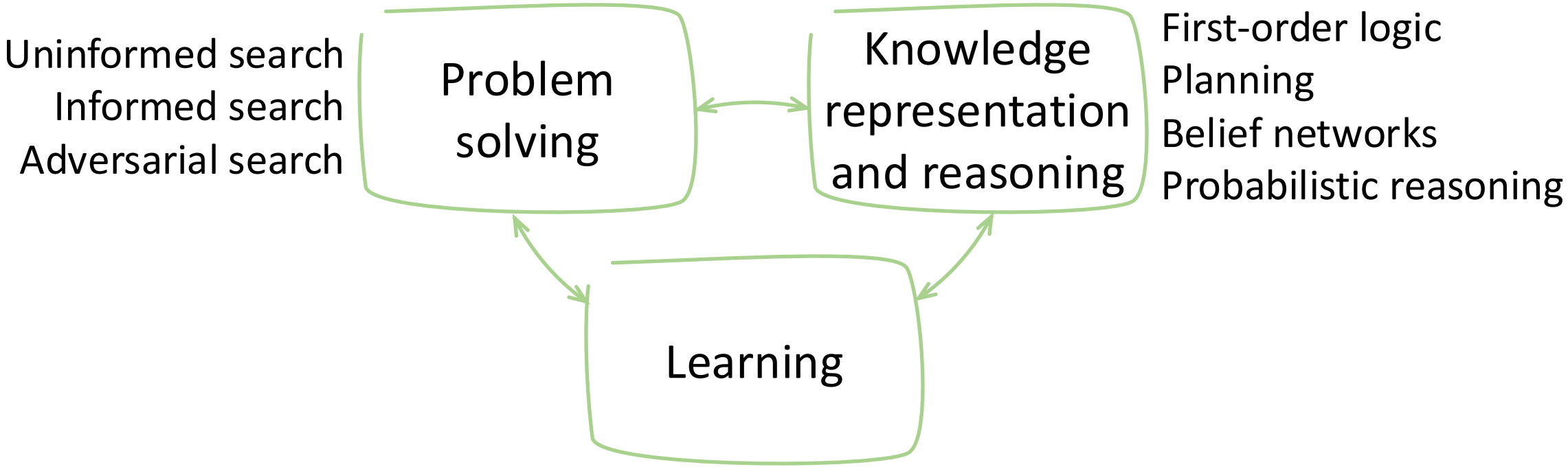
The Big Picture



Agents with learning

Agents without learning

The Big Picture



Goals

- To have fun.
- To teach you some ideas of AI.
- To be grounded despite the hype.
- To introduce you a set of key techniques and algorithms from AI.
- To get you thinking about how AI can be applied to a variety of real problems.
- To inspire your research.
- It is not about vision, natural language processing, machine learning, … it is an entry level course

Important This Week

- Read the page on the Canvas carefully.
- Python tutorial is out.