Machine Intelligence

1. A Brief Introduction to AI and Autonomous Agents

What is AI, Anyway?

Álvaro Torralba

AALBORG UNIVERSITET

Fall 2023

Course Organization

Course Homepage

Can be found under Moodle:
`https://www.moodle.aau.dk/course/view.php?id=48261`

Times

- **Usually** on Mondays at 8.15 but some of them in other days (mostly Wednesdays at 12:30) so...
  Check the calendar!
- **Usually** in Niels Jernes Vej 08A on Mondays, and Alfred Nobels Vej 27 on Wednesdays, but still...
  Check the calendar!
- Lectures consist of:
  - Classical Lecture: Fist half of the lecture (usually 8.15–10.00)
  - Exercises: Second half (in the group rooms): (usually 10.15–12.00).
- **Extended exercise sessions:** (4 sessions, usually Wednesdays 12.30–16.15 in the group rooms):

Exam

Written exam (as a Moodle Questionnaire) in January

Team

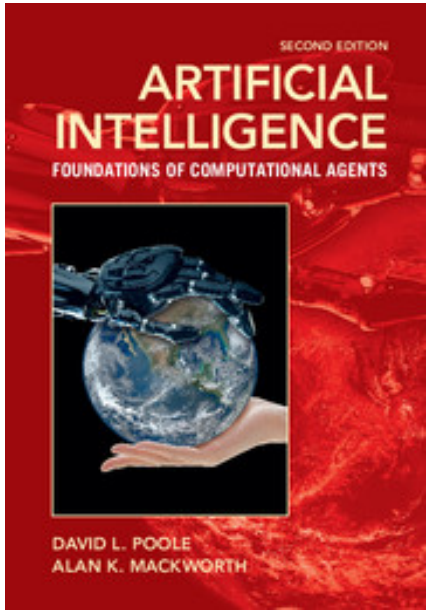
Teacher

Álvaro Torralba
e-mail: alto@cs.aau.dk

Tutors

- Mads Corfixen
- Thomas Heede
- Astrid Ipsen
- Raffaele Pojer

Literature



Main book:

- *D. Poole and A. Mackworth:* **Artificial Intelligence. Foundations of Computational Agents** (3rd edition).
- Web-version: `http://artint.info/`

Another reference book:

- *Stuart Russell and Peter Norvig* **Artificial Intelligence: A Modern Approach** (4th edition).
- ...but no need to buy it!

...but this book is not our "Bible":

- It's great to get intuitions on the basics of what we'll discuss.
- The book basically is "broad but shallow": An immense breadth of AI sub-areas is being covered. **Thus the book does not cover many important recent developments, and it often lacks formality.**

→ **The "ground truth" for this course are the post-handouts.**

(Some slides are just for reference, so don't stress about the slide number)

Course Material

Each chapter/lecture will have the following material:

- **Background** (Only in some chapters): Slides to read on your own before the lecture.
  →**Important**: Refresh basic concepts and notation that will be used in the lecture, so it is important to read and understand it before the lecture.
- Slides: three separate files:
  - **Pre-handouts**: The slides with some hidden parts (e.g. answers to questionnaires). Useful in case you want to take notes, but not really necessary.
  - **Slides**: Slides used in the lecture. Useful in case you want to see some "animation".
  - **Post-handouts**: Full slides, format to print.
    →**Important**: Use as study material after the lecture
- **Extra Reading**: Optional reading if you want to go beyond what we teach in the lecture.

Exercises and Questions

Exercises:

→Objective: Understand and apply concepts from the course.

→Solving the exercises is **very important for preparing for the exam**

→Advice: Solve the exercises **individually**, and then discuss the solution with your colleagues!

→To solve during the exercise sessions and afterwards

Additional resources that are optional, not exam relevant:

- Additional Material sections: Extra material that could be interesting to know about, but has been cut out of the course.
- Practical Exercises: Experience with AI modeling languages and tools.

Technical questions about course content/exercises:

- **Moodle Forum.** (Read by everybody)
- Also, of course: **tutors** during exercise sessions.
- **Other questions: Álvaro Torralba.**
  - Come to the front directly after the lecture.
  - Or alto@cs.aau.dk .

Our Agenda for This Chapter

- **AI?** What does this term even mean?
  → **Spoiler: Nobody knows.**
- **AI History** and **AI Today:**
  → **Brief research field overview.**
- **Intelligent Agents**
  → **What are they and where I can find one?**
- **Representation**
  → **How to define our problem?**
- **Environments**
  → **What the world looks like?**

What is Intelligence?

Nobody knows:

- **Being good at maths? (what else?)**
- **Being good at Chess or Go? (what else?)**
- **Ability to think? (what does this mean?)**
- **Ability to learn? (what does this mean?)**
- **Creativity? (what does this mean?)**
- **Passing an IQ test with high marks? (go away!)**

$\rightarrow$ This question has been debated in Philosophy since centuries ...

What's Artificial Intelligence?: Your thoughts

Mimicking Human Intelligence

- AI is a **simulation of human intelligence** in a computer.
- **Making a machine think like humans**, humans being the intelligence and machines thinking being the artificial part
- AGI an actually thinking machine, that can search for info like LLMs but unlike those, **it can actually reason like a human** or a dog.
- Advanced algorithms that **mimicks the human feature of intelligence**.
- Machines that show **human-like intelligence traits**.
- A machine that **can think human-like thoughts**.
- Artificial Intelligence is where one **attempts to mimic human like intelligence** with the use of computers.
- Artificial intelligence is a way for a computer to **try and simulate the human brain** in problem solving.

→Are humans really the pinacle of intelligence? Definitively not in chess!
→Do Airplanes fly in the same way birds do?

What's Artificial Intelligence?: Your thoughts

AI = Machine Learning/Neural Networks

- Some software which is able to improve it's algorithms when it receives more data **(it is able to "learn")**
- A program or algorithm **that is capable of learning** and improving and refining itself based on its inputs and computations.
- Artificial Intelligence is a machine **that can learn from data** it gets as input - that is it is the part of dataology that has to do with machines that can "think". It can learn supervised or unsupervised. Some AI can learn how to handle new situations by searching the data it already has available, if the new situation is similar to already familiar situations.
- Machines **learn from data**.
- Something that has 'intelligence' i.e. **it is able to learn on its own.**
- A computer program **able to 'learn'** and get better by itself
- Artificial Intelligence is an intelligence based upon **a model that has been trained on data** (and the more data, the more intelligent).

- Artificial Intelligence is **a neural network**, that learns from a large data set.
- AI makes computers able to make decisions based on complex functions. It **imitates a brain by using neural networks**. . . .

→Neural Networks are a very successful technique within AI. But definitively not the only one! AI is much broader!
→Learning is not the only form of intelligence! We also have reasoning and problem solving, for example!

What's Artificial Intelligence?: Your thoughts

AI = Data Analysis

- ...**"based on patterns in datasets"**...
- "Artificial intelligence is when you **give the computer a dataset** and some parameters and based on these the computer will make assumptions"
- "A computer or program who can think on its own by **using recognition and previous data**?"
- "... some kind of fake "human" intelect that basically **is a bunch of data** that can be used by a computer when the data is processed."
- "... **Uses models that are trained on data.** ..."
- "AI is using a **computer model based on data to make decisions** without help from a person."
- "Artifical Intelligence is a decision making entity which can decide what to do **given prior training through data sets**."
- "AI is **programs trained on large datasets** to recognise, predict or otherwise utilize the data to propose a answer based on information. AI is to create a machine which mimic intelligence."
- "Artificial Intelligence is the use of algorithms to **find and interpret patterns in the data** and give an outcome (an action or result) out of it."

→Learning from datasets is even a more limited view of AI. For example, you can also learn from interacting with the world!

What's Artificial Intelligence?: Your thoughts

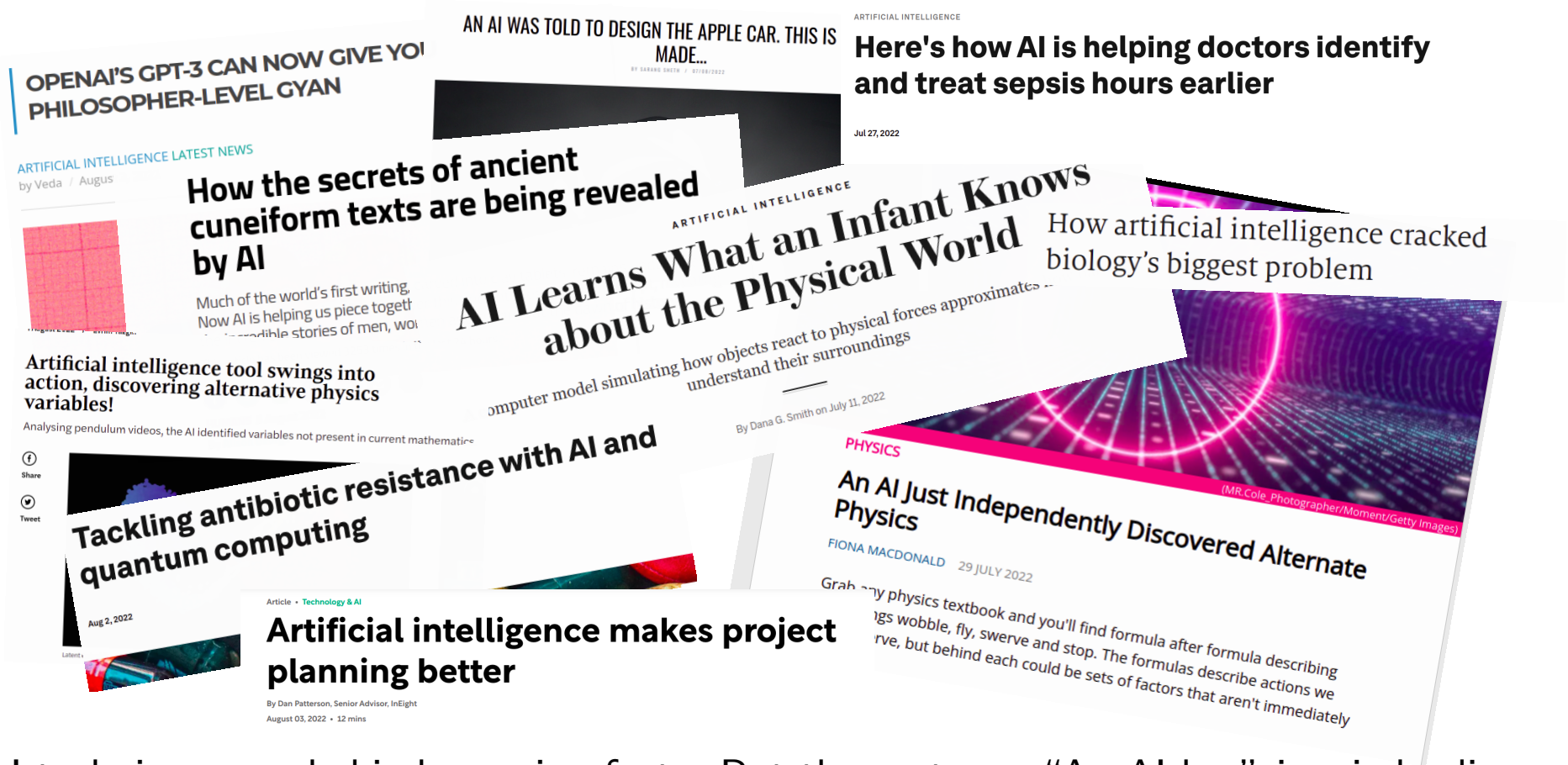
No Human Interaction

- "...is the capability of software or hardware to partially or entirely operate on its own **without needing of any human interception** ..."
- "... learn about and solve problems using **low or no human involvement ...**"
- "AI is using a computer model based on data to make decisions **without help from a person**."

→Social Intelligence, Explainable AI and Human-Computer Interaction are part (or at least related to) AI too!

What is Artifical Intelligence?

Let's ask the news:

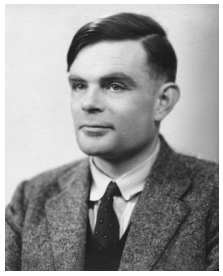


AI techniques are behind amazing feats. But the sentence "An AI has" is missleading.
→A highly-skilled team has developed an application using AI techniques/tools

- Very significant implementation effort!
- + Large computational overhead

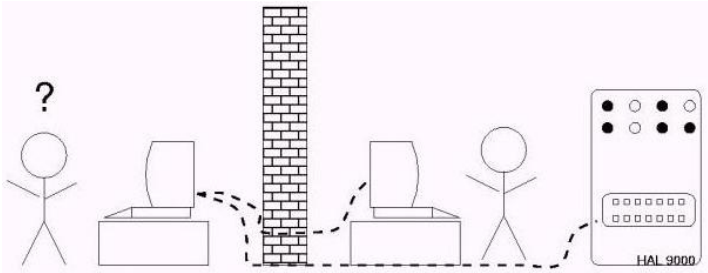
What is *Artificial* Intelligence?

The Turing Test



A.M. Turing: "Computing Machinery and Intelligence", Mind Vol.59 (1950) proposes an *imitation game*, which (slightly modified) has become known as the *Turing test*:
"I PROPOSE to consider the question, 'Can machines think?"'

An *interrogator* is connected via one terminal to a real person, and by another terminal to a computer (both in another room). The interrogator does not know which terminal is connected to the machine. He or she can perform on both terminals a (natural language) dialogue with whatever is at the other end of the line. The machine passes the Turing test, if the interrogator is not able to identify, which terminal is connected to the machine.



→The Turing test tests observable behavior, not cognitive processes!

Towards the Turing Test ...

Loebner Prize

- (Non-scientific) competition until 2020 for computer systems performing under Turing Test conditions.
- Recent serial winner: **Kuki (formerly Mitsuku)**

Missconceptions/Limitations of Turing Test

- It is not enough to chit/chat. People are easy to trick (see fake news).
  →Q: I have K at my K1, and no other pieces. You have only K at K6 and R at R1. It is your move. What do you play? A: (After a pause of 15 seconds) R-R8 mate.
- Human's are not the pinacle of intelligence! →What if we can distinguish the machine because it is too intelligent? (e.g. too good at chess)

Winograd scheme

- Scientific alternative to test "common sense"

The city councilmen refused the demonstrators a permit because **they** feared**advocated** violence.

1. **The city councilmen**
2. **The demonstrators**

- "Solving Winograd schemas is not a surrogate for the ability to do commonsense reasoning, let alone for intelligence." Kocijan *et al.* (2022)

What is Artifical Intelligence?

Achievements: Deep Blue

- 30 IBM RS/6000 processors
- 480 custom chess processors
- able to examine 200 million moves per second
- database of 700.000 grandmaster games
- endgame database (covering all 5 piece positions)

Results:

1996:	Kasparov 4	Deep Blue 2
1997:	Kasparov 2.5	Deep Blue 3.5

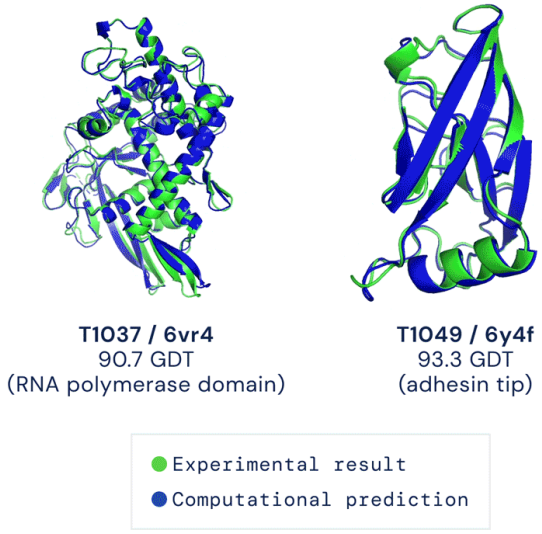
What is AI?

... two decades later (Google Deepmind)

In 2016 AlphaGo played a 5-game match against Lee Sedol with a score of 4-1 in favor of AlphaGo.

In 2017, AlphaZero defeats(**) world-champion chess program after 24 hours of training!(*)

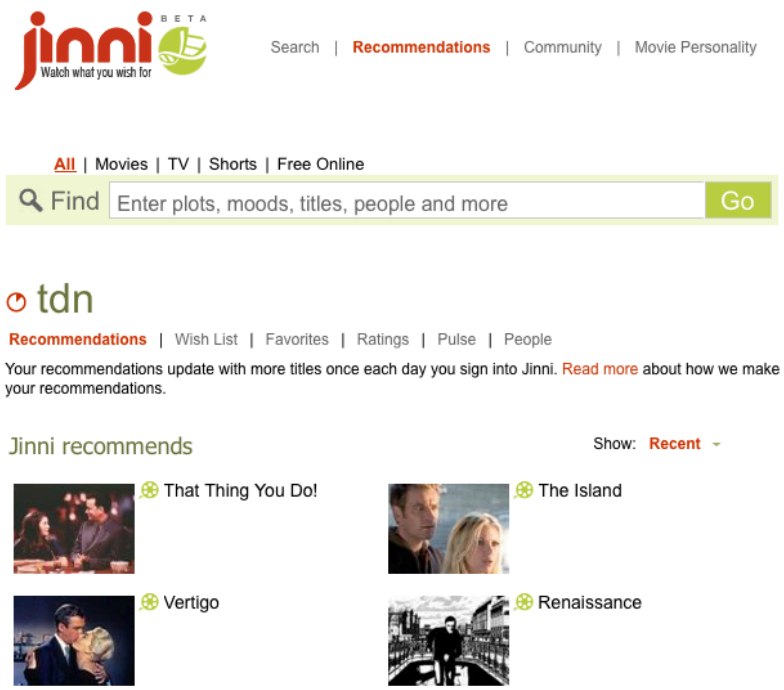
...and in 2020 AlphaFold starts to tackle the protein folding problem.



(*) in a huge cluster with specialized hardware.
(**) unclear if conditions were "fair".

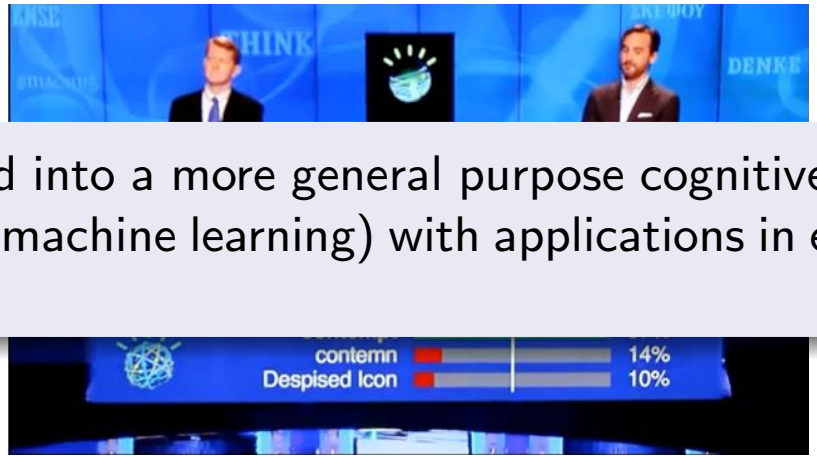
Recommender systems

Systems for recommending items that users are likely to find interesting



Jeopardy!: Watson

In 2011, Watson beat Brad Rutter, the biggest all-time money winner on Jeopardy!, and Ken Jennings, the record holder for the longest championship streak (75 days)



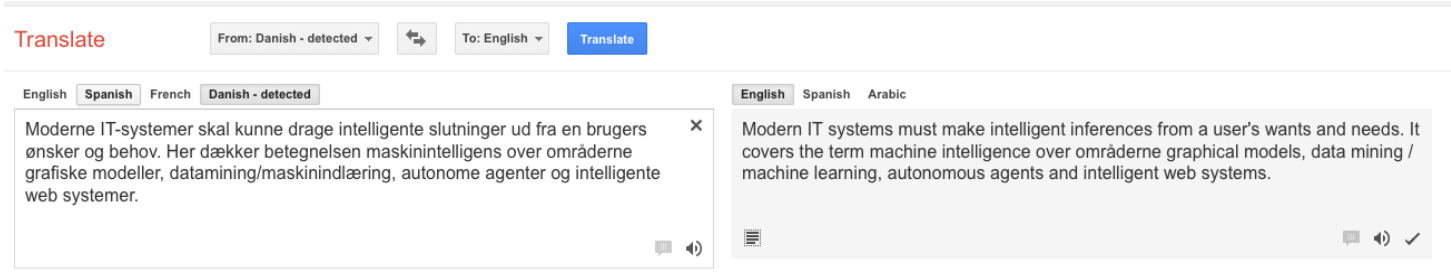
Watson has since evolved into a more general purpose cognitive platform (using natural language processing and machine learning) with applications in e.g. health care, cooking, weather forecast, ...

- Application of natural language processing, information retrieval, knowledge representation and reasoning, and machine learning
- Made up of a cluster of 90 IBM Power 750 servers with a total of 2880 POWER7 processor cores and 16 Terabytes of RAM. Each Power 750 server uses a 3.5 GHz POWER7 eight core processor, with four threads per core.

Automatic Translation

Achievements: Automatic Translation

Google translate:



General language models

GPT3: Language model from openai

Title: United Methodists Agree to Historic Split
Subtitle: Those who oppose gay marriage will form their own denomination
Article: **After two days of intense debate, the United Methodist Church has agreed to a historic split - one that is expected to end in the creation of a new denomination, one that will be "theologically and socially conservative," according to The Washington Post. The majority of delegates attending the church's annual General Conference in May voted to strengthen a ban on the ordination of LGBTQ clergy and to write new rules that will "discipline" clergy who officiate at same-sex weddings. But those who opposed these measures have a new plan: They say they will form a separate denomination by 2020, calling their church the Christian Methodist denomination.**

Part of the GPT-3 generated news article that humans had the greatest difficulty distinguishing from a human written article (accuracy: 12%).

(Brown et al., Language Models are Few-Shot Learners, 2020)

Autonomous driving

Achievement: DARPA grand challenge 2005

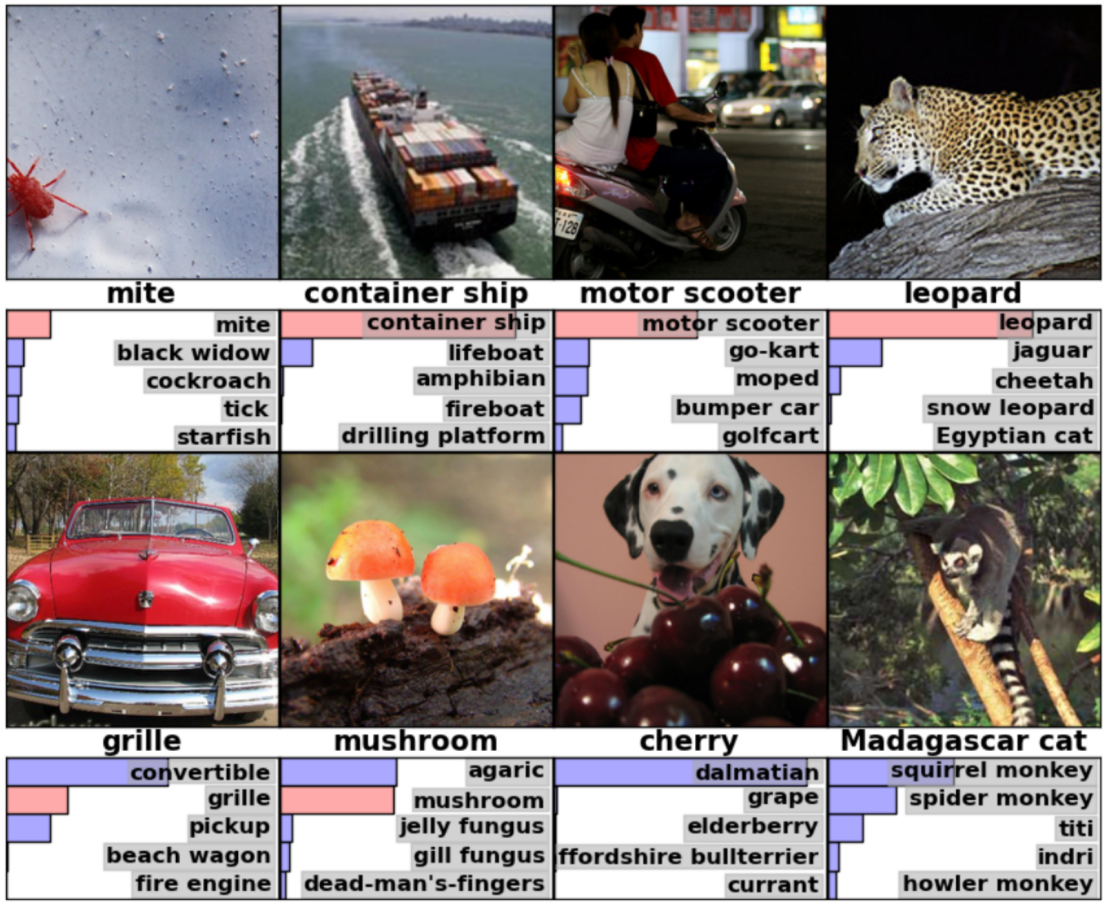
Competition for autonomous vehicles: navigate 132 miles through desert terrain (route
mpleted the task.
).

... commercial self-driving taxi service in Phoenix, Arizona.

nera pair, 1 monocular
d.

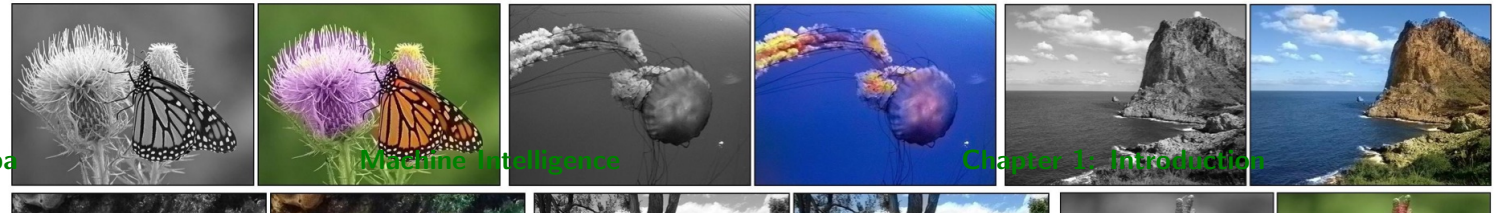
Image analysis/processing

Image classification



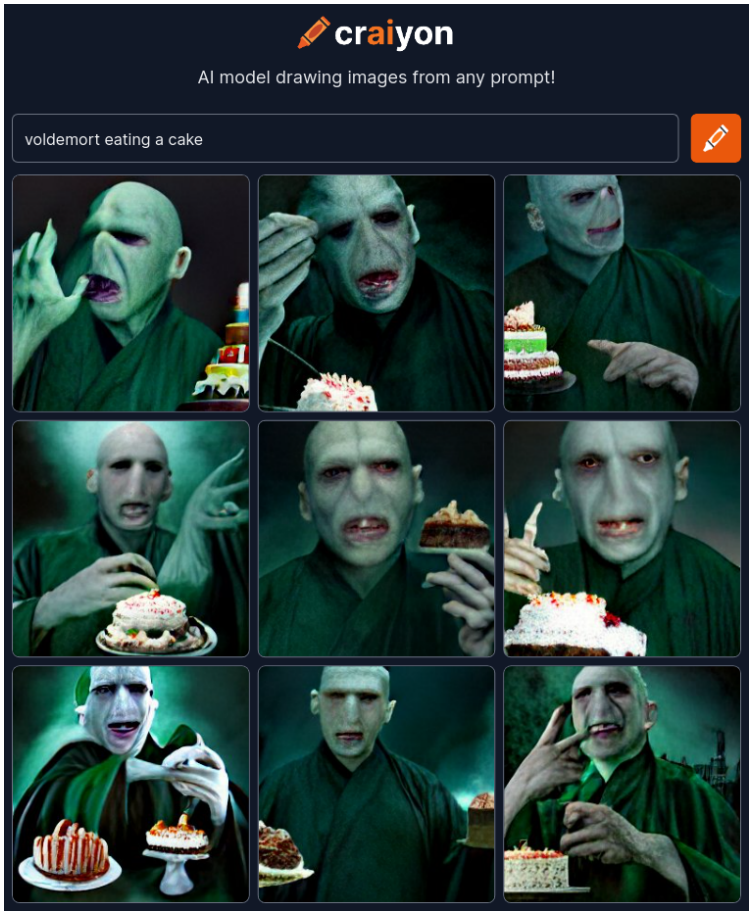
Krizhevsky et al., ImageNet Classification with Deep ConvolutionalNeural Networks, NIPS 2012

Colorization of images



Creation of images

Creation of Images from Natural Language



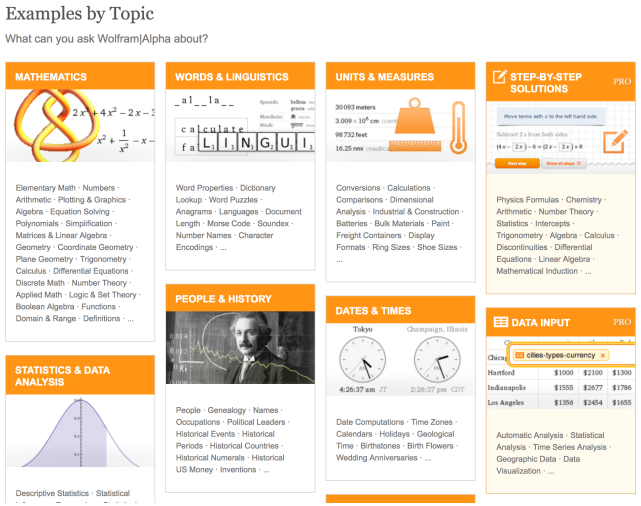
Voldemort eating a cake

Voldemort cake

Wolfram Alpha

Wolfram Alpha

"Think about that for a minute. It computes the answers. Wolfram Alpha doesn't simply contain huge amounts of manually entered pairs of questions and answers, nor does it search for answers in a database of facts. Instead, it understands and then computes answers to certain kinds of questions." [Nova Spivack on `twine.com`]



Some queries to try:

- What is the highest mountain in the united states?
- What is the highest mountain in america?
- What is the highest mountain in south america?
- What is the highest mountain in california?
- What is Mt. Whitney?
- What is the temperature in Aalborg?
- Is the temperature in Aalborg higher than 30 degrees?
- What is the prime decomposition of 74936?
- What are the prime factors of 74936?
- What is the largest prime factor of 74936?

Other application areas

- Medical diagnosis and advisory systems
- Information processing and filtering,
- Display of information for time-critical decisions
- Spam filtering
- Optical character recognition
- Profiling/Credit scoring: profiling customers
- Bioinformatics
- Real estate: Prediction of house prices
- Computer networks: intrusion detection
- Alert and monitoring systems
- Speech recognition
- Face recognition, image annotation
- Action recognition (in video sequences)
- ...

So... back to the question: What is *Artificial* Intelligence?

Quoting the PM book: "Artificial intelligence, or AI, is the field that studies the synthesis and analysis of computational agents that act intelligently."

- An agent is judged solely by how it acts. Agents that have the same effect in the world are equally good.
- Intelligence is a matter of degree. The aspects that go into an agent acting intelligently include:
  - what it does is appropriate for its circumstances, its goals, and its perceptual and computational limitations
  - it takes into account the short-term and long-term consequences of its actions, including the effects on society and the environment
  - it learns from experience
  - it is flexible to changing environments and changing goals.

So... back to the question: What is *Artificial* Intelligence?

Take 2: Let's try to be systematic here ...

	Humanly	**Rationally**
Thinking	Cognitive Science Neural Networks? Certainly not yet!	Logics Machine Learning
Acting	Turing Test	**APPLICATIONS**

→ Note: Thinking is (sometimes?) a prerequisite for acting ... logics and machine learning are motivated by, and very useful in, applications!

The Four Categories: Summary

Acting Humanly: **Turing Test**. Not much pursued otherwise.

$\approx$ Aeronautics: "Machines that fly so exactly like pigeons that they can even fool other pigeons".

Thinking Humanly: **Cognitive Science**. How do humans think, how does the human brain work.

$\rightarrow$ Neural networks are an (extremely simple, so far) approximation.

Thinking Rationally: **Logics** (formalization of knowledge and deduction). **Machine Learning (ML)** (mathematical formulation of learning).

Acting Rationally: **How to make good action choices?**

$\rightarrow$ Is what we're interested in, in practice. Encompasses logics and ML (in particular neural networks) as methods to take rational decisions.

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**:

"We propose that a 2-month, 10-man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.".

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

70's: How to scale from microworlds to real applications? $\rightarrow$ **Knowledge-based systems**, knowledge provided by humans.

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

The History of AI, ctd.

Late 80's: Expert systems prove less promising than imagined (difficult to update/maintain, cannot learn, brittle). $\rightarrow$ **"AI Winter"**.

90's–00's: **Formalization of AI techniques and increased use of mathematics** in the field. Quote from [Russell and Norvig (1995)]:

"A better understanding of the problems and their complexity properties, combined with increased mathematical sophistication, has led to workable research agendas and robust methods."

10's: **Re-advent of neural networks (NN)**.

NN have decade-old roots. Almost forgotten in the 90's and 00's. "Sudden" success in image classification end of 00's, way better than human-coded rules. Since then, **rapid successes and hype**. "Deep" NN = several layers (how many? next question please).

Enablers: advanced NN architectures; lots of data; hardware.

20's: Computers rule the world? Next AI Winter?

AI Today: Sub Areas of Artificial Intelligence

Human-AI Interaction

Explainability

Ethics/ Thrustworthy AI

AI Today: Sub-Areas and Tentative Course Schedule

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).

→ **Chapter 2**

Reasoning

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

→ **Chapter 3**

KR: Knowledge representation and reasoning (logic and deduction).

→ **Chapters 3 and 4**

Uncertainty: Reasoning about uncertain knowledge.

→ **Chapters 4-5**

AI Today: Sub-Areas and Tentative Course Schedule, ctd.

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

Learning

ML: Machine Learning: How to learn from experience?

→ **Chapters 7- 10,14**

Decision-Making

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

→ **Chapters 11-13**

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

→ **Chapter 12**

Others

Robotics: How to control/design robots?

→ Not considered here.

Vision: How to interprete/analyze camera inout?

→ Not considered here (separate Courses).

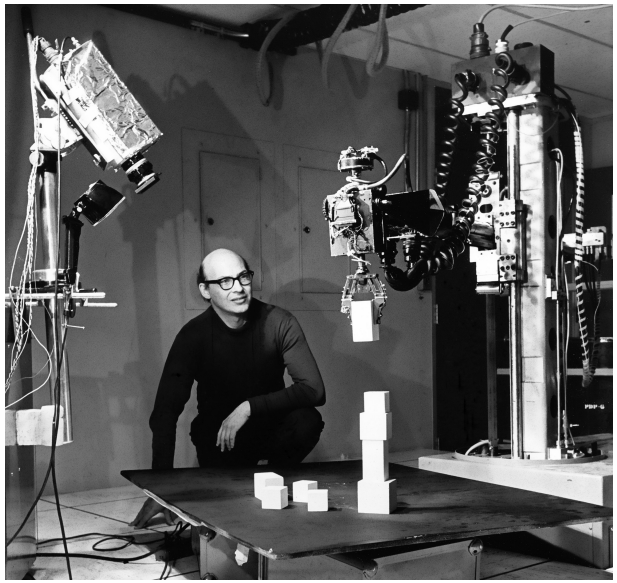
→ **Intimate relations to many other areas of CS. Logic Programming, Databases, Verification, Game Theory, . . .**

Symbolic vs. Subsymbolic AI

1956 Dartmouth College Summer Workshop

- "Complex Computer Applications"
- They coin the term: "Artificial Intelligence"
- Two ways of thinking:

John McCarthy "Standford School"
$\rightarrow$ thinking rationally
"Symbolic"



Marvin L. Minsky "MIT School"
$\rightarrow$ thinking humanly
"Subsymbolic"

Symbolic vs. Subsymbolic AI

Symbolic representations

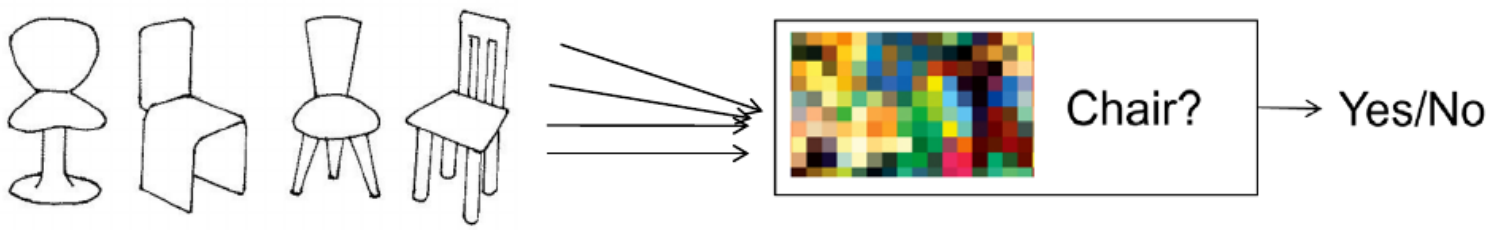
A Chair

- is a portable object
- has a horizontal surface at a suitable height for sitting
- has a vertical surface suitably positioned for leaning against

Find a definition

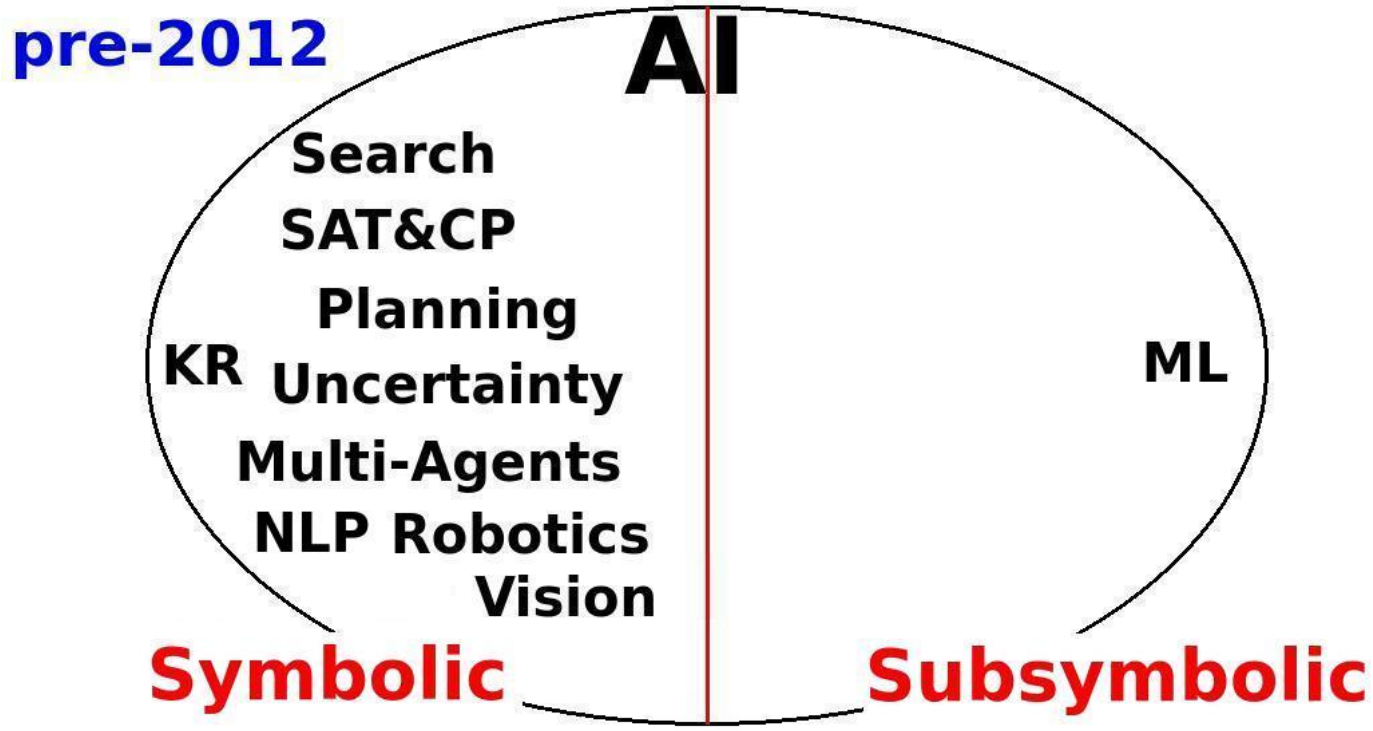
- using symbols, concepts, rules, some formalism
- apply automated reasoning procedures

Subsymbolic representations



- Use many different (arbitrary) features to describe the object
  - low-level inputs bits, encoding of neurons
  - show examples to the system and let it learn a generalization pattern
- If the patter is correct, the system has "learned" the concept without using an explicit definition.

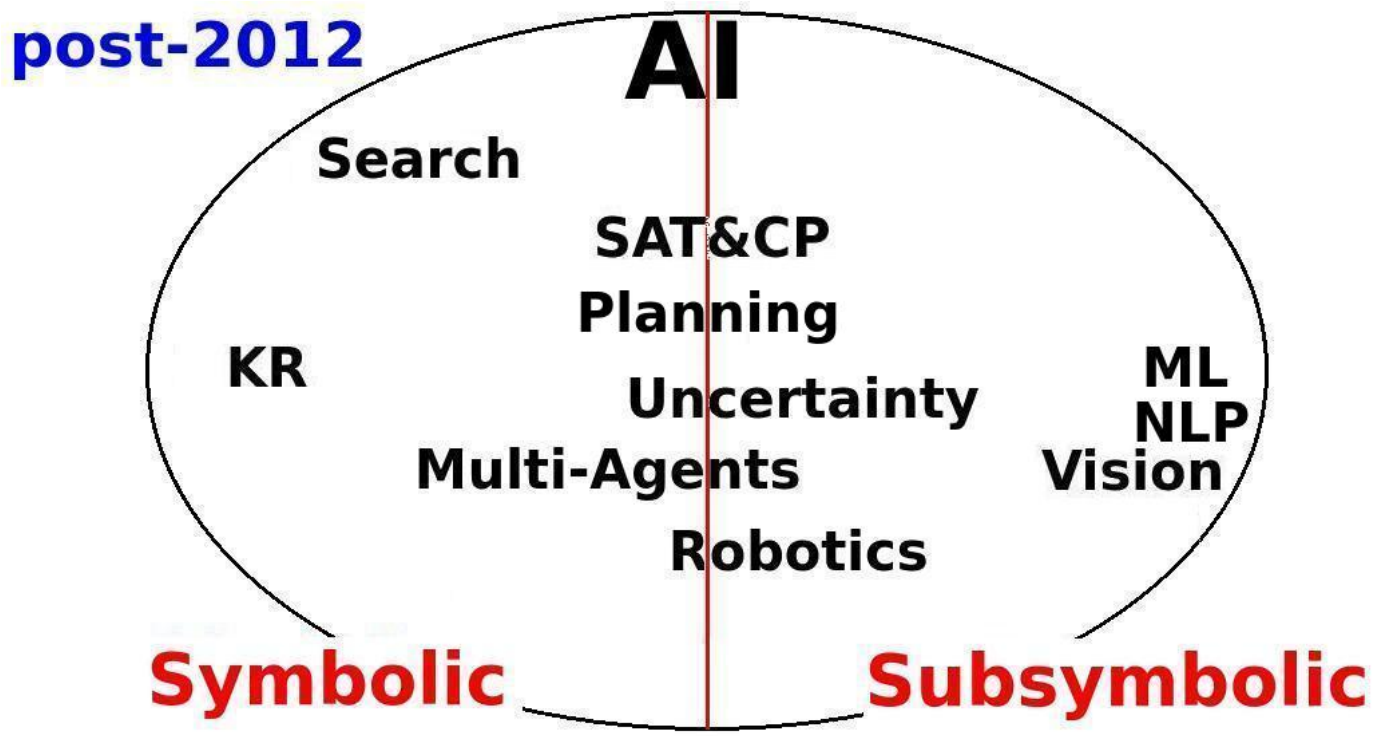
Symbolic vs. Subsymbolic AI



Symbolic: Conceptual, human-readable, formalization (model) of world behavior

Subsymbolic: Fitting of function parameters to data (world behavior observations)

Symbolic vs. Subsymbolic AI



Symbolic: **Conceptual, human-readable, formalization (model) of world behavior**

- Pros: instant performance, verifiability, explainability
- Cons: modeling can be costly or impossible, complexity of reasoning

Subsymbolic: **Fitting of function parameters to data (world behavior observations)**

- Pros: highly performant, able to tackle problems elusive for conceptual modeling
- Cons: learning curve, opaque, hyperparameters difficult to set

→ **How to combine the two?**

Questionnaire

Question!

How many scientific articles were submitted to the 2022 International Joint Conference on Artificial Intelligence (IJCAI) in Vienna?

(A): ca. 100 **(B): ca. 1000**

(C): ca. 2000 **(D): ca. 4000**

→ Answer (D) is correct (4225).

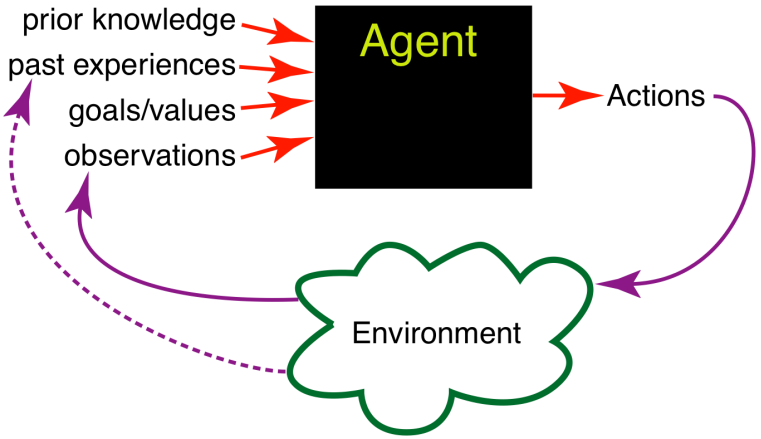
Questionnaires:

- At end of section/at start of **break**.
- You get 2/5/7/10 minutes.
- You're free to make noise (e.g., discuss with your neighbors).

What is an "Agent" in AI?

AI is the field that studies the synthesis and analysis of computational agents that act intelligently. [PM, p.3]

A coupling of perception, reasoning, and acting comprises an agent. [PM, p.10]



Agents:

- Perceive the environment through **sensors** ($\rightarrow$ **percepts**).
- Act upon the environment through **actuators** ($\rightarrow$ **actions**).

$\rightarrow$ **Examples?** Humans, animals, robots, software agents (softbots), ...

Some special flavors:

- Autonomous agents
- Intelligent agents
- Software agents
- Multi-agent systems

What is *not* an agent?

- perception $\sim$ input
- reasoning $\sim$ computation
- acting $\sim$ output
- $\rightsquigarrow$ "agent" a design metaphor, not a strict technical concept

Rational Agents . . .

. . . do “the right thing”!

→ **Meaning of “do the right thing”:** Rational agents select their actions so as to maximize a **performance measure**.

→ **What’s the performance measure of an autonomous vacuum cleaner?**

- m^2 per hour.
- Level of cleanliness.
- Energy usage.
- Noise level.
- Safety (behavior towards hamsters/small children).

→ But what if the vacuum cleaner’s sensors are not good enough to recognize the difference between a hamster and a shoe?

Actually, Rational Agents . . .

. . . ATTEMPT to do "the right thing"!

→ The hypothetical best case ("the right thing") is often unattainable.

→ The agent might not be able to perceive all relevant information. (Is there dirt under this bed? Is this a hamster or a shoe?)

Rationality vs. Omniscience:

- An **omniscient agent** knows everything about the enviroment, and knows the actual effects of its actions.
- A rational agent just makes the best of what it has at its disposal, *maximizing expected performance given its percepts and knowledge*.

→ **Example?** I check the traffic before crossing the street. As I cross, I am hit by a meteorite. Was I lacking rationality?

So, What *Is* a Rational Agent?

Mapping your input to the best possible output:

Performance measure M $\times$ **Percepts** P $\times$ **Knowledge** K $\rightarrow$ **Action** a

- An **agent** has a performance measure M and a set A of possible actions. Given a percept sequence P, as well as knowledge K about the world, it selects an action $a \in A$.
- The action a is **optimal** if it maximizes the expected value of M, given the evidence provided by P and K. The agent is **rational** if it always chooses an optimal a.

$\rightarrow$ If the vacuum cleaner bumps into the Hamster, then this can be rational in case the percept does not allow to recognize the hamster.

$\rightarrow$ Note: If **observation actions** are required, they are elements of A, i.e., the agent must *perceive actively*. **Example:** *"truck-approaching"*$\notin P$ *but I didn't look to check* $\Longrightarrow$ *I am NOT being rational!*

So, a Rational Agent is an Optimal Action Choice Function?

Practical limitations:

- Our definition captures limitations on percepts and knowledge.
- It does not capture computational limitations (often, determining an optimal choice would take too much time/memory).

  $\rightarrow$ In practice, we often *approximate* the rational decision: **bounded rationality**.

Examples of Agents: PEAS Descriptions

Agent Type	Performance Measure	Environment	Actuators	Sensors
Chess/Go player	win/lose/draw	game board	moves	board position
Medical diagnosis system	accuracy of diagnosis	patient, staff	display questions, diagnoses	keyboard entry of symptoms
Part-picking robot	percentage of parts in correct bins	conveyor belt with parts, bins	jointed arm and hand	camera, joint angle sensors
Refinery controller	purity, yield, safety	refinery, operators	valves pumps, heaters displays	temperature, pressure, chemical sensors
Interactive English tutor	student's score on test	set of students, testing agency	display exercises, suggestions, corrections	keyboard entry

Domain-Specific vs. General Agents

vs.

Solver specific to a particular problem ("domain").	**vs.**	Solver based on *description* in a general problem-description language (e.g., the rules of any board game).
More efficient.	**vs.**	More *intelligent*.

Questionnaire

Question!

Which are agents?

(A): James Bond.

(B): Your dog.

(C): Vacuum cleaner.

(D): Thermometer.

Question!

Who is rational?

(A): James Bond, crossing the street without looking.

(B): Your dog, crossing the street without looking.

(C): Vacuum cleaner, deciding to clean under your bed.

(D): Thermostat, deciding to cool down your fridge.

Questionnaire Answers

First Question: Which are agents?

→ (A) and (B): Definite yes.

→ (C): Yes, if it's an autonomous vacuum cleaner. Else, no.

→ (D): No, because it cannot do anything. (Changing the displayed temperature value could be considered an "action", but that is not the intended usage of the term.)

Second Question: Who is rational?

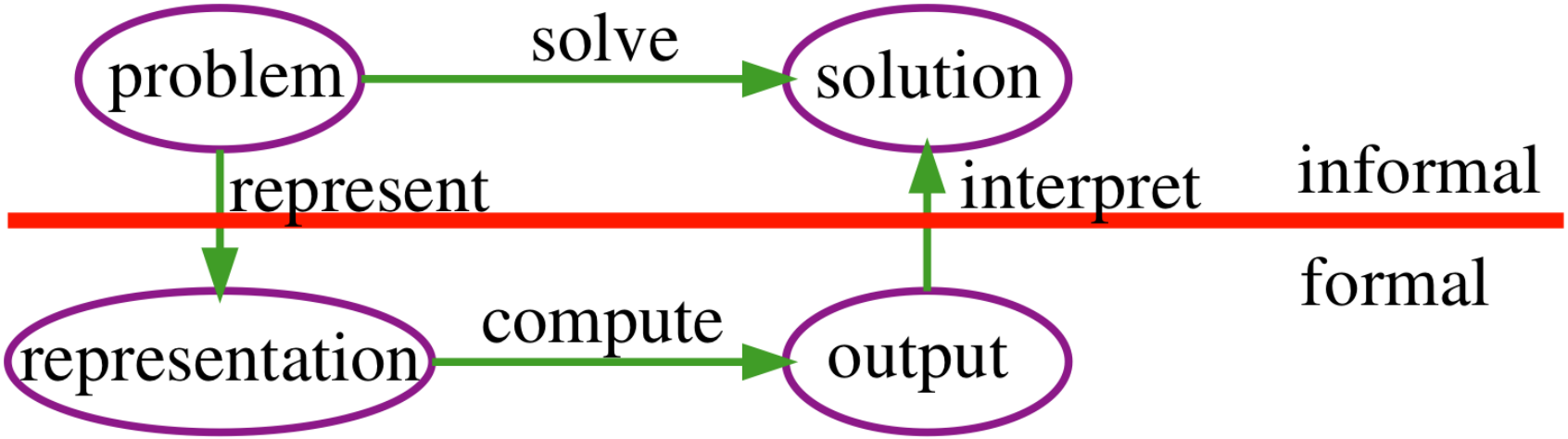
→ (A): Depends on whether safety is part of his performance measure.

→ (B): Depends on whether or not we consider dogs to be able to check the traffic. If they can't, then just running over could be optimal (e.g. to meet fellow dogs or grab a sausage).

→ (C): Yes. (Hypothetical best-case if it's dirty under your bed, and you're not currently sleeping in it.)

→ (D): Not clear whether a thermostat is an agent. On the one hand, in difference to the Thermometer, the Thermostat takes an action. On the other hand, in a classical Thermostat the "action decision" is just a physical reaction (like a solar panel that produces electricity if the sun shines).

Represent and Compute



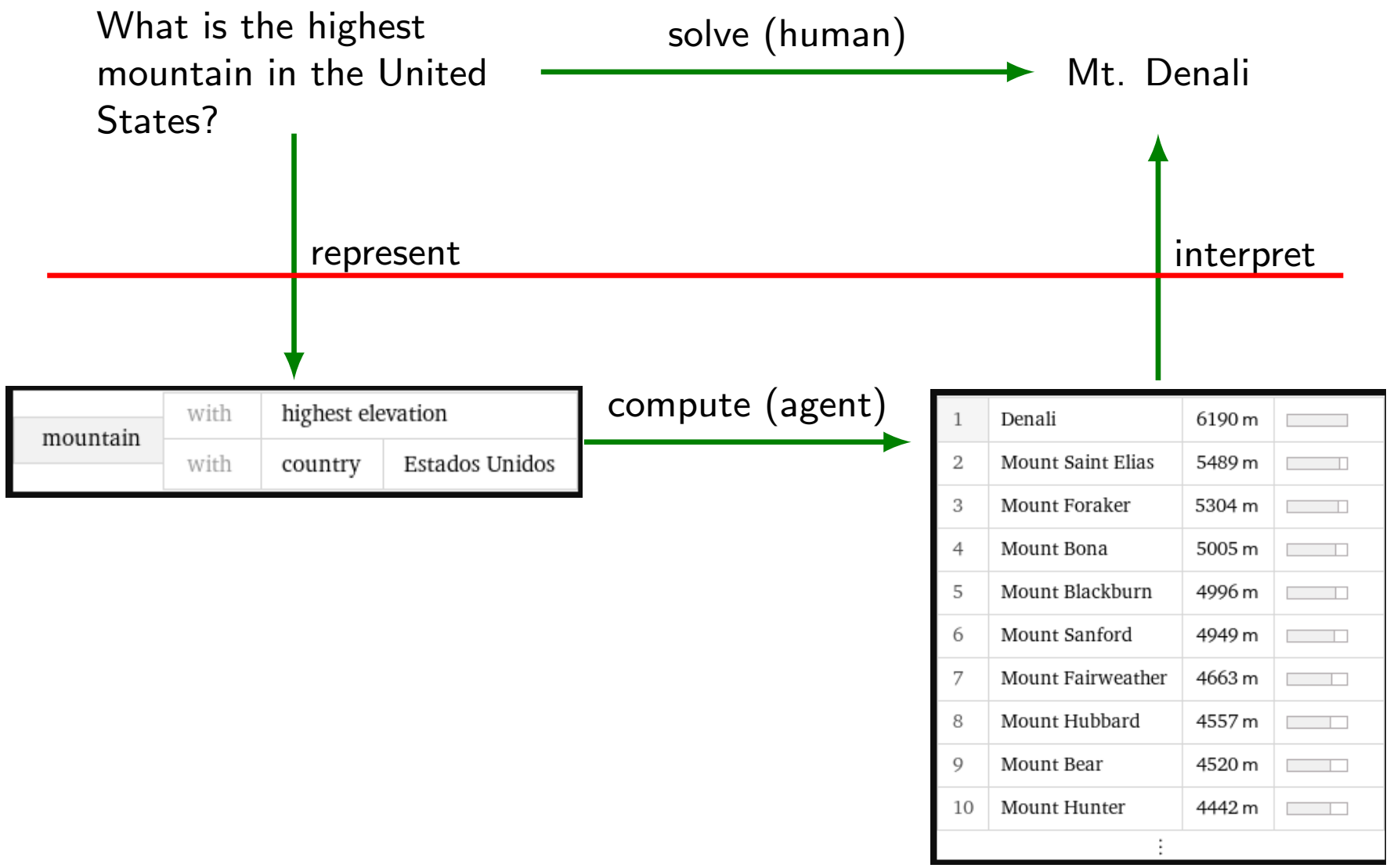
A *representation* should be:

- Sufficiently rich to encode the required knowledge.
- Be "close" to the problem.
- Amenable to efficient computation.
- Able to be acquired from people, data, or experience.

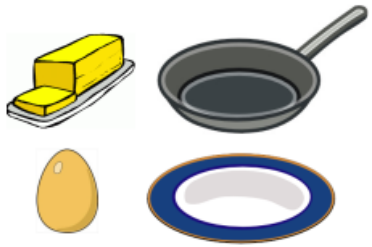
Questions to be considered:

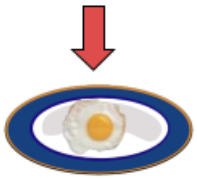
- What is a solution and how good should it be (optimal, satisficing, approximately, probable)?
- How can the problem be represented?
- How can an output be computed (what properties should a solution have)?

Represent and Compute



Represent and Compute





solve (human)

- Heat Butter in Pan
- Break Egg into Pan
- Fry at moderate heat
- Put egg on plate

represent

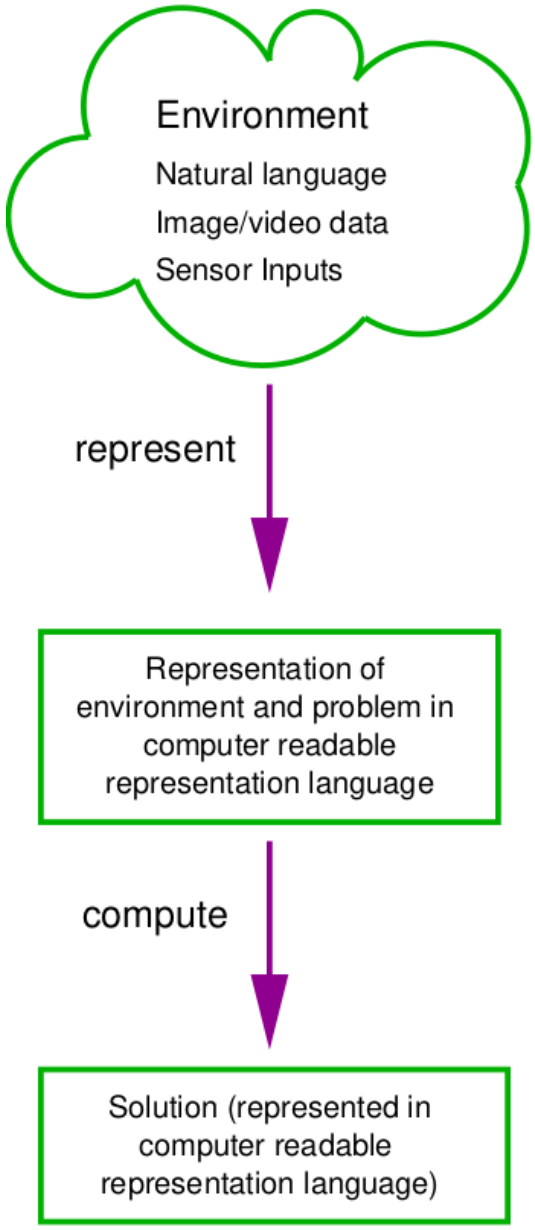
interpret

01110100 00100000 01101001 01110011
01101110 01101001 01101110 01100111
01100001 01111001 00101110 00100000
01110101 01110011 01110100 00100000

compute (agent)

01001001 01110100 00100000 01101001 01110011 00100000 01110010
01101001 01101110 01101001 01101110 01100111 00100000 01110100
01100100 01100001 01111001 00101110 00100000 01001001 01110100
01101101 01110101 01110011 01110100 00100000 01100010 01100101
01110111 01101001 01101110 01110100 01100101 01110010 00100000
01101110 00100000 01010110 01100001 01101110 01100011 01101111

Problem Formalization



1. How to interpret input/observations from environment?

Natural Language Processing
Computer Vision
...

2. What formal representation languages can be used?

Knowledge Representation

3. How to solve problems in the given representation language?

Problem Solving
Automated Reasoning

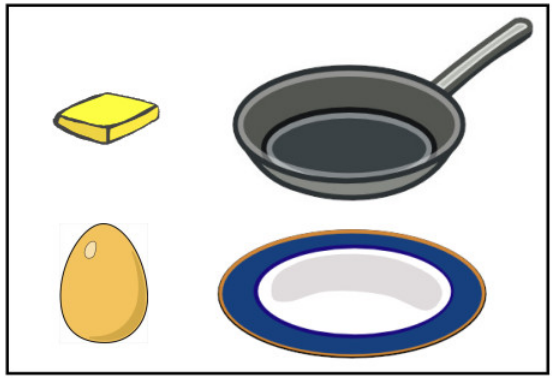
In this course the focus is on 2. and 3.!

Levels of Representation

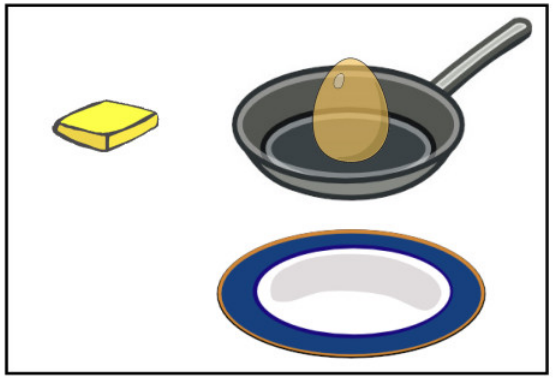
We consider 3 representation schemes

- State based
- Feature based
- Relational

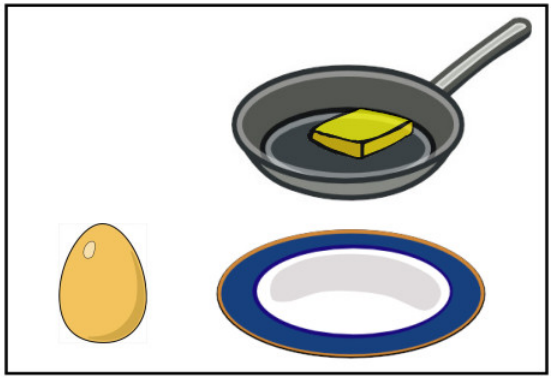
State based



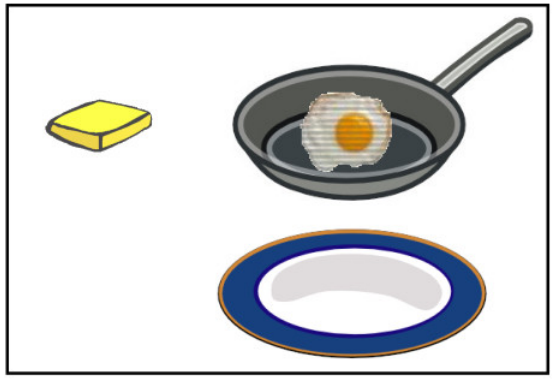
State 01



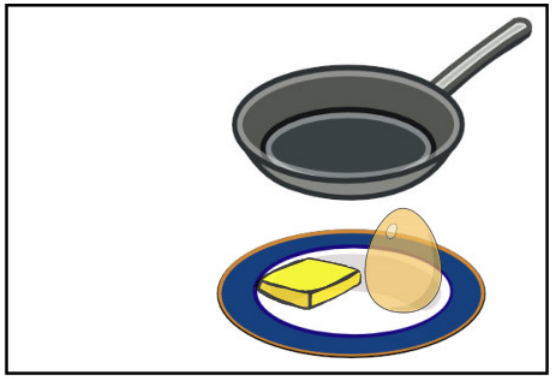
State 04



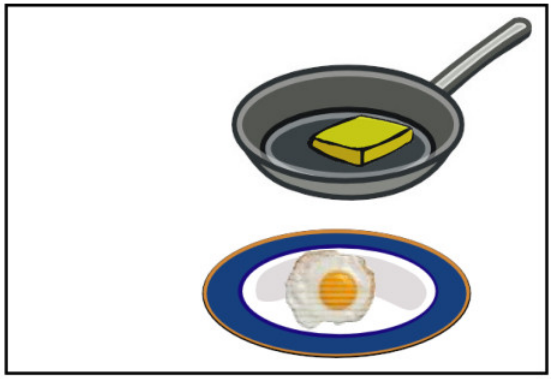
State 08



State 12

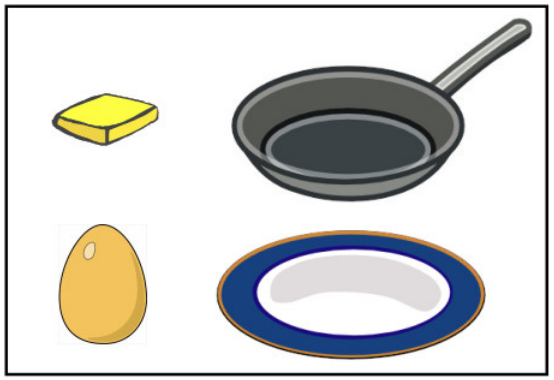


State 14

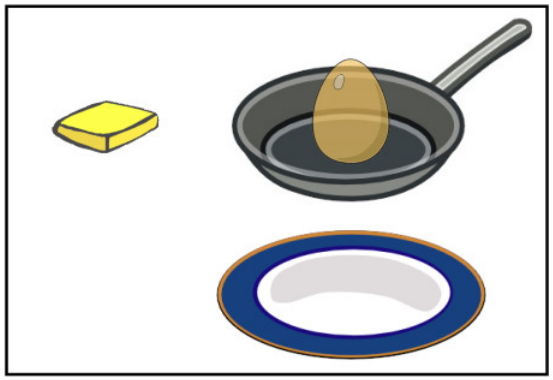


State 18

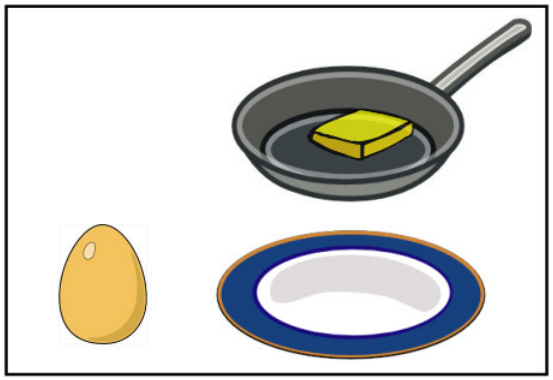
Feature based



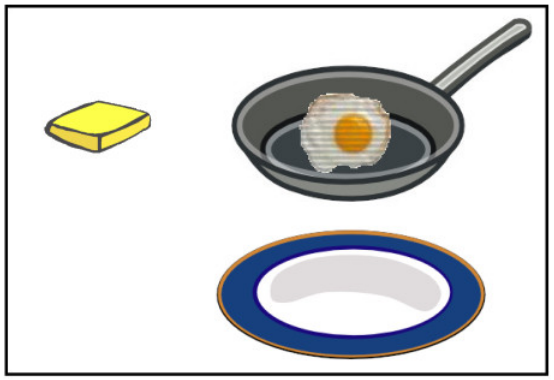
egg=whole,
butter_in=table,
egg_in=table



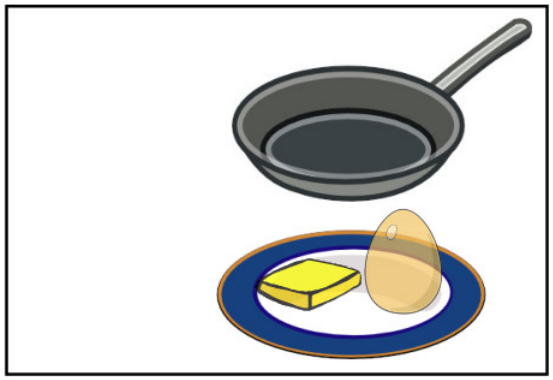
egg=whole,
butter_in=table,
egg_in=pan



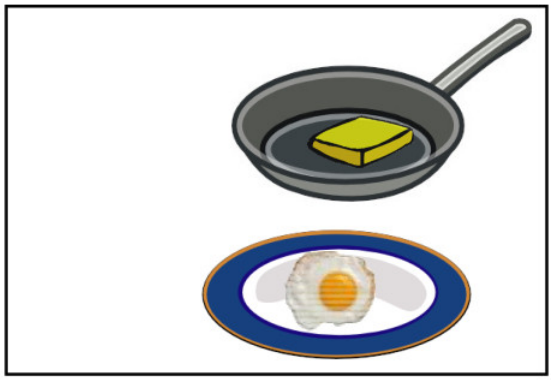
egg=whole,
butter_in=pan,
egg_in=table



egg=broken,
butter_in=table,
egg_in=pan



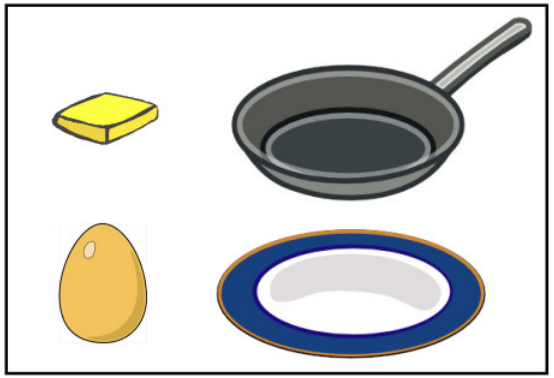
egg=whole,
butter_in=plate,
egg_in=plate



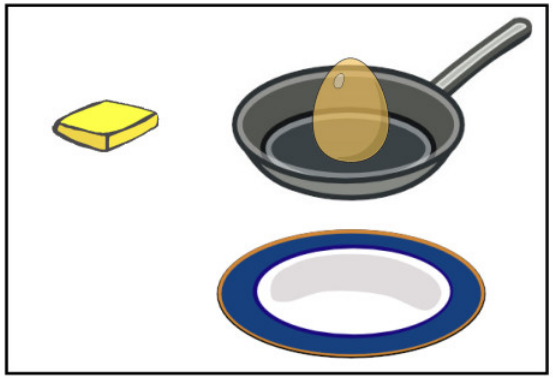
egg=broken,
butter_in=pan,
egg_in=plate

30 binary features represent $2^{30} = 1.073.741.824$ states.

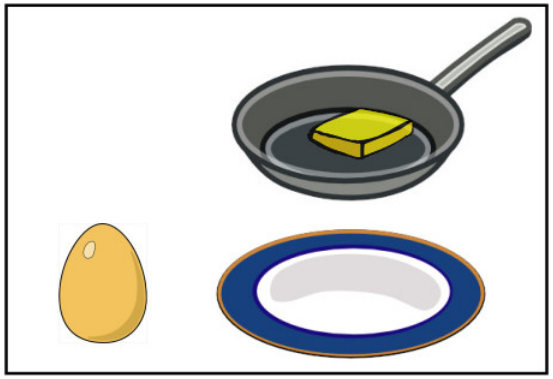
Relational



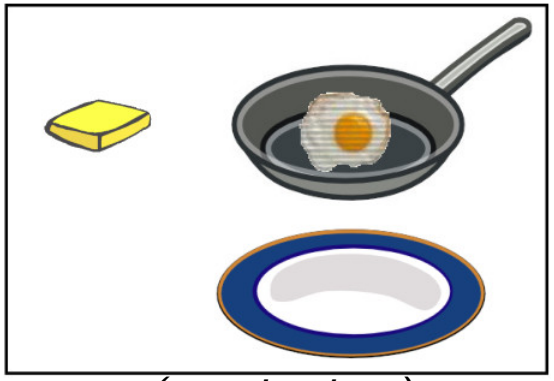
state(egg,whole), in(butter,table), in(egg,table)



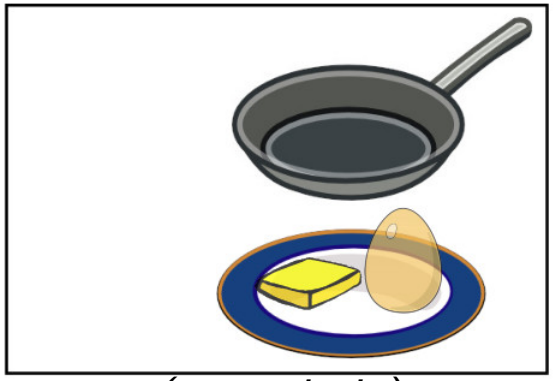
state(egg,whole), in(butter,table), in(egg,pan)



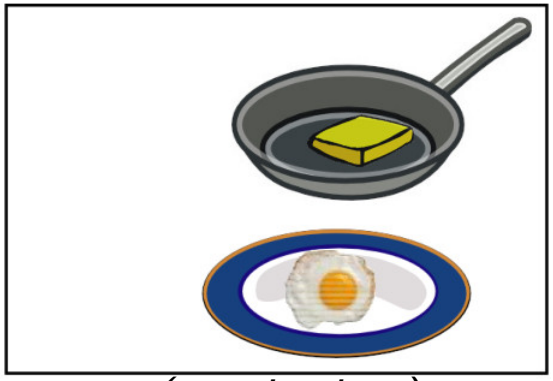
state(egg,whole), in(butter,pan), in(egg,table)



state(egg,broken), in(butter,table), in(egg,pan)



state(egg,whole), in(butter,plate), in(egg,plate)

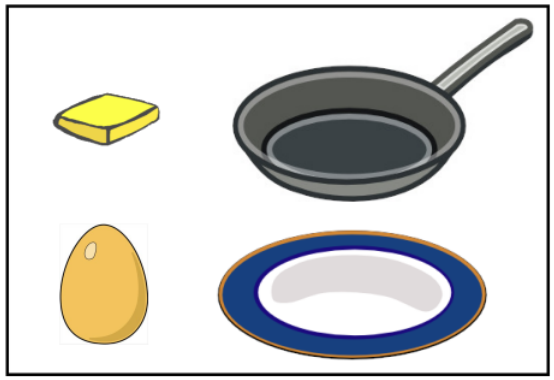


state(egg,broken), in(butter,pan), in(egg,plate)

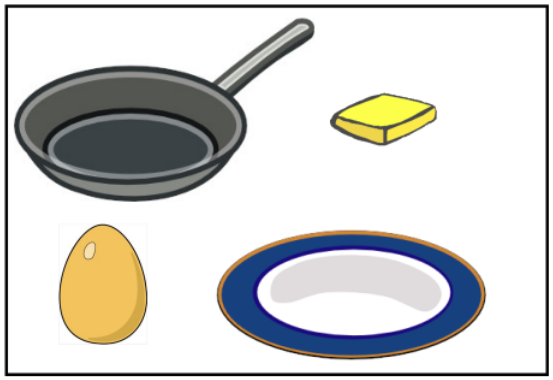
1 binary relation and 100 individuals give $100^2 = 10.000$ boolean features, or 2^{10000} states.

Levels of Detail

Do we need to distinguish states



egg=*whole*,
butter_in=*table*,
egg_in=*table*,
butter_position_to_pan=left



egg=*whole*,
butter_in=*table*,
egg_in=*pan*
butter_position_to_pan=right

- Not at "recipe level"
- Yes at robot control level: "*move arm to left, grab butter, . . .*"
- $\rightsquigarrow$ may need hierarchical description of state space to reason at different levels of abstraction.

Other dimensions of complexity

Modularity

- Flat, modular, hierarchical

Planning horizon

- Non-planning, finite, indefinite, infinite

Preferences

- Achievement goals, maintenance goals.
- Complex ordinal or cardinal preferences.

Learning

- Knowledge is given at design time or learned from experience.

Computational limits

- Perfect or bounded rationality

The Environment of Rational Agents

- **Accessible** vs. **inaccessible** (**fully observable** vs. **partially observable**)

  Are the relevant aspects of the environment accessible to the sensors?

- **Deterministic** vs. **stochastic**

  Is the next state of the environment completely determined by the current state and the selected action?

  If the only non-determinism are actions of other agents, the environment is called **strategic**.

- **Episodic** vs. **sequential**

  Can the quality of an action be evaluated within an episode (perception + action), or are future developments decisive?
  $\rightarrow$Related to the planning horizon of the agent

The Environment of Rational Agents, ctd.

- **Static** vs. **dynamic**

  Can the environment change while the agent is deliberating?

  If the environment does not change, but the agent's performance score changes, the environment is called **semi-dynamic**.

- **Discrete** vs. **continuous**

  Is the environment discrete or continuous?

- **Single agent** vs. **multi-agent**

  Is there just one agent, or several of them?

  There are **competitive** and **cooperative** multi-agent scenarios.

Examples of Environments

Task	Observable	Deterministic	Episodic	Static	Discrete	Agents
Chess/Go without clock	fully	strategic	sequential	static	discrete	multi
Poker	partially	stochastic	sequential	static (?)	discrete	multi
Car driving	partially	stochastic	sequential	dynamic	continuous	multi
Medical diagnosis	partially	stochastic	episodic	dynamic	continuous	single
Image analysis	fully	deterministic	episodic	semi	continuous	single
Part-picking robot	partially	stochastic	episodic	dynamic	continuous	single
Refinery controller	partially	stochastic	sequential	dynamic	continuous	single
Interactive English tutor	partially	stochastic	sequential	dynamic	discrete	multi

$\rightarrow$ These properties may depend on the design: E.g., if the medical diagnosis system interacts with skeptical staff then it's multi-agent, and if we take into account the overall treatment then it's sequential.

Questionnaire

Question!

James Bond's environment is?

(A): Fully Observable. **(B): Episodic.**

(C): Static. **(D): Single-Agent.**

Question!

Your own environment is?

(A): Fully Observable. **(B): Episodic.**

(C): Static. **(D): Single-Agent.**

Questionnaire Answers

First Question: James Bond's environment is?

→ (A) Fully Observable: Definitely not! Else Bond would always know immediately what the bad guys are up to.

→ (B) Episodic: Definitely not. Every one of Bond's "actions" would be "rewarded" separately and independently. The "film plot" would consist of saving/not-saving the world about every 2 minutes.

→ (C) Static: Definitely not. Just imagine Bond standing there, thinking, while the bad guys release Godzilla in NYC (or whatever else they may be up to).

→ (D) Single-Agent: Definitely not. A Bond film without bad guys would be boring.

Second Question: Your own environment is?

→ (A) Fully Observable: No. E.g., you don't know what the exam questions will be.

→ (B) Episodic: No. E.g., it takes more than one action to complete your studies.

→ (C) Static: No. E.g., if you take a year to decide how to prepare for the exam, it'll be over by the time you're done.

→ (D) Single-Agent: No. Apart from your family etc., for example at some point you will compete for the same job with somebody else.

Classifying AI Areas

Many sub-areas of AI can be classified by:

- Domain-specific vs. general.
- The environment.
- (Particular agent architectures sometimes also play a role, especially in Robotics.)

→ The same is true of the sub-topics in this course. The focus is on general methods (a bias in much of the AI field), and simple environments (after all, it's an introductory course only).

→ In the annex, only for reference: A rough classification of some of the topics of this course, in these terms.

Summary

- An **agent** is something that perceives and acts. It consists of an architecture and an agent program.
- A **rational agent** always takes the action that maximizes its expected performance, subject to the percept sequence and its environment knowledge.
- Some **environments** are more demanding than others . . .

  . . . your own, and that of James Bond, are the most difficult.

Reading

- *Chapter 1: Artificial Intelligence and Agents* from the book "Artificial Intelligence:Foundations of Computational Agents (2nd edition)
- *Chapter 2: Intelligent Agents* from the book "Artificial Intelligence:A Modern Approach (4th edition) [Russell and Norvig (2010)].

Further reading:

- Turing's Article: *Computing machinery and intelligence*, A. M. Turing. Mind, Volume LIX, Issue 236, October 1950, Pages 433–460.
  →Really nice philosophical discussion. Despite being > 70 years old, it's not hard to read and most ideas are still up to date.

- Subbarao Kambhampati's articles on recent AI trends:
  Language Imitation Games and the Arrival of Broad and Shallow AI **(Link)**
  →On the capabilities and limitations of new trends in AI

  Polanyi's Revenge and AI's New Romance with Tacit Knowledge **(Link)**
  →Should AI learn everything from scratch or should we try to give explicit knowledge?

References I

Vid Kocijan, Ernest Davis, Thomas Lukasiewicz, Gary Marcus, and Leora Morgenstern. The defeat of the winograd schema challenge. *CoRR*, abs/2201.02387, 2022.

Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Prentice-Hall, Englewood Cliffs, NJ, 1995.

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

Classifying Some AI Sub-Areas

Classical Search (Chapter 2)

Environment:

- **Fully observable** vs. partially observable.
- **Deterministic** vs. stochastic.
- Episodic vs. **sequential**.
- **Static** vs. dynamic.
- **Discrete** vs. continuous.
- **Single-agent** vs. multi-agent.

Approach:

- **Domain-specific** vs. general.

Classifying Some AI Sub-Areas

Planning (Chapter 11)

Environment:

- **Fully observable** vs. partially observable.
- **Deterministic** vs. stochastic.
- Episodic vs. **sequential**.
- **Static** vs. dynamic.
- **Discrete** vs. continuous.
- **Single-agent** vs. multi-agent.

Approach:

- Domain-specific vs. **general**.

→ Planning formalisms and approaches exist also for any and all of partial observability, and stochastic/dynamic/continuous/multi-agent settings.

Classifying Some AI Sub-Areas

Adversarial Search (Chapter 12)

Environment:

- **Fully observable** vs. partially observable.
- **Deterministic** vs. stochastic.
- Episodic vs. **sequential**.
- **Static** vs. dynamic.
- **Discrete** vs. continuous.
- Single-agent vs. **multi-agent**.

Approach:

- **Domain-specific** vs. general.

→ Adversarial search formalisms and approaches exist also for partial observability and stochastic settings.

Classifying Some AI Sub-Areas

General Game Playing (Not considered here)

Environment:

- **Fully observable** vs. partially observable.
- **Deterministic** vs. stochastic.
- Episodic vs. **sequential**.
- **Static** vs. dynamic.
- **Discrete** vs. continuous.
- Single-agent vs. **multi-agent**.

Approach:

- Domain-specific vs. **general**.

→ General game playing formalisms and approaches exist also for partial observability and stochastic settings.

Classifying Some AI Sub-Areas

Constraint Satisfaction & Reasoning (Chapter 3)

Environment:

- **Fully observable** vs. partially observable.
- **Deterministic** vs. stochastic.
- **Episodic** vs. sequential.
- **Static** vs. dynamic.
- **Discrete** vs. continuous.
- **Single-agent** vs. multi-agent.

Approach:

- Domain-specific vs. **general**.

Classifying Some AI Sub-Areas

Probabilistic Reasoning (Chapter 13)

Environment:

- Fully observable vs. **partially observable**.
- Deterministic vs. **stochastic**.
- **Episodic** vs. sequential.
- **Static** vs. dynamic.
- **Discrete** vs. continuous.
- **Single-agent** vs. multi-agent.

Approach:

- Domain-specific vs. **general**.

Classification of Agent Designs

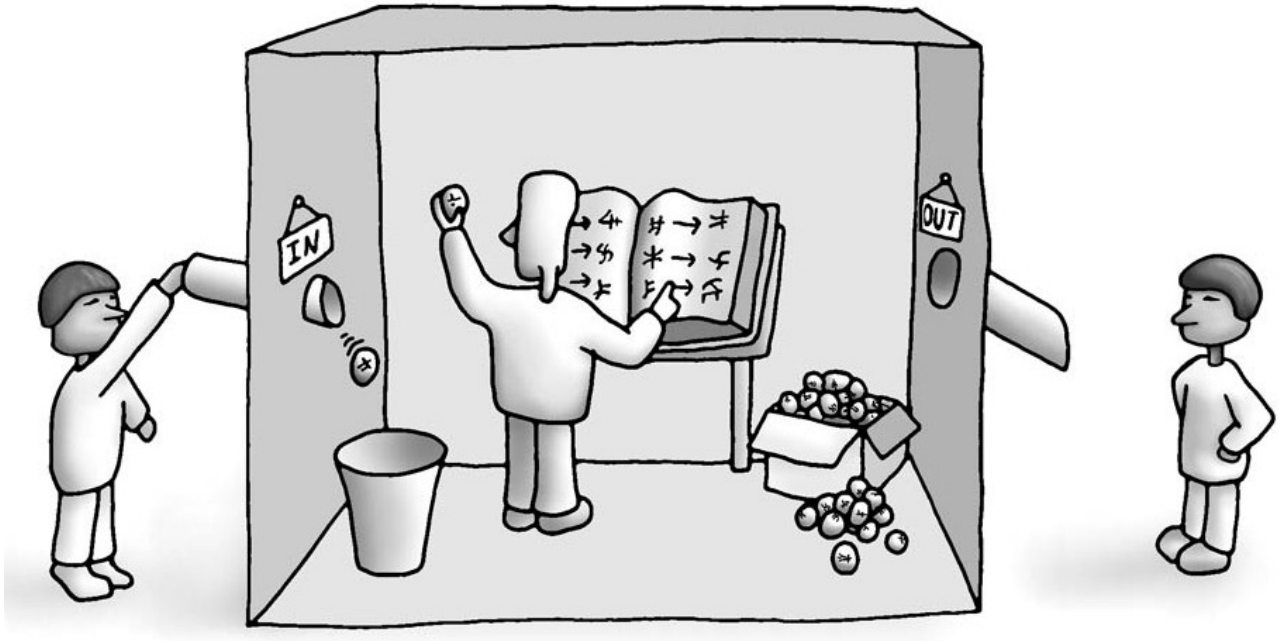
There are a variety of agent designs:

- **Reflex agents** respond to percepts by condition-action rules.
- **Goal-based agents** work towards goals.
- **Utility-based agents** make trade-offs using a utility function.
- **Learning agents** improve their behavior over time.

Table-Driven Agents

function TABLE-DRIVEN-AGENT(*percept*) **returns** an action
  persistent: *percepts*, a sequence, initially empty
    table, a table of actions, indexed by percept sequences, initially fully specified

  append *percept* to the end of *percepts*
  $action \leftarrow$ LOOKUP(*percepts*, *table*)
  return *action*

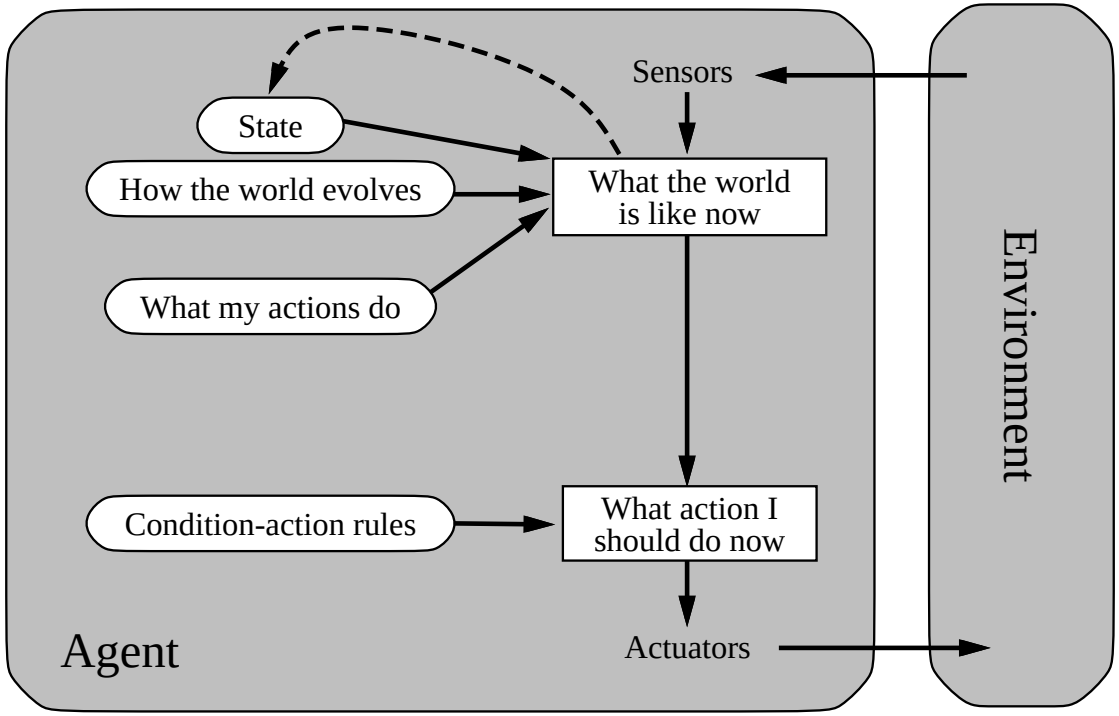


↪ **Why is this not a good idea?**

- The table is way too large. (Even when limiting to n binary percepts whose order

Reflex Agents

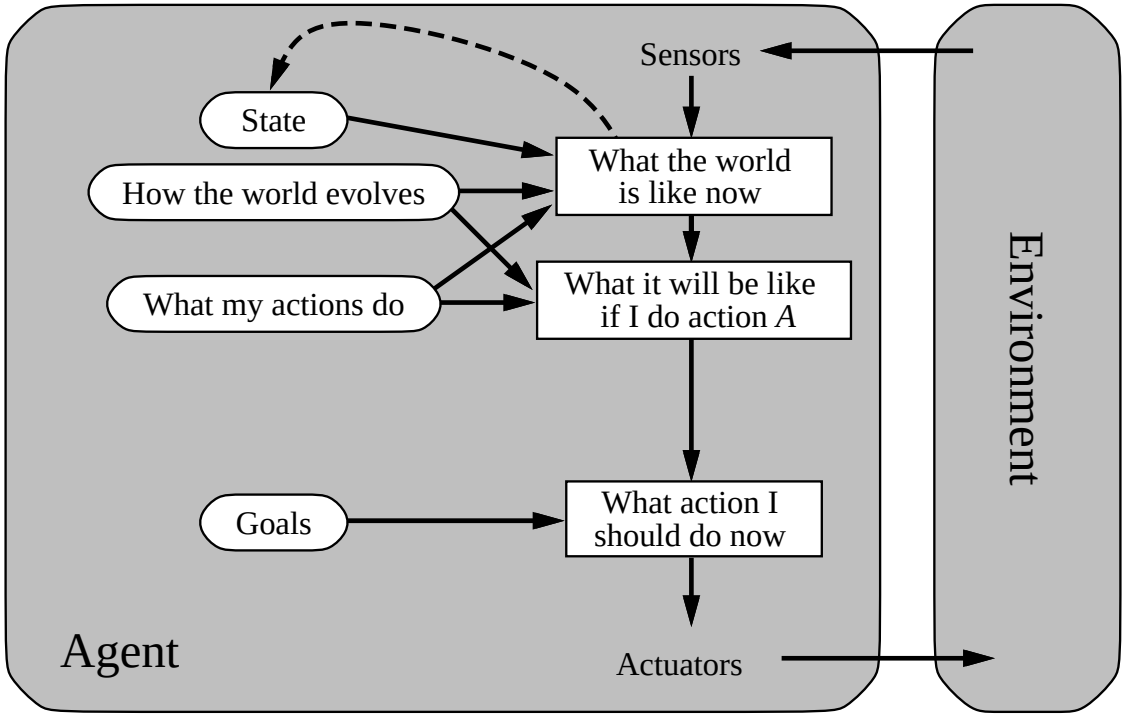
More useful, but still very simple, method for choosing actions: **Condition-Action Rules** (note: raw sensor data *interpreted*, using a *world model*, prior to evaluating the rules)



→ **Example?** Vacuum cleaner: If it's dirty where you are right now, clean; otherwise, move somewhere else randomly.

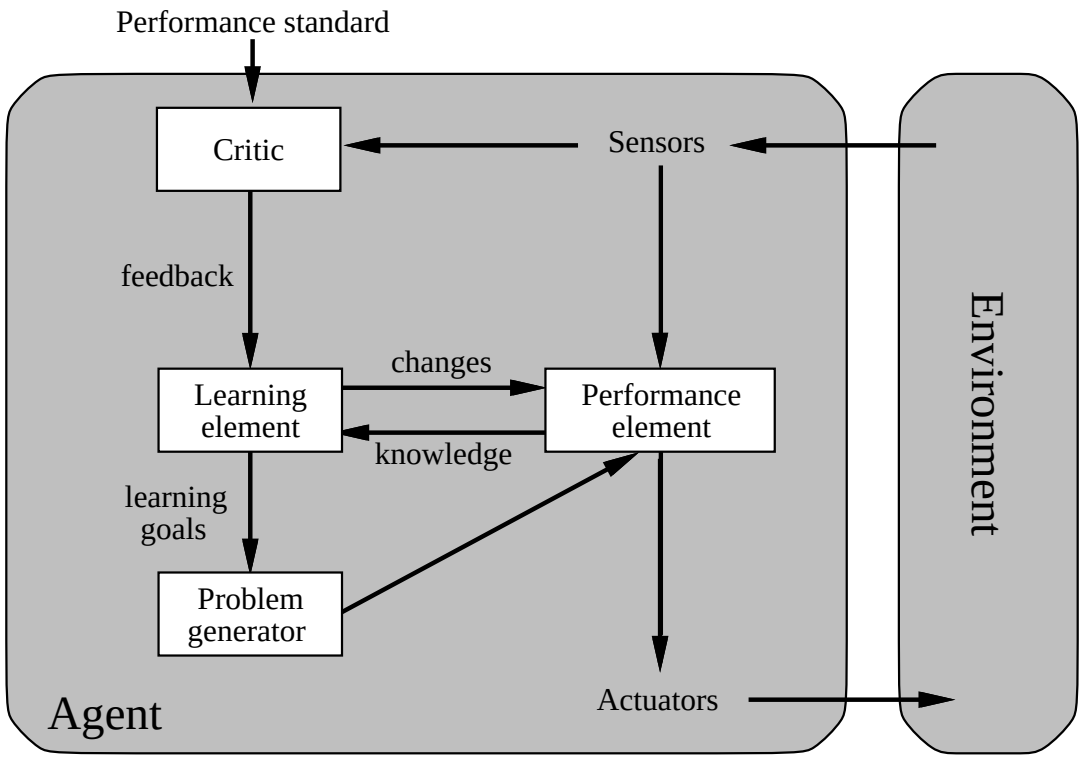
Goal-Based Agents (Belief-Desire-Intention)

Often, doing the right thing requires considering the future:



→ **Example?** If you're driving a car, then, at any one crossing you get to, whether you go left/right/straight (or U-turn) depends on where you want to get to.

Learning Agents



- **Critic**: Measures performance.
- **Learning element**: Learns new knowledge.
- **Performance element**: Selects actions (RL: **exploitation**).
- **Problem generator**: Suggests actions favoring informative learning experiences (RL: **exploration**).