TDT4136 Introduction to Artificial Intelligence

Lecture 1: Introduction
(Chapter 1 and 2 in the textbook)

Pinar Öztürk

Norwegian University of Science and Technology
Fall 2024

TDT4136 Teaching Team

Coordinator: Pinar Öztürk

Lecturers:
Ole Christian Eidheim
Xavier Fernando Cuauhtémoc Sánchez Díaz
Keith Downing
Ahmed Abouzeid
Pinar Öztürk

Teaching assistants:
Espen Berntzen Høgstedt
Xavier Fernando Cuauhtémoc Sánchez Díaz

Student assistants: 11 assistants.

Syllabus

- Text book: Artificial Intelligence - A Modern Approach, Russel & Norvig, 4th edition (Global)
- Lecture slides
- See blackboard ("Syllabus") for more information.

Lecture slides and streams/videos: Published weekly on Blackboard.

Exam

- The final exam counts 100 %.
- The *mandatory assignments* shall be deliver and passed in order to take the exam. .
- Exam date and form: Written exam. The date is/will be announced central by NTNU . The date will also be announced on BB later.
- Exam language: Questions only in English, you can answer in English or Norwegian

Assignments

- There will be 3 mandatory assignments that must be delivered and "passed".
- You may deliver the assignments individually or may work with one another student.

  You both will have to write your names when delivering your assignment
  You may not copy work from anyone else
  You may not give others your solution
- Assignment lectures: There will be 3 assignment lectures during the semester.

Lab hours

- Labs start on 02.09.2024
- Labs on 3 days a week, 2-hours each.
- 2 weeks of lab for each assignment.
- You attend if you need, and as much as you need.

Main information medium and material source

During the semester, you'll find the most updated information on the Blackboard:

- Announcements about practicalities
- The schedules
- Slides for the main and assignment lectures
- Streams/videos for the main and assignment lectures
- The assignments
- Link to discussion forum

You are responsible to make yourself aware of the announcements on the blackboard

More on *mandatory* assignments

Overview

- Assignments 1 (Informed and uninformed search) requires to solve and write the anwers by hand, i.e., not programming

- Assignment 2 (constraint satisfaction problem) and assignment 3 (adversarial search) are is programming assignments
  - Consider that the student/lærings assistants know Python better!
  - Submit only the necessary files and read instructions carefully!
  - Submit **digital** solutions only: typeset PDF (Word/LaTeX) or tablet handwriting
  - Absolutely no scanned paper nor photos

Are assignments mandatory?

Yes, you need to pass 3 mandatory assignments to take the exam.

I am re-taking the course. Do I need the assignments as well?

If you had approved assignments in 2023, then you do **not** need to pass the assignments in 2024.

I have a question about an assignment

Go to the forum[1] and see if the question has been asked before. If not, post your question.

[1] See `https://tdt4136.idi.ntnu.no/`

Frequently Asked Questions

Assignments

I signed up for the wrong group!

Send an email to **Espen**[2]

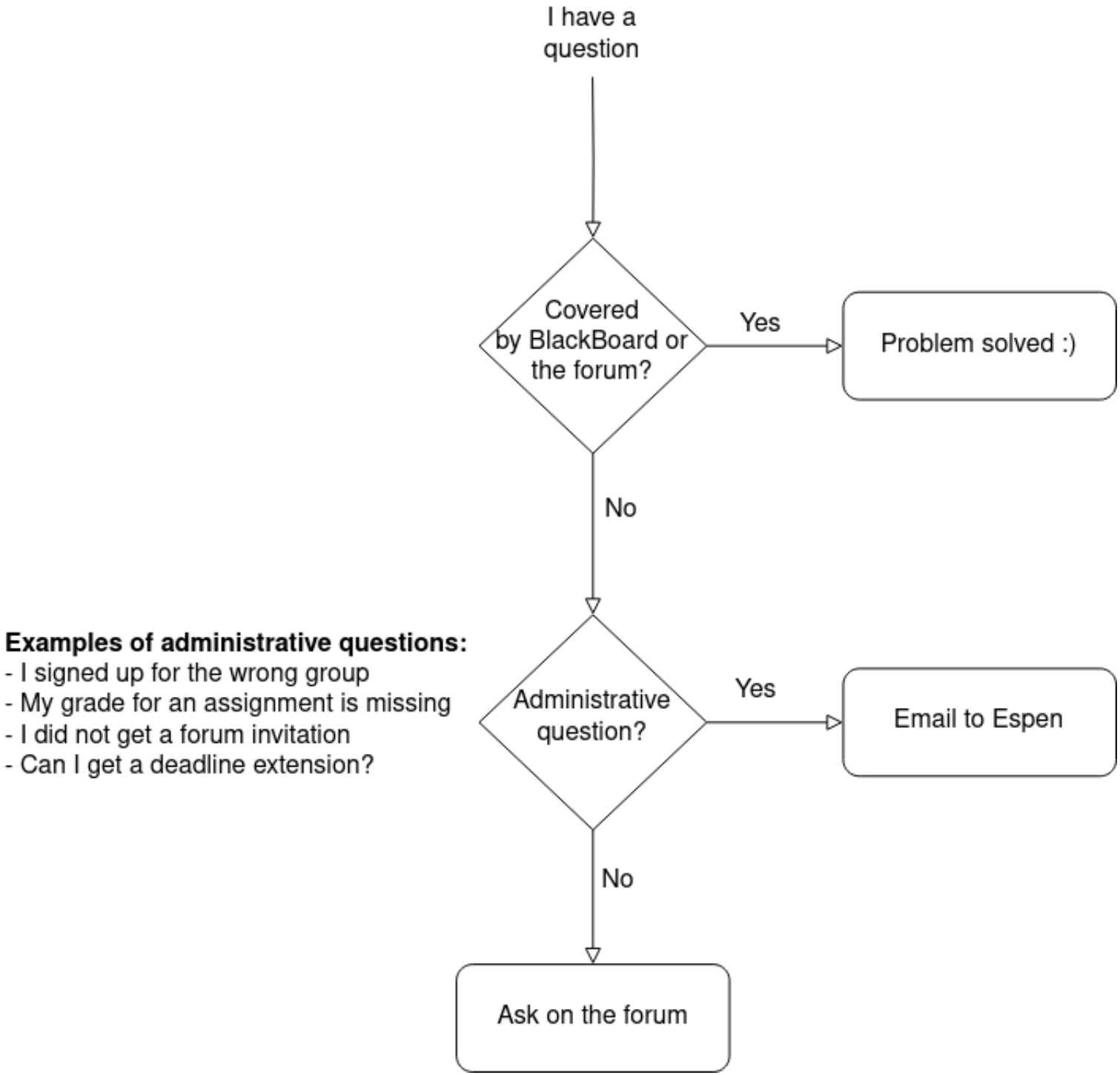
I didn't get a grade for an assignment!

Check that the grading deadline has passed. If the grade is still overdue, send an email to **Espen**

I need an extension to the deadline!

Under certain circumstances you can get an extension (e.g. w/ doctor's notice). Explain why in an email to **Espen**

[2] espen.b.hogstedt@ntnu.no



Examples of administrative questions:
- I signed up for the wrong group
- My grade for an assignment is missing
- I did not get a forum invitation
- Can I get a deadline extension?

Please prefix the topic of any email with TDT4136

For example

TDT4136 - Incorrect grade on Assignment 2

How many assignment lecturese will there be?

Three, one for each mandatory assignment.

Are assignment lectures mandatory?

No, they are not mandatory although they could be helpful.

Will assignment lectures be recorded?

Yes, they will be recorded like regular lecture.

Check the detailed FAQ in Blackboard!

Course topic overview

Intelligent agents

Problem solving by Searching

Adversarial Search

Constraint Satisfaction Problems

Logical systems

Knowledge representation

Ethical issues in AI

Planning

Game Theory

Natural Language processing

Birth of AI field

A Summer Research Project in the Dartmouth College, in 1956 was the birth of the AI research field.

(From left: Trenchard More, John McCarthy, Marvin Minsky, Oliver Selfridge, and Ray Solomonoff)

From the project proposal: "...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."

What is AI?

There are no crisp definitions. Here is one from John McCarthy, (father of the phrase *Artificial Intelligence*)

Question: What is artificial intelligence?
Answer: It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable.

Question: Yes, but what is intelligence?
Answer: Intelligence is the computational part of the ability to achieve goals in the world. Varying kinds and degrees of intelligence occur in people, many animals and some machines.

see `http://www.formal.Stanford.EDU/jmc/whatisai/`

What is AI? - cont

"AI is the science of making machines to do things that would require intelligence if done by men." Marvin Minsky

"The study of mental faculties through the use of computational models" Eugen Charniak

In short: There is no formal definition covering all aspects of intelligence

Can machines be Intelligent?

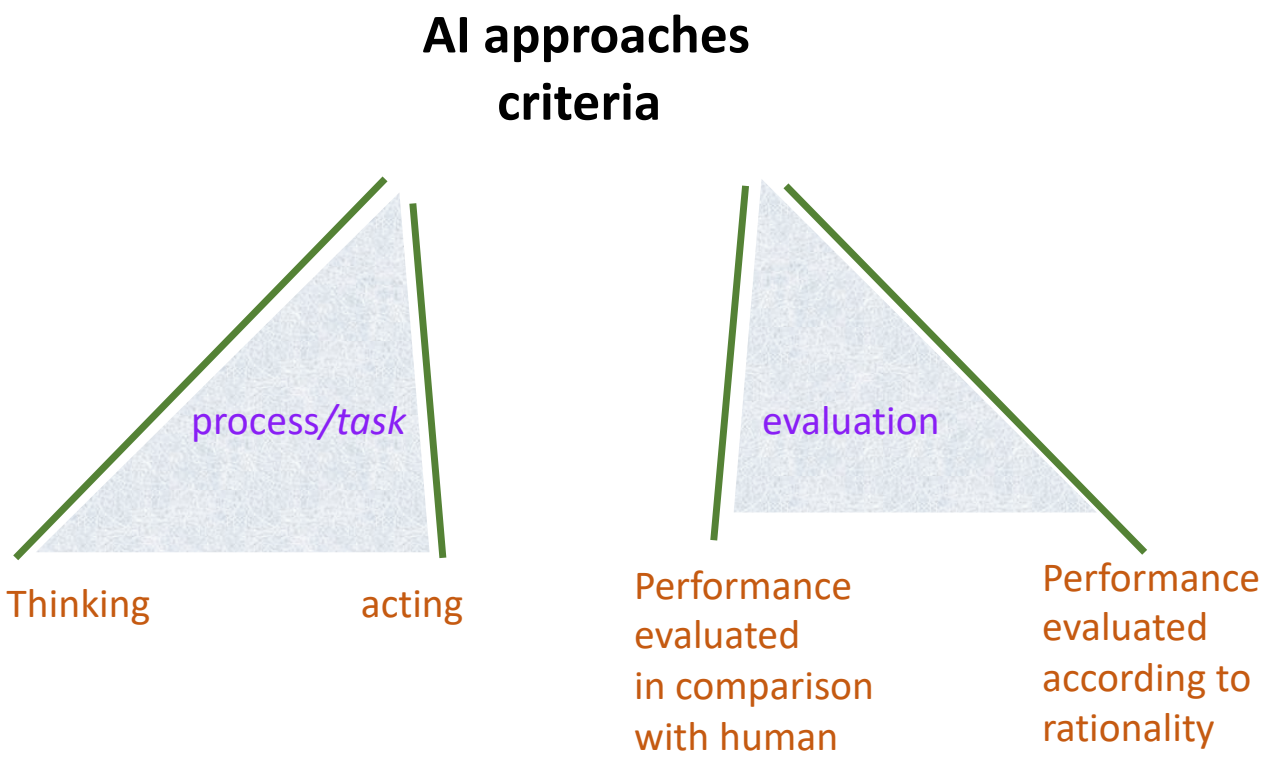
- *Symbolic system hypothesis* (Newell and Simon)
  - Intelligence is substrate neutral
  - A *physical symbol system* has necessary and sufficient means for general intelligent action.

- Biological substrate only (John Searle, philosopher)
  - Intelligence is substrate dependent.
  - The material humans are made of is fundamental for our intelligence.
  - Thinking is possible only in special machines - living ones made of proteins.

What is AI?

Russell and Norvig's definition of AI

Two dimensions:



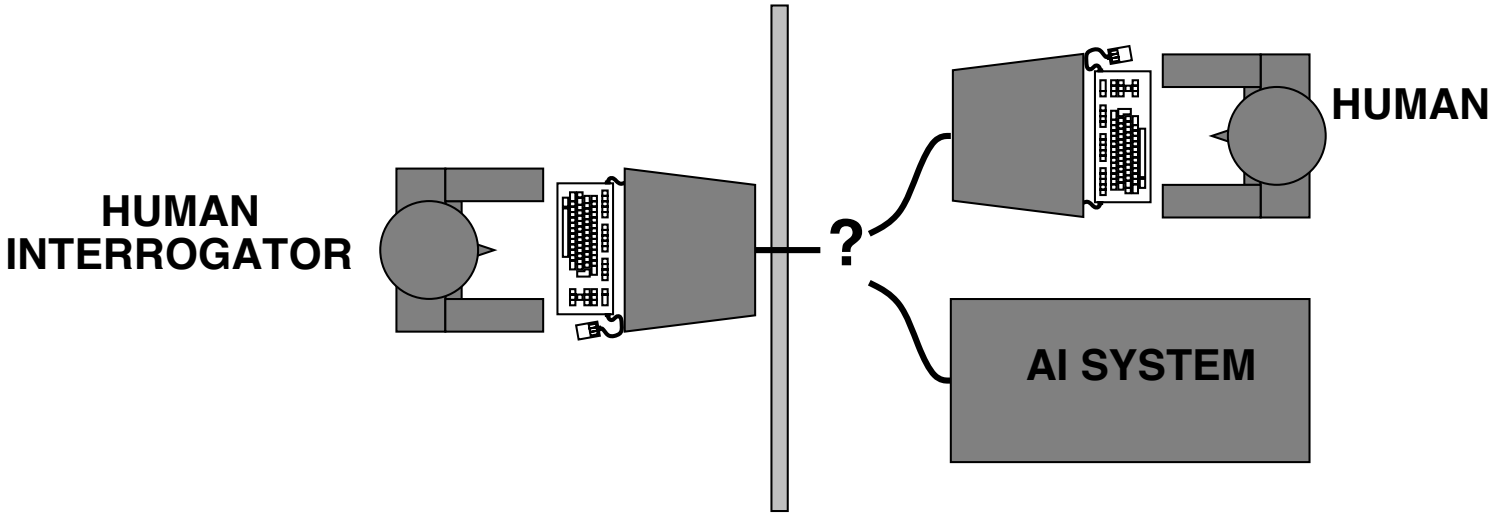
What is AI?

How R&N define AI

thinking vs. acting \ humanly vs. rationally		
	Systems that think like humans	Systems that think rationally
	Systems that act like humans	Systems that act rationally Rational Agents

Acting humanly

Alan Turing designed an operational test for intelligence -1950



- Computer would need:
  - natural language processing
  - knowledge representation
  - automated reasoning
  - machine learning
- Total Turing test requires interaction with the external world, would require:
  - computer vision and speech recognition
  - robotics

Acting Humanly - Eliza

- One of the most famous early AI programs: Eliza, the computer psychotherapist, created by Joseph Weizenbaum in 1966 at MIT.
- Eliza functions by "twisting the statements of her 'patients' back at them in the classic manner of a non-directive psychotherapist."
- The fact that it understands the subject's statements is an illusion.
- Surprisingly, many users were taking its performance quite seriously.

Acting Humanly - Eliza 2

young woman: Men are all alike.
eliza: In what way?
young woman: They're always bugging us about something specific or other.
eliza: Can you think of a specific example?
young woman: Well, my boyfriend made me come here.
eliza: Your boyfriend made you come here?
young woman: He says I'm depressed much of the time.
eliza: I am sorry to hear you are depressed.
young woman: It's true. I'm unhappy.
eliza: Do you think coming here will help you not to be unhappy?
young woman: I need some help, that much seems certain.
eliza: What would it mean to you if you got some help?
young woman: Perhaps I could learn to get along with my mother.
eliza: Tell me more about your family.
young woman: My mother takes care of me.
eliza: Who else in your family takes care of you?
young woman: My father.
eliza: Your father.
young woman: You are like my father in some ways.
eliza: What resemblance do you see?
young woman: You are not very aggressive but I think you don't want me to notice that.
eliza: What makes you think I am not very aggressive?
young woman: You don't argue with me.
eliza: Why do you think I don't argue with you?
young woman: You are afraid of me.
eliza: Does it please you to believe I am afraid of you?
young woman: My father is afraid of everybody.
eliza: What else comes to mind when you think of your father?
young woman: Bullies.
eliza: Does that have anything to do with the fact that your boyfriend made you come here?

Thinking Humanly

- Cognitive modelling approach
- Learn about how humans think
  - introspection
  - psychological experiments
  - brain imaging
- Cognitive Science promising

Thinking rationally

- Aristotle: what are correct thought processes - Laws of Thought?
  - Formalize "correct" reasoning using a mathematical model
  - Syllogism (as a law of though) codifies the patterns for argument structures.
- Theory of probability

Thinking rationally

Thinking rationally

Thinking rationally

.

Problems:

- General deductive inference is computationally intractable
- It does not generate intelligent *behaviour*

Acting Rationally: Rational Agents approach

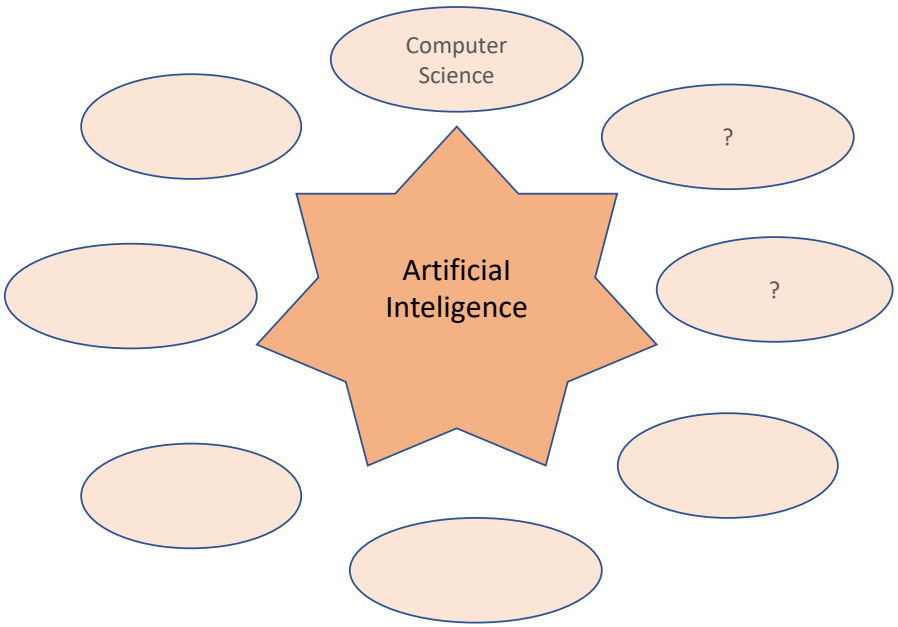
- Rational behavior: doing the right thing
- The right thing: that which is expected to maximize goal achievement, given the available information and computational abilities
- Doesn't necessarily involve thinking—e.g., blinking reflex—but thinking should be in the service of rational action
- Two advantages:

  Not limited to "laws of thought" in order to achieve rationality
  Rationality is mathematically well-defined - more operationable

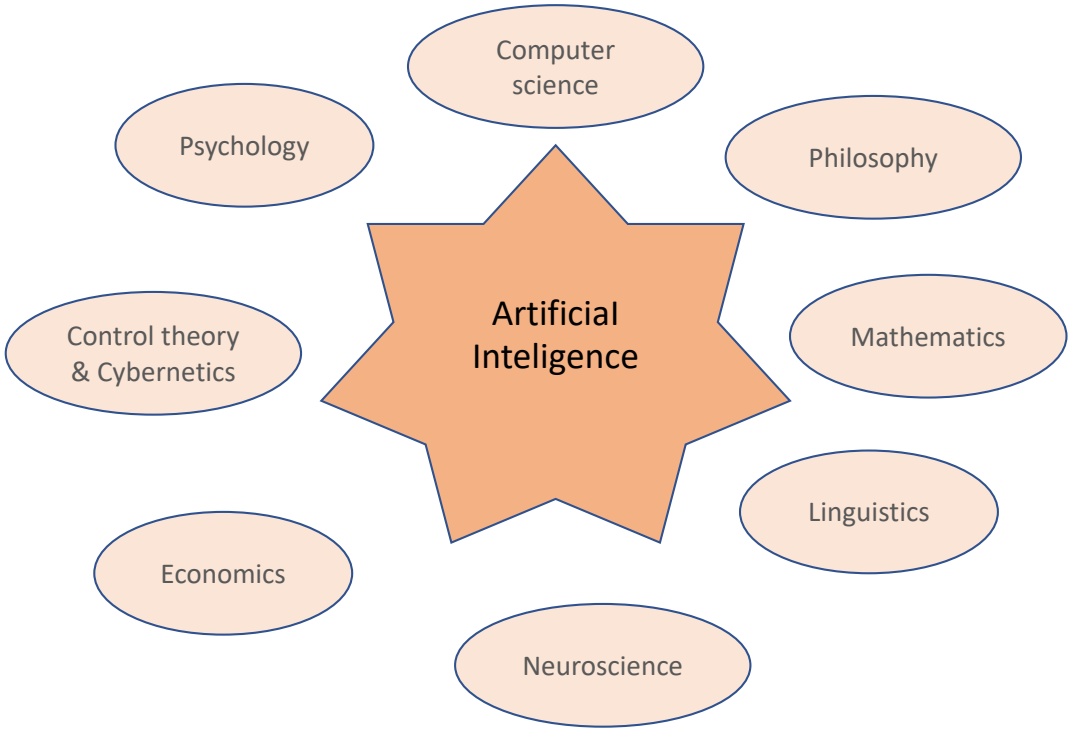
Foundations of AI

Computer Science is the main discipline underlying AI

Can we think which other disciplines AI is grounded on?



Foundations of AI



Brief history of AI

1943	McCulloch & Pitts: Boolean circuit model of brain
1950	Turing's "Computing Machinery and Intelligence"
1952-69	Early AI programs, including Samuel's checkers program, Newell & Simon's Logic Theorist, Gelernter's Geometry Engine
1956	Dartmouth meeting: "Artificial Intelligence" adopted
1962	Rosenblatt's Perceptron for training simple neural networks
1965	Robinson's complete algorithm for logical reasoning
1972	The logic programming language PROLOG is created.
1966–74	Disappointment: AI discovers computational complexity Neural network research almost disappears

Brief history of AI -cont.

1969–79	Early development of knowledge-based systems
1980–88	Expert systems industry booms
1988–93	Expert systems industry busts: "AI Winter"
1985–95	Backpropagation learning returns neural networks to popularity
1988–	Resurgence of probability; general increase in technical depth "Nouvelle AI": ALife, GA
1995–	Agents, agents, everywhere ...
2003–	Human-level AI back on the agenda
2005–2010	AI disappoints again, AI is not much appreciated
2012 -	Deep learning and a very hot AI summer since then

Example from Early History of AI

1943 McCulloch & Pitts: Boolean circuit model of brain

BULLETIN OF
MATHEMATICAL BIOPHYSICS
VOLUME 5, 1943

A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY

WARREN S. MCCULLOCH AND WALTER PITTS

FROM THE UNIVERSITY OF ILLINOIS, COLLEGE OF MEDICINE,
DEPARTMENT OF PSYCHIATRY AT THE ILLINOIS NEUROPSYCHIATRIC INSTITUTE,
AND THE UNIVERSITY OF CHICAGO

Because of the "all-or-none" character of nervous activity, neural events and the relations among them can be treated by means of propositional logic. It is found that the behavior of every net can be described in these terms, with the addition of more complicated logical means for nets containing circles; and that for any logical expression satisfying certain conditions, one can find a net behaving in the fashion it describes. It is shown that many particular choices among possible neurophysiological assumptions are equivalent, in the sense that for every net behaving under one assumption, there exists another net which behaves under the other and gives the same results, although perhaps not in the same time. Various applications of the calculus are discussed.

I. Introduction

Theoretical neurophysiology rests on certain cardinal assumptions. The nervous system is a net of neurons, each having a soma and an axon. Their adjunctions, or synapses, are always between the axon of one neuron and the soma of another. At any instant a neuron has some threshold, which excitation must exceed to initiate an impulse. This, except for the fact and the time of its occurrence, is determined by the neuron, not by the excitation. From the point of excitation the impulse is propagated to all parts of the neuron. The velocity along the axon varies directly with its diameter, from less than one meter per second in thin axons, which are usually short, to more than 150 meters per second in thick axons, which are usually long. The time for axonal conduction is consequently of little importance in determining the time of arrival of impulses at points unequally remote from the same source. Excitation across synapses oc-

Some killer apps in AI history

1991 During the Gulf War, US forces deployed an AI logistics planning and scheduling program that involved up to 50,000 vehicles, cargo, and people. Saved the US more money than spent on all AI research since 1950

1997 Deep Blue (IBM) defeated world chess champion Gerry Kasparov

2011 Watson (IBM) beat human champions on "Jeopardy"

2012 Google car obtains driver's license in Nevada, US.

2017 DeepMind's Alphago AND Elon Musk's A.I. Destroys Champion Gamer!

2022 Open AI's. ChapGPT, language model/text generation.

How humanly is Sofia (Hong Kong firm Hanson Robotics)?

https://www.youtube.com/watch?v=suRuQbDXcrc

More Videos

Computers with common sense(Doug Lenat):
https://www.youtube.com/watch?v=2w_ekB08ohU

Artificial Intelligence: The Common Sense Problem (Hubert Dreyfus)
https://www.youtube.com/watch?v=SUZUbYCBtGI

AI Boom (Rodney Brooks):
https://www.youtube.com/watch?v=ovbzjLobiDo

IBM Watson : https://www.youtube.com/watch?v=_Xcmh1LQB9I

Jeopardy : https://www.youtube.com/watch?v=WFR3lOm_xhE

State of the art AI (Minsky):
https://www.youtube.com/watch?v=aODnFdU_hds

John McCarthy talks about AI:

Chapter 2 Outline

What is an agent?

Rationality

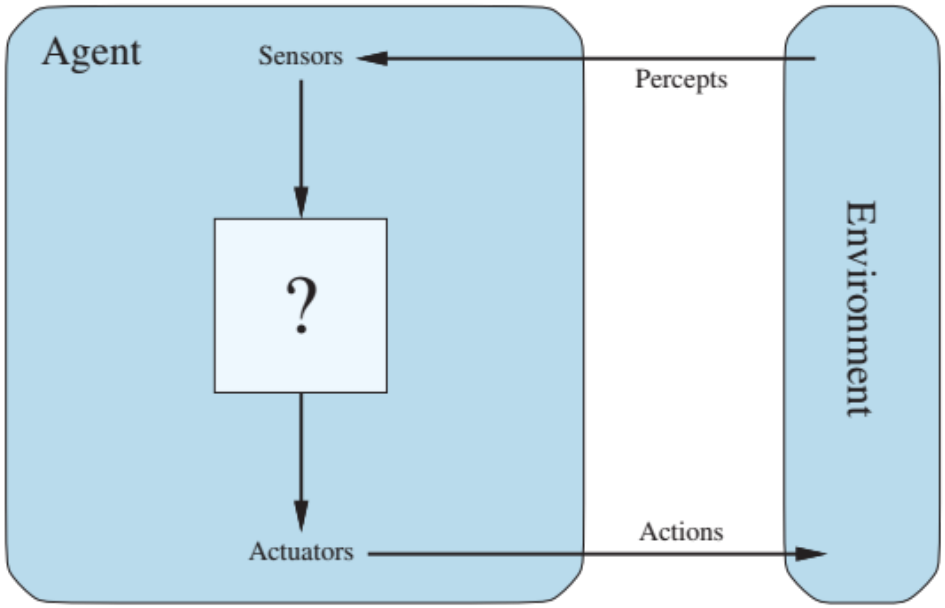
PEAS (Performance measure, Environment, Actuators, Sensors)

Environment types

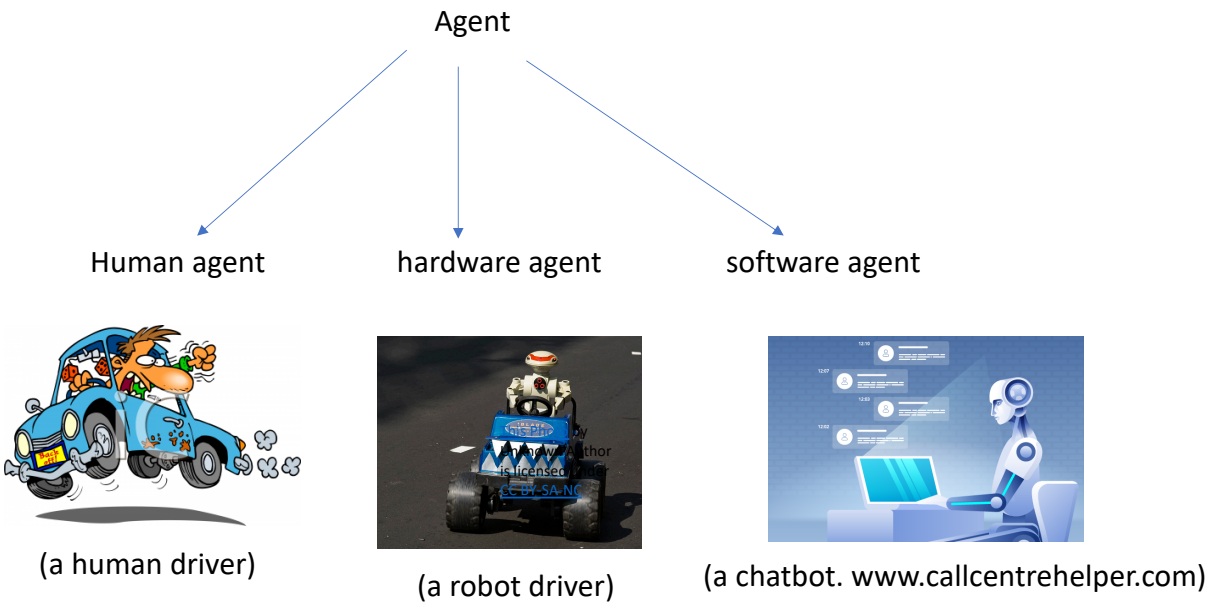
Agent types

What is an agent

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



What is an agent



robot:

perceives: with video-cameras, infra-red sensors, radar, ...
acts: with wheels, motors

softbot:

perceives: receiving keystrokes, files, network packets, ...
acts: displaying on the screen, writing files, sending network packets

. . . .

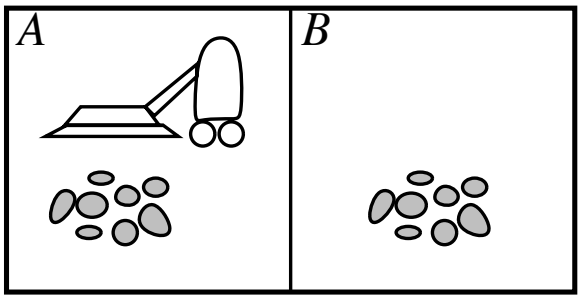
Agents behaviour

An agent´s choice of action at any given instant can depend on the entire percept sequence observed to date

Hence, agent's behaviour is described by the **agent function** that maps any percept sequence to an action.

A simple agent function

- Example: vacuum cleaner world. If the current square is dirty, then suck; otherwise, move to the other square.
- Tabulating the vacuum's agent function:



Percept sequence	Action
[*A*, *Clean*]	*Right*
[*A*, *Dirty*]	*Suck*
[*B*, *Clean*]	*Left*
[*B*, *Dirty*]	*Suck*
[*A*, *Clean*], [*A*, *Clean*]	*Right*
[*A*, *Clean*], [*A*, *Dirty*]	*Suck*
⋮	⋮

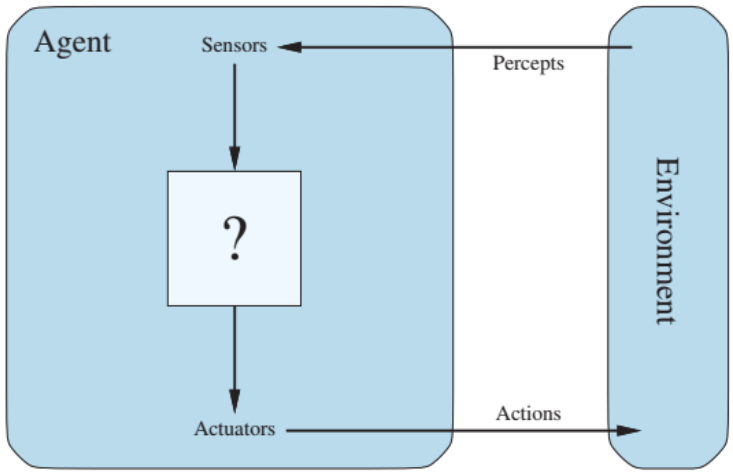
What is the right way to fill this table?

Agent Behaviour

Agent function - math description of behaviour, maps from percept sequences to actions:

$$f : \mathcal{P} \rightarrow \mathcal{A}$$

Agent program - concrete implementation of agent function, runs on the physical *architecture* to produce f



Rational agent

- A rational agent is one that "does the right thing" in context
  - but what does this really mean?
- The right action should cause the agent to be most successful
  - but how and when should the success of the agent be evaluated?
- Success needs to be evaluated with respect to an objective *performance measure*, which depends on what the agent is designed to achieve

Performance measure

Performance measure evaluates the environment sequence - *consequentialism*.

- For example, in case of the vacuum cleaner, the designer could focus only on collecting as much dirt as possible in time T
- Or, take also other factors into account, e.g :
  - amount of time taken to clean
  - amount of electricity consumed
  - amount of noise generated, etc
- e.g., one point per square cleaned up in time T, or
  - one point per clean square per time step, minus one per move
  - etc

Rational agent

For each possible percept sequence, an ideal rational agent should do whatever is expected to maximise its performance measure, on the basis of the evidence provided by the percept sequence and whatever built-in knowledge the agent has.

Is the "tabulated vacuum agent" rational under these conditions:

- performance measure gives 1 point for each clean room at each time step, over a lifetime of 100 time steps
- there are 2 rooms, the agent has location and dirt sensors. Clean room stays clean, and "suck" works properly
- available actions are right, left, suck
- no noise or error in sensors

Rationality requires learning and autonomy

- Learning agent modifies knowledge of the environment from experience

- An agent is autonomous to the extent that its behaviour is determined by its own experience
  - complete autonomy from the start is too difficult
  - the agent´s designer must give guidance in terms of some initial knowledge and the ability to learn and/or reason as it operates in its environment
- Example: After a dung beetle digs its nest and lays its eggs, it fetches a ball of dung from a nearby heap to plug the entrance....

Task Environment

The task environment is defined through "PEAS":

- Performance measure
- The agent's prior knowledge of the environment
- The actions that the agent can perform
- The agent's percept sequence to date

Task Environment

PEAS (**P**erformance measure, **E**nvironment, **A**ctuators, **S**ensors)

Agent Type	Performance Measure	Environment	Actuators	Sensors
Taxi driver	Safe, fast, legal, comfortable trip, maximize profits, minimize impact on other road users	Roads, other traffic, police, pedestrians, customers, weather	Steering, accelerator, brake, signal, horn, display, speech	Cameras, radar, speedometer, GPS, engine sensors, accelerometer, microphones, touchscreen

Medical diagnosis system agent

Performance measure??

Environment??

Actuators??

Sensors??

Medical diagnosis system agent

Performance measure : healthy patient, minimize costs, lawsuits

Environment: patient, hospital, staff

Actuators: screen display (questions, tests, diagnoses, treatments, referrals)

Sensors: keyboard (entry of symptoms, findings, patient's answers)

Environment Properties

Fully observable environments

- Relevant parts of the state of the environment can be sensed
- No need to maintain any internal state to keep track of the world
- Example: chess, image analysis

Partially observable environments

- Parts of the environment cannot be sensed
- Agent must make informed guesses about world
- Example: poker, taxi driving, medical diasgnosis

Environment Properties

Single agent

- No other agents - there may be but as a part of the environment
- Examples: medical diagnosis, image analysis

Multi-agent

- Which entities will be viewed as "other agents"?
- The environment contains other agents whose performance measure depends on my actions and vice versa
- Competitive and cooperative interactions
- Examples: poker, chess, taxi driving

Properties of environments

Deterministic environment

Any action has a single guaranteed effect, and no uncertainty/failure.

Example: chess, image analysis

Non-deterministic environment

- There is some uncertainty about the outcome of an action
- Multiple outcome alternatives
- Called "stochastic" if alternatives are quantified in terms of probabilities
- Example: poker, taxi driving, medical diagnosis

Properties of environments

Episodic environments

- The agent's experience is divided into atomic episodes.
- Each episode consists of the agent perceiving and then performing a single action
- The choice of action in each episode depends only on the episode itself
- Examples: image analysis

Sequential environments

- The current decision could affect all future decisions
- Examples: poker, chess, taxi driving, medical diagnosis

Properties of environments

Discrete

Finite number of distinct states, percepts and actions,

Examples: chess, poker

Continuous

Continuous time/state/actions

Example: taxi driving, medical diagnosis, image analysis

Properties of environments

Dynamic environment

- May change while an agent is deliberating
- Examples: taxi driving, medical diagnosis

Static environment

- The environment does not change
- Examples: poker

Semidynamic

- The world does not change but the agent's performance score may
- Examples: chess with clock, image analysis

Environment	Observable	Deterministic	Episodic	Discrete	Dynamic	Agent
Crossword	fully	deterministic	sequential	discrete	static	
Chess w/clock	fully	deterministic	sequential	discrete	semidynamic	
Backgammon	fully	stochastic	sequential	discrete	static	
Taxi driving	partially	stochastic	sequential	continuous	dynamic	
Medical diagnosis	partially	stochastic	sequential	continuous	dynamic	

Properties of environments

Known environment

- The agent's knowledge about how the environment works/evolves
- Note that a known environment (i.e., the agent knows all the rules that apply) may be only partially observable

Unknown environment

- The agent will have to learn how it works
- An unknown environment can be fully observable.

A simple vacuum cleaner agent

Table-driven agent:

Percept sequence	Action
$[A, Clean]$	*Right*
$[A, Dirty]$	*Suck*
$[B, Clean]$	*Left*
$[B, Dirty]$	*Suck*
$[A, Clean], [A, Clean]$	*Right*
$[A, Clean], [A, Dirty]$	*Suck*
$\vdots$	$\vdots$

Table-driven agent

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 ← LOOKUP(*percepts*, *table*)
  return *action*

- looks up the right response in the table
- Infeasible: if there are $|P|$ percepts and a life-time of T, then need for a look-up table of size $\sum_{t=1}^{T} |P|^t$

Agent programs

Writing down such agent functions is not feasible

The goal of AI is to write agent programs with small code which produced the desired rational behaviour

Agent types

- simple reflex agents
- model-based reflex agent
- goal-based agents
- utility-based agents
- learning agents

Simple reflex agent

- Uses only the current percept - ignores the percept sequence,
- Implemented through condition-action rules
- Large reduction in possible percept/action situations from $\sum_{t=1}^{T} |P|^t$ to $|P|$ [3]

Example:

function Reflex-Vacuum-Agent([*location*,*status*]) **returns** an action

if *status* = *Dirty* **then return** *Suck*
else if *location* = *A* **then return** *Right*
else if *location* = *B* **then return** *Left*

[3] if there are $|P|$ percepts and a life-time of T, then need for a look-up table of size $\sum_{t=1}^{T} |P|^t$

Generic Simple Reflex Agent program

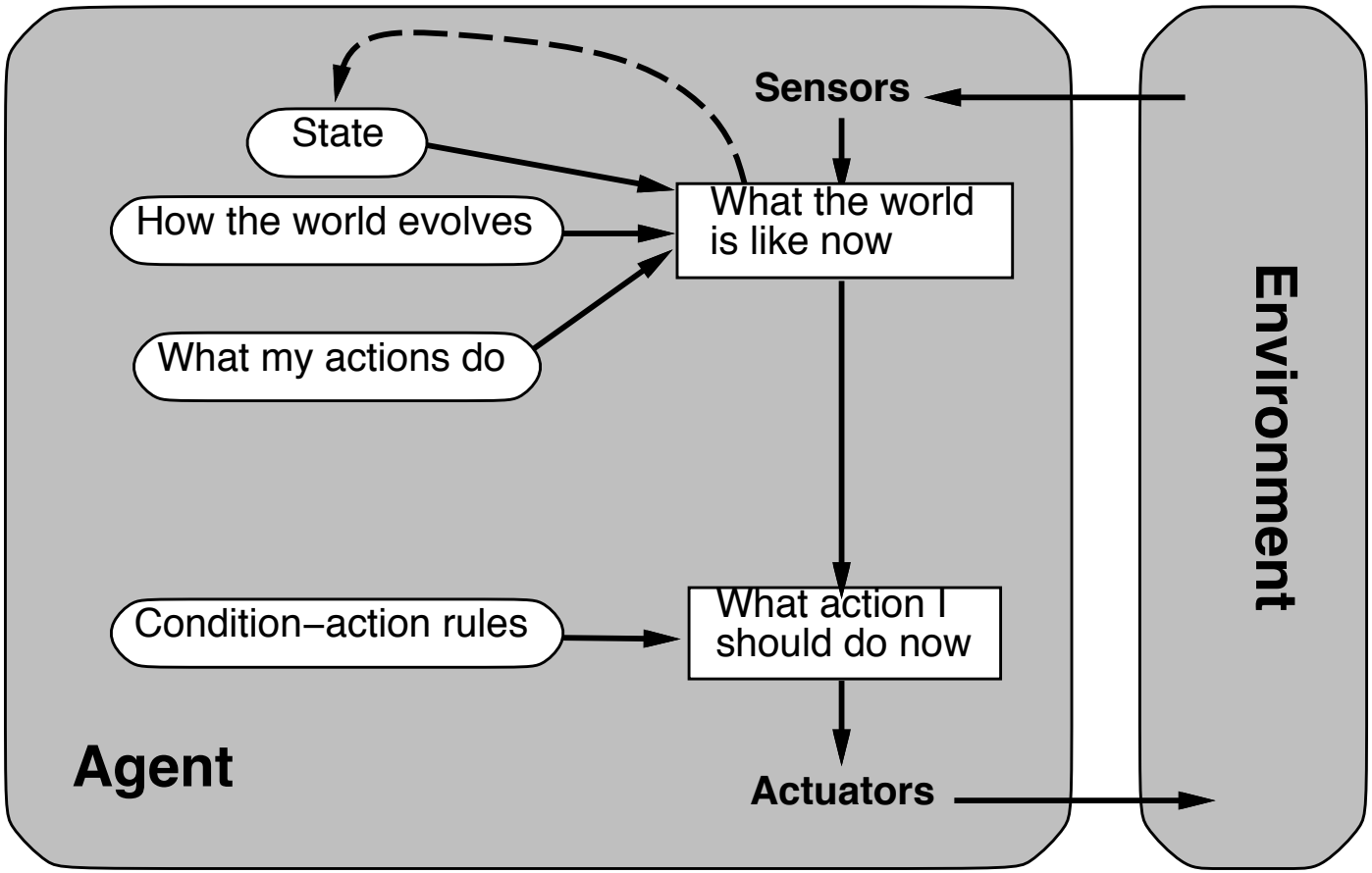
A more general simple reflex agent program:

```
function SimpleReflexAgent( percept) returns an action

persistent rules
state ← Interpret(percept)
rule ← Rule-match(state, rules)
action ← Rule-action(rule)
return action
```

- Will only work if the environment is fully observable
- everything relevant needs to be determinable from the current input
- otherwise infinite loops may occur, e.g. in the vacuum world without a sensor for the room, the agent does not know whether to move right or left
  - any possible solution: ?

Model-based Reflex agents



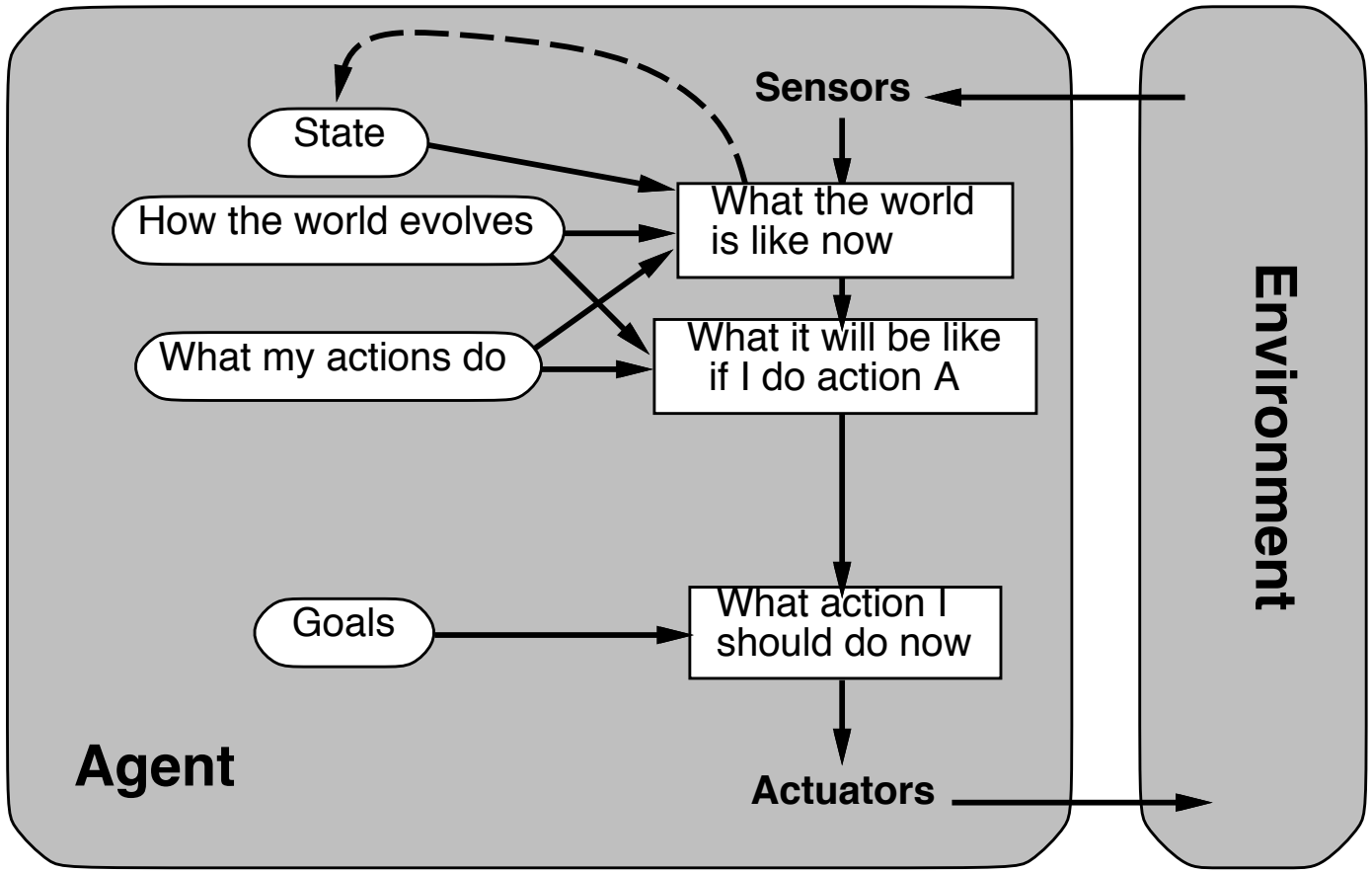
Model-based Reflex agents

function ModelBasedReflexAgent(*percept*) **returns** an action

persistent *state*, the agent's current conception of the world,
transition-model, a description of how the next state depends on
the current state and the action,
sensor-model, a description of how the current world state is reflected
in the agent's percepts
rules, a set of condition-action rules
action, the most recent action, initially none
state ← **UpdateState**(*state*,*action*,*percept*,*transition-model*,*sensor-model*)
rule ← **Rule-match**(*state*, *rules*)
action ← **Rule-action**(*rule*)
return *action*

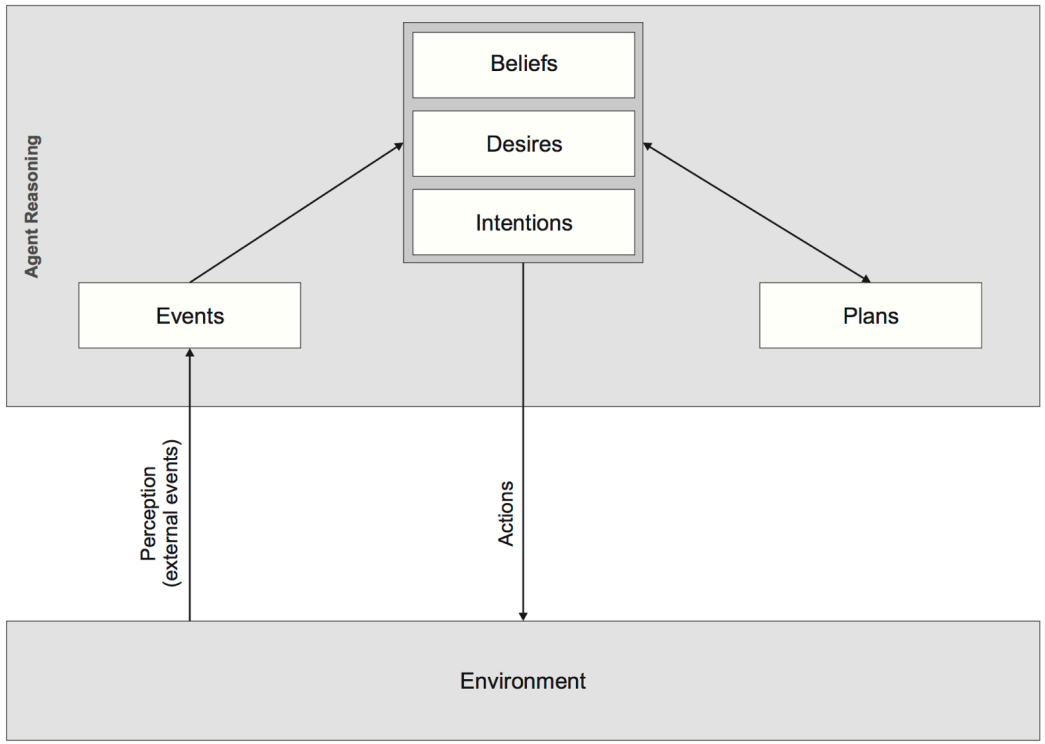
- difficult to exactly determine the current state in partially observable environments - independent from the kind of models used
- hence, may need to "guess" the current situation
- action decision in the same way as the simple reflex agent.

Goal-based agents



Example

- Belief Desire Intention agents (BDI) agents have a **mental state** as the basis for their reasoning.
- Three main mental attitudes: beliefs, desires and, intentions.
- Their reasoning is also called **practical reasoning** - e.g., contrary to deductive reasoning



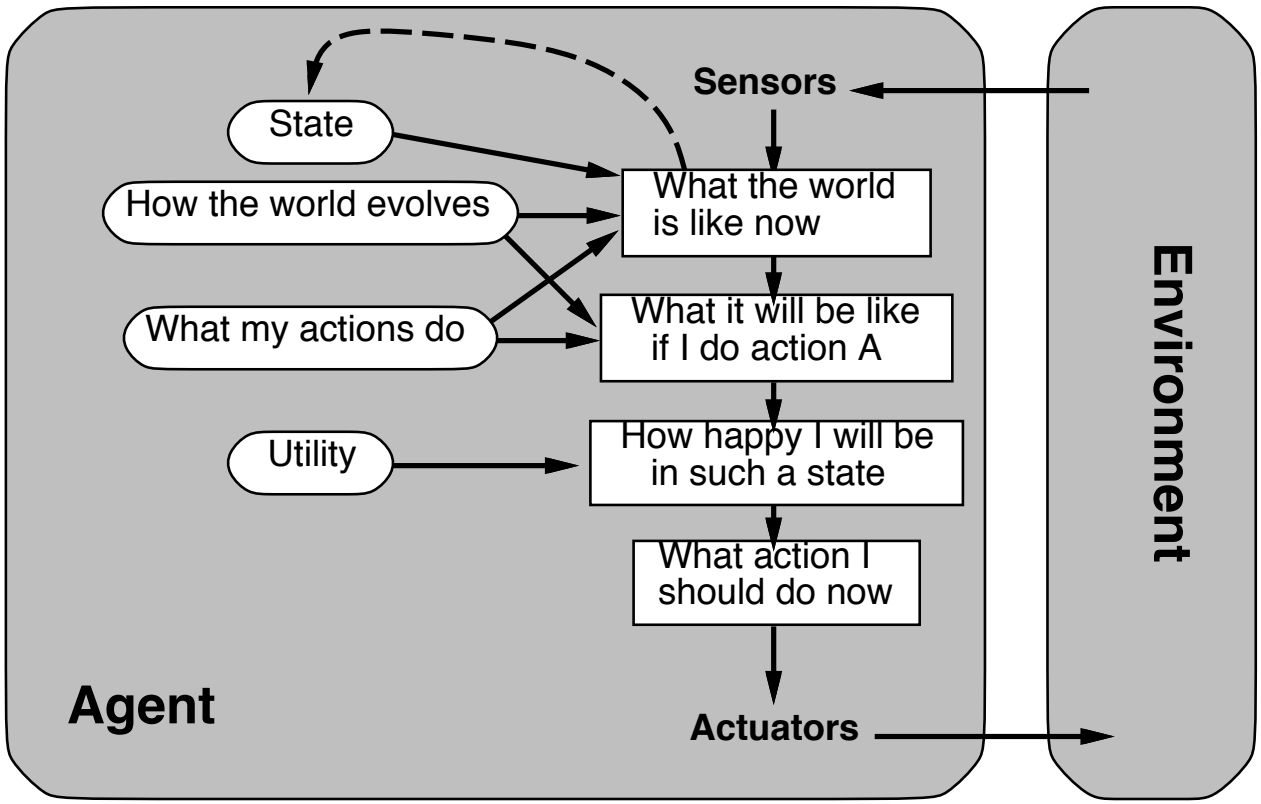
Goal-based agents

- explicit representation of goals: agent knows which states are desirable
- reasoning about goals
- more flexible since knowledge is represented explicitly and can be manipulated
- different reasoning methods than reflex - not condition-action rules and selection of rule/action
- main different from reflex agents: deliberates about future when making decision
- long sequence of actions may be needed to achieve the goal
  - e.g., agents in chapters 3-5 (search) and chapters 10-11 planning
- less efficient than reflex agents - but more flexible

Utility-based agents

Goals provide just a binary happy/unhappy disctinction while utility functions provide a continuous scale

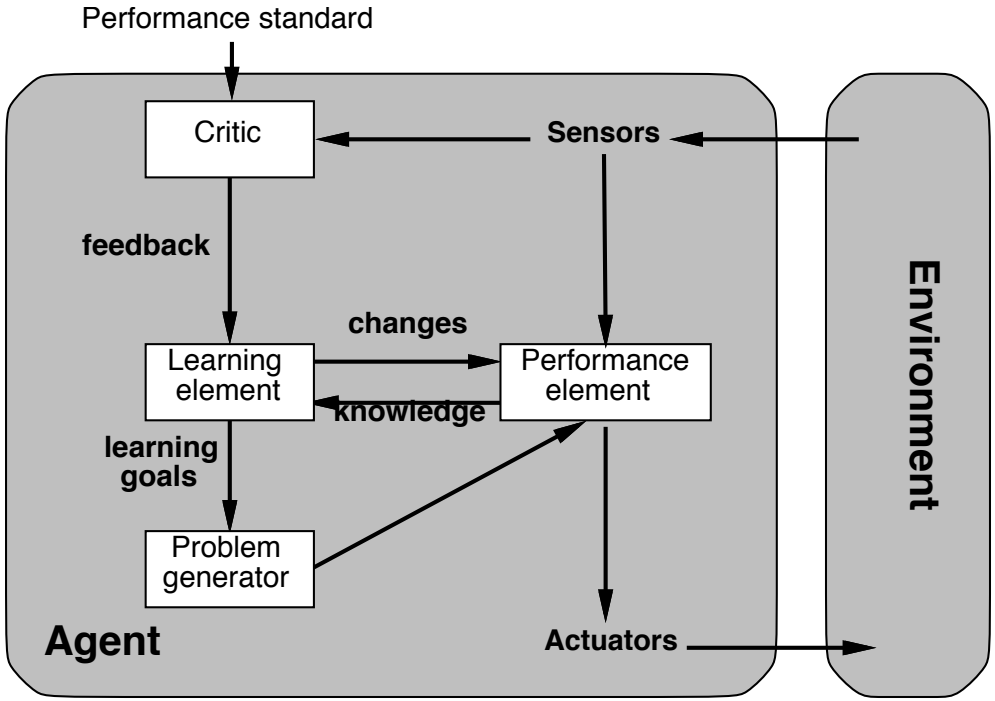
Some goals may be achieved in more than one way, with different utility values



Utility-based agents

- utilities are internalization of the performance measure
- Utility function maps a state (or a sequence of states) onto a real number
- utilities reflect agent's preferences
- not always know the utility of an action/outcome , hence "expected utility"
- agent chooses actions that maximize the expected utility of the outcomes

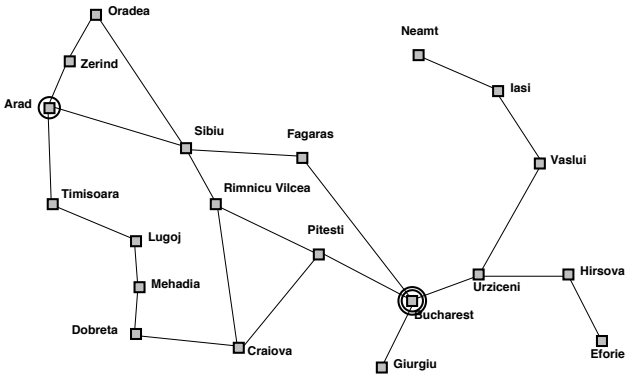
Learning agents



- Critic: evaluates current world state, determines how the performance should be modified
- Learning element: responsible for making improvements
- Problem generator: suggests explorations
- Performance element: responsible for selecting external actions

Next Week: Goal-based Agents

Goal-based agents that use search methods as Agent Function.
You are at the end of your holiday in Romania. Your return ticket is from Bucharest and you are leaving Arad for Bucharest.



- **Uninformed** Search method as Agent Function.
- In an environment:
  - Fully observable (the agent can see the current state of the world),
  - Deterministic (action has a single outcome)
  - Discrete
  - Known environment (knows which states can be reached through which actions - has a map of Romania.

Next week: Problem solving as search, chap 3.