COMP2611

Artificial Intelligence

Lecture 1

Introduction

Overview

This lecture will cover the following:

- Module staff, learning sessions and materials, assessment.
- What is intelligence?
- Approaches to Artificial Intelligence
- Modelling Intelligent Agents

Lecturers

Brandon Bennett

Bragg Building, 2.21-E

B.Bennett@leeds.ac.uk

bb-ai.net

Arash Rabbani

A.Rabbani@leeds.ac.uk

arashrabbani.com

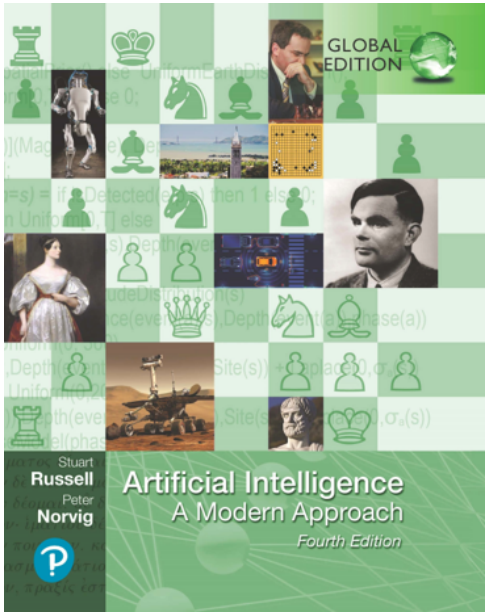
Learning Sessions (2025)

- Lectures in Conference Auditorium 1
  10-11am Mondays and 2pm Fridays

- Lab Sessions in UG Teaching Lab (Bragg 2.05):
  - Monday 3–5
  - Wednesday 11–1
  - Friday 9–11 and 3–5

  These are general teaching labs covering several Level 2 modules. Specific AI help will be given in during the week before assignments are due.

Main Textbook



The main text for this module is:
Artificial Intelligence: a modern approach
by Stuart Russell and Peter Norvig

Currently in its 4th edition (3rd edition would also be fine). Kindle version is convenient (4th ed Kindle currently listed at £36.89.)

Students intending to continue studying AI should certainly invest in this classic AI textbook. But a lot of AI information is also available online. Some pointers to that will be given.

Topics and Schedule

Wk/day	Topic	R&N 4th ed* Ch	Lecturer
1m	Introduction	1	BB
1f	Agents	2	BB
2–3	Search	3, 4	BB
4a	Adversarial Games	5	BB
4f-5	Learning from Data	19	AR
6	Uncertainty/Bayes Nets	12, 13	AR
7	Neural Networks	21	AR
8	Ethics	27	AR
9	Knowledge Represen-	7–11	BB
Easter			
10m	tation and Reasoning	7–11	BB
10f, 11f	Revision		

* Some chapter numbers different in 3rd edition.

Assessment (2025)

The schedule for your assessments is as follows:

CW1:	Search Algorithms	due 21st Feb	20%
CW2:	Machine Learning	due 14th Mar	20%
Exam	All topics	2h in exam period	60%

For CW1 you will be designing and carrying out a lot of testing experiments, a little coding and writing a short report.

CW2 is a coding task applying ML to some data.

Both courseworks will be submitted via Gradescope.

Motivations for AI

Humans have used tools and machines for their benefit for thousands of years. The construction of Intelligent Machines may bring even more benefits, such as:

- Ability to carry out complex but largely uninteresting tasks (such as housework)
- Capability to design and maintain complex infrastructure
- Design and execution of new medical techniques
- Caring for ill or frail people
- New kinds of entertainment
- Aiding creative work such as making art, music or films
- Enhancing humans' physical and/or mental capabilities

Possible Negative Affects of AI

We should also be aware that adverse affects are also possible:

- Causing unemployment
- Making life complex and stressful
- Draining resources
- Introducing unfair bias into decision making
- Manipulating human lives
- Detaching humans from reality
- Competing with, threatening or destroying humans

Are robots AI systems?

Big Dog

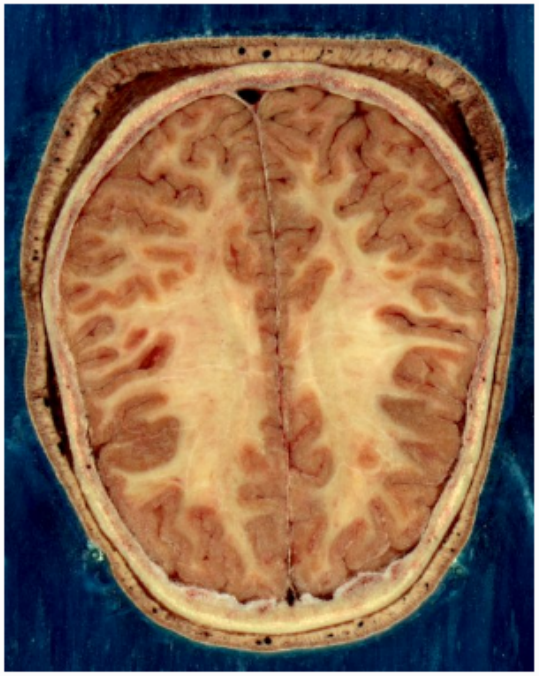
DaVinci

What is Intelligence?

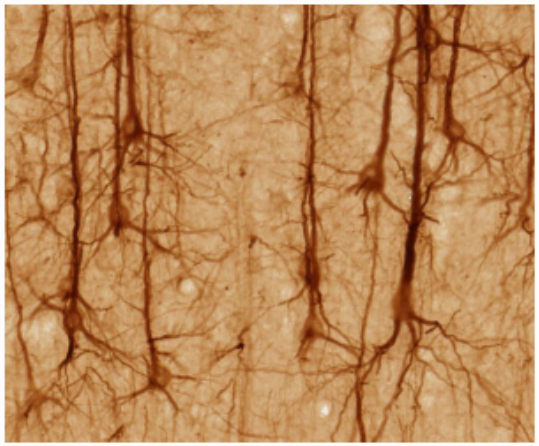
- Ability to plan tasks and solve puzzles
- Capability in mathematics and logic
- Ability to think deeply about complex ideas
- Ability to understand situations and issues
- Language and communication skills
- Problem solving ability
- Ability to act rationally
- Quick wit and humour
- Wisdom and good judgement

What are we trying to copy?

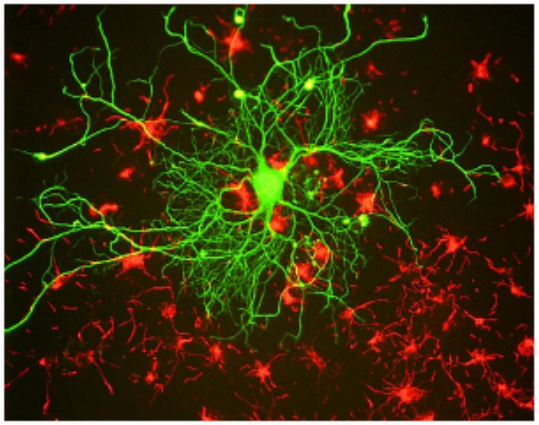
(a) Vertical cross-section of the brain



(b) Horizontal cross-section of the brain



(c) Pyramidal neurons in the cerebral cortex



(d) A cortical neuron (stained)

OR

$$(P \rightarrow Q)$$
$$P$$
$$\therefore Q$$

Levels of Intelligence

Weak (aka Narrow) AI refers to the capability of software (including some existing systems) to perform specific tasks of a kind that one would be describe as requiring intelligence if performed by humans.

Strong (aka General) AI means highly adaptive capability that can be applied to achieving a potentially unlimited range of tasks, which are regarded as involving intelligence. It is generally assumed that humans have this, whereas no current computer system does.

Super-intelligence is intelligence that goes significantly beyond that of any human being. Some theorists believe that this is possible and could potentially be dangerous for humans.

Other Attributes of the Human Mind

The human mind includes many aspects besides intelligence:

- Memory
- Knowledge
- Perception
- Thought
- Self-Awareness
- Consciousness
- Desire
- Free Will

Are these characteristics separate from, or part of, intelligence?

In what way can/could computers exhibit these attributes?

The Turing Test

In his 1950 paper *Computing Machinery and Intelligence* Alan Turing proposed the *imitation game* as a test for intelligence.

This is now known as the *Turing Test*.

The idea of the test is that if a computer behaves like a human (with respect to its linguistic abilities) then it should be considered to be as intelligent as a human:

> *"The computer passes the test if a human interrogator, after posing some written questions cannot tell whether the written responses come from a person or not."*

Thinking *vs.* Acting
Human-like *vs.* Rational

Consideration of the Turing Test brings to light a couple of other distinctions that complicate the definition of intelligence, and may be relevant to how one might go about creating AI.

- Is intelligence essentially related to some mental process of thought; or is it actually (rational) action that defines intelligence?

- Is intelligence a characteristic that is perfectly exemplified by humans; or is there an more abstract attribute of rationality that can take many different forms (and is perhaps only imperfectly present in humans))?

Arguments Against AI

- **Computers are just dumb machines following instructions.** The philosopher John Searl has argued that computational algorithms cannot implement true intelligence because they are simply following rules for manipulating symbols and have no understanding of the meaning of those rules. This view is often stated in the form of the *Chinese Room* argument.

- **The operation of the mind involves quantum mechanics.** Nobel prize wining physicist Roger Penrose is also a well known sceptic of computational AI. In his large book *The Emperor's New Mind*, Penrose argues that the human mind does not operate like a computer but depends on quantum mechanical processes.

Successes So Far

What can existing AI systems do?

- Autonomous planning and Scheduling
- Game Playing
- Autonomous Control
- Diagnosis
- Logistics Planning
- Robotics
- Language translation and generation

What can't existing AI systems do??

Sub-Problems/Fields of AI

Research over the past decades indicates that General AI includes many sub-fields that pose hard problems in themselves:

- Search and Planning
- Inference and Automated Reasoning
- Knowledge Representation
- Reasoning about Probability and Uncertainty
- Computer Vision
- Robotics
- Natural Language Processing
- Natural Language Understanding
- Machine Learning

Intelligent Agents

In *AI: a Modern Approach* (AIMA), Russel and Norvig introduce the basic conceptions of an AI system in terms of the following stages:

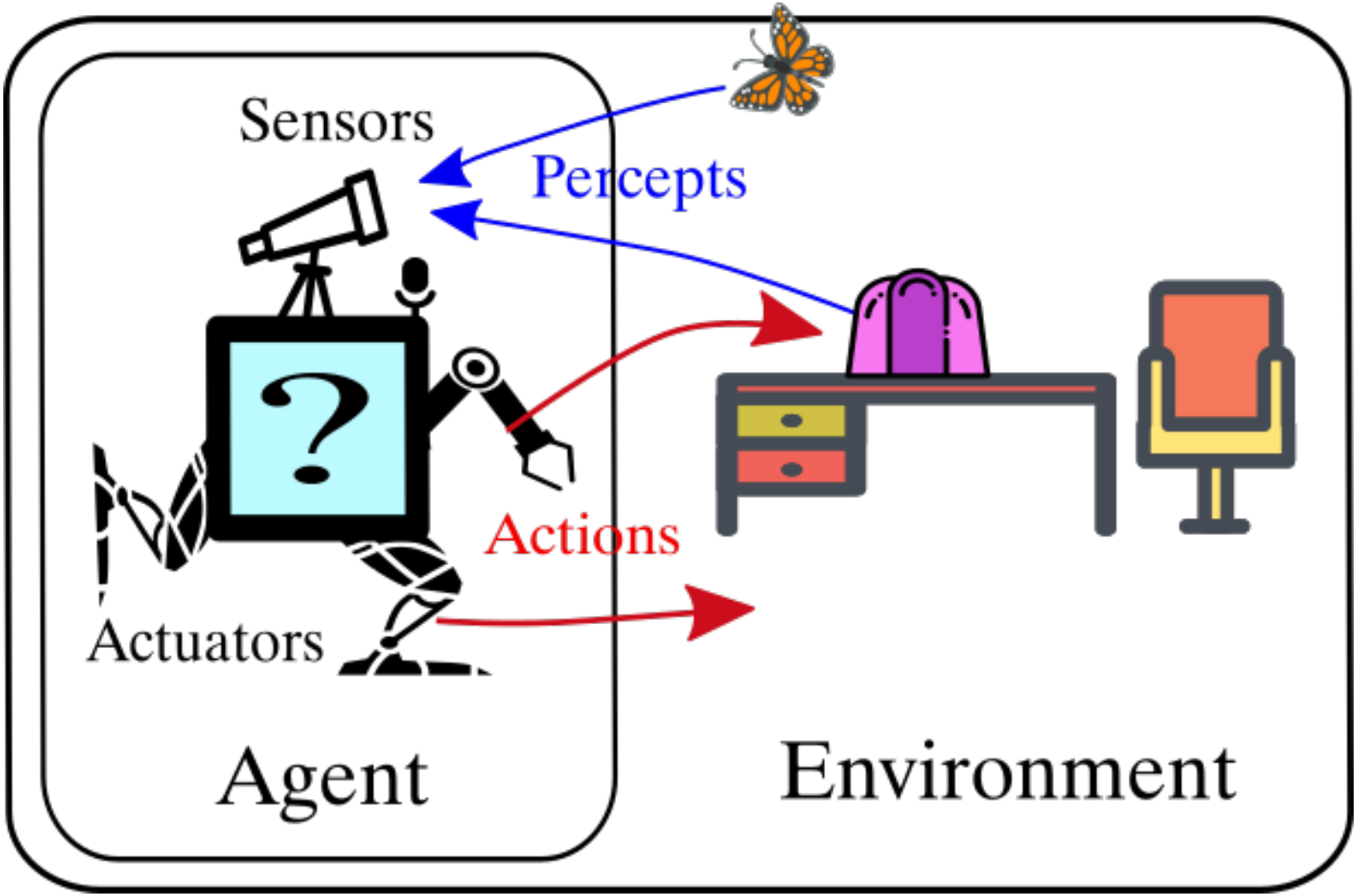
- Define an agent;
- Specify the notion of *rational* behaviour, as conception of intelligence that is subject to objective analysis and investigation;
- Classify different kinds of AI tasks in terms of environment types;
- Classify agent types in terms of the way they interact with and process information about their environment.

Agent and Environment

AIMA (Ch 2) proposes that we analyse intelligence in terms of the interactions between an agent at its environment:

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

Agent-Environment Interaction

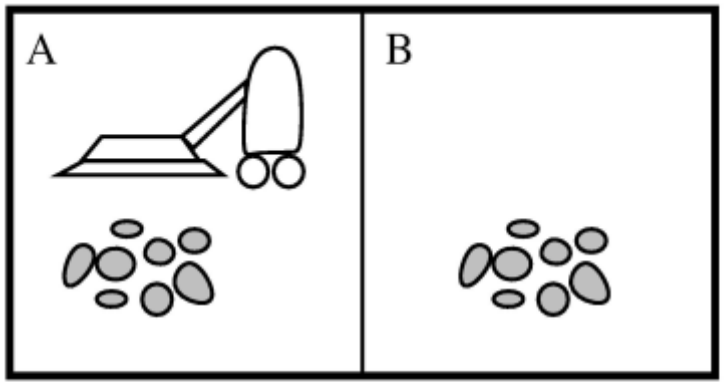


Sensors and Actuators

Different kinds of agent will have different kinds of sensors and actuators:

Agent	Sensors	Actuators
Human	eyes, ears	hands, legs
Robot	cameras, microphones	motors, lasers
Software	keyboard, disk read	display, disk write

Example: a Vacuum-Cleaner Agent



Percepts: location and contents, e.g., [A,Dirty]

Actions: Left, Right, suck, NoOp

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
⋮	⋮

The vac's action is determined by its most recent perception.

The Concept of Rationality

Assumption: A rational agent is one that does the ‘right thing’.

But what is the right thing?

- The agent should do the actions that enable it to be *most successful*
- We need to take into account its *possible actions*
- We need a way to measure success: *a performance measure*
- May be difficult to specify what is best.

A Definition of Rationality

What is rational depends on four things:

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

A *rational agent* is defined as follows:

For each possible percept sequence,
a **rational agent** *should select an action that is expected to maximize its performance measure,*
given the evidence provided by the percept sequence
and whatever built in knowledge the agent has.

Is our vacuum-cleaner agent rational?

We need to be clear about some details:

- The performance measure awards one point for each clean square at each time step over 10000 time steps.
- The geography of the environment is known. Clean squares stay clean and sucking cleans the current square.
- Only actions are left, right, suck, and NoOp
- The agent correctly perceives its location and whether that location contains dirt.

Look back at the percept-action responses that were specified for the robot vacuum cleaner. Is it rational?

Task Environments

It is often useful to identify properties of environments that may be relevant to AI algorithm design:

- Fully observable vs. partially observable.
  – Do sensors detect all aspects of the environment?
- Deterministic vs. stochastic. Is the next state determined completely by the current state and the action taken?
- Episodic vs. sequential. Episodic means that the task can be broken into separate periods that do not affect each other.
- Static vs. dynamic. Dynamic environment can change while agent is deciding what to do.
- Discrete vs. continuous.
- Single agent vs. multi-agent.

Examples

	Crossword	Poker	Medical Diagnosis	Image Analysis
Observable	Fully	partially	partially	fully
Deterministic	yes	strategic*	stochastic	yes
Episodic	sequential	sequential	sequential	yes
Static	yes	yes	dynamic	yes
Discrete	yes	yes	continuous	continuous**
n-Agents	single	multi	single	single

* Poker has a lot of randomness, but having a strategy to deal with that is central to the task.

** Although each image will be a discrete data objects, what it represents is continuous space.

Notice that not all task situations fit neatly into each of these types of classification.

Classification of Agent Types

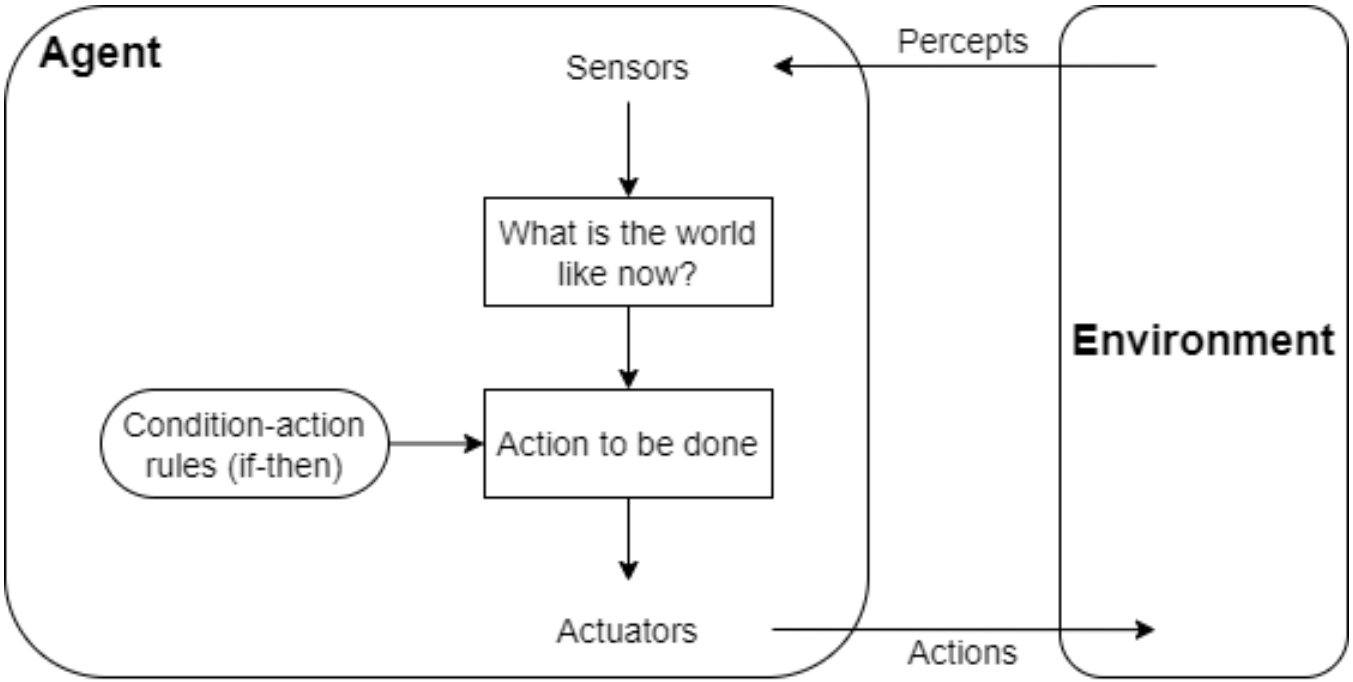
AIMA specifies 5 classes of agents, that have increasingly sophisticated ways of responding intelligently to their environment:

- Simple reflex
- Model-based
- Goal-based
- Utility-based
- Learning

For further details look up in AIMA or the article for "Intelligent Agent in Wikipedia".

Simple Reflex Agent

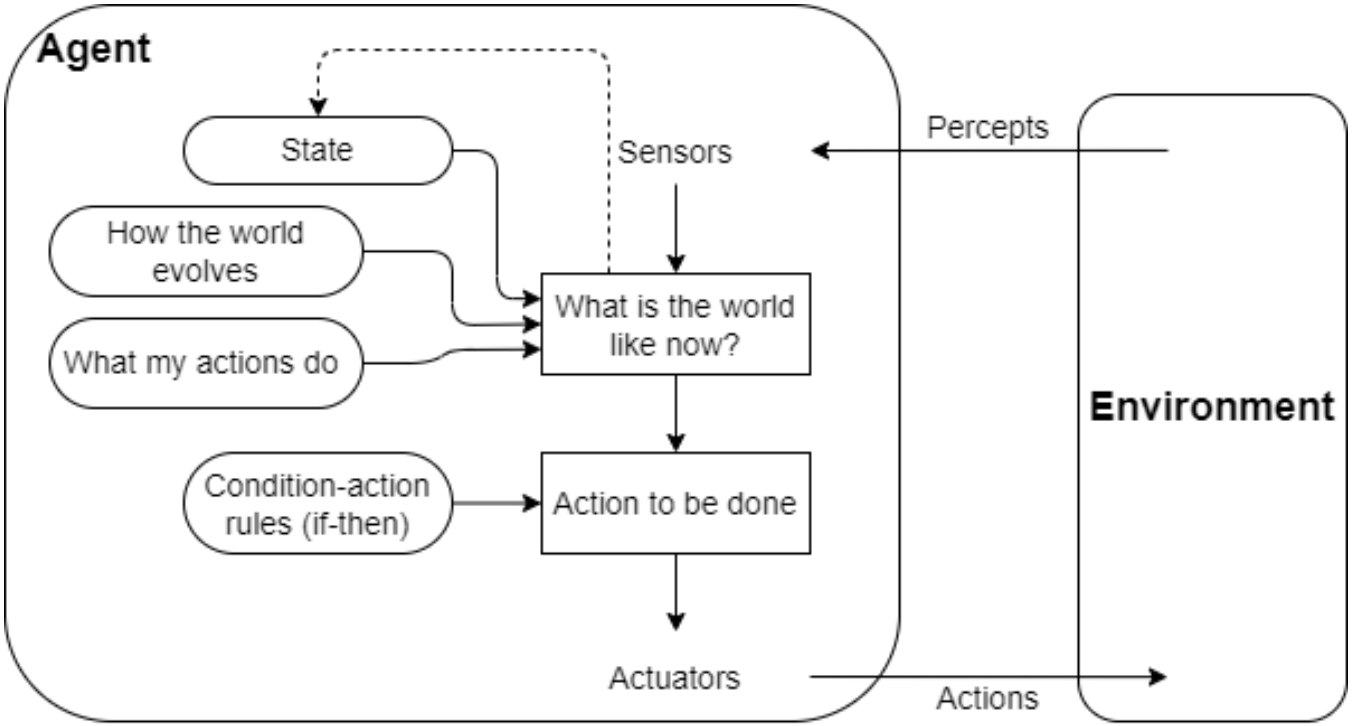
The simplest kind of agent in the AIMA classification is one that can only react to its immediate perception of the environment:



Can you think of a difficulty that a simple reflex agent might get into?

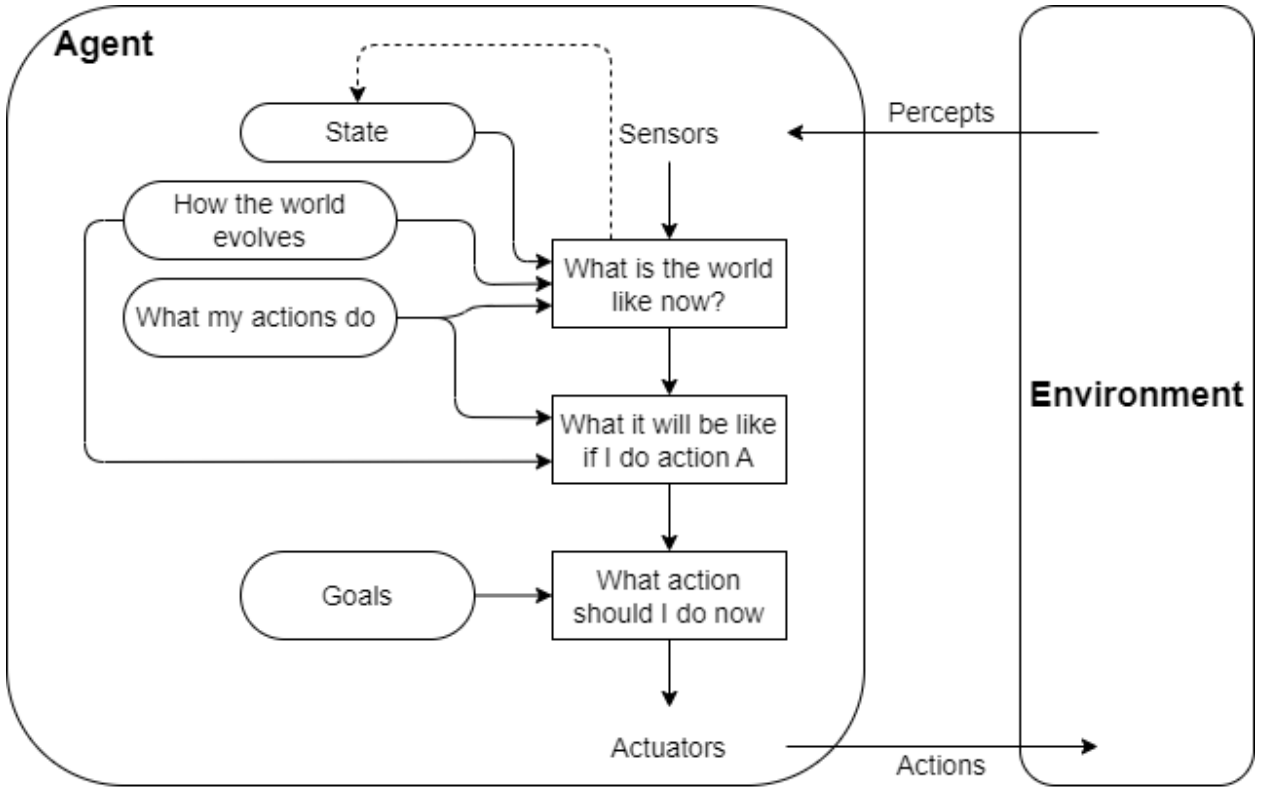
Model-Based Reflex Agent

A model based reflex agent does not just react to its most recent perceptual input. It uses perceptions to build a *model* of its environment and its action rules are based on this accumulated knowledge of the world state:



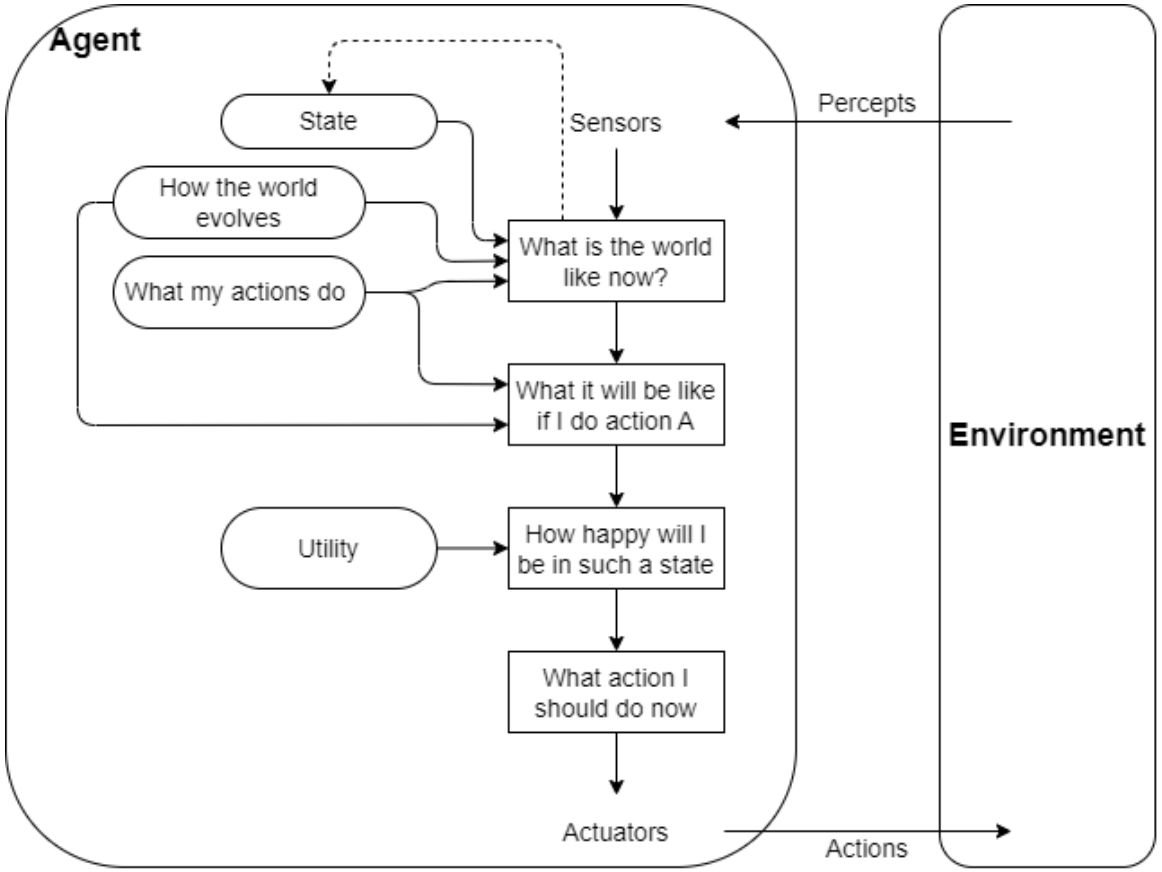
Goal-Based Agent

Instead of simply basing its actions on the current state of the system, a goal based agent also has knowledge of its goals and how they might be achieved. Thus, it can compute plans to achieve its goals:



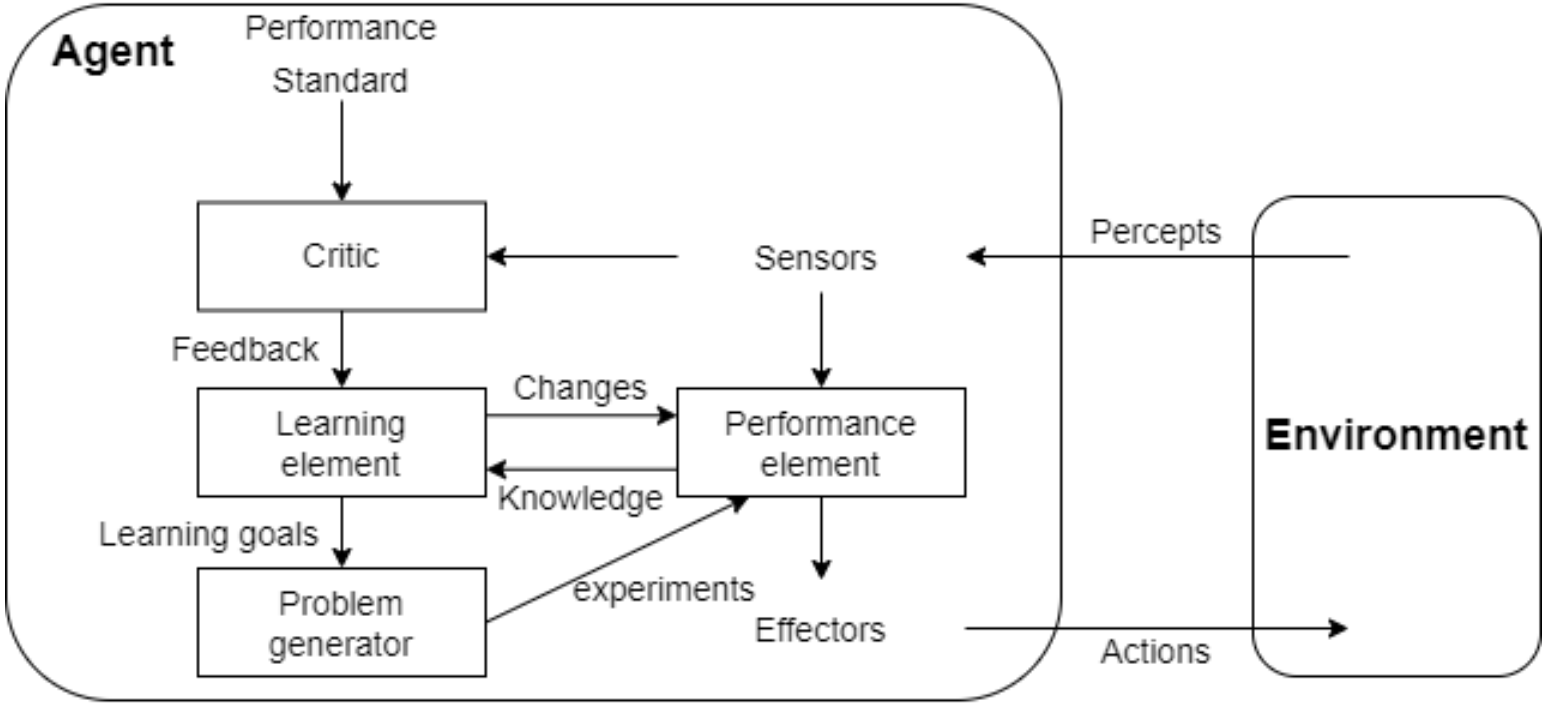
Utility-Based Agent

A *utility-based* agent does not only have knowledge about specific goals but also has more general way of evaluating the benefit of being in a situation, which it can use to guide its actions:



Learning Agent

The *learning agent* is considerably more sophisticated than the previous types. It has a *critic* module to evaluate its own performance, a *learning* module that modify its performance algorithm as well as incorporate new information. And it can even generate 'experimental' actions to help it learn.



Getting Ahead with Search

The next topic we will be studying is *Search*.

You can get some early experience by checking out the website linked below:

`https://qiao.github.io/PathFinding.js/visual/`

This provides an intuitive and very informative demonstration of different search algorithms and how they work for different kinds of path-finding problem.

An exercise using this demonstrator will be published shortly, but you can probably work out the basics just by playing around with the interface.

Summary

This introductory lecture has covered the following:

- the plan for the module: topics and assessment
- motivations, problems and definitions of AI
- the nature of Intelligence
- successes and open issues for AI
- types of task environment for AI problems
- types of AI agent