# COMP3702 - Artificial Intelligence
# Week 1 - Lecture Notes
# Module 0: Introduction

Anindya Sarup

7th August 2020

## Overview of Module 0

What will be covered in Module 0:

1. What is AI?
2. History of Artificial Intelligence
   - A very brief history of it
3. Intelligent Agents
4. Goals of Artificial Intelligence
   - Or what the purpose of developing intelligent agents is
5. Intelligent agents acting in an environment
6. Dimensions of Complexity
   - What are the dimensions of complexity of these agents at the stage of designing.

# Contents

# What do you think AI is?

The study and designs of mechanisms that...

- think like humans?
  - Build something like a brain! But how does a human brain work?
  - It is hard to build something like a brain, because we do not at a physical level know how a brain works, like spikes but we use back propagation which brain does not do.
  - Thus, modelling a brain for building AI is becoming rather hard in practice. It's more the domain of something like cognitive or neuro-science.
  - add image from lecture notes here
- think rationally?
  - Automated reasoning and logic are foundational topics in AI, that is, things that have been examined under the umbrella of AI for over 70 years.
    - But is this a complete answer? As it is unclear if logic really captures the type of knowledge that people have or need help with.
    - Plus it's really hard to search through logical statements. The logical foundations of math are really solid but using that in an automated decision support system is not necessarily going to get you really far because it is difficult to search through logical statements.
    - For example, if you have a ground set of logical statements with binary variables (similar to truth tables 0s & 1s), then all the different ways you can combine these values grows exponentially with different combinations. Thus making it very difficult to setup very complex logical systems and then ascertain whether they are doing it what you expect them to do or helping you make decisions.
    - Logic will be touched upon in Module 2.
  - Is there a need to constrain the working of AI to have the same mechanism like the brain or can we have it think like a brain?
  - Thinking ”rationally” has not been defined yet, but will be defined later on
- act like humans?
  - The Turing test: can a human tell if a computer is a computer?
  - Setting the threshold to be whether a machine can be distinguished from a human is a novel goal.
  - The Turing Test - Alan Turing (1950)
    - In the Turing test, the computer is asked questions by a human interrogator.
    - The interrogator asks questions to either a computer or a human. The questions are usually tricky, thus their correct answers are associated to being intelligent. Such questions are usually easy for a human to answer but are tricky for a machine to answer.
    - Computer passes the test if the interrogator cannot tell whether the responses come from a human or a computer.

    * The Turing test simplifies the question "is the machine intelligent" into "can the machine imitate a human?"
    * Add image from lecture slides.
  - Lets stop and think: Do we really want computers to act like humans?

- act rationally?
  - Perhaps, what we really want are machines that ***act rationally***, that make better decisions for the sort of problems that they are very good at solving. This is known as ***acting rationally***
  - Underpinning the idea of acting rationally is what is known as ***Intelligent Agents*** (approach taken in R&N and P&M texts)
  - That is, an Intelligent, rational, autonomous agent, a computational device or system that can make decisions on your behalf or control systems in the absence of a human or even provide advice on what it is you should be doing or what you might want to do in a given situation based on the way it reasons and learns about the world.
- Not sure that this truly captures the variety of AI research going on right now, but it is a good place to start.
- *In this course, we are going to be building computational agents that act rationally in given environments and can learn from those environments.*

## So what really is AI?

- According to the Association for the Advancement of Artificial Intelligence (AAAI) offers this on its home page:**AI is "the scientific understanding of the mechanisms underlying thought and intelligent behavior and their embodiment in machines"**
- Poole and Mackworth say AI is **"the synthesis and analysis of computational agents that act intelligently."** This is in-sync with what their textbook on AI is about.
- According to our lecturer Archie: **"AI is the study and development of algorithms for solving problems that we typically associate with intelligence."** This leaves the interpretation of intelligence to be open.

In general, AI is a disperse collection of topics. In this course, we address core method and models used in AI research and practice, many of which have found a wide-spread use application and as building-blocks in more sophisticated AI systems. *That's the high level mapping of AI to what we do in this course.*
All components of AI are not addressed in this course, specially topics like machine learning. Although, there is some usage of reinforcement learning and a little bit of neural networks.

# (A very brief) History of Artificial Intelligence

*For more on the history and development of AI, read chapter 1 of R&N or P&M*

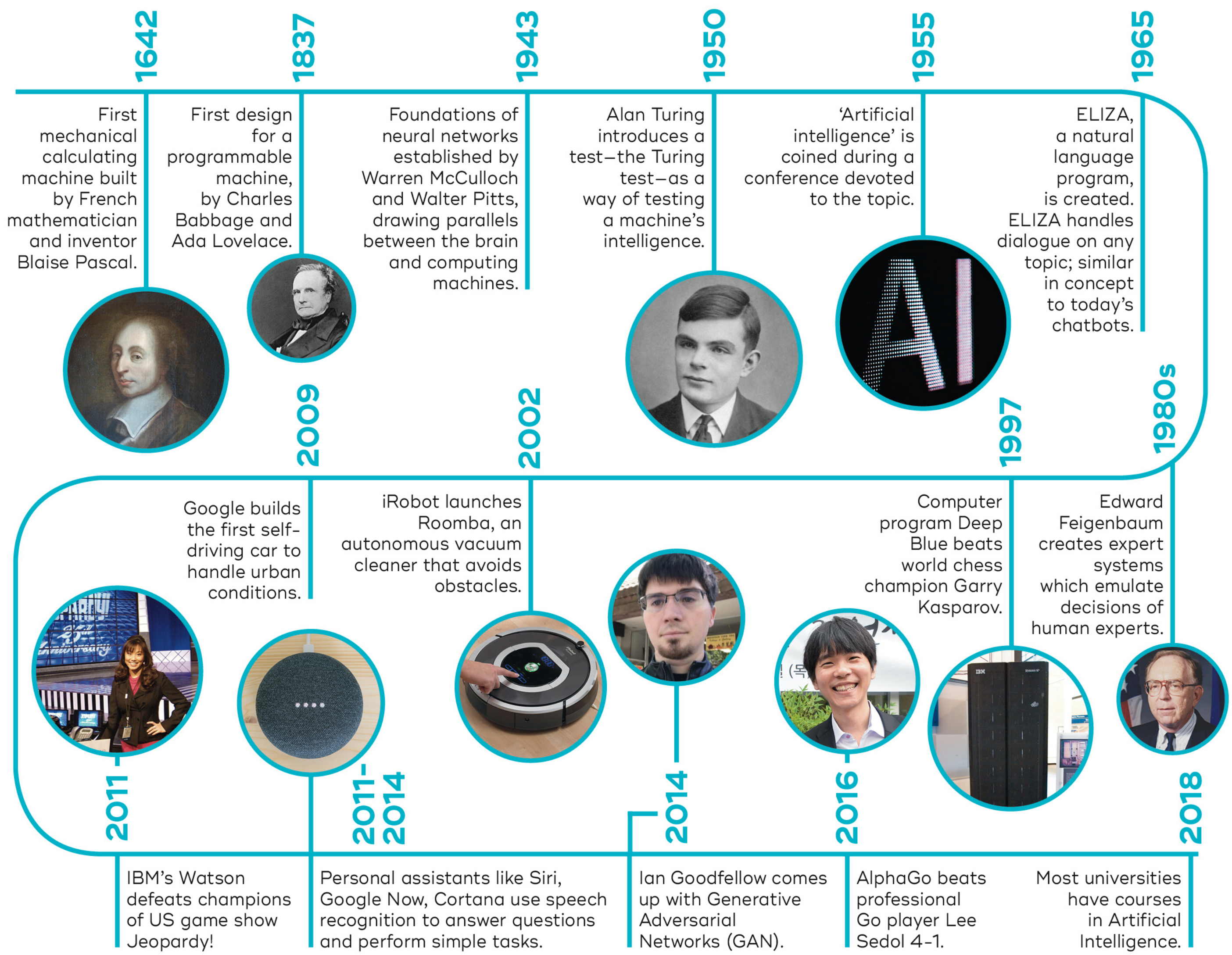


Figure 1: This diagram shows the development of the more cutting edge of AI tools proposed or deployed.

# Intelligent Agents

## What is an intelligent computational agent?

- An **agent** is something that **acts** in an environment.
  - Usually an agent has some autonomy, that is, it can follow it's own goals and it has it's own reasoning processes.
  - An agent is quite different from an object in OOP,
    * For example, an object receives commands or requests, has internal processes that it executes and then returns values.
    * Whereas the key difference between an agent and an object is that an agent has a sense of autonomy (that is, doing something for its own good).
    * So you can ask an agent to do something for you and it does not align with its goals or desires, it might not do what you requested for.
    * Thus, the sense of autonomy is the defining feature that separates an agent (or agency) from an object.
    * This difference is important because although agents will be developed using objects in python, they are not essentially objects because they will have a greater sense of autonomy and freedom to act and pursue their own goals.
- An agent acts **intelligently** if...
  - Its actions are appropriate for its goals and circumstances
    * That is, at base level it is doing something sensible
  - It is flexible to changing environments and goals
    * The idea that an agent can adapt to changing situations and different instructions.
  - It learns from experience
    * It has a way of incorporating new knowledge and new information into its decision making
  - It makes appropriate choices given perceptual and computational limitations.
    * It can actually implement actions that achieve the goals it wants to attain.
    * Even if it has poor sensing capabilities and computational power. That is shortage of time and high computational hardware.

## Examples of agents

*Not necessarily computational agents*

Things like...

- **Organisations:** Microsoft, Facebook, Government of Australia, UQ, ITEE,...
  - These all have agency.
  - They can act, learn, and reason.
  - They do what achieves their ends and not necessarily what others tell them to.

- **People:** teacher, doctor, stock trader, engineer, researcher, travel agent, farmer, waiter,...
  - Any of these classes of occupations
  - The people that are doing these jobs have agency. They choose what they do, they reason and they learn and thus try to achieve their goals.
- **Computers and devices:** air-conditioner thermostat, airplane controller, network controller, movie recommendation system, tutoring system, diagnostic assistant, robot, GPS, navigation app, Mars rover...
  - Some computers and devices, if they are appropriately setup can learn and reason from experience.
  - They can adapt to change in situations and can be set for new different goals.
- **Animals:** dog, mouse, bird, insect, worm, bacterium, bacteria,...
  - They can adapt to change in situations too, thus they have some agency as well.
  - They evolve, which is a form of learning.
- book (?), sentence (?), word (?), letter (?)
  - Can a book or article do things?
  - Convince? Argue? Inspire? Cause people to act different? Learn from experience?
  - These are not agents

# Goals of Artificial Intelligence

## What are the goals of Artificial Intelligence?

- **Scientific Goal:** To understand the principles that make intelligent behavior possible in natural or artificial systems.
  - So, essentially, what are the principles of intelligence that we can then use to engineer artificial intelligence or intelligent systems.
    * Analyze natural and artificial agents.
    * Formulate and test hypotheses about what it takes to construct intelligent agents
    * Design, build, and experiment with computational systems that perform tasks that require intelligence
- **Engineering goal:** design useful, intelligent agents.
  - Always make the "best" decision given the available resources (knowledge, time, computational power and memory)
    * The catch here, is what does "best" entail?
      · Is it maximizing the profit of the company you're working? (for example, Facebook) or are there ethical considerations that constrain what you should be doing?
      · Thus, best is a difficult term to pin down but is quite fundamental to the way we design computational agents.
  - **Best:** Maximize certain performance measure(s), usually represented as a ***utility function***.
    * A utility function is a just a way of mapping outcomes to a score, effectively.
    * In Economics, utility functions are usually associated with either actions that provide one with pleasure like eating a meal, going to see a movie, etc. weighted against the cost that you have to pay to do those things. The surplus you have here is your "utility". Its a construct, a way of thinking about things. But its really a way of mathmetising these sorts of problems so that we can use computational resources to solve them.
    * More on this throughout the semester, with increasing level of details.

## In this class...

- We are interested in building software systems (called agents) that behave rationally.
- i.e. Systems that accomplish what they are supposed to do, well, given the available resources.
  - These resources can be information, computational, or time based.
- Don't worry about how close the systems resemble humans and about philosophical questions on what "intelligence" is (not that we are not interested in this!)
- But we may use inspirations from humans or other "intelligent" beings and systems.

# Intelligent agents acting in an environment

Here we will try to get more specific and precise about what we are trying to achieve through this course.
We are trying to define an intelligent that acts in an environment. Now this contains a few terms that we need to define concretely to be able to actually code in order to automate the reasoning, acting and learning in the decision making. Now we will what we mean by **intelligent agents, acting, and an environment.**

## Recall our goal: To build a useful, intelligent agent

To start with:

- Computers perceive the world using sensors.
- Agents maintain models/representations of the world and use them for reasoning.
- Computers can learn from data.

So, to achieve our goals, we need to define our **"agent"** in a way that we can *program* it:

- The problem of constructing an agent is usually called the ***agent design problem.***
    - This design problem is the one that provides the agent sufficient autonomy to be flexible, adapt, learn, and reason. Effectively an engineering program.
- Simply, it's about defining the components of the agent, so that when the agent acts rationally, it will accomplish the task it is supposed to perform, and do it well.

***Some important things we don't address in this course***

- **User Interaction:** Making the agents interact comfortable with humans is a substantial challenge for AI developers.
- **Ethics of AI:** AI applications can impact society in both positive and negative ways.

## Agents acting in an environment: inputs and output

An agent performs an action in the environment; the environment generates a percept or stimuli. The percept generated by the environment may depend on the sequence of actions the agent has done.

The term stimuli is often interchangeable with percept or feedback.

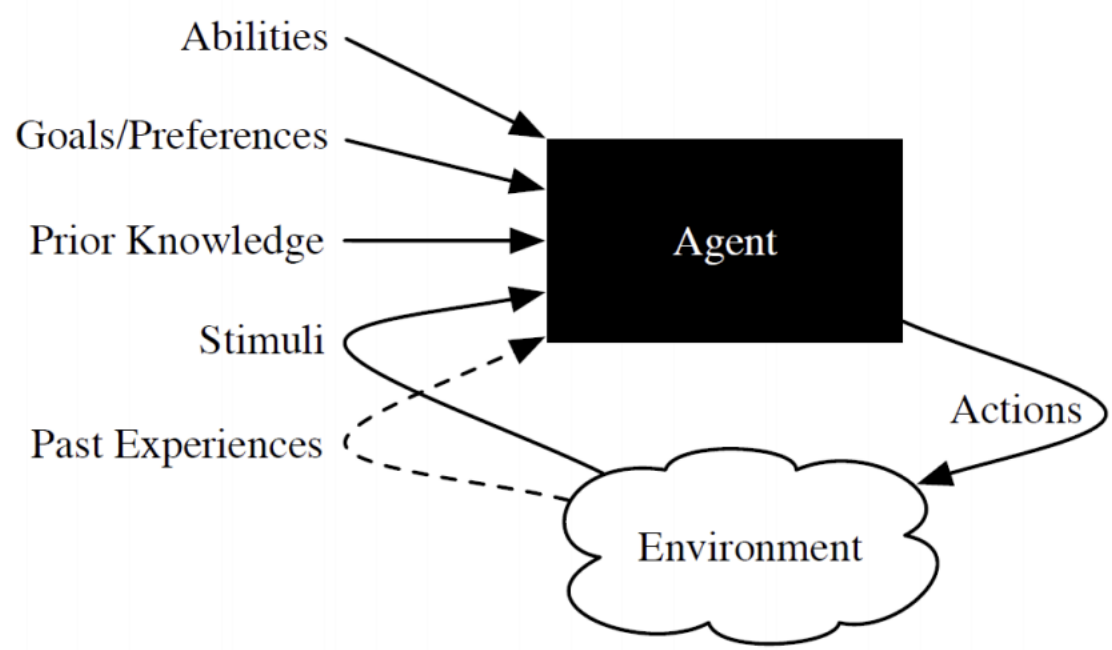


Figure 2: An agent undertaking actions which are fed into the environment with outputs from the environment

In addition to the history/past experiences and stimuli, an agent is equipped with knowing what its abilities, goals/preferences and some prior knowledge about how its environment works or how the agent itself works.
All of these are fed into the agent and incorporated into its decision making process and then agent has the opportunity to act in the environment.
From the environment it receives feedback and it adjusts its reasoning or decision making in a way that results in adjusting its actions and repeating the loop.

- **Abilities:** the set of possible actions it can perform
- **Goals/Preferences:** what it wants, its desires, its values
- **Prior Knowledge:** what it knows and believes initially, what it doesn't get from experience
- **History of stimuli:**
  - (current) **stimuli:** what it receives from environment now (observations, percepts)
  - **past experiences:** what it has received in the past

## Examples of agents

### Autonomous car

- **abilities:** steer, accelerate, brake
- **goals:** safety, get to destination, timeliness...
- **prior knowledge:** street maps, what signs mean, what to stop for...
- **stimuli:** vision, laser, GPS, voice commands...
- **past experiences:** how braking and steering affects direction and speed...

### Air-conditioner thermostat and controller agent

- **abilities:** turn air-conditioner on or of
- **goals:** conformable temperature, save energy, save money
- **prior knowledge:** 24 hour cycle, weekends
- **stimuli:** temperature, set temperature, who is home, outside temperature, rooftop PV generation...
- **past experiences:** when people come and go, who likes what temperature, building thermal dynamics...

# Dimensions of Complexity

Here we start unpacking why some problems are much harder than others.

## Dimensions of complexity in an agent design (P&M Ch 1.5)

- Research proceeds by making simplifying assumptions (to get a handle on the situation), and gradually reducing (removing) these assumptions so that your simple model moves more towards the real world while still being based on sound principles.
- Each simplifying assumption gives or represents a dimension of complexity. Moving from simpler dimensions to more complex characteristics to match the real world better.
  - Multiple values in a dimension: from simple to complex
  - Simplifying assumptions can be relaxed in various combinations
  - This results in different classes of problems. (We will go through possibly ten different classes of problems in this course).
- Much of the history of AI can be seen as starting from the simple and adding in complexity in some of these dimensions.

## From P & M Ch 1.5

| Dimension | Values |
|---:|---|
| Modularity: | flat, modular, hierarchical |
| Planning horizon: | non-planning, finite stage, indefinite stage, infinite stage |
| Representation: | states, features, relations |
| Computational limits: | perfect rationality, bounded rationality |
| Learning: | knowledge is given, knowledge is learned |
| Sensing uncertainty: | fully observable, partially observable |
| Effect uncertainty: | deterministic, stochastic |
| Preference: | goals, complex preferences |
| Number of agents: | single agent, multiple agents |
| Interaction: | offline, online |

Figure 3: Definitions of dimensions according to P&M

The example of values and dimensions above in the figure not all-encompassing but are enough to provide an idea of the types of dimensions and the values they can take on.

## Modularity

By Modularity, we are talking about the structure of the overall big problem we're trying to solve.

- Model at one level of abstraction: **flat**
  - Typically the problems we're going to address in this course are going to be flat.
  - That is, in the big problem there will only be a single sub problem
  - But for more complex systems we have a hierarchy of problems.
- Model with interacting modules that can be understood separately: **modular**
- Model with modules that are (recursively) decomposed into modules: **hierarchical**
- Flat representations are adequate for simple systems
- Complex biological systems, computer systems, organizations are all hierarchical.

**Is the environment continuous or discrete?**

- A *flat* description is typically either **continuous** (exclusive-or **discrete**.
- Hierarchical reasoning is often a hybrid of continuous and discrete

## Planning horizon

...how far the agent looks into the future when deciding what to do.

- **Static:** world does not change
- **Finite stage:** agent reasons about a fixed finite number of time steps
- **Indefinite stage:** agent reasons about a finite, but not predetermined, number of time steps
- **Infinite stage:** the agent plans for going on forever (i.e. process oriented rather than an end goal as such.)

## Representation

Much of modern AI is about finding compact representations and exploiting the compactness for computational gains.
Representation becomes a really key element to how you set up the problem that you're trying to address.
An agent can reason in terms of:

- **Explicit states:** a state is one way the world could be
- **Features** or **Propositions.**
  - States can be described using features.
  - 30 binary features can represent 230 = 1, 073, 741, 824 states.

- **Individuals** and **Relations**

    Are we dealing with an individual agent or are we dealing with relations couldn't understand the word between or are we dealing with abstract representations of agents that just comes across like an individual agent.There is a feature for each relationship on each tuple of individuals. Often an agent can reason without knowing the individuals or when there are infinitely many individuals.

## Computational Limits

- **Perfect rationality:** the agent can determine the best course of action, without taking into account its limited computational resources. That is, effectively no practical limit on the computation available to the agent.
- **Bounded rationality:** the agent must make good decisions based on its perceptual, computational and memory limitations.

## Learning from Experience

Whether the model is fully specified a priori:

- **Knowledge is given.**
- **Knowledge is learned from data or past experience.**

In practice, there is always some mix of prior (innate, programmed) knowledge and learning (nature vs nurture). For example, The way you encode the mapping from percepts to learning routines is an example of prior knowledge being incorporated into the design of the agent.

## Uncertainty

We'll learn how to deal with uncertainty using probability and how we incorporate that into decision making at the start of module 3.
There are two dimensions for uncertainty:

- **Sensing uncertainty** or noise perception
- **Effect uncertainty**
    - Uncertainty about how your action affect the environment?

In this course, we restrict our focus to **probabilistic** models of uncertainty. Why?

- Agents need to act even if they are uncertain.
- Predictions are needed to decide what to do:
    - Definitive predictions: you will be run over tomorrow
    - Point probabilities: probability you will be run over tomorrow is 0.002 if you are careful and 0.05 if you are not careful
    - Probability ranges: you will be run over with probability in range [0.001,0.34]
- Acting is gambling: agents who don't use probabilities will lose to those who do.
- Probabilities can be learned from data and prior knowledge.

## Sensing uncertainty

Whether an agent can determine the state from its stimuli:

- **Fully-observable system:** the agent can observe perfectly the state of the world. We'll stick with this type of observable system for this course.
- **Partially-observable:** there can be a number states that are possible given the agent's stimuli.

## Effect uncertainty

If an agent knew the initial state and its action, could it predict the resulting state?

The dynamics can be:

- **Deterministic:** the resulting state is determined from the action and the state
- **Stochastic:** there is uncertainty about the resulting state.

## Preferences

What does the agent try to achieve?

- **Achievement goal** is a goal to achieve. This can be a complex logical formula
- **Complex preferences** may involve trade-offs between various desiderata, perhaps at different times.
  - **Ordinal** only the order matters
  - **Cardinal** absolute values also matter

**Examples:** coffee delivery robot, medical doctor

## Number of agents

Are there multiple reasoning agents that need to be taken into account?

- **Single agent:** reasoning: any other agents are part of the environment.
- **Multiple agent:** reasoning: an agent reasons strategically about the reasoning of other agents.

Agents can have their own goals: cooperative, competitive, or goals can be independent of each other

## Interaction

When does the agent reason to determine what to do?

- **Reason offline:** before acting. For example, planning of the train schedule
- **Reason online:** while interacting with environment. For example, the roomba robot not knowing where the chairs are currently present in the room.

## Dimensions of complexity in agent design

| Dimension | Values |
|---|---|
| Modularity: | flat, modular, hierarchical |
| Planning horizon: | non-planning, finite stage, indefinite stage, infinite stage |
| Representation: | states, features, relations |
| Computational limits: | perfect rationality, bounded rationality |
| Learning: | knowledge is given, knowledge is learned |
| Sensing uncertainty: | fully observable, partially observable |
| Effect uncertainty: | deterministic, stochastic |
| Preference: | goals, complex preferences |
| Number of agents: | single agent, multiple agents |
| Interaction: | offline, online |

Figure 4: Definitions of dimensions according to P&M

## Example problem class: State-space search (Module 1)

| Dimension | Values |
|---|---|
| Modularity | flat, modular, hierarchical |
| Planning horizon | non-planning, finite stage, indefinite stage, infinite stage |
| Representation | states, features, relations |
| Computational limits | perfect rationality, bounded rationality |
| Learning | knowledge is given, knowledge is learned |
| Sensing uncertainty | fully observable, partially observable |
| Effect uncertainty | deterministic, stochastic |
| Preference | goals, complex preferences |
| Number of agents | single agent, multiple agents |
| Interaction | offline, online |

Figure 5:

### Example problem class: Deterministic planning using CSP (Module 2)

| Dimension | Values |
|---|---|
| Modularity | flat, modular, hierarchical |
| Planning horizon | non-planning, finite stage, indefinite stage, infinite stage |
| Representation | states, features, relations |
| Computational limits | perfect rationality, bounded rationality |
| Learning | knowledge is given, knowledge is learned |
| Sensing uncertainty | fully observable, partially observable |
| Effect uncertainty | deterministic, stochastic |
| Preference | goals, complex preferences |
| Number of agents | single agent, multiple agents |
| Interaction | offline, online |

Figure 6:

### Example problem class: Markov decision processes (MDPs, Module 3)

| Dimension | Values |
|---|---|
| Modularity | flat, modular, hierarchical |
| Planning horizon | non-planning, finite stage, indefinite stage, infinite stage |
| Representation | states, features, relations |
| Computational limits | perfect rationality, bounded rationality |
| Learning | knowledge is given, knowledge is learned |
| Sensing uncertainty | fully observable, partially observable |
| Effect uncertainty | deterministic, stochastic |
| Preference | goals, complex preferences |
| Number of agents | single agent, multiple agents |
| Interaction | offline, online |

Figure 7:

### Example problem class: Reinforcement learning (Module 4)

| Dimension | Values |
|---:|---|
| Modularity | flat, modular, hierarchical |
| Planning horizon | non-planning, finite stage, indefinite stage, infinite stage |
| Representation | states, features, relations |
| Computational limits | perfect rationality, bounded rationality |
| Learning | knowledge is given, knowledge is learned |
| Sensing uncertainty | fully observable, partially observable |
| Effect uncertainty | deterministic, stochastic |
| Preference | goals, complex preferences |
| Number of agents | single agent, multiple agents |
| Interaction | offline, online |

Figure 8:

### Example problem class: Classical game theory (Module 5)

| Dimension | Values |
|---:|---|
| Modularity | flat, modular, hierarchical |
| Planning horizon | non-planning, finite stage, indefinite stage, infinite stage |
| Representation | states, features, relations |
| Computational limits | perfect rationality, bounded rationality |
| Learning | knowledge is given, knowledge is learned |
| Sensing uncertainty | fully observable, partially observable |
| Effect uncertainty | deterministic, stochastic |
| Preference | goals, complex preferences |
| Number of agents | single agent, multiple agents |
| Interaction | offline, online |

Figure 9:

### The real world: Humans

| Dimension | Values |
|---|---|
| Modularity | flat, modular, hierarchical |
| Planning horizon | non-planning, finite stage, indefinite stage, infinite stage |
| Representation | states, features, relations |
| Computational limits | perfect rationality, bounded rationality |
| Learning | knowledge is given, knowledge is learned |
| Sensing uncertainty | fully observable, partially observable |
| Effect uncertainty | deterministic, stochastic |
| Preference | goals, complex preferences |
| Number of agents | single agent, multiple agents |
| Interaction | offline, online |

Figure 10: