Assignment 1

Question 1

Define artificial intelligence (AI). Find at least 3 definitions of AI that are not covered in the lecture.

- "Artificial intelligence (AI) is a wide-ranging branch of computer science concerned with building smart machines capable of performing tasks that typically require human intelligence." - Built In
- "Artificial intelligence (AI), the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings." - Britannica
- "At its simplest form, artificial intelligence is a field, which combines computer science and robust datasets, to enable problem-solving. It also encompasses sub-fields of machine learning and deep learning, which are frequently mentioned in conjunction with artificial intelligence. These disciplines are comprised of AI algorithms which seek to create expert systems which make predictions or classifications based on input data." - IBM

Question 2

What is the Turing test, and how it is conducted?

The main purpose of the Turing test (originally called the Imitation Game) is to determine whether or not a computer is capable of mimic human responses to the point where a person would not be able to answer if the respondent was a machine or human. For achieving this, the computer has to think like a human being and respond like one.

The test goes as follows. You have a questioner which interrogates the respondents, one human and one machine, within a specific subject area. The respondents take turns in answering (have to be 50/50, but not in order), and the questioner is limited to their written answers for deciding if the answers is from a human or not. The test is repeated a great amount of times, and if the questioner is able to guess correct half of the time or less, the computer is considered to have artificial intelligence because the questioner considers it as just as human as the other respondent. Accordingly, the test is subjecting whether a human can distinguish answers from a machine and a human.

Question 3

What is the relationship between thinking rationally and acting rationally? Is rational thinking an absolute condition for acting rationally?

The Turing test approach is testing whether a computer acts humanly. Some have defined intelligence in terms of fidelity to human performance, while others prefer an abstract, formal definition of intelligence called rationality - doing the "right thing". Some consider intelligence to be a property of internal thought process, while others focus on intelligent behavior, an external characterization. From these two dimensions - human vs rational and thought vs behavior - there are four possible combinations. The pursuit of human-like intelligence must be in part an empirical science related to psychology, and a rationalist approach involves a combination of mathematics and engineering.

	Humanly	**Rationally**
Thinking	Thinking humanly — cognitive modeling. Systems should solve problems the same way humans do.	Thinking rationally — the use of logic. Need to worry about modeling uncertainty and dealing with complexity.
Acting	Acting humanly — the Turing Test approach.	Acting rationally — the study of rational agents: agents that maximize the expected value of their performance measure given what they currently know.

Further on, we can discuss the relationship between thinking rationally and acting rationally. Thinking rationally can often be synonymous to thinking logically. The problem with basing intelligence on logic is that this require that everything in the world is certain - which it is not. The theory of probability fills this gap, allowing rigorous reasoning with uncertain information. In principle, it allows the construction of a comprehensive model of rational thought, leading from raw perceptual information to an understanding of how the world works to predictions about the future. What it does not do, is generate intelligent behavior. For that, we need a theory of rational action. Rational thought, by itself, is not enough.

Acting rationally is as simple as doing the right thing. The problem is that we need to make one important refinement to the standard model to account for the fact that perfect rationality—always taking the exactly optimal action—is not feasible in complex environments.

Rational thinking is not an absolute condition for acting rationally as there are ways of acting rationally that cannot be said to involve inference. For example, recoiling from a hot stove is a reflex action that is usually more successful than a slower action taken after careful deliberation.

Sure, making correct inferences is *sometimes* part of acting rationally, because one way to act rationally is to deduce that a given action is best and then to act on that conclusion. But it's not an absolute condition.

Question 4

Describe rationality. How is it defined?

> Rationality is the quality or state of being rational - that is, being based on or agreeable to reason

Being rational implies being based on or in accordance with reason or logic. Furthermore, rationality implies behaving like a rational being.

Question 5

What is Aristotle's argument about the connection between knowledge and action? Does he make any further suggestion that could be used to implement his idea in AI? Who was/were the first AI researcher(s) to implement these ideas? What is the name of the program/system they developed? Google about this system and write a short description about it.

In his book, *De Motu Animalium* ("Movement of Animals"), Aristotle argued that actions are justified by a logical connection between goals and knowledge of the action's outcome.

> "But how does it happen that thinking is sometimes accompanied by action and sometimes not, sometimes by motion, and sometimes not? It looks as if almost the same thing happens as in the case of reasoning and making inferences about unchanging objects. But in that case the end is a speculative proposition ... whereas here the conclusion which results from the two premises is an action. ... I need covering; a cloak is a covering. I need a cloak. What I need, I have to make; I need a cloak. I have to make a cloak. And the conclusion, the "I have to make a cloak," is an action." - Aristotle

He further elaborates on this topic, in the *Nicomachean Ethics* (Book III. 3, 1112b), suggesting an algorithm which can be described like in a few key points:

- We do not question about the result, but the action/method needed. A doctor does not question whether he shall heal. They assume the desired outcome and consider how and by what means it is attained, in addition to if it seems easily and best produced in that way.
- If it is achieved by one means only, they consider how it will be achieved by this and by what means this will be achieved. This goes recursive until they come to the first cause. And what seems last in the order of analysis seems to be the first in the order of becoming. Proceed to perform the first method, and work your way up.
- If we come on an impossibility, we give up the search, until a thing appears possible and we try to do it.

The algorithm described by Aristotle was implemented by Newell and Simon in their **General Problem Solver program**, 2300 years later. Today, we would have called it a greedy regression planning system. In the first few decades of theoretical research in AI, these methods based on logical planning to achieve definite goals dominated the field.

The General Problem Solver (GPS) involves means-end-analysis, that is, breaking a problem down into subcomponents (subgoals) and solving each of those. It was intended to provide a core set of processes that could be used to solve a variety of different types of problems. The critical step in solving a problem with GPS is the definition of the problem space in terms of (a) the goal to be achieved, and (b) the transformation rules. GPS would divide the overall goal into subgoals and attempt to solve each of those. Some of the basic solution rules include: (1) transform one object into another, (2) reduce the difference between two objects, and (3) apply an operator to an object.

While it was intended to be a general problem-solver, as the name implies, it could only be applied to "well-defined" problems such as proving theorems in logic or geometry, word puzzles and chess. However, this was the basis of other theoretical work for future implementations and algorithms.

Question 6

Consider a robot whose task it is to cross the road. Its action portfolio looks like this: look-back, look-forward, look-left-look-right, go-forward, go-back, go-left and go-right.

1. **While crossing the road, a helicopter falls down on the robot and smashes it. Is the robot rational?**

   Here the robot could have benefited from having the action "look-up" in its action portfolio. Whether or not the robot behaves rational, has to be based on the knowledge/information it could gather, based on the skills it holds. As perfect rationality is almost impossible to reach, and we measure the rationality of the agent based on how it maximizes *expected* performance - we can't say that it is not rational based on something it was not expected to account for. That would need perfection, maximizing *actual* performance. And the robot could have used all its skills, and still not have discovered the helicopter crashing. Therefore, we may still see the robot as rational.

2. **While crossing the road on a green light, a passing car crashes into the robot, preventing it from crossing. Is the robot rational?**

   In this example the first thing that comes to my mind is how we humans see the green light, and may lack the rationality to look-left-look-right, but just go-forward. The actions needed for discovering the incoming car was in the robot's action portfolio, but it did not discover the car as the crash happened. Therefore, one might say that the robot acted humanly (walked based on green light), and not rationally. It should have thought rationally about the possibility of the car not stopping for the green light, and acted accordingly.

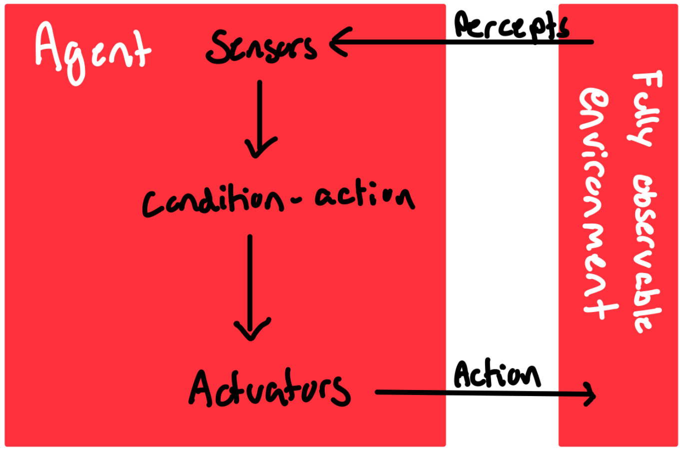
Question 7

Consider the vacuum cleaner world described in Chapter 2.1 of the textbook. Let us modify this vacuum environment so that the agent is penalized 1 point for each movement.

- **Can a simple reflex agent be rational for this environment? Explain your answer**
- **Can a reflex agent with state be rational in this environment? Explain your answer.**
- **Assume now that the simple reflex agent (i.e., no internal state) can perceive the clean/dirty status of both locations at the same time. Can this agent be rational? Explain your answer. In case it can be rational, design the agent function.**

As the simple reflex agent just takes into account its current situation/state, it will not be taking the penalties from the former outcomes of its actions into consideration. Therefore, it does not know how to act rationally, as it can move unnecessary back and forth without it taking these penalties into consideration later on. For the reflex agent with state on the other hand, it can take into consideration rewards/penalties from an actions outcome in earlier situations, allowing it to act rationally in further situations.

As the simple reflex agent now has access to a fully observable environment, it has the possibility to act rational even though the penalties won't affect the further actions of the agent. Now it does know when the other half of the environment is dirty, and can move without gaining unnecessary penalties. An agent function may look like this:




```
If current square is dirty:
  suck up dirt
else if other square is dirty:
  move to this square
else:
  do nothing
```

Question 8

Consider the vacuum cleaner environment shown in Figure 2.2 in the textbook. Describe the environment using properties from Chapter 2.3.2, e.g. episodic/sequential, deterministic/stochastic etc. Explain selected values for properties in regards to the vacuum cleaner environment.

The vacuum cleaner environment holds the following properties:

- **Fully observable**
  It is not specified in the figure nor its caption whether it is partially or fully observable. But, as it only consists of two squares placed next to each other, and we have the opportunity to give the vacuum a sensor in which is not only local to one square, I choose to refer to it as fully observable. The limitations is on the agent side, not the environment.
- **Single-agent**
  The environment is described to have one single vacuum cleaner. Nevertheless, It

could just as well have been a multi-agent environment.

- **Dynamic**
  The environment is subject to change while the agent is acting in it. It may appear dirt at random times.
- **Deterministic**
  As we have determined that the environment can be both partially and fully observable based on the sensors, we have to take a look at the property dynamic/static as it may also affect this parameter. As the next state (the outcome) is not determined by the current state or actions executed by the agent, as dirt may appear randomly - the environment is nondeterministic. We do not hold all the probabilities, which would have made the environment stochastic.
- **Sequential**
  A decision could affect all future decisions - e.g. the agents position decides which action should be taken. If it decides to move, it will affect the further decisions.
- **Discrete**
  The amount of states, percepts and actions are finite
  - It is either dirty or not
  - The agent is either in square A or B

Question 9

Discuss the advantages and limitations of these four basic kinds of agents:

- **Simple reflex agents**
- **Model-based reflex agents**
- **Goal-based agents**
- **Utility-based agents**

Simple reflex agent uses only the current percept, and does not know about the percept sequence, making it simple and having a large reduction in possible percept/action situations. It is implemented through condition-action rules, and it may depend on the environment being fully observable to work well. Randomization for exploration may also be a good solution in this model. Simple, yet effective in the right environment. But lacks understanding on former actions and its outcome, and therefore may be subject of doing the same irrational choices over and over.

This is where the model-based reflex agent comes in. It has knowledge about the world through a percept sequence, and takes this into consideration when making a new choice. The conditions in the condition-action function changes as its knowledge expands. It may be difficult to exactly determine the current state in partially observable environments, hence they may require reasoning under certainty - to "guess" the current situation. But it still has the capability not to do the same mistakes more than once.

Goal based agents has goals which is stated explicitly. It also has knowledge as the model-based reflex agents, but takes an action based on the goal stated - and the knowledge it holds about how the world works. This is a better model if the goals is possible to state explicitly, as the model knows what is the best outcome, and can work towards that goal. Not a good solution if there are not some explicit goals to be stated.

The problem with goals is that it provides only a binary happy/unhappy distinction if the goals is fulfilled or not. Utility functions provide a continuous scale, and is a better solution if a goal can be achieved in multiple ways. Utilities are internalization of the performance measure, and reflect agents preferences. Let's say we have one goal, with multiple ways of reaching this goal - we can have the utility function give the best score to the best solution. The utility of an action/outcome is not always known, but when present, the agent chooses the actions that maximizes the expected utility.