# TDT4136 Assignment 1

Erlend Paulsen Skaaden

September 7, 2022

## Question 1

There are multiple definitions of AI, based on how the AI is thinking or acting. Generally it revolves around a machines ability to act or think intelligently. It's opposed to natural intelligence you find in humans and animals. The earlier definitions defined AI as a machine being able to think or act humanly. This means to mimic the humans, especially within the concepts 'learning' and 'problem-solving'. This has since been retired in favour of a machines ability to act or think rationally, which doesn't describe how intelligence should be perceived (as opposed to the mimic definition).

## Question 2

The Turing test, proposed in 1950 by Alan Turing, is a test designed to determine (for a human) whether a machine is intelligent or not [Russell and Norvig, 2020]. The test is conducted by a human interrogator that writes questions and gives it to the interrogated. The interrogated then answers the questions, and a machine passes the Turing test, i.e considered intelligent if the human interrogator is not able to tell whether he is communicating with a human or a machine. For a machine to pass the test, it needs to be able to process natural language, store information about what it sees, automated reasoning for answering the questions and machine learning to be able to adapt and detect patterns.

## Question 3

Rationality is often translated to 'doing the right thing'. Rationality is connected to logic and reasoning - thinking or acting reasonable. One

can say that be being rational, you are guided by or based on reasons [Wikipedia, 2022].

## Question 4

In AI at least, to think rationally means to have a reasoning processes that has no holes in it, meaning it cannot be disproven. Such AIs are heavily influenced by logic, e.g:
+ All humans are mortal
+ I am a human
= I am mortal.
It therefore relies on certain information, things we know are 100% correct. In the real world this is impossible. Probability is therefore often used when thinking rationally.

To act rationally means to act so that you reach the best outcome, or when there is uncertainty, reach the best expected outcome [Russell and Norvig, 2020]. In simple words, it means that you do the correct thing. What is right is determined by constraints given to you (or that you would have to figure out yourself).

No, rational thinking is not an absolute condition for rational acting. For instance, in some cases, its best to skip the rational thinking processes and just react. For example, if something is thrown at you, its best to react and duck away, than to stop en extensively think over the situation.

## Question 5

Aristotle argued that actions are justified by a logical connection between knowledge and goals [Russell and Norvig, 2020]. In essence, you have a main problem, e.g 'I am hungry'. Then you have a set of premises, for example 'Apples makes me less hungry' and 'I have an apple'. Given the premises (our knowledge) we can conclude that eating the apple (performing the action) satisfies the goal (to be less hungry).
His idea reminds a lot of decision trees, which are broadly used in AI. For bigger problems, you are basically solving subproblems, which each has a solution/conclusion. Take the example above. Assume that we no longer have an apple, then we would have to acquire an apple, either by finding one or buying one. The latter assumes you have money and so on. This I

think can be translated into a decision tree.

### a and b

It was Newell and Simon that first implemented Aristotle's idea. Their solution is called General Problem Solver. In essence given a well-formed-formula that can generate a directed graph with multiple sources (starts) and sinks (desired conclusions), the GPS program can solve the problem.

## Question 6

### Scenario 1

It depends on the percept sequence. If for instance the robot has checked left, right, back and front and the elk suddenly appears, then the robot did a rational choice. This is called omniscience, and cannot be considered when talking about rational AI. If however the robot saw the elk from the start and still crossed the road, the robot was not rational.

### Scenario 2

Here I assume that the robot can see the car the entire time. Even though it was a green light, a rational mind should always check that it is in fact clear. Since it obviously was not, the robot has not acted rationally, and is followingly not rational.

## Question 7

### Simple reflex agent

No, it couldn't. The simple reflex agent does not keep track of earlier state, and therefore it has no chance knowing that the other tile is clean. Therefore, after it has cleaned tile A, it would notice that tile A is now clean and move to tile B. Tile B is dirty so it cleans it. But now it does not know that A also is clean, and moves back to A to check whether the tile is dirty or not. It will do this aimlessy until one of the tiles become dirty again.

### Model-based reflex agents

It depends on whether the agent knows how the world is evolving. The agent knows that it just cleaned tile A, and will therefore stay staionary

on tile B. You would have to tell the agent that for instance that a tile gets dirty approximately every 4 hours. Even then you could get situations where i.e a human spills something on tile A right after the agent has cleans tile A. The robot would then not clean tile A in before 4 hours, which is not a rational choice. Then again, you judge a agents rationality based on its action portfolio. I would say that the agent is rational with that taken into consideration.

### Simple Reflex agent with extra perceive status

This agent would be rational, since we can introduce a condition-action rule that if both tiles are clean the agent does nothing apart from checking again. It would then maximize the loss of points from moving, but still maximize points from keeping the area clean. It is also doing the right action all the time.

```
function Reflex-Vacuum-Agent([location, status]) returns an action
    if status = Dirty then return Suck
    else if status = Clean then Stand still
    else if location = A then return Right
    else if location = B then return Left
```

## Question 8

The environment I would say is partially observable, because the agent cannot check the state of the whole environment at each point in time. It can only check the state of the tile its currently on.
It's a single-agent environment because we only have a single agent that interacts with the world, and it does not compete or cooperate with other agents (if any) in the world.
I would argue that the environment is nondeterministic, because a tile getting dirty is not a direct consequence of the agents action. It is rather a natural development of the world, and could happen at any given moment.
The environment is episodic. The agents action does not rely on earlier actions. It only makes an action based on the current state of the environment.
It's a dynamic enivornment because the world can change over time. For instance, after some time, a tile will get dirty again.
The environment is discrete as it has a finite number of states and percepts and actions.

The world is so simple and you can make probable guesses on how often a tile gets dirty, so the environment is known.

## Question 9

### Simple Reflex Agents

These agents have the benefit of being simple. Some of then could be implemented in boolean circuits. However they have limited intelligence and struggle in partially observable environments. It has to make the correct decision based only off the current percept, which in many cases would fail, as it demands a fully observable environment.

### Model-based Reflex Agents

This agent allows for storing an internal state, which can in most cases deal with partially observable environments. However this agent only has a 'best-guess' for the full environment. This information can be outdated, leaving the agent with a wrong perception of the current state of the environment.

### Goal-based Agents

Goal-based Agents are more flexible than reflex agents. This is because the knowledge that supports it's decisions are explicit and can be easily modified [Russell and Norvig, 2020]. This agent is however less efficient than reflex agents, as it sometimes has to consider long sequences to achieve it's goal. It also in essence only have two states: happy and unhappy.

### Utility-based Agents

An utility Agent can handle uncertainty, as well as different paths that lead to the same goal. This however demands that the agent keeps track of the environment which can be a challenge. Its also hard to make an utility agent fully rational, mainly because of the computational constraint.

## References

[Russell and Norvig, 2020] Russell, S. J. and Norvig, P. (2020). *Artificial Intelligence: a modern approach*. Pearson, 4 edition.

[Wikipedia, 2022] Wikipedia (2022). Rationality — Wikipedia, the free encyclopedia. http://en.wikipedia.org/w/index.php?title=Rationalityoldid=1105981852. [Online; accessed 26-August-2022].