Artificial Intelligence Fundamentals and Intelligent Agents

Theoretical questions

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

1. Bellmann defines AI as *"The automation of activities that we associate with human thinking, activities such as decision-making, problem solving, learning"*[1]. This is a way of defining AI as **thinking humanly**.
2. Another definition of AI as a **thinking rationally** approach is by Charniak and McDermott; *"The study of mental faculties through the use of computational models"*[1].
3. Next, we have a definition by Rich and Knight in 1991, using the approach of **acting humanly**: *"The study of how to make computers do things at which, at the moment, people are better."*[1].
4. Finally, we have a definition by Nilsson in 1998, who uses the approach of **acting rationally**: *"AI ...is concerned with intelligent behavior in artifacts"*[1].

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

The turing test is a test designed by Alan Turing to provide a definition of intelligence. It is performed by a human interrigator, asking a computer written questions for five minutes. If the interrigator cannot tell if he was talking to a computer or human being 30% of the time, the test has passed.

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

To *think* rationally is to think by a logical set of rules. Eg. *"Paper burns. This book is made out of paper. This book will burn"*. To *act* rationally is to first *think* rationally and logically reason to a conclution and then *act* on the conclution, For example *"Humans will die if jumping from a high place. The mountain I'm on is high. I should not jump from this mountain"* and then walk down the mountain instead of jumping down. Thinkin rationally is not always a necessity for acting rationally. For example, if you touch a hot stove top, the rational thing would be to quickly take your hand of the stove. However, you do not have time to logically reason what to do. Therefore, rational thinking is not an absolute condition for acting rationally.

> Describe rationality. How is it defined?

The book defines rationality as *"Ratinonality is when a system does the "right thing" given what it knows"*. In other words - rationality is acting out of knowledge. As explained above, an example of actiong out of knowledge is *"Humans will die if jumping from a high place. The mountain I'm on is high. I should not jump from this mountain"*. I *know* that humans die when jumping from a high place, and I *know* this mountain is high. Therefore, I will not jump.

> 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.

Aristotles argument about the connection between knowledge and action is that "actions are justified by a logical connection between goals and knowledge of the action's outcome".

Yes, he made an algorithm for "how to solve problems". He said that "we deliberate not about ends, but about means". With this, he meant that we do not think about the end, but the way to get to the end. This algorithm was implemented by Newell and Simon in their GPS-program (General Problem Solver). GPS was the first machine to separate the its knowledge of problems from the strategy it has to solve them.

> 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.
>
> a. While crossing the road, a helicopter falls down on the robot and smashes it. Is the robot rational?
> b. While crossing the road on a green light, a passing car crashes into the robot, preventing it from crossing. Is the robot rational?

a. I would say yes, the robot is rational. If we are to use the book's definition of rationality, the robot is doing *the right thing* based on the knowledge the robot has. The robot cannot look upwards, and therefore, cannot have any knowledge about a helicopter crashing.
b. I think this scenario is somewhat divided. Given the ability to look-left-look-right, the robot should have anticipated the car passing and therefore waited to cross the road until the car had passed. Since it didn't, the robot is not rational. On the other hand, the light was green, and based on the knowledge that green is go and red is stop, the robot did the right thing by crossing the road and the robot was rational.

> Consider the vacuum cleaner world described in Chapter 2.2.1 of the textbook. Let us modify this vacuum environment so that the agent is penalized 1 point for each movement
>
> a. Can a simple reflex agent be rational for this environment? Explain your answer
> b. Can a reflex agent with state be rational in this environment? Explain your answer.
> c. 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

a. No, I do not think a simple reflex agent can be rational in this environment. This is because the agent cannot determine when the whole floor is clean. It would just go back and forth between the squares and be stuck in an infinite loop checking "*is the floor clean? Yes. Switch. Is the floor clean? Yes. Switch. Is the floor clean? Yes. Switch.*"
b. If I assume that the floor won't get dirty after it is cleaned, I think a reflex agent can be rational here. Let's say the vacuum cleaner is in square A and this is dirty. It will clean the square and save it's "clean"-state and move on to square B. It will then clean this square and save this state. Now, when deremining if it should switch squares, it can check the state and decide not to switch, since square A is clean.
c. Yes, if it can percieve the clean/dirty status of both locations, it can be rational.

```
const vacuumAgent = () => {
    while (dirty.left || dirty.right) {
        if (
            (dirty.left && location === 'A') ||
            (dirty.right && location === 'B')
        ) {
            suck();
        } else {
            location === 'A' ? go_right() : go_left();
        }
    }
};
```

Where `dirty` is the percieved status.

> Consider the vacuum cleaner environment shown in Figure 2.3 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 is **partially** observable, because the agent can only "see" one square at the time. It is also a **single agent**, because the vacuum cleaner is the only agent present. Further, it is **deterministic** because the next state is determined by the current state and the action executed (e.g floor is dirty, clean). The environment is **episodic**, because the agent's experience is divided into atomic episodes (See square, clean, change square, clean). The question to wether the environment is **static** or **dynamic** is based on the assumptions of the environment. If the floors can get dirty again (after cleaning), it is **dynamic**, if not, it is **static**. If the agent is penalized 1 point (as in the previos task), and the floor cannot get dirty again, it is **semidynamic**, because the agent's points would change but the environment would not. The environment is **discrete** since the floor has the distinct states "clean" or "dirty", and the agent has the distinct states "suck" or "change location". Lastly, the environment is **known**, because the outcome of all states are given (if dirty - suck, if clean - change).

> Discuss the advantages and limitations of these four basic kinds of agents:
>
> a. Simple reflex agent
> b. Model-based reflex agents
> c. Goal-based agents
> d. Utility-based agents

a. A **simple reflex agent** can be easy to implement, but have limited intelligence. Only one minor unobservability can cause trouble.
b. A **model-based reflex agent** has the advantage that it can save a state of "how the world is now". The downside is that it cannot know exactly how the world looks and may need to make an educated guess, which may not be right. (E.g a trck has stopped in front of a self-moving car and the car tries to pass it, since in the way the world looked a couple of minutes ago it was "clear" to pass)
c. A **goal-based agent** finds action sequences based on searching and planning, helping it to reach it goals. Furthermore, it is flexible in the way that knowledge that supports it's decisions is represended

explicitly and can be modified. The disantvages are that it is less efficient and more complex than the other agents discussed earlier.

d. A **utility-based agent** is a lot like a **goal-based agent**, but it differs in the way that it reaches its goal(s). A **utility-based agent** uses searching and planning to find an optimal actions, by using a utility function which evaluates "Does this make me happy?".

Bibliography

[1]: Russell, S. J., & Norvig, P. (2010). 1.1. In Artificial intelligence: A modern approach. Boston: Pearson Education.