1 Theoretical Questions

1. AI is a broad field with many interpretations, and there are thus numerous ways to define this field. One broad definition is that AI is the study and construction of agents that do the right thing. Though what is considered the right thing is also up for debate. An AI should be able to make decisions or act in a way that could mimic human intelligence, thus it should emit intelligent behavior. Moreover, what we may consider intelligent behavior is possible to define in many ways. One way to define it is that an AI should try to maximize expected utility. This means that the AI should try to achieve some goal within a given set of contraints in a way that maximizes the expected outcome (Russell and Norvig 2021, pp. 7–22).

2. The Turing test is an experiment created by Alan Turing to see if a machine could exhibit intelligent enough behaviour to such an extent that a human interacting with it, for instance posing and answering questions, could not distinguish it from another human. A new proposal of the Turing test includes interaction not only through answering and posing questions, but actual physical interaction with humans and objects in the real world, which would require for instance computer vision and robotics in addition to natural language processing (Russell and Norvig 2021, p. 23).

3. Rationality is defined as doing the right thing. Thinking rationally can be described as the ability to make predictions about the future based on a sound understanding of how the world works. Acting rationally, however, is the ability to make decisions that maximize expected utility given ones information and understanding about the world. An AI doesn't need to be able to perform rational thinking the way we humans do to act rationally. It should make decisions that are optimal given a set of goals or constraints, and may act rationally without rational thought as long as it takes the best course of action given the understanding it has of the environment around it due to its sensors, and the actions it can perform (Russell and Norvig 2021, pp. 19–22).

4. According to Aristotle, there is a connection between knowledge and action due to the fact that to reach a goal, one has to have knowledge about the outcomes of an action, and thus what actions are required to reach a specific goal. Aristotle himself proposed an algorithm which explains how his argument can be used to implement the idea in AI. One would assume the goal state and then work backwards to find a sequence of actions that would lead to that goal state, i.e. regression.

   (a) Newell and Simon were the first AI researchers to implement Aristotle's ideas.

   (b) The name of the program they developed was General Problem Solver, which is a planning system that uses a greedy algorithm together with regression to find a solution to a problem.

   (Russell and Norvig 2021, p. 25)

5. - The robot has a set of restricted actions, which include looking and going in four different directions. We can assume that the robot would look in a specific direction, e.g. forward, to see if there were any obstacles blocking it from reaching its goal, i.e. crossing the road. If there were, it would take a step left/right, and if there were none, simply go forward. It is after this action is taken that an elk suddently crashes into the robot and smashes it. Based on this, the robot made decisions based on its available actions and the information it had at the time, and thus acted rationally. The robot could not have known that an elk would crash into it, and thus could not have avoided it.
   - The robot chose to cross the road on a green light, which from using its lookForward action, would mean that it was safe to cross the road. Again, the robot could not have known that a car would drive a red light and subsequently crash into it, and thus acted rationally. Even though it didn't reach its goal in any cases, it still acted within the actions and knowledge it had at its disposal to best reach its goal.

6. - A simple reflex agent would make a decision on the next course of action based on the current percept (Russell and Norvig 2021, p. 67). Since it has no state it has no

knowledge about whether or not the other box is dirty, so it will continously suck the dust in the current box if dirty, then move to the other, move back if that one is clean, and continue this cycle indefinitely, thus not acting rationally.

- Compared to the simple reflex agent, a reflex agent with state can perceive the statuses of both boxes at once (Russell and Norvig 2021, p. 69). It will therefore know when in box A, whether or not box B is dirty or clean, and will make a decision based on this, thus making less useless (in terms of its goal) moves like the simple reflex agent, and act rationally.
- In the case of a simple reflex agent with no internal state, but with the ability to perceive the clean status of both boxes, it would always be able to make a rational decision based only on its current percepts since it knows whether or not any of the boxes are clean or dirty. Thus it could be rational. The agent function would look like this:

  If CC is the current box is clean, CD is current box is dirty, OC is other box clean and OD is other box dirty, the agent function could be defined in a table as follows:

Percept	Action
[CD, OD]	Suck
[CD, OC]	Suck
[CC, OD]	Move to other box
[CC, OC]	Do nothing

7. The original vacuum cleaner has the following properties:

   - **Partially observable:** it's sensors can only perceive the current box it is in, and not the state of the other box, thus it does not have full knowledge of the whole environment.
   - **Single-agent:** it is the only agent in the environment, and thus does not have to consider other agents when making decisions.
   - **Deterministic:** the next state of the environment is determined only by the current action performed by the agent.
   - **Episodic:** the agen't current decision won't have an impact on future decisions, thus it does not have to consider future outcomes when making decisions.
   - **Static:** since the environment does not change while the agent is deliberating and the agent is the only one impacting the environment, it is static.
   - **Discrete:** since there are two boxes who can either be clean or dirty, the environment has a finite number of distinct states.
   - **Known:** the agent knows the environment and does not have to rely on learning to know what action will result in the best outcome.

   (Russell and Norvig 2021, pp. 62–64)

8. Under are some advantages and limitations listed for the following types of agents:

   - **Simple reflex agents:** simple since they're made up of if-then rules, but therein also lies their limitations since these types of rules don't scale when you have to take into consideration an increasing number of cases which have to be put in manually (Russell and Norvig 2021, pp. 67–69).
   - **Model-based reflex agents:** can keep track of the current state of the whole environment outside of its current percept which can be utilized to make better decisions to maximize its performance measure. Though in a partionally observable environment, it can only keep track of the state to the limits of its own sensors, thus it will make a guess and base its decision on that guess (Russell and Norvig 2021, pp. 69–71).
   - **Goal-based agents:** having knowledge of an explicit goal, the agent can use this together with its knowledge about the state of the environment to make better decisions. The model is also flexible in that it can change its decisions based on what it believes the outcome of an action will be, and thus can make even better, more informed decisions

that in the end will maximize its performance measure. This model falls short however when there are conflicting goals, wherein only some can be achieved, and will have no way of knowing which one to choose since it only measures the goal in binary "good" and "bad" states (Russell and Norvig 2021, pp. 71–72).

- **Utility-based agents:** even more flexible than goal-based agents since it also takes utility into consideration when making decisions in terms of which action to take. When there are conflicting goals, it will choose an action based on an appropriate tradeoff, specified by a utility function. These models are however very complicated to design as they have to keep track of a lot of information. Knowing which utility-maximizing action to take is also not always easy, nor is defining the utility function correctly. Trying to reach perfect rationality is also not achievable either, since it is impossible to know all the possible outcomes of an action in a computationally feasible way (Russell and Norvig 2021, pp. 72–74).

Bibliography

Russell, Stuart and Peter Norvig (2021). *Artificial Intelligence: A Modern Approach, 4th Edition*. Pearson Education.