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

"The theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages." [1]

"A branch of computer science dealing with the simulation of intelligent behavior in computers." & "The capability of a machine to imitate intelligent human behavior." [2]

"artificial intelligence (AI), the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings." [3]

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

The Turing test is a testing method used to evaluate a machines ability to "act humanly". A Turing test consists of an interrogator, a human and a computer. The test is said to be successful if the interrogator cannot differentiate who is the human in the interaction and who is the machine.

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

Thinking rationally in AI is using a set of logical links and conclusions to find the correct answer for a given premise. Here the focus is on correct inference, the process of which the computer arrives at it's answer. Acting rationally in AI is a process in which the goal is to achieve the best outcome, or the best expected outcome. Over a time-period this often requires the ability to adapt to change and to create and pursue goals autonomically.

This process of achieving the best outcome (acting rationally) often uses the logical inference from thinking rationally, but it is not an absolute condition. Sometimes there are no logical answer to an action but something still needs to be done, and sometimes the need to act quickly is of greater importance than to use the time and find the logically best solution.

4. What is Tarski's "theory of reference" about?

Alfred Tarski's theory of reference describes how we can map, or relate, objects in a logic to objects in the real world

[1] The English Oxford Living Dictionary
[2] https://www.merriam-webster.com/dictionary/artificial%20intelligence
[3] https://www.britannica.com/technology/artificial-intelligence

5. Describe rationality. How is it defined?

At any given time, what is rational depends on four things (PEAS):

- The performance measure that defines the criterion of success
- The agent's prior knowledge of the environment
- The actions the agent can perform
- The agents percept sequence to date

The definition of a rational agent is the following:

"For each possible percept sequence, a rational agent should select an action that is expected to maximize its performance measure, given the evidence provided by the percept sequence and whatever built-in knowledge the agent has"

6. Consider a robot whose task it is to cross the road. Its action portfolio looks like this: look-back, lookforward, 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?

Yes, in this context the robot is rational. It has no action to evaluate the environment above it, so this event is outside it's rational universe. It can also be considered an unlikely event to have this happen so you could argue that implementing features to check for fast falling objects from the sky also would be irrational.

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

Given the context this statement is debatable. If the green light is a green light for the car then I would argue that crossing the road at all is irrational. If it's a green light for the robot, it is a rational move for the robot to try to cross the road. If the car is moving at a low or reasonable speed even if it runs a red light I would say the robot should use action to check the environments on both sides of the road, as we are often taught as humans. In this case I would say the car acted irrationally. If the car is moving at a very fast pace the answer once again becomes somewhat of a gray area.

7. 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

I argue that the simple reflex agent is not a rational choice for this environment. The agent can only evaluate the current location it is on if that location is clean or dirty (not the whole environment). It also does not have a memory and it will most likely result in driving infinitely forward and back between the two locations due to the ruleset it obeys by.

(b) Can a reflex agent with state be rational in this environment? Explain your answer.

A model-based reflex agent will be a better choice than a simple reflex agent. Due to the simplicity of the environment we might be able to accurately guess how the environment changes over time and how the actions of the vacuum cleaner effects it.

(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.

As it now can observe the full environment I will now argue that the simple reflex agent is rational. As we can add functionality to wait (do nothing) if both locations are clean.

Current agents & added agent (added agent shown in red)

if status = Dirty, then return Suck (assumes this only checks our current location)

else if (status A = Clean && status B = clean), wait X*

else if location = A, then return Right

else if location = B, then return Left

* The variable X is a variable of time and should be based on observations and estimations on how long it will take in general for a location to get dirty.

8. 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 environment is episodic, as only a single action is made based on atomic (single) perceptions we have about the current state of the environment. Also actions it does doesn't affect the decisions long term (in the future).

The environment is deterministic if we assume that cleaning dirt in one location does not affect the other location. Then our next state in the environment is only determined by the current state and the agents action.

We can also note that the environment is only partially observable.

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

(a) Simple reflex agents

An advantage with the simple reflex agent is that it is simple and easier to implement than the other agents. A disadvantage is that the agent requires to be able to fully observe the whole environment, something that is not often possible in real world applications. Another disadvantage is that the agent sometimes can end up in a endless loop, where the actions recursively lead to eachother.

(b) Model-based reflex agents

An advantage here is that it is still somewhat easy to implement. This model tries to estimate how the world reacts on its actions and have more states to reflect this. This makes it possible to use this agent in a only partially observable environment.

A negative here is that since it only based its decisions on a internal model of the environment there can be a lot of uncertainty of how well the agent actually is going to act. Oversight in the implementation can also lead to unintended behavior.

(c) Goal-based agents

A positive with the goal-based agent is that it is more flexible than the reflex agents. Its process is goal driven and is subject to natural change given data it retrieves which means less work for the developers changing certain things with the AI

Some negatives are the fact that it is less efficient than the reflex agents and that we have to manually change the goal and “start over” when we want the agent to achieve a different goal.

(d) Utility-based agents

Is driven by finding an optimal set of actions to reach its goals. It has positives in terms of flexibility and learning. Perhaps the biggest positive comparatively is that it can find a “estimated optimal solution” when we have no exact solution. It also tries to use the “best” solution and not just a working solution, increasing the chance for a satisfying solution/ set of actions.

A negative here is trying to optimize must factor in a lot about the environment and states and therefor is a lot more complex.