Name: Amer Kalac

Assignment 1 – TDT4136

1) The definition of AI:

It is hard to stick to one specific general definition for AI, since it is such a broad field of study, and there are many ways to view intelligence. If I were to define it, I would probably try to involve most of the types of approaches like described in the textbook, however it would still be little specific and probably still not be representative for the whole field. I think the definition from the textbook is a good way to define it, since they also divide AI into four categories.

Some definitions found online, not mentioned in the lecture:

- From Wikipedia: "Articifial intelligence is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans. Colloquially, the term Artificial Intelligence is often used to describe machines (computers) that mimic cognitive functions that humans associate with the human mind, such as learning or problem solving."

- Demis Hassabis, CEO of DeepMind: "Artificial intelligence is the science of making machines smart."

- IBM's definition: "By artificial intelligence we mean anything that makes machines act more intelligently."

- PwC's definition: "AI, shorthand for artificial intelligence, defines technologies emerging today that can understand, learn, and then act based on that information."

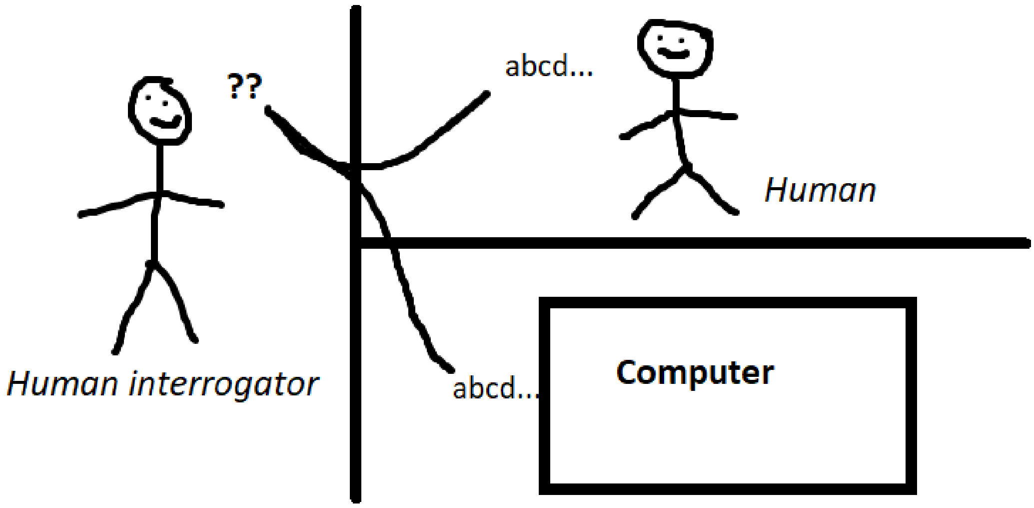
- McKinsey's definition: "Artificial intelligence is the ability of machines to exhibit human-like intelligence."

- Deloitte's definition: "Artificial intelligence is getting computers to do tasks that would normally require human intelligence." (similar to Minsky's definition).

Name: Amer Kalac

2) The Turing test:

The Turing test was proposed by Alan Turing in 1950, designed to test if one can separate an AI system from a human. The test involves a human, an interrogator, and a computer. The computer passes the test if a human interrogator cannot tell whether the responses are from the human or the computer, after asking a set of written questions.



3) Thinking and acting rationally:

Aristotle was one of the first who attempted to systemize “right thinking”, or how human reasoning works. These Laws of Thought initialized the study of logic, which later were transferred to AI, where the goal is to make programs think rationally (logically) and create intelligent systems.

Acting rationally means acting to achieve the best outcome, or best possible outcome, when acting towards a goal. In AI we can refer to rational agents, where we can see that making correct inferences is often part of being a rational agent, or acting rationally. This is because one way to act rationally is to use logical reasoning (rational thinking), to find an action that will take you to your goal.

However, there is not always a clear correct thing to do in a situation, but some action must still be done. So there are ways of acting rationally without requiring making logical inferences. For example, if you step on a sharp object, you will instantly move your foot as a reflex action, which is probably more rational than making a inference in your head before acting. Therefore, rational thinking is not an absolute condition for acting rationally.

4) Tarski's theory:

Alfred Tarski's (1902-1983) theory of reference shows how to relate the objects in logic to objects in the real world.

5) Rationality:

Rationality can be thought of as making the right decision in a situation, often based on logical inferences. One must however know how to define what the "right thing" is, like having a goal or a task which one may wish to reach. We can therefore see rationality as making a decision which will achieve the best outcome, or best expected outcome (if there is uncertainty), to reach or solve our task, which often is backed by logical reasoning to do said actions.

6) Robot:

a) Yes, I would say so. Because there is no way, with the given action portfolio, that the robot could have perceived the falling helicopter. There is no look-up action. So, it did not do anything irrational based on the knowledge available. It would not really be rational for humans to check the skies before crossing a road either, I guess.

b) You could argue for both sides here, but I would say the robot is not rational. The light is green, which the robot perceives, but it should perceive the environment to the right and left to minimize risk of being ran over, like we humans do. It can still be debated, since one should not expect to get hit by a car while crossing on a green light, but for us humans it would be rational to look right and left.

Name: Amer Kalac

7) Vacuum cleaner 2.2.1:

a) It will check if there is dirt, and suck if there is. Otherwise it will move to the side, and do the same there. It will end up moving back and forth during the rest of its lifetime, however it would still be gaining points. A problem that most likely will occur however, is the robot being stuck in an infinite loop of tying to go through a wall. This would happen if the agent were to be deprived of its location sensor, only to be left with a dirt sensor, and therefore it might not be rational.

b) Yes, it would be rational. With an internal state, it will be able to make a model of how the world works based on percept history. Thus, it will be able to handle this partially observable environment even if the location sensor were to stop working.

c)

```
agentFunction (perception) {
if( perception == B_Dirty){
        if( WallIsRight ){      //if it is on B and B is dirty
                suck(); }
        if(A_Clean) {      //if it is on A and A is clean, but B is dirty
                moveRight(); }
}
if( perception == A_Dirty){
        if( walltoLEft) {          //if it is on A and A is dirty
                suck();
        if( B_clean) {        // if it is on B but B is clean
                moveLeft(); }
}
If( perception == A_clean and B_clean){
 Stop_cleaner(); }      //ideally stop it after both are clean to not lose points


}
```

We have to take into account that we don't know the starting position. However, the vacuum cleaner can know the states of both tiles. Also, we must avoid the infinite loop which might happen. After both are clean, the cleaner should stop moving.

8) Vacuum cleaner 2.3:

The environment is:

- **partially observable**, since one cannot sense which tiles are clean or dirty everywhere. Only know if the block you are on is clean or not.
- **deterministic**, because the next state is completely determined by current state and the agent's action (dirty block becomes clean).
- **Episodic**, because the agent's experience is divided into atomic episodes. Each episode (on each block) consists of perceiving if there is dirt, and then performing suck or nothing. The decision does not affect future decisions.
- **Discrete**, because there is a finite number of distinct states, percepts and actions. The blocks are either dirty or clean, the agent can choose to suck, move left or right, etc.
- **Single agent**, as there are no other agents.
- **Known**, since the outcomes for all actions are given (clean after suck).
- **Dynamic,** because the world changes each time the robot sucks dirt.

9) Agents:

a) *Simple reflex agents:* These are the simplest kind of agents and rather easy to implement, which is a big advantage. In turn they turn out to have limited intelligence. In more complicated worlds, they might not act rationally. A big problem are infinite loops which might appear if the world is not fully observable, and therefore they will not work. A possible solution to this is to randomize actions to get out of the loop.

b) *Model-based reflex agents:* This type keeps a internal state stored inside, maintaining knowledge, a model, of how the world works. It can describe, or assume, parts of the world which are not observable, depending on percept history. It can therefore handle partially observable environments, which is an advantage over the simple reflex agent. One disadvantage to this is that you can rarely make an exact determination of the environmental state, which means that uncertainty may be unavoidable.

c) *Goal-based agents*: Difference from the reflex agents is that goal-based agents look to the future when making decisions. These agents have goals which describe desired situations. Knowledge is represented explicitly, and can be manipulated, and the agent can use searching and planning to find the best action sequences for the goal. These agents are therefore more flexible than reflex agents, however they are more complex and less efficient.

d) *Utility-based agents:* While goals just say if something makes the agent happy/unhappy in a binary way, the utility-based agents have a continuous way of seeing it. With utility functions (internalizing performance measures), one can see the utility (happiness) values when a goal can be achieved in more ways. These agents therefore see multiple possibilities and choose the best one, based on more things than just reaching the goal). These agents can be very complex, and perfect rationality is usually unachievable in practice because of computational limitations. It has to deal with uncertainty as well.