Assignment 1: AI Fundamentals

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

Merriam-Webster defines artificial intelligence as “the capability of a machine to imitate intelligent human behavior” (Merriam-Webster, n.d.).
Stanford Encyclopedia of Philosophy defines artificial intelligence as “[...] the field devoted to building artificial animals (or at least artificial creatures that – in suitable contexts – appear to be animals) and, for many, artificial persons (or at least artificial creatures that – in suitable contexts – appear to be persons)” (Bringsjord & Govindarajulu, 2018).
Britannica defines artificial intelligence as “[...] the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings” (Copeland, 2020).

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

The Turing test is a test for artificial intelligence, conducted physically - sidestepping the need for a strict definition of intelligence. The test is conducted by having one human - the interrogator - ask questions to two players - one human and the machine being tested - by text. The task of the interrogator is to figure out which player is human and which player is a machine. The machine (or artificial intelligence) has passed the Turing test if the test is repeated a number of times and the proportion of interrogators unable to distinguish the two is high enough, at which point the machine should be considered an intelligent entity.

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

The idea of thinking rationally being able to find the correct (or optimal) solution to a given problem, and can be represented by formal logic. While many humans aspire to think rationally (and, perhaps, believe they do), most human behavior is less rational and more random - most humans make mistakes and miscalculate problems (if not, then most people would receive perfect scores on every exam). Acting rationally means that an agent works to achieve its goals, and to maximize the chance of doing so.

While many cases use rational thinking as a starting point for rational acting (say, to find an optimal goal - then working towards that goal), there are examples showing that one can act rationally without thinking. Say, for instance, a human puts his or her hand on a hot stove top. One’s first reaction would be to remove the hand from the stove, as taking time to rationally figuring out a best solution would give worse results than simply acting. This is also the case for computers - if an agent has an already set goal, then it would simply need to act (rationally) to achieve that goal, without the need to figure it out by itself (thus removing the need for thinking).

4) **Describe rationality. How is it defined?**

While different fields have slightly different definitions of rationality, the core of rationality seems similar independent of fields. In short, rationality is basing one's thoughts and/or decisions on logic, meaning a rational decision is correct based on a series of logical assumptions and the current knowledge of the environment.
Cambridge Dictionary defines rationality as "the quality of being based on clear thought and reason, or of making decisions based on clear thought and reason" (n.d.).

5) **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.**

Aristotle argued that actions come about as a result of the connection between one's goals and knowledge of the outcome of an action, meaning that a person with a goal (such as feeling hungry and wanting to feel full) and the knowledge that an action will lead to a goal (eating will lead to a person feeling full), that person will act (eat).

Aristotle came up with an algorithm for deciding whether or not to act (the example is for a medical decision):

a) Assume the end result (healing a patient)
b) Consider how and by what means the end result can be achieved (figure out what procedures and medication is needed to heal the patient)
c) If achieved by multiple means:
    i) Find the easiest method giving the best result
d) When finding one way to do it:
    i) Consider how the result will be achieved by the chosen method
    ii) Consider how this method will be achieved
        1) … Recurse until finding a first step
e) Start working from the bottom up - the last step in analysis is the first step in solving

When met with an impossibility - cancel the search.

This algorithm was first implemented by Allen Newell, J. C. Shaw and Herbert Simon in a program called GPS (General Problem Solver). GPS was able to solve any problem that can be represented as a directed graph with one or more sources and sinks (representing axioms and solutions, respectively), and was the first language to separate its knowledge (the system's rules, given as input data) from its solution strategy. While the program was, in theory, able to solve all sufficiently formalized problems, most real-life problems became computationally impossible, due to the complexity of the problems (and resulting graph).

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?**

While being smashed by a helicopter in no way makes a previously irrational robot rational, it also does not remove rationality. This means that a previously rational robot is still rational if a helicopter falls down on its head. It is not feasible for a person (or robot) to take every detail into consideration when crossing the road, including helicopters flying above. This is approaching omniscience, which is impossible to achieve in real life.
That being said, if the robot had the ability to look up (and take in the information) and assuming the helicopter is falling slowly, standing still when a helicopter comes straight for you is not at all rational. This is not at all relevant to the rationality of a robot, however.

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

This greatly depends on the robot and its design. Rationality is defined as doing the right thing given current knowledge, and walking in front of a passing car can hardly be considered the right thing to do. The question, then, depends on what the robot knows. If a robot is designed with only a traffic-light sensor, meaning it only has knowledge of whether the light says to stay or to go, and it gets hit by a car speeding through a red light, then it could still be considered rational - the only knowledge it has is that it's allowed to cross, thus making crossing the correct choice. In the given case, however, the robot has the ability to look left and right, thus giving it a way to get more knowledge than simply whether it's allowed to cross. This means that the robot should know that a car is coming (as it should know to look for speeding cars), making the robot irrational.

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**

No. If the agent is penalized for each movement, then the rational action would be to:

a) If the floor does not get dirty again: Clean each square once, then stop
b) If the floor does get dirty again: Clean each square once, then rest and clean the current square, while periodically changing the current square

As a simple reflex agent does not take into consideration whether or not it has visited the other square, the agent will be stuck "bouncing" back and forth between the squares, losing one point each move.

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

Yes. As explained before, rational actions would be either (a) og (b) given above. (a) can be solved by having an internal state that says which squares has been visited, and (b) can be solved by having an internal state that says how long it's been since the agent last changed square.

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.

Yes. If it can check both locations at the same time, then the system is fully observable, meaning a simple reflex agent can be rational. This can be achieved by cleaning squares as they become dirty, moving only when necessary.

The agent can perceive: [Current square, status A, status B]. Example: [A, clean, dirty]

```
if Current square = A and status A = dirty => Clean
else if Current square = A and status B = dirty => Right
else if Current square = B and status B = dirty => Clean
else if Current square = B and status A = dirty => Clean
else => Do nothing              // If both squares are clean
```

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:

- Partially observable → The agent can only see if one square is dirty, can't see both
- Single-agent → There's only one agent (vacuum robot) involved
- Deterministic → Actions have clearly defined consequences - cleaning always cleans up dirt, and moving left/right always leads to the agent being in the left/right square.
- Episodic → The agent needn't plan ahead, as its next action depends solely on the current state of the environment
- Static/dynamic
    - If we assume the floors won't get dirtier over time, then the system is static
    - If we assume the floors will get dirtier over time (say, by use), then the system is dynamic
- Discrete → The system has a finite amount of possible states, time in discrete periods, etc.
- Known → We assume that the agent / designer knows the system - whether the floors get dirtier over time or not, and that cleaning removes dirt

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

a) Simple reflex agents

Advantages:

+ Simple system, easy to design, efficient
+ Can be rational in fully observable systems
+ No need to store data (such as an internal state)

Limitations:

- Only works if the system is fully observable
- Real life situations are usually not fully observable, limiting the realistic use of simple reflex agents
- Often gets stuck in infinite loops (especially in partially observable environments)
- Any change in environment requires changing the internal rules of the agent

b) Model-based reflex agents

Advantages:

+ Can handle partially observable environments, as its internal model tells it “how the world should work”
+ Still efficient

Limitations:

- Any change in the environment requires changing the internal rules of the agent
- Has no way to express goals

c) Goal-based agents

Advantages:

+ More flexible than agents mentioned above, as one can easily change the agent’s goals

Limitations:

- Not good at choosing different possible paths to its goal
- Not good at dealing with conflicting goals

d) Utility-based agents

Advantages:

+ More flexible than reflex agents, as one can easily change the agent’s goals
+ Can prioritize and select from different possible paths, depending on expected results and probability
+ Can prioritize and balance conflicting goals to achieve a good result

Limitations:

- Utility function must match actual performance measure
- More setup required, harder to design

References

Bringsjord, S. & Govindarajulu, N.S. (2018). Artificial Intelligence. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Summer 2020 Edition)*, Retrieved August 26, 2020, from https://plato.stanford.edu/archives/sum2020/entries/artificial-intelligence

Cambridge Dictionary. (n.d.) Rationality. Retrieved August 28, 2020, from https://dictionary.cambridge.org/dictionary/english/rationality

Copeland, B. J. (2020) Artificial intelligence (AI). In Encyclopædia Britannica. Retrieved August 26, 2020 from https://academic.eb.com/levels/collegiate/article/artificial-intelligence/9711

Merriam-Webster. (n.d.). Artificial intelligence. In Merriam-Webster.com dictionary. Retrieved August 26, 2020, from https://www.merriam-webster.com/dictionary/artificial%20intelligence