**Unit 4 Algorithmics**

**Week 12 Submit Tasks**

1. Rishi says that the Turing Test is a waste of time because we already know that machines cannot think. Explain the error in his reasoning.

How does he know that machines cannot think? The ability to “think” is a relatively vague concept, and to be able to establish a test that provides a definitive way to measure how well a computer can “think” like a human is useful. His error lies in the assumption that computers could never be able to think, but we cannot definitively prove this.

(2 marks)

1. Briefly describe three shortcomings in the Turing Test’s aim to establish whether or not machines can exhibit human-like behaviour.

* **Language-centric Focus**: the Turing Test focuses purely on a machine’s ability to mimic human language and fails to address other aspects of what we attribute to intelligence, such as perception, problem-solving, decision making and the ability to understand/respond to emotion.
* **Lack of Semantic Understanding**: the Turing Test measures a machine’s ability to *imitate* a human, not its ability to understand the responses. This is the main shortcoming addressed by John Searle’s Chinese Room Argument.
* **Dependence on Interrogator**: the test’s effectiveness heavily depends on the skill of the human interrogator, who will likely produce a non-deterministic response. If the interrogator lacks the necessary skills to ask complex questions, the computer may easily pass. Along with this, the test is subjective and up for debate due to this element.

(3 marks)

1. State the ‘systems response’ to the Chinese Room Argument, and a standard reply.

The system’s response to the Chinese Room Argument concedes that the person themselves does not understand Chinese, but the system as a whole (the person, the books, the bin, etc) understands Chinese collectively, effectively stating that a CPU in a computer could not have understanding, but the entire system of a computer including all its components could.

Searle’s response to this is to consider the scenario where the person simply memorised the rules completely. In this case, the system and the person are identical because the person encompasses the system, but the person still does not understand Chinese. Hence, it can be concluded that the system does not understand Chinese either, because we are only left with one object: a person who doesn’t understand Chinese.

(2 marks)

1. Can we ever prove that machines can be conscious?

It would likely not be possible to ever prove if machines can be conscious. This is because our perception of consciousness is purely limited to our interactions with the machine, and since we cannot *be* the machine ourselves, we cannot tell if they understand anything. This is similar to how we can’t tell if dogs are conscious because we are not dogs.

(2 marks)

1. Suppose Google constructs an algorithm which is able to predict, with 99% accuracy, the most appropriate marriage partner within a 100km radius for whoever runs the program. It does this based on data gathered over a long time period, and gives you nothing but a name and email address (so there is no doubt you’ve found the right person!)

Discuss how this raises ethical issues in the following areas:

* Transparency

Whenever a computer system takes over a social function previously performed by humans, transparency of the process is extremely important. People would like to know *why* a partner is suitable, and such an algorithm would likely not be able to provide that information. Similarly, it would be an issue if a user already had a partner and the algorithm chose a different partner; if the original partner sued Google for their algorithm, there would likely not be a way of justifying the algorithm’s decision-making process.

* Bias

Another issue that such an algorithm could raise would be the bias that could be introduced to influence large populations of people. This could be the codification of existing biases that the algorithm has cemented and will continue to act upon or an emerging bias based on its training data. Such biases may further racial segregation due to the selected partners, and if the algorithm is not transparent enough, it could be a hard problem to fix. Similarly, if certain biases do exist, humans could try to give themselves conventionally better features to selectively influence the output of the algorithm.

* Human dignity

It is devaluing to ascertain that humans can be boiled down to a list of predefined characteristics that can be compared in a deterministic fashion. To say that someone is the best marriage partner implies that they will remain the best marriage partner for the rest of their lives, causing humans to feel like they cannot change or improve to be more compatible with a potential partner. Further, these relationships may have a stigma surrounding them due to their predetermined nature, where humans feel belittled due to their emotionless categorisation.

(5 marks)