HUL381/ELL457: Assignment 1

Maximum Marks: 6

Submission deadline: **31 January, in class**

Instructions: Please submit a *handwritten* response to the following question. There is no word limit, but as a rough guideline you should aim for 500–1000 words. You should refer to the assigned readings, and may also consult any other materials you find useful. However, your response should be written entirely in your own words. Any direct copying detected will be regarded as plagiarism and dealt with accordingly. If you refer to sources outside of the assigned readings, you must mention them in a list of references at the end. Please remember to write your name and entry no. on your submission.

The Turing test (`https://en.wikipedia.org/wiki/Turing_test`) was a proposed test for intelligence. The idea was simple but very powerful; the question being how we can characterise intelligence, and check whether a machine possesses it or not. We will discuss this test in detail a bit later in the course, but the basic outline for now is as follows: There are three rooms. In one room, we have an interrogator, in one a human, and in the third a machine (running some kind of language-processing software). The only form of communication permitted between the rooms is asking questions and getting answers via a terminal in a language that is mutually understandable. There is no other way for the interrogator to see inside another room. How can the interrogator come to know in which room there is a machine and in which a human? Turing proposed that if the interrogator cannot make out from the answers which respondent is human and which a machine, then the machine has shown itself to be 'intelligent'. See the setup in this schematic: `https://en.wikipedia.org/wiki/File:Turing_test_diagram.png`.

Now, the question for you is: Using a similar setup where in one room there is an interrogator (you) and in the other two rooms, in one a machine and in the other a human, can you develop a test for *consciousness*? If you cannot make out from the answers (depending on whatever test you develop) whether they are coming from a machine or a human, then by analogy to the Turing test the machine would have shown itself to be 'conscious'!

Please both describe your test (*i.e.*, which questions you would be asking, and how you think the responses would be indicative of consciousness), and also explain your reasoning for why only a conscious being would be able to answer the questions correctly or appropriately.