CS 486 – Assignment #1
Justin Franchetto
20375706

Section I – Turing's Imitation Game

1) *The Imitation Game* is a game is which a man (A) and a woman (B) sit in a room, and an interrogator sits in a separate room. The goal of the interrogator is to ask A and B questions and try to determine which of them is the man and which is the woman. The goal of A is to try and deceive the interrogation, while B is to try and aid the interrogator. To make the game more interesting, the role of A is replaced by a computer, and the interrogator is to determine which one is the human and which is the machine.

2) Turing suggests that by trying to imitate an adult mind, we will inherently come to spend a lot of time examining the human mind and the process which brings the adult mind to its current state; namely three components: the initial state at birth, education, and other experiences. Turing further suggests that instead of trying to simulate an adult mind, it might be better to simulate a child's mind and subject it to different forms of education to find which is most effective. Additionally, by preparing a machine to play The Imitation Game via this process of education and teaching, human fallibility will be omitted naturally instead of the machine being taught specifically to do so.

3) Turing raises several objections in his paper, two of which are:

    i) *Theological objection:* Since God gave souls to men and women (not machines and animals), and thinking is the function of the soul, then machines cannot think.

    Response: Turing says that this places a heavy restriction on the omnipotence of God and questions whether God could grant a soul to an elephant if he saw fit. It would be expected that God would do so if he also gave the elephant a brain capable of fulfilling the needs of the soul. Turing then says that the exact same argument can be made for machines being given a soul, but it is harder for people to accept.

    ii) *Argument from continuity in the nervous system:* The nervous system is not a discrete-state machine and a small error in information about an incoming impulse may cause a large error in an outgoing impulse. Since this is the case, we cannot mimic the nervous system with a discrete state machine.

    Response: Indeed, a state system is different from a continuous machine, but in the Imitation Game, an interrogator would not be able to take advantage of this. Turing then gives an example of the differential analyzer (a continuous system) in which the computer is able to mimic it effectively, making it difficult for the interrogator to distinguish between the two machines.

4) Some of the objections raised are still relevant today, notably:

    i) *Head in the sand objection:* The consequence of machines thinking would be too dreadful, let us hope and believe that they cannot do so.

Explanation: This is still relevant today because there are many people who fear a scenario in which robots will overtake and eliminate the human race, or something similar. Such a scenario is unlikely in the near future and it would be foolish to ignore AI, a tool with multidisciplinary implications and uses, on the grounds of potentially "dreadful" consequences.

ii) *Arguments from various disabilities:* You have made a machine that can do several tasks, but you will never be able to make one that can do task X.

Explanation: This argument is especially relevant today, as we are making rapid progress in both hardware and software. There are always tasks which people claim computers cannot do, such as translation of language or creative tasks. Yet, there are already programs that can compose music or write a newspaper article. Since these and other historically "impossible" tasks are being completed by machines, it seems silly to claim that there are tasks in which a machine could not completed. Given sufficient time, there will be new advances and new machines to perform even more difficult tasks in the future.

5) Katrina LaCurts addresses the concern that the Turing Test does not provide us with a gradient of intelligence. It is suggested that because the test is binary in nature, it does not allow for the interrogator to give a level of intelligence to a machine. LaCurts provides two responses to this criticism:

   i) If machines learn in a similar way and at a similar rate as humans, such that a machine could be as intelligent as a five year old but not an adult, then the machine should simply have to play against a five year old in The Imitation Game.

   ii) In the event that machines learn differently from humans (or more quickly), then we may not even need such a scale of intelligence. LaCurts uses the comparison of the vocabulary of a five year old to that of an adult; while it takes humans many years to acquire a large vocabulary, once a machine reaches the point of intelligence, it may become trivially fast for it do so.

   Citation:
   Katrina LaCurts; Criticisms of the Turing Test and Why You Should Ignore (Most of) Them (http://people.csail.mit.edu/katrina/papers/6893.pdf, page 3)

Section II – Searle's Chinese Room

1) Weak AI suggests that the value of a computer in the study of the mind is only that it is a useful and powerful tool, while Strong AI says that the computer is not only a tool, but that a properly programmed computer really is a mind itself and can be said to understand and have cognitive states.

2) The Chinese Room is an experiment Searle proposes in which he is placed in a room, without knowledge of the people outside the room. First, he is given a batch of Chinese writing, of which he has no understanding of. Second, he is provided with another batch of writing in addition to a formal set of rules (in English) which are used to connect the first batch of writing with the

second. Finally, a third set of Chinese writing, along with English instructions which explain how to connect the third set with the first two, and how to return Chinese characters in response to symbols in the third set.

Unknown to Searle, the given Chinese writings are: a script, a story, questions and the response returned by Searle are considered to be answers to those questions. The question here is if the responses given by Searle are indistinguishable from those given by a native Chinese speaker, can we say that Searle understands Chinese?

3) Searle thinks that the Chinese Room invalidates Strong AI because it does not fulfill the two main claims of Strong AI, which requires that the machine understands the story (and provides answers) and that the program helps to explain how the human mind understands the story and provides answers. Searle claims that the person (or machine) in the Chinese Room does not understand Chinese, the story or the answers it gives, and that the operation of the program does not help to explain how a human understands a Chinese story or provides answers about it.

4) Searle discusses a few replies, two of them are as follows:

   i) *The systems reply: "*While it is true that the individual person who is locked in the room does notunderstand the story, the fact is that he is merely part of a whole system, and the system does understand the story"

      Critique: Searle suggests that even if the person in the room where to memorize all of the elements of the system, including the rules and the Chinese symbols and performs all operations in his head (therefore encompassing the entire system), that the person still understands nothing of Chinese. Furthermore, since the person has everything the system has in their mind, that the system also understands nothing of Chinese.

   ii) *The combination reply:* Combine the other three replies to obtain a robot with a brain-shaped computer in its head, which is programmed with all of synapses of the human brain and behavior indistinguishable from human behavior. At this point, we would have to say that the system has intentionality.

      Critique: Searle says that in such a case, it would be rational to accept that the robot had intentionality, but only if we knew nothing further about it. However, he says that this is not really of any help to Strong AI because according to Strong AI, by instantiating a formal program with the right input and output is constitutive of intentionality. In the case of the robot, the attribution of intentionality has nothing to do with formal programs and is only based on the idea that if the robot looks and acts like us, then we assume it has mental states like ours that cause such behavior, unless proven otherwise.

5) Searle gives a number of differences between human cognition and AI:

   i) People claim that human minds do something called "information processing" and similarly, a computer does the same thing by using a program. Searle says that the notion of information is different in these cases. Humans process information by

reflecting on problems or reading and answering questions. Contrary, computers and their programs only manipulate formal symbols.

ii) Though it is tempting to guess that a computer has mental states similar to human states if it has I/O patterns similar to that of a human, we can ignore the impulse to do so once we that the system can have such capabilities without having intentionality. That is to say, although a computer may mimic the behavior of a human mind, it does so without understanding or intentionality.

iii) Searle says that Strong AI only makes sense if we use the dualistic assumption that the brain does not matter where the mind is concerned. It is the programs that matter and they are independent of their realization in their machines. He then states that the procedures in a human mind are dependent on chemical properties of the human brain, which would not be possible to mimic with an AI if the mind is to be independent from the brain.

6) I disagree with the idea that he understands nothing after becoming proficient with operating on the symbols to the point where he is indistinguishable from a native speaker. I believe that by being able to use the underlying rules of the language with such proficiency, he does have knowledge of the language, even though he might not be able to speak it. Similarly, if a program contains an algorithm that can produce indistinguishable answers, I believe that the algorithm can provide some level of insight into understanding how humans produce such answers.

<u>Section III – Hawking</u>

1) This article is concerned with the fact that we cannot predict what will happen when human intelligence is magnified by the tools that AI will provide, and the creation of such AI could prove to be disastrous if we do not learn how to avoid and manage the risks associated with it.

2) The scenarios proposed by Hawking are quite plausible, especially in that it is conceivable that a program may take data about the world as input, process and produce a new version of itself that was improved upon based on the input data. Additionally, there are already programs and algorithms used in financial industries to help determine the correct course of action. With such a large history of financial data at its disposal, it is more than possible that a machine would learn how to be more successful than financial experts. Finally, while it is possible that a program might design better or more deadly weapons, I think it is unlikely that it would manage to do so uninterrupted and there would be human intervention before it was successful.

3) In order to avoid the risks of the scenarios proposed by Hawking, it is important to consider these sorts of risks, not only when designing new systems or programs, but also when deciding the level of responsibility we assign to them. While these scenarios are not likely to happen soon, they are also plausible, and even more so if limitations on responsibility are not placed on such programs. If a program is allowed to have free reign, or is given access to too much information, then the future consequences might be troublesome.