

Qthulu AI Essay

Leif Schelin, Björn Persson Mattsson, Chen YuTing

1 Introduction:

It all started with this question: “Can machines think?” We as human civilisation had been creating tools to solve problems since beginning of our time. As we grow better at creation of tools, collection of tools forms into machines, more and more complexity of machines we built upon each other. Then it comes with computer, a machine first designed as tool for computation aid, later we gave them “programable” ability to improve themselves. Inevitably, we started to feel the uncertainty of what can this leads to.

Turing himself also avoided answering this question by introduced another imitation game experiment. Can machines behaves indistinguishably as human from another human observer’s perspective? This experiment itself raised another discussion among computer scientists and philosophers. As researches in various areas improved from time of Turing, instead of asking if machines think, we would like to explore the question of whether machines can have minds.

1.1 Strong versus weak AI:

Before we dive into the discussion regarding to machines with their minds, a clear framework should be illustrated in order to achieve most fruitful discuss. We often hear “strong AI” and “weak AI” in domain of artificial intelligence research, but these phrases can have actually two sets of meaning. Since the overlapping on terms, one often accidentally associated with mixture of both concepts. We would like to clear the distinctions here:

- Strong / Weak AI used in computational theory of mind:
John Searle is often regarded as the first person to use “strong AI” term in this definition, in his “Minds, brains, and programs” published in 1980. According to John Searle, “Strong AI” claims that programs (or programmed computers) which reacts correctly according to right input with right outputs, should be regarded as possessing a mind just

alike humans possessing our minds. And “weak” AI claims this reaction is merely an imitation of mind, which cannot be accounted as real mind. Searle’s famous Chinese room thought experiment, which we will discuss more in detail in following section, serves as refutation to strong AI claims. In which he introduced “intentionality” is not possible to be included in any programs, or to say “intentionality” is not possible to be modeled. A man in a Chinese room can perform perfect task following the instructions and rules while he/she still knows absolutely nothing about Chinese, no intentionality is at all involved in tasks carried out. Another way of understanding the difference is by looking into the actual actions carried out in Chinese room experiment, one can describe the instruction-following guy only access the syntax rules of Chinese, not the semantics of Chinese he/she been given.

- Strong / Weak AI used to describe ability to solve general problems within single agent:

Sometimes also referred as full AI, strong AI in this branch can be defined as agent which can demonstrate ability to perform any intellectual task human capable of (Kurzweil, 2005). That is an agent can solve not-predetermined tasks matching to human level. If there exists one artificial intelligence agent which can pass the Turing test, one can classified it as strong AI or full AI because it demonstrated ability to perform tasks in terms of natural language dialoge equivalent (or indistinguishable from a third person’s observation) to human level. Passing Turing test might not suggest the agent can solve “any” task human can, which can be too vague too achieve (infinite list of task). However, general agreements on general intelligence must accomplish following:

1. reasoning
2. learning
3. planning
4. natural language communication
5. Knowledge representation (KR)

We refer frist definition of strong AI as Searle’s idea on strong AI and later one as Turing’s. Although discussion in either one of these branch often leads to some common arguments from another branch, it is better to clarify at beginning.

1.2 What is a mind ?

- Does simulated minds count as real mind?

As we mentioned in Searle's work, he refuted the claim of simulation of mind itself should be counted as real mind. Till today this argument still remain topic of debate. So why do people waffle about whether a simulation of mind can be attributed as real mind or not? Another famous question of philosophy could be linked to this simulation problem, the mind-body problem. Rene Descartes is commonly acknowledged to be the first person who formalised this problem: "how the immaterial mind (or soul) could influence the material body?" and from where is the boundary of material crossing to immaterial world if one regard information flow as input and mind as processer of information. Like most philosophical problems, answer to mind-body problem seems to depend largely on one's personal believe.

- Border between mindful creation and not?

- (Maybe a way to separate): Purpose-driven OR Meaning driven ?

2 Discussion:

2.1 How to test if one creation possess mind or not?

Since it is quite hard to know what a mind is in general we instead attempt to answer the question of whether a machine can possess a human mind, that is think like a human. To determine if machines can possess minds, one can go through some formulated tests:

- Turing test: Alan Turing proposed a game of imitation to test if a program can mimic human behaviour in his classical paper [REF here]. If a program can answer a set of written arbitrary questions that requires abstract reasoning and be mistaken for being human, Turing concludes such a program to be .
- College test: Another test was proposed by Goertzel [REF here]: if a machine could get a college degree, then it must be .
- Nagel Bat argument: The two tests above only test the behaviour of an agent, that is AGI, the "first" version of strong AI. When it comes to minds, consciousness and understanding others, Nagel made an interesting

point [REF here]. He argues that a human can never understand what it is like to be a bat; we can only imagine what it would be for us to be a bat which is different from understanding what it is like for a bat to be a bat. A related example would be how does one's native tongue (ex Swedish) sound like to a foreign person (ex Chinese), we can only imagine what it would be like because we only know what it sounds like to a Swedish person. If we accept Nagel's argument, one can draw the parallel to computers. Thus a bat expert might be able to predict and even imitate a bat while not being able to understand it internally, a computer could possibly imitate and predict human behaviour whilst not understanding them emotionally. For example, things like laughter, grief and art could be predicted but not understood.

- Chinese Room Experiment: In any case, even if a machine would successfully understand all abstract parts of being human, it could be very hard to tell from just a smart imitating machine. Searle illustrated this in his famous thought experiment: the Chinese room [REF here]. If an agent that speaks English natively but doesn't understand any Chinese is sealed away in a room together with a list of very thorough instructions of how to manipulate some input characters (presumably) written in Chinese, he might be able to produce some Chinese text but it doesn't mean he understand what he is doing. However, the people on the outside might not be able to distinguish that agent from a native Chinese speaker, assuming the instructions are sufficiently detailed and intelligently designed. Likewise, if a computer successfully could understand humans, we couldn't really tell the difference of that from a list of intelligent instructions executed by a thorough but non-understanding agent.

2.2 How to create simple mind ?

Some examples of what might be considered simple minds can be found both in animals and in robotics. (Judging whether something is a mind or just an intelligent mechanical/deterministic behaviour is definitely non-trivial...)

- Slime molde
Slime mold are unicellular amoeba which, when food is in short supply, can form into networks that coordinate and have behaviours similar to animals with simple brains. An experiment showed that a mass of slime mold was able to generate a rail network over Tokyo that was strikingly similar to the network actually used [REF].
- Bee drones
In robotics, a research team at Harvard is aiming to create microrobots

with a size and behaviour similar to bees [REF]. The researchers hope that these robots will be able to efficiently collaborate as a “colony” in order to act as robotic pollinators. Another research team from the universities of Sheffield and Sussex have made progress in simulating parts of a bee’s brain and uploading it to a UAV (drone), allowing it to fly and use simple navigation to avoid running into obstacles [REF].

3 Conclusion

3.1 Does it matter ?

- Before we start, you have to know they don’t care!
Simulation of intelligence or real intelligence is in many sense not important issue for AI researchers working on delivering programs that finish designated tasks. As much as you wouldn’t care if a screwdriver is actually a simulated solution or a real solution as long as it fixes your Ikea beds. AI researches should therefore in absolute need as they produce more agents solving our problems in life, and some of which are practically unsolvable without these agents. We will return to this point when we discuss ethical perspectives of developing technology in later, but now leaving you a nice question to ponder, screwdriver is handy tool but in wrong hand could be weapons to harm others, did the inventor of screwdriver thought of this perspective and wouldn’t it be more safe to not release this potentially dangerous version of screwdriver and keep refining until he/she come up with a harmless version?

Well, since AI researchers are not so interested in debating mindfulness of machines, it shouldn’t matter so much for discussion right? Why not consider following question: “If a computer, provided with correct software, can be considered as a creation with mind (and in some degree, consciousness), would it be crime of murder if I pull off its plug?”

- Assuming a computer with proper software accounts as mindful creation, we wont be able to find out!
- Does physical differences impose different level of mutual understanding?

- If we regard programs as form of intelligence, machine intelligence shouldn't be confined (or defined) by limitations of humans'