Future Artificial Intelligence Philosophy and Sentience

Abstract

This paper outlines: the ethics behind the development of advanced, and sentient artificial intelligence, the historical AI philosophy's of Marvin Minsky, and the Chinese Room argument, the hopes and fears for the outcomes of future artificial intelligence, the current efforts by IBM to create a computer that mimics the human brain, and by Google to create true artificial intelligence through the use of machine learning, and the final topic covered is the concern of how we must go along determining sentience in things, and in this case, computers.

Introduction

Artificial intelligence is a growing trend in computer science, and with it comes the need for new information about it. We need to philosophize the possible outcomes of advanced AI , we need to have expectations, and we need to have an idea of what we are actually working to build in AI. We need to explore the possibility of creating advanced, and possibly sentient AI, the ethics of creating such AI, the possible things that the systems will be used for, such as warfare. Finally, we need to decide whether it is actually helpful to us as humans to develop something artificial that has the same abilities as we do to process information in real time and experiencing subjectively.

Sentience is essentially what it means to be a living, thinking, creature, person, or thing. We as humans struggle to recognize sentience in creatures or things, so we must determine what sentience actually is in terms of computers. There are numerous obstacles we face in creating sentience in computers, and there are many unpredictable variables and situations that may take place with the creation of advanced AI. We have to think deeply about the morality, ethics, and the possible positive, or negative outcomes of creating an artificial intelligence that is capable of feeling, perceiving, and experiencing real life events subjectively. Important aspects to keep in mind and hypothesize about future AI could be; what problems will we possibly face in the near future of AI, how we can possibly ensure that AI continues to be consistent with human aims and goals, how we will we determine morality of machines themselves, and how can we address the possible issue of creating an AI that is more intelligent than humans, and ensuring that this AI does not abuse its intelligence.

I. Ethics Behind Artificial Intelligence

An important aspect of the possibility of developing intelligent or sentient machines is the ethics of creating them, and the possible threats these machines may pose. If a machine is given the capability to understand natural language such as English, there is possibility for it to be used for malicious intent to read every email, text, and listen to any phone call it pleases; with the fear of being spied on already being in the minds of people, especially citizens of the United States (NSA), the thought and fear of

machines being capable of fully understanding humans is warranted.

Computers are already being used to replace humans in doing specific jobs and tasks, but some argue that computers should never be used in jobs that require respect and care for others. Joseph Weizenbaum was a computer scientist who believed that there were jobs that should never be replaced by machines. He believed the most crucial of these jobs were; soldiers, judges, police officers, therapists, elderly care workers, and customer service representatives.

An obvious fear of advanced AI is the use of them in warfare, but there is also a hope among many that autonomous AI will be in fact used in warfare. A fully autonomous weaponized robot could make decisions more effectively than a human, there would be no grey area of morality or decision making as there is with human soldiers. Some fear that AI weaponry will be extremely detrimental to society and humans, and possibly even cause the downfall of mankind. An outspoken believer of such a possibility is the great Stephen Hawking, who believes that computers will indeed cause mankind to stumble. Along with the weaponization of AI comes the area of fictional, and science-fiction based AI hypotheses. Movies such as The Matrix or Bicentennial Man both deal with the idea of sentient artificial life.

II. Historical AI Philosophy

We need to study and understand more closely what sentient AI and artificial consciousness would imply. There are multiple aspects to consciousness; awareness of one's surroundings, ability to maintain a memory, capability of learning, natural anticipation of events, and experiencing subjectively. Some of these things have been studied and hypothesized through recent history. Marvin Minsky, a computer scientist in the mid 1900's, laid out the essential elements that a system must have in order to be an AI.

Igor Aleksander, a professor at Imperial College in London, England, suggests that there are actually 12 aspects of computer consciousness, and suggests that we need to define how these aspects should be engineered into a computer or machine to ensure its success. Some of the more essential principles in Aleksander's list are; awareness of self, will, instinct, emotion, and self reliance, or in a computer's case, auto programming and debugging. There are debates and theories for, and against whether a computer can even have a mind or mental state of its own. John Searle, an American philosopher offers us the "Chinese room argument," which argues that no physical symbol system, such as a computer, or any other non-human object, can have a mind. Before John Searle's "Chinese Room Argument," a German man from the late 1600's, named Gottfried Leibniz, had a very similar, but more basic explanation of why a system cannot have a mind or mental state of its own. We will be focusing on Marvin Minsky's work in Computer Science, and a brief explanation of the "Chinese Room argument."

A. Marvin Minsky - Artificial Intelligence Pioneer

Marvin Minsky was the first person to coin the term "Artificial Intelligence," and he defined and explained exactly what it was. Minsky is famous for his computer science accomplishments in the mid 1900's, and he is most notably recognized for his 1961 paper,

"Steps Towards Artificial Intelligence." In this paper, Minsky outlines the 4 essential skills a computer must have in order to be classified as an artificial intelligence, and in some ways, the 4 problems that must be solved in order to obtain artificial intelligence.

Search - The first skill is the means by which a computer, or even a human, searches for solutions to a proposed problem. It is a tedious process to investigate and research to solve a problem, but if a machine can be designed to produce answers and reduce the amount of time to search for solutions it could be of value.

Pattern Recognition - The second skill is the ability to recognize complex patterns to reduce the amount of time required to solve a problem. In order to be resourceful, a machine of artificial intelligence must be capable of separating different problem situations into categories to maximize the effectiveness of the machines methods and capabilities. This is a very complicated and difficult issue. The simplest way that this can work is for the machine to compare the objects being recognized against things it already knows, or standards. When it comes to more complex recognition, the problem of not having enough systems to compare against, and the ability to produce such systems arises.

Learning - The third skill is the ability to learn and apply new systems and functions. Humans naturally learn from past experiences, and are capable of applying methods to situations that they have learned from previous situations naturally. In computers, AI must also use a "basic learning heuristic," or the capability to use past experience in the present. Minsky states that the way that AI needs to do this is to use "success-reinforced decision models," and to break down problems into smaller problems that it already knows it can solve. And the final problem computers face in doing this is the ability to credit the success of solving a problem to its different methods.

Problem Solving and Planning - The fourth skill is the ability to solve problems and to plan how to solve such problems. When solving complex problems, a plan is typically required to solve it, and the problem solver becomes involved with a large amount of subproblems. These subproblems must be separated and solved separately to solve the larger problem, and AI must be able to do this by using, as previously stated, heuristics. AI must be able to break down a problem and decide how it should be split into subproblems before it actually begins solving.

These 4 skills are still very relevant today. Minsky outlined what it essentially means for computer to be "intelligent." Minsky writes in the "Induction and Models" section of his paper that he is "confident that sooner or later we will be able to assemble programs of great problem-solving ability from complex combinations of heuristic devices - multiple optimizers, pattern-recognition tricks, planning algebras, recursive administration procedures, and the likes."

B The Chinese Room

The idea of "The Chinese Room" is to first imagine yourself as a monolingual English speaker who is locked in a room, and is then given numerous Chinese writings, Chinese scripts, and a set of rules and instructions in English to compare all of the Chinese symbols so that you as the person locked in the room has the ability to produce Chinese symbols as an output to questions that you are asked, or procedures that you are tasked to complete. Searle argues that

this is exactly how a computer, or AI works. The experiment states that the person in the room does not actual understand the Chinese symbols, therefore it is not cognitive or sentient on the material, but is only given instructions on how to interpret them. It is implied that this means that a computer cannot truly become sentient or human-like because it does not truly understand, just as the person in the room does not understand, but instead it is what Searle refers to as "Strong AI."

III. Hopes and Fears of Developing AI

A. Hopes

When we think about what we "hope" the outcome of advanced AI will be, we tend to think about how it will benefit us as humans, and how it will contribute to society and culture. The most common ideas of positive outcome of AI are obvious; increased problem solving abilities by computers, and the usage in developing cures and medicines for human disease. AI will help us solve complex problems that we are simply incapable of solving with current technology, mostly in the fields of science. We will most notably be capable of studying quantum physics at a more in depth level, and possibly simulating things that we cannot actually observe as humans, and the simulation of life itself. Computers will hopefully help us develop medicines and procedures that will help save millions of lives; artificial intelligence and machine learning may even lead to the cure for the world's most devastating illnesses such as HIV/AIDS and cancer.

B. Fears

Now, to consider some common fears and concerns of the development of AI. The use of AI in service based industries or processes could be extremely beneficial due to optimized efficiency, but we must also consider whether it is moral to replace these jobs with computers while there are still many people who depend on these jobs. As mentioned briefly in the introduction, there are some jobs that could be replaced with AI, but the question we must ask is, should these jobs be replaced with AI. Science fiction has always portrayed AI as dangerous, militaristically unstable, and willing to betray the trust of humans, and abuse their intelligence. This portrayal of evil intelligence is a concern we must actually consider. There needs to be a guaranteed safety from the abuse of AI in warfare by a society, the possible ability for an AI to act on its own in a militaristic manner, and the possibility of the AI surpassing the general intelligence of humans. These are just a few concerns we need to consider, and the solutions to these concerns are necessary to ensuring the success of such AI.

IV. Current Efforts Towards, and the Achievability of Developing AI

It is hypothesized that in order for us as humans to build sentient AI, and artificial consciousness, we must understand more deeply how the human brain actually works in creating consciousness, and how we can build a machine to mimic these functions in a quick, efficient manner (hopefully with the coming of quantum computing, speed will not be an issue). There are numerous theories of the most likely way we will succeed in

building an autonomous, self reliant, sentient artificial intelligence in the 21st century. The one's that will be quickly discussed through the examples of IBM and Google's success are through the development of an artificial brain that will likely consist of a multitude of individual processors that do the same processes as individual neurons in a brain, and the use of neural learning and networking.

A. IBM

IBM, the American technology company, is well known for their development of artificial intelligence. They built the AI recognized as "Watson;" Watson is a platform that uses natural language processing and machine learning to solve problems, and extract information from unstructured documents and data sets very quickly. Watson is arguably one of the most successful attempts at artificial intelligence, yet, it is not sentient, but IBM is making progress towards creating a cognitive computer that has been developed to mimic the neurons in a brain. The idea behind this theory of developing AI is to actually recreate a human brain using artificial neurons that spike randomly to store and access information just as a human brain does.

B. Google

Google has two major teams working on future AI and sentience in a sense; the Google Brain team, and the Deep Mind team. Google Brain has recently delved into a project where the goal is for two artificial intelligences, given nothing but a similar cryptographic code, to develop their own cryptographic algorithm, neural network, and language that they then send messages with to each other. All while a third artificial intelligence was created to try to crack the two other AI's code. What is incredible is the fact that the AI's were not given explicit instruction on how to create the neural network, or even what crypto techniques they could use, and the system is fairly successful in establishing a neural network.

The systems are not fail proof, but the fact that they can successfully establish communications via machine learning at all is incredible, let alone at a rate where it works most of the time it works. The machines are not connected in any way other than the two communicating AI's having the same starting "key" to work with. Other than the key they are given when they are booted up, no human instruction is given to the computers to develop this neural network. This development of machine learning shows exactly how advanced we are in the current age; these computers are actually capable of communicating with each other in a cryptographic manner that the human developers can't even figure out.

V. The Challenge of Determining Sentience

A part of sentience is is the warranting of moral consideration, and whether we should owe moral consideration to something that is only a thing, not a living human, or creature. Things cannot be hurt, they have no personal interests or concerns, and they have no preferences. But if we give a "thing" the ability to feel and perceive for itself, an important aspect we need to approach is how we should go about offering moral consideration towards these things.

Humans are not naturally good at recognizing sentience, we have made considerable changes in our beliefs about

sentient life forms throughout our recent history (17th century - present). Descartes, a famous philosopher claimed that animals were not sentient, but were merely biological machines that acted by reflex. Following Descartes's claim, a machinist named Malebranche claimed that animals were incapable of feeling pain, but finally in 2012, a group of scientists at Cambridge declared that humans are not the only living creatures that feel pain and experience subjectively, and in fact, animals can as well. An argument as to why we have trouble detecting sentience in other creatures is because we as humans are so different from the things, or creatures we are trying to determine sentience in. So how can we truly determine if a machine is sentient if it is so different from a human?

The "Turing Test" is a test that Alan Turing, an English computer scientist from the early 1900s, developed to test a machine's ability to portray intelligence equivalent to that of a human. The Turing Test has been used to a great extent to test the intelligence, and in turn, sentience of many systems ever since Turing defined the test in his paper, "Computing Machinery and Intelligence" in 1950. The test is criticized for not being entirely accurate, meaning that non-intelligent systems can trick the tester into believing that it is thinking. There are many problems with the Turing test, and we must consider alternatives to determine sentience in modern AI.

Conclusion

The need to understand the potential ethics, outcomes, and viability of artificial intelligence is ever growing. The potential of creating something that is capable of experiencing subjectively as a human can is very complex. There are many aspects to consider, the following of which were researched in this paper; the ethics and repercussions of creating such machines, past AI philosophies and dreams, the hopes and fears of creating AI, and finally how we as humans can determine sentience in something other than humans. We as humans naturally fear things we do not understand, or have not experienced, and in the case of AI, there is fear even from the things we create. We learn from past mistakes and successes, and we can learn from those who come before us. Marvin Minsky's pioneering paper on artificial intelligence is still relevant today and can be used to philosophize the necessities for AI to be successful. AI is seen by some as dangerous, and possibly a weapon; while others see AI as an opportunity for greatness, an opportunity to create something that has been dreamt of for years that will benefit the common good. There are potential pros and cons of creating advanced, and possibly sentient AI, and we must consider these possibilities closely while we continue to make leaps and bounds such as IBM and Google are, towards the pinnacle of computer science, the development of true artificial intelligence.