Module	Assessment Type
Concepts and Technologies of AI	Report

ARTIFICIAL INTELLIGENCE AND ITS ETHICS

Student Id : 2049867

Student Name : Roshan Parajuli

Section : L5CG3

Module Leader : Mr. Prakash Gautam

Submitted on : 19-02-2021

Acknowledgment

I would like to express my deepest gratitude to my Artificial Intelligence and its ethics lecturer Sachin Kafle and module leader Prakash Gautam for giving me this opportunity to research Artificial Intelligence and its ethics. Their guidance and continuous motivation helped me a lot in completing this report. I have taken efforts in this project. However, it would not have been possible alone. I would like to thank all of them and my colleagues for setting up the competitive environment for research and knowledge gain.

I would like to extend my thankfulness to the University of Wolverhampton as well as Herald College for providing the module contents and good cooperation during the whole semester and for providing a great topic for research which helped me learn a lot about the topic.

Table of Contents

1. Artificial Intelligence

1.1. Introduction

The definition of Artificial Intelligence, or AI, is very uneasy to build because there is no shared definition that is agreed upon by everyone. Two approaches to the definition of AI exist. One is the human-centered approach and another is the rationalist approach. By no means, one is better than the other. The human-centered approach incorporates the assumptions and theories about human behavior while a mixture of mathematics and engineering is used in the rationalist approach (Russell & Norvig, 2010). Artificial Intelligence can vaguely be defined as a computer science division that aims to mimic or simulate human intelligence on a machine to allow computers to perform tasks that usually involve human intelligence (O'Carroll, 2017). The ability to carry out tasks under control and learn to improve performance is usually essential to AI and its primary purpose is to accomplish the replication of a human being's cognitive ability.

1.2. Levels of Artificial Intelligence

There are three levels of Artificial Intelligence if classified in terms of the intelligence embedded into it: ANI, AGI, and ASI.

i. **ANI (Artificial Narrow Intelligence)** - The name justifies the characteristic of this artificial intelligence. It has a very narrow range of skill sets. All the artificial intelligence we have accomplished effectively to date is narrow. Even if the skill set is great, it focuses on only one task. The perfect example would be of an IBM computer named Deep Blue who beat the world chess champion after a six-game match 24 years ago (IBM, 1997). Deep Blue was great at its job but its intelligence was narrow in the sense that playing chess is the only thing it did. Other notable examples of artificial narrow intelligence are google search, Facebook newsfeed, email spam filters, and even self-

driving cars and so much more. It is either reactive or has a limited memory capacity. It is used almost everywhere in the present context.

ii. **AGI (Artificial General Intelligence)** – It is also called Strong AI or Deep AI. It is a level upper than the ANI as it can reach and even pass the intelligence level of a human being. In any given situation, AGI can think, comprehend, and behave in a manner that is indistinguishable from that of a human being. Alan Turing, an English mathematician, logician, and cryptographer who is also called the father of modern computing and also the pioneer of machine learning during the 1940s and 1950s proposed a test called the Turing test in which he mentioned that the machine is only said to be intelligent if the questioner in a different room, who is handed the role of asking the same questions to the human and the AI, cannot decide which the AI is and which the actual human answering their questions is. It is not achieved at present because AI researchers would need to make it conscious by programming a complete range of cognitive skills. It has the potential to reason, plan, solve challenges, consider complicated concepts, learn quickly from experience with the help of machine learning. Once the functionality of a human brain is properly comprehended, the growth is AGI will be rapid.

iii. **ASI (Artificial Super Intelligence)** – It is a hypothetical AI that is much cleverer than the smartest human brain in intellectual ingenuity, general knowledge, and social skills and almost on any subject. The future it holds is still unknown. It might be dystopian as all the science fiction movies portray or it can be utopian and humans and ASI will be in peace. Most people including Elon Musk fear ASI because of the self-awareness it has. Its impact on civilization, its survival, and our way of life is a pure conjecture. (Strelkova, 2017)

2. History and Leaps of Artificial Intelligence

The idea of Artificial Intelligence started with the simple “heartless” Tinman from “The Wizard of Oz” and continued with many other movies to follow. The 1950s had a rising number of scientists, mathematicians, and philosophers having the concept of Artificial Intelligence inside their heads. Alan Turing loved to explore the possibility of artificial intelligence (Jacobson, 2014). In 1950, he wrote a paper on Computing Machinery and Intelligence where he addressed how intelligent machines are designed and how their knowledge is tested. Five years later, a historic conference was initialized which concluded the fact that the AI was achievable. Around the 1980s, AI became an industry in itself (Harvard, 2017). In the present context, AI is widespread and its applications are massive which include Robot-Assisted Surgery, Virtual Nursing Assistants, Administrative Workflow Assistance, Fraud Detection, Dosage Error Reduction, Connected Machines, Clinical Participant Identifier, Preliminary Diagnosis, Automated Image Diagnosis, and Cybersecurity. From robots beating humans in August 2001 in a virtual rivalry in financial exchange by making 7 percent more cash than the human in the finance industry to the replacements of human opponents in games, AI has leaped where it originally started. Even after all this, we have just reached a very premature and dull age through a roller coaster of success and setbacks.

3. Current Advancements in AI

NASA's latest Mars rover named Perseverance landed on Mars on February 18, 2021, following a six-month flight. Once it landed on the Martian surface, it extensively is using artificial intelligence as it starts to travel across the landscape. According to NASA, Perseverance would use X-Rays to hunt fossils from millions of years ago. With a challenging road before it, the journey of perseverance would not have been possible without the AI-powered device called the Planetary Instrument for X-ray Lithochemistry or PIXL in short. (NASA, 2021)

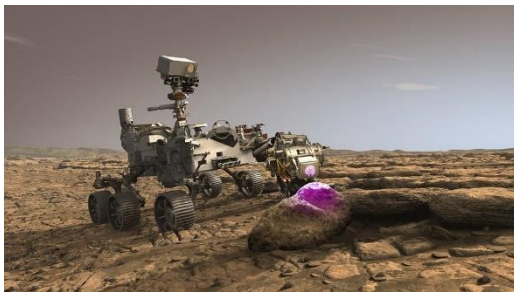

Figure 1: Perseverance using PIXL

In late 2018, AI artwork was sold for $432,400 which signaled the arrival of AI art on the world auction stage. The way that the algorithm made such a painting that is indistinguishable from a human one is by its composition – The Generator and the Discriminator. The job of the generator is, as described in its name, to generate a painting after being fed around fifteen thousand paintings that were painted around the 14th century, and the job of the discriminator, also self-explanatory, is to look for the difference between a human-made painting and the generated painting.

Figure 2: Portrait of Edmond Belamy by Generative Adversarial Network

A piece of news on ScienceDaily states that the robots can detect a range of physical interactions using cameras and shadows without even relying on touch at all. That concludes that the robots with better and better artificial technology are getting better and better at feeling human touch even though they may not be in touch with human feelings as of now. (ScienceDaily, 2021)

Not only that, there are negatives sides in the field as well. In 2016, Microsoft's AI chatbot turned out to be racist in less than 24 hours which is a concerning factor on

how AI can do unpredictable things even if manufactured with utmost precision. Also, On 2-18 a fatal Uber crash in Tempe, Arizona was widely believed to the world’s first death by a self-driving car. From a funny incident of AI following the referee's head throughout the game thinking it was a ball to the Sophia robot cheerfully saying that she would destroy humans, a lot of uncertainty awaits the human race.

4. Ethics of Artificial Intelligence

Artificial Intelligence would offer substantial and varied socioeconomic benefits from increased productivity and efficiency to solving certain daunting global issues, including climate change, illness, and war. As it is going to be responsible for the development of humanity shortly, it is very important to be concerned about its system and the risks it brings with itself and also how it would be possible to control them effectively. There is a great dilemma on the way of developing as well as using this technology to not only make people's life better but also be ethically and morally accepted. Algorithms are very capable of reinforcing present prejudices and discriminate. Not only that, but it can also jeopardize our life, exploit us, and have deadly consequences. AI ethics concerns itself with how the authorities and operators should behave to mitigate the ethical risks that AI could generate in society from the unintentional or intentional misuse of the technology. Isaac Asimov, an American writer, and professor at Boston University proposed three laws of robotics. In short, they are about how a robot should not injure a human being, obey the orders except the ones which would conflict with the first law, and protect its existence as long as it obeys the first and the second law which would very much be applicable in terms of artificial intelligence (Cambridge, 2011). Once we have understood the technology well, it would be a great time to shape the societal response which includes certain regulations and laws.

A poorly designed superintelligence is conceivable to take over the flawed objectives that it has been provided and applied. To tackle the problems that might be faced, different institutions have contributed to the creation of an AI ethical system. Even though their opinions vary, they have a growing consensus. New research on the global landscape of AI ethics guidelines reveals that AI ethics have converged on five values at a surprisingly rapid pace which includes non-maleficence, accountability, transparency, justice, and respect for human rights (Jobin, et al., 2019). Since the ethical issues related to AI are numerous, sometimes it is very frustrating to see why a particular problem is caused because of its lack of transparency. That is the exact reason why it is very important to create potent and scalable algorithms that are also transparent for inspection. Alongside transparency, other important factors include

accountability and being open source. Unfortunately, the open-source code does not mean that the AI implementation would be transparent as it learns on its own and can be unpredictable to a great extent. AI systems can be very biased depending on the data that is being fed to them and is being increasingly innate in facial and speech recognition and are very vulnerable to biases and faults that the developers introduce. An example of it would be Amazon completely halting its AI hiring software because the algorithm preferring male applicants over women as the data it was fed contained mostly of the male applicants (BBC, 2018). For the decrement in the bias, data that is used to feed to the machine is needed to be properly researched and well documented, and more awareness should be brought up about biased AI. The economic incentive can override the commitment to ethical principles and this suggests that AI systems are created and implemented for reasons other than those of social principles and fundamental rights such as profit, non-maleficence, fairness, and explanatory existence (Hagendroff, 2019). There is also a big question raised on who would take the blame if the AI system miserably fails at its tasks - either the programmers, the end-users, the person who came up with the idea of the system, or any other person directly or indirectly involved in it. As a machine does not have a feeling of sentience (capacity to feel pain and suffering) and sapience (being self-aware and a reason-responsive agent), the factor of responsibility is still being researched. In the context of self-driving cars, the driver is needed to pay full attention to the road as the cars are semi-autonomous and have been involved in multiple accidents in the past. New policies are needed to be introduced regarding the same for the technological advancement in the field (Morris, 2020).

Apart from that, the ethical concerns of technical safety, malicious use & capacity for evil, unemployment concerns, socio-economic inequality, environmental effects, dependency on AI, and the effects of the human spirit need to be addressed before it is too late. Innovations are developed for the positive and AI gives us a great set of capabilities to support people and enrich the environment. Yet, we must decide to do so following principles to make the planet a safer place. With the help of the collective efforts of many people and organizations, we would hope that AI technologies can coordinate with us to make this world a better place to live in.

5. Conclusion

The possibility of superhuman intellect and superhuman expertise in AIs creates an unprecedented obstacle for the provision of an algorithm culminating in super ethical behavior. New design requirements on transparency and predictability are very much in need because of the AI portending complications due to its rapid movement towards a more humanistic way of thinking. Not only that, new kinds of safety assurance and engineering should be handled with ethical consideration. Although there is a lot of time for all the fears of AI to come true and the extensive research has been going on in this field, this issue should not be taken lightly because AI is our future and it is inevitable. The AI should move towards well-being and safety for a prosperous future following a proper set of guidelines. AI should be taught to deal with moral and ethical dilemmas so that even its factor of unpredictability would not harm the human race and continue to establish a utopian society for years to come from now.

6. References

BBC, 2001. *Robots beat humans in trading battle.* [Online] Available at: http://news.bbc.co.uk/2/hi/business/1481339.stm [Accessed 25 January 2021].

BBC, 2018. *Amazon scrapped 'sexist AI' tool.* [Online] Available at: https://www.bbc.com/news/technology-45809919 [Accessed 1 February 2021].

Cambridge, 2011. *The Unacceptability of Asimov's Three Laws of Robotics as a Basis for Machine Ethics.* [Online] Available at: https://www.cambridge.org/core/books/machine-ethics/unacceptability-of-asimovs-three-laws-of-robotics-as-a-basis-for-machine-ethics/D58C8BAD402DF52AD2785C17A68431EB [Accessed 1 February 2021].

Hagendroff, T., 2019. *The Ethics of AI Ethics: An Evaluation of Guidelines,* s.l.: s.n.

Harvard, 2017. *The History of Artificial Intelligence.* [Online] Available at: http://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/ [Accessed 10 January 2020].

IBM, 1997. *IBM100 - Deep Blue.* [Online] Available at: https://www.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/ [Accessed 10 January 2021].

Jacobson, R., 2014. *8 things you didn't know about Alan Turing.* [Online] Available at: https://www.pbs.org/newshour/science/8-things-didnt-know-alan-turing [Accessed 18 February 2020].

Jobin, A., Ienca, M. & Vayena, E., 2019. *The global landscape of AI ethics guidelines,* s.l.: Natural Machine Intelligence.

Morris, D. Z., 2020. *Who's liable when a self-driving car collides with another vehicle?.* [Online] Available at: https://fortune.com/2020/11/21/self-driving-car-crash-liability-who-is-liable/ [Accessed 10 February 2021].

NASA, 2021. *NASA's New Mars Rover Will Use X-Rays to Hunt Fossils.* [Online] Available at: https://www.nasa.gov/feature/jpl/nasas-new-mars-rover-will-use-x-rays-

to-hunt-fossils
[Accessed 19 February 2021].
O'Carroll, B., 2017. *A guide to narrow, general, and super artificial intelligence.* [Online]
Available at: https://codebots.com/artificial-intelligence/the-3-types-of-ai-is-the-third-even-possible
[Accessed 20 January 2020].
Russell, S. & Norvig, P., 2010. Introduction to AI. In: *Artificial Intelligence: A Modern Approach.* Upper Saddle River: Prentice Hall, pp. 1-5.
ScienceDaily, 2021. *Robots sense human touch using camera and shadows.* [Online]
Available at: https://www.sciencedaily.com/releases/2021/02/210208173054.htm
[Accessed 19 February 2021].
Strelkova, O., 2017. *THREE TYPES OF ARTIFICIAL INTELLIGENCE.* [Online]
Available at: https://conf.ztu.edu.ua/wp-content/uploads/2017/05/142.pdf
[Accessed 2 January 2021].