|  |  |  |
| --- | --- | --- |
| **Academic**  **Year** | **Module** | **Assessment Type** |
| A19 | Concept and Technologies of AI | Report |

AI and Ethics of AI

Student ID: 2049822

Student Name : Sashank Dulal

Section : L5CG12

Module Leader : Mr. Prakash Gautam

Lecturer : Mr. Sachin Kafle

Submitted on : 12-Feb-2021

Contents

[Introduction 1](#_Toc63242954)

[Phases of AI 2](#_Toc63242955)

[Data Preparation 2](#_Toc63242956)

[Training 2](#_Toc63242957)

[Optimization 2](#_Toc63242958)

[Evaluation 2](#_Toc63242959)

[Types of AI 3](#_Toc63242960)

[Ethical Concerns Surrounding AI 4](#_Toc63242961)

[Responsibility 4](#_Toc63242962)

[Privacy 4](#_Toc63242963)

[Transparency 4](#_Toc63242964)

[Unemployment and Distribution of wealth 5](#_Toc63242965)

[Bias and Singularity 5](#_Toc63242966)

[Conclusion 6](#_Toc63242967)

[Bibliography 7](#_Toc63242968)

Table Of Figures

[Figure 1: Types of AI (SOFTWARETESTINGHELP, 2020) 4](file:///Y:\Downloads\AI%20level%205\coursework\report.docx#_Toc63242574)

[Figure 2: Camera doing wrong identification (G/O Media Inc, 2021) 8](file:///Y:\Downloads\AI%20level%205\coursework\report.docx#_Toc63242575)

# Introduction

Understanding reasoning and perception of human beings and trying to replicate it to a form of machinery or intelligent agents has been a continuous effort of scientists all over the globe. The efforts goes way back to 1940s and the term was coined as academic field in 1956 by John McCarthy in Dartmouth conference (www.javatpoint.com, 2018) . In the book “Artificial Intelligence – A Modern Approach”, the authors define AI considering four different aspects based on behavior and rationale, thinking humanly, thinking rationally, acting humanly and acting rationally. (Russell & Norvig, 2010). So generally, AI can be defined as a machine or a system that can replicate human behavior or thinking capabilities making them able to make decisions on their own and the definition complies with the question raised by Alan Turing on thinking machines (www.javatpoint.com, 2021). Since the concept of thinking autonomous machines came into being, huge leap have been achieved such as neural networks and deep learning, reinforcement learning, data-mining to produce more reliable results, speech recognition, natural language processing, self-driving cars and so on (Russell & Norvig, 2010).

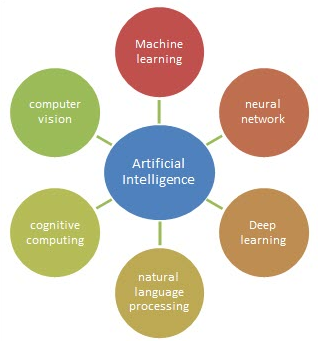
Artificial Intelligence in itself is a vast collection of sub-fields such as machine learning where a machine is fed with a bunch of unlabeled data and the machine deduces relation between data and the output. Deep learning is a field in which the data goes through various computation layers to get to a conclusion and in neural networks data is processed just like the human neurons does by passing information from one neuron to another. Similarly, natural language processing deals with trying to interpret and decode human language. Computer vision is another field which is used for image related tasks such that the machine interprets the image for classification and makes better output based on the images observed and cognitive computing is a field which deals with replicating a human brain by analyzing how human brain responds to text, objects and speech (Great Learning, 2020).

Figure : Types of AI (SOFTWARETESTINGHELP, 2020)

With the development of high speed computing and big data, AI is booming and the advent of any new academia requires huge deliberation on the future prospects, effects and the ethical concerns it poses over the course of time. As human like thinking is the objective of Artificial General Intelligence (AGI), and humans have their own opinion as well as biases. Such example cannot be found in today’s machines so it is completely uncharted territory for us humans. So much of researches, discussions and debates need to be shifted towards the ethical aspects of having a human like thinking machines or superior intelligent systems (Beyond Limits Inc., 2021).

# Phases of AI

## Data Preparation

It is the first and foremost part of any AI implementation except for reinforcement learning. in… where data is collected, cleaned, normalized to scale and finally the most import features that play a vital role in the prediction is selected for training the models. In this step, any defective data is dealt with so that no complications rise during further processes. Similarly, the dataset is split into train, test and validation sets such that the model has sufficient data for training and it gets some new data for testing purposes. (nexoya ltd, 2021)

## Training

Appropriate model according to the problem scenario is selected and the cleaned data is fed into the learning model in order to produce predictions. Selection of model is an important aspect of this process. For instance, if the target of our AI application is to predict whether a person has cancer or not, then we might select classification models and so on (nexoya ltd, 2021).

## Optimization

After our model is trained, the important parameters of our algorithm are tuned so that the prediction gets better and the error gets minimized (nexoya ltd, 2021).

## Evaluation

The final process in any AI application development is the evaluation of the model itself. Different accuracy measurement techniques and error calculation methods are used in order to numerically evaluate the overall performance of the algorithm. The use of test and validation sets is also done to ensure that the algorithm can react to new data and provide similar results like that of the training set (nexoya ltd, 2021).

## Types of AI

Although artificial intelligence can be classified into types being based on many criterions but the types of AI we are discussing are categorized on the basis of the extent to which the AI can replicate human behavior and thought process. Namely, Artificial Narrow Intelligence (ANI), Artificial General Intelligence (AGI) and Artificial Super Intelligence (ASI).

* Artificial Narrow Intelligence (ANI)
  + ANI refers the AI system that is capable of doing a specific task. This includes every AI application till date such as translation engines, self-driving vehicles and so on. Even the most complex AI using neural networks and deep learning lies in the category of narrow intelligence as it has a narrower range of abilities.
* Artificial General Intelligence (AGI)
  + AI agents with AGI are equipped with the perception and capability of learning and understanding the environment as well as their human counterparts. These AI systems would be able to learn on their own, deduce conclusion out of different scenarios and make generalizations in various fields. Agents with AGI are said to have intelligence and multitasking capabilities like that of human beings.
* Artificial Super Intelligence (ASI)
  + This would be the gold standard of artificial intelligence as the systems equipped with ASI would be far more intelligent than the most intelligent beings on the earth. The features that would make it possible are availability of huge amount of memory, fastest data processing and the best decision-making abilities.

# Ethical Concerns Surrounding AI

Intelligence has defined human beings as superior organisms and has affected every nook and corner of our society. Literally everything humans do constitutes intelligence and consciousness of the surrounding. Since the digital revolution, much of the human activities have been affected by the digital phenomena such as recommender systems suggesting new product or digital content, election interference from foreign nation, real time translation, spellcheck and so on (Dubber, et al., 2020). These technologies have significantly challenged the prevalent norms and values which indicates the need for a widespread debate on what challenges it may pose to mankind and the whole ecosystem as well as the mitigation measures for the problems. We must hope for the best being prepared for the worst. The major ethical issues of artificial intelligence is discussed below.

## Responsibility

As AI is still in the developing phase, many errors occur during the testing and implementation. If an AI makes mistake and the consequence is fatal, who will be held responsible for the AI, the founder or the AI itself? It is a major concern that needs to addressed beforehand. In the case of Elaine Herzberg who was hit by an uber while she was riding bicycle, the safety driver of the car, Rafael Vasquez was charged with negligent homicide while uber was not charged. The driver was streaming television show at the time of the accident being fully assured that the system would work and it did not. Similarly, police report labelled same “fatal crash” as “avoidable”. According to the report, the system failed to identify the victim as an immediate collision danger (BBC, 2021). It raises a question that if the users of the system be punished or the ones who commercialize those systems and make profit out of it.

## Privacy

Most of the recommender systems require huge amounts of personal data collected over time in order to correctly recommend anything to the user. The control over the data used is not well defined and the mechanism through which it is processed to get recommendation is also hidden from the user. This along with the many incidents of widespread data collection by the private companies and individuals for their personal needs has endangered the privacy of any individual. If there is a major leak, then the data is all over the web, ready to be misused by anyone according to their personal and professional needs (The Metaphysics Research Lab, 2020).

## Transparency

Machine learning is the basis for many AI systems in use till now and neural networks is the backbone of machine learning. Neural networks is used to find the patterns with the help of series of calculations between layers and the layers are independent of one another. It behaves just like the neurons inside of human brain by passing information from one neuron to the other. This makes the algorithm opaque, meaning that any of the computations done inside the algorithm is completely unknown to the user and even the programmer. The result of this could be that the input data is generalized in a different sense, giving completely irrelevant output/result. And even if the result is right, the basis through which the input data is processed to correct result is unknown. This black-box like behavior of the system is often difficult to debug and trace back to errors as well as maintain a standard.

## Unemployment and Distribution of wealth

The ultimate goal of artificial intelligence is to automate all the labor intensive tasks which ultimately results in many people losing jobs. According to “technology.org”, about one quarter of the jobs in the US would be affected by AI and automation. For an instance, trucking industry employs millions of individuals in the US but what will happen after the self-driving trucks replace those drivers? And another question is how the wealth generated by using AI systems be distributed and how will the unemployed persons be taken care of? Use of AI in such industries could help create heavy concentration of wealth for their owners while making the people who lost their jobs to AI economically weak (World Economic Forum, 2021).

## Bias and Singularity

Despite of the fast processing of the data to give meaningful result, AI can make biased output since the humans who made it are mostly biased and have opinion of their own. An instance of this is the camera “Nikon S630” identifying the user who was smiling at the time as blinking (G/O Media Inc, 2021).Similarly, another incident of such bias occurred in 2015 when google photos mistakenly labelled black people as gorillas (New York Times Company, 2017). On the other hand, recent firing of Timnit Gebru, a renowned AI ethics researcher at google because of the paper she co-authored that found out the lesser accuracy of facial recognition algorithms at identifying people of color. This has raised some eyebrows over the actions taken by google, one of the biggest firms in AI research (MIT Technology Review, 2021). Instead of trying to resolve these types of bias, such pioneers found to be suppressing the findings, in no time the whole AI community could end up being careless of the reality and move towards achieving their personal or professional goals.

Figure : Camera doing wrong identification (G/O Media Inc, 2021)

Another main concerns of major scientists and innovators like Elon Musk, Late Stephen Hawking, Stuart Russell is the AI takeover. They along with many other scientists have signed an open letter stating the need for regulations and controlling measures in order to make AI that does only what humans tell it to do. The main objective of the letter is to focus on researching about how to get lots of benefits from AI and avoiding the potential drawbacks (Future of Life org, 2021).

## Conclusion

With new technology comes new fields which can help mankind. AI has a wide range of applicability such as healthcare, industry, security and so on. There are many unanswered questions on how AI will be in future. Some predict it as taking over humanity and some say it will be a boon for humankind. The ethical aspects of this technology needs to be deliberated and discussed upon so that a standard framework is defined for all AI implementations. A governing body should be looking after and enforcing the use of AI for the goodwill and social as well as economic equality should be maintained in case of unemployment caused and in case of recognition algorithms.

# Bibliography

BBC, 2021. *BBC.* [Online]   
Available at: https://www.bbc.com/news/technology-54175359  
[Accessed 21 01 2021].

Beyond Limits Inc., 2021. *Beyond Limits.* [Online]   
Available at: https://www.beyond.ai/news/ai-our-world-the-importance-of-ethical-ai/  
[Accessed 20 01 2021].

Dubber, M. D., Pasquale, F. & Das, S., 2020. The oxford Handbook of Ethics of AI. In: *Ethics Of AI.* New York: Oxford University Press, pp. 1-12.

Encyclopædia Britannica, Inc., 2021. *Artificial Intelligence-Reasoning | Britannica.* [Online]   
Available at: https://www.britannica.com/technology/artificial-intelligence/Is-strong-AI-possible  
[Accessed 16 01 2020].

Future of Life org, 2021. *Future of Life.* [Online]   
Available at: https://futureoflife.org/ai-open-letter/  
[Accessed 27 01 2021].

G/O Media Inc, 2021. *GIZMODO.* [Online]   
Available at: https://gizmodo.com/camera-misses-the-mark-on-racial-sensitivity-5256650  
[Accessed 25 01 2021].

Great Learning, 2020. *What is Artificial Intelligence? How does AI work, Types and Future of it?.* [Online]   
Available at: https://www.mygreatlearning.com/blog/what-is-artificial-intelligence/  
[Accessed 17 01 2021].

MIT Technology Review, 2021. *MIT Technology Review.* [Online]   
Available at: https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/  
[Accessed 26 01 2021].

New York Times Company, 2017. *The New York Times.* [Online]   
Available at: https://bits.blogs.nytimes.com/2015/07/01/google-photos-mistakenly-labels-black-people-gorillas/  
[Accessed 25 01 2021].

nexoya ltd, 2021. *The process of building an AI model.* [Online]   
Available at: https://www.nexoya.com/blog/the-process-of-building-an-ai-model/  
[Accessed 18 01 2021].

Russell, S. J. & Norvig, P., 2010. *Artificial Intelligence: A Modern Approach.* 3rd ed. New Jersey: Pearson Education Inc..

SOFTWARETESTINGHELP, 2020. *What is AI:Definition and Sub-fields of AI.* [Online]   
Available at: https://www.softwaretestinghelp.com/what-is-artificial-intelligence/  
[Accessed 17 01 2021].

The Metaphysics Research Lab, 2020. *Stanford Encyclopedia of Philosophy.* [Online]   
Available at: https://plato.stanford.edu/entries/ethics-ai/#MainDeba  
[Accessed 25 01 2021].

World Economic Forum, 2021. *Top 9 ethical issues in artificial intelligence.* [Online]   
Available at: https://www.weforum.org/agenda/2016/10/top-10-ethical-issues-in-artificial-intelligence/  
[Accessed 25 01 2021].

www.javatpoint.com, 2018. *Applications of AI-javatpoint.* [Online]   
Available at: https://www.javatpoint.com/application-of-ai  
[Accessed 16 01 2021].

www.javatpoint.com, 2018. *History of Artificial Intelligence-javatpoint.* [Online]   
Available at: https://www.javatpoint.com/history-of-artificial-intelligence  
[Accessed 16 01 2021].

www.javatpoint.com, 2021. *Javatpoint.* [Online]   
Available at: https://www.javatpoint.com/artificial-intelligence-tutorial  
[Accessed 16 01 2021].