Unleashing The Power of AI Under The Ethical Sphere

Ritvik Voleti[1], Ch V Raju[2], Keshav Kumar Jha[3], Bappaditya Jana[4]
[1, 3, 4]KCC Institute of Technology and Management, Gr Noida
[2]Sharda University, Gr Noida

Abstract- *AI has become an irreplaceable entity of our lives with each and every decision-making being dependent on AI machines. The engine of these machines is the embedded algorithms which provide them quick and efficient decision-making ability. This usually results in keeping a blind eye over factors like understandability, meaningfulness and only worrying about effectiveness and usability. The future possibilities of AI (Super AI) based machines that will be better and smarter than human intelligence have the potential of catastrophic issues for humanity. This article focuses on key issues in Super AI machines and puts an emphasis that these machines must be built on an ethical basis. The paper moreover analyses the causes, impactsand problems surrounding the futuristic AI technologies which would need immediately resolving. Lastly, it aims on providing different realistic solutions in enabling a more-safer AI-based future thereby making ethical secure prospects for humanity that would safeguard us from a dystopian society.*

Keywords- Artificial Intelligence, Ethics, Machine Learning, Robotics

I. INTRODUCTION

Ethical AI is on the lips of all modern researchers and philosophers due to the controversies and entails of it. The world of AI started with its cameos and features in science fiction movies, and all of a sudden within a blink, it is now the most influential part of our lives. It is inbuilt within our smartphones, assisting the users with its smart music preferences and many uses but the most unique application is its ubiquitous nature. The impact of AI is influencing each and every industry be it IT, healthcare, science, education, and many more. Most importantly its influence on human life is unheard of. People are no longer just merely drawn into the power of AI in enhancing the efficiency, cost reduction, and R&D accelerations of programs but fear the societal threat that this complex entity possesses if it gets the freedom to take its decision-making autonomously.

Ethics are a necessity, as it is composed of rules, guidelines, moral obligations, and principles that provide a barrier between good and bad in the creation and designing of intelligent robots. The concept of Ethical AI has gained more relevance due to the threats of poor designing, misuse, and dangerous results of Artificial Intelligence systems. It is based on interconnection and interdependence between humanity and technology, the ambitious technological developments and the reality check on the human side is what ethical AI is all about.

In 1949, Gabriel Marcel a French Philosopher was wary of the harm that technology might cause in blindly focusing on only solving our life's tasks and not thinking about its consequences. According to Karen Mills if society is not conscientious and thoughtful then the end is inevitable. It is difficult to imagine a future of AI supporting humanity, which potentially would support society, and the most dangerous part of it all is its uncertainty surrounding the dangers which it can cause for the world when used improperly [1]. According to Professor Dan Weld, Artificial Intelligence is seeing unparalleled growth and its applications are influencing all parts of our lives. Although this also raises many challenges along with the successes which must be overcome [2]. Humans have a huge role in the futuristic technologies impact and their sustainability in society. This paper mainly aims in observing the influence of Artificial Intelligence on human life and the environment, its problems, and its answers. But before all this, the paper observes the causes behind it.

II. ETHICAL ISSUES RELATED TO UNLEASHING AI

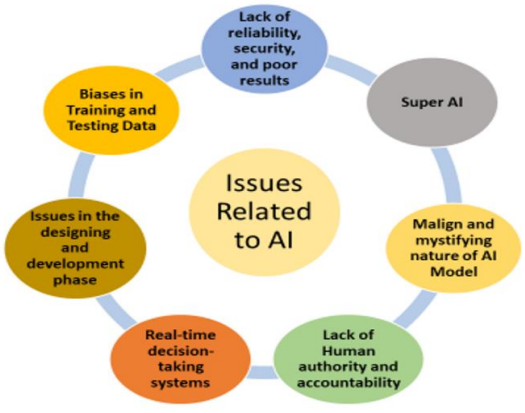

Figure 1. Ethical Issues related to Unleashing AI

A. Lack of reliability, security, and poor results

The increasing mismanagement of data, negligible production procedures in designing, and uncertain deployment activities have led to rapid unreliability, lack of safety, and poor results. Furthermore, the impact can possibly harm human welfare and ultimately cause damage to society. This results in a lack of public faith that these Artificial Intelligence technologies would be used keeping in the health and safety of mankind.

B. Super AI

An enormous future threat "which could possibly have huge consequences on the beginning of the end for the life on Earth and could also lead to its natural growth to be stopped permanently." But at the same time, the counterargument of a possibility of super AI systems safeguarding Earth's life-supporting system and its future potentials is also prevalent. It is imperative to realize that these systems have their fair share of risks and prospective rewards. The direct results will be that AI will not just be smarter than human beings due to its cognitive knowledge but at the same time will continuously evolve with its self-awareness and conscious behaviour [3].

C. Malign and mystifying nature of AI Model

Most Artificially Intelligent systems are dependent on ML algorithms in deep networks whose role is to analyse the dataset's patterns and nature either labelled or unlabelled data i.e unsupervised, supervised, and semi-supervised learning. The model focuses on finding key data patterns, and the data mapping is performed in a manner that is ideal for decision making, but while this is happening at the back end the AI programmer finds it a tedious and a huge obstacle in gaining insights into what patterns have been chosen by the model. This limited understanding is exploited as a vulnerability by hackers worldwide, through their powerful malware, spyware, and zombie machines in harming computer devices for meeting their ultimate goal of either attacking systems or sometimes for casual fun [4]. This is what the black box problem is all about [5]. The high dimensionality mapping in AI systems overwhelms the human-understandable reasoning capacity.

D. Real-time decision-taking systems

The main merit of Artificial Intelligence is that it autonomously takes accurate decisions, thereby reducing human stress and errors. But, these real-time automated systems bring with them a huge flaw in security for society, as it blindly follows the makers of the machine. If the makers safeguard society then it is nothing to worry about but on the other hand, if these machines are encouraged by evil minds bad things are bound for humanity. One of the main factors is its autonomous nature; the results of which are human beings are not acceptable to give up all their powers to these systems as they have the potential to override us [6]. It is because of the dangers looming on societies and individuals, that we are only encouraged in building ethical machines which support human dignity, and also ensure that relations between the humans and systems remain the same [7].

E. Issues in the designing and development phase

AI models are constructed on the phases of data structuring, and data processing (big data). Large volumes of data are firstly collected, then processed, and lastly used in making AI models. Data these days is extracted and analysed without any provision of the data owner's consent. It is thereby comprising the confidentiality, integrity, and authentication of user data. We must have morality to not build such systems which pose a fear to humanity and not only worry about profits and performance [8].

F. Biases in Training and Testing Data

The modern systems are still dependent on labelled data which motivates training, modifying the dataset in order to enhance its AI performance [9]. The model is based on assisting in the process of data mining and providing the freedom to designers in selecting the features of the system like its metrics & structures, etc. So, the smart systems are thereby affected by replication for the biases and their preconceptions made by their creators. Our human errors will only increase if we ignore to keep an eye on the upcoming AI systems and will be dangerous for future generations [10]. Lack of identity and clear unbiased ethical protocols has to lead to programmers finding it hard in keeping a machine completely ethical be it the training phase or the testing phase.

G. Lack of Human authority and accountability

The AI activities ranging from classifications, decisions, and predictions have a huge impact on human beings. Slowly and steadily the AI systems are gaining more knowledge base and are improving autonomously, therefore a big giant leap of intelligence is expected [11]. There are numerous cases where humans find it hard to explain and contemplate their roles in taking accountability for the actions performed by the intelligent machine with regards to the final outcome. Human responses is mostly of confusion and anxiety as they fail to give any valid justification for the failings of the machine and blindly blame the machine for the results, and to rub salt into wounds feel that they have left no stones unturned

in improving the situation for the better. While programmers and users utilize these AI systems, there are several factors that lead to the shortcomings of the agent including its programmed source code, input data, poor functionality, etc.

III. AI's IMPACT IN DIFFERENT AREAS

Apart from the safety and security aspect of AI its impact is felt in many parts of the industry and excluding financial aspects, humanity is closely linked to AI. In the coming years this impact is expected to increase even further. It has its fair share of usefulness and controversies which needs to be addressed.

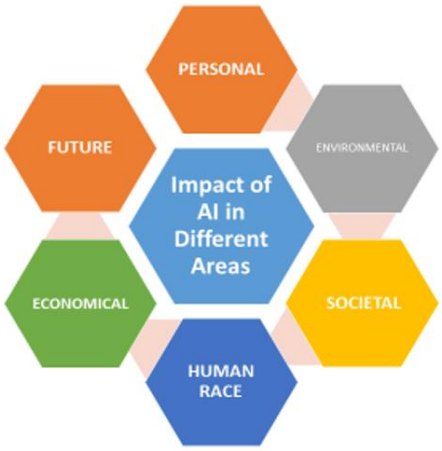

Figure 2. Impact of AI in Different Areas

IV. ECONOMICAL IMPACT OF AI

A. Financial Disparity between the have's and the have not's

McKinsey has predicate that, automation will fully replace about 400-800 million present jobs by the year 2030, which would mean that at least 375 million employed people must change and adapt according to the demands of the market for survival. There has been a giant transformation in the labour market due to Artificial Intelligence. Artificial Intelligence and other latest technologies will potentially lead to an unemployment drought due to the possibilities of smart machines replacing humans and ultimately taking over their jobs [12]. It is because of the rapid technological advancements caused by Artificial Intelligence. It is estimated that this is at such a high scale to worry, that an additional1 robot can compete performance-wise with the likes of a thousand laborers which ultimately will be a profitable option for the enterprise but will have a massive employment reduction in the job market [13].

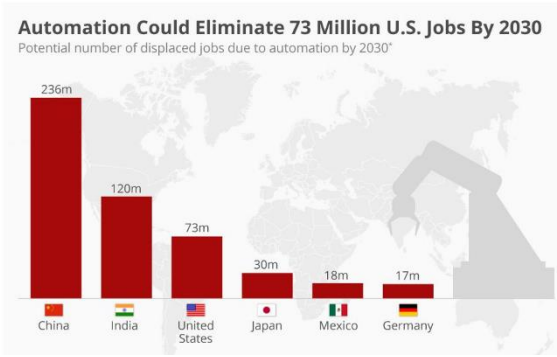

Figure 3. Jobs lost to Automation
Source :McKinsey

These developments are largely isolated to favour only businessmen CEOs of big companies as it only provides them a fair share of luck, rewards, and other benefits. It on a global scale would largely affect the developing countries of the world.

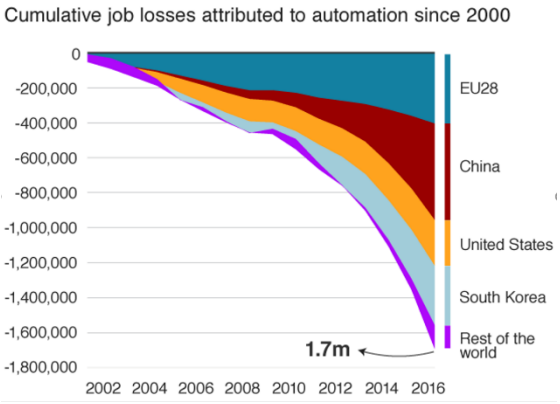

Figure 4. Impact of Job loss in Different Countries
Source : Oxford Economics

V. SURVIVAL OF HUMAN RACE

A. Conflict between man and a machine

According to many experts, AI could spark a technological war and whoever adapts according to it will win or survive whilst others will perish. For instance, in the future, there is growing possibilities of autonomous defence system where there could be little or no time left for humans to evade future attacks which may lead to widespread casualty and certainly- a disastrous situation. Humans will have no warning to prevent these attacks and will be dependent on machines to safeguard humanity which looks unrealistic [14]. Technologies like AI agents like robots, self-driving cars, drones, underwater vehicles are the mediums that will spark this IT war. On the other hand, there are positive features of robots being more quick, effective, powerful along with precisely obeying the orders given to them makes it is an

excellent choice for security but this almost it. The inference attacks maliciously gain unauthorized access of a valid user which could activate numerous evasion attacks on their systems. Data poisoning attacks harm the targeted AI models mainly during the training period for sufficing their malevolent intentions.

VI. SOCIETAL

A. The disparity in Socioeconomic factors

According to MIT's professor Erik Brynjolfsson, his observation from statistics and data analysis shows that technology is the single most negatively impactful driving force towards the increasing social inequality [15]. The AI-based developments are leading to more and more disparity within the society, especially within the financial and socially backward communities. Further continual declination of the society will sadly result in more volatility, bringing the risk factor at a high and many more dangerous consequences. The main cases are clearly visible in biased machines that have established built-in biases towards age, gender, ethnicity, and other impairments [16]. This issue is and will be a rash that needs immediate treatment as soon as possible to avoid further chaos and discrimination.

VII. ENVIRONMENTAL

A. Wastage of Energy

ML models release large quantities of energy during the training phase, and not just nature this also affects businesses financially in able to run these machines which are of a value of as many as millions. Since building and using these devices needs fossil fuels, it is environmentally non-viable and could become a major reason for climate change. Overall if AI is used wisely it can a boon for society [17], otherwise, it has some threatening future linked with it for mankind.

VIII. PERSONAL

A. Large scale surveillance and harmful technological threats

The direct applications of AI have led to a host of privacy issues experienced by both users and large-scale enterprises. Most countries are keeping constant surveillance on their general public with regards to the Internet world as it is affecting all aspects of our life. Some have even passed some harsh laws on data privacy to protect the rights of their citizens. At least 75 countries of the total are utilizing smart systems powered by AI for surveillance needs against internal and external attacks.

B. Moral Deskilling and its weaknesses

There are programmers with limited knowledge and experience blindly observing the instructions performed by the systems and thereby it undermines and showcases the weaknesses of the programmers. This is what moral deskilling is all about. The more we provide autonomous decision-making abilities to AI the more chances are of human-driven innovations and skills to be limited until we change this phenomenon and try to make systems that value human endeavours. Our uniqueness and innovative ideas will dwindle and become a rarity if computers are allowed to take initiatives on their own and alongside this, humanity will always be at continuous risk. Deskilling is a threat for mankind and will make human life less interactive, understandable, dangerous, limited, and very complicated [18]. Henceforth, it must be avoided at all means to stop this vulnerability to expand.

C. Limited Purpose

Humans increasing dependent nature for doing almost any human activity through AI-based systems has become a huge question mark over whether they are competent enough to do any task on their own. This is not just a financial concern that is based on the vanishing of jobs in the market but also human life is becoming increasingly void, dull, and miserable life. It is a mental health threat for unemployed people as they will start searching for new job opportunities and focus on gaining more impetus into the realization of what their life is all about. Also realizing how one can enhance the society to make it better for the coming generations is something to be considered as well. It is a challenge to prevent and protect society as rising dissatisfaction, frauds, mismanagements, robbing, bitterness has made the world a feared place to live peacefully. So, this has left humans puzzled whether or not they prefer to give AI-complete support and be dependent on it or hinder its growth and continue where we are now [19]. According to Danahar finding this balance is key for survival otherwise it would not be long when we will be struggling for survival and will lose our identity.

D. Isolated and lonely from the world

In the last 5-8 years there has been a reduced interaction observed between humans and also with nature, and their problem-solving ability will also take a big hit all because of rising automation. An algorithmically trained

personalized model is seeing giant strides in satisfying the demands placed by the customer, however, this also makes us narrow-minded and leads to selfish short-sighted nature as we fail to recognize the feelings of the world, and in a way, humans will form socially isolated relationships. Society is slowly emerging as a victim of the human being's lonely behaviour. It is depressing and soul-crushing as loneliness breaks a person from within [20]. Therefore, there has to be a collective emphasis placed by all, in ensuring that while AI is transforming the world, there has to be a tightly knitted affection, trust, understanding among all human beings.

E. Compulsive gadgets usage

Humans are more and more becoming addicted to machines largely due to the 21st century AI revolution. This addiction affects our personality and ultimately one becomes highly dependent. There is rising human-machine communication at such a continuously high level that it is resulting in more problems; this is when one fails to realize that he/she is technologically addicted [21]. AI-driven systems are trained in a manner to understand and interpret human feelings at a very deep level such that they are able to exploit our vulnerabilities like making one indulge in gambling, crime, greed, etc. Addiction is not just weakening and destroying our in-built personality but most importantly makes us naïve and blind to ignore our support system i.e our education, money, and community.

IX. POTENTIAL OUTCOMES

A. Need to safeguard the rights of futuristic robots

Some analysts believe that smart AIs becoming self-conscious if created, would need to be recognized and given the value they deserve similar to humans. Philosophically and morally predicting, people more or less anticipate robots depicting identical behaviour like humans, so in the future, there could be a possibility for good robots being rewarded for their actions and goodwill but this is still all in the air. If in the upcoming 10-25 years we see upcoming robots by virtue of their actions depicting morality, then humans will be inclined to give them their moral status. Even if they get the moral status or not, what is important is all humans must take care of robots and never abuse them.

B. Super Intelligent machines and AGI

AI community worldwide is keeping an eye on the prospects and threats linked with Artificially Intelligent Machines for mankind if encouraged in the coming years [22]. When an AI system will fulfil its ultimate goal and be as intelligent as human intelligence then it will be referred to as AGI. Historically, whenever new species had their encounter with humans, they were either merged genetically or in worst-case scenarios become extinct. Lastly, as AI has become a vital tool for survival one must carefully ensure to keep a balance between humans and machines so that they remain ethical.

C. Reliance on AI systems

Human beings are heavily reliant on innovation. From the stone age to the present there is one constant, which is the dependence of humans on technology. It is in a sense a key part or closely linked to our character. As time is passing by, we are seeing more dependency on AI systems and this will keep on increasing. There is no coming back once we jump into this technological river of future advancements which shows how much we rely on technology. The main question is whether humans can adapt and survive in this AI-dependent environment [23].

X. SUGGESTIONS AND CONSIDERATIONS

To be successful in this competitive technological warfare one must establish new regulations in order to not just survive but to thrive in it. These rules must be created and adhered to from the start till the maintenance phase of the model. On an organizational level when multiple companies are working together, they must agree to what they feel is ethical to perform and not to do so, to avoid any future complications. This will ensure a sense of confidentiality, integrity, and authentication within the companies. Every modern enterprise values its data to the fullest and invests highly in increasing that and protecting it. Some key suggestions need to be adopted for maintaining data quality, and integrity in an organization.

- One must realize quickly as to what type of data be it structured, unstructured, or semi-structured data is ideal and apt for their organization. The faster one realizes this, the chances of development rely largely on this factor.
- There has to be a plan to decide where the data is best to be stored (warehouses) and which techniques are ideal for using this data appropriately.
- Data must be regularly cleaned for ensuring the data remains error-free and its significance level also remains intact. Not keeping data clean results in more confusion and incorrect decisions being taken by companies which leads to massive losses.
- Companies must be motivated in keeping data decisions to be taken by teams built specifically to

handle this data. The more-sharper and experienced the team, the better results the organization expects and receives.

- After the successful establishment of companies there is a sense of complacency seeking into the management as they neglect to monitor part with time and it affects them in the long run. Proper monitoring must be made a must.

Some key responsibilities must be followed by all the members of the organization from the top to the bottom.

- There must be well thought and agreed-upon governance framework ensuring the protection of human rights while AI systems are running in full flow.
- Proper legally accountable policies provide necessary protection against future AI threats, which could weaken the company and its workers to drastic levels, and in worse circumstances, the right mitigation steps must be followed in order to withstand and remove those vulnerabilities.
- Management must be able to implement long-term regulations for maintaining a transparent and secure environment where each byte of data must be thoroughly checked and shared at various organizational levels once it is acceptable for applications.
- Every enterprise must focus on making its members realize the expectations and the work values to be adhered to. These established values ensure more coordinated and responsible practices performed by the employees and the decision-makers.

Current IT giants like IBM, Google, Amazon, and Microsoft, are driven towards ensuring AI developments be practiced on an ethical basis. Singapore's Monetary Organisation in the year 2018 had collaborated with Amazon's AWS and Microsoft and built the FEAT principles which ensured that there must be a fair, ethical, accountable, and transparent environment for the key decision-makers as well as the users by encouraging ethical practices during key developmental steps.

XI. CONCLUSION

It is imperative to understand human ethics deeply, once one has a good knowledge of real-world ethics, then only one can develop AI ethical laws for society. Before developing any AI system, there has to be predefined-ethical laws for a system and should not be considered as an afterthought. It is the responsibility of the programmers for establishing these regulations for the AI systems. Most importantly they must be embedded in the autonomous systems. The problem arises when different organizations have their own ethical restrictions and which make it hard for the machines to be having high performance and being righteous at the same time. Therefore, it is highly encouraged that there has to be a worldwide consensus on ethical issues as to the limits to which technological developments could be allowed. New technologies are coming and what is a novelty now may be old in 2-3 years so, in the same way, ethics have to be regularly examined and monitored to keep them up to date. There is a growing possibility of expecting super AI's in the future that will be either on par or even more intelligent than humans but this raises the challenge of protecting our ethical and general rights to ensure security and integrity. If humans are not ready or not up to the challenge then there is a dangerous pathway lying ahead and there must be a collective effort to prevent such catastrophe from happening.

REFERENCES

[1] See : https://openai.com/blog/introducing-openai/
[2] Dan Weld, Beneficial AI 2017, https://futureoflife.org/2017/01/29/dan-weld-interview/
[3] Torrance S. Super-intelligence and (super-)consciousness. Int J Mach Conscious. 2012;4:483–501. doi: 10.1142/S1793843012400288.
[4] P. A. Taylor, Hackers: Crime in the digital sublime. ISBN 9780415180726 Psychology Press, 1999
[5] Wachter, S., Mittelstadt, B. and Russell, C. (2018). Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR. Harvard Journal of Law & Technology, 31(2), 841–87
[6] Floridi L, Cowls J (2019) A unified framework of five principles for AI in society. Harvard Data Science Review. https://hdsr.mitpress.mit.edu/pub/l0jsh9d1
[7] Coeckelbergh, M. (2020). Artificial intelligence, responsibility attribution, and a relational justification of explainability. Science and Engineering Ethics, 26(4), 2051–2068.
[8] Bryson, J. (2010). Robots Should Be Slaves. In Y. Wilks (Ed.), Close Engagements with Artificial Companions, 63–74. Amsterdam: John Benjamins.
[9] Ritvik Voleti, 2020, Data Wrangling- A Goliath of Data Industry, International Journal of Engineering Research & Technology (IJERT) Volume 09, Issue 08 (August 2020),
[10] The Partnership on AI to Benefit People and Society, Inaugural Meeting, Berlin, Germany, October 23-24, 2017.

[11] Kurzweil R. The singularity is near. London: Gerald Duckworth & Co; 2006. [Google Scholar]
[12] Danaher, J. (2019a). Automation and Utopia. Cambridge, MA: Harvard University Press.
[13] Acemoglu, D. and Restrepo, P. (2017b). Robots and Jobs: Evidence from US Labor Markets. NBER Working Paper no. 23285. National Bureau for Economic Research.
[14] Horowitz, M., & Scharre, P. (2015). An introduction to autonomy in weapon systems. CNAS Working Papper. https://www.cnas.org/publications/reports/an-introduction-to-autonomy-in-weapon-systems.
[15] David Rotman, MIT tech Review, Technology and inequality, https://www.technologyreview.com/2014/10/21/170679/technology-and-inequality/
[16] Kraemer, F., Van Overveld, K., and Peterson, M. (2011). Is There an Ethics of Algorithms? Ethics and Information Technology, 13, 251–60.
[17] Leila Scola, "AI and the Ethics of Energy Efficiency," Markkula Center for Applied Ethics website, May 26, 2020, available at: https://www.scu.edu/environmental-ethics/resources/ai-and-the-ethics-of-energy-efficiency/
[18] Vallor, S. (2015). Moral Deskilling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character. Philosophy & Technology, 28(1), 107–24.
[19] Danaher, J. Will Life Be Worth Living in a World Without Work? Technological Unemployment and the Meaning of Life. Sci Eng Ethics 23, 41–64 (2017). https://doi.org/10.1007/s11948-016-9770-5
[20] Julianne Holt-Lunstad, Timothy B. Smith, Mark Baker,Tyler Harris, and David Stephenson, "Loneliness and Social Isolation as Risk Factors for Mortality: A Meta-Analytic Review," Perspectives on Psychological Science 10(2) (2015): 227–237, available at: https://journals.sagepub.com/doi/full/10.1177/1745691614568352
[21] Griffiths, M.D. Technological addictions. Clin. Psychol. Forum 1995, 76, 14–19.
[22] Müller, V. C., & Bostrom, N. (2016). Future progress in artificial intelligence: A survey of expert opinion. In Vincent C. Müller (Ed.), Fundamental issues of artificial intelligence (pp. 555–572). Cham: Springer International Publishing.
[23] Ritvik Voleti, Unfolding the Evolution of Machine Learning and its Expediency, International Journal of ComputerScience and Mobile Computing, Vol.10 Issue.1,January- 2021, pg. 1-7, DOI: 10.47760/ijcsmc.2021.v10i01.001