iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 9 **Issue:** XI **Month of publication:** November 2021

**DOI:** https://doi.org/10.22214/ijraset.2021.38992

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429*
*Volume 9 Issue XI Nov 2021- Available at www.ijraset.com*

# Governance and Ethics of AI

Atharv Jangam[1], Nikhil Suryawnshi[2], Asst. Prof. Megha Adhikrao Patil[3]
[1, 2, 3]*Bharati Vidyapeeth College of Engineering, Lavale*

***Abstract:*** ***In this article we discuss ways AI can be fruitful and inimical at the same time and consider hurdles in implementing ethics and governance of AI. We conclude with presenting solutions to overcome this issue. Artificial intelligence (AI) is a technology that allows a computer system to mimic the human mind. AI, like humans, is capable of learning and developing itself through doing tasks such as planning, organizing, and executing numerous activities.***
***However, as we develop and expand our understanding of AI, there are a few advantages and downsides that should be addressed.***
***Privacy and security are vital, but they conflict with the advancement of AI technology since computers and AI require a large quantity of data to Comprehend and anticipate outcomes. With the advancement of technology, we should be able to maximize security and eliminate the current drawbacks.***
***Keywords: Artificial Intelligence (AI), Autonomous, Machine Learning (ML), Governance, Ethics, Deepfake***

## I. INTRODUCTION

In today's world, the applications of Artificial Intelligence are becoming widely known. Most of us are unaware that technology is being utilized in our daily lives. AI is a technology that uses a computer system to imitate the human mind. AI is capable of learning and improving itself through various tasks such as planning, organizing, and performing various activities just like humans. Some of the industries that use AI are finance, healthcare, gaming, and
Autonomous vehicles. Researchers are still finding out the novel capabilities of AI every day.
AI is capable of handling various tasks which are physically impossible for humans. AI can perform various tasks without requiring much effort and time. However, it can also handle various tasks with impeccable efficiency.

Some cons of AI are:

1) Lack of thinking out of the box,
2) High cost of implementation,
3) Making humans lazy.
4) Lack ethics and morals.

Machine learning is highly data-dependent and exhibits unpredictable results, the model can demonstrate diverse results to similar inputs which makes it difficult to conclude the right output and confirm safety in advance. This kind of behavior is uncontrollable by humans, therefore, developing many challenges as the liability of damage caused by the model is not well specified. Under strict product liability, AI decisions are not expected to be predicted or considered by manufacturers and software designers. This uncertainty could imply that many of the decisions made by AI are outside the control of those involved. It is also widely Acknowledged that excessive liability risks are constraining the development and use of new technologies. This issue highlights the need for governments to develop effective liability frameworks that can promote innovation while protecting society from these risks.
Out of these, the most crucial issue is the lack of governance of Artificial intelligence which might become detrimental in future. As the absence of ethics will welcome invasion of privacy and algorithm discrimination etc. And end up superseding the benefits. As a result, will seed a lack of trust among mankind.
ML algorithms learn through the data provided to them and make decisions accordingly, this could lead to an incompetent model due to inequality and discriminatory patterns present in the data. If characteristics such as race and gender are erroneously related will yield disparate outcomes than expected, would disbenefit certain groups of people.
Proper governance and regulations will not only help detect loopholes but also create a positive environment.

©IJRASET: All Rights are Reserved

## II. WILL ARTIFICIAL INTELLIGENCE ENDANGER OUR PRIVACY AND SECURITY?

The insatiable demand for training data in AI models, on the other hand, puts pressure on developers to acquire and store data they don't necessarily need. Should corporations or ungoverned employees inside them fail to exercise control, this circumstance poses a threat to anonymity and privacy.

Due to the increasing popularity of smart products such as virtual assistants (Amazon's Alexa, Apple's Siri, Google Assistant) and smartwatches, the demand for these products has increased significantly.

Through machine learning technology, these products can collect and interpret the data that they consume. No regulation prohibits the use of these electronic devices for nefsarious purposes. These electronics use technology such as face recognition, voice recognition, and fingerprint scanner even though these technologies make our lives convenient but can also lead to disastrous consequences.

Although AI-based technology is beneficial in terms of cybersecurity, it is also a weapon that can be abused by hackers. They can easily steal our sensitive information and misuse it. The most common repercussion of using this technology is our sensitive data being leaked online. This heinous act of hackers, leads to depression, mental illness in the worst cases even the suicide of the victim.

## III. AUTONOMOUS CARS

Self-driving cars have a higher risk of accidents than human-driven automobiles, despite Claims suggesting the contrary, but the casualties are less severe. Self-driving auto accidents Occur at a rate of 9.1 per million miles travelled, compared to 4.1 per million miles for ordinary vehicles.

The recent tragic death of two persons in a Tesla self-driving car incident in the United States has revived debate over the capability and safety of today's "self-driving" technologies.

For fear of limiting innovation, regulators in the United States and abroad have maintained a hands-off approach so far. Existing safety regulations may be reduced, according to the US

National Highway Traffic Safety Administration. Pedestrians and human drivers, on the other hand, have not agreed to be test subjects. Another serious accident in 2019 may cause regulators to change their minds.

The AI is designed to perform something good, but it devises a damaging strategy to accomplish its goal: This can happen if we don't fully connect the AI's goals with ours, which is a difficult task. If you order an obedient intelligent car to drive you to your destination as quickly as possible, it may get you there chased by cops and covered in vomit, doing exactly what you asked for. If a super-intelligent machine is tasked with a large-scale geoengineering project, it may cause chaos in our ecosystem as a side consequence, and hence human attempts to stop it as a danger that must be defeated.

This issue was brought to light after a self-driving automobile exceeded the speed limit on a highway, and it was unclear to whom the ticket should be sent.

## IV. ALGORITHM FOR PEACE

When Google employees realized that their business was selling technology to the US Air Force for classifying drone photos last year, an AI peace movement arose. Workers were concerned that this was a crucial step toward supplying technology for automating lethal drone strikes. As a result, the firm shelved Project Maven in favor of establishing an AI code of ethics.

A campaign to outlaw the deployment of autonomous weapons has gained support from academics and industry leaders. However, the military's use of AI is only growing, and corporations like Tesla and Amazon have shown no reluctance to assist.

(October 20, 2020, Washington, DC) — In a report released, the International Human Rights Clinic argues that a treaty banning fully autonomous weapons, also known as "killer robots," is both necessary and feasible. "The only viable method to avoid the delegation of Life-and-death decisions to robots is to create a new international treaty."

AI is programmed to do something amazing: Whereas nuclear weapons are a specific sort of weapon, AI is a technology that may be applied to a wide range of weapons and support systems. A nuclear missile, for example, may be fitted with an AI system that would allow it to locate and destroy a specified target. If fallen in the hands of the wrong person, these weapons could easily cause mass casualties. To avoid being foiled by the enemy, these weapons would be engineered to be incredibly difficult to simply "switch off," allowing humans to lose control in a case like this. This risk exists even with narrow AI, but it becomes more prevalent as AI intelligence and autonomy rise.

This problem arises because AI-enabled machines make judgments on their own, making it difficult to discern whether a bad decision is due to programme errors or the AI-enabled machines' unsupervised deliberations.

©IJRASET: All Rights are Reserved

## V. A PORTMANTEAU OF DEEP LEARNING AND FAKE:

Last year, a slew of "deepfake" movies demonstrated how simple it is to create fake clips using AI. This might imply fake celebrity pornography, bizarre movie mashups, and potentially vicious political defamation operations.

Artificial neural networks (ANNs) are computer systems that recognize patterns in data and are used in deepfakes. Creating a deepfake photo or video usually entails sending hundreds or thousands of photos into an artificial neural network, which is then "trained" to recognize and rebuild patterns—usually faces.

Nvidia recently demonstrated how ANNs can produce lifelike faces of any ethnicity, gender, or age. Deepfakes are currently primarily used for pornography. According to research published in June 2020, 96 % of all deepfakes on the internet are for pornographic purposes, with over 100 % of those cases involving women.

Many people, on the other hand, believe that online disinformation is a strong antecedent to the development of deepfakes. The Conservative Party doctored a video of now-Labour leader Kier Starmer in 2019 to make it appear like he did worse in an interview than he did.

People will most likely be misled by deepfakes this year as they improve. DARPA will put new deepfake detection algorithms to the test. But, because AI is involved, it'll be a cat-and-mouse game.

The panel determined that deep fakes — a technology that currently exists and is spreading posed the greatest threat, based on their potential harm. Deepfakes, the truth is, can trigger a slew of serious problems around the world. These deepfake videos do not have the approval of celebrities or adult film producers. Not only might this violate copyright laws, but it could also cause the celebrity involved significant emotional and mental distress.

## VI. IS ARTIFICIAL INTELLIGENCE A THREAT TO PEOPLE'S JOBS?

Artificial intelligence (AI) has the potential to disrupt every major business and will likely aff ect every aspect of our life. Outside of computer science departments, people must become AI-fluent, or at the very least AI-aware. Gone are the days when a few brilliant data s cientists could be counted on to steer the ship ahead. AI/ML literacy and data-driven decision-making will certainly become professional table stakes in the same way that computer literac y did a generation ago.

Artificial intelligence offers several advantages when used in daily life, such as making work easier and faster. However, it is also becoming a source of joblessness. As a result, the most difficult task would be to implement it in such a way that it would not increase unemployment.

Many individuals feel that artificial intelligence will be the leading cause of job losses shortly. Do they have a point? ... According to the research, AI would dispense with or destroy 75 million employment by 2022, with a net rise of 133 million new jobs. In 2024, it is anticipated that automation would increase by 12%. Previously, we needed a human to test any software code. However, because automated testing is being used, no human tester is required. As a result, the demand for IT professionals in the industry is reduced.

While technology improvements have traditionally resulted in the creation of new jobs, there are worries that employment possibilities are unevenly distributed across industries and skill levels.

According to studies, extremely regular and cognitive activities, which are common in many middle-skilled occupations, are at a high risk of becoming automated. However, Manual activities in low-skilled, service jobs that demand flexibility and "physical adaptation," as well as high-skilled positions in engineering and research that require creative thinking, have a reduced risk of automation. Automation could worsen income and societal inequities as high- and low-skilled occupations benefit from higher wage premiums while middle-skilled employment is phased off. This ultimately creates inequalities in incomes.

## VII. ALGORITHM DISCRIMINATION

Last year, bias was identified in several commercial technologies. Women and people of colour were unable to be recognised by vision algorithms trained on uneven data sets, and recruiting programmes based on historical data were found to perpetuate existing discrimination.

The lack of diversity within the AI sector is linked to the issue of bias—and therefore more difficult to address. At most, women hold 30% of industry employment and less than 25% of teaching positions at elite institutions. There are also a disproportionately small number of black and Latino scholars.

What to look forward to in 2019: We'll look at algorithms that can provide unbiased outcomes from skewed data, as well as approaches for identifying and minimising bias.

©IJRASET: All Rights are Reserved

Because African scientists researching bias issues may have difficulty obtaining visas to go to other locations, the International Conference on Machine Learning, a prominent AI conference, will be hosted in Ethiopia in 2020. Other events may also shift. Some examples of these incidents are:

1) *Healthcare:* Researchers discovered in October 2019 that an algorithm was used on over 200 million people in US hospitals to predict which patients will likely require additional medical care preferred white patients over black patients. While the race was not a variable in this method, another variable that was substantially connected with race, healthcare expense history, was. The idea was that a person's healthcare demands are summarised by their cost. For a variety of factors, black patients with the same diseases had lower healthcare costs on average than white patients with the same ailments. Fortunately, researchers collaborated with Optum to reduce prejudice by 80 per cent. However, AI bias would have continued to discriminate harshly if they hadn't been interrogated in the first place.
2) *COMPAS:* (Correctional Offender Management Profiling for Alternative Sanctions) The COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) algorithm, which is used in US court systems to forecast the possibility of a defendant becoming a recidivist, is arguably the most well-known example of AI bias. Because of the data collected, the model is chosen, and in the general process of developing the algorithm, the model predicted twice as many false positives for recidivism for black offenders (45 %) as it did for white offenders (23 % ).
3) *Amazon Hiring Process:* Amazon is one of the world's most powerful technology companies. As a result, it's no wonder that machine learning and artificial intelligence are heavily used by them. Amazon discovered in 2015 that their algorithm for hiring staff was skewed against women. The explanation for this was that the algorithm was trained to prefer men over women based on the number of resumes submitted over the previous 10 years, and because the majority of the candidates were men.

## VIII. WHAT'S STOPPING US TO HAVE GOVERNANCE OVER AI ?

1) Governments face numerous obstacles in establishing and executing effective AI policy due to the high level of ambiguity and complexity in the AI ecosystem. Many of the difficulties posed by AI are due to the nature of the problem, which is highly unexpected, intractable, and nonlinear, making it impossible for governments to articulate precise policy objectives.
2) The inherent opacity and unpredictability of machine learning systems provide technological hurdles for governments in establishing AI accountability. Tobegin with, the opacity of complex ML algorithms remains a major barrier to AI governance because it limits the extent of transparency, explainability, and accountability that can be achieved in AI systems (Algorithms are often intentionally opaque by their developers to prevent cyberattacks and to safeguard trade secrets, which is legally justified by intellectual property rights, and the complexity of ML algorithms is a major barrier to AI governance because it limits the extent of transparency, explainability, and accountability that can be achieved in AI systems).
3) The difficulty of assigning liability and accountability for harms resulting from software defects is exacerbated by the lack of human controllability over AI systems' behaviour, as manufacturers and programmers are often unable to predict the inputs and design rules that could result in unsafe or discriminatory outcomes.
4) Existing governance and regulatory frameworks are also ill-equipped to deal with the societal issues brought on by AI, due to a lack of information needed to understand the technology and regulatory gaps. Google, Facebook, Microsoft, and Apple, among other major technology companies and AI developers, have significant informational and resource advantages over governments in regulating AI, which significantly outweighs governments' traditional role in distributing and controlling resources in society.
5) Legislators may be deterred from defining specific rules and responsibilities for
6) algorithm programmers to allow for future experimentation and code modifications to improve the software but doing so allows programmers to avoid responsibility and accountability for the system's societal behaviour. These challenges show how the four governing resources traditionally used by governments for regulation are insufficient to manage the risks posed by AI, highlighting the need for governments to find new ways of gathering information and developing effective policies that can adapt to the changing AI landscape.
7) The unfettered influence of companies and major political players in the AI landscape could increase power imbalances and socioeconomic inequities since the ideology and interests of a small group of people could manifest themselves in AI design and the decisions they make in society . More research is needed to examine the key actors, their roles, the dominant ideas, and values promoted in AI policies, whether there is a global convergence in these values across countries, and the degree to which these values reflect society's interests or are politically motivated, to ensure greater inclusivity and diversity in AI governance.

©IJRASET: All Rights are Reserved

## IX. CONCLUSION

Trust is essential for AI solutions to be revolutionary. Integrity, explainability, fairness, and resilience are the four main anchors on which this trust is built. These four guidelines assist businesses in ensuring that algorithms are properly governed.
Integrity entails algorithm integrity as well as data validity, which includes lineage and the appropriateness of data use.
Explainability — gaining transparency through comprehending the algorithmic decision-making process in layman's terms.
Fairness entails ensuring that AI systems are ethical, free of bias and prejudice and that protected characteristics are not employed.
Resilience refers to your AI's technical robustness and compliance, as well as its adaptability across platforms and resilience to attack.
AI dangers in future include using driverless cars as weapons, using machine learning to tailor phishing messages, disrupting machine-controlled systems, writing fake news, deepfake, and harvesting online information for blackmail.
Who will be entrusted with the task of regulating and incentivizing?
Policymakers and international organisations are already collaborating to create circumstances for AI applications in the form of standards, protocols, and regulations.
Regulatory bodies could impose restrictions on AI auditability, capability, and applications, as well as impose financial penalties on offenders. On the other hand, these same bodies, as well as private companies and philanthropists, may choose to reward people who create and deploy AI in a way that is consistent with societal standards.
Who is in charge of ensuring that artificial intelligence is created and applied in a fair, safe, and ethical manner?
Before wide-scale implementation, leaders and developers must seek to discover and prevent "algorithmic accidents" to reduce the risk of societal biases and prejudices sabotaging outcomes. Moving away from unpredictable "black-box" algorithms and toward transparent, auditable models that promote accountability is the first step in this approach.
If AI is just permitted to accelerate the occurrence of undesirable behaviours, it will not be a force for good in the future.
As a crucial forum for executive and parliamentary debate and dialogue, it is the ideal venue for discussing the future of digital governance and responding to one of the world's most pressing dangers and challenges. There isn't yet a single correct solution for the optimal AI roadmap, but there are various possibilities. We must collaborate to determine which path will benefit the most people. By participating in this discussion and driving the dialogue, the G20 may become the backbone of a new architecture for the twenty-first century, assuring a brighter future for all.

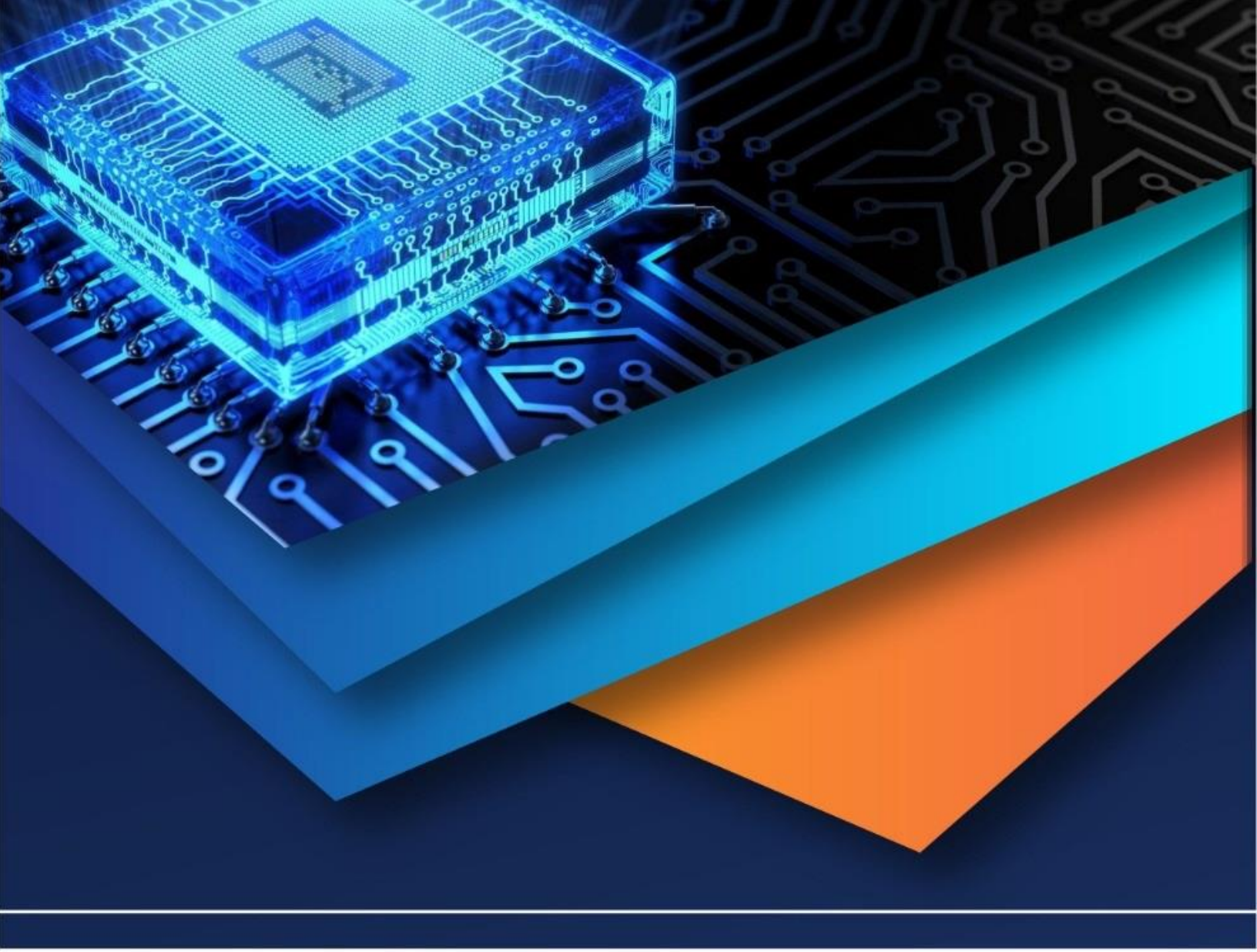
## X. ACKNOWLEDGEMENTS

I thank the following individuals for their expertise and assistance throughout all aspects of our study and for their help in writing the manuscript.

## REFERENCES

[1] Cath Corinne. 2018 Governing artificial intelligence: ethical, legal and technical opportunities and challenges. Phil. Trans. R. Soc. A. 376: 20180080. 20180080. http://doi.org/10.1098/rsta.2018.0080
[2] Julia Pomares, María Belén Abdala [last access 2020] The future of AI governance The G20's role and the challenge of moving beyond principles https://www.global-solutions-initiative.org/wp-content/uploads/2020/04/GSJ5_Pomar es_Abdala.pdf
[3] Ethics and governance of artificial intelligence for health: WHO guidance ISBN 978-92-4-002920-0 (electronic version) ISBN 978-92-4-002921-7 (print version) © World Health Organization 2021 Some rights reserved. This work is available under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 IGO licence (CC BY-NC-SA 3.0 IGO; https://creativecommons.org/licenses/by-nc-sa/3.0/igo).
[4] LARSSON, Stefan. (2020). On the Governance of Artificial Intelligence through Ethics Guidelines. Asian Journal of Law and Society. 7. 1-15. 10.1017/als.2020.19.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)