



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 9, September 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.524

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Artificial Intelligence Ethics: Investigating Ethical Frameworks, Bias Mitigation, and Transparency in AI Systems to Ensure Responsible Deployment and Use of AI Technologies

Archana Todupunuri

Fidelity Information Services, USA

ABSTRACT: Advancements and usefulness of artificial intelligence (AI) have been transforming the ways in which companies reach decision making and ethical adherence in competitive market. This study has focused on secondary data collection and thematic analysis for developing understanding about the mitigation of bias and transparency in the use of AI technologies. This study has been responsible for stating about different methods that has been used for gaining proper outcomes. Apart from that, preparation of thematic analysis has been effective to reach the research objectives effectively. Finally, this study has also established strategic recommendations through which organisation can imply proper ethical principles for better utilisation of AI technology.

KEYWORDS: Artificial intelligence (AI), technology, bias, transparency, RAI.

I.INTRODUCTION

Background

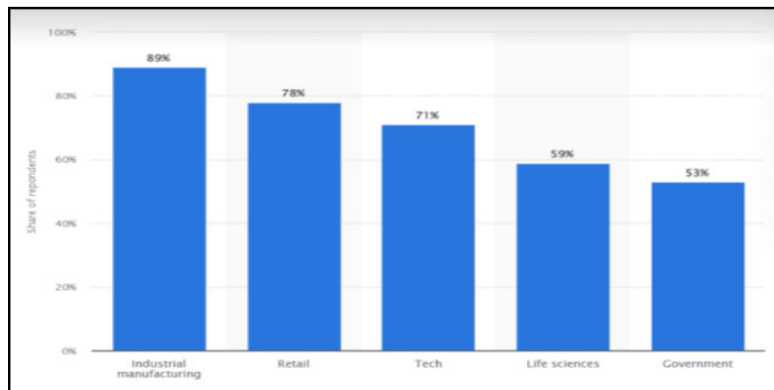


Figure 1: “Ethical Policies for AI” in organisations

“Artificial Intelligence (AI)” is responsible for making remarkable development as well as rapid changes within businesses operations for the last few decades. Different AI technologies that organisations are using for the welfare of their businesses are natural language processing, and “Machine Language (ML)” [1]. In this respect AI ethics is also considered to be one of main concerns for organisations in order to maintain depletion of AI technologies. In figure 1, it can be noticed that “89% of respondents” from the manufacturing industry agreed that proper ethical policies in AI usability are used in organisations [18]. On the other hand, AI measurements and development is accountable for helping organisations in building awareness, managing risks, improving individual’s well-being and provide actionable insights [2]. Finally, in this respect, principles and ethical guidelines regarding “use of AI technologies” by organisations is considered as profitable aspect in overcoming ethical issues related to “use and deployment” of AI technologies.

Problem statement

Main ethical challenges that organisations face in using AI technologies and are being highlighted in this study are security issues and privacy leakage.

Aim

Main aim of this study is to focus on an ethical framework that organisations can use for maintaining artificial intelligence ethics, transparency and bias mitigation for ensuring usability and deployment of AI technologies.

Objectives

- To comprehend concepts of artificial intelligence ethics and AI systems
- To analyse factors influencing organisations in maintaining artificial intelligence’s “ethical framework, transparency and bias mitigation” through using AI technologies for using and deploying it
- To evaluate impact of using AI technologies and maintaining AI related “ethical framework, transparency and bias mitigation” for ensuring its usability and deployment
- To identify challenges that organisations face while supporting artificial intelligence’s “ethical framework, transparency and bias mitigation” in order to use and deploy AI technologies
- To provide appropriate recommendations to organisations regarding using AI technology’s “ethical framework, transparency and bias mitigation” ethically for better deployment and usability

Questions

1. What are concepts of artificial intelligence ethics and AI systems?
2. What are factors influencing organisations in maintaining artificial intelligence’s “ethical framework, transparency and bias mitigation” through using AI technologies for using and deploying it?
3. What are impact of using AI technologies and maintaining AI related “ethical framework, transparency and bias mitigation” for ensuring its usability and deployment?
4. What are the challenges that organisations face while supporting artificial intelligence’s “ethical framework, transparency and bias mitigation” in order to use and deploy AI technologies?
5. What are appropriate recommendations to organisations regarding using AI technology’s “ethical framework, transparency and bias mitigation” ethically for better deployment and usability?

II. LITERATURE REVIEW

A. Concept of ethics in artificial intelligence (AI) systems

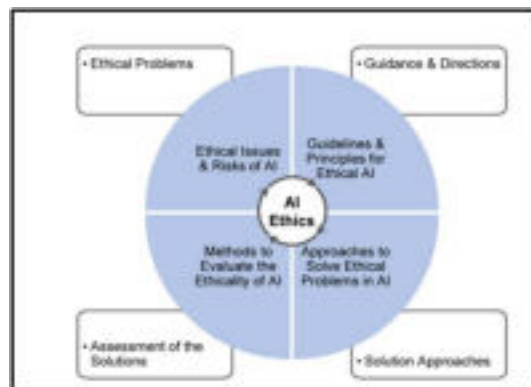


Fig. 2. Ethics in AI

AI ethics can be defined as the set of guidelines and practices that need to be incorporated at the time of responsible use of AI technology. A storing mode of ethics in AI systems can include the activities of avoiding bias, management of privacy of the users along with mitigation of the environmental risks [3]. In context with this, the approaches which can reshape the maintenance of ethics in AI system applications involve “government-led regulatory frameworks” and “codes of ethics in companies”. These approaches have become beneficial for the maintenance of responsible regulation of AI in the companies, covering all the national and international issues associated with the use of AI systems. On the other hand, the ethics in AI systems has become beneficial as this technology is aimed as augmenting or replacing human intelligence and lack of ethics can lead to biased results [4]. Therefore, the development of the AI system with biased or inaccurate data can create serious consequences for the individuals or the groups.

B. Ethical challenges in the use of AI system

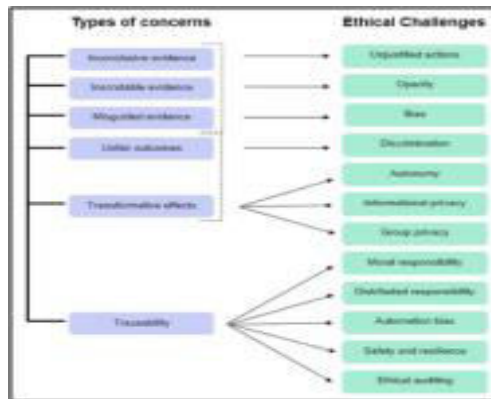


Fig. 3. Ethical challenges in the use of AI system

The operational parameters of decision-making algorithms of the AI system have identified a number of ethical challenges including bias, discrimination and lack of privacy or transparency. The AI system is dependent on making decisions based on the provided data and the decisions of AI can result in selection bias in the case of inaccurate data and biased sampling [5]. Accordingly, the biased result produced by AI systems can contribute to discrimination within the companies that depended on AI based decision making. The biased nature of decision making related to the AI systems has a major challenge for companies which further might create the issue of legal risk for companies due to increasing discrimination. Conversely, lack of transparency and privacy is another significant ethical challenge in AI systems which make it difficult for the users to understand the decision making process of the technology [6]. The lack of visibility to the users at the time of using AI leads to accelerating the issue of transparency. As a result, the transparency issue creates concern regarding trust to the information provided by AI at the time of using the technology.

C. Strategies for responsible use and deployment of AI technologies

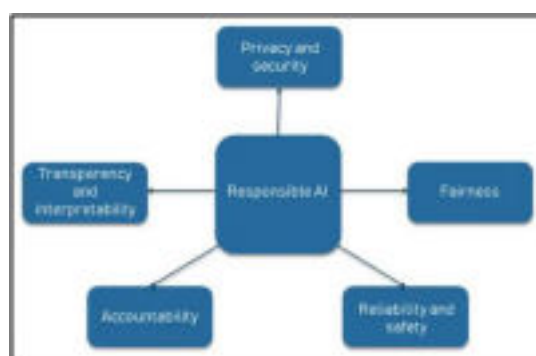


Fig. 4. Responsible use of AI technologies

AI and machine learning (ML) have played crucial roles in transforming the aspect of lives along with revolutionising the industries. The implementation of the algorithms for the detection and mitigation of bias can be beneficial for reducing bias in the use of AI technologies [7]. In this context, the techniques and algorithms that help in AI bias mitigation includes adversarial training, bias audits along with the rectification of bias in the models. However, the maintenance of transparency and privacy in the use of AI technologies can include activities like the allocation of quality data [8]. Companies needs to be more focused on ethical sourcing of the information along with regular analysis of the data to ensure that transparency management was considered adequately. Therefore, the activities of data quality assurance and the selection of unbiased information are directly associated with maintaining transparency at the time of using AI technologies.

D. Theoretical underpinning Technology acceptance model

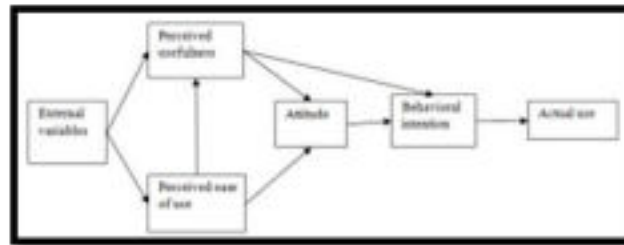


Fig. 5. “Technology acceptance model (TAM)”

“Technology acceptance model (TAM)” states that the perceived usefulness and perceived ease of use of technology leads to decisions about the acceptance of advanced technology. The principles of TAM reflect that the behavioural intention of individuals can play a significant role in generating understanding on the adoption of technology [9]. Accordingly, the application of this model can be beneficial in understanding the attitude and behaviours of the individuals about the benefits and challenges of advanced technologies like AI. However, the ethical issues associated with the use of advanced technologies like AI has resulted in creating negative attitudes and behaviour among the individuals. Therefore, the integration of the TAM is beneficial in the development of the understanding about the contribution of technology like AI in human life along with the behaviour of the individuals in adopting the technology.

E. Literature gap

Previous literature has mentioned about the ethical challenges of AI including bias and the aspect of transparency. However, the other ethical issues associated with the use of AI like autonomy, informational privacy are not covered in the existing studies. Existing literature has not critically discussed the strategy that are utilised by the organisations for ensuring responsible use of AI technologies with the integration of real life examples. Therefore, this study has focused on mitigating the gap in previous studies by covering diverse areas of ethical issues with the use of AI along with the strategies to respond to the issues.

III. METHODS

Research Onion

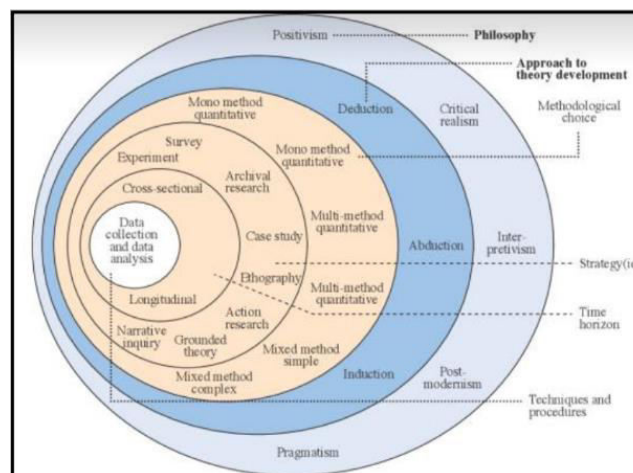


Figure 6: Research Onion

Research onion was applied to gain precise ideas regarding different stages of research methodology in research work which include "philosophy strategy data analysis approach and data collection" [10].

Research philosophy

Research philosophy is divided into three parts which are "pragmatism interpretivism and positivism research philosophy". Selected research philosophy for this study is "interpretivism research philosophy" due to its advantageous characteristics of helping in gaining better insights about artificial intelligence insights.

Research approach

Three types of research approaches are "inductive adductive and deductive research approaches". "Inductive research approach" was chosen for the study as it helped in collecting data regarding artificial intelligence's "ethical framework, transparency and bias mitigation".

Research strategies

Different variants of "research strategies" are "action research strategy grounded theory case study experiment and survey". In this study incorporation of "archival strategy" was done which included qualitative study in order to collect information about ethical behaviour of employees working in organisations that implement AI technology for betterment of business operation.

Data collection method

Data collection process is considered as one of the most necessary requirements in order to perform research work accordingly and gain clear understanding regarding concerned research topic. Two fundamental forms of data collection methods are "primary and secondary "data collection methods [11]. Moreover, this study has utilised the "secondary data collection method" to ensure a seamless and structured collection of data related to ethical application of AI technology in organisations. Information about artificial intelligence ethics were collected from different secondary databases which were "Research gate, IEEE, and Google consideration of a secondary method for collecting data was done in order to gain valuable secondary data regarding deployment of AI technology in an ethical manner in order to me ticket its ethical issues and maintain transparency accordingly.

Search strategy

One of the main requirements in secondary research is proper way of searching "pertinent information" as it helps in executive research work in a proper manner. In the study different keywords that were used include "artificial intelligence, AI technologies, ethical framework in AI and transparency within AI system". Finally, for conducting this research successfully key phrases were also used which include "deployment of AI technologies, ethical framework for deploying AI technology, and bias mitigation within AI systems to using ethical frameworks".

Inclusion and exclusion criteria

"Inclusion and exclusion" criteria within research work is an interconnected research strategy that further helps in assessing necessary information about artificial intelligence ethics in order to facilitate creation of "authentic research results". Secondary general and articles that were published on and from 2020 considered and once that only provided abstract sections were excluded.

Data analysis

"Thematic and statistics" are two forms of data analysis methods used for research work. "Qualitative thematic analysis" was considered for this study in order to develop themes to emphasise on research objectives which assisted in gathering necessary and relevant information regarding artificial intelligence that are maintained by organisations. Finally, thematic data analysis was chosen for the study as it helped in gaining huge information about artificial ethics in regard to ethical framework bias mitigation as well as transparency maintenance in respect to AI technology deployment in organisation.

Ethical consideration

"Reliability and validity" of research work is interconnected with ethical consideration of study as it assists in acquiring proper research results. Guidance for collecting data from secondary sources was gained through the following "Copyright, Designs, and Patents Act 1988 (c.48)".

IV. DATA ANALYSIS

Theme 1: Organisations prefer using emerging technologies such as AI as data-driven learning systems for maintaining transparency and mitigating bias.

Use of AI within business operations is highly praised by organisations in order to have zero human supervision or involvement in decision-making processes. Application of AI within is expected to be helping organisations in different ways such as taking data-driven decisions for welfare of businesses [12]. Through including AI within decision making system organisations are available to make a positive impact on human life which include healthcare along with maintaining transparency as well as mitigating bias. On the other hand, organisations in the agriculture sector also use AI in their business activities for using information as well as communication technology effectively [13]. Precision agriculture which is known to be a management strategy is accountable for using artificial intelligence and mission learning simultaneously increasing cost, climate variation and soil degradation.

Theme 2: Organisations are using AI that is designed through using human-centred approach along with explain ability through maintaining ethical principles

In recent times, “Human-centred AI (HCAI)” is in high demand as organisations are using it for improving their decision-making process which includes betterment of multi-stakeholder engagement. Use of “Human-centred AI (HCAI)” helps employees of organisations in reframing technology-centric and traditional approaches used in their supply chain [14]. On the other hand, HCAI is known to be an “emerging discipline” for caretaking AI systems that organisations use in order to augment and amplify in place of displacing human abilities in business operations [14]. Implementation of HCAI also further assists organisations in preserving human control for providing beneficial outcomes by AI along with maintaining ethical principles. Therefore, through using HCAI, it is possible for firms to use both AI systems and humans simultaneously rather than replacing humans with AI systems.

Theme 3: Positive impact of ethical framework of AI in healthcare is indulgence of “Deep Learning (DL) and Machine Learning (ML)”

Applying an ethical framework has become an effective and necessary application in healthcare which can be facilitated by utilising AI while providing better health care services to the public. Through an ethical framework in regard to AI technology implementation, it has become easier for organisations of different sectors of healthcare to use AI in different domains which include governance and transportation [17]. On the other hand, through fastening open dialogues within stakeholder’s organisations are able to help them in creating dynamic ethical framework which would further assist in adopting AI applications. Proper use of ethical framework regarding AI technology usability and deployment is also expected to be allowing different stakeholders such as employees in promoting AI development [17]. Finally, AI technology usability and deployment is also expected to be for improving their performances.

Theme 4: Different challenges that organisations face while using AI in their businesses are vague principles, “lack of ethical knowledge” and poor technical understanding

Organisations face difficulties in following ethical principles and guidelines that need to be followed for proper maintenance of AI systems due to different causes. Lack of technical understanding, ethical knowledge as well as vague principles are considered to be some of main reasons behind poor use of AI systems [16]. Contrastingly, due to lack of training it becomes difficult for employees to have proper understanding about ethical challenges they might face while using AI technology in their performances within what place. Therefore, it is the responsibility of the organisation to pay attention towards challenges that employees are facing due to inappropriate ethical practices in regard to AI technology implementation.

Theme 5: Companies, organisations and governments are responsible for releasing ethical principles that can be used for tackling “Responsible Artificial Intelligence (RAI)”

Use of RAI in companies can posit both positive and negative consequences to the services or facilities expected by the marketplace. For example, unless RAI is deployed appropriately through adhering to all ethical standards, it becomes difficult for organisations to apply it in business as a positive aspect. Algorithm level solutions are designed by organisations and government as AI ethic principles which mainly focuses on fairness and privacy for mitigating respective challenges [15]. Therefore, through developing algorithm level solutions it becomes easier for organisations using RAI to maintain privacy as well as in their business operation in collaboration with using AI.

V. DISCUSSION

AI technologies are potentially making changes in different aspects of business operations which include the decision making process. This study explored that inclusion of AI in business has been a beneficial integration in organisations which helps to remove human interference that is viewed as be one of the main reasons that lead to ethical challenges. Existence of ethical challenges in AI technology usability occurs due to presence of differences within law makers and practitioners in organisations [16]. In addition to this, magnitude of aspects combined with AI technology helps to simplify complexity in bias management that leads to argue more systematic approach [20]. On the other hand, it is difficult for organisations to mitigate ethical challenges without support of government. Moreover, majority of governmental bodies and organisations have incorporated AI technology in business process for endorsing ethical principle into firm performance optimisation [21]. In this context both organisations and government are found to be responsible for creating ethical principles for proper use of AI technology and in order to mitigate ethical related challenges appropriately.

VI. CONTRIBUTION AND LIMITATION

Business natural language processing becomes easier for organisations by the application of AI. In addition, the decision making capacity of an organisation increased by the application of AI. competitive advantages of an organisation get improved by the application of AI through better application of AI. employees failed to access AI due to lack of training of employees. Company failed to make proper utilisation of AI due to poor investment over employees training. Hence, it can be concluded that organisational performances get improved by the use of AI through maintaining privacy at the workplace, chances of threats of competitors get reduced by the use of AI.

This study has experienced main limitation of choosing a new type of context such as “ethical framework that organisations can use for maintaining artificial intelligence ethics for ensuring usability and deployment of AI technologies”. Application of secondary data collection method has created barriers for researchers increasing clarity in the study due to lack of financial support. Avoidance of use of primary data collection method has reduced scope of this study in terms of not including responses of participants related to this topic. On the other hand, time limitation has created barriers for the choosing of secondary articles. Unique topic has created a barrier for researchers in choosing accurate journal articles for this study.

VII. CONCLUSION AND RECOMMENDATION

In conclusion, it can be stated that ethical standard management provides positive contribution to the companies which apply the AI systems in decision making. It has been found that AI has helped in revolutionising industries all across the globe by transforming the way of decision making in place of human intelligence. The ethical issues associated with the use of advanced technologies like AI are identified in this study which includes bias, lack of transparency and privacy. This article provides a comprehensive overview of AI ethics with analysing and summarising ethical risks as per the demands of the research objectives.

The improvement in the quality of data provided to the AI system can be recommended for mitigating the ethical issues like biased decisions making. Code of ethics creation at the time of designing AI systems is one of the first major steps for maintaining ethical practices in AI technologies [19]. Accordingly, the development of code of practices can be recommended which can be followed by an AI system at the time of maintaining its performance. Therefore, the strategies like the improvement in quality data, responsible deployment along with the creation of code of ethics are significant aspects for mitigating bias and transparency sites in the use of AI technologies.

REFERENCES

- [1] Huang, C., Zhang, Z., Mao, B. and Yao, X., 2022. An overview of artificial intelligence ethics. *IEEE Transactions on Artificial Intelligence*, 4(4), pp.799- 819.
- [2] Schiff, D., Ayesh, A., Musikanski, L. and Havens, J.C., 2020, October. *Ieee 7010: A new standard for assessing the well-being implications of artificial intelligence*. In 2020 IEEE international conference on systems, man, and cybernetics (SMC) (pp. 2746-2753). Ieee.
- [3] Wang, J.Z., Zhao, S., Wu, C., Adams, R.B., Newman, M.G., Shafir, T. and Tsachor, R., 2023. Unlocking the

Emotional World of Visual Media: An Overview of the Science, Research, and Impact of Understanding Emotion Drawing Insights From Psychology, Engineering, and the Arts, This Article Provides a Comprehensive Overview of the Field of Emotion Analysis in Visual Media and Discusses the Latest Research, Systems, Challenges, Ethical Implications, and Potential Impact of Artificial Emotional Intelligence on Society. Proceedings of the IEEE.

[4] Chamola, V., Hassija, V., Sulthana, A.R., Ghosh, D., Dhingra, D. and Sikdar, B., 2023. A review of trustworthy and explainable artificial intelligence (xai). IEEE Access.

[5] Nazar, M., Alam, M.M., Yafi, E. and Su'ud, M.M., 2021. A systematic review of human-computer interaction and explainable artificial intelligence in healthcare with artificial intelligence techniques. IEEE Access, 9, pp.153316-153348.

[6] Madhavan, R., Kerr, J.A., Corcos, A.R. and Isaacoff, B.P., 2020. Toward trustworthy and responsible artificial intelligence policy development. IEEE Intelligent Systems, 35(5), pp.103-108.

[7] Sanderson, C., Douglas, D., Lu, Q., Schleiger, E., Whittle, J., Lacey, J., Newnham, G., Hajkowicz, S., Robinson, C. and Hansen, D., 2023. AI ethics principles in practice: Perspectives of designers and developers. IEEE Transactions on Technology and Society.

[8] Cihon, P., Kleinaltenkamp, M.J., Schuett, J. and Baum, S.D., 2021. AI certification: Advancing ethical practice by reducing information asymmetries. IEEE Transactions on Technology and Society, 2(4), pp.200-209.

[9] Omrani, N., Rejeb, N., Maalaoui, A., Dabić, M. and Kraus, S., 2022. Drivers of digital transformation in SMEs. IEEE Transactions on Engineering Management.

[10] Saunders M., Thornhill, A. and Lewis P. (2019) Methods for Business Students (8th edition). Person Education: London.

[11] Cvitić, I., Perakovic, D., Gupta, B.B. and Choo, K.K.R., 2021. Boosting-based DDoS detection in internet of things systems. IEEE Internet of Things Journal, 9(3), pp.2109-2123.

[12] Rawal, A., McCoy, J., Rawat, D.B., Sadler, B.M. and Amant, R.S., 2021. Recent advances in trustworthy explainable artificial intelligence: Status, challenges, and perspectives. IEEE Transactions on Artificial Intelligence, 3(6), pp.852-866.

[13] Singh, R.K., Berkvens, R. and Weyn, M., 2021. AgriFusion: An architecture for IoT and emerging technologies based on a precision agriculture survey. IEEE Access, 9, pp.136253-136283.

[14] Schoenherr, J.R., Abbas, R., Michael, K., Rivas, P. and Anderson, T.D., 2023. Designing AI using a human-centered approach: Explainability and accuracy toward trustworthiness. IEEE Transactions on Technology and Society, 4(1), pp.9-23.

[15] Lu, Q., Luo, Y., Zhu, L., Tang, M., Xu, X. and Whittle, J., 2023. Developing responsible chatbots for financial services: A pattern-oriented responsible AI engineering approach. IEEE Intelligent Systems.

[16] Khan, A.A., Akbar, M.A., Fahmideh, M., Liang, P., Waseem, M., Ahmad, A., Niazi, M. and Abrahamsson, P., 2023. AI ethics: an empirical study on the views of practitioners and lawmakers. IEEE Transactions on Computational Social Systems.

[17] Nasir, S., Khan, R.A. and Bai, S., 2024. Ethical Framework for Harnessing the Power of AI in Healthcare and Beyond. IEEE Access, 12, pp.31014-31035.

[18] Statista, 2024, Implementation rate of ethics policies for artificial intelligence (AI) in organizations worldwide in 2021, by industry Available at: <https://www.statista.com/statistics/1232659/worldwide-artificial-intelligence-implementation-ethics-policies/> [Accessed on: 24th May, 2024]

[19] Narang, N.K., 2024. Mentor's Musings on Concerns, Challenges & Opportunities for Generative AI at the Edge in IoT. IEEE Internet of Things Magazine, 7(3), pp.6-11.

[20] Larsson, S. and Heintz, F., 2020. Transparency in artificial intelligence. Internet policy review, 9(2).

[21] Ashok, M., Madan, R., Joha, A. and Sivarajah, U., 2022. Ethical framework for Artificial Intelligence and Digital technologies. International Journal of Information Management, 62, p.102433.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details