AI Ethics: Bridging Innovation with Morality

Saroj Singh[1], Priyanka Gupta[2], Amit Kumar Sachan[3], and Sonam Singh[4]

[1, 2] Assistant Professor, Department of Computer Science & Engineering, Babu Banarasi Das Engineering College, Lucknow
[3]Assistant Professor, Department of Computer Science & Engineering, Babu Banarasi Das Institute of Technology and Management, Lucknow
[4]Assistant Professor, Department of Computer Science & Engineering, Babu Banarasi Das University, Lucknow

Abstract: With advancements in face recognition, self-driving cars, and medical diagnostics, artificial intelligence (AI) is having an increasing impact on people and society. AI has the amazing potential to significantly improve human wellbeing and advance both socially and economically. However, because to obstacles such limited viability, data biases, data security, and some privacy concerns, AI also presents significant risks. To reduce hazards, users and developers need to be aware of these drawbacks and take appropriate action. To anticipate and lessen negative effects, it is essential to identify possible areas of danger across the AI life cycle. Even though "machine ethics" research started in 2006, AI ethics is still a relatively new area of study within ethics more broadly. Establishing appropriate ethical principles, norms, standards, rules, and regulations requires an understanding of AI ethics and how to develop ethical AI. AI that is ethical acts and behaves in a morally upright way.

There are many different perspectives on AI, and in order to properly identify AI, developers, scholars, business leaders, legislators, and the general public must all contribute. This paper examines ethical difficulties and prejudices in AI, emphasizing injustice and bias as pressing problems from a sociotechnical standpoint. Uncertain standards and a lack of ethical expertise are the primary barriers in AI ethics.

Keywords: AI ethics, Bias; Data Security; Sociology, Robo ethics.

1. INTRODUCTION

AI is not a new field; rather, it is just coming out of one of its many "winters" and is starting a new "spring" [1,2]. AI's development is transforming society in ways that were unthinkable just a few years ago. Even though AI has been around for a while, its pervasiveness in our daily lives is frequently invisible. Whether it's for reading emails, conducting Google searches, recommending movies and music, or even providing driving directions, artificial intelligence has become a vital part of our lives. The COVID-19 epidemic has increased demand for AI's help and highlighted our reliance on it. AI is becoming more and more common in a number of sectors, including communications, education, healthcare, transportation, and agriculture. It is necessary to regularly interact with AI-powered applications in order to survive in the current world. [3, 4].

A viewpoint on the early machine age was offered by considering technology as a tool. However, advancements driven by AI are altering how people interact and exist in the current second age of machines. Governments, corporations, and other organisations frequently utilise AI-based algorithms to make decisions that have a big influence on people and society. Everyone, everywhere, and at any time could be impacted by their choices. They might provide answers to issues that come up in a range of occupations or in daily life, but they might also carry risks, including being turned down for a job or medical care. The discriminatory effect that AI-based decision-making has on some demographic groups has been shown by prior incidents. The company's ad engine, which has been shown to show far fewer high-paying job ads for women than for males, is one instance of gender bias in Google's targeted advertising [5]. These incidents have increased public anxiety about the potential effects of AI on our day-to-day lives.

Since it is normal for most people to be ignorant of the experiences of other groups, prejudice is not a new occurrence. Instead, it dates back to the earliest days of human civilisation [6]. However, AI-based decision-making has the potential to establish completely new categories, norms, and biases of a completely different kind, which could reinforce pre-existing biases. These ever-growing worries have prompted a reassessment of AI-based systems in favour of new approaches that also consider how equitable their judgements are.

Many academics and industry professionals believe that since AI is still far from being on par with humans, there is no urgent need to think about ethical issues. However, when paired with other intelligent technologies like robotics, AI has already shown promise in a variety of industries, including business, healthcare, and transportation. Applications such as autonomous cars bring up significant ethical issues, and AI is already having an impact on civilisation and society. Autonomous vehicles, for example, have the potential to eliminate a wide range of jobs, which would significantly impact related businesses like transportation. For instance, allowing

people to work and sleep while travelling in autonomous vehicles will have an impact on roadside lodging services and short-distance flights. It is commonly known that biases acquired from training data cause AI systems, especially hiring tools, to harbour human assumptions. There is a chance of social unrest and revolutions when wealth disparities arise as a result of the varying rates of return on capital and labour. Future employment and humankind may be impacted by AI, so preparations and actions must be taken. In order to navigate these difficult issues, the creation of ethical AI is not only essential, but also critical. Despite its importance, creating moral AI is still a very difficult and complex undertaking.

It is only normal for individuals in the ruling majority to be unaware of the struggles endured by other groups, considering the lengthy history of prejudice in human civilisation. However, there's a significant probability that AI-powered decision-making may add new biases of its own and reinforce pre-existing ones by creating new categories and standards. These growing concerns have led to a reassessment of AI-based systems in favour of new approaches that incorporate decision-fairness as well.

The application of AI benefits society, but it also poses social and ethical risks. The renowned physicist Stephen Hawking has noted the significance of assessing these hazards for humanity's future. Because of its superiority in a small number of specific fields, the current version of AI is typically referred to as narrow or weak AI. The programming and training data, which are intimately related to humans and vast amounts of data, affect how well weak AI performs. Therefore, human variables also provide ethical issues with limited AI.

AI typically uses data that has been supplied by people (like user-generated content) or acquired using technology that has been created by people. Unfortunately, the complex nature of sociotechnical systems like the web exacerbates people's preconceptions, which are subsequently incorporated into our systems. According to Karimi et al. [7], algorithms may perpetuate or even exacerbate underlying discrimination or inequality. particular social groups may experience disadvantage in society, which can frequently lead to "institutional bias," which is the propensity of particular institutions' practices and policies to favour some social groups while damaging others. This may not be the result of deliberate discrimination, but rather of the majority following the rules. In addition to being a part of pre-existing (biassed) organisations and processes, algorithms can also induce or worsen bias. This is due to their propensity to give more weight to easily assessed aspects of human behaviour than to more complicated or even challenging ones. This issue is made worse by the possibility that some data is more reliable and simpler to obtain than others, as demonstrated by the exaggerated involvement of Twitter in many social events [8]. When implemented, algorithm-based systems encourage the expansion of highly specialised data gathering methods and infrastructures, including tracking and surveillance, which alter or worsen power dynamics [9]. Thus, algorithms impact and mould societal systems and potential solutions. The goal of AI and Ethics is to encourage thoughtful debate about the moral, legal, and political ramifications of AI developments. It will concentrate on the advancements in AI methods, instruments, and technologies while taking into account potential future directions for these discoveries.

The ethical advantages of using AI are described in Section II, the specifics of what constitutes bias in AI are covered in Section III, the difficulties in creating ethical AI are covered in detail in Section IV, the ethical issues are highlighted, and the conclusion is included in Section V.

2. COMPREHENDING THE ETHICAL : ADVANTAGES OF AI

A collection of moral precepts and guidelines known as AI ethics are intended to direct the creation and appropriate application of AI technology. AI ethics are beginning to change as AI has become a crucial component of services, goods, and organisations. An AI ethical statement, also known as a problem of value, is a formal policy declaration that outlines the contribution of artificial intelligence to human welfare and advancement. An AI ethics handbook is intended to serve as a reference for anyone making moral decisions about the use of AI.

The use of AI in contemporary life has encouraged the use of data-driven algorithms in high-stakes enterprises. But it's becoming harder and harder for individuals to comprehend how these intricate and mysterious algorithms operate and why they reach the conclusions they do. Since different people have varied opinions about AI, it is critical to address the main ethical concerns before AI may be used to support significant decisions in order to avoid misconceptions that could negatively affect people. To effectively define AI, it is crucial to bring together the diverse viewpoints of the many groups of developers, researchers, corporate executives, legislators, and individuals.

At the moment, AI-powered technologies have the potential to improve morality. The International Risk Governance Centre (2018) emphasises AI's analytical capabilities, or its capacity to delve into sources and volumes of data that are beyond human

comprehension. AI has the ability to link data, spot patterns, and provide insights in a variety of academic fields and disciplines. AI could relieve humans of monotonous duties since it is more consistent than humans and can swiftly adjust to new data inputs. Since they allow us to better understand a variety of phenomena, these technological developments are often seen as improving human well-being. Reducing travel times and improving email spam filters are two real-world examples of how AI may improve daily life [10].

There is sometimes an implicit presumption that morally reprehensible topics are being discussed while talking about the ethical concerns surrounding AI. It is important to remember that AI has a lot of promise, and many policy documents extensively emphasise the predicted financial benefits of better production and efficiency. However, these ethically dubious discoveries are the centre of the AI debate and require discussion. AI is included in the advanced technology ethics, which focusses on artificially intelligent things like robots, because they improve quality of life, reflect ethical ideals, and give higher wealth levels and welfare—possibly the crucial elements for human flourishing. This area of study can be essentially divided into two categories: robo-ethics and machine ethics. Robo-ethics examines the moral conduct of persons who create, use, and interact with AI agents as well as the effects of robots on people and society. Combining the social sciences and artificial intelligence results in a "fruitful arrangement." Despite the fact that the humanities encourage the development of AI techniques and socially beneficial structures, sociological research benefits from advances in AI.

Philosophy, sociology, and other academic fields may help us understand AI more thoroughly. A sociological perspective on AI adds to our understanding of the broad ramifications AI will have on society and is essential for developers and designers who want to anticipate unexpected consequences. The dearth of persuasive empirical research in the social sciences is the reason for the necessity for a sociological perspective on AI. Instead of philosophically anchored research issues, the majority of current research is driven by technological potential applied to social and economic events [11]. More precisely, AI systems might be able to inform and advance technological advancement, but in the former case inequality, diversity, and power hierarchies they are more likely to be unaware of social complexity.

If our objective was to develop ethics for people using technology in this setting, then we urgently need to talk about ethics for robots. AI differs from previous technologies in that it aims to mimic human behaviour, which causes a change in the way ethics are examined. According to an instrumental perspective, machines are merely tools used by humans, and as such, the decision of what is proper and inappropriate use rests with the people who use them [8]. However, what if all decisions were made by machines? While improving our lives, the developments brought about by AI are changing how people connect and prosper [12]. This must be emphasised in order to address this issue; yet, ethics is one of the aspects of human life that requires a re-examination.

AI may make it possible for technology to completely replace its creators rather than merely coexist with them, in contrast to earlier innovations that had a significant social impact [13]. AI is characterised as non-biological intelligence that embodies the technological age, the pinnacle of advancement, when software and hardware may be freely created. The majority of users of contemporary technology are impacted by AI, but society at large will surely be impacted to some extent [14]. However, these developments raise a number of societal issues, including how customers are exposed to particular products, loans, and employment opportunities, given how technology and its algorithms quietly impact our lives [15]. Even significant repetitive medical duties have been replaced by automation in sectors like education that are vulnerable to it. [16]. These are a few instances of the crucial assessments that were previously discussed, including life-or-death medical choices.

The belief that future advancements will lessen issues like poverty, natural catastrophes, conflict, and inequality justifies allowing these intrusions into our life. Therefore, the topic of discussion is our future rather than technology. This background knowledge would be sufficient to understand the ethical implications of AI, but knowing more about its history would guarantee that the issue is not being ignored.

Ethical concerns about human interaction will only get worse when AI systems in the future are able to assess their own morality. Weak or narrow AI is the term used to describe current AI technologies. Consequently, an AI system must be regarded as an object with rights comparable to or equal to those of humans, not as a machine [17]. It is crucial to take an ethical approach to AI technologies seriously and to abide by existing rules in order to prevent the misuse of the technology. This implies that it's critical to remember that ethics allow companies to capitalise on the innovative and socially acceptable prospects that artificial intelligence presents. Ethics also help companies avoid expensive mistakes. AI systems won't be widely accepted unless their results are regarded as significant and low risk. The public's use of AI technologies, their transparency regarding their operation, and the ease with which AI systems may be understood by others are all critical to their success.

3. SIGNALS OF AI BIASES

AI has the power to significantly raise people's standard of living and revolutionise a wide range of businesses. Nonetheless, bias continues to be a major barrier to the advancement and uptake of AI systems. It happens when decision-making processes contain systematic flaws that produce unfair outcomes. Inductive bias, for instance, explains the assumptions that a certain model makes. Numerous academic fields, including psychology, anthropology, law, and others, have examined the different facets of human prejudice. AI bias can come from a number of places, including human interpretation, algorithmic design, and data collection techniques. A subset of AI systems called machine learning models can identify and reinforce biased patterns in training data, producing unfair or discriminatory results.

Bias is the propensity of an AI system to render biassed conclusions, either in support of or against a particular person or group, particularly when those decisions are thought to be unjust. Human bias is attributed to technology growth, according to a number of studies [18, 19]. AI and robotic advancements are not immune to cognitive biases, which impact human decision-making in all its manifestations [20]. Understanding how human biases inadvertently infiltrate artificial intelligence design is crucial, especially considering how often marketers and consumers use humanised technology [21]. This makes sense because data may be transferred from humans to AI during the coding and programming process, which makes racism and prejudice worse. AI biases could result from this. Therefore, it is riskier for businesses to use AI algorithms now that bias in these algorithms is more well known. This is due to the fact that algorithms powered by AI are by nature less transparent and intelligible than their equivalents that do not use AI. The MYCIN decision-making process is one example of an expert system that has failed throughout AI history; it was not always straightforward or simple for humans to comprehend.

Since the actual reasoning behind the algorithm's suggestions was never disclosed, they were commonly viewed as "black-box" decisions. Because of the possibility that biases in the system's decision-making process may go unreported due to a lack of transparency, medical professionals regard the MYCIN undesirable. Even with close examination of the algorithms and data sets, AI systems like MYCIN and other AI systems might not be able to completely eliminate all unwanted biases because they are trained on historical data, which preserves biases from the past. The various biases in AI systems and their applicability according to significance scores are graphically shown in the bar chart (Figure 2). On a scale of one to ten, each form of bias is given a score; higher scores denote greater significance for AI ethics and system development. This aids in comprehending and ranking options for mitigating prejudice.

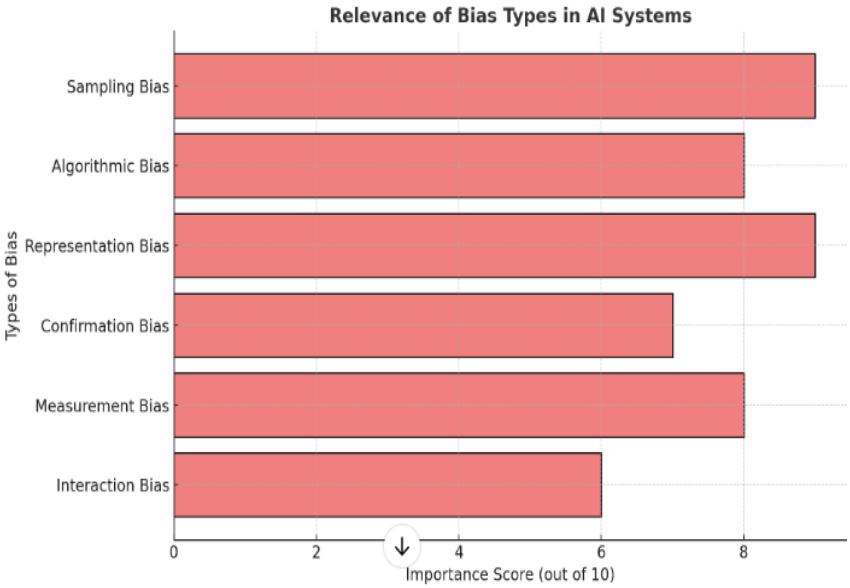

Figure 1: Relevance of Bias in AI System Type

Bias in artificial intelligence has the potential to significantly harm individuals and society. Discrimination is a major cause for concern because biassed AI systems have the potential to both exacerbate and maintain current inequities [22]. When the criminal

justice system employs biassed algorithms, people of colour are more likely to experience unjust treatment, which increases the possibility that they may be wrongfully convicted or given harsher terms [23]. AI bias may negatively affect a person's ability to obtain basic services like banking and healthcare. According to Dwork et al. [24], certain groups, such as people of colour or those from low-income backgrounds, may be under-represented in credit scoring systems due to biassed algorithms, which could make it more difficult for these groups to obtain credit. All things considered, bias in AI may have detrimental effects that are pervasive and have a big influence on people and society. To guarantee that AI systems are fair, equal, and meet the demands of all users, it is imperative to recognise and reduce bias in the technology.

4. AI'S ETHICAL CONCERNS AND IMPLEMENTATION DIFFICULTIES

All-purpose technology is changing how we live, work, and communicate. The world is expected to change at a faster rate than it is now, and artificial intelligence has made great strides in the last five to ten years. As a result, protections against the dangers AI presents to humans have been developed. An AI ethics framework is essential because it describes the benefits and drawbacks of AI technologies and establishes standards for their moral use. The outdated instrumental perspective of ethics about human usage is being replaced by AI that integrates ethical standards into its development. Ethics must now govern how robots interact with humans and other machines as they are now the ones making these choices and exhibiting these behaviours [12]. The industry and interested parties must consider important social issues and, ultimately, the question of what makes humans human in order to establish moral standards and ethical procedures for the appropriate use of AI. Despite its many positive uses, AI technology has the potential to reproduce prejudice and discrimination seen in the real world, deepening social divisions and jeopardising basic human rights and liberties.Human biases, such as racial and gender biases, must be addressed when training AI systems because these biases can be carried over from the datasets. For AI to be developed effectively and fairly, it is imperative that algorithms be trained free from these biases. [17].

All results, whether good and bad, involving machines that are seen as tools are ultimately the responsibility of humans. It implies that they are the basis for all moral conundrums and trade-offs, which simplifies the process of establishing accountability. Since AI is a cultural artefact and a result of human intelligence, its ethics differ from those of other sentient beings [25]. Commitment to using AI for the benefit of people and society at large is being urged in order to enhance freedom and well-being and manage any dangers in a responsible and acceptable manner. Developers have seen that there is currently a great chance for success in both the economy and technology. Using RAG models is one method of resolving the problem of data gathering containing biases that resulted from past events, scenarios, or circumstances during data collection. One kind of AI model architecture called retrieval-augmented generation (RAG) blends aspects of generation and retrieval in tasks involving natural language processing. RAG models improve their generation capabilities by extracting pertinent information from a big corpus of text prior to producing replies, in contrast to typical language models that only produce text based on learnt patterns. This method allows RAG models to offer more contextually relevant and accurate outputs, especially in settings requiring deep understanding or specific information retrieval. These systems are especially helpful in applications like dialogue and question-answering systems where the relevance and correctness of the content produced are crucial. Many people who believe that the current state of affairs isn't working have suggested solutions, and for a variety of reasons, ethical standards may not have a stronger influence on judgements made in the real world. claims that the difficulty of applying current ethical principles to AI development and design is one of the main reasons ethical standards have had such a limited impact on AI design [26]. In many cases, ethical education for engineers and technologists has focused on morality as a personal quality [27]. Deep learning is used in a variety of industries, including the computer and automotive sectors, and it both exacerbates and creates new societal problems that some have dubbed "exponential." The 2016 ProPublica report that exposed racial prejudice in the Compas recidivism evaluation system was among the first to garner notice [28]. Although thousands of events are known to have happened, many more are yet unknown. The following are some of the challenges in resolving AI's ethical issues, as illustrated in Figure 2:

- Discrimination against persons based on bias is without a doubt the primary problem. The most common cause of this is biases in the data, but they can also originate from the algorithm that is learning or from the feedback loop that exists between the system and the user. When training AI systems, human bias—including racial and gender prejudices—must be taken into account. Since AI systems require a large amount of data in order to learn efficiently, human selection is done for the datasets. Prejudices may be transferred to machine learning systems as they develop throughout this process. Consequently, it is essential to train algorithms without human biases [17]. To make judgements, OpenAI's GPT-3 AI model uses 175 billion parameters.

- Privacy and data protection are two crucial ethical issues that are regularly brought up. Although privacy and data protection are not the same thing, data privacy is the main privacy concern in AI ethics [29]. One way to maintain informational privacy is through data protection. AI powered by machine learning poses a number of data security risks. Training necessitates access to large datasets, which may present data protection concerns. The potential for privacy issues to develop from AI's capacity to recognise patterns—even in situations where direct access to personal data is impractical—is more intriguing and uniquely related to AI.

- Living in the digital age comes with a number of real-world problems, starting with financial ones. Of these, (un)employment is perhaps the most significant. It has long been acknowledged that AI-related technologies have the potential to cause a new wave of automation and perhaps job displacement [30]. It is certain that this would result in unemployment on a scale where the current recession and even the depression of the 1930s will seem like little jokes, as Norbert Wiener predicted that computers competing with people for occupations would have major employment consequences [31].

- The majority of ethical dilemmas in the digital age fall under the category of freedom. It is clear how parole judgements influenced or made by AI could impact a person's freedom. However, the impact of AI on freedom is more complex and all-encompassing. Our daily interactions with technology limit the range of things we can perform by providing and removing access to information. This phrase does not necessarily imply a conscious attempt to mislead or deceive; rather, it emphasises the technical mediation of our social reality and the consequence impact of this mediation. The results of a web search engine could serve as an illustration. Search engines employ AI extensively. They also organise what people can see, what they will value as a result, and how they respond to goods, services, and circumstances. This is how search engine providers generate money: they use it to make sponsored content more visible and encourage people to buy.

- Transparency was the openness-related ethical issue that attracted the most attention. Transparency is the capacity to understand how an AI system makes decisions [32]. Because most AI techniques are opaque, deep learning in particular lacks openness into the inner workings of its algorithms [32-35]. To be more precise, the black-box's decision-making process and fundamental analytical stages are opaque [32, 36]. These concerns are amplified when algorithms are trained on biased data or overlook certain demographic characteristics [32]. An AI system used in the United States to predict future recidivism rates assigned higher risk ratings to African Americans with fewer criminal records than to White Americans with more criminal records [37]. In the long run, transparency will be essential to the public, patient, and physician

confidence in AI in healthcare [38-40]. Transparency is frequently questioned when dealing with sensitive data like medical records [41].

- Explainability is a challenge that is closely related to transparency, the primary ethical concern. Similar to transparency, explainability is associated with the opaque character of machine learning and artificial intelligence algorithms, which makes it difficult to comprehend and elucidate the relationship between input data and outputs. To properly grasp this, we can use DENDRAL's decision-making mechanism as an example. Similar to modern AI systems, this method could not have been completely obvious or intelligible to human users. The algorithm's suggestions might have been perceived as "black-box" decisions because the reasoning behind them was not disclosed. This lack of transparency could raise concerns about accountability, equity, and trust in the system's output. Consequently, incorporating explainability can aid in preserving or boosting consumer trust. As a result, adding Explainability can support preserving or boosting user confidence in AI systems [42]. Confidence in the AI system's use in the medical area is increased by how well it explains things [42, 43]. Medical personnel will struggle to maintain the legitimacy of the system and promote confidence in a judgement if they are unable to properly explain it to patients or other medical professionals [43].

- The issue of defining fairness itself is another. Individuals and groups may have different ideas on what fairness is, and this definition may evolve over time [24]. This implies that it would be difficult to create AI systems that are seen fairly by all parties involved.

- Autonomy is one of the most prevalent traits of AI systems that behave socially or resemble humans. It also has to do with the ability to change its own behaviour, but only under specific circumstances. Despite its uneven use, the term “objective” is typically linked to the capacity to avoid interfering, which is the best approach to conceptualise control [44]. Even though AI seems to be making decisions on its own, the results are clearly a reflection of external factors, such as inputs from the surrounding environment and human-developed code [45].

- The legal and regulatory frameworks for the regulation of AI vary among the many countries and regions, reflecting historical, political, and cultural contexts. While some countries may adopt a more laissez-faire approach or rely on industry self-regulation, others may enact strict legislation and monitoring protocols to combat AI bias and ethics. Cultural views on individual rights, regulation, and government participation impact the development and application of legislation relating to artificial intelligence. Global initiatives to address AI bias and ethics are influenced by international collaboration, alliances, and agreements. Global standards and best practices for responsible AI development and deployment are influenced by cultural variations in human rights norms, ethical principles, and AI governance.

Furthermore, given the issues and challenges already discussed, the implementation of biassed and immoral AI systems may erode public trust in technology, which could result in a decline in the adoption or even rejection of new developments. People may not be able to take advantage of AI if they lack confidence in it or perceive it as a discriminatory instrument, which could have a negative effect on the economy and society. To avoid negative consequences and encourage positive outcomes from societal ramifications like economic inequality, job relocation, and ethical governance, a proactive approach is required. To prepare the workforce for the evolving nature of work due to AI, reskilling initiatives and lifelong learning courses are essential [46]. International cooperation and ethical governance are necessary to establish consistent standards and promote moral AI practices. We must encourage corporate cooperation, advocate for strong legislative frameworks, and participate in continuing education if we want to guarantee that AI research is conducted ethically in the future. The development of AI systems that are consistent with societal values and advance society requires the incorporation of various points of view, stakeholder participation, and an unwavering dedication to moral principles. It's also critical to have conversations about how AI will affect society and how people can be empowered to participate ethically. The relevance scores given to different ethical issues in AI are displayed in the bar chart (Figure 3). On a scale of one to ten, each issue is assessed according to its importance in relation to the development and use of AI. Along with other urgent issues like prejudice, privacy, and justice, the chart emphasises how important openness and explainability are.

Building on these ideas, it is crucial to put in place safeguards that guarantee AI systems behave morally and fairly. Building public and stakeholder trust in AI development and implementation requires transparency and accountability. By putting Explainable AI (XAI) frameworks into practice, biases can be reduced and comprehension improved. Open data exchange encourages teamwork and guarantees that different viewpoints are taken into account. Furthermore, it is crucial to have ongoing discussions on the societal

effects of AI and to enable people to engage in ethical behaviour. By adopting these strategies, we can properly negotiate the challenges of AI adoption and make sure it upholds social values and fosters beneficial results in a variety of societies.

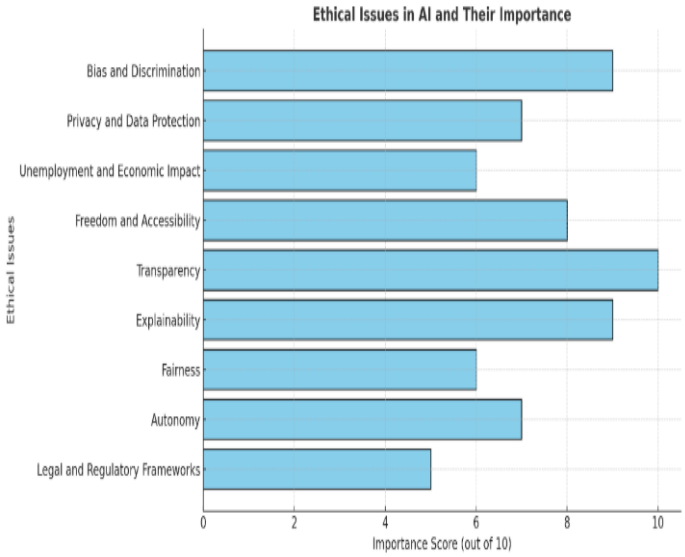

Figure 3: Ethical Issues in AI & their Importance

5. CONCLUSION

AI and social science cooperation is a "fruitful arrangement." The development of socio-beneficial systems and even AI approaches has been aided by the social sciences, while sociological study has also benefited from AI research. A thorough grasp of artificial intelligence may be possible when combined with other academic fields like sociology and philosophy. Sociological research on AI contributes to the body of knowledge on the broader societal ramifications and is crucial for developers and designers who are attempting to foresee unforeseen outcomes.

AI practice is intrinsically interdisciplinary and can profit from a variety of viewpoints on technological inputs, as we have been highlighting in our work. Fairness, impartiality, and correctness are examples of technical considerations that shouldn't supersede other kinds of knowledge that could help bring about social change. The largest social and technological obstacles to AI's advancement are prejudice and bigotry, and they have existed throughout human history. Intelligent robots replicate and amplify pre-existing inequities since they are initially taught on biassed datasets. By automating and exacerbating preexisting inequities, they act as magnifying glasses. Because they have so many options, especially when it comes to validity, machines are frequently referred to as "black boxes."

AI research prioritises inclusivity in addition to efficiency and accuracy since it bridges the gap between technological and social studies. AI should be inclusive in addition to being responsible. High-quality data that takes into account variables like gender, ethnicity, educational attainment, and any other pertinent social or economic distinctions that can occasionally make it impossible to mine inequality should serve as its foundation. Good data considers each individual's privacy as well as the goals and concerns of society at large. Data facilitates the stringent control of models and algorithms through enforced standards. AI systems may eventually take on societal responsibilities, necessitating the use of novel design concepts like transparency and predictability. Highly generalised AI systems might not function as intended in predicted situations, which would call for the development of fictitious moral conundrums and new security measures. Though they would differ greatly from current humans and be governed by other laws, AI systems with sophisticated mental states—or the right kinds of states—would have moral standing and some might even be regarded as persons.

Last but not least, most moral and ethical dilemmas surrounding AI are still unidentified and in its infancy. AI ethics is more complex than "right or wrong," "virtue and vice," or "good or bad." It is undoubtedly not an issue that can be resolved by a small group of

people. But there are serious ethical and moral problems with AI that need to be addressed right away. This study aims to demonstrate how important it is for different stakeholder groups to take AI agents' ethics and morals into account. We will better understand and apply human ethics, improve our relationships with AI, and get a deeper understanding of current ethical notions as we define AI ethics.

REFERENCES

[1] M. Boden, "AI: Its nature and future", Oxford University Press, 2016.

[2] L. Floridi, "AI and its new winter: From myths to realities", Philosophy & Technology, 2020.

[3] R. Bendici, "Rise of the machines", Univ. Bus, Vol. 21(10), pp. 53–54, 2018.

[4] M. Iman, H.R. Arabnia, R. and M. Branchinst, "Pathways to artificial general intelligence: a brief overview of developments and ethical issues via artificial intelligence, machine learning, deep learning, and data science", paper presented at International Conference on Artificial Intelligence (ICAI), Las Vegas, Nevada, USA, 2020.

[5] A. Datta, M.C. Tschantz, and A. Datta, "Automated experiments on ad privacy settings", Privacy Enhancing Technologies, Vol. (1), pp. 92–112, 2015.

[6] Fei-Fei Li, Chief-Scientist for AI at Google and Professor at Stanford.

[7] F. Karimi, M. Génois, C. Wagner, P. Singer, & M. Strohmaier, "Homophily influences ranking of minorities in social networks", Scientific Reports, 2018.

[8] Z. Tufekci, "Big questions for social media big data: Representativeness, validity and other methodological pitfalls", In ICWSM, The AAAI Press, 2014.

[9] L. Introna, & D. Wood, "Picturing algorithmic surveillance: The politics of facial recognition systems", Surveillance & Society, Vol. 2(2/3), pp.177–198, 2004.

[10] D. Faggella, "Everyday examples of artificial intelligence and machine learning", Emerj, Boston, MA, 2020.

[11] Z. Liu, "Sociological perspectives on artificial intelligence: A typological reading", Sociology Compass, Vol. 15, e12851, 2021.

[12] M. Anderson, and S.L. Anderson, "Machine ethics: creating an ethical intelligent agent", AI Mag, Vol. 28, pp. 15–26, 2007.

[13] E. Brynjolfsson, and A. McAfee, "The Second Machine Age: Work, Progress, and Prosperity in A Time of Brilliant Technologies", New York: W. W. Norton, 2016.

[14] K. Franklin, and W. M. Ramsey, "Introduction" in the Cambridge Handbook of Artificial Intelligence, ed. K. Frankish and W. Ramsey (Cambridge, United Kingdom: Cambridge University Press), 2014.

[15] K. Martin, "Ethical implications and accountability of algorithms", J. Bus. Ethics 160, pp. 835–850, 2019.

[16] J. Hooker, and T. W. Kim, "Ethical implications of the fourth industrial revolution for business and society", Vol. 3, pp. 35–63, 2019.doi: 10.1108/ s2514-175920190000003002

[17] K. Siau, W. Wang, "Ethical and moral issues with AI - a case study on healthcare robots, emergent research forum (ERF)", 2019. https://www.researchgate.net/publication/325934375_Ethical_and_Moral_Issues_with_AI.

[18] M.M.A. de Graaf, S.B. Allouch, "The influence of prior expectations of a robot's lifelikeness on users' intentions to treat a zoomorphic robot as a companion", International Journal of Social Robotics, Vol. 9 (1), pp. 17-32, 2017.

[19] D. Kuchenbrandt, F. Eyssel, S. Bobinger, and M. Neufeld, "When a robot's group membership matters" International Journal of Social Robotics, Vol. 5 (3), pp. 409-417, 2013. 10.1007/s12369-013-0197-8

[20] K. Letheren, R. Russell-Bennett, L. Whittaker, "Black, white or grey magic? Our future with artificial intelligence", Journal of Marketing Management, Vol. 36 (3–4), pp. 216-232, 2020.

[21] A. Jobin, M. Ienca, E. Vayena, "The global landscape of AI ethics guidelines", Nature Machine Intelligence, Vol.1, pp. 389-399, 2019.

[22] L. Sweeney, "Discrimination in online ad delivery", Communications of the ACM, Vol. 56(5), pp. 44-54, 2013.

[23] J. Angwin, J. Larson, S. Mattu, and L. Kirchner, "Machine bias. In Ethics of data and analytics", Auerbach Publications, pp. 254-264, 2016.

[24] C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. Zemel, "Fairness through awareness", Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, pp. 214-226, 2012.

[25] J. J. Bryson, "Robots should be slaves", Artif. Models Nat, Intell, Vol. 8, pp. 63–74, 2010.

[26] B. Mittelstadt, "Principles alone cannot guarantee ethical AI", Perspectives. Nat. Mach. Intell, Vol. 1, pp. 501–507, 2019.

[27] R. Howard, "Ethical distinctions for building your ethical code," in Next Generation Ethics: Engineering a Better Society, ed. A. Abbas (Cambridge: Cambridge University Press), pp. 7–16, 2019.

[28] J. Larson, S. Mattu, L. Kirchner and J. Angwin, "How We Analyzed the COMPAS Recidivism Algorithm", ProPublica, 2016.

[29] G. Buttarelli, "Choose humanity: putting dignity back into digital", In: Speech at 40th international conference of data protection and privacy commissioners, Brussels, 2018.

[30] H. Collins, "Artificial experts: social knowledge and intelligent systems" MIT Press, Cambridge, MA, 1990.

[31] N. Wiener, "The human use of human beings", Doubleday, New York, pp. 164, 1954.

[32] T.A. D'antonoli, "Ethical Considerations for Artificial Intelligence: An Overview of the Current Radiology Landscape", Diagn. Interv. Radiol, Vol. 26, pp. 504, 2020.

[33] G. Currie, K.E. Hawk, E.M. Rohren, "Ethical Principles for the Application of Artificial Intelligence (AI) in Nuclear Medicine", Eur. J. Nucl. Med. Mol. Imaging 2020, Vol. 47, pp. 748–752, 2020.

[34] M.E. Fenech, and O. Buston, "AI in Cardiac Imaging: A UK-Based Perspective on Addressing the Ethical, Social, and Political Challenges", Front. Cardiovasc. Med., Vol. 7, pp.54, 2020.

[35] D. Sen, R. Chakrabarti, S. Chatterjee, D. S. Grewal, and K. Manrai, "Artificial Intelligence and the Radiologist: The Future in the Armed Forces Medical Services, J. R. Army Med. Corp , Vol. 166, pp. 254–256, 2020.

[36] M. Ananny, and K. Crawford, "Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability", new media & society, Vol. 20(3), pp. 973-989, 2018.

[37] J.K. Paulus and D.M. Kent, "Predictably Unequal: Understanding and Addressing Concerns That Algorithmic Clinical Prediction May Increase Health Disparities", NPJ Digit. Med., Vol. 3, pp.99, 2020.

[38] G.C.K. Yew, "Trust in and Ethical Design of Care bots: The Case for Ethics of Care", Int. J. Soc. Robot, Vol... 13, pp. 629–645, 2021.

[39] O. Akinrinola, C.C. Okoye, O.C. Ofodile, and C.E. Ugochukwu, "Navigating and reviewing ethical dilemmas in AI development: Strategies for transparency, fairness, and accountability", GSC Advanced Research and Reviews, Vol. 18(03), pp. 050–058, 2024.

[40] B. Seroussi, K. F. Hollis,and L. F. Soualmia, "Transparency of Health Informatics Processes as the Condition of Healthcare Professionals' and Patients' Trust and Adoption: The Rise of Ethical Requirements", Yearb. Med. Inform, Vol. 29, pp. 7–10, 2020.

[41] T. Wangmo, M. Lipps, R.W. Kressig, and M. Ienca, "Ethical Concerns with the Use of Intelligent Assistive Technology: Findings from a Qualitative Study with Professional Stakeholders", BMC Med. Ethics, Vol. 20, pp. 98, 2019.

[42] C. G. Walsh, B. Chaudhry, P. Dua, K. W. Goodman, B. Kaplan, R. Kavuluru, A. Solomonides ,and V. Subbian, "Stigma, Biomarkers, and Algorithmic Bias: Recommendations for Precision Behavioral Health with Artificial Intelligence", JAMIA Open 2021,Vol. 3, pp. 9–15, 2021.

[43] A. Vellido, "A. Societal Issues Concerning the Application of Artificial Intelligence in Medicine", Kidney Dis., Vol. 5, pp. 11–17, 2019.

[44] J. M. Beer, A. D. Fisk, and W. A. Rogers, "Toward a framework for levels of robot autonomy in human-robot interaction" J. Human-Robot Interact", Vol. 3, pp. 74–99, 2014.

[45] A. Etzioni, O. Etzioni, "Incorporating ethics into artificial intelligence", J Ethics, Vol. 21, pp. 403–418, 2017.

[46] S. Tatineni, "Ethical Considerations in AI and Data Science: Bias, Fairness, and Accountability", International Journal of Information Technology & Management Information System (IJITMIS), Vol. 10(1), pp. 11-20, 2019.