iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

ETHICAL CHALLENGES OF AI INDECISION MAKING: A SYSTEMATIC REVIEW

Ravindra Prabad Suraweera
suraweera.rp@gmail.com

ABSTRACT
Nowadays, Artificial Intelligence (AI) has played an important role in decision making in different fields such as healthcare, finance and public administration. However, AI brings a lot of efficiency and scale at the same time and that implies a ton of ethical challenges as well. The goal of this systematic review is to pinpoint and assess the ethical issues with AI placed decisions. Since our PRISMA guidelines we have chosen only peer reviewed articles and trusted sources from 2015 – 2025. Algorithmic bias, lack of transparency, accountability, data privacy and human autonomy are among the key ethical challenges. These findings highlight the importance of developing strong ethical frameworks, interdisciplinary cooperation and constant oversight when deploying AI.

Keywords:
Artificial Intelligence, Ethical Challenges, Decision Making, Algorithmic Bias, Transparency, Accountability, Data Privacy, Human Autonomy

1. INTRODUCTION

Artificial Intelligence or AI is changing the world at a rapid pace and altering the way by which decisions are being made in various fields of life. But as an engineer of decision support systems, these systems are being used to assist or even fully automate decision making in sectors such as the healthcare, education, finance, law and public services. Russell & Norvig (2021) indicate that these are systems which can process large amounts of data quickly, identify complex patterns and predict with high accuracy. For instance, AI may assist doctors with the diagnosis of diseases, financial institutions with processing of loan applications and law enforcement in prediction of crime hotspots. This brings great potential to enhance efficiency, accuracy and results in diverse areas (Vinuesa et al. 2020).

Now, as machines are increasingly working hand in hand with human beings in making decisions which directly impact people's lives, serious ethical issues have surfaced. Algorithmic bias is the major concern of AI systems that unintentionally prefer one group over another due to biased data sets or defects in the design. In other words, it can cause an unfair treatment as it happens in areas like, hiring, lending or even in criminal justice (O'Neil, 2016, Obermeyer et al., 2019). For instance, if an AI tool is taught with information from the past that signifies unfair treatment, it can continue to discriminate against minorities and women.

Another is transparency. One thing is that many of today's AI systems run as 'black boxes' – this refers to those which employ deep learning, so it's hard to understand how they come to their decisions. The problem with this lack of explain ability is particularly important for critical applications such as healthcare or legal judgements where people have a right to know why decisions which ultimately affect them, are made (Burrell, 2016). However, if AI systems falter to make errors, there often is uncertainty of who should bear the responsibility of it; the developers, users or the system (Morley et al., 2020).

A major ethical challenge is data privacy as well. Most AI systems need large quantities of personal data to work well. The collection or use of this data without consent can result in numerous serious violations of privacy (Floridi et al. 2018). Also, human autonomy is being lost in the midst of AI taking more and more decisions on behalf of humans. If the users rely too heavily on the AI system, there is a chance that they lose their responsibility or critical thinking skills since they simply reliterate the system's recommendations (Calo, 2017).

This leads to issues of concern which necessitate looking into the ethical aspect of use of AI to make decisions. Although there are many previous studies discussing these issues in isolation, there is a need for a systematic review

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

whereby it brings together existing research, discusses shared topics and points out knowledge gaps. This could be a valuable resource for those who are in the business of designing, implementing (as has been argued) and how AI should be used responsibly (as has also been argued) because such a review can provide policymakers, technology developers and end users with insights into how to make informed decisions. (Jobin et al., 2019) The ethical challenges of AI in decision making has not been adequately researched for lack of a good literature. This paper attempts to fill this gap by reviewing and analyzing the existing literature. This paper will identify key ethical concerns, examine how this challenges are being addressed and suggest recommendations for future practice and research. The paper, through this systematic review, hopes to assist in the construction of ethical and trustworthy AI systems that respect human values and ensure fair, transparency and accountability.

2. LITERATURE REVIEW

Artificial Intelligence (AI) today is known as a fundamental technique supporting and automating decision making in sectors like healthcare, criminal justice, finance, education and in human resources (Russell & Norvig, 2020). AI development is further gaining moral issues, especially regarding the way the systems decide or impact individuals and society. In this literature review, the main ethical challenges addressed in academic and policy based studies are discussed in five areas: bias and fairness, transparency and explainability, accountability and responsibility, privacy and data security and human autonomy.

2.1 Bias and Fairness in AI Decision-Making

Algorithmic bias is one of the most talked about ethical challenges in AI. The data that we put into the training of an AI system are reflective of past patterns, including social inequalities. These systems, as a result, can just replicate or amplify existing biases (Barocas and Selbst, 2016). As an example, Buolamwini and Gebru (2018) showed that commercial facial recognition systems have strong racial and gender biases where they are trained worse on dark skinned and female faces than light skinned male faces.

It is not just technical but social and ethical problem, too. According to Noble (2018), algorithmic bias is a display of systemic racism and sexism on the part of AI applications made for hiring, policing or lending decisions etc. Biases of this kind, if present, may result in treatment of individuals and social groups that is unfair, in violation of principles of justice and equality (Binns, 2018). Mehrabi et al. (2021) suggest that this bias mitigation also needs to start from the data collection and curation point together with regular audits and a diverse design team.

2.2 Transparency and Explainability

The latter concept also captures two sub categories, transparency and explainability which describes how comprehensible and understandable an AI's processes are to a human observer (Doshi-Velez & Kim, 2017). It is important to know these concepts, for it helps in developing trust and making decisions of the AI to be fair and transparent.

Often, black box models especially deep learning models are not explainable. The problem is that this creates trouble in sectors like healthcare or criminal justice where decisions have to make sense (Lipton, 2018). For instance, even when an AI system refuses a loan or suggests a medical procedure, the stakeholders should know the reasons we got to that outcome (Guidotti et al., 2018).

This is tackled by researchers, for example in the context of Explainable AI (XAI). Gunning et al. (2019) states that XAI is the process of making AI models interpretable while maintaining performance. However, the question still lies in how to balance the accuracy and interpretability especially when the user wants a simple explanation for complex decisions (Miller, 2019).

2.3 Accountability and Responsibility

One of the ethical issues around AI is determining whose job it is when AI systems do harm. Most common conceptions of responsibility entail that a human can be culpable for a decision. Autonomous systems help diffuse responsibility when decisions are made or when they are made in such a way (Calo, 2015). As a result, we have the so called responsibility gap (Matthias 2004).

In the case of AI driven decisions, there can be several actors who are: developers, data scientists, users or organizations. This creates legal and ethical accountability complications (Santoni de Sio & Van den Hoven 2018).

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

For example, if a self-driving car accident happened then it's difficult to figure out who is fault, the programmer, the manufacturer, the user, the AI system itself.

Some scholars, for example, suggest "meaningful human control" (Nyholm, 2018)—humans are involved with decision-making, especially when applications are at high-risk. In other words, other researchers theorize that regulatory frameworks should be created to allocate responsibility when AI developers fail to follow ethical standards (Cath, 2018).

2.4 Privacy and Data protection

Learning and functioning of AI systems are dependent on large amounts of data. Despite that, collection, storing and use of personal data are of a major privacy concern (Zuboff, 2019) AI systems use the data the user provides health records, financial and social media behavior — and this data is extremely sensitive and if not protected properly, will be the data to be misused.

To counter this, regulations such as the General Data Protection Regulation (GDPR) in the European Union have handed over rights to individuals with their data and laid stringent provisions to govern the way businesses process data (Voigt & Von dem Bussche, 2017). It also follows from Article 22 of the GDPR that people have the right not to be subject to automated decision making without human intervention.

Such regulations are in place, yet there are problems. Take for example, data breaches, it increases surveillance and AI face profiling purposes for targeted advertising or if policing (Eubanks, 2018). The solution comprises privacy by design and data minimization practices, stronger regulatory enforcement and public education.

2.5 Human Autonomy and Moral Agency

The reliance on AI systems presents another ethical issue that is the potential human autonomy erosion. Users have a tendency known as 'automation bias'; when AI systems make or recommend decisions they tend to defer to them (Goddard et al., 2012). In particular, this is dangerous for areas that are crucial to life like aviation, medicine or military applications.

Shneiderman (2020) further warns that over automation might result to loss of human skills and moral judgment. Human decision – making should not be replaced by AI, but rather augmented by it, so that people can reason according to their moral or ethical values. In democratic societies, human values, accountability and consent have to be preserved and this is where this is particularly important.

Other researchers suggest a model known as "human in the loop", whereby AI is a support tool that never cuts humans out of the equation (Rahwan, 2018). In contrast, others propose that AI systems should be designed with the inclusion of the respect of human dignity, agency and freedom of choice (Jobin et al., 2019).

2.6 Ethical Frameworks and Governance

However, to meet these challenges, international organisations, government and institutional agencies have proposed a number of ethical frameworks and guidelines. For instance, in the case of AI design, the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (Perez Alvarez, Havens, & Winfield, 2017) has defined the following as ethical principles, viz., transparency, accountability and alignment with human values.

For instance, High Level Expert Group on AI (2019) of European Commission put forward seven principles for trustworthy AI that involve human agency and oversight, technical robustness, privacy, transparency, fairness, ethical values and societal well being and accountability and governance. They use these frameworks trying to smoothen the way, for AI developers and users to create ethical and socially beneficial technologies.

Nevertheless, critics claim that several of these guidelines are unenforceable (Hagendorff, 2020) and overly broad. Some are calling for a stronger legal basis, ethical training for developers and inclusive governance structures where different voices, including the voices of marginalized groups affected by AI decisions, can be part of the process.

3. METHODOLOGY

The method of systematic review was employed to examine, analyze and summarize the main ethical challenges of Artificial Intelligence (AI) in the context of decision making. In this section, this process of how the literature search has been carried out, how the studies have been selected and how the data analysis has been done will be explained.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

3.1 Search Strategy

Structured and comprehensive search using multiple academic databases was done to gather relevant literature. The selected databases were Google Scholar, IEEE Xplore, PubMed, Scopus and Web of Science. The reason for choosing these platforms is that they cover all of AI, ethics and interdisciplinary research.

A search of studies was conducted from January 2015 through June 2020, thus ensuring the current understanding of recent developments in the field of AI ethics. In order to maximize the results we used the following keywords and phrases in all possible combinations using Boolean operator ("AND", "OR") and so on.

- "Artificial Intelligence"
- "Ethics"
- "Decision Making"
- "Algorithmic Bias"
- "AI Transparency"
- "Accountability in AI"
- "Data Privacy"
- "Human Autonomy and AI"

Example search strings included:

- "Artificial Intelligence" AND "Ethics" AND "Decision Making"
- "AI" AND "Algorithmic Bias" AND "AI Transparency"
- "AI" AND "Ethical Challenges" AND "Autonomy"

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) was followed to find our results (Page et al., 2021). Tool reference management as Zotero, Endnote were used to remove duplicate articles.

3.2 Inclusion and Exclusion Criteria

To be able to select the quality and relevant literature the following inclusion and exclusion criteria were mentioned:

Inclusion Criteria

- In the case of peer reviewed journal articles, systematic reviews and credible institutional reports (i.e., OECD, UNESCO, EU AI policy paper reports) as well as primary data analysis in the context of the development of a theory, design or organizational statements.
- The publications that focus on ethical issues in AI driven decision making in particular.
- English articles written and published.
- Were studies that at least considered at least one core ethical issue (bias, transparency, accountability, privacy or autonomy).

Exclusion Criteria

- Non AI ethics or decision-making focus articles.
- Studies covering AI solely in terms of its purely technical or engineering counterpart, adopting AI without ethical values accounted for.
- Non-English publications.
- The most common type of source is an editorial, opinion piece, blog post or any other source that is not a peer-reviewed source.

The filtering process from the study ensured a guarantee of high-quality and relevant literature for analysis, thereby adding credibility to the review.

3.3 Data Extraction and Analysis

After identification and screening of the articles, a standardized data extraction form was used to get information systematically and organize it. The following were recorded from each selected article that was reviewed:

- Author(s) and year of publication
- Objectives of the study
- Research methodology
- Key ethical Challenges discussed

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

- Proposed solutions or frameworks
- Conclusions and recommendations

To make it easy to compare across studies I organized this information in an Excel spreadsheet.
A thematic analysis of Braun and Clarke (2006) was used to analyze the collected data. The reason for choosing thematic analysis is that it is a flexible method that identifies, analyse and report patterns (themes) within data. In particular, this was very useful for exploring numerous ethical problems in this mission across a broad spectrum of decision situations.
The process involve following steps:

1. Reading and re-reading through the selected articles in order to familiarize with the data.
2. Then, I coded content line by line to have identifiers indicating the references about ethical challenges.
3. Initial themes were generated like algorithmic bias, transparency, accountability, privacy and autonomy.
4. A review and refitting of the themes to make sure these themes reflected the data that had been collected.
5. Creating a structured framing of its results by defining and naming themes.

An ethical theme was assigned to each article which was organized under one or more of the five major ethical themes that were in the literature. These were:

- Algorithmic Bias and Fairness
- Lack of Transparency and Explainability
- Accountability and Responsibility
- Data Privacy and Consent
- Autonomy and Human Oversight

In the course of the analysis, special consideration was given to how various sectors (healthcare, finance, criminal justice, etc.) addressed ethical challenges, as well as how various regions or authors, proposed solutions (Jobin, Ienca and Vayena, 2019).

3.4 Quality Assessment

A modified version of the CASP (Critical Appraisal Skills Programme) checklist was used to ensure the study outcome met quality standards set for the studies to be included. Articles were assessed for:

- Clarity of research questions
- Appropriateness of methodology
- Ethical consideration in data collection
- Depth of ethical analysis

Only these articles were retained if they had scored satisfactorily on these measures.

3.5 Limitations of the Methodology

Nevertheless, it was difficult to be 100 percent thorough and unbiased and this methodology did present a few shortcomings.

- It may have limited the language to only English thus excluding relevant international research.
- It may be that inclusion of Google Scholar caused some gray literature to be viewed, but only peer reviewed content was selected.
- Further, while focusing on publications post 2015, it may have eliminated discussions of earlier foundational research, but it allowed the focus to be on current developments.

Even with these limitations, the methodology served as a well-defined and transparent means of taking a structured and transparent look at the current set of ethical concerns associated with AI decision making.

4. RESULTS

4.1 Study Selection

The literature search identified a total of 1200 articles across five major databases. After removing duplicate entries, the number was reduced to 920. These articles were then screened based on titles and abstracts. After applying the

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

inclusion and exclusion criteria, 85 articles were selected for a full-text review. Upon further assessment, 50 articles met all the criteria and were included in final systematic review.
These articles covered a variety of fields including healthcare, finance, criminal justice, education and public administration with a focus on ethical issues in AI based decision making systems.

4.2 Identified Ethical Challenges

The thematic analysis of the 50 selected articles revealed five primary ethical challenges associated with AI in decision making:

- Algorithmic Bias and Fairness
- Transparency and Explainability
- Accountability and responsibility
- Data privacy and Security
- Erosion and Human Autonomy

Each challenge was addressed in multiple studies and emerged as a recurring theme across various decision making contexts. These themes are discussed in detail in the next section.

5. DISCUSSION

5.1 Algorithmic Bias and Fairness

Algorithmic bias was cited as one of the most common problems. AI bias happens when systems make unfair decisions based upon things that are not really there in the training data or model. For instance, biased datasets can reflect present society inequalities and, in case they are not adjusted, AI can exacerbate such biases and potentially reproduce them. Especially so in areas like hiring, lending or criminal justice (RSM Global, 2024).
However, many studies revealed that AI recruitment tools rather preferred male candidates instead of female candidates or had a tendency to disregard minority groups. We encourage researchers and developers to use diversity in training data, bias detection tools and fairness aware machine learning models in an effort to reduce bias. And there are regular audits and ethical reviews that can also detect and address potential biases before systems are deployed.

5.2 Transparency and Explainability

AI lacks transparency, especially when it comes to deep learning networks, but this is a big challenge. However, these systems are often “black boxes,” and users cannot so easily understand why a certain conclusion was reached through the AI. However when AI lacks explainability, it’s difficult for stakeholders to trust or question AI decisions.
This problem is something which is being addressed in the field of Explainable AI (XAI) i.e., where the models offer clear, understandable reasoning for their decisions. Where to surface the risk, reasoning and limitations of AI recommendations in order to help a user such as a doctor, judge or financial analyst, better understand what they are being recommended. In addition, Transparent AI conveys trust and promotes more informed decision making by humans.

5.3 Accountability and Responsibility

It is not easy to ascertain who is responsible when harm arises or AI systems make the wrong decisions. In traditional decision making, the responsibility is on a human. Yet in the world of AI, developers, users, AI all produces an overlap of responsibility.
For instance, if a self-driving car causes an accident or an AI system unfairly denies someone a loan, it is not clear who is at fault, the programmer, the company or more seriously the machine itself. Numerous scholars demand clear legal and ethical frameworks for assignment of responsibility. Additionally bringing in human-in-the-loop decision making, where humans are always there in seeing these critical, large impacts of AI.

5.4 Data Privacy and Security

Large volume of personal data that AI systems depend on is a huge point of privacy concern. Some of the major risks are related to data breaches, unauthorized access or misuse of information. For example, in healthcare, patient data should be held in confidence but AI models trained on such data may by mistake leak sensitive data.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

To solve this, one needs to comply with the data protection laws such as the General Data Protection Regulation (GDPR). Data anonymization, encryption and user consent protocols should be used by developers to keep privacy. Even more, good cybersecurity practices can be sustained and data collection can be minimized to only needed data.

5.5 Erosion of Human Autonomy

Yet another key ethical concern is humans loss of autonomy when they hand over too much of their control to AI systems. In other cases, like medicine or criminal justice, professionals may end up relying much on the recommendations of AI that may not be accurate. The phenomenon is called automation bias in which people trust automatic systems over their judgment. Such dependency may lead to reduced critical thinking and thus generating unethical or even harmful outcomes. In the second place, it may result in a scenario where people have a lesser role to play in decisions that impact their lives directly. Otherwise, there needs to be a balance of AI assistance to humans and human oversight to stop them. In life altering situations people should be empowered to question or override AI decision when needed.

6.RECOMMENDATIONS AND FUTURE DIRECTIONS

It is difficult to address the ethical challenges of Artificial Intelligence (AI) in decision making. This is an interdisciplinary work that involves cooperation of different stakeholders, including researchers, developers, policymakers, organizations and the public. The next section features the following recommendations on practical steps on how to ensure the development and use of AI technologies are in compliance with ethical principles to promote fairness in the use of AI technologies, transparency and accountability.

6.1 Ethical Framework Development

Inclusion of strong ethical frameworks is one of the most important steps in guaranteeing ethical AI. Early in the AI development process these frameworks must be created and must be maintained throughout a system's lifecycle. An ethical framework should in general be concerned with key principles like fairness, accountability, transparency and respect for the autonomy of human beings.

If these principles are applied during the design of AI systems at early stages, they can be used to pinpoint potential ethical risks so they can be treated before becoming serious rebellion. As an example, for instance, the algorithms need to be tested for biases prior to being deployed and fairness metrics should be used to assess outcomes. Moreover, ethical frameworks should be updated on a regular basis as the technology and society evolves.

A great example is already being practiced by many leading organizations who are already using ethical guidelines. Having a model, for instance, the "Ethics Guidelines for Trustworthy AI" published by European Commission (2019) that addresses ethical issues mitigating against human agency, technical robustness and societal wellbeing. This way, following such standards will help organizations ensure their AI solutions not only works but the way it works is ethically responsible.

6.2 Interdisciplinary Collaboration

AI is a field that involves many (human) areas, not only technical one and that will affect the law, society, economics, healthcare, education, etc. As a result, ethical AI cannot be constructed by computer scientists. A wide range of professionals can provide input in successful development and deployment of ethical AI.

Technologists, ethicists, legal experts, policy makers and even professionals from within the industry are able to understand AI's effects holistically when genius comes together. For instance, ethicists can detect risks missed by engineers and legal experts can assist the AI to respect the laws and prioritise the rights of individuals.

In addition to that, having people from different cultural and social backgrounds involved in developing AI systems helps in reducing the cultural biases and making them more inclusive. Put simply, it is imperative that interdisciplinary collaboration be used in designing AI systems that reflect on these ethical, legal and societal expectations.

6.3 Regulatory Oversight and Standards

For a major part, ethical guidelines are useful but legal regulations are the ones we need to enforce ethical practices, particularly in high risk AI applications like healthcare, finance or criminal justice. Clear rules and standards that should govern AI usage should be created by governments and international organizations.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

They should have to make it transparent, accountable and auditable. For example, if an AI system, say, for people's job applications or loan approvals applies to people's rights, then it should have to explain how it made its decisions. It should also provide systems by which individuals can appeal or challenge decisions made by AI and thus keep DAI accountable to humans.

Useful models such as the General Data Protection Regulation (GDPR) in Europe, require organizations to have informed consent and limit what personal data is used. Moreover, similar regulations should be extended to algorithmic decision making to preserve fairness and to privacy.

Also, industry wide standards for AI testing, validation and certification should be developed as well. This would create consistency and establish some kind of trust with users and the public.

6.5 Public Engagement and Education

Social media feeds, as well as healthcare decisions, are becoming more and more influenced by AI. But still there are many people who do not understand how it works and how it changes them. For a society that can respond to AI's (ethically challenging) use, public engagement and education are needed.

Along with that, governments and universities need to conduct awareness campaigns, along with community discussions, online courses and offline workshops on AI ethics. So individuals will know their rights, be able to recognize potential harms and be able to advocate for AI that is fair and ethical.

School and university curriculums should also integrate education so that the next generation is ready to manage AI technologies at its best. We can ensure that people have knowledge to make decisions about AI, in a democratic way and in a way that matches the values of the society that it works for.

7.CONCLUSION

Modern decision-making processes have been taken to a powerful tool in the hand of modernists by Artificial Intelligence. Diseases are diagnosed, banks approve loans and courts make their recommendations all via AI. But such rapid growth is tied to serious ethical challenges which cannot be ignored.

Five major ethical issues are summarized from the literature in this systematic review through: algorithmic bias (e.g., potential for unfair treatment, exclusion), lack of transparency (e.g., predictive analytics mechanisms and outcomes obscured), accountability dilemmas (e.g., responsibility unknown), data privacy (e.g., fair but involuntary information use) and effect on human autonomy (e.g., influence on individual decision making). If these challenges are not addressed, this may result in unfair treatment of individuals, infringements of rights or lack of public trust in the functioning of the AI system.

In order to overcome these challenges, they need to be approached from multiple dimensions. In the first instance, ethical frameworks should be part of the development of AI itself and guide responsible design and use. Secondly, interdisciplinary cooperation ensures diverse input with respect to ethical problems solving by including diverse views and expertise. Thirdly, regulatory oversight of such apps must be guarded to keep users safe and also to command responsibility from developers who can result in dangerous outcomes. Finally, public education and engagement create awareness, trust and the informed participation in AI governance.

Making ethical AI is more than a technical problem; it is a problem of humanity and society. How we will design, implement and regulate AI today will influence its impact on lives tomorrow. So all the stakeholders, developers, businesses, governments and citizens have to join together to ensure the AI being built works for the common good. In order to truly unlock the promise of AI in a rights respecting, dignified and autonomous way we must confront these ethical challenges head on.

REFERENCES

1) Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1), 1–12. https://doi.org/10.1177/2053951715622512
2) Calo, R. (2017). Artificial intelligence policy: A primer and roadmap. *U.C. Davis Law Review*, *51*(2), 399–435. https://lawreview.law.ucdavis.edu/issues/51/2/Symposium/51-2_Calo.pdf

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

3) Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Schafer, B. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5
4) Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2
5) Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*, *26*(4), 2141–2168. https://doi.org/10.1007/s11948-019-00165-5
6) Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453. https://doi.org/10.1126/science.aax2342
7) O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishing Group.
8) Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson Education.
9) Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., ... & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications*, *11*(1), 233. https://doi.org/10.1038/s41467-019-14108-y
10) Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671–732.
11) Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. In *Proceedings of the 2018 Conference on Fairness, Accountability and Transparency* (pp. 149–159).
12) Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 77–91.
13) Calo, R. (2015). Robotics and the lessons of cyberlaw. *California Law Review*, 103(3), 513–563.
14) Cath, C. (2018). Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A*, 376(2133), 20180080.
15) Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.
16) Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
17) European Commission. (2019). Ethics guidelines for trustworthy AI. Retrieved from https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
18) Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association*, 19(1), 121–127.
19) Guidotti, R., Monreale, A., Ruggieri, S., et al. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys*, 51(5), 1–42.
20) Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G. Z. (2019). XAI—Explainable artificial intelligence. *Science Robotics*, 4(37), eaay7120.
21) Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, 30(1), 99–120.
22) Perez Alvarez, M. A., Havens, J. C., & Winfield, A. F. T. (2017). *Ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems*.
23) Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399.
24) Lipton, Z. C. (2018). The mythos of model interpretability. *Queue*, 16(3), 30–57.
25) Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, 6(3), 175–183.
26) Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6), 1–35.
27) Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1–38.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

28) Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.
29) Nyholm, S. (2018). Attributing agency to automated systems: Reflections on human–robot collaborations and responsibility-loci. *Science and Engineering Ethics*, 24(4), 1201–1219.
30) Rahwan, I. (2018). Society-in-the-loop: Programming the algorithmic social contract. *Ethics and Information Technology*, 20(1), 5–14.
31) Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A modern approach* (4th ed.). Pearson.
32) Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human-Computer Interaction*, 36(6), 495–504.
33) Voigt, P., & Von dem Bussche, A. (2017). *The EU General Data Protection Regulation (GDPR): A practical guide*. Springer.
34) Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.
35) Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
36) Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2
37) Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71
38) RSM Global. (2024). *The ethical dilemma of artificial intelligence: Addressing AI bias in decision-making*. Retrieved from https://www.rsm.global
39) Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3), 973–989.frontiersin.org