The Ethics of Understanding: Exploring Moral Implications of Explainable AI

Balkrishna Rasiklal Yadav

Independent Researcher
Email: *yaarkrishna[at]gmail.com*

Abstract: ***Explainable AI (XAI) refers to a specific kind of artificial intelligence systems that are intentionally built to ensure that their operations and results can be comprehended by humans. The main objective is to enhance the transparency of AI systems' decision - making processes, allowing users to understand the rationale behind certain judgements. Important elements of XAI include transparency, interpretability, reasoning, traceability, and user - friendliness. The advantages of Explainable Artificial Intelligence (XAI) include trust and confidence in the system's outputs, ensuring accountability and compliance with regulations, facilitating debugging and refinement of the model, promoting greater cooperation between humans and AI systems, and enabling informed decision - making based on transparent explanations. Examples of XAI applications include healthcare, banking, legal systems, and autonomous systems. Healthcare guarantees that AI - powered diagnosis and treatment suggestions are presented in a straightforward and comprehensible manner, while finance offers explicit elucidations for credit score, loan approvals, and fraud detection. Legal frameworks promote transparency in the implementation of AI applications, therefore assuring equity and mitigating the risk of biases. As artificial intelligence becomes more embedded in society, the significance of explainability will persistently increase, guaranteeing responsible and efficient utilization of these systems. The study of explainable AI is essential as it tackles the ethical, sociological, and technical difficulties presented by the growing use of AI systems. The level of transparency in AI decision - making processes has a direct influence on accountability, since systems that are not transparent might hide the reasoning behind the judgements. Explainability is crucial for detecting and reducing biases in AI systems, so preventing them from perpetuating or worsening social injustices. The objective of the study is to ascertain significant ethical concerns, comprehend the viewpoints of stakeholders, establish an ethical framework, and provide suggestions for policies. The incorporation of Explainable AI into different industries has a significant and far - reaching effect on both technology and society. This includes potential benefits such as increased trust and acceptance, adherence to regulations, improved AI development and troubleshooting, ethical AI design, empowerment and equal access, advancements in education and collaboration, changes in skill requirements, and the establishment of new ethical guidelines.***

Keywords: Explainable AI (XAI), Transparency, Interpretability, Reasoning, Traceability

1. Introduction

The term "explainable AI" (XAI) describes artificial intelligence models and systems that are created with human comprehension of their operations and results in mind. Making AI systems' decision - making processes visible is the main objective of XAI, allowing users to understand how and why certain choices are made. Establishing confidence, guaranteeing responsibility, and promoting efficient human - AI cooperation all depend on this. Explainable AI's main features are as follows:

Transparency: The data utilized, the algorithms employed, and the decision criteria followed are all made available to consumers so they may comprehend the inner workings of the AI model.

Interpretability: The AI system's capacity to provide consumers understandable, succinct explanations. Translation of intricate model operations into human - readable representations is the main goal of interpretability.

Justification: By offering explanations or justifications for its choices, XAI helps users comprehend the thinking behind certain results. This is crucial for applications like the legal or healthcare systems where choices have big consequences.

Traceability: The capacity to link certain data points and model parameters to the decision - making process. This aids in spotting biases or mistakes and comprehending the behavior of the model.

User - Friendliness: XAI systems are made to be friendly to a range of user groups, such as stakeholders, non - experts, and domain experts, so explanations are adapted to the appropriate degree of skill.

DOI: https://dx.doi.org/10.21275/SR24529122811

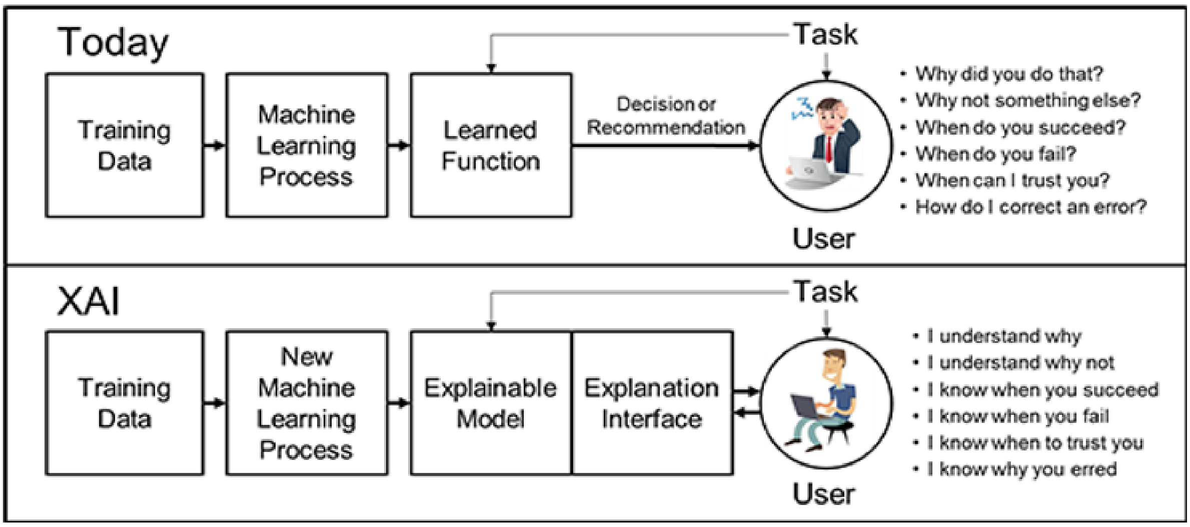

Figure 1: Explainable AI
(Source: https: //www.darpa. mil/program/explainable - artificial - intelligence)

1.1 Benefits of Explainable AI

a) **Trust and Confidence:** If users comprehend how AI systems operate and can discern the logic behind their actions, they will be more inclined to trust and use them.
b) **Accountability and Compliance:** In regulated industries, XAI helps meet legal and ethical standards by providing clear documentation and explanations of AI decisions.
c) **Debugging and Improvement:** Developers can identify and correct errors, biases, or unintended consequences in AI models more effectively when the decision - making process is transparent.
d) **Improved Cooperation:** XAI facilitates improved cooperation between AI systems and human specialists, enabling more successful and efficient decision - making procedures.
e) **Informed Decision - Making:** Users can make more informed decisions by understanding the strengths and limitations of AI recommendations.

Login
CODE OF CONDUCT
Reducing Impact of Model Biasing
Responsibility and Accountability
Governance
Code Confidence
Code Compliance

Figure 2: Benefits of XAI
(Source: https: //www.birlasoft. com/articles/demystifying - explainable - artificial - intelligence)

1.2 Techniques and Methods in Explainable AI

a) **Model - Agnostic Methods:** Strategies that may be used with any machine learning model to explain specific predictions, such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model - agnostic Explanations). Creating models with built - in interpretability, such as rule - based systems, decision trees, or linear regression, where the decision - making process is clear - cut and easy to comprehend.
b) **Post - Hoc Explanations:** Techniques applied after the model has made a decision to explain its behaviour. This includes visualization tools, feature importance measures, and example - based explanations.
c) **Counterfactual Explanations:** Providing alternative scenarios to illustrate how different inputs could lead to different outcomes, helping users understand the model's decision boundaries.
d) **Visual Explanations:** Utilizing visual aids such as heatmaps, attention maps, and graphs to illustrate which parts of the input data were most influential in the decision - making process.

1.3 Applications of Explainable AI

a) **Healthcare:** Ensuring that AI - driven diagnoses and treatment recommendations are transparent and understandable to healthcare providers and patients.
b) **Finance:** Providing clear explanations for credit scoring, loan approvals, and fraud detection to comply with regulatory requirements and build customer trust.
c) **Legal Systems:** Enhancing the transparency of AI applications in legal decision - making, ensuring fairness, and avoiding biases.
d) **Autonomous Systems:** Making the decision - making processes of autonomous vehicles and drones transparent to enhance safety and public acceptance.

e) Explainable AI is a critical area of research and development aimed at making AI systems more transparent, interpretable, and trustworthy. As AI becomes increasingly integrated into various aspects of society, the importance of explainability will continue to grow, ensuring that these systems can be used responsibly and effectively.

2. Literature Review

To properly comprehend the ethical implications of explainable artificial intelligence (AI), which form an important intersection of technology, ethics, and social influence, a great deal of academic study is required. Explainable AI (XAI) aims to make AI systems' decision - making processes transparent and understandable to humans, therefore addressing significant ethical concerns related to accountability, trust, and fairness.

Self and Barocas's (2018) research highlights the need of transparency in AI decision - making, contending that opaque AI systems make it harder for stakeholders to challenge or understand decisions that affect them, exacerbating worries about accountability and justice. According to Lipton (2018), in critical industries like healthcare, law enforcement, and financial services, where decisions may have far - reaching consequences, openness is essential for both user trust and the moral application of AI systems.

Moreover, explainability and fairness are closely related concepts in AI. According to Binns (2018), explainability may be used to detect and lessen biases in AI systems. It is feasible to identify discriminatory trends and make sure AI applications follow Rawls's (1971) distributive justice ideas by gaining insight into the inner workings of AI systems. This is essential to stop AI systems from escalating or maintaining current societal injustices.

Considerations of ethics also include the autonomy and dignity of users. Explainable AI, according to Floridi et al. (2018), respects users' autonomy by giving them the knowledge needed to comprehend and maybe challenge choices made by AI. This is consistent with the ethical philosophy of Kant, which emphasizes the value of seeing people as ends in and of themselves rather than only as means to a goal (Kant, 1785). Therefore, explainable AI preserves human dignity by making sure that users are not exposed to opaque or capricious decision - making processes.

Furthermore, it is believed that explainable AI may improve the ethical accountability of companies and developers using AI technology. Making AI systems explainable places a greater burden of duty on developers to make sure their systems are not just technically sound but also morally good, according to Danks and London (2017). Mittelstadt et al. (2016) share this viewpoint, arguing that an ethical design of AI systems must take into consideration the consequences of their use and the possibility of damage, hence requiring a commitment to openness and responsibility.

Explainability and ethical concerns regarding the usage and security of personal data collide in the field of data privacy. Explainable AI may help consumers understand how their data is used and processed, as Wachter, Mittelstadt, and Floridi (2017) pointed out. This can improve informed consent and compliance with data protection laws like the General Data Protection Regulation (GDPR). This is especially important to make sure AI systems respect people's rights to privacy and control over their personal data.

Reliability and trust are further aspects of the ethical framework for explainable AI. Explainability, as shown by Ribeiro, Singh, and Guestrin (2016), may greatly increase users' confidence in AI systems by offering comprehensible justifications for AI - driven results. Because it affects users' desire to interact with and depend on AI technology, trust is essential to the ethical deployment of AI (Mayer, Davis, & Schoorman, 1995). Therefore, building trust between AI systems and their users is the moral imperative for explainable AI.

In conclusion, there are a variety of moral concerns surrounding explainable AI, including those related to openness, justice, autonomy, accountability, privacy, and trust. Scholarly research emphasizes how morally imperative it is to provide explanations for AI systems in order to guarantee that they are consistent with the values of fairness, dignity for persons, and the well - being of society. This corpus of work emphasizes how important explainable AI is to developing an ecosystem of AI that is morally upright.

3. Significance of Study

Examining the moral ramifications of explainable artificial intelligence (XAI) is important because it tackles important moral, technical, and social issues brought up by the growing use of AI systems. This field of study is essential for a number of reasons:

First, accountability is directly impacted by the openness of AI decision - making processes. Selbst and Barocas (2018) stress that judgements made by opaque AI systems may be difficult for impacted parties to comprehend or challenge since the reasoning behind them may be hidden. Legal foundations and democratic norms that depend on accountability and the right to seek redress are undercut by this lack of openness.

There are important ramifications for justice and fairness in the research of XAI's moral implications. AI systems are often used in situations where judgements may have a significant impact on people's lives, such lending, employment, and law enforcement. In order to ensure that AI algorithms do not reinforce or worsen already - existing social disparities, Binns (2018) emphasizes the need of explainability in recognizing and reducing biases in these systems. Thus, research in this field helps to design AI systems that are consistent with Rawls's (1971) distributive justice concepts.

Furthermore, investigating the ethical aspects of XAI helps to protect people's autonomy and dignity. Explainable AI, according to Floridi et al. (2018), respects users' autonomy by giving them the knowledge necessary to comprehend and maybe contest AI judgements. This is consistent with the ethical theory of Kant, which holds that people ought to be seen as ends in and of themselves rather than as means to a goal (Kant, 1785). By encouraging openness, XAI protects users' dignity by making sure they aren't exposed to capricious or mysterious decision - making procedures.

Enhancing the moral responsibility of AI creators and deploying organizations is another important goal of studying XAI. According to Danks and London (2017), developers are held to a higher ethical standard when creating explainable AI systems, since they must make sure that their creations are both technically and morally sound. Mittelstadt et al. (2016), who contend that ethical AI design must take into consideration the wider implications of deployment and possible damage, calling for a commitment to openness and responsibility, bolster this point of view.

XAI is essential to maintaining ethical data use and regulatory compliance in terms of data protection. Explainable AI has the potential to improve users' comprehension of how their data is handled, which may lead to more informed consent and compliance with data protection regulations such as the General Data Protection Regulation (GDPR), as discussed by Wachter, Mittelstadt, and Floridi (2017). Maintaining people's rights to privacy and control over their personal information depends on this.

XAI also significantly influences other important elements like trust and dependability. Explainability, as shown by Ribeiro, Singh, and Guestrin (2016), may greatly increase user confidence in AI systems by offering understandable, concise justifications for AI - driven results. Users' propensity to depend on AI systems is influenced by trust, which is crucial for the ethical adoption of AI technology (Mayer, Davis, & Schoorman, 1995). As a result, XAI plays a crucial role in creating and maintaining a trustworthy connection between people and AI systems.

Lastly, research on the moral implications of XAI adds to the larger conversation on the moral application of AI in society. It offers a structure for creating rules and regulations that guarantee AI technologies are used in a reasonable, equitable, and human rights - abiding manner. The development of an ethical AI ecosystem that puts people's welfare and the welfare of society at large first is supported by this study.

In conclusion, since explainable AI has the potential to improve privacy, autonomy, responsibility, accountability, and justice in AI systems, it is important to research the moral implications of this technology. In order to ensure that AI technologies are created and used in a manner that is consistent with moral standards and social norms, and eventually contribute to a more fair and equitable society, this study is crucial.

4. Proposed Methodology

The research design, data collecting strategies, data analysis approaches, and ethical considerations required to examine the moral implications of Explainable AI (XAI) will all be covered in the methodology section. A mixed - methods approach will be used in the study to fully comprehend the ethical aspects involved.

4.1 Research Design

The research will use a mixed - methods strategy, integrating qualitative and quantitative techniques to get a comprehensive comprehension of the ethical ramifications of artificial intelligence.

4.2 Research Questions

- What are the primary moral concerns associated with the use of XAI?
- How do stakeholders perceive the ethical implications of XAI?
- What frameworks can be developed to address these ethical concerns?

4.3 Data Collection Methods

a) **Literature Review**
A comprehensive fiction review will be conducted to identify existing research on the moral and ethical implications of XAI. Sources will include academic journals, conference papers, books, and reputable online publications.

b) **Surveys**
Quantitative data will be collected using structured surveys. The survey will target various stakeholders, including AI developers, users, ethicists, and policymakers. The survey will include Likert scale questions to gauge attitudes and perceptions regarding the ethical implications of XAI.

c) **Interviews**
Qualitative data will be gathered through semi - structured meetings with key stakeholders. This will allow for an in - depth understanding of their perspectives on the moral concerns associated with XAI. Interviewees will include AI researchers, industry experts, ethicists, and representatives from regulatory bodies.

d) **Case Studies**
Case studies of organizations implementing XAI will be analyzed to understand practical ethical issues and solutions. These case studies will provide real - world examples of how XAI is being used and the associated moral implications.

4.4 Data Analysis

a) **Quantitative Analysis**
Statistical techniques will be used to analyse the survey data. While inferential statistics (such as chi - square tests and t - tests) can find significant differences and correlations between variables, descriptive statistics will provide an overview of the replies.

b) **Qualitative Analysis**
Thematic analysis will be used to examine data from case studies and interview transcripts. To find recurrent themes and patterns regarding the moral ramifications of XAI, the data will be coded.

4.5 Ethical Considerations

- **Informed Consent:** Every participant will get information about the goals, methods, and free withdrawal policy of the research at any time.

- **Confidentiality:** To maintain confidentiality, participant names will be safeguarded and data will be anonymized.
- **Bias and Objectivity:** The research will strive to minimize bias by employing triangulation—using multiple data sources and methods to cross - verify findings.

4.6 Framework Development

Based on the findings from the data analysis, a framework for addressing the moral implications of XAI will be developed. This framework will provide guidelines for developers, users, and policymakers to ensure ethical practices in the deployment and use of XAI.

4.7 Validation

The proposed framework will be validated through expert reviews and a pilot implementation in selected organizations. Feedback from these validations will be used to refine the framework.

4.8 Conclusion

The methodology outlined above aims to provide a complete understanding of the moral implications of XAI and develop practical solutions to address these ethical concerns. This research will contribute to the responsible expansion and deployment of XAI technologies.

This methodology provides a structured approach to investigating the ethical dimensions of XAI, ensuring a thorough and balanced examination of the topic.

5. Limitations and Future Implications

When researching the "Moral Implications of Explainable AI, " several limitations could potentially influence the study's outcomes and its general applicability. Addressing these limitations upfront can help in managing expectations and in framing the results appropriately:

a) Subjectivity in Qualitative Data

Interpretation Bias: Since the study heavily relies on qualitative methods such as interviews and case studies, The interpretation of data is inherently subjective. From the same data collection, various researchers may come to different findings.

Interviewer Bias: The dynamics between the interviewer and the interviewee can affect how questions are answered, possibly skewing data towards socially desirable responses rather than honest opinions.

b) Limited Generalizability

Sample Diversity: The purposive sampling method, while effective for obtaining detailed insights from specific groups, may not represent all stakeholders involved with XAI. This limits the generalizability of the findings to other contexts or populations.
Case Study Selection: The case studies selected may not cover all possible applications or scenarios where XAI is used, which might result in a partial picture of the moral implications.

c) Evolving Nature of AI Technologies

Rapid Technological Changes: The domains of AI and XAI are developing quickly. The results of this research might be swiftly superseded by new techniques and technology.
Regulatory Lag: There might be a lag in ethical and regulatory guidelines catching up with technological advancements, making some of the ethical considerations discussed either too speculative or soon outdated.

d) Ethical and Moral Complexity

Cultural Relativity: Ethical norms and moral judgments can vary significantly across different cultures and societies. A framework that is developed based on one cultural context might not be applicable universally.
Conflicting Ethics: There might be conflicting ethical principles involved in the use of XAI. For instance, transparency might conflict with privacy. Balancing these ethical principles can be challenging and context - dependent.

e) Access to Proprietary Technologies

Confidentiality Constraints: Some organizations may use proprietary XAI technologies and may not be willing to share detailed information for case studies due to confidentiality or competitive reasons. This can limit the depth of analysis possible for real - world applications.

f) Scope of Study

Concentration on XAI: Focusing specifically on explainable AI may overlook broader ethical issues related to AI that are not directly linked to explainability but are equally important, such as bias in data or AI misuse.

g) Dependence on Existing Literature

Research Bias in Sources: The existing literature may have its biases, especially if it is dominated by authors from specific regions or institutions. This can influence the literature review process and subsequently the framing of the entire study.

By acknowledging these limitations, researchers can tailor their analysis and discussions to provide a clearer, more reliable interpretation of the data and its implications for the field of XAI ethics.

The integration of Explainable AI (XAI) into various sectors poses a transformative impact on both technology and society. As AI systems become more prevalent, the demand for transparency and understandability in these systems is increasing. Below are some of the potential future implications of XAI across different domains:

a) Enhanced Trust and Adoption

Implication: As AI systems become more explainable, users and stakeholders are likely to develop greater trust in these technologies. This could lead to wider adoption across critical sectors such as healthcare, finance, and legal services, where understanding AI decisions is crucial.

Example: In healthcare, doctors could more readily incorporate AI diagnostics into their practice if they can understand how the AI reached its conclusions, thereby improving patient outcomes through augmented decision - making.

b) Regulatory Compliance

Implication: Regulatory bodies worldwide are beginning to demand greater transparency in AI operations, particularly in how data is used and decisions are made. XAI can help organizations meet these regulatory requirements by making AI processes more transparent.

Example: An example of this would be the right to explanation provided by the General Data Protection Regulation (GDPR) of the European Union, which allows anyone to request an explanation of any algorithmic decision made about them.

c) Improved AI Development and Debugging

Implication: Explainability aids developers in understanding and improving AI models more effectively, particularly in identifying and correcting prejudices or errors in AI behavior.

Example: XAI can reveal if a model is using irrelevant features (like race or gender) for making decisions, allowing developers to adjust the model to avoid unethical outcomes and improve its accuracy.

d) Ethical AI Design

Implication: With a better understanding of how AI models make decisions, designers and developers can create more ethical AI systems that bring into line with human values and ethical standards.

Example: XAI could ensure that AI lending systems do not discriminate against convinced groups by making the basis of their credit scoring transparent and adjustable.

e) Empowerment and Democratization

Implication: Explainable AI can democratize AI technology by making it accessible and understandable to non - experts, thereby empowering more people to utilize and scrutinize AI technologies effectively.

Example: Small business owners could use XAI tools to understand customer data and behavior predictions, enabling them to make informed decisions without needing specialized knowledge.

f) Educational and Collaborative Advancements

Implication: XAI has the potential to become a powerful educational tool, helping students and researchers understand complex AI models and facilitating interdisciplinary collaboration.

Example: In academia, XAI could be used to help students in fields like psychology or sociology understand how AI can be applied in their disciplines, fostering new research opportunities and insights.

g) Shift in Skill Requirements

Implication: As AI systems become more explainable and adopted across various sectors, there will be a shift in the job market, with an increasing demand for professionals who can interpret and work with AI outputs effectively.

Example: Future jobs might require employees to understand and interact with AI recommendations, integrating these into their workflows and decision - making processes.

h) Creation of New Ethical Standards

Implication: The expansion of XAI will likely lead to the development of new ethical standards and best practices specifically tailored to the deployment and use of explainable AI systems.

Example: Professional organizations and ethical bodies may establish guidelines on how XAI should be implemented to ensure fairness, accountability, and transparency in AI - driven decisions.

These implications highlight the vast potential of XAI to influence technology development and social norms, emphasizing the importance of integrating explainability in AI systems to foster an ethical, transparent, and inclusive future.

6. Expected Outcome

The study on Explainable AI's moral implications aims to identify key ethical issues, understand stakeholder perspectives, develop an ethical framework, and generate policy recommendations. It will catalogue and elucidate primary ethical concerns, such as privacy, bias, accountability, and transparency, providing a comprehensive overview of the challenges faced by stakeholders in implementing AI systems. The research will also provide insights into how developers, users, ethicists, and policymakers perceive the moral implications of XAI, including attitudes towards transparency, trust in AI decisions, and trade - offs between explainability and other AI performance metrics. The study will also generate recommendations for policymakers on regulating XAI applications, such as legislation or industry standards. The research will also promote broader adoption of ethical AI practices across various sectors, increasing trust in and effectiveness of AI technologies. Future research directions will identify gaps in current research and propose areas for further investigation.

References

[1] Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency, 149 - 159.

[2] Danks, D., & London, A. J. (2017). Algorithmic bias in autonomous systems. In Proceedings of the Twenty - Sixth International Joint Conference on Artificial Intelligence, 4691 - 4697.

[3] Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V.,. . . & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28 (4), 689 - 707.

[4] Kant, I. (1785). Groundwork for the Metaphysics of Morals. Cambridge University Press.

[5] Lipton, Z. C. (2018). The mythos of model interpretability. Communications of the ACM, 61 (10), 36 - 43.

[6] Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. Academy of Management Review, 20 (3), 709 - 734.

[7] Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3 (2), 2053951716679679.

[8] Rawls, J. (1971). A Theory of Justice. Harvard University Press.

[9] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135 - 1144.

[10] Selbst, A. D., & Barocas, S. (2018). The intuitive appeal of explainable machines. Fordham Law Review, 87, 1085.

[11] Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision - making does not exist in the general data protection regulation. International Data Privacy Law, 7 (2), 76 - 99.

[12] Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency, 149 - 159.

[13] Danks, D., & London, A. J. (2017). Algorithmic bias in autonomous systems. In Proceedings of the Twenty - Sixth International Joint Conference on Artificial Intelligence, 4691 - 4697.

[14] Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V.,. . . & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28 (4), 689 - 707.

[15] Kant, I. (1785). Groundwork for the Metaphysics of Morals. Cambridge University Press.

[16] Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. Academy of Management Review, 20 (3), 709 - 734.

[17] Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3 (2), 2053951716679679.

[18] Rawls, J. (1971). A Theory of Justice. Harvard University Press.

[19] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135 - 1144.

[20] Selbst, A. D., & Barocas, S. (2018). The intuitive appeal of explainable machines. Fordham Law Review, 87, 1085.

[21] Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision - making does not exist in the general data protection regulation. International Data Privacy Law, 7 (2), 76 - 99.