Frontiers in Environmental Research
ISSN: 2960-026X
DOI: 10.61784/fer3005

EXPLAINABLE AI FOR TRANSPARENT EMISSION REDUCTION DECISION-MAKING

Jeng-Jui Du, Shiu-Chu Chiu*
College of Science and Engineering, Flinders University, Clovelly Park, SA 5042, Australia.
Corresponding Author: Shiu-Chu Chiu, Email: shchiuflin234@gmail.com

Abstract: This paper examines the critical role of Explainable AI (XAI) in enhancing transparency in emission reduction decision-making processes. As climate change poses an urgent global challenge, effective strategies for reducing greenhouse gas emissions are essential for mitigating its impacts. Artificial Intelligence has emerged as a powerful tool in environmental management, facilitating data analysis and optimizing emission reduction efforts. However, the increasing reliance on AI raises concerns about transparency and accountability, which are vital for gaining public trust. This paper defines XAI and explores its methodologies, emphasizing their potential to improve stakeholder engagement and decision-making in environmental policy. By synthesizing existing literature and case studies, we highlight the importance of explainability in fostering trust among stakeholders and ensuring effective and accountable emission reduction strategies. The findings contribute to the ongoing discourse on the ethical and practical implications of AI in environmental governance and underscore the necessity of incorporating XAI into future emission reduction initiatives.
Keywords: Explainable AI; Emission reduction; Transparency

1 INTRODUCTION

Climate change represents one of the most pressing challenges of our time, with far-reaching consequences for ecosystems, human health, and global economies[1-5]. The scientific consensus underscores the urgent need for significant reductions in greenhouse gas emissions to mitigate the worst effects of climate change [6]. Governments, corporations, and civil society organizations are increasingly recognizing the necessity of transitioning to low-carbon economies and implementing effective emission reduction strategies [7-10]. The Paris Agreement, adopted in 2015, set forth ambitious targets to limit global warming to well below 2 degrees Celsius, necessitating immediate and sustained action [11-13].

Artificial Intelligence has emerged as a powerful tool in various sectors, including environmental management, by enhancing data analysis, forecasting, and decision-making capabilities [14]. AI systems can process vast amounts of data and uncover patterns that may not be immediately apparent to human analysts [15]. In the context of emission reduction, AI can optimize energy consumption, predict emissions, and evaluate the efficacy of different strategies [16]. However, the increasing reliance on AI in decision-making raises critical questions regarding transparency, accountability, and public trust [17-20].

Explainable AI refers to a set of processes and techniques designed to make the outputs of AI systems understandable to human users [21]. As AI models become more complex, the need for explainability becomes paramount, particularly in high-stakes domains such as environmental policy [22]. XAI aims to provide insights into how AI systems arrive at their conclusions, thereby facilitating better decision-making and fostering trust among stakeholders [23-24].

This paper aims to explore the role of Explainable AI in enhancing transparency within emission reduction decision-making processes. By examining the intersection of XAI and environmental policy, we seek to understand how explainable AI methodologies can improve stakeholder engagement, foster trust, and ultimately lead to more effective and accountable emission reduction strategies. The findings will contribute to the ongoing discourse on the ethical and practical implications of AI in environmental governance.

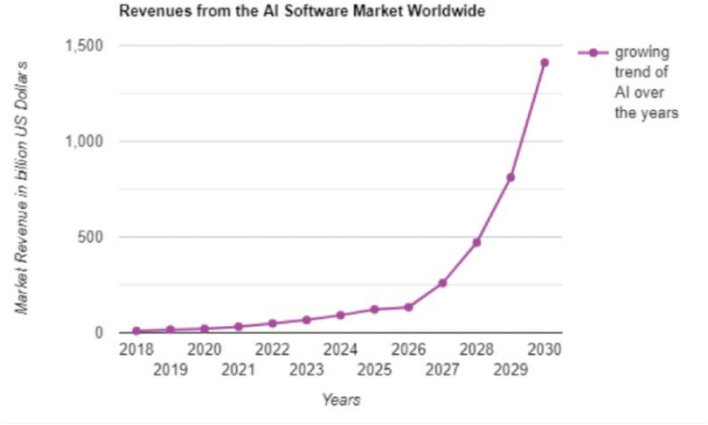

Figure 1 Worldwide AI Revenue and Growth

2 LITERATURE REVIEW

A robust body of literature has emerged around the themes of climate change, artificial intelligence, and explainability, highlighting the critical need for transparency in decision-making processes related to emission reduction. This literature review synthesizes key findings from various studies to provide a comprehensive overview of the current state of research in these interconnected fields [25-30].

The urgency of addressing climate change has prompted extensive research into effective emission reduction strategies [31]. Scholars have identified a range of approaches, including renewable energy adoption, carbon pricing, and improvements in energy efficiency [32]. For instance, studies have shown that transitioning to renewable energy sources, such as solar [33] and wind power [34], can significantly lower greenhouse gas emissions while fostering economic growth [35]. Additionally, carbon pricing mechanisms, such as cap-and-trade systems and carbon taxes, have been advocated as effective tools for incentivizing emission reductions among corporations and industries [36]. Furthermore, enhancing energy efficiency in buildings, transportation, and industrial processes has been recognized as a cost-effective strategy for reducing emissions and mitigating climate change impacts [38-41].

The role of diverse stakeholders—governments, corporations, non-governmental organizations , and the public—in shaping these strategies has been emphasized in the literature [42-46]. Collaboration among these stakeholders is essential for achieving meaningful reductions in emissions, as it fosters the sharing of knowledge, resources, and best practices [47]. Research has highlighted successful case studies where multi-stakeholder partnerships have led to innovative solutions and significant emission reductions, demonstrating that collective action is vital in the fight against climate change [48-50].

AI's potential to transform environmental decision-making has been explored in various contexts. Research indicates that AI can enhance predictive modeling [51], optimize resource allocation [52], and facilitate real-time monitoring of emissions [53]. For example, machine learning algorithms have been employed to analyze vast datasets related to energy consumption and emissions, providing insights that enable policymakers to make informed decisions [54].

Additionally, AI-driven tools can optimize the deployment of monitoring system in emissions, ensuring that energy supply aligns with demand while minimizing emissions [55]. However, the integration of AI into environmental policy raises concerns regarding transparency, as many AI models operate as "black boxes," making it difficult for stakeholders to understand their outputs and the rationale behind them [56].

The concept of Explainable AI has gained traction as researchers seek to address the opacity of AI systems. XAI encompasses a variety of techniques aimed at making AI decision-making processes more transparent and interpretable. Studies have shown that explainability can enhance user trust and facilitate better decision-making, particularly in high-stakes applications such as healthcare and finance [57]. In the context of environmental policy, XAI can play a pivotal role in ensuring that AI-generated recommendations are interpretable and actionable, thereby empowering stakeholders to implement effective emission reduction strategies [58].

Recent studies have highlighted several applications of XAI in emission reduction efforts. For instance, researchers have demonstrated the effectiveness of XAI in emission forecasting and modeling, enabling policymakers to make informed decisions based on transparent data [59, 60]. By providing insights into the factors influencing emissions over time, XAI can help stakeholders identify trends and assess the impact of various interventions. Additionally, XAI has been employed to evaluate carbon offset programs, providing insights into their effectiveness and potential improvements. By clarifying how offsets are calculated and the assumptions underlying these calculations, XAI enhances accountability and encourages more effective carbon management practices [61-63].

Despite the promise of XAI, several challenges remain. Technical difficulties, such as balancing accuracy and explainability, pose significant obstacles to the widespread adoption of XAI techniques. Many advanced AI models, particularly those based on deep learning, excel in predictive accuracy but are often criticized for their lack of interpretability. This trade-off raises questions about the appropriateness of using such models in critical areas like environmental policy, where understanding the reasoning behind decisions is essential.

Furthermore, ethical considerations, including bias in AI models and privacy concerns, must be addressed to ensure responsible AI deployment. AI systems can inadvertently perpetuate existing biases present in training data, leading to inequitable outcomes in emission reduction strategies. It is crucial for researchers and practitioners to implement strategies that mitigate bias and promote fairness in AI applications. Additionally, the use of personal or sensitive data in AI models raises significant privacy issues, necessitating robust data protection measures and compliance with regulations such as the General Data Protection Regulation (GDPR).

The literature on climate change, AI, and explainability reveals a complex interplay between these fields, highlighting the critical need for transparency in emission reduction efforts. As research continues to evolve, it is essential for stakeholders to address the challenges associated with AI integration while leveraging its potential to drive meaningful change in environmental policy. By fostering collaboration, enhancing explainability, and prioritizing ethical considerations, stakeholders can work towards more effective and equitable emission reduction strategies that align with global sustainability goals. The ongoing dialogue in this area will be vital for advancing our understanding of how AI can be harnessed responsibly in the context of climate change.

3 UNDERSTANDING EXPLAINABLE AI

3.1 Definition and Key Concepts of XAI

Explainable AI encompasses a range of methods and techniques designed to make the outcomes of artificial intelligence systems interpretable and understandable to human users. The primary aim of XAI is to provide insights into the decision-making processes of AI models, which is crucial for fostering trust among stakeholders and facilitating informed decision-making. As AI technologies become increasingly integrated into high-stakes domains such as healthcare, finance, and environmental policy, the need for XAI has become more apparent. In these contexts, understanding the rationale behind decisions can have significant consequences, impacting not only individual lives but also broader societal outcomes.

The importance of XAI is underscored by the growing complexity of AI models, particularly those based on deep learning. While these models often achieve impressive accuracy, their intricate architectures can obscure the pathways through which they arrive at specific predictions. This opacity can lead to skepticism and reluctance among users who are required to rely on these systems for critical decisions. By providing clear explanations, XAI serves to demystify AI processes, enabling users to engage with the technology confidently and effectively.

3.2 Types of Explainability

3.2.1 Global vs. local explainability

Explainability can be categorized into two main types: global and local explainability. Global explainability refers to the understanding of the overall behavior and decision-making patterns of an AI model across the entire dataset. This type of explainability is essential for stakeholders to grasp how the model functions as a whole, including the factors that influence its predictions and the general trends it identifies within the data. Global explainability can help organizations assess the reliability and robustness of a model, ensuring that it aligns with their objectives and ethical standards.

Table 1 Some Efficiency Parameters for GKE/GAKE Protocols Claimed to be Quantum-Resistant

Protocol	# Rounds	Avoids PQ-Sign.	# Broadcast Messages	# PtP Messages
n-UM [1]	1	Yes	n	0
BC n-DH [1]	1	Yes	n	0
Apon et al. [2]	3	Yes (but is unauth.)	$2n+1$	0
STAG [4]	3	No	$2n+1$	0
Pers. et al. [5]	3	No	n	$2n$
Gonz. et al. [6]	2	Yes	n	n^2-n
This work	4	Yes	$2n$	$2n$

In contrast, local explainability focuses on providing insights into individual predictions or decisions made by the model. This type of explainability is crucial for validating specific outcomes, as it allows users to understand the reasons behind particular predictions. For instance, in the context of emission reduction strategies, local explainability can help stakeholders evaluate the rationale behind a model's recommendation for a specific policy or intervention. By understanding the factors that led to a particular decision, users can assess its relevance and appropriateness in their specific context.

Both global and local explainability are important for a comprehensive understanding of AI models. While global explainability helps stakeholders develop a broad understanding of a model's capabilities and limitations, local explainability provides the detailed insights necessary for informed decision-making at the individual level.

3.2.2 Model-agnostic vs. model-specific techniques

Another important distinction in the realm of XAI is between model-agnostic and model-specific techniques. Model-agnostic techniques are designed to be applicable to any machine learning model, regardless of its underlying architecture. These techniques focus on generating explanations that can be interpreted across different types of models, making them versatile tools for practitioners. Examples of model-agnostic techniques include Local Interpretable Model-agnostic Explanations and SHapley Additive exPlanations. Both methods provide insights into how individual features contribute to a model's predictions, allowing users to gain a better understanding of the decision-making process.

On the other hand, model-specific techniques are tailored to particular algorithms and leverage the unique characteristics of those models to generate explanations. For instance, decision trees inherently provide a level of interpretability due to their straightforward structure, making them easier to explain compared to more complex models like neural networks. Techniques such as feature importance scores or visualization tools can be used to elucidate the workings of specific models, providing stakeholders with insights that are directly relevant to the algorithms they are

using.

3.3 Importance of XAI in AI Applications

3.3.1 Enhancing user trust

Trust in AI systems is critical for their acceptance and effective use. As AI technologies become more prevalent in various sectors, including emission reduction strategies, the need for transparency has never been more crucial. XAI enhances user trust by providing transparent explanations that allow users to understand and validate AI-generated recommendations. When stakeholders can comprehend the reasoning behind a model's predictions, they are more likely to feel confident in the technology's reliability and accuracy. This trust is particularly important in high-stakes scenarios, where the consequences of decisions can have far-reaching impacts on environmental sustainability, public health, and economic stability.

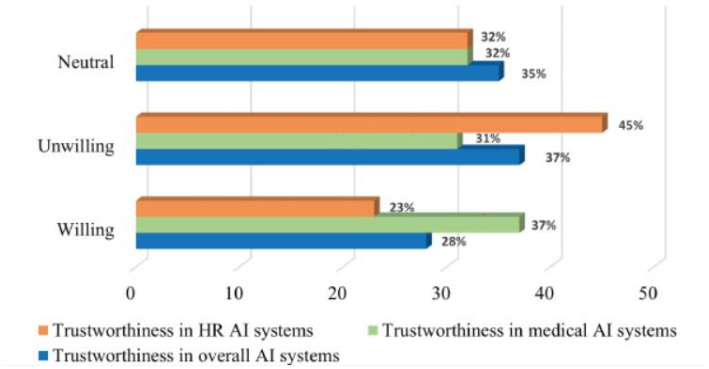

Figure 2 Trustworthiness in AI systems

The relationship between trust and explainability is reciprocal; as users gain confidence in AI systems through clear explanations, they are more inclined to rely on these technologies for decision-making. This dynamic is especially pertinent in the context of emission reduction, where stakeholders must navigate complex data and competing interests. By fostering trust through XAI, organizations can encourage broader adoption of AI-driven insights, ultimately leading to more effective and impactful emission reduction strategies.

3.3.2 Facilitating better decision-making

In addition to enhancing trust, XAI plays a critical role in facilitating better decision-making. By offering interpretable insights into the workings of AI models, XAI enables stakeholders to make more informed and effective decisions. In the context of emission reduction, this can lead to the development of more effective policies and strategies that are grounded in data-driven insights. When stakeholders understand how various factors influence model predictions, they can assess the relevance of these insights to their specific circumstances and make adjustments as necessary.

Moreover, XAI can help identify potential biases or shortcomings in AI models, prompting stakeholders to critically evaluate the data and assumptions underlying their decisions. This scrutiny can lead to more equitable and just outcomes, as organizations are better equipped to address disparities and ensure that their emission reduction strategies benefit all segments of society. By promoting transparency and accountability, XAI ultimately contributes to the creation of more robust and effective environmental policies that align with broader sustainability goals.

In summary, understanding Explainable AI is essential for harnessing its potential in various applications, particularly in the realm of emission reduction. By defining key concepts, exploring different types of explainability, and highlighting the importance of XAI in enhancing user trust and facilitating better decision-making, stakeholders can better navigate the complexities of AI technologies. As the integration of AI into critical domains continues to expand, the role of XAI will be increasingly vital in ensuring that these systems serve the interests of society effectively and equitably.

4 FRAMEWORK FOR IMPLEMENTING XAI IN EMISSION REDUCTION DECISION-MAKING

A comprehensive framework for implementing XAI in emission reduction decision-making consists of several key components:

4.1 Key Components of an XAI Framework

Effective XAI implementation begins with robust data collection and preprocessing. This involves gathering relevant data from diverse sources, such as satellite imagery, sensor data, and historical emissions data. Data preprocessing steps, including cleaning, normalization, and feature selection, are essential to ensure high-quality inputs for AI models.

Selecting the appropriate AI model is crucial for achieving accurate predictions. The choice of model should consider the complexity of the problem, the nature of the data, and the need for explainability. Once a model is selected, it should be trained on the preprocessed data, with performance metrics evaluated to ensure reliability.

After training the model, explanation generation methods should be employed to provide insights into the model's

predictions. This can include model-agnostic techniques like LIME and SHAP, as well as model-specific techniques tailored to the chosen algorithm.

4.2 Stakeholder Engagement and Collaboration

Engaging diverse stakeholders is crucial for the successful implementation of XAI in emission reduction strategies. Policymakers, scientists, and the public can provide valuable insights and feedback that enhance the effectiveness of AI systems. Collaboration can also help ensure that the needs and concerns of all stakeholders are addressed.
Addressing complex environmental challenges requires interdisciplinary collaboration. Involving experts from fields such as environmental science, data science, and social sciences can lead to more comprehensive and effective XAI solutions.

4.3 Tools and Technologies for XAI

Various software tools and platforms are available to support the implementation of XAI. These include open-source libraries like LIME, SHAP, and ELI5, which facilitate the generation of interpretable explanations for machine learning models.
Best practices for implementing XAI include ensuring transparency throughout the model development process, regularly engaging with stakeholders, and continuously evaluating and refining the model based on user feedback.

5 CHALLENGES AND LIMITATIONS OF XAI IN EMISSION REDUCTION

Despite the potential benefits of Explainable AI in enhancing transparency and accountability in emission reduction strategies, several challenges and limitations must be addressed to fully realize its potential. These challenges span technical, ethical, and organizational dimensions, each posing unique obstacles to the effective implementation of XAI.

5.1 Technical Challenges

One of the primary challenges in implementing XAI is the inherent trade-off between model accuracy and interpretability. Advanced models, particularly those based on deep learning architectures, often achieve higher accuracy through their ability to capture complex patterns in data. However, this complexity comes at a cost; such models are frequently criticized for their lack of interpretability, making it difficult for stakeholders to understand the rationale behind their predictions. Conversely, simpler models, while more interpretable and easier to understand, may sacrifice predictive accuracy, leading to suboptimal decision-making outcomes . This dichotomy presents a significant challenge for practitioners who must balance the need for accurate predictions with the necessity for clear explanations.

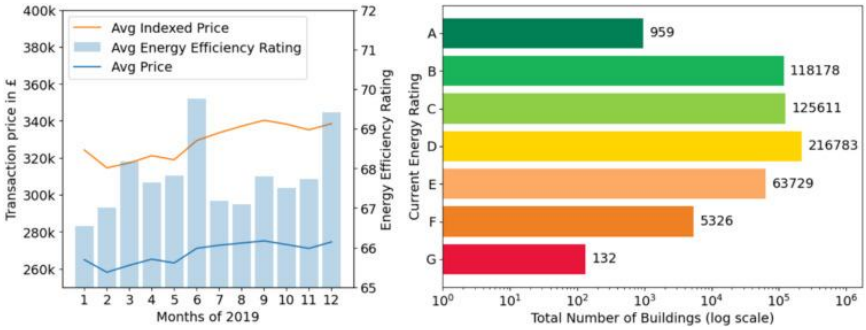

Figure 3 Average Energy Efficiency Rating and Transaction Price Per Month (Left) and Distribution of Energy Rating (Right)

Current XAI methods may not always provide sufficient explanations, particularly when applied to highly complex models. For instance, many existing XAI techniques focus on local explainability, which provides insights into individual predictions but may fail to capture the broader context necessary for effective decision-making. This limitation can hinder stakeholders' ability to understand how various factors interact within the model, potentially leading to misinterpretations and misguided actions. Moreover, the dynamic nature of emission reduction strategies often requires a holistic view of the system, which may not be achievable through localized explanations alone.

5.2 Ethical Considerations

Ethical considerations are paramount when deploying AI models in emission reduction efforts. One significant concern is that AI models can inherit biases present in the training data, leading to unfair or discriminatory outcomes. For example, if historical data reflects systemic inequalities in energy access or pollution exposure, AI models trained on such data may perpetuate these biases, resulting in emission reduction strategies that disproportionately benefit certain

groups over others. Therefore, ensuring that XAI methods effectively address these biases is crucial for promoting fairness and equity in emission reduction strategies. This requires ongoing vigilance and a commitment to evaluating the ethical implications of AI applications.

Protocol	Assumption Type	Model	FutQ/PostQ	Authent.
n-UM [1]	Isogeny	QROM	PostQ	Yes
BC n-DH [1]	Isogeny	ROM	PostQ	Yes
Apon et al. [2]	Lattice	ROM	PostQ	No
STAG [4]	Lattice	ROM	PostQ	Yes
Pers. et al. [5]	Compiler	No RO added	PostQ	Yes
Gonz. et al. [6]	Compiler	No RO added	FutQ	Yes
This work	Lattice	QROM	PostQ	Yes

Table 2 Security of GKE/GAKE Protocols Claimed to be Quantum-Resistant

Additionally, the use of personal or sensitive data in AI models raises significant privacy concerns. As organizations increasingly rely on data-driven insights, ensuring that data protection measures are in place is essential. Compliance with regulations such as the General Data Protection Regulation is not only a legal requirement but also a moral obligation to protect individuals' rights and privacy. Organizations must implement robust data governance frameworks that prioritize transparency and accountability in data handling practices to build trust among stakeholders.

5.3 Resistance to Change

Resistance to change within organizations can pose a significant barrier to the adoption of XAI practices. Many organizations have established workflows and processes that may not readily accommodate the integration of new technologies or methodologies. This inertia can be particularly pronounced in sectors that are traditionally risk-averse or heavily regulated, where stakeholders may be hesitant to embrace unfamiliar approaches. Overcoming this resistance requires strong leadership and a commitment to fostering a culture of innovation that encourages experimentation and learning.

Effective implementation of XAI necessitates comprehensive training and education for stakeholders at all levels. Providing resources and support for understanding XAI concepts and techniques can facilitate acceptance and effective use. Organizations must invest in capacity-building initiatives that empower employees to engage with XAI tools and methodologies confidently. This includes not only technical training but also fostering an organizational mindset that values transparency, collaboration, and continuous improvement.

In summary, while XAI holds great promise for enhancing decision-making in emission reduction efforts, addressing the challenges and limitations outlined above is essential for its successful implementation. By navigating the technical complexities, ethical considerations, and organizational resistance, stakeholders can leverage the power of XAI to create more effective and equitable emission reduction strategies. As the landscape of climate action continues to evolve, the integration of explainable AI will play a critical role in ensuring that stakeholders can make informed decisions that align with global sustainability goals. Continued research and dialogue in this area will be vital to overcoming these challenges and unlocking the full potential of XAI in the fight against climate change.

6 CONCLUSION

This paper has thoroughly explored the pivotal role of Explainable AI in enhancing transparency and accountability within emission reduction decision-making processes. As the urgency to combat climate change escalates, the integration of advanced technologies such as AI has become increasingly prevalent in formulating strategies aimed at reducing greenhouse gas emissions. However, the complexity and often opaque nature of traditional AI models pose significant challenges to stakeholders who rely on these insights for critical decision-making. By defining key concepts of XAI, outlining a comprehensive framework for its implementation, and discussing the various challenges and future directions in this field, it is evident that XAI possesses the potential to significantly improve decision-making processes in the context of climate change.

The significance of XAI lies primarily in its ability to foster trust among stakeholders, including policymakers, scientists, and the general public. In an era where decisions regarding climate action can have far-reaching implications, the ability to understand and interpret the rationale behind AI-driven insights is crucial. By providing clear explanations of how models arrive at specific conclusions, XAI can mitigate skepticism and enhance the credibility of AI applications in emission reduction strategies. This transparency is vital for promoting informed decision-making, as stakeholders can better assess the implications of various strategies and make choices that align with sustainability goals.

Moreover, the adoption of XAI practices is essential for ensuring accountability in the deployment of AI technologies. As organizations increasingly rely on AI-driven insights to inform their emission reduction strategies, it is imperative that these systems are not only effective but also transparent and justifiable. XAI can help address the opacity associated with traditional AI models by elucidating the underlying mechanisms of decision-making, thereby enabling stakeholders

to hold systems accountable for their predictions and recommendations. This accountability is particularly important in the context of climate change, where the consequences of decisions can have profound effects on environmental sustainability and public health.

To harness the full potential of XAI, a collaborative effort among policymakers, researchers, and industry leaders is necessary. Such collaboration can foster a deeper understanding of the unique challenges posed by climate change and the role of AI in addressing these challenges. By investing in training programs that equip stakeholders with the skills needed to interpret and utilize XAI effectively, organizations can enhance their capacity to implement data-driven solutions for emission reduction. Furthermore, fostering interdisciplinary collaboration between AI experts, environmental scientists, and policymakers can lead to the development of more robust and contextually relevant XAI frameworks that address the specific needs of different sectors.

Promoting transparency is another critical aspect of maximizing the benefits of XAI in emission reduction efforts. By advocating for open data practices and the sharing of methodologies, stakeholders can create an environment conducive to trust and collaboration. Transparency not only enhances the credibility of AI-driven insights but also encourages the sharing of best practices and lessons learned, ultimately leading to more effective and innovative emission reduction strategies.

In conclusion, the integration of Explainable AI into emission reduction decision-making processes marks a significant advancement in the quest for effective climate action. By addressing the challenges associated with traditional AI models and fostering a culture of transparency and accountability, XAI has the potential to empower stakeholders to make informed decisions that contribute to global climate goals. As the world continues to grapple with the pressing challenges of climate change, the collaborative efforts of policymakers, researchers, and industry leaders will be essential in realizing the full promise of XAI, paving the way for a more sustainable and resilient future. The ongoing exploration of XAI's capabilities will not only enhance emission reduction strategies but also set a precedent for the responsible use of AI in addressing complex global issues.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Gunning, D, Aha, DW. DARPA's explainable artificial intelligence program. AI Magazine, 2019, 40(2): 44-58.
[2] Murdoch, WJ, Singh, C, Kumbier, K, et al. Definitions, methods, and applications in interpretable machine learning. Proceedings of the National Academy of Sciences, 2019, 116(44): 22071-22080.
[3] Fuss, S, Canadell, JG, Peters, GP, et al. Betting on negative emissions. Nature Climate Change, 2014, 4(10): 850-853.
[4] Wang, X, Wu, YC, Zhou, M, et al. Beyond surveillance: privacy, ethics, and regulations in face recognition technology. Frontiers in big data, 2024, 7, 1337465.
[5] Ma, Z, Chen, X, Sun, T, et al. Blockchain-Based Zero-Trust Supply Chain Security Integrated with Deep Reinforcement Learning for Inventory Optimization. Future Internet, 2024, 16(5): 163.
[6] Wang, X, Wu, YC, Ma, Z. Blockchain in the courtroom: exploring its evidentiary significance and procedural implications in US judicial processes. Frontiers in Blockchain, 2024, 7, 1306058.
[7] Wang, X, Wu, YC, Ji, X, et al. Algorithmic discrimination: examining its types and regulatory measures with emphasis on US legal practices. Frontiers in Artificial Intelligence, 2024, 7, 1320277.
[8] Chen, X, Liu, M, Niu, Y, et al. Deep-Learning-Based Lithium Battery Defect Detection via Cross-Domain Generalization. IEEE Access, 2024, 12, 78505-78514.
[9] Liu, M, Ma, Z, Li, J, et al. Deep-Learning-Based Pre-training and Refined Tuning for Web Summarization Software. IEEE Access, 2024, 12, 92120-92129.
[10] Li, J, Fan, L, Wang, X, et al. Product Demand Prediction with Spatial Graph Neural Networks. Applied Sciences, 2024, 14(16): 6989.
[11] Asif, M, Yao, C, Zuo, Z, et al. Machine learning-driven catalyst design, synthesis and performance prediction for CO2 hydrogenation. Journal of Industrial and Engineering Chemistry, 2024. DOI: https://doi.org/10.1016/j.jiec.2024.09.035.
[12] Lin, Y, Fu, H, Zhong, Q, et al. The influencing mechanism of the communities' built environment on residents' subjective well-being: A case study of Beijing. Land, 2024, 13(6): 793.
[13] Sun, T, Yang, J, Li, J, et al. Enhancing Auto Insurance Risk Evaluation with Transformer and SHAP. IEEE Access, 2024, 12, 116546-116557.
[14] Srivastava, N, Hinton, G, Krizhevsky, A, et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
[15] Le, QV, Ranzato, MA, Monga, R, et al. Building High-level Features Using Large Scale Unsupervised Learning. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015, 8595-8599.
[16] Shapley, LS. A value for n-person games. Contributions to the Theory of Games, 1953, 2(28): 307-317.

[17] Vinuesa, R, Azizpour, H, Leite, I, et al. The role of artificial intelligence in achieving the Sustainable Development Goals. Nature Communications, 2020, 11(1): 1-10.
[18] Doshi-Velez, F, Kim, B. Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608. 2017. DOI: https://doi.org/10.48550/arXiv.1702.08608.
[19] Guo, W, Zhao, Y, Lu, H. Big data analytics for concept drift detection in non-stationary data streams. In 2019 IEEE International Conference on Big Data (Big Data), 2019, 2362-2371.
[20] Creutzig, F, Roy, J, Lamb, WF, et al. Towards demand-side solutions for mitigating climate change. Nature Climate Change, 2018, 8(4): 260-263.
[21] Obermeyer, Z, Powers, B, Vogeli, C, et al. Dissecting racial bias in an algorithm used to manage the health of populations. Science, 2019, 366(6464): 447-453.
[22] Gass, V, Schmidt, J, Strauss, F, et al. Assessing the economic wind power potential in Austria. Energy Policy, 2013, 53, 323-330.
[23] Kroll, JA. The fallacy of inscrutability. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2018, 376(2133): 20180084.
[24] Lundberg, SM, Lee, SI. A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA. 2017, 4768-4777.
[25] Eckstein, D, Künzel, V, Schäfer, L, et al. Global Climate Risk Index 2020. Bonn: Germanwatch. 2019.
[26] Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 2019, 1(5): 206-215.
[27] Gilpin, LH, Bau, D, Yuan, BZ, et al. Explaining explanations: An overview of interpretability of machine learning. In 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA) (pp. 80-89). IEEE. 2018.
[28] Carleton, TA, Hsiang, SM. Social and economic impacts of climate. Science, 2016, 353(6304): aad9837.
[29] Adadi, A, Berrada, M. Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI). IEEE Access, 2018, 6, 52138-52160.
[30] Gilvary, C, Madhukar, N, Elkhader, J, et al. The missing pieces of artificial intelligence in medicine. Trends in Pharmacological Sciences, 2020, 41(8): 555-564.
[31] Hao, K. The AI gurus are leaving Big Tech to work on climate change. MIT Technology Review. 2018.
[32] Papernot, N, McDaniel, P. Deep k-nearest neighbors: Towards confident, interpretable and robust deep learning. arXiv preprint arXiv:1803.04765. 2018. DOI: https://doi.org/10.48550/arXiv.1803.04765.
[33] Bauer, N, Calvin, K, Emmerling, J, et al. Shared socio-economic pathways of the energy sector–quantifying the narratives. Global Environmental Change, 2017, 42, 316-330.
[34] Wang, X, Wu, YC. Balancing innovation and Regulation in the age of geneRative artificial intelligence. Journal of Information Policy, 2024, 14. DOI: https://doi.org/10.5325/jinfopoli.14.2024.0012.
[35] Hastie, T, Tibshirani, R, Friedman, J. The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media. 2009. DOI: https://doi.org/10.1007/978-0-387-84858-7.
[36] Kaur, H, Nori, H, Jenkins, S, et al. Interpreting Interpretability: Understanding Data Scientists' Use of Interpretability Tools for Machine Learning. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 2020, 1-14. DOI: https://doi.org/10.1145/3313831.3376219.
[37] Zednik, C. Solving the black box problem: A normative framework for explainable artificial intelligence. Philosophy & Technology, 2019, 1-24.
[38] Langer, M, Oster, D, Speith, T, et al. What do we want from Explainable Artificial Intelligence (XAI)?–A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research. Artificial Intelligence, 2021, 296, 103473.
[39] Lipton, ZC. The mythos of model interpretability. Queue, 2018, 16(3): 31-57.
[40] Montavon, G, Samek, W, Müller, KR. Methods for interpreting and understanding deep neural networks. Digital Signal Processing, 2018, 73, 1-15.
[41] Rolnick, D, Donti, PL, Kaack, LH, et al. Tackling climate change with machine learning. ACM Computing Surveys, 2019, 55(2): 1-96. DOI: https://doi.org/10.1145/3485128.
[42] Barredo Arrieta, A, Díaz-Rodríguez, N, Del Ser, J, et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 2020, 58, 82-115.
[43] Lapuschkin, S, Wäldchen, S, Binder, A, et al. Unmasking Clever Hans predictors and assessing what machines really learn. Nature Communications, 2019, 10(1): 1-8.
[44] Mitchell, M, Wu, S, Zaldivar, A, et al. Model cards for model reporting. In Proceedings of the Conference on Fairness, Accountability, and Transparency, 2019, 220-229. DOI: https://doi.org/10.1145/3287560.3287596.
[45] Strobelt, H, Gehrmann, S, Pfister, H, et al. LSTMVis: A tool for visual analysis of hidden state dynamics in recurrent neural networks. IEEE Transactions on Visualization and Computer Graphics, 2018, 24(1): 667-676.
[46] Guidotti, R, Monreale, A, Ruggieri, S, et al. A survey of methods for explaining black box models. ACM Computing Surveys, 2018, 51(5): 1-42.
[47] Holzinger, A, Biemann, C, Pattichis, CS, et al. What do we need to build explainable AI systems for the medical domain? arXiv preprint arXiv:1712.09923. 2017. DOI: https://doi.org/10.48550/arXiv.1712.09923.
[48] Arrieta, AB, Díaz-Rodríguez, N, Del Ser, J, et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies,

opportunities and challenges toward responsible AI. Information Fusion, 2020, 58, 82-115.
[49] Vaughan, J, Wallach, H. A human-centered agenda for intelligible machine learning. Machines We Trust: Perspectives on Dependable AI. MIT Press. 2020. DOI: https://doi.org/10.7551/mitpress/12186.003.0014.
[50] Srivastava, M, Heidari, H, Krause, A. Mathematical notions vs. human perception of fairness: A descriptive approach to fairness for machine learning. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, 2459-2468. DOI: https://doi.org/10.1145/3292500.3330664.
[51] Samek, W, Wiegand, T, Müller, KR. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. arXiv preprint arXiv:1708.08296. 2023. DOI: https://doi.org/10.48550/arXiv.1708.08296.
[52] Das, A, Rad, P. Opportunities and challenges in explainable artificial intelligence (XAI): A survey. arXiv preprint arXiv:2006.11371. 2024. DOI: https://doi.org/10.48550/arXiv.2006.11371.
[53] Zuo, Z, Niu, Y, Li, J, et al. Machine Learning for Advanced Emission Monitoring and Reduction Strategies in Fossil Fuel Power Plants. Applied Sciences, 2024, 14(18): 8442.
[54] Cheng, HF, Wang, R, Zhang, Z, et al. Explaining decision-making algorithms through UI: Strategies to help non-expert stakeholders. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 2019, 559, 1-12. DOI: https://doi.org/10.1145/3290605.3300789.
[55] Miller, T. Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence, 2019, 267, 1-38.
[56] Ribeiro, MT, Singh, S, Guestrin, C. "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, 2016, 1135-1144. DOI: https://doi.org/10.1145/2939672.2939778.
[57] Dubey, A, Naik, N, Parikh, D, et al. Deep learning the city: Quantifying urban perception at a global scale. In Proceedings of the European Conference on Computer Vision (ECCV), 2016, 9905, 196-212. DOI: https://doi.org/10.1007/978-3-319-46448-0_12.
[58] Vafa, K, Naidu, S, Blei, D. Text-based ideal points. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, 5345-5357. DOI: https://doi.org/10.48550/arXiv.2005.04232.
[59] Wang, D, Yang, Q, Abdul, A, et al. Designing theory-driven user-centric explainable AI. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 2019, 601, 1-15. DOI: https://doi.org/10.1145/3290605.3300831.
[60] Wachter, S, Mittelstadt, B, Russell, C. Counterfactual explanations without opening the black box: Automated decisions and the GDPR. Harvard Journal of Law & Technology, 2017, 31(2): 841-887.
[61] Yang, K, Qinami, K, Fei-Fei, L, et al. Towards fairer datasets: Filtering and balancing the distribution of the people subtree in the ImageNet hierarchy. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 2020, 547-558. DOI: https://doi.org/10.1145/3351095.3375709.
[62] Díaz-Rodríguez, N, Lamas, A, Sanchez, J, et al. EXplainable Neural-Symbolic Learning (X-NeSyL) methodology to fuse deep learning representations with expert knowledge graphs: The MonuMAI cultural heritage use case. Information Fusion, 2022, 79, 58-83.
[63] Sundararajan, M, Taly, A, Yan, Q. Axiomatic attribution for deep networks. In Proceedings of the 34th International Conference on Machine Learning, 2017, 70, 3319-3328. DOI: https://doi.org/10.48550/arXiv.1703.01365.