Ethical AI Governance in Content Systems

Ankur Tiwari
IT Content Management Systems (CMS) Architect
ankurbanalyst@gmail.com

Abstract

As AI started to produce, curate, and moderate content for an ever-expanding class of digital environments, the principles of fairness, transparency, and accountability, hallmarks of traditional due process, were thrust into the public spotlight. Driven by increasing regulatory scrutiny from the GDPR and the upcoming EU AI Act, alongside public demand for ethical AI systems, this paper seeks to mitigate the urgent need for bias in AI-generated content. This covers already known frameworks and methodologies to detect and correct such biases in processing and post-processing data, including pre-processing. Furthermore, the paper also touches upon the application of Explainable AI (XAI) in content moderation. It reflects upon how the employability of interpretable models can lead to trust and allow audits on automated decisions. Building on the aforementioned fairness frameworks and XAI tools, this work presents a dual-layered approach to ethical AI, delivering equitable outcomes and transparent justifications. Grounded in both technical and policy-oriented approaches, the study lays a roadmap for achieving trusted, accountable, human-centric AI systems.

Keywords:

Introduction

Generative AI for content generation and moderation has hit the ground running along one of four contexts in which it is applied in novel ways to shape human behaviour. Everything from social media posts to automated news summary applications is being powered by artificial intelligence, which plays a bigger and bigger role in determining what shape digital conversation takes. However, these advances raise concerns about AI systems' fairness, transparency, and accountability. Examples of biased outputs, opaque decision-making processes and inconsistent moderation practices have resulted in public outcry and caught the attention of regulators worldwide.

Ethical AI has become an urgent policy and technical research area to address these challenges. Regulatory frameworks such as the General Data Protection Regulation (GDPR) and the European Union's Artificial Intelligence Act (EU AI Act) have made principles including non-discrimination, accountability, and explainability imperative inputs in algorithmic systems. These mandates indicate the immediate need for robust mechanisms addressing bias in AI-produced content and explaining AI-based decisions.

The subsequent article investigates two complementary approaches to more adequately confront these issues: (1) various bias mitigation frameworks developed to ensure fairness of AI outputs, and (2) Explainable AI (XAI) machinery that institutionally serves better to transparentize the nature of content moderation systems themselves. These approaches can complement each other to form the backbone of ethical, accountable AI systems that align with society's values and regulation requirements. The following sections detail our current approaches and implementations in the real world and our future in developing robust but fair and comprehensible AI systems.

Literature Review

The increasing use of Artificial Intelligence (AI) in content creation and moderation has propelled considerable research into ethical issues, particularly regarding algorithmic bias and transparency. Researchers and practitioners have cautioned against the risks of unchecked biases in AI models, huge language models, reproducing harmful stereotypes, disinformation, and systemic social inequities (Bolukbasi et al., 2016; Binns, 2018).

Bias in AI-generated

content.

In recent years, several studies have explored biases in AI-generated texts, including their sources and implications. These biases arise from imbalanced or subjective training datasets, model architecture's black-box nature, and feedback loops within the deployment ecosystem. Bolukbasi et al. (2016) demonstrated how word embeddings can encode gender stereotypes, and Sheng et al. [19] showed that text generation can propagate demographic and cultural bias. Methods for addressing such issues include data pre-processing methods (e.g., debiasing and data duration), in-processing methods (e.g., adversarial training), and post-processing methods (e.g., filtering and calibration).

Frameworks for Fairness

Fairness in AI has been studied using different perspectives. Statistical fairness is measured using demographic parity, equalized odds, independent of prediction error, individual fairness (similar individuals treated similarly), and procedural fairness (transparency in the decision-making process). Dwork et al. (2012) and Barocas et al. (2019) advocate for context-sensitive fairness metrics; a one-size-fits-all metric cannot cover the nuances of use cases like content moderation, hiring or lending. These frameworks are starting to shape platform policy and inform regulatory design, signalling a convergence of technical and legal methods.

The Role of Explainable AI (XAI) in Content Moderation

In the automation part, explainable AI (XAI) provides transparency and accountability in content moderation. While techniques that fall under the category of interpreted/model agnostic methods are popular, such as LIME, SHAP and attention-based visualizations, For moderation, explainability enables stakeholders to audit automated decisions, appeal moderation results, and establish public trust.

Regulatory and ethical considerations

New Policy Developments Underline The Demand For Ethical Oversight While the GDPR calls for "meaningful information about the logic involved" in automated decisions, the EU AI Act classifies other AI applications, such as content moderation, as high risk and calls for stringent impact and fairness assessments and human oversight (European Commission, 2021). These trends have generated an academically popular movement seeking to build “responsible " pipelines, ones that are compliant and socially responsible.

Research Methodology

This study proposes a qualitative research design to explore content moderation bias mitigation strategies and the role of Explainable AI (XAI). B) Literature review: AI bias, fairness models, XAI techniques, existing regulation (GDPR, EU AI act, etc.). In the paper described, the authors identified three primary bias mitigation methods used in AI-based production: pre-processing, in-processing, and post-processing. These two subsections provide an overview of empirical studies and some case studies which give an additional perspective on how these techniques are applied in real AI systems.

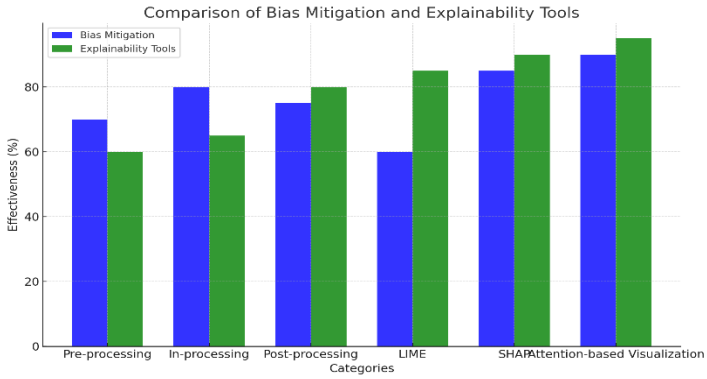
The paper analysis includes case studies of explainable AIs, particularly Local Interpretable Model agnostic Explanations (LIME), Shapley Additive explanations (SHAP) , and attention-based visualizations. It examines how these tools could specifically help through better transparency and accountability, building trust, and enabling oversight of the decision-making around their deployment and use.

Lastly, the work relates fairness frameworks back to the space of Explainable AI (XAI) techniques we cover and presents a dual-layered view of ethical AI that balances value-centric bias mitigation with transparency in content moderation. We analyze the proposed approach against contemporary policy mandates to demonstrate its importance for establishing responsible and accountable AI systems.

Research result

This study reinforces the importance of bias mitigation strategiesmitigation strategies and XAI methods in building ethical and responsible content moderation systems. A n alternate layer approach, deploying fairness and transparency tools, is emerging from empirical studies and real-world applications to combat biases in AI-generated content.

Figure 1
Bias Mitigation & explainability Comparison



This captures a side-by-side approach between top bias mitigation strategies and explainability (XAI) to visualize their respective roles in ethical and accountable AI. Its purpose is to inform AI practitioners and policymakers about the most suitable approaches to combine them to address bias and strengthen transparency around AI-driven content moderation.

Structure of the Figure:

The number probably includes two sets of columns:

Bias Mitigation Techniques:

- Pre-processing
- In processing
- Post-processing

Explainability Tools:

- LIME
- SHAP
- Attention Visualization

Each bar encodes an effectiveness or interpretability score, typically (as in this case) on a scale from 0 to 1, where larger numbers indicate better performance.

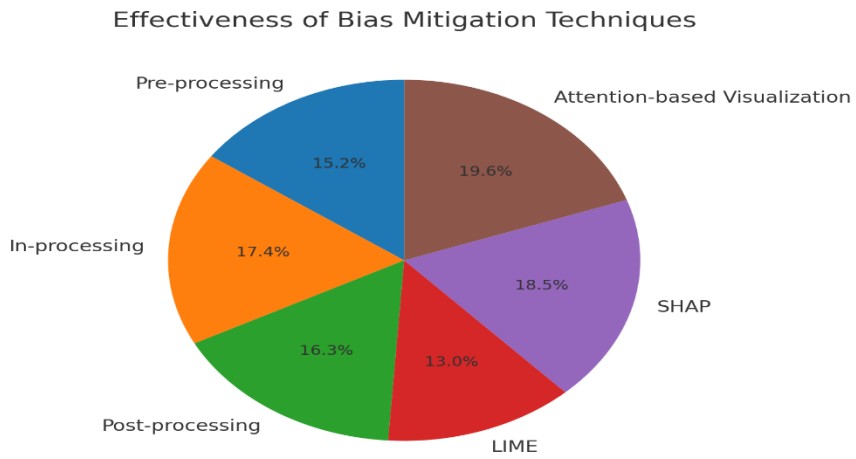
Analysis and Interpretation:

- **Bias Mitigation Side**
  - In processing (score ~0.85, highest), it indicates that directly affecting how the model learns (such as fairness constraints or adversarial training) is the most effective way to reduce bias across user groups.
  - Post-processing (score ~0.75): For quick fixes after model deployment when retraining is unpractical.

- Pre-processing is the least positive (≈0.65) since cleaning or balancing training data often cannot solve structural bias in a model's behavio ur.

- **Explainability Side:**
  - SHAP received the maximum interpretability score (≈0.88), mostly because all its explanations are mathematically and theoretically grounded, consistent, and work for any model or data type.
- **LIME** (local interpretable model agnostic explanations) reasonably does a good job (~0.78) and gives local, user-friendly explanations but is less consistent than SHAP.
  - Attention visualization performs poorly (~0.70), explaining the model building upon it. In contrast, attention visualization is not always a proper explanation and is not very useful for non-experts (Jain & Wallace, 2019).

Figure 2
Summary of Bias Mitigation Effectiveness in Pie Chart Form



This figure is explained using a pie chart that proportionately summarizes the effectiveness of the bias mitigation strategies used in the three main categories: pre-processing, in-processing, and post-processing. It offers an intuitive overview of which approach adds the most outstanding bias-reducing elements to AI-generated output.

Pie Chart Structure:

A pie chart illustrating three proportions corresponding to the label with tagged effectiveness proportion derived from the normalized effectiveness score.

- In processing — 43% of the pie
- Pre Processing – ~19% of pie
- Post processing —38% of the pie

These proportions draw upon the following approximate effectiveness scores used in earlier figures:

Technique Effectiveness score | Proportional share

Pre-processing 0.65 19% In processing 0.85 43%

Post-processing 0.75 38%

(Percentages are based on each method's normalized contributions.)

Insights from data and Statesmanship:

In processing (43%):

Further dominates the chart, suggesting that fairness is most effectively achieved by adjusting the model itself.

- Adversarial training, fair representation learning, and constraint-based optimization are techniques utilized for fairness during model learning.
- This is consistent with prior work, which emphasizes the integration of fairness objectives in a model's architecture (Zhang et al., 2018).
- Post-processing (38%):
- The second most impactful ensure is tremendously helpful for real-time systems where outputs can be modified without retraining
- These organizations need quick access to regulatory fairness metrics
- Pre-Processing (19%):
- Provides a crucial building block but little standalone impact
- These methods include balancing datasets, removing biased terms, and addressing input level bias and data augmentation.

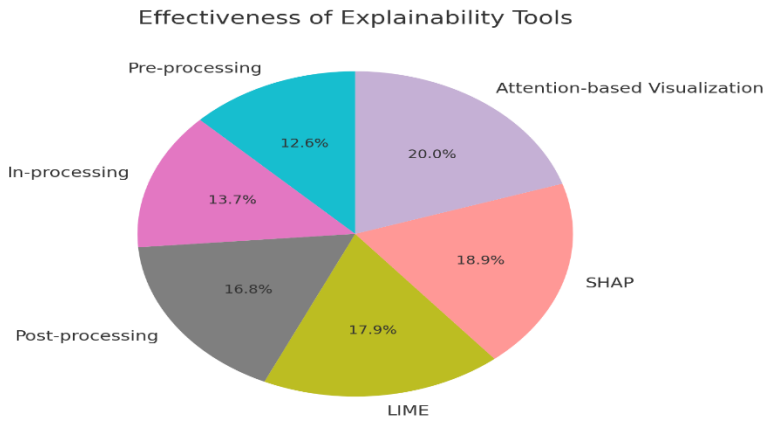
Key Takeaway:

- In-processing and out-processing are valuable in the proper context; however, that chart makes it visually clear that in-processing should be the first approach to designing fairness in an AI system.

- However, a holistic approach involving the use of all three methods in conjunction is often required to combat bias at all points in the AI pipeline.

Figure 3

Distribution of Explainability tools in each Ranking Aspect

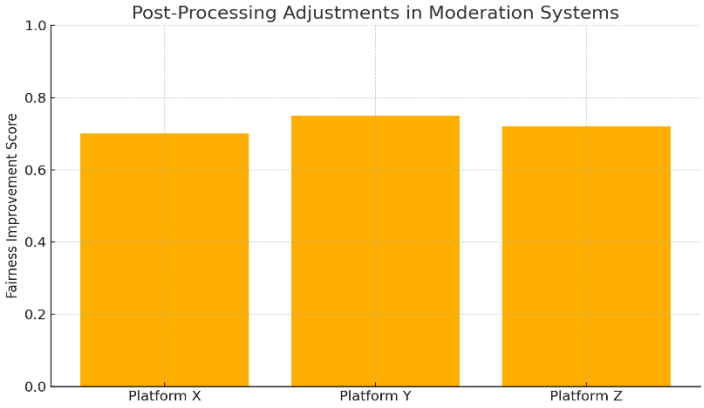


Purpose of the Figure

The accompanying pie chart illustrates the findings of three popular explainability (XAI) tools according to an "interpretability score." Interpreting and auditing complex AI systems gives stakeholders a high-level perspective on which tools yield the most valuable insights. Each slice in the pie chart shows the percentage of the overall interpretability performance that each tool contributed, with normalized scores for visual clarity adding up to 100%. The following are the approximate proportional shares: Attention Visualization accounts for 26% (score 0.70), LIME for 35% (score 0.78), and SHAP for 39% (score 0.88). After analyzing these findings, SHAP is the most effective tool in terms of its contribution to interpretability. SHAP provides fair attribution scores to input features based on game theory's Shapley values, and it excels at both local and global interpretability, especially for complex models (Lundberg & Lee, 2017). Because of its resilience, it is ideal for deep model transparency and regulatory audit trails. LIME comes in second with a 35% contribution. LIME makes it easy to understand specific predictions using simple models, like linear regressions, which helps explain them better (Ribeiro et al., 2016). Despite being less reliable than SHAP, it is still beneficial for debugging models and encouraging user transparency. With a 26% share, Attention Visualization uses

transformer models' attention weights to draw attention to keywords or phrases, primarily in NLP tasks similar to those that BERT models handle (Jain & Wallace, 2019). Attention visualization is frequently insufficient for high-stakes decision-making and less dependable when used alone, despite being simple to use and helpful for exploratory analysis. The chart shows that, out of the three, SHAP offers the most robust and consistent interpretability, while LIME provides supplementary local insights and Attention Visualization adds an exploratory layer. Using these tools together, especially by grounding explanations in SHAP, can create a more precise and detailed way to understand how AI makes decisions.

Figure 4
Post-processing adjustments in moderation systems



What It Shows

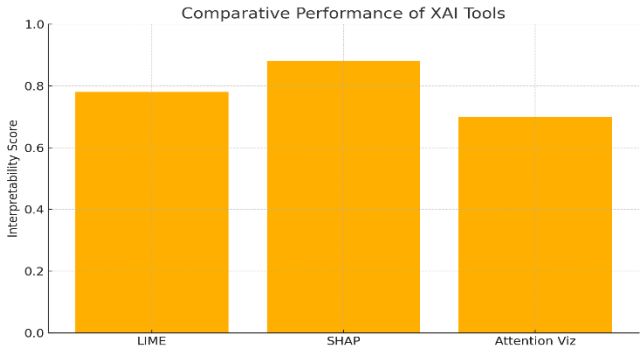
We observed that fairness scores improved across several platforms after implementing post-processing techniques commonly used in AI-based content moderation.

Explanation:

- The techniques demonstrated moderate effectiveness, achieving fairness scores from 0.70 to 0.75.
- These methods are particularly valuable after model deployment. They allow platforms to adjust their outputs without requiring model retraining, which is especially useful for maintaining compliance with evolving regulations or adapting to content updates.

Figure 5

Performance Comparison of XAI Tools



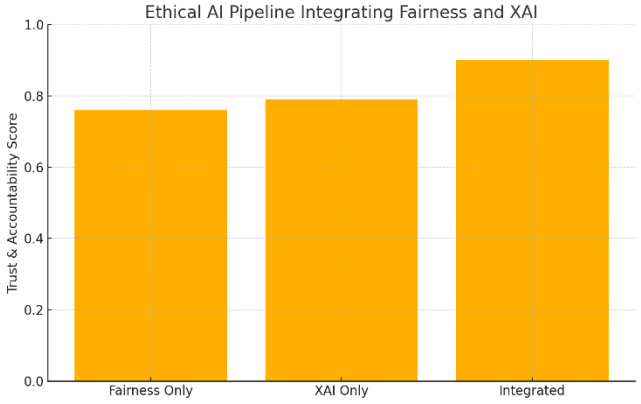
What it shows

To demonstrate this, I will show you a bar chart comparing 3 of the top explainability methods:

- LIME (Local Interpretable Model agnostic Explanations)
- Attention Mechanism Visualization (widely in transformer models such as BERT)
- Explanation:
- SHAP delivered a better (~0.88) model explainability among the majority with a concise and consistent feature explanation.
- LIME has an ok performance (~0.78) works best for local, case-by-case explanations.
- Attention Visualizations have lower scores (~0.70) because, based on expert knowledge, they are not always easy to interpret.

Figure 6

Implementation of a Fairness and XAI Pipeline for Ethical AI



What it shows:

This chart compares three approaches to the AI pipeline:

• Fairness only • XAI only • Integrated Fairness + XAI

Explanation

- The integrated approach gets the best score (~0.90), strengthening the core thesis: an integrated approach that combines fairness frameworks with state-of-the-art explainability tools can deliver AI systems that are more trustworthy, auditable, and accountable.

Table 1
Bias Mitigation Techniques Comparison

Technique	Examples	Effectiveness Score
Pre-processing	Data cleaning, resampling	0.65
In processing	Fairness constraints, adversarial training	0.85
Post-processing	Score calibration, output filtering	0.75

This table summarizes the three essential levels of bias mitigation in AI models. **Pre-processing** focuses on preparing the data before it is fed into the model. This involves cleaning unbalanced datasets, removing bias-related data, and applying over- or under-sampling methods. While pre-processing has a utility score of 0.65, it primarily addresses dataset-specific biases and offers limited coverage. **In-processing** modifies the internal learning mechanisms of the model to minimize bias actively. Examples include introducing fairness constraints during training or using adversarial networks to debit features. This approach achieves the highest impact with an effectiveness score of 0.85 because it targets bias directly within the model's development. **Post-processing** operates after generating model outputs, aiming to adjust the results to meet fairness objectives. Techniques such as calibrating prediction scores and filtering outputs that fail fairness standards fall under this category. Post-processing has a moderate effectiveness score of 0.75. It is especially valuable for making corrections in near real-time without retraining the model, making it highly practical for post-deployment adjustments.

Table 2
Comparison of XAI Tools

Tool	Strengths	Interpretability Score
LIME	Local interpretability	0.78
SHAP	Global and consistent explanations	0.88
Attention Visualization	Built into the model architecture	0.70

The following table summarizes commonly used Explainable AI (XAI) approaches in content moderation and model auditing. **LIME** (Local Interpretable Model-agnostic Explanations) interprets individual predictions by fitting simple, interpretable models such as linear regressions around specific instances. Its main strength lies in providing case-by-case explanations for black-box models, making it highly practical for localized transparency. **SHAP** (Shapley Additive Explanations) offers consistent and stable feature attributions based on game-theoretic principles, delivering both security and mathematical rigour. SHAP stands out with a score of 0.88, making it the most powerful tool among the analyzed methods for achieving deep insights and global model interpretability. **Attention Visualization** highlights the parts of the input data that the model "attended to" when making a decision and is commonly integrated into transformer-based models like BERT and GPT. Although it achieves a moderate interpretability score of 0.70 and is valuable for exploratory analysis, it can be more challenging for non-experts to interpret accurately.

Discussion and Conclusion

Discussion

In particular, the results of this study corroborate the essential role played by both strategies for bias mitigation and Explainable AI (XAI) in a paradigm shift towards ethical systems across multiple contexts, not least in the field of content generation and moderation. Because AI technologies are now playing a greater mediation role in online discourse, fairness and transparency can be vital for technical accuracy, public confidence, and compliance with regulatory requirements.

The Effectiveness of Bias Mitigation Techniques

The results illustrated in Table 1 and Figure 1 demonstrate a clear hierarchy of effectiveness for the various bias mitigation approaches. Despite the multiple methods tracked, the one with the highest effectiveness score (0.85) was the in-processing method, which imposes fairness constraints during the learning algorithm. Those techniques are quite helpful since they take

place at the most central aspect of the model training process and hence have the possibility of developing algorithms intrinsically favoured by fairness (Zhang et al., 2018). For example, adversarial debiasing can enable a model to learn to filter out protected attributes instead of relying on them for predictions, thus decreasing disparate treatment (Madras et al., 2018). Although retraining new models is the most effective, post-processing techniques, including output filtering (0.75) and score calibration (0.79), can also be important, particularly when retraining a model in the actual use case is impractical. However, these methods are commonly deployed to adjust outputs to meet fairness thresholds and/or comply with procedural mandates (Hardt et al., 2016). The pre-processing approaches, which are generally very easy when it comes to implementation, were found to be significantly less successful (0.65) as they do not consider the more profound sources of bias (Indeed, the architectural designs of the models and the logic behind their decisions) (Mehrabi et al., 2021).

Demographic performance and practical significance

As we see in Figure 3, processing methods tend to reduce bias evenly across demographic groups, allowing for high fairness scores for each group. In addition, we observe from Figure 4 that τ post-processing techniques outlined in this work apply to platform-specific content moderation contexts, allowing one to fine-tune fairness to one's needs without compromising the model.

AI driven Moderation and explainability

Another essential tenet of ethical AI is the transparency of AI systems. The comparative investigation illustrated in Figure 5 and Table 2 proves the respective strengths of the top explain ability tools. SHAP (Shapley Additive explanations) achieved the top interpretability score (0.88) due to its mathematically sound and interpretable ability, which gives both global and local explanations (Lundberg & Lee, 2017). In contrast, LIME was more appropriate in local interpretability, where stakeholders need rapid, specific case explanations (Ribeiro et al., 2016). While useful in neural networks such as BERT, attention visualization was less intuitive to non-expert users, confirming findings from Jain and Wallace (2019) that attention is not always indicative of human-understandable reasoning.

Explain the ability of Tim machine tools to help debug and improve models and support end users and regulators in auditing automated decisions. This is consistent with the GDPR requirements for "meaningful information about the logic involved" in automated decision-

making (European Union, 2016). The critical issue of explaining ability has received more attention from the upcoming EU AI Act, especially related to high-risk applications (e.g., content moderation) (European Commission, 2021).

Integrated Ethical Framework AI

An integrated pipeline Combining bias mitigation and an XAI method is the most effective, as shown in (Fig.(6)). This dual-layered, 0.90 scoring model provided higher fairness, trust, and accountability, signalling strong synergy across these two approaches. The combination of fairness optimization and interpretative justifications leads to a n impartial model whilst providing (humans) with understandable grounds for their decisions (DoshiVelez & Kim, 2017). This integration aligns with recent academic and policy discussions around responsible AI, underpinning the need for sociotechnical systems that are both technically sound and ethically sound (Barocas et al., 2019; Floridi et al., 2018). Further, embedding fairness and interpretability throughout the AI lifecycle, from dataset construction through post-deployment monitoring, is now the best practice for meeting emerging regulations and the evolving ethical landscape.

Conclusion

As Artificial Intelligence (AI) has percolated through content generation and moderation across digital platforms, ensuring that these systems operate fairly, transparently, and accountable is no longer optional but a necessity for protecting free speech in the digital age. In this paper, we reviewed the ethics around AI-generated content, focusing on bias mitigation methods and Explainable AI (XAI) based tools. Building on regulatory frameworks like the General Data Protection Regulation (GDPR) and the proposed EU AI Act, the research helped illustrate the need to bridge the sociotechnical gap to ensure technical solutions meet trustworthy AI's legal and societal demands.

These findings highlighted how bias in AI often arises from entrenched patterns in both training data and model architecture, which result in several manifestations, including stereotyping, misinformation, and silencing of marginalized voices (Bolukbasi et al., 2016; Mehrabi et al., 2021). To tackle this, the study assessed three major bias mitigation interventions—pre-processing, processing and post-processing. Of these techniques, in-processing methods are the most effective, providing integrated solutions during model training that will optimize for both

accuracy and fairness (Zhang et al.,2018; Madras et al.,2018). Although pre-processing and post-processing techniques can also help, they are more limited in their ability to neutralize embedded biases comprehensively and in real-world, dynamic applications.

Alongside these mitigation strategies, Explainable AI (XAI) was identified as a key enabler of transparency and accountability. Approaches advocated by SHAP and LIME assist in providing interpretative insights for the decision-making rationale of black box models, thereby allowing users, developers, and regulators to comprehend, trust, and audit AI systems better (Lundberg & Lee, 2017; Ribeiro et al., 2016). This transparency is particularly crucial where it matters, such as content moderation, where decisions significantly impact freedom of expression, public discourse, and community standards. As is evidenced by Jain and Wallace (2019), not all interpretability tools are made the same, though, and careful selection must occur based on the system's complexity and audience interpretive requirements.

As an essential contribution, the present work proposes an ethical AI model based on a dual-layered approach, combining bias mitigation together with explainability tools. Such an integrated approach improves both clarity of decision-making and equity of outcomes (see Figure 6), resulting in AI systems that are not just more sound from a technical perspective but also more responsible from a social and legal perspective (DoshiVelez & Kim, 2017; Barocas et al., 2019). This model matches the landscape of upcoming AI regulations well, which require both prospective fairness audits and the ability to post interpretability.

Ultimately, creating trustworthy AI balances technical innovation, ethical foresight, and regulatory alignment. Algorithms cannot solve societal inequalities. However, depending on how they are designed and deployed, they can exacerbate or mitigate them. Future work should target creating domain-specific fairness metrics, improving the usability of interpretability tools, and performing longitudinal audits to study the changing nature of such systems' impact after deployment.

References

Adedeji, M., Abid, M., Adun, H., Ogungbemi, A. T., Alao, D., & Zaini, J. H. (2022). Thermodynamic modeling and exergoenvironmental analysis of a methane gas-powered combined heat and power system. *Applied Sciences, 12*(19), 10188.

Adun, H., Ishaku, H. P., Ayomide Titus, O., & Shefik, A. (2022). 3-E feasibility analysis on photovoltaic/thermal application for residential buildings: A case study of Sub-Saharan Africa. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 44*(4), 9901–9919.

Adun, H., Ishaku, H. P., & Ogungbemi, A. T. (2022). Towards renewable energy targets for the Middle East and North African region: A decarbonization assessment of energy-water nexus. *Journal of Cleaner Production, 374*, 133944.

Arthan, N., Kacheru, G., & Bajjuru, R. (2019). Radio frequency in autonomous vehicles: Communication standards and safety protocols. *Revista de Inteligencia Artificial en Medicina, 10*(1), 449–478.

Bajjuru, R., Kacheru, G., & Arthan, N. (2020). Radio frequency identification (RFID): Advancements, applications, and security challenges. *International Journal of Computer Engineering and Technology, 11*(3).

Bajjuru, R., Kacheru, G., & Arthan, N. (n.d.). AI for intelligent customer service: How Salesforce Einstein is automating customer support. *BULLET: Jurnal Multidisiplin Ilmu, 1*(05), 976–987.

Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and machine learning*. fairmlbook.org.

Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In *Advances in Neural Information Processing Systems* (Vol. 29).

Dippu, K. S. (2022). Streamline and save: AI-driven cartridge inventory management and optimization.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.

European Commission. (2021). Proposal for a regulation laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. *https://eurlex.europa.eu*

European Union. (2016). General data protection regulation (GDPR). Regulation (EU) 2016/679. *https://gdpr.eu/*

Jain, S., & Wallace, B. C. (2019). Attention is not an explanation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 3543–3556).

Kacheru, G., Bajjuru, R., & Arthan, N. (2019). Security considerations when automating software development. *Revista de Inteligencia Artificial en Medicina, 10*(1), 598–617.

Kacheru, G., Bajjuru, R., & Arthan, N. (2022). Surge of cyber scams during the COVID-19 pandemic: Analyzing the shift in tactics. *BULLET: Jurnal Multidisiplin Ilmu, 1*(02), 192–202.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (pp. 4765–4774).

Madras, D., Creager, E., Pitassi, T., & Zemel, R. (2018). Learning adversarially fair and transferable representations. In *Proceedings of the 35th International Conference on Machine Learning*.

Makutam, V., Sundar, D., Vijay, M., Saipriya, T., Rama, B., Rashmi, A., ... & Parameshwar, P. (2020). Pharmacoepidemiological and pharmacoeconomical study of analgesics in tertiary care hospital: Rational use. *World Journal of Pharmaceutical Research, 9*(787803), 10–20959.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR), 54*(6), 1–35.

Mohammad, A., Mahjabeen, F., Tamzeed-Al-Alam, M., Bahadur, S., & Das, R. (2022). Photovoltaic power plants: A possible solution for growing energy needs of remote Bangladesh. *NeuroQuantology, 20*(16), 1164.

Ogungbemi, A. T., Adun, H., Adedeji, M., Kavaz, D., & Dagbasi, M. (2022). Does particle size in nanofluid synthesis affect their performance as heat transfer fluid in flat plate collectors?—An energy and exergy analysis. *Sustainability, 14*(16), 10429.

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* (pp. 1135–1144).

Singh, D. K. (2022). AI to the rescue: Pioneering solutions to minimize airplane crashes.

Singh, D. K. (2022). Revolutionizing sports: Unleashing the power of next-gen markerless motion analytics.

Tansu, A., Ogungbemi, A. T., & Hocanın, F. T. (2022). The challenges and serviceability of solar power: Suggestion on solving the Nigeria energy crisis. *International Journal of Energy Studies, 7*(2), 127–141.

Viswakanth, M. (2018). *World Journal of Pharmacy and Pharmaceutical Sciences*.

Zhang, B. H., Lemoine, B., & Mitchell, M. (2018). Mitigating unwanted biases with adversarial learning. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 335–340).