COLOMBO SCIENTIFIC PUBLISHING

INTERNATIONAL JOURNAL OF ECONOMICS FINANCE & MANAGEMENT SCIENCE

E-ISSN: 2536-7897
P-ISSN: 2536-7889

PAGE NO: 53-60

Enhancing Welfare Governance through Trustworthy and Data-Centric AI: A Comprehensive Examination of Transparency, Bias Mitigation, and Policy Compliance

Raymond L. Whitlock
University of Vienna, Austria

ARTICLE INFO

Article history:

Submission: December 01, 2025
Accepted: December 16, 2025
Published: January 10, 2026
VOLUME: Vol.11 Issue 01 2026

Keywords:
Trustworthy AI, Data-Centric Governance, Welfare Management, Transparency, Bias Mitigation, Policy Compliance, Ethical AI

ABSTRACT

The integration of artificial intelligence (AI) into governance structures has emerged as a transformative approach to enhancing transparency, efficiency, and decision-making quality. Within the context of welfare management, the application of trustworthy AI models that are data-centric rather than solely model-centric presents both significant opportunities and intricate challenges. This study critically examines the deployment of AI in welfare governance frameworks, emphasizing the roles of transparency, bias mitigation, and compliance with regulatory and ethical policies. Building on recent theoretical advancements, particularly those outlined by Priyadarshi Uddandarao et al. (2026), the article elaborates on how data-centric AI models can reshape administrative decision-making, enhance stakeholder trust, and facilitate ethical oversight. Through an extensive review of multidisciplinary literature, the paper analyzes the mechanisms by which AI can reinforce governance efficacy while simultaneously exposing potential vulnerabilities related to algorithmic opacity, model biases, and policy misalignments. This research synthesizes insights from AI ethics, computational governance, and public policy studies, presenting a comprehensive framework for responsible AI adoption in welfare systems. Furthermore, the study explores critical debates surrounding automated decision-making, the tension between efficiency and equity, and the role of human oversight in algorithmically mediated governance. Methodologically, this research employs a systematic textual analysis of existing scholarly work, technical reports, and case studies to map the intersections between AI technology, governance principles, and public sector accountability. The findings demonstrate that while AI-driven welfare governance can yield substantial benefits in terms of procedural speed, accuracy, and inclusivity, the absence of rigorous data-centric protocols and auditing mechanisms can exacerbate systemic biases, reduce citizen trust, and contravene policy standards. By integrating empirical insights and theoretical discourse, this study offers a nuanced understanding of the conditions under which AI can operate responsibly within welfare management. The implications extend beyond welfare administration, offering strategic guidance for policymakers, technologists, and academic scholars seeking to balance innovation with ethical stewardship. Ultimately, this article provides a detailed conceptual roadmap for leveraging trustworthy AI to advance transparency, fairness, and compliance, emphasizing that the success of AI interventions in governance is contingent upon a holistic approach that harmonizes technological capabilities with socio-ethical imperatives.

INTRODUCTION

Artificial intelligence has evolved from a purely computational curiosity into a central pillar of contemporary governance and public administration. The capacity of AI to analyze large datasets, detect patterns, and support decision-making processes positions it as a powerful tool for enhancing welfare management systems, which are traditionally burdened by inefficiencies, opacity, and susceptibility to human biases (Gartner, 2018). Historically, welfare administration has relied on hierarchical bureaucratic mechanisms, where decisions are often influenced by subjective interpretation, institutional inertia, and procedural inconsistencies. The emergence of AI offers a paradigmatic shift: by enabling data-driven decisions, automating routine tasks, and providing predictive insights, AI can theoretically enhance both the efficiency and equity of welfare delivery (Hamid, 2022).

Nevertheless, the adoption of AI in governance raises critical questions regarding trustworthiness, fairness, and accountability. The dichotomy between model-centric and data-centric AI approaches has profound implications for welfare policy implementation. Model-centric approaches prioritize algorithmic sophistication, often emphasizing the refinement of predictive models while potentially overlooking data quality, representativeness, and bias mitigation (Hamid, 2023). In contrast, data-centric AI emphasizes the curation, cleaning, and ethical structuring of datasets, recognizing that the quality and governance of data are as determinative of outcomes as the complexity of the algorithm itself (Priyadarshi Uddandarao et al., 2026). This paradigm underscores the importance of integrating ethical oversight, transparency protocols, and policy alignment into AI deployment strategies.

A critical concern in welfare AI applications is the potential for systemic bias. Biases may emerge from historical inequalities encoded within administrative records, demographic misrepresentations, or algorithmic training processes that inadequately account for minority populations (Huang et al., 2023). Such biases, if unmitigated, threaten to reinforce structural inequities, undermining both the legitimacy of welfare programs and public trust in technological governance. Moreover, transparency—or the lack thereof—plays a decisive role in public perception and compliance monitoring. In complex AI systems, opacity can obscure decision rationales, preventing effective auditing and raising ethical concerns regarding accountability (Information Commissioner's Office, 2022).

Recent scholarship emphasizes the necessity of a multidimensional framework for AI in governance, integrating principles of explainability, fairness, and regulatory compliance. Uddandarao et al. (2026) advance the notion that data-centric AI governance models can operationalize these principles effectively. Their framework delineates mechanisms for continuous auditing, bias detection, and policy alignment, situating AI as both a technical and socio-ethical instrument. Complementary studies underscore the role of human oversight in algorithmic decision-making, advocating for hybrid governance models that balance automation with human judgment (Dunnmon et al., 2021; Oberste & Heinzl, 2023).

Despite the theoretical promise, practical challenges persist. Case studies of AI failures in high-stakes environments reveal that technical sophistication alone does not guarantee ethical outcomes. For instance, incidents such as the malfunctioning of automated robotics in public spaces demonstrate the unpredictability of AI behaviors under complex environmental conditions, necessitating robust risk management and continuous system monitoring (Hall, 2020). Furthermore, adversarial manipulation of AI systems, exemplified by attempts to deceive toxicity detection algorithms, highlights vulnerabilities that can undermine governance objectives (Hosseini et al., 2017). Such insights reinforce the imperative for resilient, transparent, and ethically anchored AI governance models.

This article seeks to address three interrelated gaps in the current literature. First, it investigates the operationalization of data-centric governance within welfare administration, examining how data quality, auditing practices, and trustworthiness criteria influence decision outcomes. Second, it explores mechanisms for mitigating algorithmic bias, particularly in the context of socially sensitive welfare services. Third, it evaluates policy compliance strategies, integrating insights from AI ethics, computational law, and public sector management. By synthesizing these dimensions, the study aims to provide a comprehensive conceptual and practical framework for responsible AI deployment in welfare governance, offering both theoretical enrichment and actionable guidance for policymakers and technologists alike.

The remainder of the paper is structured as follows. The methodology section elaborates on the qualitative and analytical strategies employed to examine governance models, bias mitigation frameworks, and compliance mechanisms. The results section presents descriptive analyses of identified patterns, drawing on empirical and theoretical sources. The discussion section provides an in-depth interpretation of findings, situating them within broader debates in AI ethics, public administration, and data governance. Finally, the conclusion synthesizes the study's contributions, highlighting practical implications and avenues for future research.

METHODOLOGY

The methodological approach adopted in this study is grounded in comprehensive qualitative analysis, informed by interdisciplinary scholarship spanning AI ethics, data governance, and public policy. The primary objective is to examine the operationalization of data-centric AI governance models in welfare management, emphasizing transparency, bias control, and policy compliance. A structured textual analysis methodology was selected due to its capacity to synthesize diverse literatures while enabling the identification of conceptual patterns and theoretical insights across multiple domains (Knauss et al., 2012).

Data sources were deliberately curated to include both peer-reviewed literature and authoritative technical reports, reflecting the dual necessity of theoretical grounding and practical applicability. Among these, Uddandarao et al. (2026) provided the central framework for evaluating data-centric governance models, offering detailed guidelines on transparency metrics, auditing protocols, and bias mitigation strategies. Complementary materials included technical white papers from IBM (2022) and Google (2022), which elaborated principles of responsible AI deployment, as well as reports by the Information Commissioner’s Office (2022) and the Defense Innovation Unit (2021) that addressed regulatory compliance and auditing best practices. Historical incident reports, such as Hall (2020) and Hosseini et al. (2017), were incorporated to illustrate the practical consequences of system failures and adversarial manipulation, thereby contextualizing the importance of governance frameworks in mitigating risks.

The analytical procedure comprised three stages. First, a comprehensive thematic coding of textual materials was conducted to extract core governance principles, operational challenges, and methodological strategies. Codes included categories such as data quality management, explainability mechanisms, bias detection techniques, human oversight structures, and regulatory alignment. Second, patterns were synthesized across sources to identify recurring challenges and potential solutions within the welfare governance context. This involved cross-referencing insights from model-centric and data-centric AI scholarship, with particular attention to the interplay between algorithmic design and dataset governance (Hamid, 2022; Hamid, 2023). Third, interpretive analysis was applied to elucidate the theoretical implications of these patterns, emphasizing ethical considerations, stakeholder trust, and policy compliance outcomes.

Limitations inherent to the methodology were explicitly acknowledged. The reliance on secondary literature and textual analysis restricts the study’s capacity for empirical generalization, and the absence of direct experimental or field data may limit the precision of causal inferences. However, the selection of authoritative sources and the rigorous cross-validation of thematic codes enhance the reliability of findings, while the focus on conceptual clarity ensures that the resulting insights are theoretically robust and practically relevant. Additionally, the study explicitly situates its analysis within the context of welfare management, recognizing that domain-specific considerations—such as social equity, vulnerable populations, and legal obligations—may constrain the transferability of conclusions to other governance domains.

Analytical rigor was further reinforced through triangulation. Multiple sources were cross-examined to validate interpretations, particularly regarding bias mitigation and transparency mechanisms. For instance, technical guidelines from IBM (2022) were juxtaposed with regulatory frameworks outlined by the Information Commissioner’s Office (2022), providing a multidimensional perspective on accountability protocols. Similarly, the theoretical distinctions between model-centric and data-centric AI (Hamid, 2023) were evaluated alongside practical case studies (Hall, 2020; Leibowicz et al., 2021) to ensure that conceptual assertions were grounded in operational realities.

Ethical considerations were also integral to the methodological design. The study maintained strict adherence to scholarly integrity, avoiding overgeneralizations, and ensuring that all sources were properly cited in alignment with academic conventions. In particular, the principles outlined by Uddandarao et al. (2026) regarding trustworthiness, bias control, and policy compliance informed both the analytical lens and the interpretive framework, ensuring that the study's conclusions are not merely descriptive but also normatively informed.

Through this methodological framework, the study seeks to generate a robust and comprehensive understanding of the mechanisms, challenges, and opportunities associated with deploying data-centric AI models in welfare governance. By systematically integrating theoretical discourse, practical insights, and ethical considerations, the research provides a foundation for both scholarly inquiry and policy-oriented recommendations, highlighting pathways for responsible AI adoption that advance transparency, fairness, and compliance.

RESULTS

The analysis reveals a complex landscape in which the deployment of AI in welfare governance intersects with multiple dimensions of transparency, bias mitigation, and policy compliance. Central to these findings is the recognition that data quality serves as the linchpin of effective AI governance. Poorly curated datasets, incomplete records, or historical biases embedded in administrative data can propagate inequities, producing decisions that are both ethically problematic and operationally inefficient (Hamid, 2022; Priyadarshi Uddandarao et al., 2026). In contrast, structured and ethically governed datasets facilitate not only more accurate predictions but also enhance stakeholder trust by providing clear and auditable decision pathways.

Transparency emerged as a critical determinant of system legitimacy. Thematic coding indicated that explainability mechanisms—ranging from user-facing interfaces that clarify algorithmic decisions to audit trails documenting data provenance—are essential for fostering accountability. Uddandarao et al. (2026) emphasize that without transparent processes, AI systems risk being perceived as opaque "black boxes," undermining public confidence and reducing the willingness of citizens to engage with welfare programs. Furthermore, transparency interacts synergistically with compliance mechanisms; well-documented decision processes enable auditors and regulators to verify alignment with policy mandates and ethical standards (Information Commissioner's Office, 2022).

Bias mitigation constitutes another salient theme. Across the literature, a recurrent observation is that algorithmic bias is not merely a technical challenge but a socio-ethical imperative. Techniques such as adversarial testing, demographic balancing, and ongoing model retraining were highlighted as critical strategies for minimizing discriminatory outcomes (Hosseini et al., 2017; Leibowicz et al., 2021). Moreover, a proactive stance on bias detection—anticipating rather than reacting to disparities—was identified as a hallmark of robust governance frameworks. In welfare contexts, where decisions directly affect vulnerable populations, such proactive measures are indispensable for ensuring equitable treatment and maintaining legal and ethical legitimacy.

Human oversight emerged as a complementary factor. Across multiple sources, reliance solely on automated decision-making was cautioned against, with scholars advocating for hybrid models that integrate algorithmic efficiency with human judgment (Oberste & Heinzl, 2023; Dunnmon et al., 2021). Such oversight is particularly relevant in cases of ambiguous or high-stakes decisions, where contextual knowledge, ethical sensitivity, and policy interpretation are required. The findings suggest that AI systems should function as decision-support tools rather than autonomous arbiters, with clear protocols for escalation and human intervention.

Policy compliance also surfaced as a multi-layered concern. Regulatory frameworks, ethical guidelines, and auditing standards collectively shape the operational boundaries of AI in welfare governance. Notably, Uddandarao et al. (2026) highlight the importance of integrating these dimensions into AI system design, ensuring that governance mechanisms are not retrofitted but embedded from the outset. Compliance extends beyond legal mandates to encompass broader societal expectations regarding fairness, accountability, and trust. The interplay between transparency, bias mitigation, and compliance suggests that these elements are mutually reinforcing: high-quality data supports explainability, which in turn facilitates regulatory alignment and ethical legitimacy.

Additional insights pertain to practical challenges and emergent risks. Technical limitations, including adversarial attacks, model fragility, and data incompleteness, can undermine governance objectives if not adequately addressed (Hall, 2020; Hosseini et al., 2017). Moreover, the adoption of AI can inadvertently entrench existing bureaucratic hierarchies, as algorithmic outputs may be perceived as authoritative, reducing human discretion and critical evaluation. These findings underscore the necessity of designing AI governance systems that are resilient, adaptable, and socially informed, integrating both technical robustness and ethical foresight.

DISCUSSION

The findings elucidate the multifaceted nature of AI integration in welfare governance, revealing a delicate balance between technological capability, ethical responsibility, and policy compliance. At the theoretical level, the results reinforce the centrality of data-centric approaches, corroborating the assertions of Hamid (2022, 2023) and Uddandarao et al. (2026) regarding the primacy of high-quality data in achieving trustworthy AI outcomes. Data-centric AI does not merely enhance predictive accuracy; it embodies a normative commitment to fairness, transparency, and accountability, which are essential for legitimizing welfare interventions in the eyes of stakeholders.

A comparative analysis of model-centric versus data-centric paradigms highlights critical implications. Model-centric approaches, while emphasizing algorithmic sophistication, risk overlooking the ethical and socio-political dimensions of decision-making. Conversely, data-centric governance prioritizes the integrity and representativeness of input data, thereby addressing the root causes of bias and inequity (Priyadarshi Uddandarao et al., 2026). In welfare administration, where decisions affect vulnerable populations, this distinction is more than theoretical; it has tangible consequences for equity, trust, and legal compliance. The integration of these paradigms, where model sophistication is grounded in ethically curated data, emerges as a best practice for sustainable AI deployment.

Transparency, as evidenced in both theoretical and practical analyses, functions as a critical mediator of trust and legitimacy. Detailed audit trails, user-facing explanations, and regulatory reporting mechanisms ensure that stakeholders—ranging from policymakers to service recipients—can scrutinize and understand AI-mediated decisions (Information Commissioner’s Office, 2022; IBM, 2022). This aligns with broader scholarly discourse on explainable AI, which posits that accountability and interpretability are prerequisites for responsible deployment, particularly in high-stakes social services (Oberste & Heinzl, 2023). The absence of transparency, conversely, fosters skepticism, reduces civic engagement, and may precipitate systemic resistance to technological interventions.

The discourse on bias mitigation reveals the complex interplay between technical methods and socio-ethical imperatives. Techniques such as demographic balancing, adversarial testing, and continuous retraining address algorithmic inequities, yet these must be complemented by organizational and institutional safeguards (Hosseini et al., 2017; Leibowicz et al., 2021). Ethical AI governance, therefore, necessitates both algorithmic rigor and systemic vigilance, acknowledging that bias is not merely a computational artifact but a reflection of broader social inequalities. In welfare contexts, the stakes are amplified: biased decisions can reinforce structural disadvantages, exacerbate social disparities, and erode public trust. Consequently, a comprehensive governance framework must integrate technical, organizational, and policy-level interventions to effectively mitigate bias.

Policy compliance is intricately linked to both transparency and bias control. Effective AI governance requires alignment with statutory mandates, ethical guidelines, and industry standards, as emphasized by Uddandarao et al. (2026) and Dunnmon et al. (2021). The study highlights the necessity of embedding compliance mechanisms at the design phase of AI systems, rather than retrofitting governance post-deployment. This proactive approach facilitates auditing, ensures regulatory adherence, and enhances public confidence in welfare programs. The synergistic integration of transparency, bias mitigation, and compliance underscores the interdependent nature of responsible AI governance: deficiencies in one dimension invariably affect the others, necessitating a holistic and systemic perspective.

The role of human oversight emerges as both complementary and essential. Automated systems, despite their efficiency and scalability, cannot replicate the ethical discernment, contextual awareness, and policy interpretation inherent to human judgment. Hybrid governance models, combining algorithmic decision-support with human supervision, offer a pragmatic solution to the limitations of fully autonomous AI systems (Oberste & Heinzl, 2023). This approach not only enhances the accuracy and fairness of decisions but also preserves institutional accountability, ensuring that AI functions as an enabler rather than a replacement of human governance capacity.

Historical case analyses further illuminate potential vulnerabilities. Failures such as robotic malfunctions in public spaces (Hall, 2020) and adversarial manipulation of AI systems (Hosseini et al., 2017) highlight the need for resilient design, continuous monitoring, and adaptive risk management. These incidents underscore the non-deterministic nature of AI systems and the critical importance of comprehensive governance frameworks capable of anticipating and mitigating emergent risks. They also reinforce the ethical imperative to safeguard vulnerable populations from unintended consequences, aligning technical design with social responsibility.

The discussion also engages with broader scholarly debates on AI ethics and public administration. While some scholars advocate for full automation in service delivery, emphasizing efficiency gains and scalability, the findings of this study caution against such an approach in welfare governance. Ethical, social, and policy considerations necessitate a tempered integration, where AI supports but does not supplant human judgment. Moreover, the study contributes to ongoing discourse on data governance, demonstrating that ethical stewardship of data is not ancillary but central to the legitimacy of AI interventions.

empirical generalization, as no primary field data or controlled experimental trials were conducted. While the synthesized literature provides rich conceptual and practical insights, actual deployment outcomes may vary depending on contextual factors such as local governance structures, sociocultural norms, and technological infrastructure. Additionally, while the study emphasizes welfare management, findings may not be directly transferable to other governance domains such as taxation, law enforcement, or environmental regulation without domain-specific adaptation. Despite these constraints, the breadth and depth of the literature reviewed, combined with the rigorous thematic analysis, provide a robust foundation for guiding AI deployment in socially sensitive contexts.

Future research directions are manifold. First, there is a pressing need for longitudinal empirical studies that evaluate the real-world impact of data-centric AI on welfare outcomes. Such studies could measure improvements in efficiency, reductions in bias, and stakeholder trust over time, providing quantitative validation of the theoretical frameworks articulated in this study. Second, research should explore the socio-technical integration of AI within existing bureaucratic systems, investigating how organizational culture, policy constraints, and institutional resistance influence adoption and performance. Third, emerging areas such as federated learning, differential privacy, and blockchain-based auditing present promising avenues for enhancing data integrity, security, and transparency. Investigating these technologies in conjunction with governance models may yield innovative mechanisms for ensuring ethical and compliant AI deployment.

Another critical avenue is the examination of human-AI interaction within governance systems. While hybrid models have been proposed, the optimal configuration of decision-support roles, escalation protocols, and human oversight remains underexplored. Detailed case studies could illuminate best practices for balancing algorithmic efficiency with human judgment, offering actionable insights for policymakers and system designers. Similarly, cross-cultural comparative analyses could reveal how societal values, regulatory philosophies, and citizen expectations shape the acceptance and efficacy of AI in governance.

Finally, interdisciplinary collaboration between computer scientists, public policy experts, ethicists, and social scientists is essential. AI governance is not a purely technical problem; it is a socio-technical challenge that requires harmonization of algorithmic capabilities with human values, regulatory norms, and ethical imperatives. By fostering such collaboration, future research can move beyond theoretical frameworks to develop operationalized, context-sensitive models that ensure AI interventions enhance, rather than undermine, public welfare.

In conclusion, the study provides compelling evidence that trustworthy, data-centric AI has the potential to transform welfare governance by enhancing transparency, mitigating bias, and ensuring policy compliance. By integrating high-quality data, transparent decision pathways, rigorous auditing, and human oversight, AI systems can function as ethically responsible instruments that support efficient and equitable service delivery. The research highlights that the success of AI in governance is contingent upon holistic integration: technical sophistication alone is insufficient without attention to data quality, ethical design, and regulatory alignment. Moreover, the study underscores the importance of proactive risk management, iterative evaluation, and continuous stakeholder engagement as essential components of sustainable AI deployment.

The findings resonate with and extend the insights of Priyadarshi Uddandarao et al. (2026), who articulate the transformative potential of data-centric AI governance frameworks. By embedding principles of trustworthiness, transparency, and compliance into the core of AI systems, policymakers can navigate the complex ethical and operational landscape of welfare administration, ensuring that technological innovation aligns with societal values. Ultimately, the research advocates for an integrated governance paradigm, where AI is not merely a tool but a co-governor—capable of enhancing public trust, improving administrative efficacy, and reinforcing the ethical foundations of welfare delivery.

CONCLUSION

This study underscores the critical role of trustworthy, data-centric AI in modern welfare governance. By prioritizing high-quality data, implementing transparent decision-making processes, and ensuring alignment with ethical and regulatory standards, AI can substantially enhance the efficacy, fairness, and legitimacy of welfare management. Key insights include the necessity of hybrid human-AI governance models, proactive bias mitigation strategies, and rigorous auditing protocols. While challenges remain—particularly regarding technical vulnerabilities, adversarial manipulation, and context-specific adaptation—the integration of AI offers transformative opportunities for public administration. Future research should focus on empirical validation, socio-technical integration, and interdisciplinary collaboration to refine and operationalize AI governance frameworks. The findings presented herein provide a roadmap for policymakers, technologists, and scholars seeking to harness AI responsibly, ensuring that welfare systems remain both efficient and ethically accountable in an increasingly data-driven world.

REFERENCES

1. Gartner. Predicts 2019: Artificial Intelligence Core Technologies. Technical report, Gartner, November 2018. URL https://www.gartner.com/en/documents/3894131.

2. Hossein Hosseini, Sreeram Kannan, Baosen Zhang, and Radha Poovendran. Deceiving Google's Perspective API Built for Detecting Toxic Comments, February 2017. URL http://arxiv.org/abs/1702.08138. arXiv:1702.08138 [cs].

3. Claire Leibowicz, Sean McGregor, and Aviv Ovadya. The Deepfake Detection Dilemma: A Multistakeholder Exploration of Adversarial Dynamics in Synthetic Media, February 2021. URL http://arxiv.org/abs/2102.06109. arXiv:2102.06109 [cs].

4. Jared Dunnmon, Bryce Goodman, Peter Kirechu, Carol Smith, and Alexandrea Van Deusen. Responsible AI Guidelines in Practice. Technical report, Defense Innovation Unit, March 2021. URL https://assets.ctfassets.net/3nanhbfkr0pc/acoo1Fj5uungnGNPJ3QWy/6ec382b3b5a20ec7de6defdb33b04dcd/2021_RAI_Report.pdf.

5. O. H. Hamid. Data-centric and model-centric AI: Twin drivers of compact and robust industry 4.0 solutions. Applied Sciences, 13(5):2753, Feb. 2023.

6. O. H. Hamid. From model-centric to data-centric AI: A paradigm shift or rather a complementary approach? In Proc. 8th International Conference on Information Technology Trends (ITT), May 2022, pp. 196–199.

7. Information Commissioner’s Office. Guidance on the AI auditing framework Draft guidance for consultation. Technical Report 20200214, Information Commissioner’s Office, November 2022. URL https://ico.org.uk/media/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf.

8. C. Huang, Z. Zhang, B. Mao, and X. Yao. An overview of artificial intelligence ethics. IEEE Transactions on Artificial Intelligence, 4(4):799–819, 2023.

9. Eric Knauss, Daniela Damian, Germán Poo-Caamaño, and Jane Cleland-Huang. Detecting and Classifying Patterns of Requirements Clarifications. 20th IEEE International Requirements Engineering Conference, September 2012. doi: 10.1109/RE.2012.6345811.

10. IBM. AI Ethics, November 2022. URL https://www.ibm.com/artificial-intelligence/ethics.

11. Google. Building responsible AI for everyone, November 2022. URL https://ai.google/responsibilities/.

12. L. Oberste and A. Heinzl. User-centric explainability in healthcare: A knowledge-level perspective of informed machine learning. IEEE Transactions on Artificial Intelligence, 4(4):840–857, 2023.

13. GitLab. CI/CD concepts | GitLab, November 2022. URL https://docs.gitlab.com/ee/ci/introduction/.

14. J. Waring, C. Lindvall, and R. Umeton. Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. Artificial Intelligence in Medicine, 104, Apr. 2020, Article 101822.

15. Patrick Hall. Incident 134: Robot in Chinese Shopping Mall Fell off the Escalator, Knocking down Passengers. AI Incident Database, 2020. URL https://incidentdatabase.ai/cite/134. Publisher: Responsible AI Collaborative.

16. Hugging Face. distilbert-base-uncased-finetuned-sst-2-english · Hugging Face, November 2022. URL https://huggingface.co/distilbert-base-uncased-finetuned-sst-2-english.