Mednarodno inovativno poslovanje
Journal of Innovative Business and Management

# A domain-independent review of explainable AI's role in facilitating innovation and creativity in organizations

**Miha Lavrič*, Aurora Lea Lavrič****

***Abstract***: Explainable Artificial Intelligence (XAI) offers transparency and interpretability in AI systems, addressing the opacity of traditional models. This review examines how XAI fosters innovation and creativity in organizations by enhancing decision-making, trust, and collaboration across diverse domains such as healthcare, manufacturing, and agriculture. Thematic analysis of relevant literature reveals that XAI builds stakeholder confidence, promotes ethical practices, and bridges gaps between technical and non-technical teams, encouraging inclusive problem-solving. While this study highlights significant short-term benefits, future longitudinal research is necessary to explore XAI's long-term impact. This paper provides insights for academics, practitioners, and policymakers, emphasizing XAI's potential to foster innovation.

***Keywords:*** explainable AI, innovation, creativity, decision-making, organizational behavior
***JEL classification:*** O31

# Domensko neodvisen pregled vloge razložljive umetne inteligence pri spodbujanju inovacij in ustvarjalnosti v organizacijah

***Povzetek:*** Razložljiva umetna inteligenca (XAI) zagotavlja preglednost in razložljivost v sistemih umetne inteligence ter naslavlja nepreglednost tradicionalnih modelov. Ta pregled preučuje, kako XAI spodbuja inovativnost in ustvarjalnost v organizacijah z izboljšanjem odločanja, zaupanja in sodelovanja na različnih področjih, kot so zdravstvo, proizvodnja in kmetijstvo. Tematska analiza relevantne literature razkriva, da XAI krepi zaupanje deležnikov, spodbuja etične prakse ter premošča razkorak med tehničnimi in netehničnimi ekipami, kar spodbuja vključujoče reševanje problemov. Čeprav ta študija poudarja pomembne kratkoročne koristi, so za razumevanje dolgoročnih vplivov XAI potrebne nadaljnje longitudinalne raziskave. Prispevek ponuja vpogled za akademike, strokovnjake in oblikovalce politik ter poudarja potencial XAI pri spodbujanju inovacij.

***Ključne besede:*** razložljiva umetna inteligenca, inovacije, ustvarjalnost, odločanje, organizacijsko vedenje

**dr.,
DOBA Business School, Maribor, Slovenia*
miha.lavric@gmail.com

***Gymnasium and Vocational Secondary School for Chemistry and Pharmacy Ruše, Slovenia*
lavricaurora@gmail.com



*DOI 10.32015/JIBM.2025.17.1.8*

*Mednarodno inovativno poslovanje = Journal of Innovative Business and Management*

*ISSN 1855-6175*

# 1 INTRODUCTION

## 1.1 Background

Artificial Intelligence (AI) has revolutionized decision-making in organizations by enabling data-driven insights, optimizing processes, and enhancing efficiency across diverse domains. Despite its advantages, traditional AI systems often operate as “black boxes,” obscuring their inner workings and leading to stakeholder mistrust (Guidotti et al., 2019). This lack of transparency limits the potential of AI to foster collaboration and creativity, as decision-makers and technical teams struggle to understand, validate, and trust AI-generated outcomes. Also, this opacity raises fairness concerns, as stakeholders cannot verify whether AI decisions align with ethical and equitable principles (Binns, 2018).

Explainable Artificial Intelligence (XAI) addresses this issue by providing transparency and interpretability in AI systems. Techniques like saliency maps, which undergo rigorous sanity checks, are crucial in elucidating model behavior and ensuring reliability (Adebayo et al., 2018). XAI’s ability to generate human-understandable explanations for AI models fosters trust and bridges the gap between technical experts and organizational stakeholders. As a result, XAI is emerging as a critical tool for enhancing organizational innovation and creativity by enabling diverse teams to collaborate effectively and derive actionable insights from AI systems (Samek et al., 2017). To better understand the myriad applications of XAI across various industries, Table 1 summarizes its key use cases and associated benefits.

Table 1 (Portrait). Summary of XAI Applications Across Industries

| Industry | XAI Applications | Innovation & Creativity Contribution | Key Benefits | Key Reference |
|---|---|---|---|---|
| **Healthcare** | Diagnostic Tools | Enhances trust and interpretability, allowing clinicians to co-create personalized treatment plans and innovate care models. | Improved Patient Outcomes | Carvalho et al., 2019 |
| **Manufacturing** | Predictive Maintenance | Promotes collaborative innovation by enabling cross-functional teams to anticipate and creatively solve production issues. | Reduced Downtime | Gunning & Aha, 2019 |
| **Agriculture** | Precision Farming | Empowers farmers to make informed, sustainable choices and experiment with techniques, stimulating bottom-up innovation. | Sustainable Practices | Samek et al., 2017 |
| **Finance** | Fraud Detection | Facilitates creative design of risk models and regulatory strategies through understandable anomaly explanations. | Operational Efficiency | Rai, 2020 |
| **Education** | Tailored Content | Supports innovative curriculum development by explaining AI-driven personalization, enhancing educator creativity. | Improved Learning Outcomes | Miller, 2019 |
| **Recruitment** | Fair Hiring Practices | Encourages ethical innovation in HR by revealing decision logic, promoting inclusivity in candidate selection. | Increased Inclusivity | Dwivedi et al., 2023; Floridi et al., 2018 |
| **Retail** | Transparent Recommendation Systems | Enhances marketing innovation by clarifying consumer behavior, supporting creative | Increased Customer Trust and Engagement | Rai, 2020 |

| | | | | |
|---|---|---|---|---|
| | | customer segmentation strategies. | | |
| **Energy** | Energy Management | Enables innovative sustainability planning by explaining complex energy optimization models to diverse stakeholders. | Improved Efficiency, Reduced Environmental Impact | Hassija et al., 2024 |
| **Transportation** | Edge Case Decision Explanation | Fosters trust-driven innovation in autonomous systems by clarifying decisions in exceptional scenarios. | Enhanced Regulatory and Public Trust | Zhou et al., 2021 |
| **Marketing** | Reasoning For Ad Performance Metrics | Inspires creative campaign adjustments by clarifying how AI attributes success to specific variables or audience segments. | Strategy Refinement | Kaur et al., 2022 |

Table 1 (Landscape). Summary of XAI Applications Across Industries

| Industry | XAI Applications | Innovation & Creativity Contribution | Key Benefits | Key Reference |
|---|---|---|---|---|
| **Healthcare** | Diagnostic Tools | Enhances trust and interpretability, allowing clinicians to co-create personalized treatment plans and innovate care models. | Improved Patient Outcomes | Carvalho et al., 2019 |
| **Manufacturing** | Predictive Maintenance | Promotes collaborative innovation by enabling cross-functional teams to anticipate and creatively solve production issues. | Reduced Downtime | Gunning & Aha, 2019 |
| **Agriculture** | Precision Farming | Empowers farmers to make informed, sustainable choices and experiment with techniques, stimulating bottom-up innovation. | Sustainable Practices | Samek et al., 2017 |
| **Finance** | Fraud Detection | Facilitates creative design of risk models and regulatory strategies through understandable anomaly explanations. | Operational Efficiency | Rai, 2020 |
| **Education** | Tailored Content | Supports innovative curriculum development by explaining AI-driven personalization, enhancing educator creativity. | Improved Learning Outcomes | Miller, 2019 |
| **Recruitment** | Fair Hiring Practices | Encourages ethical innovation in HR by revealing decision logic, promoting inclusivity in candidate selection. | Increased Inclusivity | Dwivedi et al., 2023; Floridi et al., 2018 |
| **Retail** | Transparent Recommendation Systems | Enhances marketing innovation by clarifying consumer behavior, supporting creative customer segmentation strategies. | Increased Customer Trust and Engagement | Rai, 2020 |

| | | | | |
|---|---|---|---|---|
| **Energy** | Energy Management | Enables innovative sustainability planning by explaining complex energy optimization models to diverse stakeholders. | Improved Efficiency, Reduced Environmental Impact | Hassija et al., 2024 |
| **Transportation** | Edge Case Decision Explanation | Fosters trust-driven innovation in autonomous systems by clarifying decisions in exceptional scenarios. | Enhanced Regulatory and Public Trust | Zhou et al., 2021 |
| **Marketing** | Reasoning For Ad Performance Metrics | Inspires creative campaign adjustments by clarifying how AI attributes success to specific variables or audience segments. | Strategy Refinement | Kaur et al., 2022 |

Table 1 highlights domain-specific XAI applications that not only enhance decision-making but also stimulate organizational creativity by fostering transparency, trust, and cross-functional collaboration. These cases demonstrate how XAI enables stakeholders—technical and non-technical alike—to understand, challenge, and creatively adapt AI-driven outputs. For instance, in agriculture, XAI supports iterative experimentation with farming inputs, while in education, it informs the design of personalized learning paths based on explainable student data; embodying the paper's central argument: that explainability is not merely a technical feature but a critical enabler of innovation and inclusive problem-solving.

## 1.2 Literature review

### 1.2.1 Explainable AI and Trust

Research indicates that transparency is a fundamental driver of trust in AI systems. Methods such as saliency maps, validated through sanity checks to ensure meaningful explanations, exemplify this approach (Adebayo et al., 2018). XAI provides interpretability by elucidating the reasoning behind AI-generated decisions, thereby reducing user skepticism and resistance (Zhou et al., 2021). For example, in healthcare, explainable models have been shown to improve physicians' confidence in diagnostic tools, enabling faster adoption and integration into clinical workflows (Carvalho et al., 2019). Methods such as feature importance scores and surrogate models have been pivotal in elucidating the reasoning behind AI-generated decisions (Linardatos et al., 2020). Recent advancements in explainable AI, including its integration with large language models, highlight its potential to enhance user trust and system interpretability (Cambria et al., 2024).

### 1.2.2 Creativity and Innovation in Organizations

Innovation thrives in environments where diverse teams can collaborate and challenge conventional approaches. XAI facilitates this by providing accessible explanations of complex AI processes, allowing non-technical stakeholders to engage meaningfully in decision-making (Kaur et al., 2022). Integrating XAI with large language models opens new avenues for creativity and innovation, particularly in computational social science (Ziems et al., 2024). Guidelines proposed by Amershi et al. (2019) emphasize the importance of designing AI systems that enable intuitive human interaction. These guidelines underscore the role of explainability in ensuring that users across technical and non-technical domains can effectively collaborate, fostering creativity and informed decision-making. Furthermore, explainable systems enable organizations to identify and mitigate biases in AI algorithms, fostering inclusive and ethical innovation (Binns, 2018; Gunning & Aha, 2019). In human resource management, XAI facilitates innovative practices by bridging the gap between technical and non-technical teams, as demonstrated in AI capability frameworks (Chowdhury et al., 2023).

### 1.2.3 Gaps in the Existing Literature

While studies emphasize the technical and ethical benefits of XAI, limited research investigates its direct impact on organizational creativity. Existing literature lacks insights into how XAI fosters cross-functional collaboration, improves problem-solving capabilities, and drives innovative outcomes in industries like manufacturing, agriculture, and healthcare (Hassija et al., 2024). This review seeks to address these gaps by synthesizing findings from multiple domains.

### 1.3 Problem Statement, Research Gap, and Objectives

While explainable artificial intelligence (XAI) has been extensively examined from technical, ethical, and accountability standpoints (Dwivedi et al., 2023; Floridi et al., 2018), its potential to foster innovation and creativity within organizations remains significantly underexplored (Blatch-Jones et al., 2024). Existing literature primarily addresses how XAI enhances transparency and trust but offers limited insights into its role in driving creative problem-solving, improving team dynamics, and supporting innovative decision-making (Ali et al., 2023).

Innovation and creativity often thrive in collaborative environments where transparency and shared understanding are essential. The opacity of traditional AI systems can hinder collaboration, inclusivity, and trust among stakeholders, thus impeding innovation. In contrast, XAI has the potential to support fairness, improve communication, and enable mutual understanding across teams, creating fertile ground for innovation (Binns, 2018; Ali et al., 2023).

This review seeks to fill this gap by exploring the domain-independent role of XAI in fostering innovation through improved decision-making, accountability, and collaboration. The overarching research question guiding this study is:

How does XAI enhance decision-making processes and collaborative efforts in diverse organizational contexts to foster innovation and creativity?

To address this question, the study aims to:

- Synthesize academic and industry insights into the relationship between XAI and organizational innovation;
- Examine how XAI influences team interactions, creative problem-solving, and inclusive decision-making;
- Provide a holistic understanding of XAI's transformative potential beyond technical applications, with a focus on organizational growth and creativity.

## 2 METHODOLOGY

### 2.1 Research Design

This study employs a qualitative literature review methodology to explore the role of Explainable Artificial Intelligence (XAI) in fostering innovation and creativity across diverse organizational contexts. The approach is designed to synthesize findings from academic literature, industry reports, and case studies, providing a domain-independent perspective on XAI's applications. By focusing on secondary data sources, the study captures insights from various industries, including healthcare, manufacturing, agriculture, and finance, where XAI has shown potential to influence decision-making and collaboration (Gilpin et al., 2018; Rai, 2020).

### 2.2 Data Collection

The data for this study were sourced from peer-reviewed journals, conference proceedings, white papers, and industry reports published between 2017 and 2024. This timeframe was chosen to ensure relevance, as XAI has gained significant attention in recent years due to advances in machine learning interpretability techniques. Databases such as Scopus, Web of Science, IEEE Xplore, and Google Scholar were utilized to gather comprehensive and credible sources.

Inclusion Criteria:

- Studies focusing on the implementation or theoretical discussion of XAI.

- Literature addressing innovation, creativity, and decision-making in organizations influenced by AI technologies.
- Cross-domain analyses of XAI's impact in sectors such as healthcare, manufacturing, agriculture, and finance.

Exclusion Criteria:

- Studies emphasizing purely technical aspects of XAI without organizational implications.
- Outdated sources published before 2017, unless foundational to understanding XAI (e.g., seminal works like Samek et al., 2017).
- Non-peer-reviewed or opinion-based content lacking empirical support.

The keywords used in the search included: Explainable Artificial Intelligence, XAI and creativity, XAI in organizational decision-making, and XAI in innovation. Boolean operators (AND, OR) and truncations were applied to refine the results and ensure the inclusion of relevant studies.

While the PRISMA guidelines were considered, they were not applied in this study. Given the focused scope and conceptual nature of this review, a comprehensive systematic review was not intended. Instead, this paper employed purposive selection to identify literature of high relevance to the research question. The selected works were then analyzed thematically, allowing for deep, qualitative insight into recurring patterns regarding XAI's role in organizational innovation and creativity.

Despite efforts to ensure a comprehensive and balanced review, potential biases in literature selection must be acknowledged. The selection process may have been influenced by the researcher's access to specific databases, familiarity with certain academic communities, and the predominance of English-language publications. Additionally, the inclusion of peer-reviewed and high-impact sources could unintentionally privilege certain perspectives while overlooking valuable insights from emerging regions or non-traditional formats. To mitigate these risks, a deliberate effort was made to diversify source types (e.g., journals, industry reports, case studies) and domains. However, future research would benefit from incorporating multilingual sources and broader access to grey literature to further enhance representativeness and reduce selection bias.

#### 2.2.1 Literature Categorization Approach

To ensure thematic relevance and focus, the selected literature was categorized into two distinct groups:

(1) papers directly exploring XAI's impact on organizational innovation and creativity (15 papers), and
(2) technical, methodological, or foundational papers supporting a broader understanding of XAI principles (11 papers).

This classification helped guide the thematic analysis by distinguishing between application-driven insights and foundational support. The categorization also reflects the balance between practical organizational implications and theoretical or technical grounding. A full breakdown of this categorization is provided in Appendix A to ensure transparency in the inclusion criteria and to demonstrate alignment with the research objectives.

## 2.3 Data Analysis

A thematic analysis was employed to identify recurring patterns and themes in the selected literature. Thematic analysis is well-suited for qualitative research, allowing researchers to extract meaningful insights across diverse data sources (Braun & Clarke, 2006).

The following steps were undertaken:

- Familiarization with Data: The literature was read and reread to identify relevant sections pertaining to XAI's influence on innovation, creativity, and decision-making.
- Coding: Key concepts, such as trust, collaboration, interpretability, and ethical implications, were coded systematically to manage and organize qualitative data effectively.

- Theme Development: The codes were categorized into broader themes, including XAI's role in improving decision-making, fostering collaboration, and promoting ethical innovation.
- Cross-Domain Comparison: The themes were compared across industries to evaluate whether the impact of XAI was consistent or varied depending on the organizational context.

## 2.4 Validation and Reliability

To ensure the reliability and validity of the thematic analysis, the study adopted rigorous methodological practices tailored for single-researcher qualitative research:

The following steps were undertaken:

- Self-Reflexivity and Audit Trail: The researcher maintained a detailed audit trail, documenting every step of the coding and analysis process. This included memos on decisions made during data coding, thematic development, and the interpretation of results. This practice ensured transparency and allowed for consistent re-evaluation of themes to reduce potential biases (Nowell et al., 2017).
- Triangulation of Sources: Validation was achieved by triangulating findings across multiple sources, including peer-reviewed journal articles, industry reports, and case studies. This approach ensured that the themes and conclusions were supported by diverse and independent pieces of evidence, enhancing their credibility (Patton, 1999).
- Iterative Coding Process: The coding process was iterative, involving repeated cycles of data immersion, theme refinement, and re-analysis. This approach minimized errors and ensured consistency in the thematic development across the dataset.
- Comparison with Seminal Studies: The results were cross-referenced with seminal literature on XAI and organizational innovation. This ensured that the themes aligned with well-established theoretical frameworks and findings in the field, providing additional credibility.

The methodological framework is summarized in Table 2, which highlights the sources, inclusion criteria, and exclusions that guided the data collection and analysis process.

Table 2. Methodological Overview

| Data Source | Inclusion Criteria | Exclusion Criteria |
|---|---|---|
| **Journals** | Published After 2017 | Technical-Only Focus |
| **Reports** | Peer-Reviewed | Non-Peer Reviewed |
| **Case Studies** | Relevant to XAI | Published Before 2017 |

# 3 RESULTS

## 3.1 Enhancing Decision-Making through Explainable AI

Explainable Artificial Intelligence (XAI) improves decision-making processes in organizations by making AI models transparent and interpretable. Unlike traditional "black-box" models, XAI provides users with clear justifications for predictions or outcomes (Zhou et al., 2021). Validated techniques such as saliency maps contribute to this transparency, especially in high-stakes industries where clear explanations are essential (Adebayo et al., 2018). In healthcare, XAI is used in diagnostic tools to present the reasoning behind AI recommendations (Carvalho et al., 2019). In the financial sector, XAI models help identify fraudulent transactions by highlighting patterns of anomalous behavior (Rai, 2020).

The literature also reports applications of interpretable models in manufacturing, where XAI supports predictive maintenance systems by identifying and explaining potential machinery failures (Gunning & Aha, 2019).

## 3.2 Promoting Ethical and Inclusive Innovation

XAI is associated with detecting bias in AI algorithms and supporting ethical decision-making (Blatch-Jones et al., 2024). Studies describe how the integration of fairness principles into AI design enhances accountability and inclusivity (Binns, 2018). For example, XAI has been used to audit hiring decisions by revealing influential factors in candidate selection, aiding in the identification of discriminatory patterns in training data (Floridi et al., 2018; Dwivedi et al., 2023).

XAI models also simplify complex AI outputs, making decision-making tools more accessible to non-technical users. In agriculture, XAI-driven precision farming systems present explanations for decisions such as irrigation scheduling or crop selection, enabling broader adoption (Samek et al., 2017). Table 3 provides an overview of the primary challenges addressed by XAI, illustrating its potential to mitigate barriers to AI adoption and integration.

Table 3. Key Challenges Addressed by XAI

| Challenge | XAI Solution | Example Applications |
|---|---|---|
| **Lack of Trust** | Interpretability | Healthcare Diagnostics |
| **Ethical Concerns** | Bias Detection | Recruitment |
| **Communication Barriers** | Simplified Explanations | Cross-Team Collaboration |

## 3.3 Facilitating Collaboration Across Teams

The literature indicates that XAI enhances collaboration between technical and non-technical teams by creating a shared language for understanding AI systems. In manufacturing, for example, XAI-powered quality control systems explain deviations in product quality by identifying contributing factors (Molnar, 2024; Chowdhury et al., 2023).

These explanations are often delivered through visualizations and user-friendly interfaces, which help align technical and operational perspectives across different roles and departments (Kaur et al., 2022).

## 3.4 Driving Innovation and Creativity

XAI is frequently associated with exploratory approaches to problem-solving. In product development, XAI tools allow designers to assess how variables such as material properties and cost constraints influence performance outcomes (Rai, 2020).

In the education sector, XAI platforms provide transparency in personalized learning pathways by explaining how individual metrics inform system-generated recommendations (Miller, 2019).

## 3.5 Cross-Domain Implications

The literature shows that XAI applications are consistent across multiple industries, though use cases and impact vary. In healthcare, XAI supports trust and transparency in diagnoses and treatments (Carvalho et al., 2019). In agriculture, it contributes to sustainable practices by explaining complex analyses of soil and weather data (Samek et al., 2017). In finance, XAI is applied to ensure regulatory compliance and reduce risk (Rai, 2020).

Differences in implementation are often linked to organizational culture and levels of readiness, suggesting that XAI's effectiveness may depend on the structural and collaborative characteristics of each setting (Hassija et al., 2024).

## 3.6 Summary of Findings

Thematic analysis revealed the following key themes:

- Improved Decision-Making: XAI enhances transparency in organizational decisions.
- Ethical Practices: XAI supports bias detection and fairness in AI systems.
- Collaboration: XAI facilitates alignment across technical and non-technical stakeholders.
- Innovation and Creativity: XAI enables exploratory design and problem-solving.
- Cross-Domain Consistency: XAI applications are widespread but shaped by contextual factors.

A summary of the key findings and their practical implications is provided in Table 4.

Table 4. Key Findings and Implications

| Theme | Key Findings | Implications |
| --- | --- | --- |
| **Decision-Making** | Improved Trust | Better Decisions |
| **Collaboration** | Enhanced Teamwork | Improved Innovation |
| **Ethical Practices** | Mitigated Biases | Inclusive Practices |

# 4 DISCUSSION

## 4.1 Significance of Findings

This study highlights the transformative potential of Explainable Artificial Intelligence (XAI) in fostering innovation, creativity, and collaboration within organizations. By addressing the opacity of traditional AI systems, XAI—through validated interpretability techniques such as saliency maps (Adebayo et al., 2018)—empowers stakeholders with interpretable insights, thereby enhancing trust, decision-making, and inclusivity. The use of inherently interpretable models, especially in high-stakes environments, aligns with calls for transparent and accountable AI (Rudin, 2019; Molnar, 2024).

The findings indicate that XAI plays a crucial role in improving decision-making by supporting transparency across sectors such as healthcare, finance, and manufacturing (Zhou et al., 2021; Carvalho et al., 2019; Rai, 2020). In such settings, the ability to verify and understand AI-driven outputs leads to greater stakeholder confidence and improved outcomes.

The literature also consistently points to XAI's contribution to ethical and inclusive innovation. By enabling the detection of bias and offering auditability in AI processes, XAI supports accountability in domains such as recruitment, resource allocation, and risk management (Floridi et al., 2018; Dwivedi et al., 2023; Binns, 2018).

Collaboration is another major theme: XAI bridges the communication gap between technical and non-technical teams by making AI outputs understandable across diverse user profiles. This has been documented in manufacturing, agriculture, and HR contexts, where shared interpretability fosters better teamwork and alignment (Gunning & Aha, 2019; Chowdhury et al., 2023).

In terms of innovation and creativity, the reviewed literature describes how XAI supports exploratory work such as "what-if" scenarios and iterative design improvements, particularly in product development and education (Rai, 2020; Miller, 2019). This suggests that transparency not only supports risk mitigation but also acts as a catalyst for ideation.

Based on the findings, Table 5 outlines actionable recommendations for integrating XAI into organizational decision-making frameworks, emphasizing the potential outcomes and steps to achieve them.

Table 5: Recommendations for XAI Integration

| Recommendation | Practical Steps | Expected Outcomes |
|---|---|---|
| **Training Programs** | Educate Teams | Enhanced Adoption |
| **Regular Audits** | Ensure Compliance | Reduced Risks |
| **Stakeholder Workshops** | Improve Engagement | Better Collaboration |

## 4.2 Implications for Theory and Practice

*Theoretical Implications*

The findings expand the theoretical understanding of XAI by placing it within the broader context of organizational behavior and innovation. Previous research has primarily focused on the technical and ethical dimensions of XAI (Zhou et al., 2021). The findings also expand upon machine learning interpretability methods, building on the techniques outlined by Linardatos et al. (2020). This study extends that foundation by demonstrating how XAI fosters collaboration and creativity across diverse organizational contexts. It also emphasizes the interplay between XAI and organizational culture, suggesting that cultural readiness is a critical factor in realizing the benefits of XAI.

*Practical Implications*

For practitioners, integrating XAI into decision-making frameworks presents opportunities for improved trust, collaboration, and responsible innovation. In manufacturing, for instance, XAI improves efficiency by enabling operational teams to interpret predictive outputs effectively (Gunning & Aha, 2019). In education and HR, XAI enables broader stakeholder involvement by translating AI logic into accessible formats (Miller, 2019; Chowdhury et al., 2023).

The findings also suggest that XAI enhances creativity by empowering users to explore alternatives and adapt solutions based on AI-derived insights (Rai, 2020). This makes XAI not only a risk-management tool but a driver of ideation and experimentation.

Ethical considerations are also central to practical implementation. By adopting fairness-oriented design and auditability features, organizations can align AI systems with inclusive and transparent decision-making processes (Floridi et al., 2018; Binns, 2018).

## 4.3 Limitations and Future Directions

Despite its contributions, the study has limitations. Its reliance on secondary data constrains the ability to provide firsthand empirical validation. The focus on current implementations also limits insight into the long-term organizational transformation XAI might foster.

Future research should include longitudinal studies that trace how XAI influences decision quality, collaboration, and innovation over time. Industry-specific case studies could further illuminate how contextual factors such as organizational readiness, team composition, and regulatory environments affect XAI adoption (Nowell et al., 2017; Patton, 1999).

Moreover, future studies should examine the integration of XAI with emerging technologies, such as large language models, blockchain, and IoT (Cambria et al., 2024). These combinations may yield new applications in computational social science, real-time monitoring, or decentralized systems (Ziems et al., 2024), potentially amplifying both the benefits and challenges of explainability.

# 5 CONCLUSION

Explainable Artificial Intelligence (XAI) is emerging as a vital component in the responsible integration of AI within organizations. By increasing the transparency and interpretability of complex models, XAI supports critical organizational functions—such as decision-making, collaboration, ethical accountability, and creative exploration—across diverse domains.

This review has synthesized literature demonstrating that XAI improves stakeholder trust by offering clear justifications for AI outputs, mitigates bias through enhanced auditability, and fosters cross-functional collaboration by translating technical insights into accessible formats. Furthermore, XAI contributes to innovative and exploratory processes by enabling iterative experimentation with AI-generated insights.

While the benefits of XAI are consistently observed across sectors such as healthcare, finance, manufacturing, agriculture, and education, their practical impact is shaped by organizational readiness, culture, and implementation strategies. These contextual variables must be considered when developing and deploying explainable systems.

Looking ahead, XAI is poised to play an increasingly central role in the evolution of ethical and creative AI use in business and public sectors. Policymakers, industry leaders, and researchers must collaborate to establish robust frameworks that ensure explainability remains a cornerstone of AI governance, innovation, and inclusivity.

## References


Adebayo, J., Gilmer, J., Muelly, M., Goodfellow, I., Hardt, M., & Kim, B. (2018). Sanity checks for saliency maps. Advances in neural information processing systems, 31. https://dl.acm.org/doi/10.5555/3327546.3327621

Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M., Confalonieri, R., & Herrera, F. (2023). Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. Information Fusion, 99, 101805. https://doi.org/10.1016/j.inffus.2023.101805

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... & Horvitz, E. (2019, May). Guidelines for human-AI interaction. In Proceedings of the 2019 chi conference on human factors in computing systems (pp. 1-13). https://dl.acm.org/doi/abs/10.1145/3290605.3300233

Binns, R. (2018, January). Fairness in machine learning: Lessons from political philosophy. In Conference on fairness, accountability and transparency (pp. 149-159). PMLR. https://proceedings.mlr.press/v81/binns18a.html

Blatch-Jones, A. J., Church, H., & Crane, K. (2024). Exploring the potential benefits and challenges of artificial intelligence for research funding organisations: a scoping review. *medRxiv*,2024-09. https://www.medrxiv.org/content/10.1101/2024.09.26.24314280v1

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Cambria, E., Malandri, L., Mercorio, F., Nobani, N., & Seveso, A. (2024). Xai meets llms: A survey of the relation between explainable ai and large language models. *arXiv preprint arXiv:2407.15248*. https://arxiv.org/abs/2407.15248

Chowdhury, S., Dey, P., Joel-Edgar, S., Bhattacharya, S., Rodriguez-Espindola, O., Abadie, A., & Truong, L. (2023). Unlocking the value of artificial intelligence in human resource management through AI capability framework. *Human resource management review*, *33*(1), 100899. https://www.sciencedirect.com/science/article/abs/pii/S1053482222000079

Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. Electronics, 8(8), 832. https://doi.org/10.3390/electronics8080832

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

Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., & Ranjan, R. (2023). Explainable AI (XAI): Core ideas, techniques, and solutions. ACM Computing Surveys, 55(9), 1–33. https://dl.acm.org/doi/abs/10.1145/3561048

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), 80–89. https://doi.org/10.1109/DSAA.2018.00018

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. ACM Computing Surveys (CSUR), 51(5), 1-42. https://dl.acm.org/doi/10.1145/3236009

Gunning, D., & Aha, D. W. (2019). DARPA's explainable artificial intelligence program. AI Magazine, 40(2), 44–58. https://doi.org/10.1609/aimag.v40i2.2850

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., ... & Hussain, A. (2024). Interpreting black-box models: a review on explainable artificial intelligence. *Cognitive Computation*, *16*(1), 45-74. https://link.springer.com/article/10.1007/s12559-023-10179-8

Kaur, D., Uslu, S., Rittichier, K. J., & Durresi, A. (2022). Trustworthy artificial intelligence: A review. ACM Computing Surveys, 55(2), 1–38. https://dl.acm.org/doi/abs/10.1145/3491209

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable ai: A review of machine learning interpretability methods. *Entropy*, *23*(1), 18. https://www.mdpi.com/1099-4300/23/1/18

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence, 267, 1–38. https://doi.org/10.1016/j.artint.2018.07.007

Molnar, C. (2024). Interpretable machine learning: A guide for making black box models explainable. https://christophm.github.io/interpretable-ml-book/

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. International Journal of Qualitative Methods, 16(1), 1–13. https://doi.org/10.1177/1609406917733847

Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. Health Services Research, 34(5), 1189–1208. https://pubmed.ncbi.nlm.nih.gov/10591279/

Rai, A. (2020). Explainable AI: From black box to glass box. Journal of the Academy of Marketing Science, 48(1), 137–141. https://doi.org/10.1007/s11747-019-00710-5

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, *1*(5), 206-215. https://www.nature.com/articles/s42256-019-0048-x

Samek, W., Wiegand, T., & Müller, K. R. (2017). Explainable artificial intelligence: Understanding, visualizing, and interpreting deep learning models. arXiv preprint arXiv:1708.08296. https://arxiv.org/abs/1708.08296

Zhou, J., Gandomi, A. H., Chen, F., & Holzinger, A. (2021). Evaluating the quality of machine learning explanations: A survey on methods and metrics. Electronics, 10(5), 593. https://doi.org/10.3390/electronics10050593

Ziems, C., Held, W., Shaikh, O., Chen, J., Zhang, Z., & Yang, D. (2024). Can large language models transform computational social science?. *Computational Linguistics*, *50*(1), 237-291. https://direct.mit.edu/coli/article/50/1/237/118498/Can-Large-Language-Models-Transform-Computational

## Appendix A. Categorized Literature Review Sources

To provide transparency in literature selection and thematic relevance, the reviewed sources were categorized into two groups:

- Group 1: Papers directly related to XAI's role in fostering innovation and creativity in organizations
- Group 2: Technical, methodological, or foundational works supporting broader understanding of XAI

**Group 1: Organizational Innovation & Creativity (15 sources)**

Ali, S. et al. (2023) – Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence.
Amershi, S. et al. (2019) – Guidelines for human-AI interaction.
Binns, R. (2018) – Fairness in machine learning: Lessons from political philosophy.
Blatch-Jones, A. J. et al. (2024) – AI for research funding organisations: a scoping review.
Cambria, E. et al. (2024) – XAI meets LLMs: A survey of the relation between explainable AI and large language models.
Chowdhury, S. et al. (2023) – AI in HRM through AI capability framework.
Carvalho, D. V. et al. (2019) – Machine learning interpretability: A survey on methods and metrics.
Dwivedi, R. et al. (2023) – Explainable AI (XAI): Core ideas, techniques, and solutions.
Floridi, L. et al. (2018) – AI4People: An ethical framework for a good AI society.
Gilpin, L. H. et al. (2018) – Explaining explanations: Overview of interpretability.
Gunning, D. & Aha, D. W. (2019) – DARPA's explainable artificial intelligence program.
Kaur, D. et al. (2022) – Trustworthy artificial intelligence: A review.
Miller, T. (2019) – Explanation in artificial intelligence: Insights from the social sciences.
Rai, A. (2020) – Explainable AI: From black box to glass box.
Ziems, C. et al. (2024) – Can large language models transform computational social science?

**Group 2: Foundational, Technical, or Methodological Literature (11 sources)**

Adebayo, J. et al. (2018) – Sanity checks for saliency maps.
Braun, V. & Clarke, V. (2006) – Using thematic analysis in psychology.
Doshi-Velez, F. & Kim, B. (2017) – Towards a rigorous science of interpretable machine learning.
Guidotti, R. et al. (2018) – A survey of methods for explaining black box models.
Hassija, V. et al. (2024) – Interpreting black-box models: A review on explainable artificial intelligence.
Linardatos, P. et al. (2020) – Explainable AI: A review of ML interpretability methods.
Molnar, C. (2024) – Interpretable machine learning: A guide for making black box models explainable.
Nowell, L. S. et al. (2017) – Thematic analysis: Trustworthiness criteria.
Patton, M. Q. (1999) – Enhancing the quality and credibility of qualitative analysis.
Rudin, C. (2019) – Stop explaining black box ML models: Use interpretable models instead.
Samek, W. et al. (2017) – Explainable artificial intelligence: Visualizing and interpreting deep learning models.