

Responsible and Trustworthy Artificial Intelligence in Mining, Oil and Gas, and Manufacturing: A Systematic Review

Didula Dissanayaka

ddissanayaka@mun.ca

Memorial University of Newfoundland

Thumeera R. Wanasinghe

Memorial University of Newfoundland

Raymond G. Gosine

Memorial University of Newfoundland

Peter J. Warrian

University of Toronto

Research Article

Keywords: Artificial Intelligence, Trustworthy AI, Explainable AI, Mining Industry, Oil and Gas, Manufacturing

Posted Date: June 3rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9717960/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Responsible and Trustworthy Artificial Intelligence in Mining, Oil and Gas, and Manufacturing: A Systematic Review

Didula Dissanayaka^{1*}, Thumeera R. Wanasinghe¹,
Raymond G. Gosine^{1,2}, Peter J. Warrian²

^{1*}Faculty of Engineering and Applied Science, Memorial University of
Newfoundland, St. John's, NL, Canada.

²Innovation Policy Laboratory, Munk School of Global Affairs and
Public Policy, University of Toronto, Toronto, ON, Canada.

*Corresponding author(s). E-mail(s): ddissanayaka@mun.ca;
Contributing authors: thumeerawa@mun.ca; rgosine@mun.ca;
peter.warrian@utoronto.ca;

Abstract

The integration of artificial intelligence (AI) into mining, oil and gas, and manufacturing industries has created opportunities to improve productivity, efficiency, safety, and reliability. However, the use of AI also raises concerns related to trust, transparency, and ethical considerations. This study presents a systematic literature review of 87 peer-reviewed studies published between 2010 and 2025, examining how the integration of AI affects trust, transparency, and ethical considerations across mining, oil and gas, and manufacturing industries. Following PRISMA 2020 guidelines, relevant publications were identified, screened, and classified based on industry, research focus, methodological approach, applications, and maturity of deployment.

The analysis reveals clear differences between industries. Manufacturing demonstrates the highest level of maturity, with strong emphasis on trust, human-robot collaborative system design, and operational transparency, alongside increasing adoption of explainable AI (XAI) methods. Oil and gas shows notable use of XAI driven by safety assurance and risk management requirements, but limited adoption of AI-specific governance and ethical evaluation. Mining remains at an earlier stage of AI adoption, focusing primarily on performance reliability and safety-critical applications. Overall, the review identifies persistent gaps, including reliance on post-hoc explainability, limited validation of explanations,

and underdeveloped regulatory alignment, highlighting the need for integrated approaches that link trust, transparency, ethics, and governance in industrial AI systems.

Keywords: Artificial Intelligence, Trustworthy AI, Explainable AI, Mining Industry, Oil and Gas, Manufacturing

1 Introduction

The integration of AI into industries such as mining, oil and gas, and manufacturing is transforming how operations are planned, monitored, and executed. In this work, the term “production industries” refers collectively to the mining, oil and gas, and manufacturing industries. From predictive maintenance and autonomous systems to process optimization and decision support, AI has the potential to significantly increase efficiency, safety, and profitability in production industries [1–5]. Currently, governments and industries are increasingly prioritizing AI as a key driver of competitiveness and sustainability, accelerating its adoption into traditionally conservative industrial domains [6, 7].

However, alongside these opportunities, there are critical challenges surrounding trust, transparency, ethics, human-AI collaboration, and governance. Industrial AI systems can be deployed in high-stakes, safety-critical environments, such as drilling operations, large-scale manufacturing systems, or resource exploration processes, where opaque decision-making can undermine human trust and operational safety. Ethical concerns such as bias in predictive models, worker displacement, data privacy, and accountability gaps further complicate their broader integration [8–10]. Moreover, effective human-machine interaction (HMI) and regulatory frameworks are essential for ensuring technical success as well as acceptance and responsible deployment.

A growing body of literature examines aspects of AI integration across production industries, but the focus and maturity vary significantly by industry. In the mining industry, research remains largely conceptual or pilot-scale, with an emphasis on safety monitoring, exploration and resource estimation, and operational optimization, but with limited attention to explainability or governance [1, 2, 8, 9, 11–17]. Oil and gas has a stronger emphasis on risk management, predictive modeling, and ethical considerations, particularly regarding safety and environmental risks, but formal adoption of Explainable AI (XAI) methods and structured HMI frameworks remains scattered [3, 6, 18–25]. In manufacturing, there is broader coverage of AI, including XAI techniques, collaborative robotics, and emerging regulatory discussions. However, even in manufacturing, implementations are fragmented and often technology-driven rather than human or governance-centered [5, 7, 10, 26–56].

Across production industries, XAI consistently emerges in the literature as a key strategy to improve trust and transparency in AI systems. Many studies propose or discuss XAI methods such as model-agnostic explainers, visualization techniques, or interpretable model architectures as potential solutions to bridge the gap between complex AI models and human understanding. Despite this, most applications remain

at the conceptual or early experimental stage, with limited evidence of systematic or industry-wide integration.

Five major themes can be identified across the literature.

1. Trust and transparency: While widely acknowledged as essential, most approaches rely on technical performance or post-hoc explanations, rather than embedding explainability and interpretability into system design.
2. Ethics: Ethical discussions remain conceptual and lack actionable frameworks, with limited industry-specific adaptation.
3. Explainable AI: XAI appears consistently in the literature as a proposed means to enhance transparency and trust, yet it remains a conceptual addition rather than a systematically integrated component of industrial AI systems.
4. Human-machine interaction: HMI is largely underexplored outside of manufacturing, with mining and oil and gas focusing on automation rather than collaboration.
5. Regulation and governance: There is an absence of clear regulatory alignment or governance mechanisms, especially in mining and oil and gas, where deployment often outpaces oversight.

Despite a growing interest in responsible AI, no systematic review currently synthesizes how trust, ethics, XAI, HMI, and governance are addressed across the mining, oil and gas, and manufacturing industries. Existing reviews primarily focus on individual technologies, single industries, or narrow ethical dimensions [10, 25, 57]. As a result, there is limited understanding of cross-industry challenges, transferable solutions, and industry-specific gaps.

This paper addresses this gap by conducting a systematic review of academic literature on AI integration across these three production industries, focusing on trust/transparency, ethics, XAI, HMI, and regulations. By analyzing over 80 papers related to AI and production industries published since 2010, this review provides:

- A structured synthesis of current practices in each industry;
- Identification of gaps and limitations in how responsible AI aspects, including XAI, are deployed;
- A comparative analysis highlighting differences and commonalities across industries; and
- Recommendations for future research and policy directions to support responsible and trustworthy adoption of AI in the production industries.

2 Background

The application of AI across production industries has grown substantially over the past decade, driven by advances in automation, data analytics, and sensing technologies. However, the extent and maturity of AI adoption differ across these industries due to their distinct operational risks, data environments, and regulatory contexts. Although AI promises greater safety, efficiency, and sustainability, the recent literature emphasizes that these benefits can only be realized if the issues of trust, transparency, and ethical accountability are addressed [1, 2].

2.1 AI in the Mining Industry

AI research in mining has focused mainly on predictive analytics, resource estimation, and safety monitoring [1, 11, 14, 58]. These applications aim to lower operational risks and improve efficiency through data-driven insights. However, explainability and transparency have not been explored in these studies, and the development of responsible AI frameworks is at a very early stage. A few recent works have begun to discuss ethical implications such as data privacy and the social effect of automation [9, 17], but systematic approaches to governance or human–AI collaboration remain limited.

2.2 AI in the Oil and Gas Industry

In the oil and gas industry, AI has been widely used for predictive maintenance, reservoir modeling, and risk assessment [3, 19]. As these applications directly affect safety-critical operations, several studies highlight the importance of transparency and reliability [3, 19, 22, 59]. Furthermore, ethical and environmental considerations are also increasingly discussed in connection with AI-based decision support [6, 21]. Although XAI has been identified as a potential tool to enhance operator confidence, practical implementation remains limited, and regulatory frameworks for AI usage are at an initial stage.

2.3 AI in the Manufacturing Industry

Among the production industries studied, manufacturing demonstrates the deepest and most mature integration of AI. Applications include predictive quality control, process optimization, and collaborative robotics [26, 31, 60]. In this industry, explainability and transparency have gained traction, with multiple studies in which interpretable models and human-in-the-loop frameworks have been investigated [36, 37, 61–63]. The discussion of ethics and regulations is more evident in this industry, particularly in relation to HMI and the challenges of standardization to operational workflows and AI-integration [46, 64]. Despite this progress, research is still somewhat fragmented, and comprehensive governance frameworks are still underdeveloped.

While each industry has explored AI integration to different extents, the broader understanding of how trust, ethics, explainability, HMI, and governance are put into practice remains incomplete. Existing studies tend to focus on a specific issue or a single industry, providing limited insight into transferable lessons or cross-industry patterns. This gap is indicative of the need for a systematic review that consolidates current evidence and compares the maturity, challenges, and opportunities for responsible AI adoption across mining, oil and gas, and manufacturing industries.

3 Methodology

This systematic review follows a structured and transparent process to identify, select, and analyze scientific literature addressing the impact of AI integration within three production industries: mining, oil and gas, and manufacturing.

3.1 Study Design

To conduct this review, we applied a clearly defined multi-step process led by established guidelines for systematic literature reviews [65] and aligned with the PRISMA 2020 reporting framework [66] to ensure transparency and reproducibility. All steps were carried out between May 2025 and January 2026. The overall design is summarized below:

1. **Research Questions:** We formulated eight research questions addressing trust and transparency, ethics, explainable AI, human–robot interaction, governance frameworks, maturity and real-world deployment, future opportunities and challenges, and the geographical and institutional distribution of research.
2. **Literature Search Strategy:** A systematic search strategy was developed to identify relevant peer-reviewed academic publications in two major databases (Scopus and Web of Science), using industry-specific keyword combinations.
3. **Data Cleaning and Eligibility Criteria:** Inclusion and exclusion criteria were defined to ensure that only credible sources were considered. These criteria covered publication type, language, time frame, industrial domain, and thematic relevance.
4. **Screening and Relevance Identification:** Both manual and AI-assisted screening (using the ASReview tool [67]) were applied to remove duplicates, exclude irrelevant papers, and identify the most relevant studies for further analysis.
5. **Data Extraction:** Data were systematically extracted using a structured template capturing thematic elements such as trust/transparency vs. ethics, XAI usage, HMI, governance, maturity levels, real-world implementations, and opportunities/challenges.
6. **Data Analysis:** Descriptive and thematic analyses were performed to identify recurring patterns, trends, and gaps across the three industries, enabling a deeper understanding of how AI integration is conceptualized and operationalized in different contexts.

The following sections provide detailed descriptions of each step.

3.2 Research Questions

The review focuses on understanding how the integration of AI affects issues of trust/transparency, ethics, and related governance mechanisms across the selected industries. Specifically, the research is guided by the following questions:

1. RQ1: How does the integration of AI technologies affect public trust or transparency in mining, oil and gas, and manufacturing industries and what factors influence them?
2. RQ2: What ethical considerations arise from AI deployment in these production industries, and how are they discussed in the literature?
3. RQ3: To what extent is XAI adopted to support transparency, trust, or ethical decision-making in production industries?
4. RQ4: How is HMI discussed in relation to building trustworthy human–AI collaboration?

5. RQ5: What governance or regulatory frameworks are discussed or proposed for AI use in production industries?
6. RQ6: What is the maturity level (e.g., conceptual, proposed, pilot-tested, implemented) of AI, XAI, and governance approaches in these industries, and what real-world implementations or evidence of effectiveness have been reported?
7. RQ7: What opportunities and challenges are identified for responsible and trustworthy future AI integration across these industries?
8. RQ8: How is AI-related research in mining, oil and gas, and manufacturing distributed over time, across countries, and among different institutional affiliations?

This structured set of research questions facilitates a comprehensive assessment of both the theoretical and practical dimensions of AI integration, as well as forward-looking implications for industry adoption and policy development.

3.3 Literature Search Strategy

The literature search targeted peer-reviewed English-language publications published since 2010 in two leading bibliographic and citation indexing databases: Scopus and Web of Science.

The search strategy combined three sets of terms to capture relevant literature across domains:

1. AI-related terms: “artificial intelligence”, “AI”, “machine learning”, “deep learning”, “autonomous systems”
2. Societal and governance terms: “trust”, “trustworthiness”, “public perception”, “transparency”, “ethics”, “credibility”
3. Industry-specific terms: “mining”, “oil and gas”, “manufacturing”, “production”

An example search query for the oil and gas industry is shown below:

(“artificial intelligence” OR “AI” OR “machine learning” OR “deep learning”) AND (“trust” OR “trustworthiness” OR “public perception” OR “ethics” OR “credibility”) AND (“oil” AND “gas”)

Similar structures were applied for mining and manufacturing, with mining-related searches excluding unrelated domains such as “data mining”, “text mining”, “bitcoin mining” to identify related publications.

The search returned a total of 3,327 publications across the two indexing databases, with the following breakdown shown in Table 1:

Table 1: Number of publications identified from Scopus and Web of Science

Topic	Scopus	Web of Science
Mining	221	101
Oil and Gas	62	19
Manufacturing	1,972	952
Total	2,255	1072

3.4 Data Cleaning and Eligibility Criteria

After keyword-based search, the resulting publications underwent a two-step filtering process:

1. Duplicate and retraction removal: Duplicates between Scopus and Web of Science were eliminated. Retracted papers were excluded for academic integrity.
2. Eligibility criteria: Inclusion was limited to peer-reviewed journal articles, conference papers, or book chapters that were written in English, published from 2010 onward, and discussed AI integration within at least one of the three industries.

3.5 Screening and Relevance Identification

The overall screening process, including record identification, duplicate removal, eligibility assessment, and final inclusion, followed the PRISMA 2020 guidelines. The complete workflow is illustrated in Figure 1, which summarizes the progression from the initial dataset of 3,327 records to the final set of 87 studies included in the review. Screening was conducted in two different methods, depending on the size of the retrieved dataset in each industry:

- Manual screening: Used for oil and gas and mining literature. Titles, abstracts, and where necessary full texts were reviewed to assess relevance to the research questions.
- AI-assisted screening using ASReview: For the large manufacturing and production dataset, the ASReview active learning tool was employed to efficiently prioritize relevant studies. The system iteratively ranked papers based on manual relevance labels until the stopping criterion was met (20 consecutive irrelevant papers in this occasion).

For the manufacturing dataset, ASReview was used to accelerate the screening process, since manual review alone would have been significantly more time-consuming. The procedure began with the manual labelling of a small set of papers as relevant or not relevant. Using these initial labels, ASReview ranked the remaining papers based on their estimated likelihood of relevance. The highest-ranked paper was then manually reviewed and assigned a relevance label, after which the system automatically updated the ranking. This iterative manual–AI hybrid process continued until 20 consecutive papers were identified as irrelevant, at which point all remaining papers were excluded from further screening.

To ensure the reliability of the ASReview process, we validated its performance using the oil and gas and mining datasets, which were initially screened manually. These same datasets were later processed through ASReview, and the tool’s selections closely matched the manual decisions. This confirmed that ASReview performed acceptably and could be confidently applied to the much larger manufacturing dataset.

The total number of relevant papers identified after screening is summarized in Table 2:

Table 2: Screening results across industries

Industry	Total Papers	After Cleaning	Relevant Papers	Full text accessed
Mining	322	195	14	13
Oil and Gas	81	56	17	12
Manufacturing/Production	2,924	2,015	89	62
Total	3,327	2,266	120	87

3.6 Data Extraction

Data extraction was conducted using the systematic review management platform DistillerSR [68]. Three distinct extraction forms were used:

1. Bibliographic and contextual information: The first form captured basic metadata for each paper, including title, authors, publication year, source type (journal, conference, or book chapter), and country.
2. Core thematic content: The second form focused on the core contribution and thematic relevance of each paper. Each paper was coded according to whether its primary focus was on trust/transparency or ethics. Additional coding captured whether the paper discussed or applied XAI methods, addressed HMI, or included regulatory or governance perspectives. Follow-up questions were used to document the nature and depth of these discussions, such as the type of XAI techniques used, the framing of ethical issues, or the scope of regulatory frameworks considered.
3. Maturity, opportunities, and challenges: The third form assessed the maturity level of the AI applications or frameworks described (conceptual, proposed, pilot-tested, or implemented) and extracted any explicit mentions of future opportunities or challenges related to trust, ethics, XAI, HMI, or governance. This allowed for the systematic identification of trends across different levels of technological and organizational adoption.

The use of multiple customized forms enabled a structured, layered approach to data capture, ensuring that both descriptive and analytical dimensions were systematically covered. All papers were reviewed and coded, with ongoing consistency checks to ensure stable interpretation and classification throughout the extraction process.

3.7 Data Analysis

Data analysis followed a combination of descriptive and thematic evaluation approaches. First, descriptive statistics were compiled to provide an overview, including trends in publication over time, industry distribution, and the prevalence of themes related to trust, ethics, XAI, HMI, and governance, answering the last research question (RQ8) in Section III.B. This provided a quantitative foundation for understanding how the focus of research varies between industries and how it varies over time.

Second, thematic analysis was performed to identify recurring patterns and conceptual trends across and within the three industries. The thematic coding was directly aligned with the research questions one to seven (RQ1 to RQ7) outlined in Section III.B.

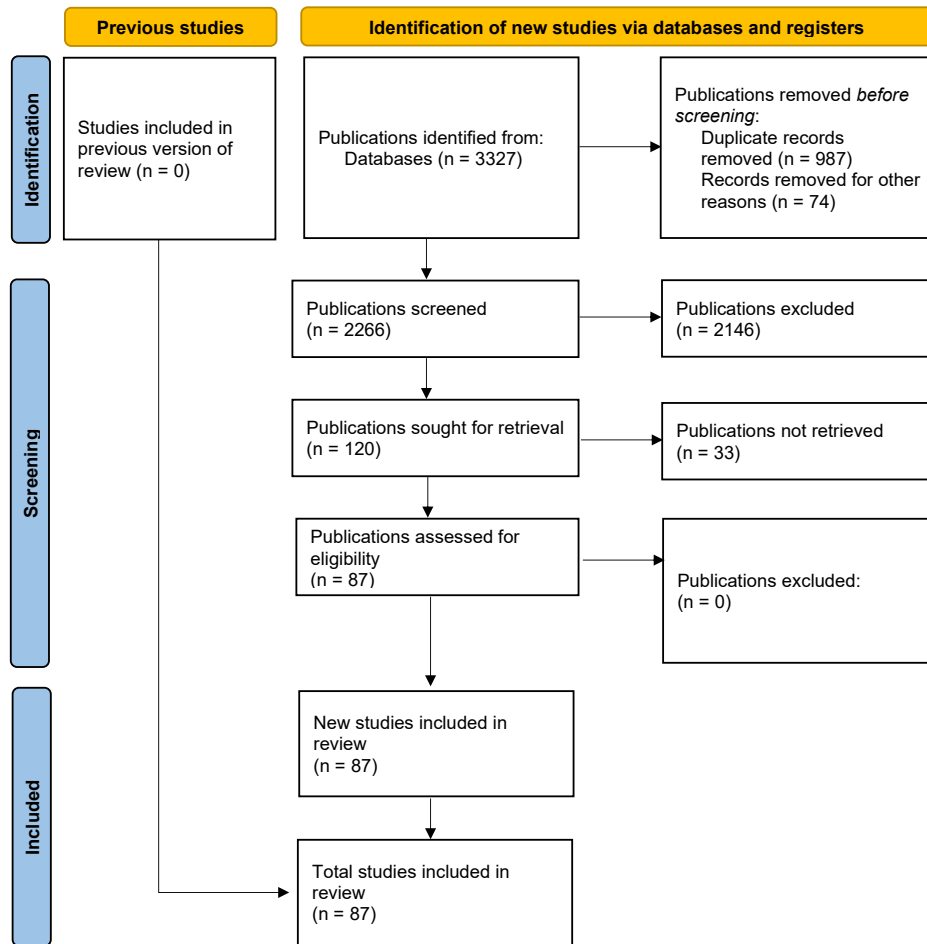


Fig. 1: PRISMA 2020 flow diagram for systematic reviews which included searches of databases and registers only

Comparative thematic mapping was then conducted to highlight similarities and differences between mining, oil and gas, and manufacturing. This involved identifying both common cross-industry trends (e.g., widespread conceptual focus on XAI but limited real-world implementation) and industry-specific gaps (e.g., limited HMI discussions in mining). These thematic insights form the basis of the results and discussion sections, providing a structured overview of the state of research regarding responsible AI integration in production industries.

4 Comparative Analysis Results

This section presents a comparative analysis of AI integration within the three major production industries across five thematic dimensions: trust and transparency, ethics, XAI, HMI, and regulations and governance.

To visualize the distribution of these themes, a Sankey diagram is presented in Figure 2. The diagram illustrates how research emphasis flows across industries, indicating that manufacturing contributes the largest share of studies related to trust and HMI, oil and gas demonstrates the strongest concentration of XAI-focused work, and mining shows a narrower but safety-driven thematic profile. The width of each flow corresponds to the number of papers addressing each theme, making visible the imbalance and fragmentation of AI research across industries. In the figure, the total count across themes is greater than 87 because a single paper can be categorized under multiple themes. For example, if a study addresses both trust and XAI, it is counted once in each category.

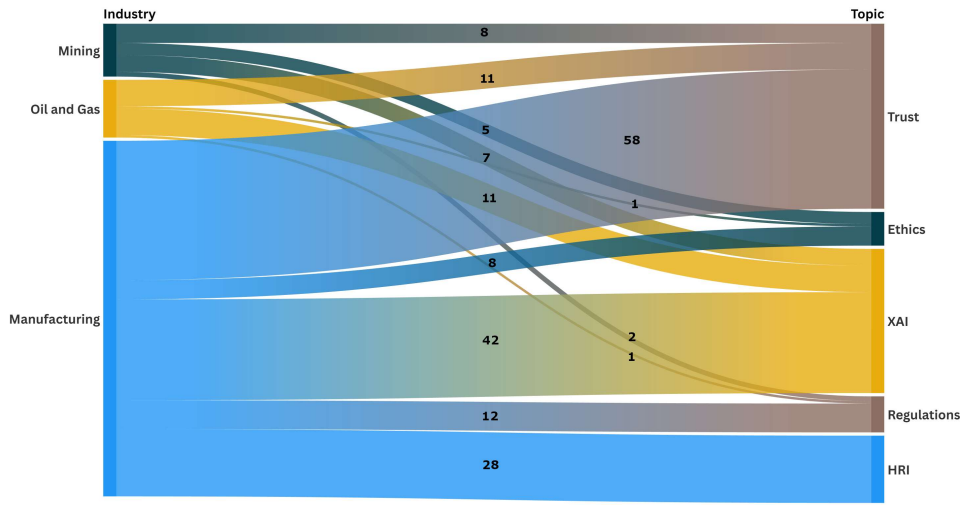


Fig. 2: Sankey diagram for main themes

4.1 Trust and Transparency

Trust and transparency represent the foundation for responsible AI adoption across all production industries, yet their meaning and implementation differ according to each industry's operational environment. Quantitatively, trust and transparency are discussed in 62% of mining papers, 92% of oil and gas papers, and 94% of manufacturing papers. Despite this broad attention, the way these concepts are conceptualized and operationalized differs considerably from one industry to another.

In the mining industry, discussions of trust primarily focus on safety assurance and the reliability of predictive systems used for ground stability monitoring, ore-grade estimation, and haul-truck automation. Many studies emphasize the need for uncertainty quantification, anomaly detection, and interpretable alerts to improve confidence in model outputs [1, 8, 9, 11, 15, 69, 70]. Transparency, in this context, is often defined in technical terms, such as traceability, auditability, and parameter sensitivity, rather than in relation to human interpretability or user trust [8, 9, 15, 70]. The focus remains on how well autonomous systems perform under operational uncertainty, indicating that trust is still treated largely as a function of reliability rather than interactive understanding.

The oil and gas industry presents a different narrative. Here, trust is mainly considered from a risk-focused and compliance-oriented perspective. Studies link trustworthy AI to accountability mechanisms within safety-critical applications, including drilling control, leak prediction, and asset integrity management [3, 19, 23, 59, 71]. Transparency is used as a regulatory instrument rather than a feature of user-facing design, enabling verification of AI-generated decisions through traceability logs, data lineage records, and validation protocols [19, 23, 59]. In oil and gas, explainability is explicitly connected to verification and audit processes, supporting both internal governance and external compliance requirements.

Manufacturing demonstrates a more developed and practically applied approach to trust and transparency. Across a wide range of applications, including predictive maintenance, quality inspection, scheduling, and collaborative production systems, transparency is increasingly embedded into operator-facing decision-support tools. Explanations are integrated into dashboards and interactive interfaces to support real-time oversight, diagnosis, and intervention [7, 32, 47, 72–76]. Trust is treated as a dynamic attribute that evolves through continuous interaction, feedback, and demonstrated system behavior, rather than as a static consequence of model accuracy alone [32, 61]. This reflects a shift toward human-centered transparency mechanisms that support effective human–AI collaboration in production environments.

Across the three industries considered, a number of recurring factors consistently shape and influence trust and transparency. These include the stability and consistency of explanations, the perceived reliability of models under changing conditions, and the extent to which decision-making processes align with human mental models and established safety case frameworks. Representative examples include uncertainty-aware dashboards in mining [11], management of change explanation logs and audit trails in oil and gas [23, 71], and saliency-based visual feedback systems and interactive explanation interfaces in manufacturing [7, 32]. Although manufacturing increasingly evaluates these factors empirically, mining and oil and gas continue to approach trust mainly through performance metrics and compliance requirements.

Overall, mining frames trust largely in terms of technical reliability, oil and gas aligns it with regulatory accountability and risk management, and manufacturing treats trust and transparency as design attributes necessary for effective human–AI collaboration. Together, these perspectives illustrate a gradual shift across industries

from performance-based trust toward more transparent, participatory, and human-aligned AI systems, while also revealing uneven progress in how trust is operationalized and validated in practice.

4.2 Ethics

Ethical considerations represent an essential dimension of responsible AI deployment but remain unevenly addressed across production industries analyzed. Quantitatively, 39% of mining papers, 8% of oil and gas papers, and 13% of manufacturing papers discuss ethical issues explicitly. Where present, ethical analyses tend to be mostly theoretical, focusing on moral principles and values rather than on measurable evaluation frameworks or applied audits.

In the mining industry, ethics are frequently tied to community impact, environmental protection, and responsible resource extraction [2, 15]. Several studies discuss how AI-driven automation and predictive models influence worker safety, employment conditions, and the broader social-environmental footprint of mining operations. Ethical considerations also include transparency in decision-making related to exploration, extraction, and safety-critical monitoring, particularly where AI outputs may affect community trust or regulatory approval processes [70]. However, the discussion often ends with acknowledging these implications rather than implementing them. Practical frameworks such as fairness metrics, bias detection procedures, or ethical impact assessments integrated into AI pipelines are largely absent, resulting in a persistent gap between ethical intent and technical implementation.

In the oil and gas industry, ethical considerations are addressed far less frequently and are typically embedded within broader discussions of safety, environmental protection, and operational risk. While studies frame moral responsibility implicitly through accident prevention, environmental stewardship, and risk mitigation under uncertainty [3, 24], a small subset of work explicitly engages with ethical aspects of AI-supported decision-making, particularly in relation to accountability, transparency, and the social implications of automated control systems [59]. Nevertheless, ethical reasoning in this industry remains closely aligned with existing safety and compliance frameworks rather than stand-alone ethical evaluations.

In the manufacturing industry, the development of ethical AI is increasingly discussed within the context of human-centered automation. Studies examine how AI systems can support worker well-being, improve workplace ergonomics, and enable fair and efficient task allocation between humans and machines [72, 73, 77, 78]. Human oversight, accountability, and the preservation of operator autonomy are recurring themes, with multiple papers emphasizing that humans should retain ultimate control over automated decisions, particularly in collaborative and safety-sensitive environments [32, 47]. However, similar to mining and oil and gas, these discussions rarely include measurable indicators such as bias testing, transparency audits, or explainability benchmarks, keeping the ethical dimension largely aspirational rather than operational.

Across all three industries, common ethical themes emerge, including prevention of harm, fairness, accountability, and human oversight. Privacy-by-design and data minimization principles are frequently cited as desirable attributes of AI systems but

Table 3: XAI algorithms used across industries, frequency of implementation, and corresponding references

Industry	XAI Algorithm	Count	Reference(s)
Mining	SHAP	4	[9, 11, 16, 69]
	LIME	1	[12]
Oil and Gas	SHAP	3	[3, 6, 18]
	LIME	1	[20]
	Counterfactual Reasoning	2	[19, 20]
	SHAP	14	[51–53, 56, 62, 72, 79–86]
Manufacturing	LIME	6	[46, 62, 72, 79, 86, 87]
	Gradient-Based Methods	1	[81]
	Grad-CAM	1	[83]
	Decision Trees	1	[61]
	Saliency Maps	1	[88]
	Class Activation Maps	1	[26]
	GSInquire	1	[5]

are rarely implemented in practice. Collectively, these findings suggest that while awareness of AI ethics is increasing across production industries, ethics is still largely treated as a guiding philosophy rather than as a formalized, evidence-based component of AI system design, deployment, and governance.

4.3 Explainable Artificial Intelligence (XAI)

XAI plays a central role in building transparency and trust in AI-driven decision systems across production industries. However, its adoption and level of integration vary substantially among industries. Quantitatively, XAI is discussed in 54% of mining papers, 92% of oil and gas papers, and 68% of manufacturing papers. Among these, applied XAI occurs in 47% of mining, 50% of oil and gas, and 39% of manufacturing papers, while theoretical or conceptual discussions account for 15% in mining, 30% in oil and gas, and 20% in manufacturing.

In the mining industry, explainability is emerging primarily through post-hoc visualization methods such as Shapley additive explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME), applied to safety prediction, ground stability analysis, and ore-grade estimation tasks [11, 15, 16, 69]. These methods aim to identify influential geological or environmental variables affecting model outputs and to support safety-oriented decision-making. However, most mining studies stop at visual explanation without formal validation by human experts, leaving the interpretability process largely untested in real operational contexts. Consequently,

explainability in mining is typically treated as an analytical supplement rather than an operational requirement.

In the oil and gas industry, explainability is more mature and closely aligned with risk management objectives. Model-agnostic techniques such as SHAP, LIME and counterfactual reasoning are used to enhance transparency in asset integrity monitoring, fault detection, and safety compliance [3, 6, 18, 19]. Here, the purpose of XAI extends beyond model debugging; it contributes to verification, audit documentation, and communication of risk-related insights to engineers and regulators. This integration of explainability into assurance processes reflects the industry’s broader focus on traceability and accountability.

Manufacturing exhibits the most integrated and application-oriented use of XAI, combining visual and statistical interpretability techniques across production diagnostics, process control, quality inspection, and predictive maintenance [26, 29, 46, 47, 51–53, 56, 62, 79–84, 88]. Implemented methods include SHAP, LIME, gradient-based attribution, decision trees, saliency maps, Grad-CAM, Class activation maps, and GSInquire, which are often deployed through operator-facing dashboards to support real-time decision-making [89, 90]. In several implementations, inherently interpretable models such as decision trees are preferred for tasks including fault classification and quality assessment, allowing explanations to be derived directly from the model structure rather than through post-hoc approximation [61, 74]. These studies highlight explainability as a key facilitator of human–machine collaboration, allowing operators to understand and act upon AI-generated insights. However, the assessment of the quality of explanations in real-time, for example, their stability, consistency, or ease for humans, remains underdeveloped.

The distribution of XAI algorithms across industries, summarized in Table 3, highlights the predominance of post-hoc and localized explanation techniques that interpret model outputs after decisions are made rather than embedding interpretability at the design stage. Moreover, SHAP dominates XAI implementations across production industries, particularly in manufacturing, while other methods appear primarily in isolated or task-specific deployments. Very few studies assess the robustness of explanations under data drift, conduct user-centered evaluations of interpretability effectiveness, or integrate XAI outputs into formal governance mechanisms such as model cards, sign-off templates, or audit checklists. These gaps indicate that while XAI adoption is expanding across production industries, its practical validation, standardization, and alignment with human and regulatory requirements remain at an early stage of development.

4.4 Human–Machine Interaction (HMI)

HMI represents a key component of AI adoption in production industries, shaping how humans collaborate, supervise, and coexist with intelligent systems. However, its level of maturity differs significantly across industries. Quantitatively, HMI is discussed in 45% of manufacturing papers and is not presented in the reviewed mining or oil and gas literature.

In the manufacturing industry, HMI research is relatively mature and closely linked to the growing presence of collaborative robots, intelligent workstations, and

AI-assisted decision-support systems on production floors [38, 51, 61, 79, 86, 88, 91]. Studies frequently examine how human operators and robotic or AI-enabled systems share physical and cognitive workspaces, coordinate tasks, and exchange information through multi-modal interfaces. Key topics include trust calibration between humans and autonomous agents, workload management, adaptive role allocation, and safety in close-proximity operations. Several papers investigate human-in-the-loop control strategies, where operators supervise, intervene, or override AI-driven decisions during quality inspection, assembly, or fault diagnosis tasks [62, 79, 88, 92].

Beyond physical human-robot collaboration, a subset of manufacturing studies extends HMI concepts to human-AI interaction in supervisory and decision-support settings. These works emphasize interface design, explanation timing, and feedback mechanisms that allow operators to understand model reasoning, assess confidence, and adjust their reliance on automated recommendations [61, 80, 81]. Explainability features—such as visual indicators of system intent, confidence cues, and predictive alerts—are highlighted as essential for maintaining situational awareness and reducing operator anxiety [93]. Collectively, these studies demonstrate how transparency, interaction design, and ergonomics are increasingly treated as integral components of human-centered automation in manufacturing.

By contrast, mining and oil and gas remain in the early stages of HMI integration. Most automation in these industries operates through pre-programmed or supervisory control systems with minimal direct interaction between humans and AI agents [6, 14]. In mining, where operations are increasingly shifting toward remote and semi-autonomous modes, the lack of interactive feedback mechanisms poses challenges for safety monitoring, task supervision, and anomaly response. Similarly, in oil and gas, autonomous systems are typically embedded within process control or inspection pipelines, offering limited opportunities for collaborative or adaptive decision-making. As these industries progress toward more digitalized and interconnected operations, incorporating structured HMI frameworks such as adaptive autonomy, explainable alerts, and operator-in-the-loop safety mechanism will become critical to ensuring both trust and operational resilience.

Overall, the contrast among industries highlights how manufacturing is transitioning toward interactive, human-aware automation, while mining and oil and gas remain dominated by isolated or fully autonomous systems. Bridging this gap could involve transferring HMI design principles from manufacturing such as adaptive autonomy, explainable interfaces, and trust calibration mechanisms, into safety-critical mining and oil and gas field operations, paving the way for safer and more transparent AI collaboration in industrial contexts.

4.5 Regulations and Governance

Regulatory and governance frameworks are fundamental to accountability in AI-driven industrial systems. However, their development and formal integration vary widely across production industries. Quantitatively, governance is discussed in 15% of mining papers, 8% of oil and gas papers, and 19% of manufacturing papers. Although all three industries operate under strict safety and compliance mandates, few studies explicitly extend these frameworks to AI-specific governance or algorithmic transparency.

In the mining industry, governance discussions are primarily rooted in occupational safety, environmental protection, and regulatory oversight of resource extraction [1, 2, 13, 70]. These studies highlight the potential of AI to strengthen monitoring, hazard prediction, and policy compliance but note a lack of standardized procedures for model accountability. Governance considerations are often framed in terms of compliance with existing safety regulations rather than proactive AI assurance. Most initiatives remain reactive, addressing compliance through safety audits rather than embedding governance mechanisms within the design and deployment of AI systems. Consequently, formal structures for algorithmic traceability, documentation, or certification are still underdeveloped.

The oil and gas industry maintains robust regulatory systems for operational safety, integrity management, and environmental control, and governance-related discussions are typically embedded within broader safety and risk management frameworks. Most studies treat these established protocols as implicitly supporting trustworthy AI operations, particularly through mechanisms such as risk management documentation, safety barrier verification, and change-control audits [3, 25]. Only a limited subset of the literature explicitly engages with regulatory alignment in the context of AI-enabled decision-making [71], highlighting the absence of AI-specific governance structures despite the industry’s mature compliance environment. This suggests that although governance frameworks are well established in oil and gas operations, they have not yet been systematically adapted to accommodate the evolving and adaptive behavior of AI systems.

Manufacturing shows the most visible progress toward integrating AI with established regulatory and governance frameworks. Several studies reference alignment with international standards, particularly ISO and IEC guidelines for machine safety, quality assurance, and industrial automation [10, 46, 50, 54]. These studies recognize the need to extend conventional compliance models to account for AI-driven decision logic, data dependency, and adaptive system behavior [52, 73, 85, 87]. Moreover, recent studies emphasize integrating AI governance into existing quality management systems through documentation practices, version control, traceability requirements, and explainability audits [94–96]. However, most standards remain focused on mechanical and process safety, leaving gaps in algorithmic assurance, validation of learning-based systems, and continuous monitoring of AI performance.

Across all industries, regulatory engagement with AI remains fragmented and often indirect. While mining and oil and gas continue to rely heavily on conventional safety compliance and risk-based governance, manufacturing has begun to bridge the gap between mechanical regulation and algorithmic accountability. To address these gaps, the development of dedicated AI assurance frameworks covering model documentation, risk assessment, and interpretability will be essential to ensure transparent, auditable, and responsible deployment of AI systems in industrial operations.

4.6 Maturity and Deployment

Across the reviewed literature, the level of AI maturity and deployment varies considerably across industries. As summarized in Table 4, the manufacturing industry stands out as the primary driver of AI development and implementation, demonstrating

substantial progression from conceptual research toward pilot testing and real-world deployment. In total, manufacturing includes twenty-two theoretical studies, thirty pilot-tested systems, and nine real-world implementations, reflecting increasing operational adoption.

In contrast, the oil and gas industry remains more conservatively positioned, with one theoretical study, six pilot-tested systems, and three real-world deployments reported, indicating selective but high-impact adoption in safety-critical contexts. The Mining industry shows gradual advancement, with four conceptual or proposal-stage studies, five pilot-tested systems, and three real-world deployments identified.

4.7 Cross-Industry Gaps

Despite notable progress across the three industries, several systemic gaps continue to hinder the development of transparent, ethical, and accountable AI systems. These gaps appear consistently in mining, oil and gas, and manufacturing, although their severity and implications vary according to the maturity of AI integration in each industry. A consolidated overview of these cross-industry gaps and associated best-practice recommendations is presented in Table 5.

Gap 1 - Transition from post-hoc to design-level explainability: While 60% to 90% of papers across all industries discuss XAI and applied implementations are increasingly reported, most implementations remain post-hoc, providing retrospective justification rather than influencing model design. None of the industries demonstrates widespread integration of explainability during algorithm development or system architecture design.

Gap 2 - Lack of explanation validation, and stability testing: Fewer than 10% of papers across all industries assess explanation consistency under model drift or validate interpretability results with human experts, even in studies reporting pilot testing or real-world deployment. This reduces the reliability of XAI outputs, especially in safety-critical applications.

Gap 3 - Underdeveloped governance and accountability structures: Across all three industries, explicit governance discussions appear in less than 20% of the reviewed literature, indicating limited attention to accountability and oversight. Currently, there are no unified audit templates, certification mechanisms, or AI-specific reporting standards in these industries.

Gap 4 - Limited transfer of HMI practices across industries: Although HMI is addressed in 45% of manufacturing papers, it is entirely absent in mining and oil and gas papers. As these industries move toward remote and semi-autonomous operations, the absence of collaborative and explainable human-AI interaction frameworks remains a significant gap.

Gap 5 - Insufficient socio-technical and cultural diversity: Across all domains, datasets, validation studies, and evaluation scenarios remain predominantly Western and simulation-based. Only a small fraction of the papers include real-world, cross-cultural, or multilingual deployments, which limits the generalizability and inclusivity of the model.

Table 4: Challenges and opportunities across industries

Industry	Key Challenges	Opportunities and Future Directions
Mining	• Trust discussed in 62% of studies; primarily framed through system reliability rather than transparency. • Moderate XAI adoption (54%); limited validation of explanations with domain experts. • Weak governance references (15%); no dedicated AI compliance models. • Minimal attention to HMI or operator feedback in automated systems.	• Develop interpretable and uncertainty-aware models for safety-critical tasks. • Integrate trust dashboards for real-time safety monitoring. • Connect XAI outputs to incident review and operational reporting. • Introduce HMI mechanisms for remote supervision and teleoperation. • Embed governance mechanisms during model design rather than post-deployment audits.
Oil and Gas	• High XAI presence (92%), yet limited integration with formal governance. • Governance primarily implicit, embedded within existing safety and compliance systems. • Minimal discussion of ethics and absence of structured HMI frameworks. • Selective real-world deployment due to high operational and regulatory risk.	• Utilize applied XAI for regulatory compliance, risk traceability, and verification. • Formalize algorithmic accountability within existing risk and change-management systems. • Extend counterfactual and scenario-based explanations for safety assurance. • Transfer HMI lessons from manufacturing to control room operations.
Manufacturing	• Strong trust focus (94%), but applied XAI remains moderate (39%). • Ethics discussed in only 13% of studies; fairness rarely quantified. • Partial governance adoption (19%), mainly aligned with ISO/IEC standards. • HMI emphasized (45%), yet long-term user adaptation remains underexplored.	• Expand deployment-ready XAI tools for predictive maintenance and quality inspection. • Define quantitative trust calibration metrics for human–AI collaboration. • Integrate AI governance into quality management systems and production life-cycle control. • Scale successful pilot implementations toward standardized real-world deployment.

Table 5: Cross-industry gaps and recommended best practices (Indicative coverage drawn from comparative analysis)

Identified Gap	Recommended Best Practice
Post-hoc-only explainability	Integrate explainability into system design stages using interpretable-by-construction models, and standardize design-time XAI benchmarks validated by domain experts.
Limited explanation validation	Establish explanation evaluation protocols with operator studies and stability checks to ensure interpretability and reliability under model drift.
Underdeveloped AI governance	Develop industry-specific AI assurance frameworks linked to existing ISO and IEC safety standards, including standardized evidence packs, model cards, and transparency reports.
Missing HMI outside manufacturing	Transfer human–AI collaboration practices such as trust calibration, shared autonomy, and operator-in-the-loop control from manufacturing to mining and oilfield operations.
Limited socio-cultural diversity	Promote global research partnerships to create multilingual, culturally diverse datasets and establish evaluation metrics for fairness and inclusivity.

4.8 Affiliations and Geographical Distribution of Research

This section analyzes where responsible AI-related research takes place across the mining, oil and gas, and manufacturing industries, and who is involved in carrying out this research. The focus is on publication trends, geographical distribution, and institutional affiliations.

4.8.1 Temporal Distribution of Publications

The number of studies addressing AI integration and its implications on trust, ethics, and governance in production industries has grown steadily over the past decade. Figure 3 presents the yearly publication count categorized by type—conference papers, journal articles, and book chapters. The earliest paper in the dataset appears in 2014 as a single conference publication. After a gap, activity reappeared in 2018 with a book chapter.

A clear upward trend begins in 2021, with notable growth in both conference and journal publications. The period 2023–2025 represents the most active years:

- 2023: 15 total publications (7 conferences, 8 journals)
- 2024: 22 total publications (14 conferences, 8 journals)
- 2025: 36 total publications (12 conferences, 22 journals, 2 Book chapters)

This increase aligns with the broader global rise in responsible AI and interest in explainability and governance in industrial settings.

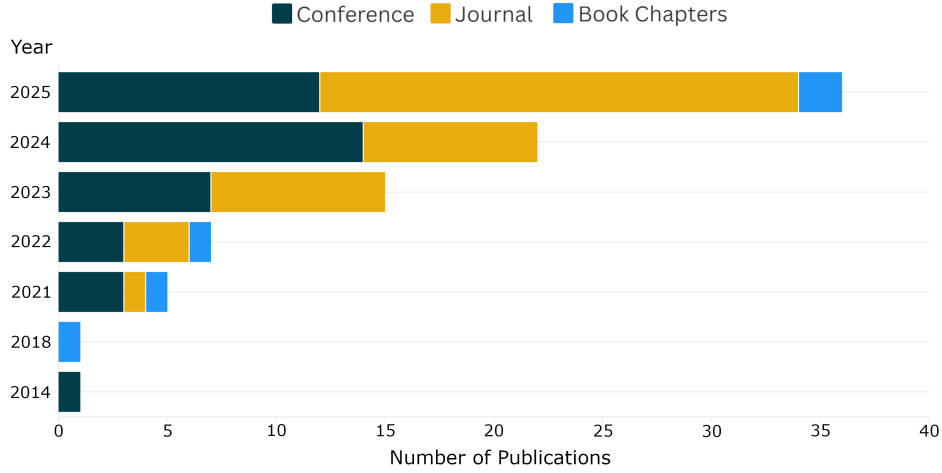


Fig. 3: Yearly distribution of publications by type.

4.8.2 Geographical Distribution of Research

Figure 4 illustrates the top fifteen countries contributing to AI-related research across the mining, oil and gas, and manufacturing industries, based on total publication counts aggregated across industries.

Manufacturing research shows the broadest geographical spread and the highest publication volume. The United States and Germany emerge as the leading contributors, each with eleven manufacturing-focused studies, followed by Italy (nine), Greece (eight), and Spain (seven). Additional contributions from European countries such as the United Kingdom (four), Cyprus (four), Portugal (three), Denmark (two) and Estonia (two) reflect strong alignment with Europe’s Industry 4.0, human-centric automation, and smart manufacturing initiatives. Beyond Europe, manufacturing research also includes notable contributions from Canada (four), the United Arab Emirates (three), India (three), Pakistan (two), and Jordan (two), indicating increasing global engagement.

Mining research remains geographically concentrated, with contributions reported from a relatively small number of countries. China leads mining-related publications with three studies, followed by Australia, Sweden, and India with two studies each, and Canada, Germany, and the United States contributing one paper each.

Oil and gas research shows moderate geographical diversity, with the United States contributing three studies, China and Canada contributing three and two studies respectively, and Germany contributing one. Additional contributions from Oman and

South Africa indicate selective engagement from regions with established oil and gas operations.

Across all three industries, the United States and China are the most consistently represented countries, contributing to mining, oil and gas, and manufacturing research. Germany also plays a significant cross-industry role, particularly in manufacturing and mining. Overall, the geographical distribution highlights a stronger concentration of research activity in North America and Europe, with a growing but still limited participation from regions in the Middle East, Asia, and Africa.

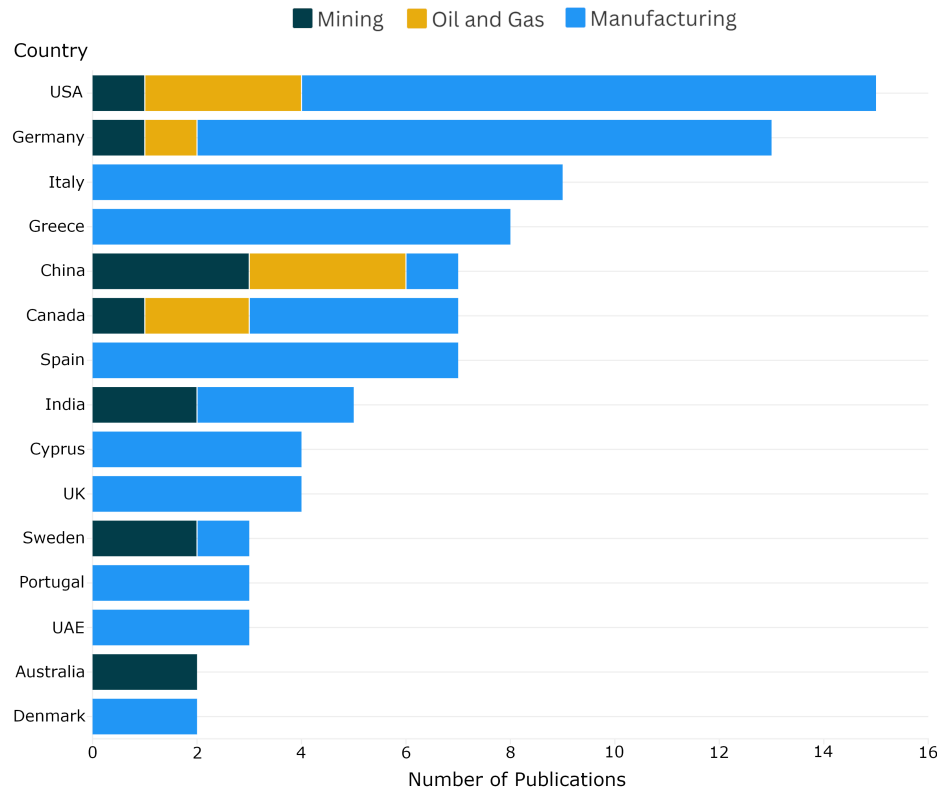


Fig. 4: Top fifteen contributing countries, categorized by industry.

4.8.3 Institutional Affiliation and Research Collaboration

Figure 5 summarizes the institutional affiliations of the authors, categorized as academic, industry, or academic-industry collaboration.

Across all three industries, academia remains the dominant contributor, accounting for 56 papers. Industry-authored works represent 14 papers, while collaborative efforts between academia and industry account for 17 papers.

Industry participation is most evident in the oil and gas industry, with four industry-only papers, reflecting the industry’s established engineering and applied-research practices. Manufacturing shows the highest number of collaborative publications, with twelve academic-industry papers, indicating active partnerships in areas such as robotics, human-machine systems, and explainable AI. Mining remains the smallest research domain but demonstrates notable collaboration, with three academic-industry papers alongside predominantly academic contributions.

These patterns suggest that while research is still largely academically driven, industrial engagement is increasing, particularly in manufacturing and oil and gas.

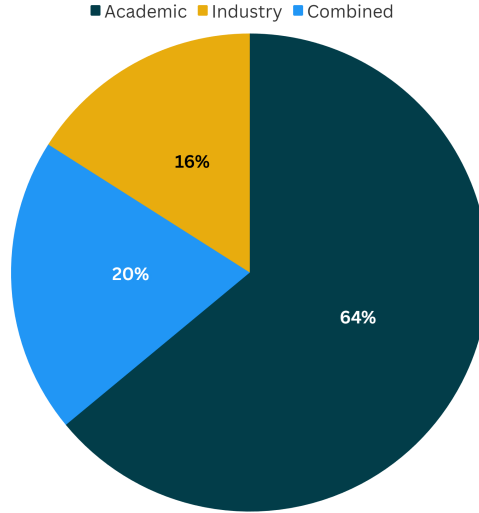


Fig. 5: Distribution of publications by institutional affiliation.

4.8.4 Cross-Industry Citation Analysis

To examine knowledge transfer and cross-industry influence, a citation analysis was conducted to assess how studies in mining, oil and gas, and manufacturing reference prior work across production industries and general AI research. Citations were categorized into four groups: mining, oil and gas, manufacturing, and general AI or other related domains.

The results show strong within-industry citation patterns across all three industries, indicating that most research builds primarily on domain-specific literature. Mining-focused studies cite mining-related work most frequently, with 113 citations, compared to 15 citations from oil and gas literature and 26 from manufacturing. A substantial proportion of mining citations (353) reference general AI and related methodological research, highlighting the industry’s dependence on foundational AI techniques rather than cross-industry operational insights.

Oil and gas papers show a similar pattern of domain concentration, citing oil and gas literature extensively (253 citations), while drawing comparatively limited references from mining (22) and manufacturing (11). At the same time, oil and gas research exhibits a strong connection to general AI literature, with 180 citations, reflecting the industry’s emphasis on established AI methods.

Manufacturing research demonstrates the largest volume of citations in general. Manufacturing papers overwhelmingly cite manufacturing-related literature (1134 citations), alongside a large number of references to general AI and related domains (1443 citations). Cross-industry citations from manufacturing to mining (6) and oil and gas (15) remain limited, indicating that knowledge transfer from other production industries into manufacturing research is relatively low despite methodological similarities.

Across industries, cross-industry citation patterns indicate that mining and oil and gas research draw to some extent from manufacturing-oriented literature, whereas manufacturing research remains largely self-referential and focused on general AI topics. This points to manufacturing as a more common source of methodological reference, while reciprocal cross-industry citation remains limited.

4.9 Conclusion

The comparative analysis reveals distinct levels of maturity in the adoption of explainable and trustworthy AI across mining, oil and gas, and manufacturing. Manufacturing shows the most advanced progress, characterized by 94% trust coverage and 45% engagement with HMI. The industry benefits from well-developed human-in-the-loop collaboration models and increasingly operational explainability tools, though applied XAI remains moderate at 39% and governance references at 19%.

Oil and gas shows the strongest uptake of XAI, with 92% of papers discussing explainability and 50% implementing applied methods. The industry also exhibits structured assurance and safety compliance systems, yet explicit AI governance mechanisms and formal ethical audits remain limited and inconsistently addressed.

Mining remains at an earlier stage of integration, with 62% of studies addressing trust and 54% employing XAI, but only 15% referencing governance frameworks. Current research focuses heavily on performance reliability, safety, and environmental risk prediction, leaving opportunities to expand into explanation validation, governance integration, and human–AI collaboration.

Together, the three industries illustrate complementary strengths and gaps. Manufacturing leads in operational trust and human-centered design, oil and gas excels in assurance and risk-based explainability, and mining advances safety-focused applications that could benefit from stronger governance structures and user-centered validation. Achieving truly trustworthy and ethical AI across production industries will require convergence between explainability, governance, and human oversight, supported by shared standards, cross-industry learning, and collaborative transparency frameworks.

Funding. This study is supported by the Hebron Project. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the Hebron Project.

Generative AI Disclosure. The authors used ChatGPT (OpenAI) to assist with improving the language and grammar of the manuscript. All technical content and conclusions are solely those of the authors.

Conflict of interest. The authors declare no conflict of interest.

Data Availability. No new datasets were generated during this study.

References

- [1] Corrigan, C.C., Ikonnikova, S.A.: A review of the use of AI in the mining industry: Insights and ethical considerations for multi-objective optimization. *The Extractive Industries and Society* **17**, 101440 (2024)
- [2] Ge, S., Xie, Y., Liu, K., Ding, Z., Hu, E., Chen, L., Wang, F.-Y.: The use of intelligent vehicles and artificial intelligence in mining operations: Ethics, responsibility, and sustainability. *IEEE Transactions on Intelligent Vehicles* **8**(2), 1021–1024 (2023)
- [3] Mkono, C.N., Shen, C., Mulashani, A.K., Carranza, E.J.M., Kalibwami, D.C., Nyangi, M.J.: A novel hybrid group method of data handling and levenberg marquardt model for estimating total organic carbon in source rocks with explainable artificial intelligence. *Engineering Applications of Artificial Intelligence* **144**, 110137 (2025) <https://doi.org/10.1016/j.engappai.2025.110137>
- [4] Mehdiyev, N., Majlatow, M., Fettke, P.: Interpretable and explainable machine learning methods for predictive process monitoring: a systematic literature review. *Artificial Intelligence Review* **58**(12), 378 (2025)
- [5] Gunraj, H., Guerrier, P., Fernandez, S., Wong, A.: SolderNet: Towards trustworthy visual inspection of solder joints in electronics manufacturing using explainable artificial intelligence. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, pp. 15668–15674 (2023)
- [6] Gurina, E., Klyuchnikov, N., Antipova, K., Koroteev, D.: Making the black-box brighter: Interpreting machine learning algorithm for forecasting drilling accidents. *Journal of Petroleum Science and Engineering* **218**, 111041 (2022) <https://doi.org/10.1016/j.petrol.2022.111041>
- [7] Dikopoulou, Z., Lavasa, E., Perez-Castanos, S., Monzo, D., Moustakidis, S.: Towards explainable AI validation in industry 4.0: A fuzzy cognitive map-based evaluation framework for assessing business value. In: *2023 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, pp. 1–9 (2023). IEEE

- [8] Cabrero-Daniel, B., Fazelidehkordi, Y., Ratuszniak, O.: Trustworthy” black-box self-adaptive systems. In: i 28th International Conference on Requirements Engineering: Foundation for Software Quality, [Https://ceur-ws. Org](https://ceur-ws.Org), vol. 3378 (2023)
- [9] Xu, J., Zheng, L., Lan, H., Zuo, Y., Li, B., Tian, S., Tian, Y.: Research on an identification model for mine water inrush sources based on the HBA-CatBoost algorithm. *Scientific Reports* **14**(1) (2024) <https://doi.org/10.1038/s41598-024-74417-1>
- [10] Żywiolek, J.: Building trust in ai-human partnerships: Exploring preferences and influences in the manufacturing industry. *Management Systems in Production Engineering* **32**(2), 244–251 (2024) <https://doi.org/10.2478/mspe-2024-0024>
- [11] Robatto Simard, S., Gamache, M., Doyon-Poulin, P.: Development and usability evaluation of VulcanH, a CMMS prototype for preventive and predictive maintenance of mobile mining equipment. *Mining* **4**(2), 326–351 (2024)
- [12] Kaul, A., Soofastaei, A.: Advanced Analytics for Ethical Considerations in Mining Industry, pp. 55–80. Springer, ??? (2022). https://doi.org/10.1007/978-3-030-91589-6_3
- [13] Yadav, R.K., C S, S.: Exploring the frontier: Space mining, legal implications, and the role of artificial intelligence. In: 2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies, pp. 1–5 (2024). <https://doi.org/10.1109/TQCEBT59414.2024.10545181>
- [14] Martinsen, M., Zhou, Y., Dahlquist, E., Yan, J., Kyprianidis, K.: Positive climate effects when AR customer support simultaneous trains AI experts for the smart industries of the future. *Applied Energy* **339**, 120988 (2023) <https://doi.org/10.1016/j.apenergy.2023.120988>
- [15] Teng, S., Li, X., Li, Y., Li, L., Xuanyuan, Z., Ai, Y., Chen, L.: Scenario engineering for autonomous transportation: A new stage in open-pit mines. *IEEE Transactions on Intelligent Vehicles* **9**(3), 4394–4404 (2024) <https://doi.org/10.1109/tiv.2024.3373495>
- [16] Ibrahim, B., Ahenkorah, I., Ewusi, A.: Explainable risk assessment of rockbolts’ failure in underground coal mines based on categorical gradient boosting and shapley additive explanations (SHAP). *Sustainability* **14**(19), 11843 (2022)
- [17] Moradi, R., Ruiz-Tagle Palazuelos, A., Lopez Droguett, E., Groth, K.M.: Toward a framework for risk monitoring of complex engineering systems with online operational data: A deep learning-based solution. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* **237**(5), 910–921 (2022) <https://doi.org/10.1177/1748006x221079964>

- [18] Mkono, C.N., Shen, C., Mulashani, A.K., Mwakipunda, G.C., Nyakilla, E.E., Kasala, E.E., Mwizarubi, F.: A novel hybrid machine learning and explainable artificial intelligence approaches for improved source rock prediction and hydrocarbon potential in the Mandawa Basin, SE Tanzania. *International Journal of Coal Geology* **302**, 104699 (2025) <https://doi.org/10.1016/j.coal.2025.104699>
- [19] Barraza, J.F., Droguett, E.L., Martins, M.R.: SCF-Net: A sparse counterfactual generation network for interpretable fault diagnosis. *Reliability Engineering & System Safety* **250**, 110285 (2024) <https://doi.org/10.1016/j.res.2024.110285>
- [20] Aditama, P., Eita, M., Koziol, T., Zhukova, K., Dillen, M., Sanchez, F.P.A., Schemmer, M., Nitsche, M., Tomina, M., Kimari, T.: Deployment of an artificial intelligence-based well integrity monitoring solution. In: *SPE Europe Energy Conference and Exhibition. 24EURO. SPE*, ??? (2024). <https://doi.org/10.2118/220003-ms>
- [21] Ershaghi, I., Ershaghi, M.A., Al-Ruwai, F.: Human factors in domain adaptation within the oil and gas industry. In: *SPE Middle East Oil and Gas Show and Conference*, pp. 031–020001 (2021). SPE
- [22] Agbaji, A.L.: An empirical analysis of artificial intelligence, big data and analytics applications in exploration and production operations. In: *International Petroleum Technology Conference*, pp. 101–043001 (2021). IPTC
- [23] Abd Rahman, M.S., Jamaludin, N.A.A., Zainol, Z., Sembok, T.M.T.: The application of decision tree classification algorithm on decision-making for upstream business. *International Journal of Advanced Computer Science and Applications* **14**(8) (2023)
- [24] Bogen, M., Ewert, C., Landenberg, A., Rilling, S., Wulff, B.: Using Interactive Visualization and Machine Learning for Seismic Interpretation, pp. 115–177. Springer, ??? (2021). https://doi.org/10.1007/978-3-030-90716-7_4
- [25] Cann, G.: The Business of AI Adoption, pp. 153–179. Elsevier, ??? (2021). <https://doi.org/10.1016/b978-0-12-820714-7.00008-x>
- [26] Goldman, C.V., Baltaxe, M., Chakraborty, D., Arinez, J., Diaz, C.E.: Interpreting learning models in manufacturing processes: Towards explainable AI methods to improve trust in classifier predictions. *Journal of Industrial Information Integration* **33**, 100439 (2023) <https://doi.org/10.1016/j.jii.2023.100439>
- [27] Sadrfaridpour, B., Saeidi, H., Burke, J., Madathil, K., Wang, Y.: Modeling and control of trust in human-robot collaborative manufacturing. In: *Robust Intelligence and Trust in Autonomous Systems*, pp. 115–141. Springer, ??? (2016)
- [28] Chu, C.-H., Liu, Y.-L.: Augmented reality user interface design and experimental evaluation for human-robot collaborative assembly. *Journal of Manufacturing*

- [29] Karbouj, B., Al Rashwany, K., Alshamaa, O., Krüger, J.: Adaptive behavior of collaborative robots: Review and investigation of human predictive ability. *Procedia CIRP* **130**, 952–958 (2024)
- [30] Campagna, G., Rehm, M.: Analysis of proximity and risk for trust evaluation in human-robot collaboration. In: 2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), pp. 2191–2196 (2023). IEEE
- [31] Campagna, G., Dadgostar, M., Chrysostomou, D., Rehm, M.: A data-driven approach utilizing body motion data for trust evaluation in industrial human-robot collaboration. In: 2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN), pp. 1984–1990 (2024). IEEE
- [32] Guidolin, M., Berti, N., Gnesotto, P., Battini, D., Reggiani, M.: Trust the robot! enabling flexible collaboration with humans via multi-sensor data integration. In: 2024 IEEE 29th International Conference on Emerging Technologies and Factory Automation (ETFA), pp. 1–8 (2024). IEEE
- [33] Żywiołek, J.: Trust-building in ai-human partnerships within industry 5.0. *System Safety: Human - Technical Facility - Environment* **6**(1), 89–98 (2024) <https://doi.org/10.2478/czoto-2024-0011>
- [34] Plathottam, S.J., Rzonca, A., Lakhnori, R., Iloeje, C.O.: A review of artificial intelligence applications in manufacturing operations. *Journal of Advanced Manufacturing and Processing* **5**(3) (2023) <https://doi.org/10.1002/amp2.10159>
- [35] Onu, P., Pradhan, A., Madonsela, N.S.: AI in smart manufacturing: An ethical and legal perspective. In: 2024 4th International Multidisciplinary Information Technology and Engineering Conference (IMITEC), pp. 290–295 (2024). IEEE
- [36] Miltiadou, D., Perakis, K., Sesana, M., Calabresi, M., Lampathaki, F., Biliri, E.: A novel explainable artificial intelligence and secure artificial intelligence asset sharing platform for the manufacturing industry. In: 2023 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), pp. 1–8 (2023). IEEE
- [37] Lampathaki, F., Agostinho, C., Glikman, Y., Sesana, M.: Moving from ‘black box’ to ‘glass box’ artificial intelligence in manufacturing with XMANAI. In: 2021 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), pp. 1–6 (2021). IEEE
- [38] Agostinho, C., Dikopoulou, Z., Lavasa, E., Perakis, K., Pitsios, S., Branco, R., Reji, S., Hetterich, J., Biliri, E., Lampathaki, F., Rodríguez Del Rey, S., Gkolemis,

- V.: Explainability as the key ingredient for AI adoption in industry 5.0 settings. *Frontiers in Artificial Intelligence* **6** (2023) <https://doi.org/10.3389/frai.2023.1264372>
- [39] Biliri, E., Lampathaki, F., Mandilaras, G., Prieto-Roig, A., Calabresi, M., Branco, R., Gkolemis, V.: Explainable artificial intelligence bundles for algorithm lifecycle management in the manufacturing domain. In: 2023 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), pp. 1–7 (2023). IEEE
 - [40] Moosavi, S., Farajzadeh-Zanjani, M., Razavi-Far, R., Palade, V., Saif, M.: Explainable AI in manufacturing and industrial cyber–physical systems: A survey. *Electronics* **13**(17), 3497 (2024) <https://doi.org/10.3390/electronics13173497>
 - [41] Nikiforidis, K., Kyrtoglou, A., Vafeiadis, T., Kotsiopoulos, T., Nizamis, A., Ioannidis, D., Votis, K., Tzovaras, D., Sarigiannidis, P.: Enhancing transparency and trust in ai-powered manufacturing: A survey of explainable AI (XAI) applications in smart manufacturing in the era of industry 4.0/5.0. *ICT Express* **11**(1), 135–148 (2025) <https://doi.org/10.1016/j.icte.2024.12.001>
 - [42] Hoffmann, R., Reich, C.: A systematic literature review on artificial intelligence and explainable artificial intelligence for visual quality assurance in manufacturing. *Electronics* **12**(22), 4572 (2023) <https://doi.org/10.3390/electronics12224572>
 - [43] Zeiser, T., Ehret, D., Lutz, T., Saar, J.: Explainable AI in manufacturing. In: 2024 IEEE International Conference on Engineering, Technology, and Innovation (ICE/ITMC), pp. 1–8. IEEE, ??? (2024). <https://doi.org/10.1109/ice/itmcc61926.2024.10794363>
 - [44] Aromaa, S., Heikkilä, P., Jurvansuu, M., Pehlivan, S., Väärä, T., Jurmu, M.: Company perspectives of generative artificial intelligence in industrial work. *Procedia Computer Science* **253**, 217–226 (2025) <https://doi.org/10.1016/j.procs.2025.01.085>
 - [45] Zelba, F., Balzereit, K., Windmann, S.: Concepts and measures towards trustworthy AI in industrial manufacturing. In: 2024 IEEE 29th International Conference on Emerging Technologies and Factory Automation (ETFA), pp. 1–4 (2024). IEEE
 - [46] Milcke, B., Dinglinger, P., Holtmann, J.: Exploring the Role of Explainable AI in the Development and Qualification of Aircraft Quality Assurance Processes: A Case Study, pp. 331–352. Springer, ??? (2024). https://doi.org/10.1007/978-3-031-63803-9_18
 - [47] Fragueiro, F.G., Martín, D.G., López, A.B., Rial, A.A., Tranchero, J.O., Lorenzo, B.C., Montenegro, J.M.F., Muiños-Landín, S.: An Integrated Active Learning Framework for the Deployment of Machine Learning Models for Defect Detection

- in Manufacturing Environments, pp. 3–14. Springer, ??? (2024). https://doi.org/10.1007/978-3-031-57496-2_1
- [48] Ostheimer, J., Chowdhury, S., Iqbal, S.: An alliance of humans and machines for machine learning: Hybrid intelligent systems and their design principles. *Technology in Society* **66**, 101647 (2021) <https://doi.org/10.1016/j.techsoc.2021.101647>
 - [49] Gkrizis, C., Dimitropoulos, N., Katsampiris-Salgado, K., Karagiannis, P., Michalos, G., Makris, S.: Intersubjective ai-driven multimodal interaction for advanced user-centric HRC applications - the jarvis approach. *Procedia CIRP* **130**, 325–330 (2024) <https://doi.org/10.1016/j.procir.2024.10.095>
 - [50] Wajid, U., Nizamis, A., Anaya, V.: Towards industry 5.0—a trustworthy AI framework for digital manufacturing with humans in control. *Proceedings http://ceur-ws.org*. ISSN **1613**, 0073 (2022)
 - [51] Gawade, V., Chen, M.: Explaining multimodal CNN-DNN model predictions for quality monitoring of porosity in laser metal deposition. *Knowledge-Based Systems* **311**, 113095 (2025) <https://doi.org/10.1016/j.knosys.2025.113095>
 - [52] Büscher, J., Zajackowski, J., Rademacher, H.-G., Tillmann, W., Deuse, J.: Trustworthiness of artificial intelligence applications for quality optimisation in metal additive manufacturing. *Procedia CIRP* **122**, 497–502 (2024) <https://doi.org/10.1016/j.procir.2024.01.072>
 - [53] Sermmany, K., Wanjantuk, P., Leelapatra, W.: Utilizing explainable artificial intelligence (XAI) to identify determinants of coffee quality. In: 2024 21st International Joint Conference on Computer Science and Software Engineering (JCSSE), pp. 696–703. IEEE, ??? (2024). <https://doi.org/10.1109/jcsse61278.2024.10613641>
 - [54] Haindl, P., Hoch, T., Dominguez, J., Aperribai, J., Ure, N.K., Tunçel, M.: Quality characteristics of a software platform for human-ai teaming in smart manufacturing. In: *International Conference on the Quality of Information and Communications Technology*, pp. 3–17 (2022). Springer
 - [55] Scarton, G., Trono, F., Trevisan, C., Formentini, M.: An Expert System for Automated Quality Control: A Case Study in a Mechatronic Manufacturing Company, pp. 630–641. Springer, ??? (2023). https://doi.org/10.1007/978-3-031-38274-1_53
 - [56] Ghezeljehmeidan, A.G., Driel, W.D., Dauwels, J.: Unveiling hidden anomalies: A hybrid approach for surface mounted electronics. In: *2024 IEEE 22nd International Conference on Industrial Informatics (INDIN)*, pp. 1–6. IEEE, ??? (2024). <https://doi.org/10.1109/indin58382.2024.10774393>
 - [57] Fernández-Miguel, A., Ortíz-Marcos, S., Jiménez-Calzado, M., Hoyo, A.P.,

- García-Muiña, F.E., Settembre-Blundo, D.: Agentic AI in smart manufacturing: Enabling human-centric predictive maintenance ecosystems. *Applied Sciences* **15**(21), 11414 (2025)
- [58] Azhari, F., Sennersten, C.C., Lindley, C.A., Sellers, E.: Deep learning implementations in mining applications: a compact critical review. *Artificial Intelligence Review* **56**(12), 14367–14402 (2023)
- [59] Fan, X., Meng, F., Deng, J., Semnani, A., Zhao, P., Zhang, Q.: Transformative reconstruction of missing acoustic well logs using multi-head self-attention birnns. *Geoenery Science and Engineering* **245**, 213513 (2025)
- [60] Chen, Y.-L., Sacchi, S., Dey, B., Blanco, V., Halder, S., Leray, P., Gendt, S.D.: Exploring machine learning for semiconductor process optimization: A systematic review. *IEEE Transactions on Artificial Intelligence* **5**(12), 5969–5989 (2024) <https://doi.org/10.1109/TAI.2024.3429479>
- [61] Catti, P., Nikolakis, N., Alexopoulos, K.: A human-centric drift controller framework for adaptive and explainable quality control in manufacturing. In: 2025 IEEE 30th International Conference on Emerging Technologies and Factory Automation (ETFA), pp. 1–8 (2025). IEEE
- [62] Alshkeili, H.M.H.A., Almheiri, S.J., Khan, M.A.: Privacy-preserving interpretability: An explainable federated learning model for predictive maintenance in sustainable manufacturing and industry 4.0. *AI* **6**(6), 117 (2025)
- [63] Moawad, A., Islam, E., Kim, N., Vijayagopal, R., Rousseau, A., Wu, W.B.: Explainable AI for a no-teardown vehicle component cost estimation: A top-down approach. *IEEE Transactions on Artificial Intelligence* **2**(2), 185–199 (2021) <https://doi.org/10.1109/TAI.2021.3065011>
- [64] Aslam, M.M., Tufail, A., Gul, H., Irshad, M.N., Namoun, A.: Artificial intelligence for secure and sustainable industrial control systems-a survey of challenges and solutions. *Artificial Intelligence Review* **58**(11), 349 (2025)
- [65] Kitchenham, B., Charters, S., et al.: Guidelines for performing systematic literature reviews in software engineering. *Proceedings* <http://ceur-ws.org>. ISSN (2007)
- [66] Page, M.J., Moher, D., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., et al.: PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *bmj* **372** (2021)
- [67] Van De Schoot, R., De Bruin, J., Schram, R., Zahedi, P., De Boer, J., Weijdem, F., Kramer, B., Huijts, M., Hoogerwerf, M., Ferdinands, G., *et al.*: An open source machine learning framework for efficient and transparent systematic

- reviews. *Nature machine intelligence* **3**(2), 125–133 (2021)
- [68] DistillerSR Inc.: DistillerSR. Software platform. Version 2.35. Accessed August 2025–January 2026 (2023). <https://www.distillersr.com/>
 - [69] Matrenin, P.V., Stepanova, A.I.: Enhancing the interpretability of electricity consumption forecasting models for mining enterprises using SHapley additive explanations. *Journal of Mining Institute* (271 (eng)), 154–167 (2025)
 - [70] Dharmadhikari, D.D., Ray, A., Shinde, B.A., Raut, S.V., Taware, R.D., Desai, S., Kurhade, A.S.: Machine learning applications in ore grade estimation and blending optimization for modern mining. *Applied Chemical Engineering* **8**(4), 5790 (2025) <https://doi.org/10.59429/ace.v8i4.5790>
 - [71] Khan, A.I., Mishrif, A.: Governance of artificial intelligence in the oil and gas industry. In: *Shaping the Future of Development in the GCC: Artificial Intelligence Policies, Challenges and Opportunities*, pp. 19–39. Springer, ??? (2025)
 - [72] Puthanveetil Madathil, A., Luo, X., Liu, Q., Walker, C., Madarkar, R., Qin, Y.: A review of explainable artificial intelligence in smart manufacturing. *International Journal of Production Research*, 1–44 (2025)
 - [73] Jawad, A.Y., Shehab, R.M., Zhumabaeva, A., Al-Sarraf, N.M., Alwan, G.M.: Ai ethics and legal regulation in textile manufacturing. *Izvestiya Vysshikh Uchebnykh Zavedenii. Tekhnologiya Tekstil'noi Promyshlennosti* **416**(2), 243–252 (2025) <https://doi.org/10.47367/0021-3497.2025.2.243>
 - [74] Parziale, M., Yeung, Y.F., Youcef-Toumi, K., Giglio, M., Cadini, F.: Anomaly characterization for the condition monitoring of rotating shafts exploiting data fusion and explainable convolutional neural networks. *Structural Health Monitoring*, 14759217241301288 (2025)
 - [75] Berteit, O., Binder, M., Henkel, S., Erfurth, C., Bliedtner, J.: Explainable artificial intelligence for predictive quality monitoring in optical manufacturing. In: *EPJ Web of Conferences*, vol. 335, p. 03015 (2025). EDP Sciences
 - [76] Alkhatib, F., Allam, M., Swarnakar, V., Alsadi, J., Maalouf, M.: Implementing and evaluating the quality 4.0 PMQ framework for process monitoring in automotive manufacturing. *Scientific Reports* **15**(1), 24742 (2025)
 - [77] Guillen, A., Teodoro, E.: Gconsumers and audiovisual platforms: An assessment of international jurisdiction (2022)
 - [78] Ahmed, Y.S., Milani, A.S.: A review of strategies, challenges, and ethical implications of machine learning in smart manufacturing. *Decision Analytics Journal*, 100591 (2025)

- [79] Farhan, M., Naeem, M.R., Almadhor, A., Bashir, A.K., Zhu, Z., Gadekallu, T.R.: Explainable ai-driven security framework for cyber-physical production systems in industry 4.0: Leveraging immersive embedded CIIoT. *IEEE Transactions on Consumer Electronics* **71**(4), 10446–10453 (2025) <https://doi.org/10.1109/TCE.2025.3602792>
- [80] Marchesano, M.G., Mattera, G., Guizzi, G., Santillo, L.C., Converso, G.: Explainable deep reinforcement learning enhancing industrial maintenance. *IFAC-PapersOnLine* **59**(10), 524–529 (2025)
- [81] Soesanto, R.P., Achmad, F.: Leveraging explainable AI for accurate production forecasting in the bag manufacturing industry. In: *Proceedings of the International Manufacturing Engineering Conference & The Asia Pacific Conference on Manufacturing Systems*, pp. 105–114 (2024). Springer
- [82] Gan, L., Qu, G., Yue, X.: Securing smart manufacturing with explainable clustering for intrusion detection. In: *International Manufacturing Science and Engineering Conference*, vol. 89022, pp. 002–20007 (2025). American Society of Mechanical Engineers
- [83] Sardari, S., Fernandes, F., Araya-Martinez, J.M., Zak, J.A., Roitberg, A.: Towards human-understandable visual recognition for nonexperts in industrial inspection: A case study for car manufacturing lines. In: *2025 IEEE 21st International Conference on Automation Science and Engineering (CASE)*, pp. 279–285 (2025). IEEE
- [84] Al Rawajbeh, M., Maria Soosai, A.J., Ramasamy, L.K., Khan, F.: Trustworthy adaptive AI for real-time intrusion detection in industrial IoT security. *IoT* **6**(3), 53 (2025)
- [85] Campos, J., Shakhovska, N.: Advancing XAI development: An agile framework for human-centered and explainable ai. In: *International Conference on Human-Computer Interaction*, pp. 22–40 (2025). Springer
- [86] Ciobotaru, A., Corches, C., Gota, D., Miclea, L.: An explainable deep learning-based predictive maintenance solution for air compressor condition monitoring. *Sensors* **25**(18), 5797 (2025)
- [87] Ahangar, M.N., Farhat, Z., Sivanathan, A.: AI trustworthiness in manufacturing: challenges, toolkits, and the path to industry 5.0. *Sensors* **25**(14), 4357 (2025)
- [88] Bitterling, C., Loske, D., Klumpp, M.: Improving visual inspection accuracy: Explainable AI for human-centric quality control. *IFAC-PapersOnLine* **59**(10), 1450–1455 (2025)
- [89] Velasco-Loera, F., Alcaraz-Mejia, M., Chavez-Hurtado, J.L.: An interpretable hybrid fault prediction framework using XGBoost and a probabilistic graphical

- model for predictive maintenance: A case study in textile manufacturing. *Applied Sciences* (2076-3417) **15**(18) (2025)
- [90] Vishwakarma, V., Galhotra, S.: Responsible defect detection with probabilistic contrastive counterfactuals. In: *Metrology, Inspection, and Process Control XXXIX*, vol. 13426, p. 134264 (2025). SPIE
 - [91] Amaliah, N.R., Tjahjono, B., Palade, V.: Human-in-the-loop XAI for predictive maintenance: A systematic review of interactive systems and their effectiveness in maintenance decision-making. *Electronics* **14**(17), 3384 (2025)
 - [92] Marimuktu, P., Subramanian, R.R.: Federated learning-based human-robot collaboration with explainable ai: A framework for secure and transparent interaction. In: *2025 International Conference on Computational Robotics, Testing and Engineering Evaluation (ICCRTEE)*, pp. 1–5 (2025). IEEE
 - [93] Mozaffar, F., Smith, L., Morkos, B.: Tunes of trust: A framework for auditory nudges in human-ai manufacturing collaboration. *Manufacturing Letters* **44**, 195–204 (2025)
 - [94] Banitalebi, B., Dwivedula, S.V.A.: Establishing trust in ai-driven data observability and quality control: A framework for reliable and scalable standards. In: *2025 IEEE Conference on Artificial Intelligence (CAI)*, pp. 1388–1393 (2025). IEEE
 - [95] Bagirov, M.S.: Industrial economy and artificial intelligence: real situation and development directions. *Universidad y Sociedad* **17**(5), 5402–5402 (2025)
 - [96] Besinger, P., Zaremba, J., Ansari, F.: Responsible AI in manufacturing: The case of accountability in AI systems. *IFAC-PapersOnLine* **59**(10), 1325–1330 (2025)