Explainability Requirements for Industry 5.0: Towards Role-Adaptive Explanation in Semiconductor Manufacturing

Amina Mević[1,2,*], Senka Krivić[1]

[1]Faculty of Electrical Engineering, University of Sarajevo, Sarajevo, Bosnia and Herzegovina
[2]Infineon Technologies AG, Villach, Austria

Abstract

The integration of artificial intelligence (AI) into semiconductor manufacturing has accelerated the development of systems for real-time quality prediction, yet adoption remains limited by challenges in interpretability, explainability, and trust. While explainable AI (xAI) research has produced a rich body of methods, the explanation needs of domain practitioners in industrial production environments remain poorly understood. This paper presents an exploratory qualitative investigation of practitioner-level explainability requirements, grounded in semi-structured expert interviews conducted at a semiconductor manufacturer Infineon Technologies. The study identifies a consistent mismatch between what current xAI methods provide and what practitioners actually require, and reveals that practitioners constitute distinct explanation user profiles depending on their operational role. These findings offer practitioner-grounded design implications for role-adaptive explainable AI systems, with relevance for explainable user modeling and personalized systems research, and represent a first step toward role-adaptive explanation interfaces for industrial domain experts.

Keywords

Explainable AI, Explainability Requirements, Industrial AI, Semiconductor Manufacturing, Role-Adaptive Explanations, Industry 5.0, Expert Interviews, Trust

1. Introduction

Artificial intelligence (AI) is rapidly moving from research prototype to operational infrastructure across high-stakes industrial sectors. In semiconductor manufacturing, one of the most capital-intensive and precision-demanding industries in the world, AI systems are being deployed to monitor processes, predict quality outcomes, and reduce reliance on expensive physical measurements [1]. A central paradigm in this context is *Virtual Metrology* (VM). VM systems use supervised or unsupervised machine learning techniques to predict product quality or physical properties directly from process sensor data (Figure 1), without the need for physical inspection [2, 3]. Unlike conventional Statistical Process Control (SPC), which monitors process reliability using physically measured values of product properties such as thin film thickness, VM systems use ML models to predict these values directly from process control data, eliminating the need for physical measurement while maintaining quality oversight. By enabling faster feedback on process outcomes and product properties and reducing measurement costs, VM has demonstrated significant potential to improve efficiency and control in semiconductor fabrication processes. Yet as ML models grow in complexity, from linear regression to deep neural networks, they increasingly function as black boxes, producing accurate predictions but offering little insight into the reasoning behind them.

This opacity is not merely a technical inconvenience. In regulated, high-stakes environments, human operators must be able to understand, validate, and take responsibility for AI-driven decisions before acting on them. Regulatory frameworks including the EU General Data Protection Regulation (GDPR) [4] and the EU AI Act [5] reinforce this requirement by establishing the right to explanation as a legal norm

Joint Proceedings of the ACM UMAP Workshops 2026, UMAP 2026, June 8–11, 2026, Gothenburg, Sweden

*Corresponding author.
amevic@etf.unsa.ba (A. Mević); senka.krivic@etf.unsa.ba (S. Krivić)
0000-0002-7442-0070 (A. Mević); 0000-0001-8045-427X (S. Krivić)

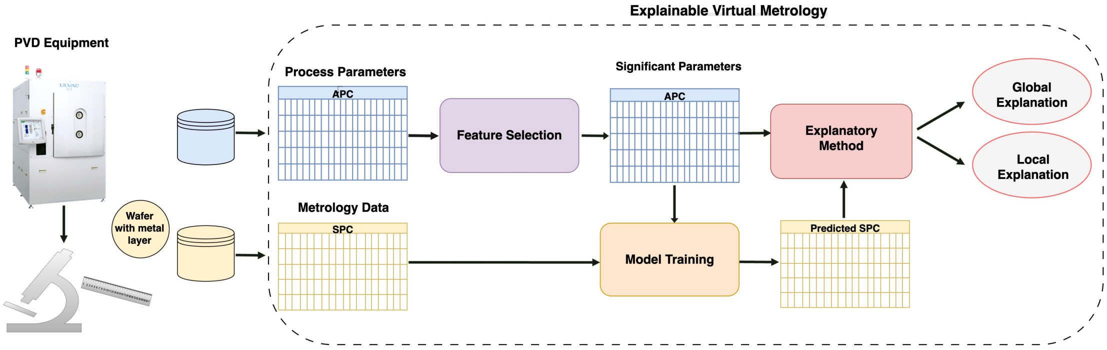

Figure 1: Explainable Virtual Metrology (xVM) pipeline. Process parameters and metrology data from production equipment (PVD) are passed through feature selection to identify significant parameters, which together with metrology data feed into a predictive model that estimates product quality (Predicted SPC). xVM extends this pipeline with an explanatory method that produces global and local explanations, making AI-driven quality decisions interpretable and auditable for domain experts.

for consequential automated decisions. The field of Explainable Artificial Intelligence (xAI) addresses this challenge by developing methods and tools that make model outputs interpretable and actionable for human users [6, 7]. xAI methods range from inherently transparent models, such as decision trees and linear regression, to post-hoc techniques applied to opaque models. Among the most widely used of which are SHAP [8], which quantifies the contribution of each input feature to a prediction. These developments align with the emerging Industry 5.0 paradigm, which places human-centricity and human-AI collaboration at the center of industrial automation [9].

Current xAI techniques produce explanations that often fail to meet the actual needs of domain practitioners [10], yet empirical evidence from real industrial deployments remains scarce. In semiconductor manufacturing specifically, most research optimizes predictive performance without equivalent attention to the practitioner-level requirements that determine whether an AI system will be trusted and adopted in production.

Explanation is fundamentally a human-centered problem. Its success depends not on the sophistication of the method, but on whether it meets the cognitive and operational needs of its intended audience [11]. This paper shifts the focus to a domain that has received comparatively little attention in the user modeling community: industrial AI systems deployed in high-stakes manufacturing environments. While much of the xAI literature examines explanation needs in consumer-facing systems such as recommender systems [12] and clinical decision support [13], the practitioner perspective in industrial production remains largely unexplored. We therefore ask: *what do practitioners actually need from AI explanations before they will trust and use these systems?* To answer this, we conducted semi-structured expert interviews with four professionals at a semiconductor manufacturer, Infineon Technologies, covering current AI usage, adoption barriers, trust requirements, accountability expectations, and specific explainability needs. Specifically, our participants represent two distinct practitioner roles: Unit Process Engineers (UPE) focused on operational reliability and Unit Process Development Engineers (UPDE) focused on process innovation. Their differing responsibilities imply systematically different explanation requirements, with important implications for explainable user modeling and the design of personalized systems.

The question of *who* receives an explanation is as important as what the explanation contains. Liao et al. [14] conducted interviews with practitioners working on AI products and found a significant gap between existing xAI algorithmic outputs and real user needs, proposing that explanation needs are best understood as questions users ask about AI systems. Ehsan et al. [11] showed empirically that users with and without AI background perceive the same explanations differently, concluding that the recipient's role and expertise fundamentally shape what constitutes a useful explanation. Nimmo et al. [15] further demonstrated that personalizing explanations to user characteristics is more complex than commonly assumed, identifying tensions between user preferences and explanation effectiveness that challenge simple personalization approaches. Most recently, Szymanski et al. [16] argued that stakeholder role and expertise must be treated as distinct dimensions in explanation design, noting that conflating the two leads to explanation systems that fail to serve domain experts and lay users equally. Despite this growing recognition, empirical work on role-specific explanation needs remains concentrated in consumer-facing and clinical settings. The practitioner perspective in industrial production environments, where users are domain experts whose operational roles directly shape their explanation requirements, remains largely unexplored.

This paper addresses the following research questions:

RQ1 What are the current patterns of AI adoption and the primary barriers to deployment in semiconductor manufacturing?

RQ2 What trust and accountability mechanisms do domain experts require before relying on AI-driven decisions in production?

RQ3 What specific explainability properties do practitioners need from AI systems, and how do these map onto existing xAI approaches?

The main contributions of this paper are:

- A qualitative investigation of practitioner-level explainability requirements derived from expert interviews in an industrial semiconductor manufacturing setting, addressing RQ1 and RQ2.
- Identification of a role-dependent mismatch between current xAI outputs and the explanation needs of different stakeholders in high-stakes production environments, addressing RQ3.
- Practitioner-grounded design implications for *role-adaptive* explainable AI systems, including the centrality of uncertainty quantification, incremental trust-building, and organizational accountability, with relevance for explainable user modeling and personalized systems research.

The remainder of this paper is organized as follows. Section 2 describes the interview methodology, while Section 3 presents the qualitative findings. Section 4 discusses implications for xAI research and the design of role-adaptive explainable user models in industrial settings and Section 5 concludes with directions for future work.

2. Methodology

To investigate AI adoption and explainability requirements in industrial practice, we conducted semi-structured expert interviews with professionals employed in semiconductor manufacturing. This qualitative approach was selected to elicit detailed, context-sensitive accounts of practitioners' experiences, perceptions, and expectations that would be challenging to capture through surveys or controlled experiments.

2.1. Participants

Four semi-structured interviews were conducted with professionals working in semiconductor manufacturing, representing varying levels of domain expertise. Participants ranged from intermediate practitioners familiar with process fundamentals to highly experienced specialists involved in process development and optimization. Ages ranged from 30 to 59 years, with company tenure spanning 1.5 to 25 years. Educational backgrounds included one high school graduate, one master's degree holder, and two doctoral degree holders. Two participants were employed as Unit Process Engineers (UPE) and two as Unit Process Development Engineers (UPDE). This diverse profile, spanning operational and development roles across a wide range of seniority, enabled the capture of perspectives from across the expertise spectrum.

2.2. Interview Protocol

The interview protocol comprised open-ended questions organized into eight thematic categories: demographics, professional experience, current AI usage, concerns about AI integration, trustworthiness, accountability, explainability requirements, and open reflections (Table 1). Questions were designed to elicit participants' experiences, perceptions, and expectations regarding AI integration into manufacturing processes. The questionnaire was approved by the Ethics Committee of the Faculty of Electrical Engineering, University of Sarajevo.

Table 1
Survey questions used for the interview on AI integration in semiconductor manufacturing.

Category	Question
Demographics	What is your age? What is your sex? (M/F/Other) What is your level of education? What is your current position?
Experience	How long have you been working at Infineon? How long in your current position? How would you describe your domain knowledge level?
AI Usage	Are there AI systems currently used in your organization? Could AI replace physical measurements after equipment testing?
Concerns	What are the minor and major concerns about AI integration?
Trustworthiness	How can AI systems be trustworthy for equipment reliability decisions?
Accountability	Who is responsible if an AI system causes financial loss?
Explainability	Would you like to know how AI systems make decisions? What information do you need about AI to trust it?
Final Thoughts	Would you like to add anything else?

2.3. Procedure and Ethical Considerations

Prior to participation, each interviewee received written information detailing the study objectives, interview procedure, confidentiality safeguards, and the voluntary nature of participation. Interviews were scheduled at mutually convenient times, conducted in person, and lasted approximately 15 minutes. All interviews were audio recorded with informed consent.

Following each interview, recordings were transcribed verbatim and shared with participants for review. Participants were given the opportunity to clarify, amend, or correct their statements before providing final approval. After approval, the corresponding audio recordings were permanently deleted. All responses were anonymised, and identifying information was removed from both transcripts and the final manuscript. Anonymised data were stored securely and accessed only by authorised members of the research team.

2.4. Data Analysis

Interview data were analysed using a thematic coding approach that combined deductive and inductive methods [17, 18]. The predefined question categories in Table 1 served as a deductive framework, providing an initial coding structure. Within each category, inductive coding was applied to identify emergent sub-themes and patterns not captured by the predefined structure.

The coding process was conducted successively by both authors. In the first stage, one author transcribed all interviews verbatim and performed an initial pass through each transcript, identifying and extracting key sentences relevant to each question category. In the second stage, the second author independently reviewed the extracted sentences and identified key phrases and recurring words within each category, organizing them into candidate codes. These candidate codes were then discussed jointly by both authors, refined iteratively, and consolidated into the thematic categories reported in the Section 3. Disagreements were resolved through discussion until consensus was reached.

The analysis focused on identifying shared perspectives across participants as well as points of divergence or tension. For each thematic category, convergent viewpoints were extracted alongside divergent or counterintuitive insights that challenged prevailing assumptions in the literature.

Although the sample consists of four participants from a single organization, it was purposively constructed to capture heterogeneity across the key dimensions relevant to AI adoption and explainability: role (operational vs. development), seniority (1.5 to 25 years of company tenure), and educational background (high school to doctoral level). The two roles represented: UPE and UPDE are the primary end-users of VM systems and the domain experts who make operational and development decisions based on VM predictions, making them the most informative group for the research questions addressed. In qualitative expert research, a small purposive sample of domain specialists can yield rich, analytically meaningful findings when participants are selected to represent the range of relevant perspectives [17], as is the case here. The findings should nonetheless be treated as exploratory and indicative rather than generalizable.

3. Findings

The qualitative findings synthesize expert perspectives on AI adoption in semiconductor manufacturing, highlighting recurring themes, points of convergence, and critical concerns related to operational use, trust, explainability, and organizational integration.

3.1. Current and Potential AI Usage

Current AI usage was reported to be only marginally integrated into daily operational work. Participants consistently indicated that routine process monitoring and parameter adjustment continue to rely on traditional statistical analysis and manual inspection, such as *"relying on statistical data and checking things manually when thickness changes require parameter adjustments"*. While several participants were aware of AI initiatives at the company level, these tools were not embedded in everyday workflows. As participants noted, *"there is an AI-based tool at the company level, but it is not used in daily work"*, or they were *"aware of AI initiatives but have not personally engaged with them"*. Only one participant reported occasional personal use of AI tools, primarily outside core production activities, describing the use of *"large language models for coding and research-related tasks on an occasional basis"*.

Regarding the potential for AI to replace physical measurements after equipment testing, participants generally viewed this as feasible for routine and standardised cases where outputs can be derived directly from machine data. Several experts highlighted that current physical measurements can be inconsistent across repeated runs, often requiring multiple tests before results align with specifications. In this context, AI-based approaches were perceived as a means to improve process stability and reduce unnecessary retesting. One participant noted that an *"AI-based tool can work since it will give outputs directly from process data"*, leading to *"savings in the testing step"* when repeated measurements are required. Participants also emphasised that AI could realistically replace *"some physical measurements, especially routine measurements"*, while supporting faster analysis and decision-making.

Despite this optimism, significant limitations were identified. Participants stressed that AI currently lacks the experiential *"feel for choosing the right direction when developing new recipes"*, making it unsuitable for recipe creation and fine-tuning. Additional challenges included machine- and chamber-specific behaviour, as *"every machine behaves differently"*, and the difficulty of fully identifying and monitoring all relevant influencing parameters. While AI was viewed as promising for measurement and trend prediction, expert judgment remained essential for complex process adjustments and innovation-driven tasks.

3.2. Concerns Regarding AI Integration

Participants raised concerns related to system safety, operational reliability, and organizational impact. A primary concern was the risk of undetected errors when AI systems operate autonomously. As one participant noted, *"if we rely on the system operating autonomously, there is a possibility that an undetected error could arise"*, underscoring the need for continuous *"system safety and validation checks"*. High production costs further amplified these concerns, with participants emphasising that even statistically reliable systems require *"some control behind the system"* in production environments.

Human oversight was repeatedly highlighted as essential, particularly for high-impact tasks such as recipe generation and parameter adjustment. Participants cautioned that *"relying strongly on AI"* in these contexts *"can be dangerous"* without expert supervision. More broadly, AI was expected to augment rather than replace human expertise.

Concerns were also raised about incorrect or misleading predictions leading to *"wrong conclusions"* and potential impacts on product quality, as well as possible displacement of routine tasks related to data acquisition and interpretation. Despite these concerns, AI was not perceived as inherently threatening. Participants emphasised its strength in pattern recognition and large-scale data analysis, while distinguishing these capabilities from human-driven innovation. As one participant stated, AI is *"strong at detecting patterns and deriving conclusions from large datasets"* but remains *"weaker at true innovation"*, reinforcing the need for a balanced integration strategy.

3.3. Trustworthiness Requirements

Trust in AI systems for equipment reliability decisions was framed as contingent on rigorous validation, continuous monitoring, and clearly defined safeguards. Participants emphasised the importance of *"defined limits or external validation to ensure the system is functioning correctly"*, along with mechanisms such as *"an email alert notifying that the system requires re-verification"*.

A common expectation was that AI systems should initially operate in parallel with existing human-supervised procedures. As one participant explained, *"I would first run AI in parallel with our current human-supervised procedures and compare the results"*, stressing that trust develops only when the system *"proves itself empirically"* over time. Trust was also linked to performance beyond routine operating conditions, with participants indicating greater confidence if AI systems could *"predict things which are outside of normally used parameter sets"*, supported by strong correlations with physical measurements. Tangible indicators such as reduced scrap rates and statistically improved performance were viewed as critical signals of trustworthiness. Overall, trust was described as emerging from demonstrable reliability, transparent operation, and sustained alignment with expert judgment rather than from model sophistication alone.

3.4. Accountability and Responsibility

Participants consistently emphasised that accountability for incorrect AI predictions cannot be assigned to the AI system itself, but should rest with the individuals or organizational units responsible for deployment and use. As one participant stated, *"we can't hold the system itself accountable"*, noting that errors already occur in existing processes and that AI systems should include *"verification to confirm whether it is functioning correctly"*.

Clear responsibility assignment prior to deployment was viewed as essential. Participants stressed that accountability *"should be defined by management"*, particularly if AI tools are used in production decision-making. Maintaining a human-in-the-loop approach was seen as a key risk mitigation strategy. Rather than focusing on blame, participants emphasised learning and continuous improvement. As one participant noted, *"mistakes can happen with humans or AI"*, and the priority should be *"learning from* *the error, avoiding repetition, and improving the process"*, with systems designed to retain institutional knowledge over time.

3.5. Explainability and Transparency Requirements

All participants identified explainability as essential for trust, usability, and long-term acceptance. The nature of their requirements, however, reflected their distinct operational roles, a divergence that is discussed in detail in Section 4. Participants emphasised the need to understand *"how the system operates, how it makes decisions, and how it 'thinks' "*, particularly when errors occur. Explainability was framed as a practical requirement rather than an abstract ideal, supporting faster diagnosis and informed intervention.

Participants desired insight into *"why something happens, and when something goes wrong, why it happened and where to look for errors"*, as well as *"basic knowledge of how the algorithms work"* to calibrate reliance on AI outputs. The opaque nature of many current AI systems was seen as a barrier, with participants describing them as *"largely black boxes"* even to experts.

Specific explainability requirements included visibility into input variables, parameter ranges, prediction evolution over time, and the impact of parameter changes on product properties. Participants also emphasised the importance of uncertainty awareness, including confidence measures and data sufficiency indicators. Explainability was considered especially critical for black-box models, where trust should be built through long-term performance evidence. As one participant noted, confidence could be established by demonstrating that *"over several years, the system correctly predicted outcomes and identified issues"*.

Participants advocated for incremental deployment, with AI initially positioned as a *"copilot"* that flags anomalies while leaving final decisions to human experts. Context awareness was also emphasised, with AI systems expected to recognise maintenance actions, process changes, and new products. Overall, transparent systems were seen as helping position AI *"as a helper rather than a threat"*, whereas opaque or overly complex tools risk rejection regardless of predictive accuracy. While these requirements were shared across participants, their interpretation and priority differed systematically across practitioner roles, as discussed in Section 4.

3.6. Organizational and Adoption Barriers

Participants identified organizational barriers including limited AI expertise, resistance to opaque models, integration complexity, and concerns over accountability. These barriers were reflected in the limited integration of AI into daily workflows, where conventional statistical methods and manual inspections remain dominant. While awareness of AI initiatives and large language models is increasing, adoption in routine operations remains low, highlighting the need for explainable, trustworthy, and well-integrated AI solutions tailored to industrial practice.

4. Discussion

The findings from our expert interviews reveal a consistent and concrete set of practitioner requirements for xAI in industrial production environments. In this section, we discuss the implications of these findings in relation to current xAI research and practice, identifying key gaps and directions relevant to the explainable user modeling community.

Explainability as a prerequisite for adoption, not an add-on. Across all interviews, explainability was not treated as a desirable feature but as a gating condition for trust and adoption. Participants described current AI systems as *"largely black boxes even to experts"*, and consistently linked their willingness to rely on AI outputs to the availability of process-level transparency. This finding aligns with broader calls in the xAI literature for explanation to be treated as a first-class design requirement rather than a post-hoc supplement [10, 13]. In the context of semiconductor manufacturing specifically, where incorrect decisions carry high financial and quality costs, the stakes of opacity are particularly severe. Explainability should therefore be designed into AI systems from the outset, rather than layered on after deployment.

The mismatch between current xAI outputs and practitioner needs. Despite the wide availability of xAI techniques, our findings suggest a significant gap between what current methods produce and what practitioners actually need. Standard post-hoc approaches such as SHAP [8] are well-established in VM applications, yet participants did not articulate a need for statistical feature attribution scores. Instead, they requested process-level explanations: why a prediction was wrong, which parameters to inspect, how outputs evolve over time, and whether the system accounts for contextual events such as maintenance or recipe changes. This points to the need for semiconductor-specific explainability approaches that account for temporal dependencies, equipment-process-product interactions, and physical constraints. Complementary methods such as counterfactual explanations [19], which communicate how specific interventions could alter outcomes, may be better suited to the process-optimization scenarios described by practitioners.

Uncertainty awareness as a first-class requirement. Participants explicitly requested confidence measures and indicators of data sufficiency, particularly for edge cases and out-of-distribution inputs. This requirement reflects a sophisticated understanding of model limitations and is consistent with growing advocacy for uncertainty-aware explanation methods in the xAI community [20]. In high-stakes production environments, knowing *when not to trust* a prediction is at least as important as understanding why a prediction was made. Uncertainty quantification should therefore be treated as a primary output of explainable VM systems, not a secondary diagnostic tool.

Incremental trust-building through parallel deployment. The trust-building process described by participants was inherently incremental. Participants did not expect to trust AI systems immediately; rather, they expected systems to prove themselves empirically over time, initially operating in parallel with human-supervised procedures before assuming greater responsibility. This copilot model of deployment, where AI flags anomalies while final decisions remain with human experts, resonates with the human-in-the-loop paradigm advocated in the literature and reflects the practical risk management strategies of high-stakes production environments [21]. Explainable user modeling frameworks intended for industrial deployment should accommodate this staged trust-building process, with transparency mechanisms that actively support comparison between AI recommendations and expert judgment.

Accountability as an organizational, not a technical, property. Participants agreed that accountability for incorrect AI predictions should rest with humans, specifically, with the organizational units responsible for deployment and use. This finding has direct implications for how xAI systems are governed and introduced in industrial settings. Explanation interfaces alone are insufficient; organizations require clear accountability structures, defined prior to deployment, that establish who is responsible for AI-driven decisions and how errors are handled. Research in explainable user modeling should engage more directly with these organizational dimensions, recognizing that trust in AI is not purely a function of model transparency but also of institutional readiness and governance.

Stakeholder-specific explanation needs and implications for user modeling. Our participants represented two distinct practitioner roles whose explainability needs, while sharing a common concern for transparency and trust, diverged in operationally meaningful ways that have direct implications for user modeling and personalized explanation design.

UPEs, responsible for day-to-day process monitoring and equipment reliability, expressed predominantly operational and action-oriented explanation needs. Their central question was not *how* the model works, but *what is wrong right now and where to look.* Concretely, they requested visibility into which input variables the system uses, whether those values fall within specified limits, and whether the system has detected relevant contextual changes such as maintenance events, recipe modifications, or product switches. Trust for UPEs is built procedurally: through scheduled validation runs, safety limit alerts, and re-verification notifications, rather than through algorithmic understanding. Abstract model knowledge was neither required nor desired.

UPDEs, whose work centers on process innovation, recipe development, and long-term system evaluation, expressed epistemic and uncertainty-aware explanation needs. Their central question was *how does this system reason, how confident is it, and can I trust it with edge cases.* Concretely, they requested basic understanding of the underlying algorithm to calibrate their own reliance, explicit confidence scores and data sufficiency indicators to assess prediction reliability, and counterfactual or parameter-sensitivity explanations to explore how changes in process conditions affect predicted outcomes. Trust for UPDEs is built empirically: through parallel operation over extended periods and longitudinal performance evidence across a range of process conditions. In direct response to RQ3, Table 2 maps each practitioner need to the limitations of current xAI approaches and to promising research directions.

Table 2
Mapping of practitioner explanation needs to xAI directions.

Practitioner need	**Interview evidence**	**Why current xAI is insufficient**	**Promising direction**
Why did this happen? Where to look? (UPE)	*"I want it to tell me where to look for the error"*	SHAP attribution scores do not localize operational faults or suggest intervention points	Process-level causal and fault-localization explanations
Is the system working correctly right now? (UPE)	*"defined limits or external validation to ensure the system is functioning correctly"*	Post-hoc methods do not monitor system health or flag re-verification needs	Validation-aware explanations with automated alert mechanisms
Did the context change? (UPE)	*"the system must recognise maintenance actions, process changes, and new products"*	Static models ignore operational context shifts	Contextual and temporal explanations sensitive to process events
How does this system reason? Can I trust it with edge cases? (UPDE)	*"basic knowledge of how the algorithms work, how decisions are made"*	Black-box models offer no insight into decision logic	Algorithm-level transparency and model documentation
How confident is the prediction? (UPDE)	*"confidence scores and data sufficiency indicators"*	Most post-hoc methods do not expose predictive uncertainty	Uncertainty quantification as a primary output
What if I change this parameter? (UPDE)	*"if we switch parameters x, y, z, can the system predict the outcome?"*	Feature attribution does not support what-if process exploration	Counterfactual and parameter-sensitivity explanations
Who is accountable if something goes wrong? (Both)	*"accountability should be defined by management"*	Explanation interfaces do not address organizational responsibility	Socio-technical governance frameworks alongside technical xAI

From a user modeling perspective, these two roles constitute distinct *explanation user profiles* that differ along four dimensions: explanation content (local fault localization vs. global model behavior), temporal scope (run-to-run differences vs. multi-year performance history), granularity (instance-level anomaly detection vs. model-level uncertainty quantification), and uncertainty presentation (procedural alerts vs. explicit confidence scores). This is consistent with recent empirical work showing that stakeholder role and expertise must be treated as distinct dimensions in explanation design [16, 11]. We propose that future xVM systems should incorporate role-adaptive explanation interfaces: an operationally-oriented interface for UPEs focused on real-time fault support, and an epistemically-oriented interface for UPDEs focused on model transparency and development-phase validation. This constitutes an initial, exploratory step toward stakeholder-aware personalized explanation design, a direction the explainable user modeling community is well positioned to advance.

Limitations. This study is based on four interviews conducted at a single semiconductor manufacturing company. While the findings are internally consistent and grounded in detailed practitioner accounts, generalizability across industrial domains, company sizes, and cultural contexts should be treated with caution. All interviews were conducted by a single interviewer, which may introduce interviewer bias. However, both authors jointly contributed to questionnaire design and data analysis to partially mitigate this risk. Expanding this study to include larger and more diverse participant samples spanning managers, data scientists, and maintenance engineers across multiple organizations would strengthen the robustness of the identified requirements.

5. Conclusion

This paper presented an exploratory qualitative investigation of AI adoption and explainability requirements in semiconductor manufacturing, grounded in semi-structured expert interviews with domain practitioners. The findings suggest a consistent mismatch between what current xAI methods offer and what practitioners actually need: process-level causal transparency, uncertainty quantification, contextual awareness of operational events, and staged deployment alongside human supervision. Given the exploratory nature of the study, these findings should be treated as indicative rather than generalizable, and as a foundation for future work with larger and more diverse samples.

Beyond individual requirements, the two practitioner roles studied exhibited distinct explanation user profiles. UPEs need operational, fault-localization explanations oriented toward real-time support, while UPDEs need epistemic, uncertainty-aware explanations oriented toward algorithmic calibration and development-phase validation. This divergence points toward role-adaptive explanation interfaces as a concrete design direction for the explainable user modeling community, and positions accountability as an organizational rather than a technical property that should be defined prior to deployment. In the longer term, this work is a first step toward personalizing explainable AI systems for industrial domain experts, adapting explanation content, granularity, and uncertainty presentation to the specific roles, goals, and decision contexts of practitioners who rely on AI predictions in production.

Ultimately, the practitioner-grounded requirements identified here provide an empirical foundation for future design and evaluation of explainable AI systems in industrial settings, contributing toward trustworthy, human-centered AI that treats explanation not as an optional feature, but as a prerequisite for meaningful human-AI collaboration in high-stakes production environments.

Acknowledgments

The work is linked to the FID activities of the IPCEI on ME (Important Project of Common European Interest on Microelectronics), which is funded by national authorities from Germany, France, Italy, the UK, and Austria.

Declaration on Generative AI

During the preparation of this work, the authors used Claude (Anthropic) and Grammarly for language proofreading and writing assistance. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] J. Moyne, J. Iskandar, Big data analytics for smart manufacturing: Case studies in semiconductor manufacturing, Processes 5 (2017) 39.

[2] G. A. Susto, A. Beghi, A virtual metrology system based on least angle regression and statistical clustering, Applied Stochastic Models in Business and Industry 29 (2013) 362–376.

[3] P. Kang, D. Kim, H.-j. Lee, S. Doh, S. Cho, Virtual metrology for run-to-run control in semiconductor manufacturing, Expert Systems with Applications 38 (2011) 2508–2522.

[4] European Parliament, Council of the European Union, Regulation (EU) 2016/679 – General Data Protection Regulation (GDPR), Technical Report L 119, Official Journal of the European Union, 2016. Available at: https://eur-lex.europa.eu/eli/reg/2016/679/oj.

[5] European Commission, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act), Technical Report COM(2021) 206 final, European Commission, 2021. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/.

[6] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, et al., Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai, Information fusion (2020).

[7] D. Kaur, S. Uslu, K. J. Rittichier, A. Durresi, Trustworthy artificial intelligence: a review, ACM computing surveys (CSUR) (2022).

[8] S. Lundberg, A unified approach to interpreting model predictions, arXiv preprint arXiv:1705.07874 (2017).

[9] H. B. Santoso, D. K. Baroroh, N. T. H. Van, Integrating xai in metaverse for operator 5.0: An analytical review, in: 2024 21st International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), IEEE, 2024, pp. 1–6.

[10] L. Longo, M. Brcic, F. Cabitza, J. Choi, R. Confalonieri, J. Del Ser, R. Guidotti, Y. Hayashi, F. Herrera, A. Holzinger, et al., Explainable artificial intelligence (xai) 2.0: A manifesto of open challenges and interdisciplinary research directions, Information Fusion 106 (2024) 102301.

[11] U. Ehsan, S. Passi, Q. V. Liao, L. Chan, I.-H. Lee, M. Muller, M. O. Riedl, The who in xai: how ai background shapes perceptions of ai explanations, in: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, 2024, pp. 1–32.

[12] N. Tintarev, J. Masthoff, Designing and evaluating explanations for recommender systems, in: Recommender systems handbook, Springer, 2010, pp. 479–510.

[13] C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nature machine intelligence 1 (2019) 206–215.

[14] Q. V. Liao, D. Gruen, S. Miller, Questioning the ai: informing design practices for explainable ai user experiences, in: Proceedings of the 2020 CHI conference on human factors in computing systems, 2020, pp. 1–15.

[15] R. Nimmo, M. Constantinides, K. Zhou, D. Quercia, S. Stumpf, User characteristics in explainable ai: The rabbit hole of personalization?, in: Proceedings of the 2024 CHI conference on human factors in computing systems, 2024, pp. 1–13.

[16] M. Szymanski, V. Vanden Abeele, K. Verbert, Disentangling stakeholder role and expertise in user-centered explainable ai, in: Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization, 2025, pp. 32–39.

[17] V. Braun, V. Clarke, Using thematic analysis in psychology, Qualitative research in psychology 3 (2006) 77–101.

[18] J. Fereday, E. Muir-Cochrane, Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development, International journal of qualitative methods 5 (2006) 80–92.

[19] S. Verma, J. Dickerson, K. Hines, Counterfactual explanations for machine learning: Challenges revisited, arXiv preprint arXiv:2106.07756 (2021).

[20] U. Bhatt, A. Weller, J. M. Moura, Evaluating and aggregating feature-based model explanations, arXiv preprint arXiv:2005.00631 (2020).

[21] E. Mosqueira-Rey, E. Hernández-Pereira, D. Alonso-Ríos, J. Bobes-Bascarán, Á. Fernández-Leal, Human-in-the-loop machine learning: a state of the art, Artificial Intelligence Review 56 (2023) 3005–3054.