From Regulation to Interaction: Expert Views on Aligning Explainable AI with the EU AI Act

Mahdi Dhaini, Lukas Ondrus, Gjergji Kasneci
Technical University of Munich, Germany
School of Computation, Information and Technology
Department of Computer Science
Munich, Germany
{firstname.lastname}@tum.de

Abstract

Explainable AI (XAI) aims to support people who interact with high-stakes, AI-driven decisions, and the EU AI Act requires that users can appropriately interpret high-risk AI system outputs (Article 13) and that human oversight prevents undue reliance (Article 14). Yet the Act offers little technical guidance on implementing explainability, leaving interpretability methods difficult to operationalize and compliance obligations unclear. To address these gaps, we interviewed eight domain experts across legal, compliance, and technical roles to explore (1) how explainability is defined and perceived under the Act, (2) the practical and regulatory obstacles to XAI implementation, and (3) recommended solutions and future directions. Our findings reveal that domain experts view explainability as context- and audience-dependent, face challenges from regulatory vagueness and technical trade-offs, and advocate for domain-specific rules, hybrid methods, and user-centered explanations. These insights provide a basis for a potential framework to align XAI methods—particularly for AI and Natural Language Processing (NLP) systems—with regulatory requirements, and suggest actionable steps for policymakers and practitioners

1 Introduction and Background

With the increasing deployment of large AI models (e.g., pretrained language models and large language models (LLMs)) in high-stakes domains, their inherent "black-box" nature offers limited transparency into decision-making processes, posing significant risks. Consequently, regulations such as the GDPR and the EU AI Act require both transparency and explainability. This has paved the way for the development of extensive research in the **Explainable AI (XAI)** field. Research in XAI has been increasing, with numerous surveys focused on critical domains such as healthcare, finance, and law (Chaddad et al., 2023; Richmond et al., 2023; Yeo et al., 2025). Moreover, information and knowledge management systems are increasingly integrating explainable AI methods to support trustworthy decision-making in different domains (Rožanec et al., 2022; Mancuso et al., 2025; Majumder and Dey, 2022; Chen et al., 2024; Brasse et al., 2023)

The **EU AI Act** (Commission, 2021) mandates transparency for high-risk AI systems via: *Article 13(1)*: systems must enable users to interpret outputs appropriately and *Article 14(4)*: human oversight must prevent undue reliance on AI decisions. However, the Act provides no technical guidance for implementing explainability methods (Panigutti et al., 2023; Gyevnar et al., 2023). Existing XAI techniques focus on algorithmic transparency but often remain opaque to non-experts and difficult to operationalize for intended users(Panigutti et al., 2023; Golpayegani et al., 2023). Moreover, the high-risk classification (Art. 6(2)) is ambiguous, creating uncertainty about compliance obligations (Golpayegani et al., 2023; Nisevic et al., 2024). Consequently, many XAI methods offer mathematical insight without satisfying legal standards for human interpretability (Panigutti et al., 2023; Kusche, 2024).

Previous research shows that no standardized framework exists to assess whether XAI methods satisfy the EU AI Act's transparency mandates (Sovrano et al., 2022; Panigutti et al., 2023). Second, legal expectations for explainability remain ambiguous, with no consensus on what qualifies as an acceptable explanation (Panigutti et al., 2023; Gyevnar et al., 2023). Third, research rarely integrates explainability techniques with compliance mechanisms, as most studies address XAI methods and regulatory frameworks in isolation (Hacker, 2023; Nisevic et al., 2024).

Despite extensive XAI research, little work has engaged experts on how current methods align with the EU AI Act's transparency and explainability

Proceedings of the Fourth Workshop on Bridging Human-Computer Interaction and Natural Language Processing (HCI+NLP), pages 230–239

requirements. With key compliance deadlines approaching (Timeline, 2024), understanding practitioners' views on the feasibility and challenges of meeting these mandates is critical yet largely unexplored. Aiming to address these gaps, we conducted an interview study with 8 domain experts from Europe, addressing the following Research Questions (RQs):

- ***RQ1****: How do domain experts define and perceive explainability in the context of the EU AI Act's transparency requirements, and what usability and technical factors shape these perceptions?*
- **RQ2**: *What practical and regulatory obstacles do practitioners face when implementing XAI to comply with the EU AI Act?*
- ***RQ3****: What solution concepts and future directions do domain experts recommend for aligning XAI methods with regulatory requirements?*

This work investigates domain experts' perspectives on the intersection of XAI and the EU AI Act, focusing on perceptions and the challenges of implementing XAI to achieve compliance. This paper provides an empirical perspective grounded in expert interviews and offers evidence-based and actionable insights for aligning XAI with regulatory demands. The remainder of the paper is structured as follows: Section 2 introduces the research methodology; Section 3 presents the results of our study; Section 4 reviews related work in the literature. Finally, Section 5 discusses our findings, presents our conclusions, and outlines future work and study limitations.

2 Research Methodology

We followed a qualitative research approach and drew on the experiences of experts by conducting semi-structured interviews (SSI), which offer flexibility while retaining a structure for data collecting and enable a full analysis of participants' experiences (DiCicco-Bloom and Crabtree, 2006). To examine the potential gaps and trends in the literature, we first conducted an initial literature review (Fresz et al., 2025; Freiesleben and König, 2023; Rožanec et al., 2022; Mancuso et al., 2025; Majumder and Dey, 2022; Chen et al., 2024; Brasse et al., 2023) to build our knowledge base and as the foundation for the questionnaire. Then, we interviewed experts and professionals to collect input on the intersection between XAI and the EU AI Act, focusing on understanding the context, interventions, mechanisms, and outcomes of XAI implementations while addressing regulatory compliance and practical challenges. We describe below the study design, data collection, and data analysis of our research.

2.1 Study Design

Drawing on the five-step process of Kallio et al. (2016), we developed our semi-structured *interview guide* as follows: *(1) Prerequisites:* Ensured that researchers possessed sufficient background in XAI, law, and policy to identify key topics and appreciate multiple stakeholder perspectives. *(2) Literature review:* Conducted a literature review to build a conceptual framework, focus on AI Act compliance issues, and identify knowledge gaps. *(3) Drafting:* Created an initial question set combining core concepts and flexible prompts to explore experts' views on XAI under the EU AI Act. *(4) Pilot testing:* Refined wording, structure, and flow through three internal iterations and a final supervisor review. *(5) Finalization:* Documented the complete guide to ensure replicability for future studies.

We reached out to experts and professionals in the fields of law, policy, and AI development, as the EU AI Act is a complex regulatory instrument to ensure domain diversity in our interviews. We assumed expertise when the participants had 4 years or more of experience in the aforementioned fields. Recruitment occurred in two waves: an initial phase (December 2024) via personal contacts, third-party introductions, and online searches, followed by a snowball phase (January 2025) based on interviewee recommendations. In total, we contacted 50 candidates between December 2024 and January 2025 via email, LinkedIn, and phone. Among the 50 candidates, 8 expressed interest in participating in an interview. We present demographic information on our participants in Table 1. All participants we interviewed were based in Europe, reflecting our focus on those subject to the EU AI Act. Except for P1, a government official, all worked in the private sector. Each had at least five years' experience in law, policy, or AI development.

Participant	Position	Expertise	Experience	Organization	Size
P1	Researcher	Legal Expert, Regulatory Compliance	10	University	Large
P2	Senior Data Engineer	Data Platforms and Governance	7	Consulting Firm	Large
P3	AI Consultant	Regulatory Technology	20	Consulting Firm	Large
P4	CEO	AI Education and Training	15	AI Company	Small
P5	Account Executive	Enterprise Data & AI Solutions	10	AI & Software Company	Large
P6	Chief Research Scientist	NLP and Explainable AI Research	10	AI Research Firm	Medium
P7	AI Governance Consultant	Strategic Innovation Management & AI Governance	5	Telecommunication Company	Large
P8	Research Scientist	NLP and Explainable AI	13	AI Research Firm	Medium

Table 1: Overview of Interviewee Demographics. Experience is measured in years.

2.2 Data Collection

We conducted semi-structured interviews via Microsoft Teams between December 2024 and March 2025. To achieve observer triangulation (Runeson and Höst, 2009), two researchers attended each session. At the outset, participants were briefed on recording procedures, anonymization protocols, and the intended use of their transcripts. The research objectives and interview structure were reiterated to ensure understanding. Although the sequence and wording of questions remained fixed, the semi-structured format permitted slight modifications in response to conversational flow. Interview prompts were drawn from themes identified during the preliminary literature review and organized into five thematic sections: (1) *Background and common basis* which included collecting foundation data about the participants like their professional background (2) *Explainability perceptions* where participants were asked on the definition of explainability and how they perceive its usability and limitation (3) *XAI implementation challenges* and (4) *Regulatory gaps and compliance risks* and (5) *Future directions and solution concepts* and finally (6) *Closing* where we invited participants to share any additional insights and to indicate their willingness for follow-up contact or to recommend other participants. To ensure reproducibility and transparency, we include the full interview questionnaire in Appendix A.

2.3 Data Analysis

The interviews were recorded and transcribed with the help of *otter.ai* with the participant's consent. We omit the full interview transcripts as the participants requested anonymity. We then systematically coded the transcriptions following the thematic approach described by (Braun and Clarke, 2006) as follows: The following guidelines were applied and adapted to our data:
(1) Familiarization with extracted data: we read and re-read the interview transcripts to identify relevant data and initial patterns. Extracted data were organized in Excel according to the questions in the interview guide.
(2) Generating initial codes: we systematically analyzed each transcript, assigning concise labels (*codes*) to meaningful text segments. Codes were entered in a dedicated Excel column to facilitate topic filtering.
(3) Searching for themes: related codes were grouped into broader categories (*themes*) that reflect the research questions. We added a "theme" column in Excel to cluster codes accordingly.
(4) Reviewing themes: we checked each provisional theme against the full dataset to ensure it accurately represented the underlying patterns and refined them for coherence.
(5) Defining and Naming Themes: final themes were precisely defined and named. Each theme's description was recorded in Excel along with its linkage to the corresponding research question.
(6) Synthesizing findings: we synthesized and presented the thematic findings using illustrative data extracts to construct a coherent narrative of our findings (Braun and Clarke, 2006).

3 Results

In this section, we present the results from our SSI study, we organize the results according to the thematic coding process explained earlier. Table 2 presents the final themes developed to address our research questions, derived from initial coding of participant responses and subsequent thematic grouping. In the next subsections, we present results of each of the thematic sections.

Main Theme	Description	Codes
Explainability Perceptions (RQ1)	How experts define explainability and perceive its usability and limitations.	Lack of established definition User-specific definitions Technical complexity
Implementation Challenges (RQ2)	Difficulties in applying XAI methods in practice.	Regulatory vagueness Accuracy trade-offs Technical expertise required
Regulatory Gaps and Compliance Risks (RQ2)	Regulatory insights, compliance challenges, and possible mitigations.	Lack of specific guidance Superficial compliance risks Lack of regulatory sandboxes
Future Directions and Solution Concepts (RQ3)	Expert recommendations on advancing XAI methodologies and regulations.	Hybrid approaches User-centered explanations Domain-specific regulations

Table 2: Key themes, descriptions, and codes relevant to the research questions

3.1 Perceptions of Explainability (RQ1)

Table 3 presents some representative answers from participants regarding their input on their perceptions of explainability. Participants emphasized the absence of a **standardized definition of explainability** and related XAI terms, noting the importance of a consistent framework. Most agreed that explainability must be context-dependent and tailored to various stakeholders: "explainability must vary based on the intended user and application domain." (P1). P3 reinforced this point, stating, "Every stakeholder understands explainability differently." Concurrently, P2 mentioned that "current methods are not easily understood by non-expert users," highlighting the inherent **technical complexity**. Together, these findings indicate a clear consensus on the need for **context-dependent and user-specific explainability**. They also reveal a substantial gap between existing theoretical constructs and practical implementation.

3.2 Implementation and Compliance Challenges (RQ2)

To answer RQ2, we asked the participants about their challenges and barriers to implementing the XAI method in practice, mainly for compliance with the EU AI Act. We present in Table 4 some representative answers from participants. Participants reported multiple challenges in implementing XAI methods. First, **regulatory vagueness** was frequently highlighted. P2 and P8 criticized the significant difficulties arising from unclear definitions of compliance under the EU AI Act, suggesting that the lack of concrete guidance hampers practical implementation. P2 stated, "Clear guidance from the EU AI Act is missing, making practical compliance difficult to achieve." P8 echoed this concern: "Clearer compliance definitions are needed to effectively operationalize these requirements." Second, **accuracy trade-offs** emerged as a prominent barrier. P4 and P6 noted an inherent conflict between interpretability and model accuracy. As P4 remarked, "When we aim for transparent explanations, predictive accuracy often suffers."

A lack of general **technical expertise** within organizations was identified as a significant hurdle. P1, P3, and P7 criticized many organizations for lacking dedicated AI teams or trained specialists in explainability and transparency P1 particularly highlighted the "the complexity of explainability tools is often underestimated" and how "teams lack the technical skills required to implement them effectively"(P1).

Conclusion on implementation challenges: unclear regulations, the trade-off between transparency and performance, and insufficient technical capacity collectively obstruct the adoption of XAI in practice for sake of compliance.

3.3 Regulatory Gaps and Compliance Risks (RQ2)

To further answer RQ2, we collected from interviewees their input on the gaps in regulations and potential risks of implementing XAI under the current AI Act from compliance, and also some possible mitigations to these challenges. In Table 5 we present results. The greatest concern regarding the AI Act was a lack of specific guidance. They expressed concern about the potential for superficial or even ineffective compliance strategies. Interview results show repeated emphasis from participants on the absence of clear standards, where "Explicit standards for compliance would greatly improve transparency efforts" (P3). P5 emphasized that "The Act needs clearer standards for transparency compliance", and P8 highlighted how a "Clearer

Code	Representative participants answers
User-specific explanations	"You need tailored explanations for different user groups." (P5) "Explanations must fit the cognitive abilities of the users." (P7)
Technical complexity	"Common explainability tools like SHAP are too technical for most end-users." (P4)
Lack of established definition	"The industry lacks a standard definition for explainability." (P6) "There's still no universally agreed definition of explainability." (P8)

Table 3: Representative answers on explainability perceptions

Code	Representative participants answers
Technical expertise	"The complexity of explainability tools is often underestimated. Teams lack the technical skills required to implement them effectively." (P1) "Few companies actually have the technical expertise in-house to leverage advanced XAI methods." (P3) "Organizations without specialized AI teams find it especially difficult to implement these tools correctly." (P7)
Regulatory vagueness	"Clear guidance from the EU AI Act is missing, making practical compliance difficult to achieve." (P2) "The EU AI Act does not specify what exactly counts as transparent or interpretable." (P5) "Clearer compliance definitions are needed to effectively operationalize transparency." (P8)
Accuracy trade-offs	"When we increase interpretability, we lose a lot in accuracy. It's a fundamental trade-off." (P4) "Achieving full transparency in AI means often sacrificing predictive accuracy." (P6)

Table 4: Representative answers on implementation challenges

guidance from the EU would help companies implement meaningful XAI." These results point to gaps in the AI Act in terms of an overarching lack of specific and operational guidance. Participants also highlighted some **risks** that could result from enforcing compliance where in the absence of precise definitions, organizations may adopt "superficial explanations just to meet regulations" (P1) or "easy but ineffective solutions to meet vague legal requirements" (P4), our study results indicate participants worry that companies may end up merely *"checking the box"* rather than genuinely implementing/adopting explainability to achieve superficial compliance with the regulation which stem from the regulatory vagueness discussed before. This also indicates the need to come up with and implement effective checkpoints to prevent this.

As for potential **mitigation strategies**, some participants emphasized on the need and value of *regulatory sandboxes*, as "regulatory sandboxes allow realistic testing of compliance strategies safely" (P1), and P3 further stressed that "explicit standards for compliance would greatly improve transparency efforts", noting the lack of specificity in current regulations. Overall, results reflect a clear demand from experts and professionals for more concrete guidance on explainability from regulators.

3.4 Future Directions for Compliance (RQ3)

We also asked participants about future directions for strategies and recommendations from an XAI and regulatory perspective to facilitate meeting regulatory requirements in a practical way. Table 6 presents a subset of participants' answers on this topic.

Domain-specific and tailored regulations were frequently mentioned, mainly as "tailored regulations for different sectors could improve practical compliance" (P5). P1 emphasized that "explainability should be regulated with sensitivity to specific industry needs". These results suggest that a *one-size-fits-all approach* is not feasible, as transparency requirements vary across AI application domains. **Hybrid approaches** also emerged as a main theme where participants emphasized hybrid methods could balance interpretability with model performance. P2 emphasized, "A blend of accuracy and interpretability is crucial", and P6

Code	Representative participants answers
Superficial compliance risks	"Companies might adopt easy but ineffective solutions to meet vague legal requirements." (P4) "Without clarity, organizations might adopt superficial solutions just for compliance." (P7)
Lack of regulatory sandboxes	"Regulatory sandboxes allow realistic testing of compliance strategies safely." (P2) "Testing in safe regulatory sandboxes helps clarify vague regulatory requirements." (P6)
Lack of specific guidance	"The Act needs clearer standards for transparency compliance." (P5)

Table 5: Representative answers on regulatory implications

Code	Representative participants answers
Domain-specific regulations	"Explainability should be regulated with sensitivity to specific industry needs." (P1)
Hybrid approaches	"Hybrid methods would allow balancing accuracy and interpretability effectively." (P7)
User-centered explanations	"Research should focus more explicitly on the usability of explanations." (P8)

Table 6: Representative answers on future directions

added, "Integrating technical detail and practical interpretability is essential," highlighting the need to combine complementary explanation techniques. The final theme concerned **user-centered explanations**, "understanding user needs should be the priority of future research" (P3), a statement which was also echoed by P4, who remarked that "user-oriented research is crucial to developing meaningful explanations." participants emphasized that XAI solutions must be tailored to stakeholders' abilities and contexts to ensure explanations are both usable and informative.

These insights reflect evolving perspectives on the interaction between XAI development and the EU AI Act, illustrating directions for tailoring explainability methods and clarifying regulation to achieve explanations that are both user-aligned and compliant.

4 Related Work

Recent studies have examined the intersection of XAI and the EU AI Act, highlighting some regulatory and technical challenges. Panigutti et al. (2023) conducted an interdisciplinary analysis of how the Act addressed black-box AI systems, reviewing existing XAI techniques and noting persistent limitations in meeting the Act's objectives. Using a case study on AI-based exam proctoring, they illustrated how transparency and human oversight could be implemented even when decisions were not inherently interpretable. Similarly, Fresz et al. (2025) analysed explainability requirements under European and German law, finding that different legal frameworks demanded distinct XAI properties and that current techniques often failed to meet expectations, particularly regarding fidelity and confidence estimates. These shortcomings reflected technical trade-offs, as many post-hoc methods lacked faithfulness guarantees, posing compliance challenges. Pavlidis (2024) explored various approaches to advance XAI and discussed the difficulties of embedding explainability into governance and policy, with attention to standard-setting, oversight, and enforcement. Hacker and Passoth (2022) further emphasized the context-dependent nature of explainability, noting that the appropriate form and degree of transparency depended on specific circumstances. They also observed that the Act's obligations were largely aimed at professional users of high-risk AI systems rather than the general public, which could result in transparency measures that overwhelm non-expert users with overly technical details.

While prior studies have examined the EU AI Act's transparency provisions from legal, policy, or interdisciplinary perspectives (Panigutti et al., 2023; Fresz et al., 2025; Pavlidis, 2024; Hacker and Passoth, 2022), they have primarily focused on conceptual analyses, legal interpretations, or technical reviews of existing XAI methods. While they provide valuable insights, they don't systematically engage with practitioners and domain experts

to understand how explainability is interpreted in real-world settings, nor how the Act's high-level mandates translate into concrete implementation practices. Our study fills this gap by using semi-structured interviews with eight experts in high-stakes AI to examine how they interpret explainability under the EU AI Act, the factors influencing their views, and the practical barriers to compliance, revealing nuances often overlooked in legal or technical analyses. Our work contributes *actionable insights by synthesizing experts' recommendations into potential solution concepts.*

5 Discussion and Conclusion

Our interview study findings with eight experts reveal that explainability is fundamentally *context- and user-specific*, challenging the EU AI Act's broad transparency mandates. In practice, organizations struggle with three main obstacles: (1) **regulatory vagueness**, which leaves compliance criteria undefined, (2) the inherent **accuracy–interpretability trade-off**, forcing a choice between performance and clarity, and (3) limited **technical expertise**, which delays effective integration of XAI tools. Together, these factors could lead to superficial "box-checking" rather than genuine explainability.

Experts propose a threefold strategy to bridge these gaps. (1) *domain-specific regulations* would adapt transparency requirements to each sector's risk profile. (2) *hybrid approaches* combining high-accuracy models with simpler explainers can reconcile predictive performance with comprehensibility. (3) *user-centered design* must guide both the development and evaluation of explanations, ensuring they meet stakeholders' cognitive and operational needs.

Our findings have several **implications** for different *stakeholders*. *Researchers* are encouraged to develop evaluation frameworks that jointly assess legal compliance and end-user comprehensibility. *Policymakers and regulators* should consider refining the EU AI Act by adding sector-tailored criteria and establishing regulatory sandboxes for safe testing. On the other hand, *practitioners* can use these results as a basis to invest in interpretability training and adopt flexible XAI pipelines. By synthesizing expert insights on definitions, challenges, and solution concepts, this work lays the groundwork for aligning XAI methods with regulatory requirements and advancing trustworthy AI in high-risk domains. Beyond regulatory alignment, our findings have implications for the design of explanation interfaces and interaction modalities, particularly those relying on natural language (Atanasova et al., 2023; Madsen et al., 2024, 2022). Incorporating user-centered design principles into explanation generation (Mishra et al., 2024; Wang et al., 2025) and presentation can help bridge the gap between legal compliance and meaningful human understanding, an intersection of growing interest in human-computer interaction and NLP research.

Unlike prior work, which has mainly offered legal or conceptual analyses of the EU AI Act's explainability requirements, our study provides an empirical perspective grounded in expert interviews. By capturing how practitioners interpret explainability, the challenges they face in achieving compliance, and their proposed solutions, we offer evidence-based, actionable insights for aligning XAI methods with regulatory demands.

We present this study as work in progress and hope it catalyzes dialogue between HCI and XAI/NLP researchers, laying a foundation for further work on explainability under regulatory mandates. With this study, we aim to advance efforts to bridge the gap between human- and regulator-oriented explainability and technical explainability in AI/NLP.

Future Work In our current sample, interviewees are primarily based in Europe, reflecting the particular relevance of the EU AI Act in this region. As future work, we plan to conduct additional interviews with experts from both European and non-European countries to expand the sample and enable comparative analyses between EU and non-EU perspectives. A complementary direction is to field a large-scale survey to validate and generalize our qualitative findings, quantify key insights, and broaden their applicability. Together, these extensions aim to support more concrete recommendations and implications for the design of explanation systems, including natural language-based approaches.

Limitations This study's modest sample size, while comparable with relevant research (Warren et al., 2025), may limit generalizability beyond the examined contexts. However, thematic saturation appeared to be reached, as no new themes emerged in later interviews. All interviews were conducted in English, and although the sample in-

cluded participants from diverse linguistic backgrounds, conducting them in additional languages could have broadened participation. We invite future research to build on and validate these findings through larger, more diverse samples. While the interviews were designed to follow a consistent and impartial approach, and reflexivity was actively maintained throughout the analysis, the findings may still carry the inherent biases and limitations of self-reported data (Donaldson and Grant-Vallone, 2002). In addition, one limitation of these types of works (e.g., prior related work we discussed and our work) could be the evolving nature of the EU AI Act, driven by the rapid development of AI research and practice. The Act has undergone several key milestones. In our study, we referred to the latest version of the Act available as of December 2024.

Acknowledgment

We thank the anonymous reviewers for their insightful and constructive feedback and the interview participants for their valuable contribution to this study. This research has been supported by the German Federal Ministry of Education and Research (BMBF) grant 01IS23069 Software Campus 3.0 (TU München).

References

Pepa Atanasova, Oana-Maria Camburu, Christina Lioma, Thomas Lukasiewicz, Jakob Grue Simonsen, and Isabelle Augenstein. 2023. Faithfulness tests for natural language explanations. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 283–294, Toronto, Canada. Association for Computational Linguistics.

Julia Brasse, Hanna Rebecca Broder, Maximilian Förster, Mathias Klier, and Irina Sigler. 2023. Explainable artificial intelligence in information systems: A review of the status quo and future research directions. *Electronic Markets*, 33(1):26.

Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101.

Ahmad Chaddad, Jihao Peng, Jian Xu, and Ahmed Bouridane. 2023. Survey of explainable ai techniques in healthcare. *Sensors (Basel, Switzerland)*, 23(2):634.

Huaming Chen, Jun Zhuang, Yu Yao, Wei Jin, Haohan Wang, Yong Xie, Chi-Hung Chi, and Kim-Kwang Raymond Choo. 2024. Trustworthy and responsible ai for information and knowledge management system. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, CIKM '24, page 5574–5576, New York, NY, USA. Association for Computing Machinery.

European Commission. 2021. Proposal for a regulation on artificial intelligence. Accessed: 27-February-2024.

Barbara DiCicco-Bloom and Benjamin F Crabtree. 2006. The qualitative research interview. *Medical Education*, 40(4):314–321.

Stewart I Donaldson and Elisa J Grant-Vallone. 2002. Understanding self-report bias in organizational behavior research. *Journal of business and Psychology*, 17:245–260.

Timo Freiesleben and Gunnar König. 2023. Dear xai community, we need to talk! In *Explainable Artificial Intelligence*, page 48–65, Cham. Springer Nature Switzerland.

Benjamin Fresz, Elena Dubovitskaya, Danilo Brajovic, Marco F. Huber, and Christian Horz. 2025. *How Should AI Decisions Be Explained? Requirements for Explanations from the Perspective of European Law*, page 438–450. AAAI Press.

Delaram Golpayegani, Harshvardhan J. Pandit, and Dave Lewis. 2023. To be high-risk, or not to be—semantic specifications and implications of the ai act's high-risk ai applications and harmonised standards. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, page 905–915, New York, NY, USA. Association for Computing Machinery.

Balint Gyevnar, Nick Ferguson, and Burkhard Schafer. 2023. Bridging the transparency gap: What can explainable ai learn from the ai act? In *ECAI 2023*, pages 964–971. IOS Press.

Philipp Hacker. 2023. The european ai liability directives – critique of a half-hearted approach and lessons for the future. *Computer Law Security Review*, 51:105871.

Philipp Hacker and Jan-Hendrik Passoth. 2022. *Varieties of AI Explanations Under the Law. From the GDPR to the AIA, and Beyond*, pages 343–373. Springer International Publishing, Cham.

Hanna Kallio, Anna-Maija Pietilä, Martin Johnson, and Mari Kangasniemi. 2016. Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *Journal of Advanced Nursing*, 72(12):2954–2965.

Isabel Kusche. 2024. Possible harms of artificial intelligence and the eu ai act: fundamental rights and risk. *Journal of Risk Research*, 0(0):1–14.

Andreas Madsen, Sarath Chandar, and Siva Reddy. 2024. Are self-explanations from large language models faithful? In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 295–337, Bangkok, Thailand. Association for Computational Linguistics.

Andreas Madsen and 1 others. 2022. Post-hoc interpretability for neural nlp: A survey. *ACM Comput. Surv.*, 55(8).

Soumi Majumder and Nilanjan Dey. 2022. *Explainable Artificial Intelligence (XAI) for Knowledge Management (KM)*, page 101–104. Springer, Singapore.

Ilaria Mancuso, Antonio Messeni Petruzzelli, Umberto Panniello, and Federico Frattini. 2025. The role of explainable artificial intelligence (xai) in innovation processes: a knowledge management perspective. *Technology in Society*, 82:102909.

Aditi Mishra, Sajjadur Rahman, Kushan Mitra, Hannah Kim, and Estevam Hruschka. 2024. Characterizing large language models as rationalizers of knowledge-intensive tasks. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 8117–8139, Bangkok, Thailand. Association for Computational Linguistics.

Maja Nisevic, Arno Cuypers, and Jan De Bruyne. 2024. Explainable ai: Can the ai act and the gdpr go out for a date? In *2024 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.

Cecilia Panigutti, Ronan Hamon, Isabelle Hupont, David Fernandez Llorca, Delia Fano Yela, Henrik Junklewitz, Salvatore Scalzo, Gabriele Mazzini, Ignacio Sanchez, Josep Soler Garrido, and Emilia Gomez. 2023. The role of explainable ai in the context of the ai act. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, page 1139–1150, New York, NY, USA. Association for Computing Machinery.

Georgios Pavlidis. 2024. Unlocking the black box: analysing the eu artificial intelligence act's framework for explainability in ai. *Law, Innovation and Technology*, 16(1):293–308.

Karen McGregor Richmond, Satya M. Muddamsetty, Thomas Gammeltoft-Hansen, Henrik Palmer Olsen, and Thomas B. Moeslund. 2023. Explainable ai and law: An evidential survey. *Digital Society*, 3(1):1.

Jože M. Rožanec, Blaž Fortuna, and Dunja Mladenić. 2022. Knowledge graph-based rich and confidentiality preserving explainable artificial intelligence (xai). *Information Fusion*, 81:91–102.

Per Runeson and Martin Höst. 2009. Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering*, 14(2):131–164.

Francesco Sovrano, Salvatore Sapienza, Monica Palmirani, and Fabio Vitali. 2022. Metrics, explainability and the european ai act proposal. *J*, 5(1):126–138.

EU AI Act–Implementation Timeline. 2024. Implementation timeline of the eu artificial intelligence act. Accessed: 27-February-2025.

Qianli Wang, Tatiana Anikina, Nils Feldhus, Simon Ostermann, Sebastian Möller, and Vera Schmitt. 2025. Cross-refine: Improving natural language explanation generation by learning in tandem. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 1150–1167, Abu Dhabi, UAE. Association for Computational Linguistics.

Greta Warren, Irina Shklovski, and Isabelle Augenstein. 2025. *Show Me the Work: Fact-Checkers' Requirements for Explainable Automated Fact-Checking*. Association for Computing Machinery, New York, NY, USA.

Wei Jie Yeo, Wihan Van Der Heever, Rui Mao, Erik Cambria, Ranjan Satapathy, and Gianmarco Mengaldo. 2025. A comprehensive review on financial explainable ai. *Artificial Intelligence Review*, 58(6):189.

A Interview Guide

Disclaimer

Before we begin, I want to inform you that this interview will be recorded for transcription purposes and used solely for a research paper. Your identity will remain confidential; no identifiable information will appear in the final document. External AI tools may be used for transcription. Could you please confirm your consent to these terms?

Introduction

This interview focuses on understanding the role of explainability in AI systems, particularly in the context of the EU AI Act. I will ask about your experiences, opinions, and insights regarding explainable AI, regulatory challenges, and practical tools. I can provide further clarification on any topic during the interview if needed.

Warm-Up Questions

- What is your current role?
- Could you briefly share your background and experience with AI and explainable AI?
- When you hear the word "explainability" in the context of AI, what does it mean to you?
- How familiar are you with the EU AI Act and its provisions related to AI transparency and explainability?

Understanding the EU AI Act

Background Note: The EU AI Act emphasizes requirements like transparency, accountability, and human oversight for high-risk AI systems. Explainability is considered essential to meet these requirements.

- Article 13 addresses transparency and provision of information to deployers. It specifies that high-risk AI systems must be designed to be transparent, where providers of high-risk AI systems must design systems to allow for traceability, ensuring that their operation and outputs are explainable to relevant stakeholders.
- Article 14 emphasizes human oversight and requires systems to provide meaningful information to enable users to understand and intervene appropriately.
- Article 15 addresses accuracy, robustness, and cybersecurity requirements, stating that high-risk AI systems must be designed to be accurate, robust, and secure.

Questions

- Which specific requirements, obligations, or guidelines in the EU AI Act do you think will have the most significant impact on explainable AI practices?
- From your perspective, how does explainability contribute to the EU AI Act's goals of transparency and accountability?
- Are there parts of the EU AI Act's explainability requirements that you find unclear or difficult to interpret? If so, which ones?
- What are the limitations of current explainability tools in meeting oversight expectations? Are current explainability tools sufficient?
- What challenges do you think recent advances in AI systems, such as Large Language Models (LLMs) like ChatGPT, pose to explainability? How might these challenges impact compliance with regulations like the EU AI Act?
- What challenges do you think ensuring explainability presents for AI systems like Large Language Models (LLMs) in meeting the EU AI Act's requirements for transparency and oversight?

Challenges in Implementation

- What challenges have you faced (or foresee) in implementing explainability under the EU AI Act?
- Do you think explainability requirements conflict with other priorities, such as innovation, intellectual property, or system performance? How can this tradeoff be balanced?

Success Factors in Implementation

- What solutions or success factors do you think are critical for addressing these challenges and successfully implementing explainability under the EU AI Act?
- Can you provide examples of strategies, tools, or frameworks that organizations have used effectively to mitigate explainability challenges?

Best Practices

- Have you seen examples of successful explainability practices in your field? What made them successful?

Future Trends

- What trends do you see shaping explainability in the next few years? Do you think they will simplify compliance with the EU AI Act?

Closing Questions

- If you could improve one aspect of the EU AI Act's explainability provisions, what would it be?
- Are there areas of explainability research you think are underexplored but critical for compliance?
- Is there anyone in your professional network you could recommend who might also provide valuable insights for this research?
- Is there anything else you want to share about explainability or the EU AI Act that we haven't covered?