A Requirements Engineering Perspective on Explainability Privacy Trade-offs in Decision Support Systems

Andressa Girotto Vargas[1][0009−0008−8192−2061] and Edna Dias Canedo[1,2][0000−0002−2159−339X]

University of Brasília (UnB), Brasília–DF, Brazil
[2]Federal University of Pernambuco (UFPE), Computer Science Center (CIn), Recife, PE, Brazil
andressagirottovargas@gmail.com,ednacanedo@cin.ufpe.br

Abstract. Context: The growing adoption of Artificial Intelligence (AI) in decision support systems operating on sensitive data has intensified demands for transparency, accountability, and privacy protection. In these contexts, explainability and privacy preservation often emerge as competing non-functional requirements. While explainable AI techniques aim to improve transparency and justification of automated decisions, privacy-preserving mechanisms restrict information disclosure to protect sensitive data, creating inherent tensions that remain insufficiently understood from a Requirements Engineering perspective. **Goal:** This study investigates how explainability and privacy requirements interact in AI-based decision support systems, focusing on identifying compatibility conditions, trade-offs, and stakeholder expectations. **Method:** We adopted a qualitative approach combining a structured literature analysis with an exploratory focus group involving eight stakeholders from technical, managerial, and governance roles. **Results:** The findings indicate that explainability and privacy-preserving techniques are compatible only under context-dependent conditions. Participants perceived global and aggregated explanations as more appropriate for privacy-constrained environments, whereas local explanations raised higher risks of information disclosure. Stakeholders also emphasized that explanations should be role-appropriate rather than maximally transparent. **Conclusion:** By conceptualizing explainability and privacy as competing non-functional requirements, this study provides a stakeholder-centered perspective that supports requirements elicitation, analysis, and negotiation in the design of trustworthy and privacy-preserving decision support systems.

Keywords: Requirements Engineering, Explainable Artificial Intelligence, Privacy-Preserving Systems, Non-Functional Requirements, Decision Support Systems.

1 Introduction

The increasing adoption of Artificial Intelligence (AI) in information systems has intensified demands for transparency, accountability, privacy protection and

Proceedings of the 29th Workshop on Requirements Engineering (WER2026), August 11-13, 2026, La Plata, Argentina.
https://doi.org/10.29327/1884517.29-9

trustworthiness, particularly in decision support scenarios involving sensitive data [5, 25]. In many domains, automated decisions influence high-impact contexts such as healthcare, finance, and public administration, where stakeholders require not only accurate predictions but also understandable and justifiable system behavior.

These expectations are reinforced by regulatory frameworks that impose obligations related to transparency and accountability in automated decision-making. Prominent examples include the General Data Protection Regulation (GDPR) [8], the Brazilian General Data Protection Law (LGPD) [6], and the European AI Act [20], which introduce requirements for transparency, documentation, and, in some contexts, explanations of automated decisions [15]. Beyond binding legislation, international governance frameworks such as the UNESCO Recommendations on Artificial Intelligence [22], the OECD Principles on AI [13], and the NIST AI Risk Management Framework [12] further emphasize explainability as a key principle for trustworthy AI systems.

From a system design perspective, these expectations introduce important implications for Requirements Engineering. Explainability, privacy protection, accountability, and trust can be understood as *non-functional requirements* that shape the behavior and governance of AI-based systems. However, these requirements may conflict in practice. While Explainable Artificial Intelligence (XAI) techniques aim to reveal information about model behavior to support transparency and auditing [5], many ML-based systems operate under strict privacy constraints and rely on privacy-preserving mechanisms such as differential privacy, federated learning, secure computation, and data anonymization [2]. These techniques intentionally restrict information disclosure, creating a fundamental tension between transparency and privacy.

Recent research shows that this tension is not merely conceptual. Explanations themselves may expose sensitive information about training data or model behavior, potentially enabling attacks such as membership inference or model extraction [3, 9]. At the same time, explainability has been widely recognized as a key factor influencing trust and adoption in decision support systems [18]. Empirical evidence suggests that explanations can improve users' confidence in ML-based decisions, although their effectiveness depends on context, explanation type, and stakeholder expertise [25]. In privacy-sensitive domains, such as healthcare and public administration, explainability is closely tied to accountability and responsibility, yet excessive transparency may conflict with data protection requirements [17].

From a Requirements Engineering perspective, these tensions reflect the broader challenge of eliciting and negotiating competing non-functional requirements. This requires making trade-offs explicit during requirements elicitation, analysis, and specification. Different stakeholders including developers, managers, privacy professionals, and decision-makers may require different forms and levels of explanation depending on their responsibilities and risk exposure. Understanding how these perspectives influence explainability and privacy requirements is therefore essential to support informed design decisions. Motivated

by these challenges, this paper investigates the interplay between explainability and privacy-preserving mechanisms in AI-based decision support systems. The study explores how stakeholders perceive these requirements and how trade-offs emerge during system design. The following research questions guide this work:

RQ1: What explainability requirements are compatible with privacy-preserving requirements in AI-based decision support systems, and what trade-offs emerge between them?
RQ2: How do different stakeholders perceive trust, accountability, privacy and explanation needs, and how can these perspectives inform requirements elicitation and negotiation?

To address these questions, we combine a structured literature analysis with an exploratory focus group involving stakeholders from technical, managerial, and governance roles. This approach integrates conceptual insights from prior research with empirical perspectives from practitioners involved in the design and operation of software-intensive systems. Accordingly, the study examines how explainability and privacy can be captured, negotiated, and documented as interacting requirements in decision support systems.

This paper makes the following contributions: 1) A synthesis of representative literature on XAI in privacy-preserving environments, structured across explanation mechanisms, privacy techniques, trade-offs, and stakeholder perspectives (Table 1); 2) Empirical insights from a multi-stakeholder focus group that reveal how practitioners perceive compatibility constraints and trade-offs between explainability and privacy; and 3) A stakeholder-centered perspective that supports the elicitation and negotiation of explainability and privacy requirements in AI-based decision support systems.

2 Background and Related Work

Privacy-preserving machine learning techniques such as Differential Privacy (DP), Federated Learning (FL), homomorphic encryption, secure computation, and anonymization aim to minimize information leakage while enabling data-driven decision-making. In contrast, XAI techniques intentionally expose information about model behavior to support transparency and accountability. This creates an inherent tension that has been increasingly recognized in recent literature.

Several studies demonstrate that explanations may inadvertently amplify privacy risks by revealing sensitive patterns, feature contributions, or decision boundaries [3, 19, 21]. In particular, *post-hoc* explanation techniques such as feature importance scores and counterfactual explanations can facilitate privacy attacks, including membership inference, model extraction, and inversion attacks when deployed without safeguards [9]. These findings highlight that explainability mechanisms must be carefully designed when applied in privacy-sensitive environments.

Recent work has also explored the practical integration of XAI with privacy-preserving techniques. For example, Senevirathna et al. [16] propose a model-agnostic development process that integrates privacy, accountability, and resilience objectives alongside model utility. Their results show that widely used XAI techniques such as SHAP, LIME, and Layer-wise Relevance Propagation (LRP) remain applicable in privacy-preserving settings such as DP and FL, but introduce measurable trade-offs. Strong privacy guarantees may reduce predictive utility and explanation fidelity, while explanations themselves can still reveal abnormal model behavior through attribution analysis.

Complementary studies investigate privacy-aware explainability under formal constraints, showing that stronger privacy protection may significantly degrade explanation precision [4, 1]. Overall, existing evidence suggests that combining XAI mechanisms with privacy-preserving techniques is feasible but inherently conditional on context-specific trade-offs. Factors such as explanation granularity, privacy budget, and system architecture influence how these techniques interact. However, existing studies remain fragmented and often focus on isolated technical aspects rather than integrated system design decisions.

Beyond technical compatibility, explainability plays an important socio-technical role in fostering trust and accountability in decision support systems. Prior research shows that explanations can increase users' confidence in automated decisions, although the effect depends on factors such as explanation type, task complexity, and user expertise [25]. In privacy-sensitive domains such as healthcare, finance, and public administration, explainability is closely linked to accountability rather than simple interpretability. Decision support systems must provide explanations that enable auditing, contestation, and justification of decisions, while still respecting data protection constraints [17, 14]. This creates a delicate balance between transparency and privacy protection.

From an organizational perspective, explainability and privacy should not be maximized independently but balanced strategically. Managerial oriented studies describe privacy-aware explainability as a multi-level construct in which different stakeholders require different levels of transparency to support trust, accountability, and regulatory compliance [10]. Despite these advances, the literature still lacks a consolidated understanding of how explainability levels influence stakeholders' perceptions of trust and accountability in privacy-preserving decision support systems. Many studies focus either on technical robustness or user trust, rarely integrating these perspectives. To consolidate the discussion, Table 1 synthesizes representative studies according to four dimensions: XAI mechanisms, privacy-preserving techniques, identified trade-offs, and targeted stakeholders. The synthesis highlights that existing work often emphasizes either technical compatibility or trust-related outcomes, but rarely examines both dimensions simultaneously.

The synthesis shows that existing studies address explainability and privacy from security-oriented, governance-oriented, or user-centered perspectives, but rarely in an integrated manner. While several works demonstrate the feasibility of combining XAI with privacy-preserving techniques, they consistently re-

Table 1. Synthesis of prior work on XAI and privacy-preserving systems

Study	XAI Mechanisms	Privacy-Preserving Techniques	Key Trade-offs Identified	Primary Stakeholders
[9]	Feature importance, counterfactual explanations	Implicit privacy constraints, exposure analysis	Explanations may enable membership inference, model extraction, and inversion attacks; increased transparency can amplify privacy leakage	Service providers, auditors
[16]	SHAP, LIME, LRP; accountability metrics (e.g., stability/consistency)	Federated learning, access control, blockchain audit trails	Improved accountability and auditability at the cost of increased system complexity and governance overhead	System operators, compliance officers
[25]	Model-agnostic and model-specific explanations	Not explicitly addressed	Explainability generally increases trust, but effects depend on explanation type, user expertise, and context; privacy aspects remain underexplored	End-users, decision-makers
[17, 14]	Post-hoc explanations for decision support	Data minimization, consent, regulatory safeguards	Need to balance meaningful explanations with strict privacy and confidentiality requirements	Clinicians, patients, regulators
[10]	Layered and role-aware explanations	Privacy-by-design strategies, access control	Explainability must be tailored to stakeholder roles; organizational and governance trade-offs dominate purely technical considerations	Managers, compliance officers
[24]	Not the primary focus	Differential Privacy	Improved privacy leads to reductions in accuracy and fairness, revealing unavoidable multi-dimensional trade-offs	ML practitioners, policy makers
[23]	Explainability cues, transparency mechanisms	Privacy and security controls	Maximizing transparency, privacy, and efficiency simultaneously is infeasible; design requires explicit prioritization	Managers, system designers
[7]	Confidence scores, uncertainty communication	Applicable in privacy-sensitive settings without exposing raw data	Confidence cues may worsen trust calibration and appropriate reliance under knowledge mismatch	End-users, decision-makers

port trade-offs affecting model utility, explanation fidelity, and system complexity. In addition, stakeholders' trust and accountability requirements are often considered only implicitly or for specific audiences. These gaps motivate the research questions investigated in this paper, which aim to analyze both the technical compatibility and stakeholder-level implications of explainability in privacy-preserving decision support systems.

3 Research Method

This study adopts a qualitative multi-method research design that combines a structured literature analysis with an exploratory focus group study [11]. The goal is to investigate explainability privacy trade-offs from both a conceptual and practitioner-oriented perspective, addressing RQ1 and RQ2. Qualitative approaches are widely recommended in empirical software engineering and information systems research when exploring complex socio-technical phenomena and stakeholder perceptions [11]. The adopted design integrates insights from the literature with practitioner perspectives, enabling both conceptual synthesis and empirical grounding. The analysis was oriented toward identifying requirements interactions, stakeholder concerns, and trade-offs relevant to Requirements Engineering.

Figure 1 summarizes the research process. The study follows a sequential design consisting of two main phases: (i) a structured analysis of representative literature on XAI and privacy-preserving techniques, and (ii) an exploratory focus group with stakeholders involved in AI-enabled decision support systems. The integration of both phases enables the identification of explainability–privacy trade-offs and their implications for trust and accountability.

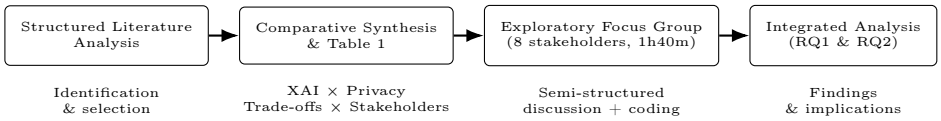

Fig. 1. Overview of the research method.

The first phase consisted of a structured analysis of representative literature on Explainable Artificial Intelligence (XAI) and privacy-preserving techniques. Rather than performing a full systematic literature review, the goal was to identify influential studies discussing technical, organizational, and governance implications of explainability under privacy constraints. Relevant studies were identified through targeted searches in bibliographic sources such as DBLP and Google Scholar using combinations of keywords related to explainability (e.g., "Explainable AI", "XAI", "interpretable machine learning") and privacy (e.g., "privacy-preserving", "differential privacy", "federated learning", "data protection"). The selection prioritized peer-reviewed publications, recent surveys, and conceptual works published from 2018 onwards addressing explainability in privacy-sensitive contexts.

For each selected study, information was extracted along four analytical dimensions: (i) XAI mechanisms, (ii) privacy-preserving techniques, (iii) reported trade-offs between explainability, privacy, and system utility, and (iv) stakeholder groups addressed. These dimensions guided the synthesis presented in Section 2 and Table 1. To complement the conceptual synthesis, an exploratory focus group was conducted to capture practitioner perspectives on explainability and privacy trade-offs. The synthesis was used as input to plan the focus group (invited roles, guiding questions, and themes), allowing for an integrated analysis directly connected to the QPs. Focus groups are widely used in empirical software engineering to elicit expert insights and collective reflections on emerging socio-technical challenges [11]. The focus group included eight participants representing different stakeholder roles associated with privacy-sensitive AI systems. Participants were selected through purposive sampling based on their professional experience with AI systems, data protection, or organizational decision-making. Recruitment was conducted through the authors' professional networks, a common practice in exploratory qualitative studies [11].

The composition of the group ensured diversity of perspectives, including technical, managerial, and governance-oriented roles, enabling the discussion of explainability requirements across different stakeholder viewpoints. A single focus group session was conducted with a duration of approximately one hour and forty minutes. The session followed a semi-structured format, allowing guided discussion while enabling participants to elaborate on their experiences and perspectives.

The discussion was organized around thematic blocks aligned with the research questions, including explainability under privacy constraints, transparency versus data protection trade-offs, stakeholder-specific explanation needs, and im-

plications for trust and accountability. A moderator facilitated the discussion, ensuring balanced participation and avoiding influence on participants' opinions, following methodological recommendations for focus group studies [11].

The session was audio-recorded with participants' consent and subsequently transcribed for analysis. The analysis followed an issue-based qualitative approach, where statements were coded and grouped into themes related to explainability mechanisms, privacy concerns, trust, and accountability. Coding was conducted iteratively. Initial open codes were assigned to meaningful excerpts and later consolidated into higher-level themes aligned with RQ1 and RQ2. Divergent interpretations were discussed among the authors until consensus was reached. The goal was to capture convergent and divergent perspectives across stakeholder roles rather than quantify responses. Participation was voluntary and only anonymized data were reported; therefore, no formal ethics committee approval was required.

4 Results

The focus group involved eight participants selected through purposive sampling based on their experience with software-intensive systems, artificial intelligence, data protection, or organizational decision-making (Table 2). Participants are identified using anonymized labels (P1–P8). The group was intentionally composed to represent diverse stakeholder perspectives relevant to privacy-preserving decision support systems. Participants had an average of approximately seven years of professional experience. Their ages ranged from 32 to 48 years, and their educational backgrounds included undergraduate degrees, master's degrees, and current master's students. This diversity supported the discussion of both technical and organizational concerns while ensuring a shared baseline understanding of the topic.

Table 2. Profile of focus group participants

ID	Role	Age (years)	Experience	Educational Level
P1	Developer	32	6 years	Undergraduate degree
P2	ML Engineer	34	7 years	Master's student
P3	Project Manager	38	6 years	Master's student
P4	Product Owner	41	9 years	Master's degree
P5	Privacy Professional	36	7 years	Master's degree
P6	Compliance Professional	39	6 years	Master's degree
P7	Decision-maker	45	8 years	Master's degree
P8	Decision-maker	48	10 years	Master's degree

The focus group was guided by open-ended questions designed to elicit participants' perceptions of explainability in privacy-preserving systems. The discussion focused on three main aspects: (i) what constitutes an adequate explanation under privacy constraints, (ii) which trade-offs emerge when explainability increases, and (iii) how different stakeholders interpret explanations in terms of trust and accountability. Table 3 summarizes the guiding questions used to structure the session.

Table 3. Guiding questions used in the focus group

Theme	Guiding Question
Explainability	What does an adequate explanation mean in systems that operate under strict privacy constraints?
Trade-offs	In which situations can increasing explainability pose risks to privacy or data protection?
Stakeholder needs	Do different stakeholders require different levels or types of explanations? Why?
Trust	What types of explanations increase or decrease your trust in privacy-preserving decision support systems?
Accountability	What information or explanations are necessary to support accountability and decision justification in such systems?
Design implications	How should organizations decide the appropriate level of explainability when privacy constraints are present?

4.1 RQ1. What explainability needs emerge in privacy-preserving decision support systems, and what trade-offs constrain them?

The focus group results indicate that explainability is perceived as necessary even in privacy-preserving decision support systems, but not as unrestricted transparency. Participants consistently argued that explanations must be useful enough to support understanding and action, while remaining bounded by privacy constraints. In this sense, explainability was framed as a context-dependent requirement shaped by technical feasibility, sensitivity of the data, and stakeholder role. This characterization is consistent with treating explainability as a non-functional requirement that must be negotiated against privacy constraints.

A first recurrent finding concerns the **conditional compatibility** between explainability and privacy-preserving techniques. Participants agreed that explainability and privacy are not mutually exclusive, but compatibility depends on the granularity of the explanation. Global, aggregated, or model-level explanations were perceived as more feasible in privacy-sensitive settings than local or instance-level explanations, since they reduce the risk of exposing information about specific records or sensitive features.

> *"You can still explain how the model behaves overall, even in a federated or privacy-preserving setting, as long as you avoid exposing individual data points or specific feature values." (P2, ML Engineer)*

At the same time, participants highlighted that more detailed explanations, although potentially useful for debugging and operational decision-making, may become privacy risks. In particular, local explanations and counterfactuals were perceived as more difficult to reconcile with regulated or sensitive environments.

> *"Local explanations can be very useful, but they also feel dangerous in sensitive systems, because you may end up revealing more about the data than you initially intended." (P5, Privacy Professional)*

A second major finding refers to the **trade-offs between explainability, privacy, and utility**. Participants repeatedly stated that stronger privacy guarantees tend to reduce explanation fidelity, actionability, and system utility. Explanations become more generic, which may preserve confidentiality but also limit their usefulness for investigation, debugging, or operational justification.

> *"When we apply stronger privacy mechanisms, the explanations become more generic and sometimes less actionable, which limits their usefulness for debugging or decision support." (P1, Developer)*

This trade-off was not interpreted as a purely technical limitation. Rather, participants described it as a **design and governance decision**, in which teams must balance legal risk, stakeholder expectations, and operational needs. In this sense, explainability was viewed less as a fixed property and more as a negotiated requirement.

> *"Deciding how much to explain is ultimately a governance decision. We have to weigh transparency against legal risk, performance, and the expectations of different stakeholders." (P4, Product Owner)*

A third result concerns **technical and organizational constraints**. Participants mentioned increased computational overhead, lower model performance under strong privacy guarantees, reduced debuggability, and the lack of clear organizational guidance on how to operationalize explainability under privacy constraints. This lack of standardization was perceived as especially problematic in regulated settings.

> *"There is no clear reference on how much explanation is enough in privacy-sensitive systems, so teams end up making isolated decisions that are difficult to justify later." (P7, Decision-maker)*

The results suggest that explainability in privacy-preserving decision support systems should be understood as a constrained requirement, shaped by trade-offs among privacy protection, explanation usefulness, and system performance.

RQ1 Summary

Privacy-preserving decision support systems still require explainability, but in bounded and context-dependent forms. Participants perceived global and aggregated explanations as more compatible with privacy constraints than local or instance-level explanations. Stronger privacy guarantees tend to reduce explanation fidelity and usefulness, revealing unavoidable trade-offs between privacy, explainability, and utility. These trade-offs were interpreted not only as technical constraints, but also as governance and design decisions that require explicit prioritization and justification.

4.2 RQ2. How do explainability privacy trade-offs affect stakeholders' trust and accountability requirements?

The findings related to RQ2 indicate that explainability affects trust and accountability in different ways depending on stakeholder role. Participants did not equate better explanations with maximal transparency. Instead, they emphasized that explanations must be appropriate to the stakeholder's responsibilities, level of expertise, and exposure to risk. Regarding **trust**, participants associated higher trust with explanations that were understandable, role-appropriate, and sufficient for practical sense-making. In privacy-sensitive environments, trust was not linked to seeing all internal details of the model, but rather to receiving an explanation that supports a credible understanding of the decision.

> *"I don't need to see all the data or the internal details of the model. What builds trust for me is understanding the logic behind the decision in a way that makes sense for my role." (P8, Decision-maker)*

Several participants also noted that too much detail may reduce trust instead of improving it, either because it increases cognitive burden or because it raises concerns about possible privacy exposure.

> *"When explanations become too detailed, especially in systems handling sensitive data, it actually makes me more uncomfortable than confident." (P5, Privacy Professional)*

With respect to **accountability**, participants stressed that explanations serve a different purpose. In this case, the main concern was not simply understanding the decision, but being able to justify it, document it, and defend it over time in audits, disputes, or compliance processes. Explanations therefore need to be stable, traceable, and defensible.

> *"In audits or legal discussions, we are not asked how the model works internally, but why a specific decision was made and whether it can be justified." (P4, Product Owner)*

> *"For accountability, explanations need to be consistent and defensible over time, not just intuitive at the moment they are presented." (P6, Compliance Professional)*

A third recurrent theme was the **stakeholder-dependent nature of explainability requirements**. Participants agreed that developers, managers, compliance professionals, and decision-makers need different forms and levels of explanation. As a result, a single explanation strategy was perceived as insufficient for simultaneously satisfying trust and accountability across all stakeholder groups.

> *"What makes sense for a developer is very different from what a manager or an end user needs. Trying to force a single type of explanation usually satisfies no one." (P2, ML Engineer)*

Taken together, these findings suggest that explainability in privacy-preserving systems should be designed as a layered and role-aware capability. The explainability level that supports trust for one stakeholder may not be the same level needed to support accountability for another.

RQ2 Summary

The findings indicate that explainability affects trust and accountability through stakeholder-specific requirements. Trust was associated with understandable and role-appropriate explanations rather than full transparency. Accountability, in turn, required explanations that were stable, traceable, and defensible over time. Participants consistently emphasized that different stakeholders require different explanation levels and formats, suggesting that privacy-preserving decision support systems should adopt layered and role-aware explanation strategies.

Discussion

The findings related to RQ1 indicate that explainability and privacy-preserving techniques are compatible only under context-dependent conditions. Both the literature and the focus group participants emphasized that global or aggregated explanations are generally more feasible in privacy-sensitive environments, whereas local or instance-level explanations introduce higher risks of information leakage. This observation aligns with prior studies showing that techniques such as differential privacy, federated learning, and secure computation constrain the granularity and fidelity of explanations [16]. Moreover, explanations themselves may act as attack surfaces when transparency is not properly governed [19, 21].

These results reinforce that explainability privacy trade-offs are structural rather than incidental. Increasing privacy guarantees often reduces model accuracy, fairness, or interpretability [24]. Participants also highlighted that these trade-offs are not purely technical but organizational decisions shaped by regulatory exposure, operational needs, and stakeholder expectations, consistent with socio-technical perspectives on trustworthy AI [10, 23]. From a Requirements Engineering perspective, explainability and privacy should therefore be treated as interacting non-functional requirements negotiated during system design. This implies the need to explicitly document rationale, priorities, and acceptable trade-offs during requirements analysis.

The results related to RQ2 indicate that explainability influences both trust and accountability in privacy-preserving decision support systems. Participants emphasized explanations that are understandable and aligned with stakeholders' roles rather than maximally detailed, supporting prior work highlighting the contextual nature of effective explainability [25]. They also distinguished between explanations supporting trust focused on sense-making and those supporting accountability, which must enable justification, auditing, and external scrutiny. Prior research shows that certain explanation cues, such as confidence scores, may not improve appropriate reliance and can even degrade decision quality [7]. The findings reinforce the stakeholder-dependent nature of explainability. Developers, managers, privacy professionals, and decision-makers require different forms and levels of explanation, supporting proposals for layered and role-aware explainability strategies [10]. From an accountability perspective, explanations must remain consistent, traceable, and defensible over time to support regulatory compliance and organizational learning [16].

6 Threats to Validity

This study is subject to limitations inherent to exploratory qualitative research [26]. **Construct validity** may be affected by the interpretation of abstract concepts such as explainability, trust, and accountability. To mitigate this risk, the focus group questions were grounded in prior literature and discussed with participants from different stakeholder roles. **Internal validity** may be influenced by researcher bias in moderation and analysis. This risk was reduced through

the use of a semi-structured protocol and iterative coding focused on recurring themes rather than isolated opinions. **External validity** is limited by the small number of participants and the use of purposive sampling, which restrict broad generalization. However, the study aims at analytical rather than statistical generalization, relating empirical findings to existing literature.

7 Conclusions

This paper investigated explainability in privacy-preserving decision support systems from a requirements-oriented perspective, combining a structured literature analysis with an exploratory focus group. The results show that explainability and privacy are not inherently incompatible, but their integration involves unavoidable trade-offs. Explanations considered more useful for operational understanding, such as local and instance-level ones, were also perceived as more privacy-sensitive, whereas global and aggregated explanations were seen as safer but less actionable. The study also showed that explainability requirements vary across stakeholders. Trust was associated with role-appropriate and understandable explanations, while accountability depended on explanations that are consistent and defensible over time. These findings suggest that explainability in privacy-preserving systems should be designed as a layered and stakeholder-aware requirement, rather than as a single uniform feature.

The main contribution of this work is to frame explainability as a socio-technical and governance concern in requirements engineering, not only as a model-level property. For practice, the results suggest the need to elicit explainability requirements together with privacy constraints, documenting explicit trade-offs and stakeholder-specific needs. Future work may extend this study through additional empirical investigations in other domains and by proposing a requirements-oriented framework to guide the specification of explainability in privacy-preserving systems.

8 Acknowledgements

This study was financed in part by the National Council for Scientific and Technological Development (CNPq), Grants No. 300883/2025-0 and No. 406266/2025-5.

References

1. Abbasi, W., Mori, P., Saracino, A.: The explainability-privacy-utility trade-off for machine learning-based tabular data analysis. In: Proceedings of the 20th International Conference on Security and Cryptography, SECRYPT 2023, Rome, Italy, July 10-12, 2023. pp. 511–519. SCITEPRESS (2023), https://doi.org/10.5220/0012137800003555

2. Adnan, M., Syed, M.H., Anjum, A., Rehman, S.: A framework for privacy-preserving in iov using federated learning with differential privacy. IEEE Access **13**, 13507–13521 (2025), https://doi.org/10.1109/ACCESS.2025.3526934
3. Allana, S., Kankanhalli, M., Dara, R.: Privacy risks and preservation methods in explainable artificial intelligence: A scoping review. CoRR **abs/2505.02828** (2025), https://doi.org/10.48550/arXiv.2505.02828
4. Allana, S., Kankanhalli, M., Dara, R.: Privacy risks and preservation methods in explainable artificial intelligence: A scoping review. Trans. Mach. Learn. Res. **2025** (2025), https://openreview.net/forum?id=q9nykJfzku
5. Arrieta, A.B., Rodríguez, N.D., Ser, J.D., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F.: Explainable artificial intelligence (XAI): concepts, taxonomies, pportunities and challenges toward responsible AI. Inf. Fusion **58**, 82–115 (2020), https://doi.org/10.1016/j.inffus.2019.12.012
6. Brazil: Brazilian general data protection law (LGPD) (2018), https://www.gov.br/anpd/pt-br/centrais-de-conteudo/outros-documentos-e-publicacoes-institucionais/lgpd-en-lei-no-13-709-capa.pdf/@@display-file/file
7. Cabitza, F., Fregosi, C., Vicente, L.: Too sure for trust. the paradoxical effect of calibrated confidence in case of uncalibrated trust in hybrid decision making. In: World Conference on Explainable Artificial Intelligence. pp. 233–254. Springer (2025). https://doi.org/10.1007/978-3-032-08317-3_11
8. European Parliament and The Council: General Data Protection Regulation (GDPR): EU Data Protection Rules (2018), https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679
9. Ezzeddine, F.: Privacy implications of explainable AI in data-driven systems. In: Joint Proceedings of the xAI 2024 Late-breaking Work, Demos and Doctoral Consortium co-located with the 2nd World Conference on eXplainable Artificial Intelligence (xAI-2024), Valletta, Malta, July 17-19, 2024. CEUR Workshop Proceedings, vol. 3793, pp. 361–368. CEUR-WS.org (2024), https://ceur-ws.org/Vol-3793/paper_46.pdf
10. Keaney, S., Berthon, P.: Balancing explainability and privacy in AI systems: A strategic imperative for managers. Business Horizons (2025). https://doi.org/https://doi.org/10.1016/j.bushor.2025.10.004
11. Kontio, J., Lehtola, L., Bragge, J.: Using the focus group method in software engineering: Obtaining practitioner and user experiences. In: 2004 International Symposium on Empirical Software Engineering (ISESE 2004), 19-20 August 2004, Redondo Beach, CA, USA. pp. 271–280. IEEE Computer Society (2004), https://doi.org/10.1109/ISESE.2004.35
12. National Institute of Standards and Technology: Artificial intelligence risk management framework: Generative artificial intelligence profile. NIST Trustworthy and Responsible AI NIST AI 600-1 pp. 1–64 (2023), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf
13. OECD: Recommendation of the council on artificial intelligence. OECD: Paris, France pp. 1–12 (2024), https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449, originally adopted on 22 May 2019; amended on 3 May 2024
14. Radanliev, P.: AI ethics: Integrating transparency, fairness, and privacy in AI development. Appl. Artif. Intell. **39**(1) (2025), https://doi.org/10.1080/08839514.2025.2463722

15. Rocha, L.D., Canedo, E.D.: Optimizing compliance: Comparative study of data laws and privacy frameworks. J. Internet Serv. Appl. **16**(1), 431–452 (2025), https://doi.org/10.5753/jisa.2025.5247
16. Senevirathna, T., Sandeepa, C., Siniarski, B., Nguyen, M., Marchal, S., Boerger, M., Liyanage, M., Wang, S.: Enhancing accountability, resilience, and privacy of intelligent networks with XAI. IEEE Open J. Commun. Soc. **6**, 8389–8409 (2025), https://doi.org/10.1109/OJCOMS.2025.3608784
17. Shah, B., Gupta, E.V.: Ethics and privacy in AI-driven healthcare decision support systems. International Journal of Research and Analytical Reviews (IJRAR) **12**, 1–11 (2025)
18. Shin, D.: The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. Int. J. Hum. Comput. Stud. **146**, 102551 (2021), https://doi.org/10.1016/j.ijhcs.2020.102551
19. Shokri, R., Strobel, M., Zick, Y.: On the privacy risks of model explanations. In: AIES '21: AAAI/ACM Conference on AI, Ethics, and Society, Virtual Event, USA, May 19-21, 2021. pp. 231–241. ACM (2021), https://doi.org/10.1145/3461702.3462533
20. Sonsini, W., Parliament, T.E.: The eu artificial intelligence act. European Union pp. 1–10 (2024), https://www.wsgr.com/a/web/qrkz1SnNzWw6nk7B3oAyDa/10-things-you-should-know-about-the-eu-artificial-intelligence-act_v2.pdf
21. Spartalis, C.N., Semertzidis, T., Daras, P.: Balancing XAI with privacy and security considerations. In: Computer Security. ESORICS 2023 International Workshops - CPS4CIP, ADIoT, SecAssure, WASP, TAURIN, PriST-AI, and SECAI, The Hague, The Netherlands, September 25-29, 2023, Revised Selected Papers, Part II. Lecture Notes in Computer Science, vol. 14399, pp. 111–124. Springer (2023), https://doi.org/10.1007/978-3-031-54129-2_7
22. United Nations Educational, Scientific and Cultural Organization (UNESCO): Recommendation on the ethics of artificial intelligence. UNESCO:Paris, France pp. 1–44 (2022), https://unesdoc.unesco.org/ark:/48223/pf0000381137.locale=en
23. Wang, Y.: Balancing trustworthiness and efficiency in artificial intelligence systems: An analysis of trade-offs and strategies. IEEE Internet Computing **27**(6), 8–12 (2023). https://doi.org/10.1109/MIC.2023.3303031
24. Wasif, D., Chen, D., Madabushi, S., Alluru, N., Moore, T.J., Cho, J.H.: Empirical analysis of privacy-fairness-accuracy trade-offs in federated learning: A step towards responsible ai. arXiv preprint arXiv:2503.16233 (2025), https://arxiv.org/abs/2503.16233
25. Wiratsin, I., Ragkhitwetsagul, C.: Effectiveness of explainable artificial intelligence (XAI) techniques for improving human trust in machine learning models: A systematic literature review. IEEE Access **13**, 121326–121350 (2025), https://doi.org/10.1109/ACCESS.2025.3575022
26. Wohlin, C., Runeson, P., Höst, M., Ohlsson, M.C., Regnell, B., Wesslén, A., et al.: Experimentation in software engineering, vol. 236. Springer (2012)