LAUTECH Journal of Computing and Informatics (LAUJCI) – ISBN : 2714-4194
Volume 5 No.1, April 2026 – www.laujci.lautech.edu.ng

EXPLAINABILITY FATIGUE IN ARTIFICIAL INTELLIGENCE: A PRISMA-GUIDED CONCEPTUAL FRAMEWORK OF COGNITIVE LIMITS IN HUMAN–AI INTERACTION

Usman Dahiru Hassan
Department of Data Science & Artificial Intelligence, Modibbo Adama University, Yola, Nigeria
dahiruusman222@mau.edu.ng.

Manunyi Godfrey
Department of Computer Science, Modibbo Adama University
Yola, Nigeria
manugodfrey02@gmail.com.

ABSTRACT
Explainable Artificial Intelligence (XAI) is increasingly recognized as essential for developing responsible and trustworthy AI, predicated on the assumption that greater transparency enhances user understanding, trust, and decision-making. Despite extensive, rigors research in explainable artificial intelligence (XAI), the integration of AI into high-stakes domains remains constrained by concerns over interpretability and trust. Current technical solutions often neglect human cognitive frameworks for interpreting complex decisions, leading to a phenomenon termed "explainability fatigue" where cognitive effort required to comprehend AI explanations outweighs perceived benefits, resulting in diminished engagement and suboptimal reliance on AI systems. This research employed a PRISMA 2020-guided conceptual systematic review to synthesize theoretical and empirical work on XAI and human cognitive constraints. Following identification, screening, eligibility, and inclusion phases, searches across major databases (IEEE Xplore, Scopus, Web of Science, ACM Digital Library, Google Scholar) yielded 32 studies from 2021-2026. Studies were systematically coded and grouped into thematic domains: cognitive load in XAI, trust calibration, interpretability techniques, and human-centered design principles. Analysis revealed that explanation complexity increases extraneous cognitive load, leading to performance degradation rather than improvement. Three key outcomes emerged: trust miscalibration (both overtrust and undertrust), degraded decision quality through cognitive overload, and accountability gaps. The synthesis identified antecedent variables (explanation complexity, volume, user characteristics, contextual constraints) that mediate explainability fatigue. The paper proposes a framework that positions explainability fatigue as a mediating factor between explanation design and responsible AI outcomes, explainable AI systems should adopt adaptive, context-aware explanation strategies aligned with human cognitive capabilities rather than pursuing maximal transparency, marking a shift toward "cognitively sustainable transparency" in responsible AI design.

Key words: Explainable Artificial Intelligence, Explainability Fatigue, Cognitive Load Theory, Human–AI Interaction, Trust Calibration, Decision Quality.

1 INTRODUCTION

The rapid development of sophisticated artificial intelligence systems has necessitated advancements in explainable AI to foster transparency and user understanding (Sanneman & Shah, 2022). However, despite extensive research in Explainable AI methods, the integration of AI into high-stakes domains remains constrained by concerns over interpretability and trust, highlighting a critical gap in current approaches (Chameera et al., 2025). Specifically, current technical solutions often neglect the intricate cognitive frameworks humans employ to interpret complex decisions, leading to user distrust and a perception of opacity (Jean & Pera, 2025). This oversight often results in a cognitive burden on the decision-maker, potentially leading to explainability fatigue and hindering the effectiveness of human-AI collaboration (Boyacı et al., n.d.). While Explainable AI endeavors to mitigate issues such as automation bias by elucidating AI decision-making processes, the efficacy of such explanations is contingent upon their alignment with human cognitive capabilities and informational requirements (Romeo &

Conti, 2025). This framework models autonomy as a continuous stochastic process influenced by information-induced cognitive load, formalizing autonomy evolution as geometric Brownian motion with information-dependent drift (Margondai & Mouloua, 2026). Furthermore, the utility of XAI is not solely determined by the availability of information but also by the human capacity to process and integrate that information effectively under varying cognitive loads and time constraints, as users often default to automatic thinking in such scenarios (Doh, 2025). This cognitive overload, stemming from an abundance of explanations, can paradoxically impair decision-making and task performance, leading to a phenomenon termed "explainability fatigue" (Maehigashi et al., 2024). This phenomenon arises when the cognitive effort required to comprehend and integrate numerous AI explanations outweighs the perceived benefit, leading to diminished user engagement, reduced trust, and suboptimal reliance on AI systems (Bertrand et al., 2022).This "transparency paradox" suggests that while transparency is often advocated as universally beneficial, an excess of information can conversely impede rather than enhance human decision-making and interaction with AI systems (Margondai & Mouloua, 2026).

2. LITERATURE REVIEW

The burgeoning integration of Artificial Intelligence into diverse high-stakes domains necessitates a comprehensive understanding of human-AI collaboration dynamics, particularly concerning the cognitive impact of AI explainability . Methodology (Romeo & Conti, 2025) A significant challenge arises from the potential for users to either excessively trust or distrust AI recommendations, thereby compromising the efficacy of human-AI collaborative systems (Boyacı et al., n.d.). This balance between appropriate reliance and critical assessment is further complicated by the increasing complexity and opacity of advanced AI systems, especially generative AI, which can obscure the underlying decision-making processes (Spitzer et al., 2025).

This interpretability challenge is further exacerbated by the natural language interfaces prevalent in many contemporary AI systems, which can foster a perception of greater AI capability than is warranted, potentially increasing over-reliance (Kaur et al., 2024). Consequently, the unintelligibility of AI explanations, information overload, and contextual inaccuracies can impose significant cognitive burdens on users, negatively impacting task performance and decision-making quality (Maehigashi et al., 2024)

This cognitive burden can manifest as "explainability fatigue," where users become disengaged or overwhelmed by the effort required to process and evaluate AI explanations, especially under time constraints or high cognitive load (Doh, 2025). This phenomenon underscores the critical need for a deeper investigation into the cognitive limits of human–AI interaction, particularly how the design and delivery of AI explanations can mitigate or exacerbate this fatigue (Chiaburu et al., 2024)

The goal is to develop a conceptual framework, guided by PRISMA, which synthesizes existing research on cognitive load, trust dynamics, and human-computer interaction to articulate the underlying mechanisms of explainability fatigue and propose design principles for resilient human-AI systems (Neyigapula, 2023). Such a framework would aim to optimize the cognitive resources allocated by human operators during interaction with AI, mitigating the detrimental effects of excessive or poorly presented explanatory information (Wang et al., 2026). This necessitates a multidisciplinary approach, drawing insights from computer science, cognitive science, and human-computer interaction to devise strategies that align AI explanations with human cognitive capabilities and workflows (Chen et al., 2023)

3. METHODOLOGY

3.1 Research Design

This study adopts a PRISMA 2020–guided conceptual systematic review to synthesize theoretical and empirical work on explainable artificial intelligence (XAI) and human cognitive constraints. The PRISMA framework (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) provides a transparent, replicable procedure for identifying, screening, and selecting relevant literature, thereby strengthening methodological rigor and limiting selection bias (Page et al., 2021).

The review process follows the four core PRISMA phases: identification, screening, eligibility, and inclusion. In the identification phase, potentially relevant studies are retrieved from major academic databases (IEEE Xplore, Scopus, Web of Science, ACM Digital Library, and Google Scholar) using structured keyword strings (e.g., "explainable AI," "human–AI interaction," "cognitive load," "trust in AI," "explainability fatigue"). During screening, duplicates are removed and preliminary inclusion criteria are applied, including publication period (2021 - 2026), peer-reviewed status, and relevance to XAI or human cognitive processes. In the eligibility phase, full texts are assessed against more stringent criteria, such as conceptual relevance, methodological robustness, and contribution to theory or design implications. The inclusion phase yields a curated corpus of high-quality studies that form the basis for synthesis.

Given the conceptual orientation of the review, the analysis prioritizes thematic synthesis and theoretical integration rather than statistical meta-analysis. Included studies are systematically coded and grouped into key thematic domains: (i) cognitive load and mental workload in XAI; (ii) trust calibration and overreliance; (iii) interpretability techniques and user comprehension; and (iv) human-centered design principles for explainable systems. This process surfaces cross-cutting patterns, tensions, and gaps in the literature, enabling the construction of a refined conceptual framework that explains how different explanation strategies interact with human cognitive limitations.

The PRISMA-guided approach also supports integration of insights from adjacent disciplines, including cognitive psychology, human–computer interaction (HCI), and machine learning, thereby grounding XAI research in a genuinely multidisciplinary evidence base. The review thus offers not only a structured synthesis of existing knowledge but also novel theoretical propositions, particularly concerning explainability fatigue and the need for adaptive, context-aware explanation strategies.

PRISMA 2020–aligned conceptual systematic review design enhances methodological transparency, theoretical depth, and analytical coherence, positioning the study to advance scholarly understanding and inform the design of more effective, human-aligned XAI systems.

3.2 Inclusion and Exclusion Criteria

Inclusion Criteria:

This review includes peer-reviewed journal articles and top conference papers published from 2023 to 2025. The chosen studies focus on topics such as explainable AI (XAI), interpretability, cognitive load, trust in AI, and human–AI interaction. To be included, studies must provide either research data or a strong theory of how users understand and judge AI explanations.

Exclusion Criteria:

Some types of literature are excluded to keep the review focused and rigorous. Editorials, commentaries, or opinion pieces without research data or a strong theory are excluded. Studies that do not deal directly with how AI explanations are processed, or that are not in English, are also excluded.

3.3 Search Strategy

Searches were conducted across Scopus, Web of Science, IEEE Xplore, ACM Digital Library, ScienceDirect, and Google Scholar (supplementary recall). Search queries include important terms related to explainability, interpretability, cognitive load, and trust in AI systems. Search strings combined keywords using Boolean operators: ("Explainable AI" OR "XAI" OR interpretability) AND ("cognitive load" OR "human cognition") AND (trust OR transparency OR decision-making).

3.4 Data Sources

The studies in this review were identified using major academic databases, including Scopus, Web of Science, IEEE Xplore, ACM Digital Library, ScienceDirect, and Google Scholar. These sources were chosen to cover fields such as AI, human-computer interaction, cognitive psychology, and decision sciences. Only peer-reviewed articles and top conference papers were included to ensure quality.

3.5 Selection Process

The PRISMA 2020 framework was employed to guide the processes of identification, screening, eligibility assessment, and inclusion. An initial database search across Scopus, Web of Science, IEEE Xplore, ACM Digital Library, ScienceDirect, and Google Scholar yielded 61 records. An additional 29 relevant records were identified through supplementary searches and cross-referencing of retrieved articles, bringing the total to 90.

After removing duplicate entries and applying the publication-year filter (2021–2025), 70 records remained. Title and abstract screening were conducted to assess relevance to explainable artificial intelligence, cognitive load, human–AI interaction, trust calibration, and interpretability. Following this screening phase, 65 records were retained for further evaluation, while 15 were excluded for lack of topical alignment.

Full-text assessments were subsequently conducted on 50 articles to evaluate conceptual relevance, methodological rigour, and consistency with the study objectives. Nine studies were excluded at this stage due to insufficient theoretical contribution or lack of direct relevance to explanation processing in AI systems.

Ultimately, 41 studies met the inclusion criteria and were incorporated into the final qualitative synthesis. Figure 1 presents the PRISMA 2020 flow diagram illustrating the selection process, and Table 1 summarises the characteristics of the included studies.

PRISMA FLOW DIAGRAM

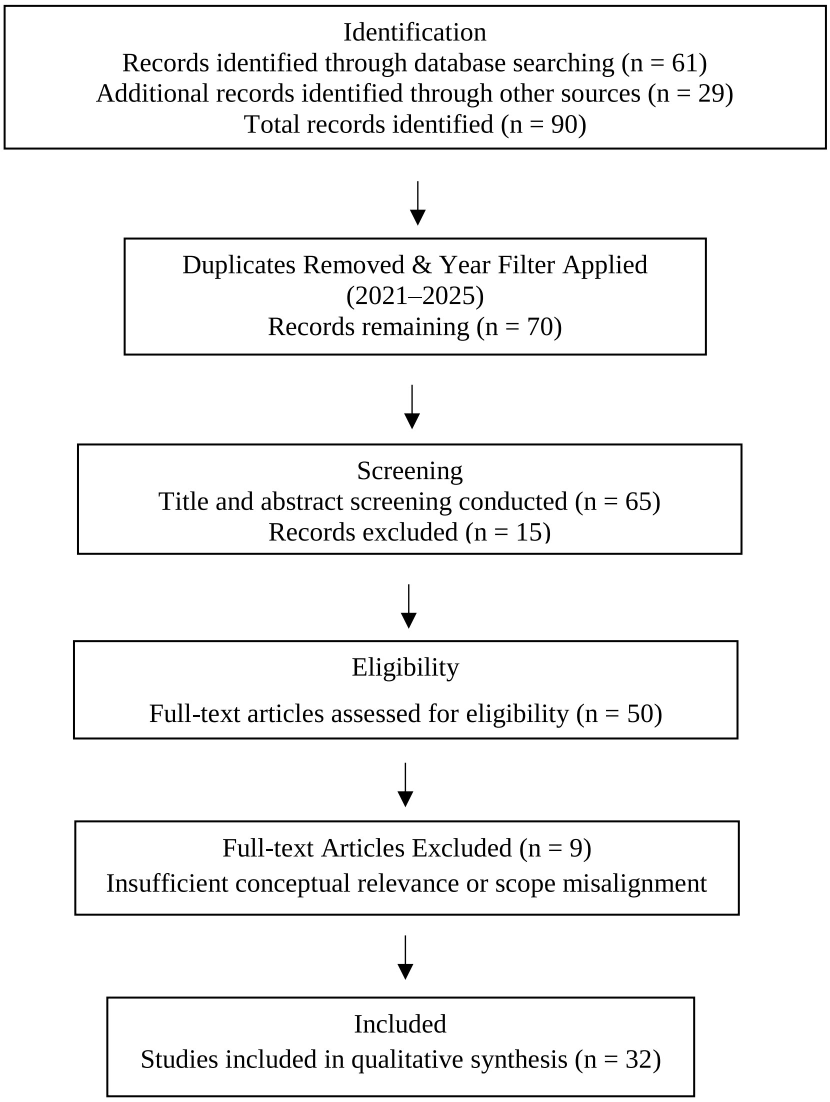

Figure 1 Prisma 2020 Flow Diagram Study Selection

3.6 Study Characteristics and Data Extraction

A structured data extraction process was conducted for all **32 included studies**. Extracted variables included publication year, study type, primary focus, methodological approach, and relevance to the explainability fatigue framework (antecedent, mediator, outcome, or governance dimension). Table 1 summarizes the characteristics of the included studies. A systematic process was used to collect data from all 32 studies, including the year of publication, study type, main research focus, methodological design, and their relationship to explainability fatigue (e.g., cause, mediator, outcome, or governance).

Table 1 summarizes the characteristics of the 32 selected studies, comprising approximately 14 empirical studies, 12 theoretical/conceptual papers, 5 review/survey studies, and 1 bibliometric/methodological study.

Table 1: Data Extraction Table of Included Studies (n = 32; 2021–2026)

S/N	Author(s) & Year	Study Type	Objective	Methodology	Key Variables	Key Findings	Relevance
1	Ahn et al. (2025)	Empirical	Study sociotechnical factors in XAI	Mixed-method	Cognitive, social factors	XAI effectiveness varies across users	Personalization need
2	Ali et al. (2023)	Review	XAI overview	Systematic review	Trust, transparency	XAI lacks standardization	Foundational
3	Alufaisan et al. (2021)	Empirical	XAI & decision-making	Experiment	Explanation detail	Too much detail reduces performance	Fatigue evidence
4	Ameen & Al-Ansari (2025)	Empirical	Trust fatigue	Survey	Trust, overreliance	Excess AI reduces judgment	Core fatigue concept
5	Bertrand et al. (2022)	Review	Bias in XAI	Systematic review	Cognitive bias	Bias distorts decisions	Bias dimension
6	Boyacı et al. (n.d.)	Theoretical	Explainability in collaboration	Conceptual	Interpretability	Explainability enhances collaboration	Framework support
7	Chen et al. (2023)	Bibliometric	XAI research trends	Quantitative	Research themes	Rapid XAI growth	Landscape
8	Chiaburu et al. (2024)	Empirical	Learning effects of XAI	Experiment	Confidence, uncertainty	Explanations affect confidence	Education angle
9	Clark & Kimmons (2023)	Theoretical	Cognitive Load Theory	Conceptual	Cognitive load	Learning limited by mental capacity	Core theory
10	Doh (2025)	Theoretical	Trust calibration	Framework	Trust calibration	Predictive models prevent overreliance	Trust control

11	Fox & Rey (2024)	Theoretical	CLT in XAI	Conceptual	Intrinsic/extraneous load	Poor explanations increase load	Core framework
12	Haque et al. (2023)	Review	User-centered XAI	Literature review	Usability, trust	Users struggle with complexity	User perspective
13	He et al. (2025)	Empirical	Conversational XAI	Experiment	Interaction style	Conversational XAI improves usability	Emerging approach
14	Herm (n.d.)	Empirical	XAI & cognitive load	Experiment	Mental workload	XAI increases cognitive demand	Fatigue evidence
15	Janssen et al. (2022)	Empirical	XAI in government	Experiment	Experience, trust	Experience moderates XAI impact	Context factor
16	Jean & Pera (2025)	Theoretical	Cognitive-AI bridge	Framework	Decision-making	Aligning cognition improves AI use	Conceptual
17	Kaur et al. (2024)	Empirical	Bounded rationality	Mixed methods	Rational limits	Users misinterpret explanations	Cognitive limits
18	Liao (n.d.)	Theoretical	Human-centered XAI	Framework	User experience	XAI must adapt to users	Design principle
19	Maehigashi et al. (2024)	Empirical	Explanation quantity	Experiment	Explanation length	Moderate detail optimal	Adaptive design
20	Margondai & Mouloua (2026)	Theoretical	Transparency paradox	Conceptual	Cognitive load	Too much transparency harms autonomy	Fatigue theory
21	Miller (2023)	Theoretical	Future of XAI	Conceptual	Decision support	Shift to evaluative AI	Future direction
22	Morandini et al. (2023)	Review	XAI & trust	Scoping review	Trust	Explainability improves trust conditionally	Trust link
23	Morrison et al. (2023)	Empirical	Explanation strategies	Experiment	Strategy type	Some explanations mislead users	Design issue
24	Neyigapula (2023)	Theoretical	Cognitive ergonomics	Conceptual	Human-AI interaction	Cognitive design improves outcomes	Ergonomics

25	Niewint-Gori (2025)	Review	XAI in education	Multi-stakeholder	Transparency	Context matters in XAI	Sector relevance
26	Romeo & Conti (2025)	Review	Automation bias	Literature review	Bias	Users over-rely on AI	Bias support
27	Sanneman & Shah (2022)	Theoretical	SAFE-AI framework	Framework	Situation awareness	Improves human-AI interaction	Design model
28	Spitzer et al. (2025)	Empirical	Misinformation effect	Experiment	Misinterpretation	Explanations distort decisions	Negative effect
29	Suresh et al. (2021)	Theoretical	Stakeholder framework	Conceptual	Stakeholders	Different users need different XAI	Personalization
30	Triki & Turki (n.d.)	Theoretical	AI fatigue	Conceptual	Cognitive overload	AI leads to mental fatigue	Core concept
31	Wang et al. (2026)	Theoretical	Cognitive load in AI	Framework	Mental workload	Overload reduces efficiency	Strong theory
32	Chiaburu et al. (2024)	Empirical	Confidence in learning	Experiment	Confidence	Explanations alter certainty	Learning effect

4. RESULTS

4.1 Study Distribution

The 41 included studies comprised empirical, conceptual, survey, and methodological works, reflecting interdisciplinary engagement across AI, HCI, and cognitive psychology.

4.2 Thematic Synthesis

4.2.1 Explanation Complexity and Extraneous Cognitive Load

Explanation complexity refers to the structural, semantic, and computational sophistication embedded within an AI explanation, including abstraction level, feature dimensionality, probabilistic reasoning, and domain-specific terminology. From a cognitive load perspective, explanations contribute to both intrinsic and extraneous load. While intrinsic load is tied to task complexity, extraneous load arises from suboptimal information presentation (Wang et al., 2026). When explanations expose internal model parameters, dense feature attribution scores, or highly technical logic without scaffolding, they increase extraneous load that does not directly enhance schema formation.

Empirical evidence shows that highly detailed explanations can impair performance rather than improve it (Alufaisan et al., 2021). This occurs because attentional resources are diverted toward parsing the explanation structure rather than evaluating the decision's relevance (Liao, n.d.). Recent studies demonstrate that increasing explanation complexity elevates cognitive load and can degrade user performance, particularly under time-constrained or high-stakes conditions (Alufaisan et al., 2021). Over time, repeated exposure to high-load explanations depletes cognitive resources, increasing the likelihood of disengagement or reliance on heuristics (Suresh et al., 2021). This aligns with contemporary perspectives on human-centered explainable AI, which emphasize the limitations of human cognitive capacity in processing complex model outputs (Liao, n.d.). Thus, complexity becomes fatigue-inducing when it exceeds the user's available cognitive capacity and fails to contribute proportionally to task-relevant understanding (Suresh et al., 2021)

4.2.2 *Trust Miscalibration*

Trust calibration refers to the alignment between system reliability and user reliance. Optimal calibration occurs when users appropriately trust accurate systems and appropriately scrutinise fallible ones. Explainability fatigue disrupts this calibration process. Cognitive depletion reduces evaluative reasoning, increasing reliance on cognitive shortcuts. In fatigued states, users may rely on automation (automation bias) because analytical processing becomes effortful (Skitka et al., 2021). Conversely, repeated exposure to frustrating or incomprehensible explanations may produce distrust, even when system outputs are accurate (Bansal et al., 2021). Thus, fatigue can generate both overtrust and undertrust depending on contextual and affective responses.

The primary mechanism is that fatigued users allocate less attention to explanations, thereby reducing their ability to assess system reliability.

4.2.3 Degradation of Decision Quality

Explainability fatigue paradoxically degrades decision quality by overwhelming users' cognitive capacity, weakening accountability mechanisms, and obscuring responsibility distinctions—requiring "cognitively sustainable transparency" rather than maximal explanation. Recent evidence supports this phenomenon across multiple dimensions. (Miller, 2023) demonstrates that people often fail to engage adequately with explanations despite their availability, undermining decision support effectiveness. (Morrison et al., 2023) showed that while causal explanation strategies can improve accuracy, they may also lead to incorrect rationalizations when AI presents correct assessments with flawed localization.

(Niewint-Gori, 2025) identifies cognitive overload from complex explanations as a critical challenge, alongside automation bias fostering over-reliance. (Ameen & Al-Ansari, 2025) document how excessive AI-driven decision support erodes managerial judgment through cognitive offloading and responsibility diffusion.

However, (Janssen et al., 2022) found that explainable AI combined with experience helps detect incorrect suggestions, though even experienced users miss some errors—suggesting explanation effectiveness depends heavily on user expertise and design.

5. DISCUSSION

5.1 Integrating XAI and Cognitive Load

The foundational claim about XAI's goals is reinforced: (Ali et al., 2023)confirm that "the need for explainable AI (XAI) methods for improving trust in AI models has arisen." However, the critical paradox you highlight is directly addressed by Osman Kaya et al., 2026, who state that "transparency alone does not guarantee trust."

Regarding negative effects of explanations, (Morandini et al., 2023)provide empirical evidence that "low-fidelity explanations, feelings of fear or discomfort, and low perceived usefulness can decrease trust." They also note that different explanation methods have varying effectiveness—"LIME appear effective at increasing user trust, while SHAP explanations perform less well."

The user-centered perspective you mention is supported by (Haque et al., 2023), who identify multiple dimensions affecting XAI effectiveness including format, completeness, and accuracy, and emphasize that XAI effects vary across users.

Cognitive Load Theory

Cognitive Load Theory's three-type framework—intrinsic, extraneous, and germane load—remains foundational (Clark & Kimmons, 2023). Regarding XAI specifically, empirical evidence demonstrates that explanation types significantly influence cognitive load and task performance: (Herm, n.d.) found distinct XAI explanation types "strongly influence end-users' cognitive load, task performance, and task time" across 271 prospective physicians.

More comprehensively, (Margondai & Mouloua, 2026) developed a theoretical framework showing that transparency effects depend on "dynamic cognitive resource depletion," with optimal policies adapting information provision based on real-time cognitive state rather than static design choices.

5.2 Defining Explainability Fatigue

Based on findings from XAI research and Cognitive Load Theory, this study defines explainability fatigue as follows:

Explainability fatigue occurs when users experience cognitive, emotional, or attentional exhaustion from AI explanations that are excessively lengthy, overly complex, or misaligned with user needs, resulting in reduced engagement, diminished understanding, or uncritical reliance on AI outputs.

Evidence supporting this concept is moderate but multifaceted. (Herm, n.d.) Empirically demonstrated that distinct XAI explanation types “strongly influence end-users’ cognitive load, task performance, and task time” across 271 prospective physicians. (Romeo & Conti, 2025)found that “overly technical, cognitively demanding, or even simplistic explanations may inadvertently reinforce misplaced trust,” particularly among less experienced users, and that “explanations may increase perceived system acceptability” but remain “often insufficient to improve decision accuracy or mitigate automation bias.” (Triki & Turki, n.d.) Identified cognitive and emotional fatigue dimensions in AI interactions through qualitative analysis of 22 users. However, (He et al., 2025)observed that even enhanced conversational XAI interfaces produced “clear over-reliance on the AI system,” suggesting explanation design alone may be insufficient to prevent fatigue-related outcomes.

6. DRIVERS OF EXPLAINABILITY FATIGUE: A SOCIO-COGNITIVE ANALYSIS

Explainability fatigue emerges not from explanations alone, but from how explanation features interact with cognitive capacity constraints and contextual demands, creating mental strain through information-induced cognitive load. (Herm, n.d.) An empirical study with 271 prospective physicians found that different XAI explanation types "strongly influence end-users' cognitive load, task performance, and task time." (Fox & Rey, 2024) applied Cognitive Load Theory comprehensively to machine learning explainability, demonstrating that "complicated and ambiguous information entails high cognitive load" that can limit understanding and learning.

Critically, (Margondai & Mouloua, 2026) developed a theoretical framework showing that "transparency effects depend on dynamic cognitive resource depletion rather than static design choices," with information provision triggering metacognitive processing that reduces perceived control when load exceeds working memory capacity. (Ahn et al., 2025) further demonstrated that explanation effectiveness varies based on "who they are, when, how, and what to explain—with different levels of cognitive load and engagement and sociotechnical contexts," emphasizing that no single explanation strategy universally mitigates fatigue.

7. CONCEPTUAL FRAMEWORK

This study suggests that explainability fatigue links the design of explanations to the outcomes of responsible AI.

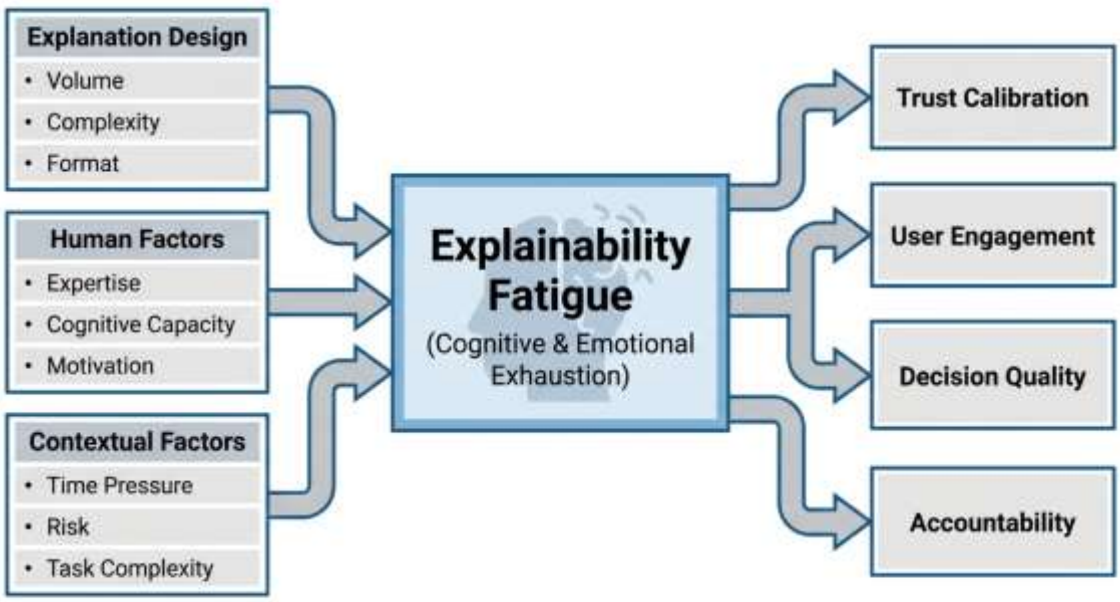

Figure 2 Explainability fatigue framework conceptual framework diagram.

Antecedent Variables

1. Explanation Complexity (length, technicality, abstraction level)
2. Explanation Volume (frequency and detail)
3. User Characteristics (expertise, cognitive capacity, motivation)
4. Contextual Constraints (time pressure, risk level, multitasking)

Mediator

Explainability Fatigue:

- Cognitive exhaustion
- Emotional disengagement
- Reduced attentional allocation

Outcomes

1. Trust Miscalibration
2. Reduced Decision Quality
3. Accountability Gaps

7.1 Research Propositions

P1: Higher explanation complexity increases cognitive load.

P2: Cognitive load positively predicts explainability fatigue.

P3: Explainability fatigue mediates the relationship between explanation design and trust calibration.

P4: Explainability fatigue negatively impacts decision quality.

P5: Contextual constraints moderate the relationship between explanation exposure and fatigue intensity.

8. Implications

Consequences for Design

The findings of this research show that explainable AI systems should adopt either consistent explanation methods or flexible, user-driven approaches that fit users' backgrounds, abilities, and situations. Recent evidence strongly supports this position. A. Mohammed (2025) demonstrated that adaptive explanations personalized to user expertise significantly enhanced understanding (+27%), trust calibration (+19%), and decision-making accuracy (+22%) compared to static approaches, grounded in cognitive load theory. Christian Meske et al. (2025) found that users consistently prefer context-sensitive, multimodal explanations over technical transparency, establishing progressive disclosure as a key design principle. The Human Factor in Explainabl (2025) emphasizes dynamic explanations tailored to diverse user needs, while San Hong & Woojin Park (2025) developed a comprehensive user-centered design framework incorporating explanation needs identification and user interface design. Critically, Sneha Roychowdhury et al. (2025) highlight that human-centered evaluation frameworks must balance technical fidelity with human interpretability, incorporating both objective and subjective metrics. Osman Kaya et al. (2026) reinforce this shift toward systems aligned with human understanding rather than algorithmic transparency alone, marking the transition to "XAI 2.0.".

9. Limitations and Future Research

This study is conceptual and does not include real-world testing. Future research should define explainability fatigue in practical terms, develop methods to measure it, and test the framework across different fields and user groups.

10. Conclusion

Explainability is central to responsible AI; however, it is not inherently beneficial at unlimited quantities or with unlimited complexity. Explainability is important for responsible AI, but more complex explanations are not always better. This review shows that human mental limits affect how well AI explanations work. By introducing explainability fatigue, the study suggests that explainability should be seen as a process with limits that needs flexible, context-aware design. Future research should define and measure explainability fatigue, and test the framework in areas such as healthcare, finance, and public-sector AI. Understanding human limits is key to ensuring explainability delivers on its promise in trustworthy AI and intelligible systems.

REFERENCES

[1] Ahn, Y., Lin, Y.-R., Alikhani, M., & Cheon, E. (2025). *Human-centered explanation does not fit all: The interplay of sociotechnical, cognitive, and individual factors in the effect AI explanations in algorithmic decision-making*. http://arxiv.org/abs/2502.12354

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

[3] Alufaisan, Y., Marusich, L. R., Bakdash, J. Z., Zhou, Y., & Kantarcioglu, M. (2021). *Does Explainable Artificial Intelligence Improve Human Decision-Making?* https://www.

[4] Ameen, A. H., & Al-Ansari, S. R. A. (2025). Algorithmic trust fatigue: How excessive AI-Driven decision support erodes managerial judgment in modern organizations. *International Journal of Research in Finance and Management*, *8*(2), 1121–1129. https://doi.org/10.33545/26175754.2025.v8.i2l.649

[5] Bertrand, A., Belloum, R., Eagan, J. R., & Maxwell, W. (2022). How cognitive biases affect XAI-Assisted decision-making: A systematic review. *AIES 2022 - Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 78–91. https://doi.org/10.1145/3514094.3534164

[6] Boyacı, T., Canyakmaz, C., & De Véricourt, F. (n.d.). *Beyond the Black Box: Unraveling the Role of Explainability in Human-AI Collaboration*.

[7] Chen, H.-Y., Sharma, C., Sharma, S., Sharma, K., & Sethi, G. K. (2023). *Intellectual Structure of Explainable Artificial Intelligence: a Bibliometric Reference to Research Constituents*. https://doi.org/10.21203/rs.3.rs-3493299/v1

[8] Chiaburu, T., Haußer, F., & Bießmann, F. (2024). *Confident Teacher, Confident Student? A Novel User Study Design for Investigating the Didactic Potential of Explanations and their Impact on Uncertainty*. http://arxiv.org/abs/2409.17157

[9] Clark, C., & Kimmons, R. (2023). Cognitive Load Theory. *EdTechnica*. https://doi.org/10.59668/371.12980

[10] Doh, W. (2025). *Calibrating Trust in Human-AI Collaboration: Systematically Predicting and Preventing Overreliance Through the Human-AI-System Concordance (HASC) Matrix and Cognitive Interventions*.

[11] Fox, S., & Rey, V. F. (2024). A Cognitive Load Theory (CLT) Analysis of Machine Learning Explainability, Transparency, Interpretability, and Shared Interpretability. *Machine Learning and Knowledge Extraction*, *6*(3), 1494–1509. https://doi.org/10.3390/make6030071

[12] Haque, A. B., Islam, A. K. M. N., & Mikalef, P. (2023). Explainable Artificial Intelligence (XAI) from a user perspective: A synthesis of prior literature and problematizing avenues for future research. *Technological Forecasting and Social Change, 186*. https://doi.org/10.1016/j.techfore.2022.122120

[13] He, G., Aishwarya, N., & Gadiraju, U. (2025). Is Conversational XAI All You Need? Human-AI Decision Making With a Conversational XAI Assistant. *International Conference on Intelligent User Interfaces, Proceedings IUI*, 907–924. https://doi.org/10.1145/3708359.3712133

[14] Herm, L.-V. (n.d.). *IMPACT OF EXPLAINABLE AI ON COGNITIVE LOAD: INSIGHTS FROM AN EMPIRICAL STUDY*.

15] Janssen, M., Hartog, M., Matheus, R., Yi Ding, A., & Kuk, G. (2022). Will Algorithms Blind People? The Effect of Explainable AI and Decision-Makers' Experience on AI-supported Decision-Making in Government. *Social Science Computer Review*, *40*(2), 478–493. https://doi.org/10.1177/0894439320980118

[16] Jean, N., & Pera, G. Le. (2025). *Bridging Human Cognition and AI: A Framework for Explainable Decision-Making Systems*. http://arxiv.org/abs/2509.02388

[17] Kaur, H., Conrad, M. R., Rule, D., Lampe, C., & Gilbert, E. (2024). Interpretability Gone Bad: The Role of Bounded Rationality in How Practitioners Understand Machine Learning. *Proceedings of the ACM on Human-Computer Interaction*, *8*(1). https://doi.org/10.1145/3637354

[18] Liao, Q. V. (n.d.). *Human-Centered Explainable AI (XAI): From Algorithms to User Experiences*. Retrieved https://arxiv.org/abs/2110.10790

[19] Maehigashi, A., Fukuchi, Y., & Yamada, S. (2024, May 11). Adjusting Amount of AI Explanation for Visual Tasks. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3613905.3650802

[20] Margondai, A., & Mouloua, M. (2026). *The Transparency Paradox in Explainable AI: A Theory of Autonomy Depletion Through Cognitive Load*. http://arxiv.org/abs/2601.13973

[21] Miller, T. (2023). Explainable AI is Dead, Long Live Explainable AI! Hypothesis-driven Decision Support using Evaluative AI. *ACM International Conference Proceeding Series*, 333–342. https://doi.org/10.1145/3593013.3594001

[22] Morandini, S., Fraboni, F., Puzzo, G., Giusino, D., Volpi, L., Brendel, H., Balatti, E., De Angelis, M., De Cesarei, A., & Pietrantoni, L. (2023). *Examining the Nexus between Explainability of AI Systems and User's Trust: A Preliminary Scoping Review*. https://orcid.org/0009-0007-3984-164X

[23] Morrison, K., Shin, D., Holstein, K., & Perer, A. (2023). Evaluating the Impact of Human Explanation Strategies on Human-AI Visual Decision-Making. *Proceedings of the ACM on Human-Computer Interaction, 7*(CSCW1). https://doi.org/10.1145/3579481

[24] Neyigapula, B. S. (2023). *Human-AI Collaborative Decision-making: A Cognitive Ergonomics Approach*. https://doi.org/10.21203/rs.3.rs-3258718/v1

[25] Niewint-Gori, J. (2025). Explainable AI in education: A multi-stakeholder approach to transparency and ethical practice. *IUL Research*, *6*(12), 117–132. https://doi.org/10.57568/iulresearch.v6i12.781

[26] Romeo, G., & Conti, D. (2025). Exploring automation bias in human–AI collaboration: a review and implications for explainable AI. *AI and Society*. https://doi.org/10.1007/s00146-025-02422-7

[27] Sanneman, L., & Shah, J. A. (2022). The Situation Awareness Framework for Explainable AI (SAFE-AI) and Human Factors Considerations for XAI Systems. *International Journal of Human-Computer Interaction*, *38*(18–20), 1772–1788. https://doi.org/10.1080/10447318.2022.2081282

[28] Spitzer, P., Holstein, J., Morrison, K., Holstein, K., Satzger, G., & Kühl, N. (2025). Don't Be Fooled: The Misinformation Effect of Explanations in Human–AI Collaboration. *International Journal of Human-Computer Interaction*. https://doi.org/10.1080/10447318.2025.2574511

[29] Suresh, H., Gomez, S. R., Nam, K. K., & Satyanarayan, A. (2021, May 6). Beyond expertise and roles: A framework to characterize the stakeholders of interpretable machine learning and their needs. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3411764.3445088

[31] Triki, M., & Turki, A. M. (n.d.). *Between Cognitive Overload and Dehumanization: Exploring the Dimensions of Consumer Fatigue with Artificial Intelligence*. https://doi.org/10.47772/IJRISS

[32] Wang, P., Liu, H., Zou, L., & Paas, F. (2026). Overloaded minds and machines: a cognitive load framework for human-AI symbiosis. *Artificial Intelligence Review*, *59*(3). https://doi.org/10.1007/s10462-026-11510-z

Author's Brief Profile

Dahiru Usman Haruna is a dedicated and resourceful IT professional, academic, and project manager with over 15 years of experience in information technology and academia. Born in 80s, He obtained his first degree in ICT from the Federal University of Technology a Master's degree in Information Management Technology from Modibbo Adama University (MAU) and a PhD in IMT with specialization in Data Warehousing and Neural Networks. Dr. Dahiru holds several professional certifications and has contributed to various IT projects in industry and academia. He is currently the pioneer and Head of the Department of Data Science and Artificial Intelligence at MAU, Yola, and is a member of professional bodies including the Computer Professionals Registration Council of Nigeria (CPN), Nigerian Computer Society (NCS), and the Academia in Information Technology Profession. I can be reached by phone on +2348062943452 and through E-mail dahiruusman222@mau.edu.ng

Manunyi Godfrey is currently a Student in the Department of Computer Science at Modibbo Adama University, Yola Adamawa State, Nigeria. I'm current pursuing my master's Degree in Computer Science. He can be reached by phone on +2347064786424 and through E-mail manugodfrey02@gmail.com