EasyChair Preprint

№ 15992

Explainable by Design: A design framework to support the design of explainable user interfaces

Kerri-Ellen Casey and Sonia Cláudia da Costa Sousa

August 7, 2025

Explainable by Design

A design framework to support the design of explainable user interfaces

KERRI-ELLEN CASEY, Tallinn University, Estonia

SONIA SOUSA, Tallinn University, Estonia

The technological advancement of artificial intelligence (AI) and large language models (LLMs) are rapidly changing what systems can do and how people interact with them. Usable and Explainable AI (XAI) is identified as a key human-centred AI (HCAI) challenge to address. Although existing research offers high-level principles and guidelines to design for AI, there is limited support in how to translate these into interface decisions, specifically for the design of explainable user interfaces (XUI). This research aims to address the gap through the development of a practical framework for user experience (UX) designers. Through the RtD methodology, the research follows a qualitative approach across four phases. In Phase 1, a scoping review and thematic analysis identified interface-level XUI guidelines. In Phase 2, the guidelines were validated and operationalised into practitioner-facing reflective design questions through two expert reviews. In Phase 3, a participatory workshop with UX designers classified UI patterns across explanation dimensions to support explainability. Finally the outputs across all phases were synthesised into a practical and flexible framework for UX designers. A set of 5 learning cards introduce the theoretical foundations, 14 XUI guidelines accompanied by reflective questions support theory in practice, and a UI pattern decision tree to guide the selection of design patterns. The result is a design artefact that bridges academic theory and design practice for the design of XUIs in AI products and systems.

CCS Concepts: • **Human-centered computing** → **Human computer interaction**; **Web-based interaction**.

Additional Key Words and Phrases: Explainable AI (XAI), Explainable user interfaces (XUI), Interaction Design, User Experience, Human-Centered design

ACM Reference Format:

Kerri-ellen Casey and Sonia Sousa. 2025. Explainable by Design: A design framework to support the design of explainable user interfaces. 1, 1 (July 2025), 7 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

In 2025, with the technological advancement of large language models (LLMs), AI is becoming more 'efficient, affordable, and accessible' [9], a transition from an emerging technology to a business imperative. The development of new algorithms is changing what systems and products can do and also how people interact with them. Nielsen (2023) [11] observes that we are now witnessing the emergence of the third user interface paradigm in sixty years. Command-based interactions have dominated for three generations of user interface design, whereby users and computers take turns, and users explicitly direct the system. The third paradigm, intent-based outcomes, shifts the locus of control, where the user tells the computer what outcome they want. This transition brings benefits and challenges 'when users don't know how something was done, it can be harder for them to identify or correct the problem' [11]. These changes also create new challenges for user experience (UX) designers to address, who must now design interfaces that operate probabilistically

Authors' Contact Information: Kerri-ellen Casey, Tallinn University, Estonia, casey@tlu.ee; Sonia Sousa, Tallinn University, Estonia, scs@tlu.ee.

Manuscript submitted to ACM

rather than deterministically. These non-deterministic behaviours produce outputs that are difficult to predict or explain [15]. UX designers can address these changes by mitigating potential risk for end-users through explainable and transparent user interfaces. As noted by Chen (2025) [4], as AI maturity grows, the demand for designers with the skills to shape these experiences will increase, and the speed at which generative AI has been deployed necessitates the need to reskill the labour force [4]. There is a growing body of HCAI research which aims to address the design challenges of AI through the contribution of design principles and guidelines [2, 3, 13]. More recent work focuses specifically on generative-AI [15, 16] and explainable interfaces [10]. While the existing literature offers high-level principles and guidelines, there is limited support to translate these principles into practical, interface-level decisions for explainable user interfaces (XUI). This lack of support leaves designers without tools to translate and apply explainability concepts during the product design process. Furthermore, a lack of practical support risks overlooking the explanation needs of end users in real-world systems. With the rapid development of AI into digital products and systems, this gap highlights the opportunity to provide support to assist designers' decision-making process.

2 Research goal and motivation

The aim of this research is to create a design framework to support the design of explainable user interfaces grounded in the literature and drawing from practice. Three research goals (RG) have been identified (1) to identify and consolidate AI design principles and guidelines in the literature to support the design of explainable user interfaces; (2) to identify UI patterns that support explainability; and (3) to develop an explainable user interface framework to support designers when designing AI systems and products.

2.1 Research methodology

This research draws on two distinct but complementary sources: design guidelines in the literature (RQ1) and UI patterns from industry practice (RQ2). Their integration forms the design framework in (RQ3). The research follows a research procedure divided into 4 phases 1.

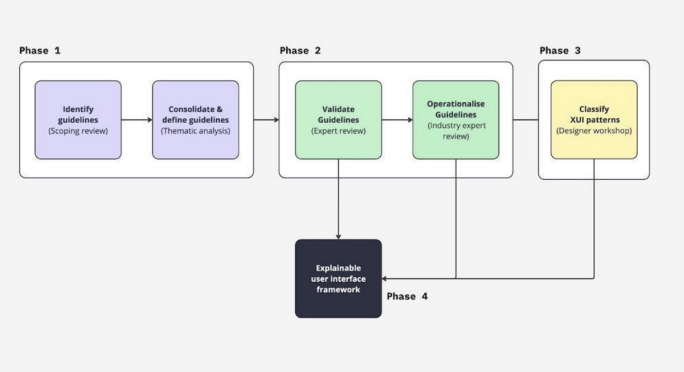

Fig. 1. Overview of the research methodology

3 Background

Explainable AI, a term coined by game developers and academic researchers van Lent et al. in (2004) [14], was used to define their 'system's capacity to explain the actions of AI-controlled entities in simulation gaming applications.' [1]. Today, it is an expansive research field with the purpose of making the results or outputs of AI systems more understandable and transparent [1]. Although there are no standardised definitions of XAI, DARPA defines XAI as 'AI systems that can explain their rationale to a human user, characterise their strengths and weaknesses, and convey an understanding of how they will behave in the future' [6]. Through a set of methods and techniques, XAI aims to address the black-box of AI to explain its inner workings and to produce more explainable models [8].

From a user perspective, we can think of XAI as 'How does it work?', "What mistakes can it make?" and "Why did it just do that?" [7]. Furthermore, from a regulatory standpoint, the European Union has established the 'right to explanation' under the General Data Protection Regulation (GDPR) [5], and for organisations in many contexts, explainability extends to this legal requirement. Although significant progress has been made in XAI research, the needs and goals of the target users can be overlooked, and there are still limitations in creating effective explanations for end users. Nguyen et al. (2024, p. 1) [10] states that "most explanations produced by XAI still lack usability, practical interpretability, and efficacy for real users". To address this, a new perspective is required to shape the future of technology so as 'to better serve human needs' [11].

The shift from an algorithm-focused approach to a humanist approach is addressed by human-centred AI (HCAI). HCAI advocates for the design and development of AI systems that are trustworthy, ethical and usable with a consideration for the individual and societal impacts. In order to foster Responsible AI, we need to go beyond what is inside the "black-box" and understand what is outside the 'black box' [12].

The study of explainable interfaces represents an emerging field that considers the design aspects of XAI systems. To enhance the user experience of XAI, the design of user interfaces is increasingly recognised as a vital component within the overall XAI application process [8]. As the term is relatively new dating back to 2021 [10] there are limited definitions for EI. Chromik and Butz (2021) [3] define the concept of an explanation user interface (XUI) as 'the sum of outputs of an XAI process that the user can directly interact with'. Almost all XAI systems will have an explainable interface, whether they are intentionally designed or not [10]. Thus, it is imperative that there are effective guidelines for designers to design AI explanations that communicate and facilitate the interaction.

4 Results

This section starts by presenting an overview of the RQ1 results (What AI design principles and guidelines exist in the literature to support the design of explainable user interfaces?. Based on the scoping review, guidelines were consolidated under six themes (Interaction, Presentation, Language, Users, Evaluation, and Ethics and trust). The themes were validated and refined with two experts (an XAI expert, an industry practitioner, and a UX educator) and aimed to provide feedback on their clarity and usefulness for designers.

4.1 The proposed AI design principles and guidelines

Understanding and transparency: support users to develop accurate mental models of the AI system. These guidelines address the fundamental need for users to understand how the AI functions, what it can do, and why it produces specific outputs.

Exploration and control: encourage designers to create meaningful user agency and exploration of AI outputs. Emphasise user interaction with the AI as well as multiple interaction types and explanation outputs to maintain appropriate control before, during, and after a user-AI interaction.

Usability and accessibility: addresses the need for usable and accessible explainable interfaces. Consider multimodal explanations to ensure AI outputs are suitable for users with a diverse set of needs and abilities, to reveal the necessary information for the task at hand through progressive disclosure, and always provide the option for users to give their feedback.

Information and presentation: focus on effective communication of AI explanations. This includes considerations for the structure of visual and textual information to reinforce understanding and learning. In totality, these guidelines address how the explanation content should be organised, formatted, and delivered to the end users.

4.2 Classification of UI patterns to support explainability in the interface

Address RQ2 ("How do designers classify explainable UI patterns across explanation dimensions?") achieved through a participatory workshop with experienced designers with the aim to identify UI patterns that can support explainability and transparency and classify them according to explanation dimensions (Explanation scope, User need for explanation, and Interaction phase).

Table 1. UI pattern classification - Explanation scope distribution

Dimension	Designer mapping	Count	UI pattern
Explanation scope	Local	9	Show the work, Input and output comparison, Demonstrated thinking, Progressive processing display, Before & after comparison, Manual overrides, Confidence status, Caveat
	Global	5	Dataset explorer, Explanation of algorithm, Token transparency, Algorithm effectiveness rating, Dataset visualisation
	Both (Global & Local)	5	Risk alert, Criteria sliders, Setting expectations, Source citation, Modal confidence

This classification was grounded in the concepts gathered from the literature. Results highlighted three classifications of explanation scope (Global, Local, Both global and local) along with the remaining dimensions, user need for explanations, and where it may occur in the human-AI interaction (Before, During, After).

4.3 XUI framework to support the designers

Through a scoping review and thematic analysis, two expert reviews, and a designer participatory workshop, the final framework was designed to balance theory and practice through learning cards, guidelines, reflective questions, and a UI pattern decision tree to support the design of explainability in the user interface. The results were addressed through RQ3 (How to develop a design framework to support the design of explainable user interfaces?).

The combination of the insights are envisioned through the framework for XUI: (1) the set of learning cards provide designers with a foundational knowledge in XUIs; (2) the guideline detail cards encourage a user-centered perspective

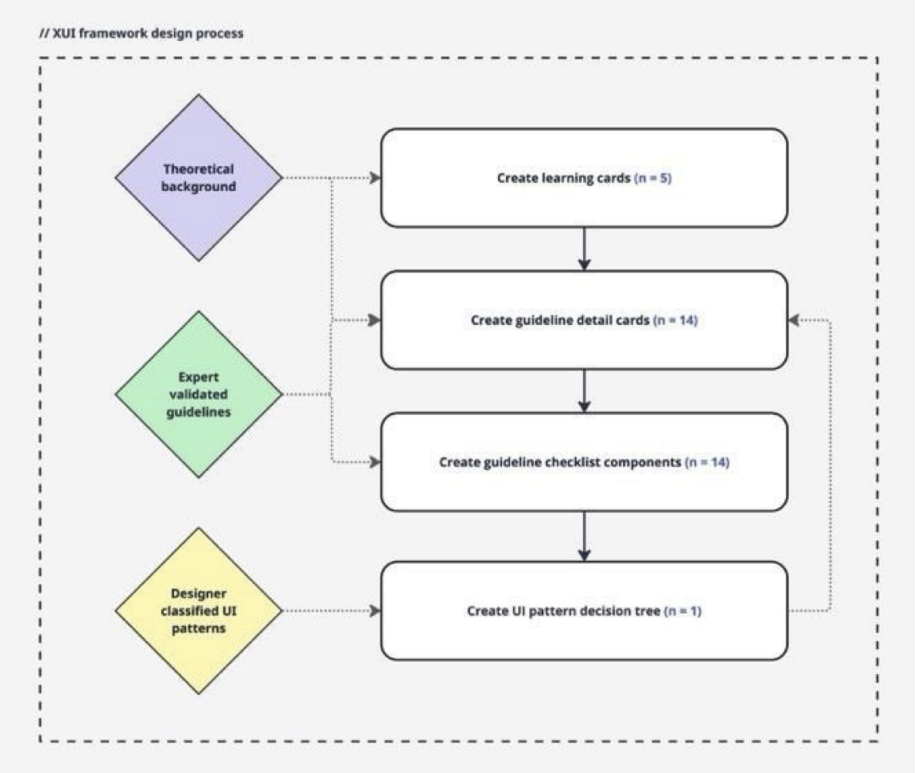

Fig. 2. XUI framework design process

when thinking about explanations and explainable interfaces to support HCAI goals; (3) the guideline detail cards formatted as reflective practitioner questions supports design reflection-in-action (Schön, 1993); and (4) the UI pattern decision tree allows for the comparison of UI patterns across explanation dimensions to bridge theory with interface-level choices.

5 Conclusions

As AI becomes more embedded in our everyday systems and products, it changes not only what systems can do but also how people interact with them. AI is no longer a feature; it is a new design material with which designers must work. It requires us to think about how we communicate, guide and empower users ethically to build and retain trust. Research-through-design aims to produce a tangible artefact and build a bridge between theory and practice. Through three core research goals:

(1) identify and consolidate AI design principles and guidelines in the literature to support the design of explainable user interfaces,
(2) classify UI patterns that support explainability and
(3) Develop an explainable user interface framework to support designers, the end result is a design artefact grounded in literature and practice.

As we enter this new interface paradigm, Explainable by Design is a framework to support the design of explainable user interfaces. While the framework will continue to evolve, it reflects the broader imperative to equip designers with the knowledge, language, and actionable steps to shape AI products that are explainable, transparent, and human-centred.

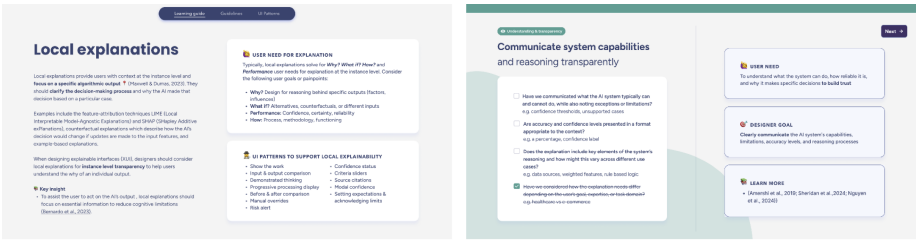

Fig. 3. Explainable by Design: Sample cards

Acknowledgments

This work was supported by the AI-Mind "Intelligent digital tools for screening of brain connectivity and dementia risk estimation in people affected by mild cognitive impairment". H2020 CORDIS (964220).

The authors wish to acknowledge the use of ChatGPT by OpenAI in writing this paper. We used the tool to assist with improving the language, refining the structure, and clarifying responses to reviewer feedback. The paper accurately represents the authors' underlying research, analysis, and novel intellectual contributions.

References

[1] Sajid Ali, Tamer Abuhmed, Shaker El-Sappagh, Khan Muhammad, Jose M Alonso-Moral, Roberto Confalonieri, Riccardo Guidotti, Javier Del Ser, Natalia Díaz-Rodríguez, and Francisco Herrera. 2023. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Information fusion* 99 (2023), 101805.

[2] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–13.

[3] Michael Chromik and Andreas Butz. 2021. Human-XAI interaction: a review and design principles for explanation user interfaces. In *Human-Computer Interaction–INTERACT 2021: 18th IFIP TC 13 International Conference, Bari, Italy, August 30–September 3, 2021, Proceedings, Part II 18*. Springer, 619–640.

[4] Michael Chui, Eric Hazan, Roger Roberts, Alex Singla, and Kate Smaje. 2023. The economic potential of generative AI. (2023).

[5] Bryce Goodman and Seth Flaxman. 2017. European Union regulations on algorithmic decision-making and a "right to explanation". *AI magazine* 38, 3 (2017), 50–57.

[6] David Gunning and David Aha. 2019. DARPA's explainable artificial intelligence (XAI) program. *AI magazine* 40, 2 (2019), 44–58.

[7] M. Kurdziolek. 2023. Explaining the unexplainable: explainable AI (XAI) for UX. User Experience. Retrieved June 12, 2025 from https://uxpamagazine.org/explaining-the-unexplainable-explainable-ai-xai-for-ux/

[8] Dian Lei, Yao He, and Jianyou Zeng. 2024. What is the focus of XAI in UI design? Prioritizing UI design principles for enhancing XAI user experience. In *International Conference on Human-Computer Interaction*. Springer, 219–237.

[9] Nestor Maslej, Loredana Fattorini, Raymond Perrault, Yolanda Gil, Vanessa Parli, Njenga Kariuki, Emily Capstick, Anka Reuel, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Juan Carlos Niebles, Yoav Shoham, Russell Wald, Tobi Walsh, Armin Hamrah, Lapo Santarlasci, Julia Betts Lotufo, Alexandra Rome, Andrew Shi, and Sukrut Oak. 2025. Artificial Intelligence Index Report 2025. arXiv:2504.07139 [cs.AI] https://arxiv.org/abs/2504.07139

[10] Thu Nguyen, Alessandro Canossa, and Jichen Zhu. 2024. How Human-Centered Explainable AI Interface Are Designed and Evaluated: A Systematic Survey. arXiv:2403.14496 [cs.HC] https://arxiv.org/abs/2403.14496

[11] J. Nielsen. 2023. AI: First New UI Paradigm in 60 Years. Nielsen Norman Group. Retrieved June 18, 2025 from https://www.nngroup.com/articles/ai-paradigm/?utm_source=empathyandai.beehiiv.com&utm_medium=referral&utm_campaign=the-seismic-shift-in-ux-cx-has-begun

[12] Sonia Sousa, José Cravino, and Paulo Martins. 2023. Challenges and Trends in User Trust Discourse in AI Popularity. *Multimodal Technologies and Interaction* 7, 2 (2023). doi:10.3390/mti7020013

[13] Sonia Sousa, David Lamas, José Cravino, and Paulo Martins. 2024. Human-centered trustworthy framework: A human–computer interaction perspective. *Computer* 57, 3 (2024), 46–58.

[14] Michael Van Lent, William Fisher, and Michael Mancuso. 2004. An explainable artificial intelligence system for small-unit tactical behavior. In *Proceedings of the national conference on artificial intelligence*. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 900–907.

[15] Justin D Weisz, Jessica He, Michael Muller, Gabriela Hoefer, Rachel Miles, and Werner Geyer. 2024. Design principles for generative AI applications. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–22.

[16] Justin D Weisz, Michael Muller, Jessica He, and Stephanie Houde. 2023. Toward general design principles for generative AI applications. *arXiv preprint arXiv:2301.05578* (2023).