

A Design Approach to Explainable Interfaces for AI Systems

Ciprian Amaritei

Faculty of Computer Science,
Alexandru Ioan Cuza University
of Iași
Iași, Romania
ciprian.amaritei@info.uaic.ro

Cosmin Irimia

Faculty of Computer Science,
Alexandru Ioan Cuza University
of Iași
Iași, Romania
irimia.cosmin@gmail.com

Adrian Iftene

Faculty of Computer Science,
Alexandru Ioan Cuza University
of Iași
Iași, Romania
adiftene@info.uaic.ro

ABSTRACT

Explainable Artificial Intelligence (XAI) research has produced valuable knowledge about different types of explanations, users' information needs, how people understand explanations, and how explainability can be presented through interfaces. However, this knowledge is fragmented across different research areas, which makes it difficult for designers to use in practice. This paper proposes a structured design repository that brings this knowledge together as reusable guidance for designing explainable interfaces. Each repository entry describes an explanation type, its design purpose, the user's information needs, its SAFE-AI level, design recommendations, and an example interface. A functional prototype of the repository was developed and demonstrated through the design of an explainable Health Behaviour Change interface mockup. A preliminary expert review with five UX/HCI professionals explored the perceived usefulness of the repository as a design support resource. Participants considered the repository a useful starting point for AI interface design and highlighted its value in making AI-specific design considerations more explicit. These findings provide preliminary evidence of perceived usefulness, while further studies are required to evaluate its effect on design outcomes.

AUTHOR KEYWORDS

Design; artificial intelligence; explainable artificial intelligence; SAFE-AI.

DOI: 10.37789/icusi.2026.24

INTRODUCTION

Artificial Intelligence (AI) systems are increasingly embedded in applications across healthcare, finance, education, and digital services [2, 3, 8]. As these systems become more complex, Explainable Artificial Intelligence (XAI) has emerged as a research field concerned with making AI decisions transparent and understandable [1, 9, 13]. As AI capabilities become increasingly integrated into consumer and enterprise applications, UX designers are expected to design interfaces that communicate AI recommendations, predictions, confidence, and reasoning. Unlike traditional interfaces, AI-powered systems introduce new design challenges related to transparency, uncertainty, and user understanding [4]. Although explainability research guides these aspects, the relevant knowledge is dispersed across multiple perspectives, making it difficult for designers to identify, select, and apply appropriate explainability concepts during interface design.

Over the past decade, XAI has produced a wide range of explanation methods, including feature importance, counterfactual

explanations, rule-based explanations, and example-based explanations, enabling AI systems to provide insights into their predictions and behaviour [5, 11, 14].

Alongside these technical advances, researchers have increasingly recognised that explainability is not solely an algorithmic challenge but also a human-centred design problem. Recent work has expanded XAI beyond explanation methods to include user information needs [10, 11], cognitive models of understanding [15], and explainable interface design [7, 12]. Together, these complementary perspectives provide valuable knowledge for designing explainable AI systems.

Despite this progress, the knowledge produced by these research directions remains fragmented. Designers must independently interpret the types of explanations, users' information needs, levels of understanding, and interface design recommendations when creating explainable user interfaces. As a result, they often lack practical guidance on what should be explained, how explanations should be represented, and how different explainability concepts can be integrated into coherent interface designs.

This paper addresses this gap by proposing a structured design repository that integrates complementary explainability knowledge into a reusable resource for interface design. Rather than introducing new explanation methods, the repository organises existing explanation types together with user information needs, levels of understanding, interface guidance, implementation examples, and practical design considerations. To demonstrate the approach, we present a prototype implementation of the repository, illustrate its application through a Health Behaviour Change (HBC) case study, and report findings from a preliminary expert evaluation investigating its perceived usefulness for supporting interface design.

The following research question guides the work:

RQ: *How can complementary explainability knowledge be organised into a reusable design resource that supports interface design?*

RESEARCH APPROACH

This work adopts a conceptual and design-oriented research approach to investigate how existing explainability knowledge can be organised into a reusable resource that supports interface design. The research was conducted in four stages.

Stage 1: Literature Review

Relevant literature from Explainable Artificial Intelligence (XAI), Human-Centred XAI, Explainable Interfaces (EI), and Human-Computer Interaction (HCI) was reviewed to identify complementary perspectives on explainability. Particular attention was given to explanation types, user information needs, cognitive models of understanding, and interface design approaches.

Stage 2: Knowledge Synthesis

The complementary perspectives identified in the literature were analysed and synthesised into a unified structure suitable for interface design. Existing explainability concepts, user information needs, SAFE-AI levels of understanding, and interface design knowledge were consolidated to identify the information designers require when implementing explainability features.

Stage 3: Repository Design

Based on this synthesis, a structured design repository was developed. For each selected explanation type, the repository captures the associated design intent, interface representations, implementation examples, benefits, and limitations, providing reusable guidance for interface design.

Stage 4: Prototype Implementation and Demonstration

The repository was implemented as a prototype using Airtable and demonstrated through a Health Behaviour Change (HBC) case study. A preliminary expert evaluation was then conducted to investigate the perceived usefulness of the proposed repository for supporting interface design.

BACKGROUND AND RELATED WORK

Explainable Artificial Intelligence (XAI) has evolved into a broad research field encompassing multiple complementary perspectives on explainability [6]. Rather than focusing on a single aspect, existing research has addressed explainability from different viewpoints, including explanation types, user information needs, cognitive models of understanding, and explainable interface design. Together, these perspectives provide the theoretical foundations for designing explainable AI systems. The following subsections briefly review each perspective before positioning the contribution of this work.

Explainable Artificial Intelligence

Explainable Artificial Intelligence (XAI) develops methods for providing information about model behaviour, predictions, and decisions so that these outputs can be more meaningfully interpreted by people. Existing research has introduced a wide range of explanation methods, including feature importance techniques, rule-based explanations, example-based explanations, and counterfactual approaches [11, 14]. These methods have significantly advanced the technical capabilities of explainability. However, they primarily focus on how explanations are generated rather than how they should be communicated to end users through user interfaces [7].

Explainability Concepts

XAI research describes a wide variety of methods for generating information that helps users and practitioners interpret model outputs, predictions, and aspects of model behaviour. Surveys such as Ortigossa et al. [14] provide comprehensive overviews of technical explanation methods, while human-centred research highlights the forms of explanation that are meaningful for end users [11, 12]. Since no universally accepted taxonomy of explanation types exists, this work focuses on a selected set of explanation types that are both frequently discussed in the literature and particularly relevant for interface design. These explanation types constitute the conceptual building blocks of the proposed repository and are intended to support designers in selecting and implementing explainability features rather than providing an exhaustive classification.

The concepts considered in this work include both explanation mechanisms and information that forms the context for an explanation. For example, an AI output or prediction is not an explanation in itself, but it represents the result that users need to perceive before additional explanations can support comprehension or action. For this reason, Output / Prediction is included as an explainability concept rather than as an explanation method.

Table 1: Selected explainability concepts relevant for interface design.

Explanation Type	Purpose
Output / Prediction	Communicate the AI recommendation or decision. <i>Example:</i> “Drink 2.3 L today.”
Confidence	Communicate the certainty of the recommendation. <i>Example:</i> “Confidence: 92%.”
Feature Importance	Explain which factors most influenced the decision. <i>Example:</i> “Yesterday’s intake contributed 45% to today’s recommendation.”
Counterfactual	Show how a different outcome could be achieved. <i>Example:</i> “If you had drunk 2 L yesterday, today’s target would be 1.9 L.”
Example-based	Explain the recommendation using similar previous cases. <i>Example:</i> “Users with similar activity levels typically require 2.2–2.4 L of water.”
Rule-based	Explain the decision through explicit rules. <i>Example:</i> “High temperature + exercise → increase hydration target.”
Data Transparency	Show which data were used to generate the recommendation. <i>Example:</i> “Weather, activity level, and hydration history were used.”

User Information Needs

While explanation types describe the information that AI systems can provide, they do not indicate which explanations are relevant in a particular interaction. Human-centred XAI research therefore shifts the focus from explanation methods to the information needs of users. Rather than asking which explanation technique should be used, designers first consider what users want to know when interacting with an AI system. Liao et al. [10, 11] proposed a question-driven perspective in which users seek explanations through a recurring set of information needs. These include understanding why a decision was made (*Why?*), why an alternative outcome was not selected (*Why not?*), how the system behaves under different conditions (*What if?*), how a decision was produced (*How?*), and whether the result can be trusted (*Can I trust this?*). Together, these questions provide a practical way of identifying the explanations that should be communicated to users.

For example, in a Health Behaviour Change application, a user receiving the recommendation “*Drink 2.3 L of water today*” may naturally ask several questions. A *Why?* question seeks the factors that influenced the recommendation, while a *What if?* question explores how the recommendation would

change if different conditions were met. A *Can I trust this?* question focuses on the confidence of the prediction, whereas a *How?* question aims to understand the reasoning process behind the recommendation.

From a design perspective, these questions provide an important starting point because they express the user's information needs independently of any particular explanation method. In the following sections, these information needs are combined with explanation types, levels of user understanding, and interface design guidance to support the practical design of explainable interfaces.

Levels of User Understanding

While user questions describe the information users seek from AI systems, they do not indicate the level of understanding that explanations should support. The SAFE-AI framework addresses this challenge by organising explainability according to three levels of human situation awareness: *perception*, *comprehension*, and *projection* [15]. Rather than focusing on explanation methods, SAFE-AI considers how explanations contribute to users' understanding and decision-making.

At the *perception* level, explanations help users recognise and interpret the current state of the AI system. Typical information includes the system output, confidence, or uncertainty. The *comprehension* level focuses on helping users understand why a particular decision was made by revealing the factors, rules, or evidence that influenced the result. Finally, the *projection* level supports anticipation and action by enabling users to explore alternative outcomes, understand potential consequences, and identify appropriate next steps.

Returning to the Health Behaviour Change example, the recommendation "*Drink 2 L of water today*" first requires users to recognise the recommendation and assess its reliability, corresponding to the *perception* level. Understanding why this recommendation was generated. For example, because of yesterday's hydration level, current weather conditions, or recent physical activity — corresponds to the *comprehension* level. Finally, exploring how different behaviours might influence future recommendations, such as increasing water intake before exercise, represents the *projection* level.

From a design perspective, these three levels provide a cognitive structure for organising explanations according to the type of understanding they support. In this work, SAFE-AI is used as an organisational principle for structuring explainability concepts within the proposed repository rather than as a taxonomy of explanation types.

Explainable Interfaces

While explanation types, user information needs, and levels of understanding describe *what* should be explained, they provide limited guidance on *how* explanations should be presented to users. Explainable Interfaces (EI) address this challenge by treating explanations as part of the user experience rather than as isolated outputs of AI models. Research in Human-Computer Interaction has therefore shifted attention from explanation generation towards explanation presentation, interaction, and evaluation [7, 12].

Explainable Interfaces investigate how explanations can be integrated into user interfaces through appropriate visualisations, interaction patterns, and user flows. Rather than exposing raw outputs produced by XAI methods, EI focuses on presenting explanations in ways that support usability,

trust, understanding, and decision-making. Typical interface patterns include confidence indicators, explanation panels, interactive visualisations, recommendation cards, timelines, expandable explanations, and contextual guidance.

Returning to the Health Behaviour Change example, the recommendation "*Drink 2 L of water today*" can be presented through multiple interface components. A confidence badge communicates the certainty of the recommendation, an expandable explanation panel presents the most influential factors, while a suggestion card can provide counterfactual guidance describing how different behaviours may influence future recommendations. These interface components transform technical explanations into information that users can interpret and act upon.

From a design perspective, Explainable Interfaces contribute practical knowledge about how explanations should be communicated, structured, and evaluated. However, existing research primarily investigates interface examples and evaluation methods rather than providing designers with a reusable resource for selecting and applying explainability concepts during interface design. This remaining gap motivates the structured design repository proposed in this work.

Positioning of the Proposed Work

The literature reviewed in the previous sections shows that explainable AI has evolved through several complementary research perspectives. Existing work has identified explanation types that AI systems can generate, user information needs that motivate explanations, cognitive models describing the levels of understanding explanations should support, and interface research investigating how explanations should be presented and evaluated. Although each perspective contributes valuable knowledge, they are typically presented independently and therefore provide limited practical support for interface design.

This work positions itself as an integration of these complementary perspectives. Rather than proposing new explanation methods, user information needs, or cognitive models, we organise selected explainability knowledge into a structured design repository intended to support interface design. The repository combines explanation types, user information needs, levels of understanding, interface guidance, implementation examples, and practical design considerations into a reusable design resource that assists designers in selecting and applying explainability concepts during the design of explainable interfaces.

STRUCTURED DESIGN REPOSITORY

The primary intended users of the repository are UX, product, and interaction designers responsible for designing user-facing AI experiences.

Repository Entry Structure

The previous section reviewed complementary perspectives on explainability, including explanation types, user information needs, levels of understanding, and explainable interface design. Although these perspectives provide valuable knowledge, they are typically presented independently, requiring designers to integrate them manually when designing explainable interfaces.

To address this gap, this work proposes a structured design repository that operationalises explainability knowledge into

reusable repository entries. Rather than introducing new explanation methods, the repository synthesises complementary theoretical perspectives into a unified design artefact that supports the selection and implementation of explainability concepts during interface design.

Each explanation type is documented as a structured repository entry using a common representation (Figure 1). The representation combines theoretical knowledge with reusable design guidance, enabling explainability concepts to be consistently documented and applied across application domains. Each repository entry contains the following elements:

- **Explanation Type** identifies the explainability concept represented by the entry.
- **Design Purpose** describes the role of the explanation from an interface design perspective.
- **Typical User Question** links the explanation type to the user information need it addresses.
- **Level of Understanding** specifies the cognitive level supported by the explanation according to the SAFE-AI framework.
- **Design Recommendations** provide reusable guidance for implementing the explanation within user interfaces.
- **Example Interface** illustrates one possible interface implementation of the explanation concept.

Example Repository Entry

Figure 1 presents an example repository entry for the *Confidence* explanation type. The example demonstrates how complementary explainability knowledge is synthesised into a reusable design artefact. While the first four elements capture complementary explainability knowledge synthesised from the literature, the design recommendations translate these concepts into actionable interface design guidance. The example interface illustrates one possible implementation without prescribing a domain-specific solution.

Repository Prototype

A functional prototype of the proposed repository was implemented in Airtable to demonstrate the feasibility of the approach. The prototype stores each explanation type as a structured repository entry, allowing designers to browse, organise, and retrieve explainability knowledge during interface design. Figure 2 presents the current implementation of the repository.

REPOSITORY DEMONSTRATION

The proposed repository was demonstrated through the design of an explainable mobile interface for a health coaching application. The application generates personalised daily activity recommendations using behavioural data collected from the user.

The repository was consulted to identify explanation types suitable for supporting user understanding throughout the interaction. Three repository entries were selected: *Confidence*, *Feature Importance*, and *Counterfactual Explanation*. Together, these explanation types support the three SAFE-AI levels of understanding by communicating the certainty of the recommendation, explaining the factors that influenced it, and illustrating how alternative user actions could improve the predicted outcome.

EXPLANATION TYPE

Confidence

PURPOSE

Communicate the certainty of an AI prediction.

TYPICAL USER QUESTION

How certain is the system?

LEVEL OF UNDERSTANDING

Perception

DESIGN RECOMMENDATIONS

- Prefer qualitative labels.
- Avoid implying certainty equals correctness.
- Use consistent visual encoding.
- Explain low-confidence predictions.
- Match detail to the user's expertise.

EXAMPLE INTERFACE

 Recommended Option

High confidence

Based on the available input information.

Figure 1: Example repository entry for the *Confidence* explanation type.

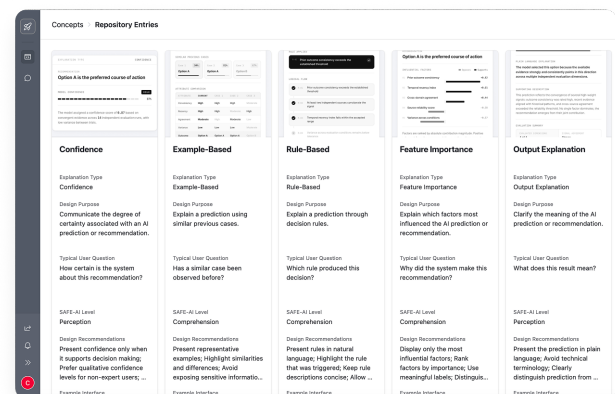


Figure 2: Prototype implementation of the structured design repository.

Figure 3 presents the resulting interface. The interface demonstrates how generic repository entries can be instantiated within a specific application context while preserving the underlying design recommendations captured in the repository.

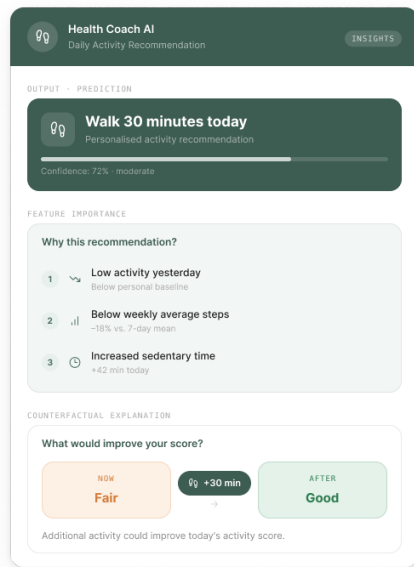


Figure 3: Explainable HBC interface designed using repository guidance that integrates three Explainability Concepts: Output, Feature Importance, and Counterfactual Explanation.

Preliminary Validation

Participants and Procedure

A preliminary expert review was conducted with five UX/HCI professionals with 6–15 years of professional experience across healthcare, finance, and enterprise software. Participants represented the intended professional audience of the repository and had experience designing digital products and user interfaces.

Each participant attended a moderated 45-minute session consisting of: (1) a brief introduction to the repository, (2) independent exploration of selected repository entries, (3) a short design-oriented reflection task, and (4) a semi-structured discussion. Participation was voluntary, informed consent was obtained, and responses were anonymised.

Data Collection and Analysis

Participants were asked to consider the following task:

“Imagine you are designing an AI interface. How would you use this repository?”

While exploring the repository, participants verbalised their thoughts and discussed where such a resource might fit within their design process. The semi-structured discussion focused on four areas: perceived usefulness, clarity of the repository structure, relevance to AI interface design, and opportunities for integration into existing design workflows. Participants were also encouraged to identify missing information or limitations. The collected feedback was reviewed to identify recurring observations across participants.

Findings

Three themes emerged from the analysis.

Repository provides a valuable starting point. Participants perceived the repository as a useful starting point when approaching the design of AI-powered interfaces. Rather than replacing existing UX methods, they viewed it as a resource

for understanding available explainability mechanisms before beginning interface design.

“I’d probably open this to understand what explanation options I have before sketching anything.” (P2)

Repository highlights the unique challenges of AI interface design. Participants reported that the repository made explicit several design considerations that they would not necessarily encounter when applying conventional interface design guidance. Confidence, uncertainty, explanation strategies, and system limitations were repeatedly identified as concerns that become particularly relevant when designing AI-powered interfaces. Participants therefore viewed the repository not only as a reference for selecting explanations, but also as a reminder of AI-specific considerations that should be addressed during design.

“It reminded me that AI interfaces require a different way of thinking.” (P4)

Integration into existing design workflows is desirable. Participants suggested that the repository would be most useful if integrated into existing design environments such as Figma or design systems.

“If this was available inside Figma or our design system, I’d definitely use it during projects.” (P1)

IMPLICATION AND LIMITATIONS

The preliminary findings provide initial qualitative support for the integration of the theoretical perspectives used to construct the repository. Participants’ description of the repository as a useful starting point aligns with the question-driven perspective of Liao et al., in which user information needs provide an entry point for selecting appropriate explanations. Within the repository, these questions are connected to explanation concepts and implementation guidance, allowing designers to move from a user information need to a potential interface representation.

Participants also responded positively to the organisation of concepts around different levels of user understanding. This suggests that the perception, comprehension, and projection levels proposed by SAFE-AI can function not only as a model of situation awareness, but also as an organising structure for designer-facing explainability knowledge.

A further observation was that the repository surfaced AI-specific design considerations, including confidence, uncertainty, and explanation strategies, that participants did not typically associate with conventional interface design guidance. This reinforces the human-centred XAI position that explainability introduces interaction concerns beyond the presentation of conventional system information.

Finally, participants’ preference for integration with tools such as Figma and organisational design systems suggests that explainability guidance may be more useful when embedded within existing design workflows rather than provided only as a standalone repository. This provides a practical direction for future development of explainability-aware design tools.

The evaluation was exploratory and assessed perceived usefulness rather than whether use of the repository improves design quality or user outcomes. The findings should therefore be interpreted as preliminary evidence of perceived relevance and applicability. The study involved five participants and a single demonstration domain. Future evaluations should compare design activities conducted with and without the

repository, examine resulting interface designs, and evaluate whether those interfaces improve user understanding in additional application domains.

CONCLUSION

This paper presented a structured design repository for explainable interfaces that translates explainability knowledge into reusable design guidance. By integrating explanation types, user information needs, SAFE-AI levels, design recommendations, and example interfaces into structured repository entries, the proposed approach bridges explainability research and interface design.

The repository was implemented as a functional prototype and demonstrated through the design of an explainable health behavior change mockup. A preliminary expert review with five UX/HCI professionals indicated that participants perceived the repository as a useful starting point for AI interface design and as a resource for making AI-specific design considerations more explicit. The evaluation assessed perceived usefulness rather than improvements in design outcomes, which remain to be investigated in future studies.

Future work will focus on expanding the repository with additional explanation types, validating the design recommendations across multiple domains, and integrating the repository into existing design environments and AI-assisted design workflows.

References

- [1] Amina Adadi and Mohammed Berrada. 2018. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* 6 (2018), 52138–52160. doi:10.1109/ACCESS.2018.2870052
- [2] Ciprian Amaritei and Adrian Iftene. 2025. VITAL: A Design Framework for Prototyping Health Behavior Change Applications. *Procedia Computer Science* 270 (2025), 2640–2649.
- [3] Ciprian Amaritei, Dragos Silion, and Adrian Iftene. 2026. Agent-Based Operationalisation of Health Behaviour Change Interventions. *International Journal of User-System Interaction (IJUSI)* 18 (2026), 24–37. Issue 1.
- [4] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, et al. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, United States, 1–13. doi:10.1145/3290605.3300233
- [5] Alejandro Barredo Arrieta et al. 2020. Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. *Information Fusion* 58 (2020), 82–115. doi:10.1016/j.inffus.2019.12.012
- [6] Vinay Chamola, Vikas Hassija, A. Razia Sulthana, Debshishu Ghosh, Divyansh Dhingra, and Biplab Sikdar. 2023. A Review of Trustworthy and Explainable Artificial Intelligence (XAI). *IEEE Access* 11 (2023), 78994–79015.
- [7] Michael Chromik and Andreas Butz. 2021. Human-XAI Interaction: A Review and Design Principles for Explanation User Interfaces. In *Human-Computer Interaction – INTERACT 2021*. Springer, Cham, 619–640.
- [8] Tim Hulsen. 2023. Explainable Artificial Intelligence (XAI): Concepts and Challenges in Healthcare. *AI 2023* 4 (2023), 652–666.
- [9] Khushi Kalasampath, K. N. Spoorthi, Sreeparvathy Sajeev, Sahil Sarma Kuppa, Kavaya Ajay, and Angulakshmi Maruthamuthu. 2025. A Literature Review on Applications of Explainable Artificial Intelligence (XAI). *IEEE Access* 13 (2025), 41111–41140.
- [10] Q. Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: Informing Design Practices for Explainable AI User Experiences. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, New York, United States, 1–15.
- [11] Q. Vera Liao, Milena Pribić, Jaesik Han, Sarah Miller, and Daby Sow. 2021. Question-Driven Design Process for Explainable AI User Experiences. *arXiv:2104.03483* 2104.03483 (2021), 1–23. arXiv:2104.03483
- [12] Thu Nguyen, Alessandro Canossa, and Jichen Zhu. 2024. How Human-Centered Explainable AI Interfaces Are Designed and Evaluated: A Systematic Survey. *arXiv:2403.14496* 2403.14496 (2024), 1–22. arXiv:2403.14496
- [13] Maria-Ecaterina Olariu and Adrian Iftene. 2025. Analysis of Current Applications of Artificial Intelligence in Cardiology Diagnosis with Focus on Explainable AI, Holter Monitoring and Blood Biomarkers. *Proceedings of the 19th International Conference on INnovations in Intelligent SysTems and Applications (INISTA 2025)*, 29-31 October, 2025 1 (2025), 1–10.
- [14] Evandro S. Ortigossa, Thales Gonçalves, and Luis Gustavo Nonato. 2024. Explainable Artificial Intelligence (XAI): From Theory to Methods and Applications. *Comput. Surveys* 12 (2024), 80799–80846.
- [15] Lindsay Sanneman and Julie A. Shah. 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 (2022), 1772–1788.