Where Are We and Where Can We Go on the Road to Reliance-Aware Explainable User Interfaces?

José Cezar de Souza Filho, Rafik Belloum, Káthia Marçal de Oliveira

Where Are We and Where Can We Go on the Road to Reliance-Aware Explainable User Interfaces?

José Cezar de Souza Filho
Univ. Polytechnique Hauts-de-France
LAMIH, UMR CNRS 8201, F-59313
Valenciennes, France
josecezar.juniordesouzafilho@uphf.fr

Rafik Belloum
Univ. Polytechnique Hauts-de-France
LAMIH, UMR CNRS 8201, F-59313
Valenciennes, France
rafik.belloum@uphf.fr

Káthia Marçal de Oliveira
Univ. Polytechnique Hauts-de-France
LAMIH, UMR CNRS 8201, F-59313
Valenciennes, France
kathia.oliveira@uphf.fr

Abstract

With the widespread use of machine learning for algorithmic decision-making, Explainable Artificial Intelligence (XAI) systems are increasingly important. However, they are not always understandable and trustworthy for end-users, who should know when to rely on AI's advice to make informed decisions. Hence, adequately presenting explanations in user interfaces (UIs) is essential. This paper investigates proposals in this direction and what is missing to support the design of explainable UIs aware of reliance issues. From a systematic literature review of 1,287 unique papers, we identified 120 secondary studies, of which we selected 22 for analysis. Our findings reveal that studies have been conducted to provide recommendations to specific application domains, and evidence regarding explanation effects on reliance remains inconclusive. We also found a lack of characterization of factors impacting reliance on XAI systems. Furthermore, we provide perspectives to foster appropriate reliance research on explainable UIs.

Key Words

human-computer interaction, explainable artificial intelligence, appropriate reliance, tertiary study

I. Introduction

Artificial Intelligence (AI) applications have gained even more value in recent years driven by advancements in Explainable AI (XAI), which seeks to make opaque models more transparent and understandable [1]–[3]. Prior works in human-centered XAI show users of these systems tend to rely on AI's advice even when it is wrong [4]. The associated explanations often contribute to this over-reliance [5]. Thus, numerous studies have emphasized the importance of calibrating trust to improve XAI-assisted decision-making (cf. [6], [7]).

In line with this, the Human-Computer Interaction (HCI) community has discussed applying human-centered approaches to design, evaluate, and provide conceptual and methodological tools for XAI systems [8]. However, there is still a lack of design concepts and HCI strategies to support appropriate reliance on XAI systems, i.e., "users should know when to trust the AI's advice and when to be cautious" [8, p. 8] to avoid over- and under-reliance, for instance, given the scenario of a loan approval process [9]. Loan officers have an AI decision aid system that supports the approval decision by providing a prediction (granted or refused), a confidence value, and a feature-based, visual explanation of that prediction. The explanation's persuasiveness may lead officers with a low need for cognition to overestimate the AI's advice, even when wrong. Also, skeptical officers can under-rely on AI's advice, even when correct. Hence, loan officers can make wrong decisions that they would have done without the XAI's help, and this financially affects bank customers.

To address this, it is essential to move forward with characterizing reliance on XAI-assisted decision-making. For this paper, *reliance-awareness* refers to the capacity of user interface design to support the appropriate reliance of users in detecting errors in XAI recommendations, thereby preventing over-reliance. It also involves recognizing correct recommendations to avoid under-reliance, ultimately supporting informed decision-making. This notion differs from the investigations on context-awareness [10] since our perspective is to identify attributes related to reliance and types of explanations that provide appropriate reliance in specific usage scenarios. Grounded on this, designers and researchers can reason about the reliance phenomenon when designing and evaluating explainable user interfaces

(UIs) so that explanations become aware of reliance. Hence, reliance-awareness considers reliance issues in design time instead of considering runtime adaptation according to environmental or user properties.

In this paper, we carried out a tertiary study, a type of Systematic Literature Review (SLR) for mapping secondary studies (e.g., reviews, state-of-the-art) to categorize them and observe trends regarding a research topic [11]. This study aims to consolidate the evidence revealed by secondary studies in the human-centered XAI field toward characterizing the role of appropriate reliance in explainable UIs. Our paper contributes a *status quo* organized into five categories: Application Domains; Design and Evaluation; Transparency and Trust; Approaches and Methods; and Human-AI Interaction. It reveals (i) over-/under-reliance, misuse, disuse, and abuse as the primary constructs of appropriate reliance; (ii) autonomously delivering the explanation, in which the system decides the timing to present it (e.g., before, with, or after prediction), and visual explanations can contribute to an appropriate reliance; and (iii) system reliability, task complexity, and task/system experience as potential reliance impact factors.

II. Overview on Explainable AI (XAI)

With the growing adoption of machine learning and black-box algorithms, a new wave of interest in XAI emerged [1]. The DARPA's initiative to fund an XAI program [2], ethical concerns, and a lack of user trust [12] have driven this resurgence. The perspective for XAI systems is that they will provide explanations (i.e., "explicitly explaining decisions to people" [12, p. 1]) to end-users so that they can understand the system's strengths and weaknesses, its behavior in the future or alternative situations, and may provide corrections for system's mistakes [2]. These systems are composed of software and hardware (e.g., sensors) that apply AI algorithms (e.g., decision trees, Deep Neural Networks (DNNs)) to make decisions, which are explained by XAI techniques (e.g., LIME, SHAP). For the end-users to easily understand it, explainable user interfaces are designed and implemented (cf. [2]).

To that end, Clement et al. [13] establish that the design of XAI systems involves two main steps: (i) *explanation design*, which covers the process of selecting one or more suitable techniques to generate appropriate explanations, depending on the requirements and the application; and (ii) *explainable user interface design* to define how to present the explanations to the application's end-users. Researchers have proposed a plethora of XAI techniques to support those steps: feature importance methods (e.g., LIME, LORE, DeepLift), white-box models (e.g., rule and tree extraction, attention network), example-based (e.g., prototypes), and visual explanations.

An important challenge for phase (ii) is providing explainable UIs that help end-users make informed decisions with appropriate confidence. In this context, researchers usually apply two terms: trust and reliance. Lee and See [14] defined *trust* in automation as an attitude and distinguished it from *reliance* as a behavior. Based on that, Scharowski et al. [15, p. 3] defined "reliance on a system as a user's behavior that follows from the advice of the system." However, trust and reliance have been usually grouped under the term trust [16].

III. Tertiary Study Planning and Execution

We structured our goal based on [17], as follows: analyze *reliance* for the purpose of *characterizing* with respect to its *explanation designs*, *concepts*, and *impact factors* in the context of *XAI-assisted decision-making* from the point of view of *HCI researchers*. Table I outlines our research protocol.

To define the *search string*, we considered terms from prior surveys [18]–[20] and followed PICO [21], searching on Explainable AI studies (*Population*), focused on appropriate reliance (*Intervention*) to get the applied explanation designs, HCI, and impact factors (*Outcome*). We did not apply *Comparison*, as there is no baseline to compare our tertiary study with other secondary studies. As *strategy*, we used Web of Science and Scopus databases due to their significantly more coverage of HCI literature and indexing of computer science databases, such as IEEE Xplore, ACM DL, Springer, and Elsevier (cf. [22]). Finally, the *inclusion and exclusion criteria* consider our focus on XAI-assisted decision-making and the RQs, limiting the search for papers published from 2014 (in line with human-centered XAI surveys [20], [23], [24]) and considering that the number of papers on this topic has been expanding since 2016 [24], including in the HCI community [1]).

One Ph.D. student and two professors working with HCI executed the planned protocol. The selection of papers made by one researcher was peer-reviewed by at least one of the other two researchers to mitigate interpretation bias. We used MS Excel spreadsheets to manage the screening of papers and data extraction. We conducted the database search[1] in digital libraries on 31 January 2024, automatically filtering by EC1, EC2, EC5 (Table I), and the computer science area to focus on computing aspects related to reliance in HCI. We found 1,604 papers (1,212 from Scopus and 392 from Web of Science), of which we excluded 317 duplicate records and included 1,287 for the subsequent two iterations of paper screening based on the defined inclusion and exclusion criteria (Fig. 1).

[1] Database search is available through an online repository [25].

TABLE I
PROTOCOL SUMMARY FOR THE TERTIARY STUDY

Research Questions	RQ1. What do secondary studies in XAI-assisted decision-making discuss on reliance?	
	RQ2. What are the types of explanation designs used to provide reliance in XAI-assisted decision-making?	
	RQ3. What are the factors impacting reliance in XAI-assisted decision-making?	
Search String	***Population***	(“explainab* artificial intelligence” OR “explainab* AI” OR explanation* OR XAI OR “transparen* artificial intelligence” OR “transparen* AI” OR “interpretab* artificial intelligence” OR “interpretab* AI” OR “intelligib* artificial intelligence” OR “understandab* artificial intelligence” OR “comprehensib* artificial intelligence” OR “explainab* system*” OR “interpretab* system*” OR “intellig* system*” OR “machine learning” OR “decision-making algorithm*”) AND
	Intervention	(reliance OR underreliance OR under-reliance OR overreliance OR over-reliance OR trust* OR distrust* OR overtrust* OR reliab* OR ”algorithm aversion”) AND
	Comparison	*Not applicable in our study*
	Outcomes	(“explanation* interface” OR “explanation* design” OR “explainab* interface” OR “explainab* design” OR “user interaction” OR “user experience” OR “UX” OR “user interface” OR UI OR “application interface” OR “human-machine interface” OR “human-machine interaction” OR “human-computer interface” OR “human-computer interaction” OR HCI OR “human-AI interface” OR “human-AI interaction” OR “interaction design” OR “user-centered design” OR “interactive system” OR “impact* factor” OR “influenc* factor” OR “human factor” OR “design factor” OR “technical factor” OR “organization* factor” OR “manag* factor” OR “cognitive bias*” OR “human bias*” OR “automation bias*”)
Sources	Database search on Scopus and Web of Science (WoS).	
Inclusion Criteria	IC1. The secondary study’s context is XAI-assisted decision-making;	
	IC2. The secondary study discusses definition(s) regarding reliance in the XAI context;	
	IC3. The secondary study discusses explanation designs regarding reliance in the XAI context;	
	IC4. The secondary study discusses factor(s) impacting reliance in the XAI context.	
Exclusion Criteria	EC1. The paper is not written in English;	
	EC2. The paper was published before 2014;	
	EC3. The paper is duplicated;	
	EC4. The paper provides a primary study;	
	EC5. Books (except conference papers published as book chapters), editorials, summaries of workshops, tutorials, and keynotes, gray literature, and other non-peer-reviewed papers;	
	EC6. The paper’s full text is not available in our institution and not available from authors;	
	EC7. The secondary study’s context is not related to XAI-assisted decision-making;	
	EC8. The secondary study’s context is related to XAI but without a focus on HCI;	
	EC9. The secondary study’s contribution is regarding AI experts;	
	EC10. The secondary study’s context is regarding physical human-robot interaction (except papers on robot-advisor).	
Extraction	Research goal, research questions, research method, time range used to select papers, amount of accepted papers, and main findings.	

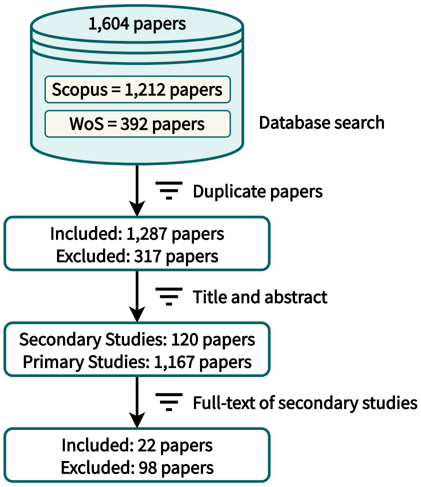

Fig. 1. Overview of papers screening for the tertiary study.

First, we analyzed only each paper’s title and abstract. Out of 1,287, we identified 120 secondary studies and excluded 1,167 papers as primary studies (EC4) for our tertiary study. Then, applying the inclusion and exclusion criteria on the full texts, we selected 22 papers for data extraction and analysis.

IV. Findings: Where We Are

Out of 22 selected secondary studies, eight were published in 2023, six in 2022, four in 2021, two in 2020, and only one in 2014. We noted a gap between 2014 and 2020, when the number of secondary studies became an increasing trend (20 publications from 2020 to 2023), indicating a resurgence of interest related to reliance on XAI systems. We also highlighted that the small number of papers for 2024 (only one) is acceptable since we did our search on January 31st.

Twelve (12) of 22 studies focus on specific application domains related to XAI (see Table II), where the most cited domains are intelligent systems and recommender systems.

TABLE II
Secondary Studies According to Application Domains

Application Domain	# Studies	References
Intelligent systems	3	[26]–[28]
Recommender systems	3	[29]–[31]
Autonomous driving	2	[32], [33]
Healthcare	2	[34], [35]
Intelligent decision support systems	1	[36]
Judgmental forecasting	1	[37]
Total	12	

Most secondary studies (15) do not explicitly define a time range and the number of analyzed papers. Thus, it is not possible to assert paper coverage across them. Fourteen (14) studies presented the number of papers selected and analyzed, varying from 13 to 97 publications.

Regarding the type of method applied, most secondary studies (16) did not follow any systematic approach; two did a SLR [28], [38]; another two conducted a SLR with snowballing procedures [39], [40]; one applied only the snowballing [27]; and one did a scoping literature review [41]. Most studies (12) were published as journal articles, followed by six conference papers, two workshop papers, and two book chapters. We also noted that most journal articles (11) are non-systematic literature reviews.

Regarding the research goals, we observed that secondary studies focus on different perspectives to navigate the XAI field, which we divided into five categories, as follows: *Application Domains* includes studies on how to apply explainability in a specific domain [29]–[31], [35]; overview of a specific domain related to XAI [32], [36], [37]; effectiveness of user interfaces in a specific domain [33]; and overview related to trust in a specific domain [34]. *Design and Evaluation* with studies regarding implementation and evaluation of explainable models [38]; design goals and evaluation methods in XAI [42]; human-centered evaluations [43] and user studies in XAI [44]; *Transparency and Trust* includes studies on elements influencing transparency in machine learning systems [45]; factors influencing trust in XAI systems [41]; and the relationship between transparency and trust in AI [46]. *Approaches and Methods* presents studies on approaches to personalizing explanations in intelligent systems [27]; existing approaches in human-centered XAI [39]; explanation delivery methods, interface modalities, and potential risks with explanations [28]; and overview on a specific kind of method in XAI [47]. *Human-AI Interaction* includes studies on users' characteristics who interact with intelligent systems [26] and AI system communication for end-users [40].

A. RQ1: Reliance-Related Discussions

Since the terms "trust" and "reliance" are interchangeably used in the literature [15], we searched for what secondary studies have discussed on trust calibration (also known as appropriate reliance). In general, the studies point out the importance of trust in XAI systems and how explanations may influence user's trust. For instance, Buschek et al. [26, p. 19] stated that "trust has been recognized as a crucial aspect of interaction with intelligent systems" and provides questions related to trust in their explainability user questions, such as "can I trust this model?" and "should I trust this prediction?".

Horvat et al. [38, p. 2] define trust as "the user's confidence that a model will act as intended." Khoozani et al. [47, p. 4] argue that it is "a multifaceted concept that includes technical knowledge, transparency, and ethical considerations." Sperrle et al. [43, p. 552] stated that the model's trustworthiness is different from the explanation's trustworthiness since "a model that performs poorly may be considered untrustworthy, but an explanation of that model may still be highly accurate and considered trustworthy." User's trust in the algorithm is associated with a reliance on the system [32], [42], and explanation capabilities can increase trust and reliance [35]. Only one secondary study discusses a definition for reliance, focusing on the context of judgmental forecasting [37, p. 103]: "the extent to which forecasters decide to follow forecast suggestions provided by computers based on statistical analyses."

Sperrle et al. [43] and Rong et al. [44, p. 2108] draw attention to the *persuasive power* of model explanations: “the capacity to convince users to follow model decisions despite its correctness.” They argue that good explanations should *calibrate user’s trust*, i.e., trust only correct advice and distrust it otherwise. Naiseh et al. [28] stated that over-trust and under-trust are potential risks coming from explanations. *Over-trust* means “a high agreement rate to wrongly made decisions,” whereas *under-trust* represents “a low agreement rate to correct decisions” [44, p. 2108]. Trust calibration is important to enhance *user engagement* and avoid misuse, disuse, and abuse of the system [41]. *Misuse* means “a higher reliance than deserved by a system”; *disuse* is defined as “a reliance in a system that is lower than appropriate given the actual system performance”; and *abuse* refers to “a poor design and implementation of automation, disregarding its effects on human operators” [37, p. 105]

Naiseh et al. [28] discuss solutions to reduce over-trust (or over-reliance), such as interactive explanations, personalized explanations based on the user’s personality, and uncertainty. However, they argue that existing solutions to reduce over-trust need further research to investigate the relationship between trust, certainty level, cognitive styles, personality, and liability, as well as to take into account usability and user experience factors, for instance, timing, level of details, user feedback, and the explanation evolution. They also see over-trust as a property that emerges over time since “users may over-trust a system due to *cognitive anchoring* and *overconfidence bias* when it proves correct in many previous occasions” [28, p. 8].

Naiseh et al. [28] suggested that under-trust is associated with the *user’s personality*, and explanations should be designed according to natural human interaction patterns to approach trust as people would in real contexts. They discuss that providing more details in an explanation may not necessarily increase trust, and explanations that use information derived about users could generate algorithm disillusionment for them. Zerilli et al. [46] also stated that *information overload* (i.e., excessive transparency) could cause under-trust, and poor or confusing explanations can generate algorithm aversion.

Take way from RQ1. Good explanations provide trust calibration to enhance user engagement and avoid misuse, disuse, and abuse. Persuasive power is a property of explanations that may lead to over-trust. Cognitive anchoring and overconfidence bias may cause over-trust, which can be approached, for instance, through interactive explanations and error presentation. Information overload may cause under-trust, which is associated with the user’s personality and can be mitigated through natural human interaction patterns.

B. RQ2: Explanation Designs and Reliance

To answer RQ2, we searched for empirical findings on the relationship between explanations and appropriate reliance. Rong et al. [44] found a mixed effect of explanations on user’s trust, in which half of the analyzed papers validate that explanations positively impact trust, but the other half cannot confirm the same hypothesis. They also found that explanations improved trust based on simulated data but not with real-world data for an autonomous driving task, as well as minimal evidence that feature-based explanations help increase appropriate trust, whereas they did not find it for counterfactual explanations. This finding reinforces the argument that empirical evidence on this topic is inconclusive [15], [48].

Morandini et al. [41] observed that *low-fidelity explanations*, low usefulness perceptions, and fear or discomfort can decrease trust. They also found that Partial Dependence Plot (PDP) and Local Interpretable Model-agnostic Explanations (LIME) have high levels of agreement among participants, revealing an increased trust, contrary to Shapley Additive Explanations (SHAP), which had less effectiveness.

Regarding concept-supported reasoning for explaining black-box decisions, Khoozani et al. [47] observed that experts or decision-makers tend to trust the model’s decisions because its reasoning matches their domain knowledge and expectations. Mohseni et al. [42] also noted that providing *explanations of facts* contributes to a higher user’s trust and reliance on a clinical decision-support system.

Laato et al. [40] found an increased trust in explanations provided by *virtual agents* compared, for instance, with only text- or voice-based explanations, and Alvarado-Valencia and Barrero [37] observed that *textual explanations* increase trust in system advice. These authors also argue that explanations are influenced by their *informative value*, which depends on the wording and format to deliver the explanation. In this sense, Naiseh et al. [28] found that *autonomously delivered explanations* (i.e., when the system has the autonomy to define the time and context to deliver the explanation) help trust calibration and avoid over- and under-trust situations.

Regarding trust calibration, Mohseni et al. [42] found that presenting model prediction confidence scores to users affects calibrating trust. Morandini et al. [41] observed that *visual explanations* help users calibrate their trust by providing additional, trusted information without over-trusting the system. Naiseh et al. [28] also noted that users with a high need for cognition faced under-trust issues with explainable recommendations, whereas explanations increased trust in users with a low need for cognition.

Take way from RQ2. Explanations delivered autonomously and visual explanations contribute to trust calibration. Textual explanations and those provided by virtual agents increase trust, whereas low-fidelity explanations can decrease it. PDP and LIME showed an increased trust compared to SHAP explanations.

C. RQ3: Reliance Impact Factors

To answer RQ3, we searched for empirical findings regarding other factors (beyond explanations) impacting appropriate reliance. In this context, Alvarado-Valencia and Barrero [37] present a set of impact factors associated with reliance in the context of judgmental forecasting, divided into factors that *increase reliance* (related to a decrease in attention paid to tasks): increased task complexity, high system reliability, increased workload, and increased system experience; factors that *decrease reliance*: low system reliability, increased human accountability, negative attitudes toward the system combined with extreme subjective norms, and experience in the task; and factors that *help an adequate reliance level*: higher computer self-efficacy and updated performance information of the system. Another study [46] observed that metainformation about low-reliability automation has a risk of generating over-trust since higher trust ratings measure it; however, meta-information about high-reliability automation indicates an opposite effect. We found only these two secondary studies that provide evidence on reliance impact factors, which focus on the automation context and do not take into account the recent advancements in the XAI field, which began in 2017 (cf. [23]).

Take way from RQ3. Research in automation suggests system reliability, task complexity, and experience, either with the system or in the task, as factors that modulate reliance, whereas providing system performance information and a higher computer self-efficacy contribute to an appropriate reliance.

V. Discussion: Where We Can Go

We found no study characterizing reliance on explainable user interfaces. Most studies (16) also do not focus on understanding what and how explanations affect appropriate reliance. Only two studies presented a research goal related to the effects coming from explanations, but regarding a broad view of trust, and focused only on empirical studies in XAI [41] or user study design for XAI research [44].

Beyond the choice of explanation method, we need to identify to what extent other factors can impact user's reliance, such as environmental, human, and technical factors, contextual information, timing, framing, and training. Despite the existing solutions, for instance, to reduce over-reliance [28], there is a lack of investigations on their suitability according to quality-in-use attributes (e.g., usability and user experience) and different explainability needs, as well as what explanation format is adequate for each usage scenario. For instance, existing studies (e.g., Szymanski et al. [49]) primarily focus on assessing user understanding and satisfaction with different explanation formats but do not investigate their impact on user's reliance. Thus, there is a need to understand how to tailor different explanation formats (e.g., static, interactive, textual, and visual [20]) to support the appropriate reliance of end-users. It is also necessary to focus on what type of explanation designs are proper for different user contexts and needs, AI tasks, and XAI systems types, providing a conceptual mapping on the interplay between explainable user interfaces and appropriate reliance.

According to Bertrand et al. [19], certain types of bias lead to over-reliance (e.g., confirmation, automation, and recognition bias). Identifying what biases are linked to under-reliance and dealing with these biases as another type of impact factor helps clarify the arguments on inconclusive evidence about appropriate reliance. Therefore, it still lacks a complete characterization of reliance concerning its concepts, related elements/attributes, explanation designs, and impact factors.

What is missing. A complete identification of reliance relation with different types of explanation designs, reliance attributes, and impact factors on reliance in order to better support the design of XAI user interfaces aware of reliance issues, as well as empirical studies regarding quality-in-use attributes and how to personalize explanations according to the end-users.

VI. Final Remarks

This paper has presented a tertiary study on reliance issues in the context of XAI-assisted decision-making systems. We concluded that although several studies have explored trust/reliance issues in XAI systems, there is still a lack of in-depth research into the particularities of reliance that should be considered in the design of explanations in user interfaces looking to provide the appropriate reliance to the non-expert user on AI.

We analyzed the threats to the validity of our study according to [50]. Formalizing an adequate protocol and data extraction by peer review and consensual meetings mitigate construct, internal, and conclusion validity. However, potential threats to *external validity* exist, as we considered only two databases (Scopus and Web of Science). These databases are widely used for SLRs and cover several publishers in computer science. Thus, we consider the results sufficiently representative and have decided to accept this threat.

Our ongoing work continues this investigation by analyzing the 1,167 primary studies looking for factors (design, human, or others) that have an impact on reliance in XAI-assisted decision-making, types of explanation designs used to provide reliance in this context, and how to evaluate explanation designs regarding reliance in XAI-assisted decision-making.

References

[1] A. Abdul, J. Vermeulen, D. Wang, B. Y. Lim, and M. Kankanhalli, “Trends and trajectories for explainable, accountable and intelligible systems: An hci research agenda,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI ’18)*. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–18.

[2] D. Gunning, E. Vorm, J. Y. Wang, and M. Turek, “Darpa’s explainable ai (xai) program: A retrospective,” *Applied AI Letters*, vol. 2, no. 4, pp. 1–11, 2021.

[3] A. Khurana, P. Alamzadeh, and P. K. Chilana, “Chatrex: Designing explainable chatbot interfaces for enhancing usefulness, transparency, and trust,” in *2021 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, 2021, pp. 1–11.

[4] D. Dos Santos Ribeiro, G. D. J. Barbosa, M. Do Carmo Silva, H. Lopes, and S. D. J. Barbosa, “Exploring the impact of classification probabilities on users’ trust in ambiguous instances,” in *2021 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, 2021, pp. 1–9.

[5] Z. Buçinca, M. B. Malaya, and K. Z. Gajos, “To trust or to think: Cognitive forcing functions can reduce overreliance on ai in ai-assisted decision-making,” *Proc. ACM Hum.-Comput. Interact.*, vol. 5, no. CSCW1, apr 2021. [Online]. Available: https://doi.org/10.1145/3449287

[6] V. L. Pop, A. Shrewsbury, and F. T. Durso, “Individual differences in the calibration of trust in automation,” *Human Factors and Ergonomics Society*, vol. 57, no. 4, pp. 545–556, 2015.

[7] M. Naiseh, A. Simkute, B. Zieni, N. Jiang, and R. Ali, “C-xai: A conceptual framework for designing xai tools that support trust calibration,” *Journal of Responsible Technology*, vol. 17, p. 100076, 2024.

[8] Q. V. Liao and K. R. Varshney, “Human-centered explainable ai (xai): From algorithms to user experiences,” 2022. [Online]. Available: https://arxiv.org/abs/2110.10790

[9] E. Purificato, F. Lorenzo, F. Fallucchi, and E. W. De Luca, “The use of responsible artificial intelligence techniques in the context of loan approval processes,” *International Journal of Human–Computer Interaction*, vol. 39, no. 7, pp. 1543–1562, 2023.

[10] P. E. Kourouthanassis, G. M. Giaglis, and D. C. Karaiskos, “Delineating the degree of ’pervasiveness’ in pervasive information systems: An assessment framework and design implications,” in *2008 Pan-Hellenic Conference on Informatics*. Los Alamitos, CA, USA: IEEE, 2008, pp. 251–255.

[11] B. A. Kitchenham, D. Budgen, and P. Brereton, *Evidence-Based Software Engineering and Systematic Reviews*, 1st ed. Boca Raton, FL, USA: CRC press, 2015.

[12] T. Miller, “Explanation in artificial intelligence: Insights from the social sciences,” *Artificial Intelligence*, vol. 267, pp. 1–38, 2019.

[13] T. Clement, N. Kemmerzell, M. Abdelaal, and M. Amberg, “Xair: A systematic metareview of explainable ai (xai) aligned to the software development process,” *Machine Learning and Knowledge Extraction*, vol. 5, no. 1, pp. 78–108, 2023.

[14] J. D. Lee and K. A. See, “Trust in automation: Designing for appropriate reliance,” *Human Factors*, vol. 46, no. 1, pp. 50–80, 2004.

[15] N. Scharowski, S. A. C. Perrig, N. von Felten, and F. Brühlmann, “Trust and reliance in xai – distinguishing between attitudinal and behavioral measures,” 2022. [Online]. Available: https://arxiv.org/abs/2203.12318

[16] T. Ueno, Y. Kim, H. Oura, and K. Seaborn, “Trust and reliance in consensus-based explanations from an anti-misinformation agent,” in *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA ’23)*. New York, NY, USA: Association for Computing Machinery, 2023.

[17] V. R. Basili, G. Caldiera, and H. D. Rombach, “Goal question metric paradigm,” in *Encyclopedia of Software Engineering*. John Wiley & Sons, 1994, vol. 1, pp. 528–532.

[18] J. J. Ferreira and M. S. Monteiro, “What are people doing about xai user experience? a survey on ai explainability research and practice,” in *Design, User Experience, and Usability. Design for Contemporary Interactive Environments. HCII 2020*, ser. Lecture Notes in Computer Science, A. Marcus and E. Rosenzweig, Eds. Cham: Springer, 2020, vol. 12201, pp. 56–73.

[19] A. Bertrand, R. Belloum, J. R. Eagan, and W. Maxwell, “How cognitive biases affect xai-assisted decision-making: A systematic review,” in *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (AIES ’22)*. New York, NY, USA: Association for Computing Machinery, 2022, pp. 78–91.

[20] A. Bertrand, T. Viard, R. Belloum, J. R. Eagan, and W. Maxwell, “On selective, mutable and dialogic xai: a review of what users say about different types of interactive explanations,” in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI ’23)*. New York, NY, USA: Association for Computing Machinery, 2023.

[21] M. Petticrew and H. Roberts, *Systematic Reviews in the Social Sciences: A Practical Guide*, 1st ed. Oxford, UK: Blackwell Publishing, 2006.

[22] L. I. Meho and Y. Rogers, “Citation counting, citation ranking, and h-index of human-computer interaction researchers: A comparison of scopus and web of science,” *Journal of the American Society for Information Science and Technology*, vol. 59, no. 11, pp. 1711–1726, 2008.

[23] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, “Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai,” *Information Fusion*, vol. 58, pp. 82–115, 2020.

[24] V. Lai, C. Chen, A. Smith-Renner, Q. V. Liao, and C. Tan, “Towards a science of human-ai decision making: An overview of design space in empirical human-subject studies,” in *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT ’23)*. New York, NY, USA: Association for Computing Machinery, 2023, p. 1369–1385.

[25] J. C. de Souza Filho, R. Belloum, and K. M. de Oliveira, “Database search for a tertiary study on appropriate reliance in xai systems,” Univ. Polytechnique Hauts-de-France, LAMIH, UMR CNRS 8201, F-59313, Valenciennes, France, Tech. Rep., Jul. 2024. [Online]. Available: https://hal.science/hal-04637901

[26] D. Buschek, M. Eiband, and H. Hussmann, “How to support users in understanding intelligent systems? an analysis and conceptual framework of user questions considering user mindsets, involvement, and knowledge outcomes,” *ACM Trans. Interact. Intell. Syst.*, vol. 12, no. 4, nov 2022. [Online]. Available: https://doi.org/10.1145/3519264

[27] M. Naiseh, N. Jiang, J. Ma, and R. Ali, “Personalising explainable recommendations: Literature and conceptualisation,” in *Trends and Innovations in Information Systems and Technologies*, Á. Rocha, H. Adeli, L. P. Reis, S. Costanzo, I. Orovic, and F. Moreira, Eds. Cham: Springer International Publishing, 2020, pp. 518–533.

[28] ——, “Explainable recommendations in intelligent systems: Delivery methods, modalities and risks,” in *Research Challenges in Information Science*, F. Dalpiaz, J. Zdravkovic, and P. Loucopoulos, Eds. Cham: Springer International Publishing, 2020, pp. 212–228.

[29] D. Afchar, A. B. Melchiorre, M. Schedl, R. Hennequin, E. V. Epure, and M. Moussallam, “Explainability in music recommender systems,” *AI Magazine*, vol. 43, no. 2, pp. 190–208, 2022.

[30] D. Jannach, M. Jugovac, and I. Nunes, *Explanations and user control in recommender systems*. Berlin, Boston: De Gruyter Oldenbourg, 2023, pp. 129–152. [Online]. Available: https://doi.org/10.1515/9783110988567-006

[31] A. Vultureanu-Albişi and C. Bădică, “Recommender systems: An explainable ai perspective,” in *2021 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, 2021, pp. 1–6.

[32] D. Omeiza, H. Webb, M. Jirotka, and L. Kunze, “Explanations in autonomous driving: A survey,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 10 142–10 162, 2022.

[33] S. Kim, R. van Egmond, and R. Happee, “Effects of user interfaces on take-over performance: a review of the empirical evidence,” *Information*, vol. 12, no. 4, p. 162, 2021.

[34] H. Lin, J. Han, P. Wu, J. Wang, J. Tu, H. Tang, and L. Zhu, “Machine learning and human-machine trust in healthcare: A systematic survey,” *CAAI Transactions on Intelligence Technology*, vol. 9, no. 2, pp. 286–302, 2023.

[35] C. Manresa-Yee, M. F. Roig-Maimó, S. Ramis, and R. Mas-Sansó, “Advances in xai: Explanation interfaces in healthcare,” in *Handbook of Artificial Intelligence in Healthcare: Vol 2: Practicalities and Prospects*, C.-P. Lim, Y.-W. Chen, A. Vaidya, C. Mahorkar, and L. C. Jain, Eds. Cham: Springer International Publishing, 2022, pp. 357–369.

[36] T. Polzehl., V. Schmitt., N. Feldhus., J. Meyer., and S. Möller., “Fighting disinformation: Overview of recent ai-based collaborative human-computer interaction for intelligent decision support systems,” in *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2023) - HUCAPP*, INSTICC. SciTePress, 2023, pp. 267–278.

[37] J. A. Alvarado-Valencia and L. H. Barrero, “Reliance, trust and heuristics in judgmental forecasting,” *Computers in Human Behavior*, vol. 36, pp. 102–113, 2014.

[38] D. Horvat, I. Botički, P. Seow, and A. Drobnjak, “Explainable ai in the real world: Challenges and opportunities,” in *The 31st International Conference on Computers in Education*, 2023, pp. 741–749.

[39] M. Suffian, I. Stepin, J. M. Alonso-Moral, and A. Bogliolo, “Investigating human-centered perspectives in explainable artificial intelligence,” in *Proceedings of the 4th Italian Workshop on Explainable Artificial Intelligence (XAI.it 2023)*, vol. 3518. CEUR Workshop Proceedings, 2023, pp. 47–66.

[40] S. Laato, M. Tiainen, A. Najmul Islam, and M. Mäntymäki, “How to explain ai systems to end users: a systematic literature review and research agenda,” *Internet Research*, vol. 32, no. 7, pp. 1–31, 2022.

[41] S. Morandini, F. Fraboni, G. Puzzo, D. Giusino, L. Volpi, H. Brendel, E. Balatti, M. De Angelis, A. De Cesarei, and L. Pietrantoni, “Examining the nexus between explainability of ai systems and user’s trust: A preliminary scoping review,” in *Joint Proceedings of the xAI-2023 Late-breaking Work, Demos and Doctoral Consortium, co-located with the 1st World Conference on eXplainable Artificial Intelligence (xAI-2023)*, vol. 3554. CEUR Workshop Proceedings, 2023, pp. 30–35.

[42] S. Mohseni, N. Zarei, and E. D. Ragan, “A multidisciplinary survey and framework for design and evaluation of explainable ai systems,” *ACM Trans. Interact. Intell. Syst.*, vol. 11, no. 3–4, sep 2021. [Online]. Available: https://doi.org/10.1145/3387166

[43] F. Sperrle, M. El-Assady, G. Guo, R. Borgo, D. H. Chau, A. Endert, and D. Keim, “A survey of human-centered evaluations in human-centered machine learning,” *Computer Graphics Forum*, vol. 40, no. 3, pp. 543–568, 2021.

[44] Y. Rong, T. Leemann, T.-T. Nguyen, L. Fiedler, P. Qian, V. Unhelkar, T. Seidel, G. Kasneci, and E. Kasneci, “Towards human-centered explainable ai: A survey of user studies for model explanations,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 4, pp. 2104–2122, 2024.

[45] D. Muralidhar, R. Belloum, K. M. de Oliveira, and A. Ashok, “Elements that influence transparency in artificial intelligent systems - a survey,” in *Human-Computer Interaction – INTERACT 2023*, J. Abdelnour Nocera, M. Kristín Lárusdóttir, H. Petrie, A. Piccinno, and M. Winckler, Eds. Cham: Springer Nature Switzerland, 2023, pp. 349–358.

[46] J. Zerilli, U. Bhatt, and A. Weller, “How transparency modulates trust in artificial intelligence,” *Patterns*, vol. 3, no. 4, 2022.

[47] Z. Shams Khoozani, A. Q. M. Sabri, W. C. Seng, M. Seera, and K. Y. Eg, “Navigating the landscape of concept-supported xai: Challenges, innovations, and future directions,” *Multimedia Tools and Applications*, pp. 1–51, 2024.

[48] R. Fok and D. S. Weld, “In search of verifiability: Explanations rarely enable complementary performance in ai-advised decision making,” *AI Magazine*, 2024.

[49] M. Szymanski, M. Millecamp, and K. Verbert, “Visual, textual or hybrid: the effect of user expertise on different explanations,” in *Proceedings of the 26th International Conference on Intelligent User Interfaces*, ser. IUI ’21. New York, NY, USA: Association for Computing Machinery, 2021, p. 109–119.

[50] C. Wohlin, P. Runeson, M. Höst, M. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in Software Engineering*. Springer-Berlin Heidelberg, 2012.