How Human-Centered Explainable AI Interface Are Designed and Evaluated: A Systematic Survey

THU NGUYEN, IT University of Copenhagen, Denmark
ALESSANDRO CANOSSA, Royal Danish Academy, Denmark
JICHEN ZHU, IT University of Copenhagen, Denmark

Despite its technological breakthroughs, eXplainable Artificial Intelligence (XAI) research has limited success in producing the *effective explanations* needed by users. In order to improve XAI systems' usability, practical interpretability, and efficacy for real users, the emerging area of *Explainable Interfaces* (EIs) focuses on the user interface and user experience design aspects of XAI. This paper presents a systematic survey of 53 publications to identify current trends in human-XAI interaction and promising directions for EI design and development. This is among the first systematic survey of EI research.

Additional Key Words and Phrases: Explainable Interface, Explainable AI, Systematic Review

1 INTRODUCTION

The research field of eXplainable AI (XAI) has emerged to make AI and machine learning (ML) more transparent and trustworthy to humans by opening the AI black-box and explaining its underlying operation[36]. This rapidly growing field has already made significant breakthroughs in *technical explainability*, producing established XAI algorithms such as LIME [84], DeepLIFT [89], LRP [9]). In comparison, XAI research has limited success producing the *effective explanations* needed by users[14, 58, 108]. As a result, most explanations produced by XAI still lack usability, practical interpretability, and efficacy for real users [1, 28, 66, 111]. A recent study found that a significant group of users (over 30%) were unable to understand the XAI explanations sufficiently well to use them even in relatively simple tasks [70].

Recently there have been growing efforts, especially from the Human-Computer Interaction (HCI) community, to adopt more human-centered approaches and rigorous empirical evaluation methods [28, 59, 73] for XAI. For example, researchers draw from cognitive science theories of how people reason [101, 103] and from social sciences of how people explain [67] and propose new paradigms of XAI that are grounded in existing understandings of human behavior.

Another growth area is *explainable interface* (EI)[69], also referred to as explanation user interface[15] or explanation format[14, 69], focusing on the user interface (UI) and user experience (UX) design aspects of XAI (e.g., structure, design, format, design, and content of the explanations). If the primary concern of XAI algorithms is to generate *what to explain*, EI research is about *how to explain* in a way that is effective for specific user groups. In other words, EI research is intrinsically user/human-centered. As a recently emerging area, EI research can shed light onto many fundamental questions about how to design XAI so that it can better meet the needs of real users. For example, there are currently no agreements over whether XAI should include all details of system logic[53, 54] or only selective important information[41, 87]). Recently, Villareale et al. [100] explored the design possibility of EIs in the format of computer games.

While there are numerous review articles on the general trends in XAI[3, 4, 19, 102], especially its technical development[18, 29, 86, 95], there has not been systematic efforts to map out the current state of EI. Even though the term EI only dates back to 2021, almost all XAI systems have an EI, whether it is explicitly designed or not.

Authors' addresses: Thu Nguyen, IT University of Copenhagen, Copenhagen, Denmark, irng@itu.dk; Alessandro Canossa, Royal Danish Academy, Copenhagen, Denmark, acan@kglakademi.dk; Jichen Zhu, IT University of Copenhagen, Copenhagen, Denmark, jichen.zhu@gmail.com.

Furthermore, there has been related work in earlier research, most notably explanation facility in recommender system research[76, 98].

Specially, we aim to identify current practices in current EI design approaches as well as promising directions that can further improve the usability, practical interpretability, and efficacy for real users of XAI. Towards these goals, we design our survey to focus on existing XAI research that involves human users, which we believe is a prerequisite of any human-centered XAI. Specifically, we seek to answer the following research questions:

- RQ1: *How do researchers involve human participants in the design and development of XAI applications?*
- RQ2: *How do XAI researchers design EIs?*
- RQ3: *How do XAI researchers evaluate EIs?*

Using the Preferred reporting items for systematic reviews and meta-analyses (PRISMA) guideline[68], we included 53 publications for analysis. The remainder of the paper is organized as follows. We first present related review papers in XAI and describe our methods for the survey. Next, we analyze our results based on the above-mentioned three research questions. Finally, we discuss the implications of our findings.

2 RELATED WORK

As the research area of XAI gain momentum, recently there have been numerous reviews about the general trends and opportunities of XAI[3, 4, 19, 102], XAI for specific ML techniques (e.g., for supervised learning[29], for natural language process[18], for deep learning[95], and for time-series data[86]), and XAI in specific use domains (e.g., XAI for medical use[47, 79, 99], for air-traffic management[22]).

Most of the XAI review articles are technical in nature, while review articles on the design aspects of XAI research are relatively rare and recent. Amongst these design articles [69, 98, 106], summarizing existing design goals of XAI is a common approach, as design goals are a good way to capture research focus. For example, Mohensi et al. [69] investigate which design goals and evaluation metrics are used in XAI research. They found that in existing XAI research, for each of the three main user groups, there are common design goals associated with them: Novice Users (Algorithmic Transparency, User Trust and Reliance, Bias Mitigation, Privacy Awareness), Data Experts (Model Visualization and Inspection, Model Tuning and Selection) and AI Experts (Model Interpretability, Model Debugging). They also classified five different types of evaluation measures: Mental Model, Usefulness and Satisfaction, User Trust and Reliance, Human-AI Task Performance, and Computational Measures. Chromik and Butz [14] survey recommender systems, a sub-area of AI where early research on explainability has concentrated, and use Hornbæk and Oulasvirta's interaction types [44] to classify different XAI design goals. For instance, "Dialogue" interaction is linked to the design goals of transparency and scrutability, while "Experience" is connected to satisfaction, trust, and persuasiveness.

Aiming to improve the practical effectiveness of XAI, a growing number of review papers attempt to bring insights from HCI/design research to this overwhelmingly of technical field. These reviews tend to focus on developing a deeper understanding of users and their needs. For instance, Suresh et al. [96] propose that XAI research should differentiate the stakeholder expertise into knowledge and contexts, and stakeholder needs into long-term goals, shorter-term objectives, and immediate-term tasks. Based on an analysis of 58 publications, they find that their framework can help researchers to design more precise application-grounded evaluations. Similarly, Liao and Varshney [58] expand the user groups of XAI to a broader range of stakeholders (Model developers, Business owners or administrators, Decision-makers, Impacted groups, and Regulatory bodies). Their survey found common disconnections 1) between technical XAI approaches and supporting users' end goals in usage contexts, 2) between assumptions underlying

technical approaches to XAI and people's cognitive processes. They argue that technical choices of XAI algorithms should be driven by users' *explainability needs*.

These existing works largely address the question of *WHAT* has been explained, leaving a gap about *HOW* to explain, especially in terms of how to design the explainable interfaces (EIs). Recent work [69] and [14] have call into attention the importance of explainable interfaces and started to develop design guidelines. Specifically, Mohensi et al. [69] proposed a design and evaluation framework where there is a separate layer dedicated to EI, including explanation format and interaction design. They call for future research to look into required features for both components of EI. Chromik and Butz's recent work [14] delves deeper into the interaction design of EI and examines different interaction styles in current XAI literature. While there is growing interest in EI design and development, there has not been a systematic review on this topic. To the best of our knowledge, this is among the first such work, and it will be complementary to the large body of algorithm-focused reviews of XAI research.

3 METHOD

3.1 Dataset: Search Procedure and Inclusion Criteria

We conducted a systematic survey to understand the current state of the art of human-centered XAI, especially in terms of the design and evaluation of explainable interfaces (EIs). We pay special attention to research that involves users since it is our belief that a user-centered design process is a prerequisite for human-centered XAI.

Given that our focus on user participation in XAI and explainable interface is at the intersection of AI and HCI, we used ACM digital library (ACM DL) as the primary source for collecting publications. It includes relevant conferences such as CHI, Intelligent User Interface (IUI), and Designing Interactive Systems (DIS), where such work is regularly published. Since EI is a recently emerged research area[69], we acknowledge that it is possible that related work may be published in other venues, especially more technical AI conferences, or in arXiv. It is a limitation of our work.

Our overall selection process is illustrated in Fig. 1.

To cast a broad net, we query the ACM DL database to find scientific publications on XAI that mentioned "participants" or "users." Since *"explainable interface"* is a new term first used by Mohensi et al. in 2021[69], we did not include it. Building on the query used by Chromik and Butz [14], we also include the term *"explanation facility"*, which the recommender systems research community uses to refer to explanations of the systems' recommendations. It is important to note that while the *term* of EI is new, almost all XAI has an interface, whether explicitly or implicitly designed, to convey the generated explanation to human participants. Finally, as XAI research became active in the last two decades, we include publications from January 2000 to May 2022. Below is our query:

"query": Abstract:("XAI" OR "explainable AI" OR "explanation facility") AND Fulltext:("participant" OR "user*") "filter": Article Type: Research Article, Publication Date: (01/01/2000 TO 05/31/2022), ACM Content: DL*
"filter": Article Type: Short Paper, Publication Date: (01/01/2000 TO 05/31/2022), ACM Content: DL

From this query, a total of 96 publications (84 research articles and 12 short publications) are found for further screening. Next, we screen these publications to select the eligible ones. Our *inclusion criteria* are

(1) publications conduct research in the area of XAI (e.g., technical research on how to explain the decisions of AI, ML, recommendation systems, and other intelligent systems; or design/HCI research on how to design effective explanations),
(2) publications that include at least one user study in the EI design process, including gathering user requirements, participatory design, and evaluation, except labeling training data sets, AND

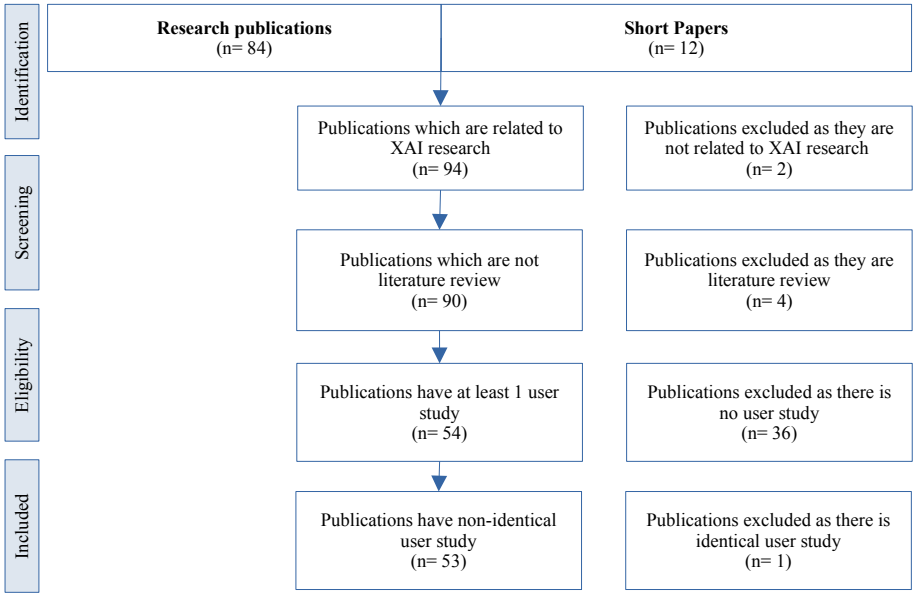

Fig. 1. Publications selection process

(3) publications are not review papers.

Our screening process is the following. We first read the abstract to identify publications to filter out those that clearly do not meet our inclusion criteria. Among the remaining ones, we go through the whole text to determine their eligibility.

Several publications are excluded because they only mention user study superficially (e.g., referring to the authors' prior user study, or mentioning user study in planned future work) without reporting any empirical results. We also excluded publications with unsuccessful user studies where the study results are not reported (e.g., [52]). In the case where there are multiple publications on the same user study, we only use the publication most relevant to EI. Out of the initial 96 publications, 53 passed our screening and are included in the data analysis.

3.2 Data Analysis Methods

We analyzed this data conducting thematic analysis [7] and grounded theory [17]. Thematic analysis is used to categorize the *required features and visual hierarchy* as well as *evaluation metrics* and *metric types*, these terms will be discussed later. Since we want to understand how EIs are currently designed and evaluated, we need to examine the properties that EIs possess. The properties of an EI, such as a graphical interface, can only be measured with qualitative method since they are concrete tangible features. Similarly, evaluation metrics which include different desired qualities and abilities that human can perform with the system, can also be accounted for with qualitative method. Thematic analysis is a qualitative method used to sort and arrange textual data into specific themes [7]. Specifically, we collect the required features found in the selected publications and group similar features into the same category. For example, it is required for the explainable interface to have multiple explanations and to have multiple data input presented, these features are

Table 1. Overview of 53 XAI publications that include human participation in the user studies

Publications	Design requirements	information architecture	static/interactive	interaction type	ISO steps	participant group	evaluation metrics	metric type
Abdul et al. [2]	F1, RF3	sequential	static	–	Evaluate	general	Cog1	COG
Arrota et al. [5]	*none*	sequential	static	–	Evaluate	domain	Des5	DES
Bansal et al. [6]	RF4	sequential	static	–	Evaluate	general	Act3	ACT
Bove et al. [10]	RF4, VH3, RF5	hierarchical	interactive	instructing	Evaluate	general	Cog1, Des5	COG & DES
Bucinca et al. [11]	VH3	sequential	static	–	Evaluate	ML/AI, general	Act3, Cog1, Des4, Des3	COG, ACT & DES
Bucinca et al. [12]	*none*	hierarchical	interactive	instructing	Evaluate	general	Des4, Des3, Cog1, Act3	COG, ACT & DES
Chromik et al.[16]	RF2, RF4	sequential	static	–	Analysis, Evaluate	domain	Act1, Des2	ACT & DES
Chromik et al. [15]	VH3, RF5	matrix	interactive	instructing	Evaluate	general		–
Das and Chernova [21]	RF2	organic	interactive	exploring	Evaluate	general	Act3	ACT
Das et al, [20]	RF7	sequential	static	–	Evaluate	general	Act3	ACT
David et al. [8]	VH1	sequential	interactive	instructing	Evaluate	general	Des4, Act4, Des5	ACT & DES
Dhanorkar et al. [24]	*not designed*	–	static	–	Analysis	ML/AI		–
Dodge et al. [26]	RF4, VH1, VH2	hierarchical	interactive	instructing	Analysis, Evaluate	domain		–
Dodge et al. [25]	VH3, RF5	matrix	interactive	instructing	Evaluate	ML/AI, general	Des4, Cog1, Act3	COG, ACT & DES
Donadello et al. [27]	RF2	sequential	static	–	Evaluate	domain	Des2	DES
Ehsan et al. [30]	RF8	sequential	static	–	Analysis, Produce, Evaluate	ML/AI		–
Flutura et al. [?]	VH4, F1	sequential	static	–	Evaluate	ML/AI	Cog1	COG
Ghai et al. [34]	*none*	sequential	static	–	Evaluate	ML/AI, general	Des4, Des5, Cog1	COG & DES
Gorski and Ramakrishna [38]	VH3	sequential	static	–	Analysis, Evaluate	domain	Cog1, Des2	COG & DES
Gou et al. [37]	F1, RF5	organic	interactive	instructing	Evaluate	general	Cog1, Des2, Act1, Act2, Des4, Des5	COG, ACT & DES
Hadash et al. [39]	RF3	sequential	static	–	Evaluate	general	Cog1	COG
Hamon et al. [40]	VH1, VH5, VH3	sequential	static	–	Produce	ML/AI, general, domain		–
H.-Bocanegra and Ziegler [42]	RF2	matrix	interactive	instructing & conversing	Analysis, Evaluate	general	Des1, Des4, Des2, Des5	DES
Hjorth [43]	VH3, RF2	organic	interactive	instructing & manipulating	Produce, Evaluate	domain, general		–
Jacobs et al. [?]	RF5, RF6, VH3	matrix	interactive	instructing	Analysis, Produce, Evaluate	domain	Des2	DES
Jesus et al. [46]	*none*	sequential	static	–	Evaluate	domain	Act3, Des2	ACT & DES
Kaptein et al. [48]	RF2	organic	interactive	instructing	Evaluate	general		–
Khanna et al. [49]	VH1	hierarchical	interactive	manipulating	Evaluate	domain	Act3	ACT
Kim et al. [50]	VH3, F1	matrix	interactive	instructing	Evaluate	ML/AI	Act3, Act4, Cog1, Des4	COG, ACT & DES
Le et al. [55]	RF7	sequential	static	–	Evaluate	general	Des3, Des2, Cog1	COG & DES
Lee et al. [56]	*not designed*	–	static	–	Produce	domain		–
Liao et al. [57]	*not designed*	–	static	–	Analysis	UI UX		–
Lima et al. [60]	*not designed*	–	static	–	Analysis	general		–
Mai et al. [64]	*none*	sequential	static	–	Analysis, Evaluate	domain	Des2, Des3	DES
Maltbie et al. [65]	RF8	sequential	static	–	Evaluate	domain	Des2, Des3	DES
Narkar et al. [71]	VH3, VH4	matrix	interactive	instructing	Analysis, Evaluate	ML/AI	Des2	DES
Nourani et al. [75]	RF4	matrix	interactive	instructing	Evaluate	general	Act3, Des2, Cog1	COG, ACT, DES
Panigutti et al. [77]	*none*	sequential	static	–	Evaluate	domain	Des2	DES
Penney et al. [80]	*not designed*	–	static	–	Analysis	domain		–
Polley et al. [81]	RF7	sequential	interactive	instructing	Evaluate	domain	Des4, Des2, Cog1	COG & DES
Qian [82]	RF5	organic	interactive	instructing	Evaluate	general	Cog1, Cog2, Des3, Des3	COG & DES
Qiu et al. [83]	*none*	sequential	static	–	Evaluate	general	Des2, Des5, Des4, Cog1	COG & DES
Sevastjanova et al. [88]	VH3, VH4, VH1	organic	interactive	instructing, manipulating & exploring	Evaluate	domain	Des2, Des4, Cog1, Act3	COG, ACT & DES
Sklar and Azhar [91]	RF2, RF5	organic	interactive	conversing & exploring	Evaluate	domain, general	Act3, Des5	ACT & DES
Slijepcevic et al. [92]	VH4, VH3	sequential	static	–	Evaluate	domain	Des2	DES
Sovrano and Vitali [93]	VH1, VH3, VH5	sequential	interactive	instructing	Evaluate	general	Des2, Des5, Act3	ACT & DES
Sun et al. [94]	*not designed*	–	static	–	Analysis	domain, ML/AI		–
Tabrez et al. [97]	RF6, RF8	organic	interactive	exploring	Evaluate	domain	Des3	DES
Wang et al. [?]	VH3, VH4	sequential	static	–	Produce, Evaluate	domain	Cog1	COG
Wang and Yin [104]	*none*	sequential	static	–	Evaluate	general	Cog1, Des4	COG, DES
Wang et al. [105]	VH4, VH3, RF8	matrix	interactive	instructing	Evaluate	general	Des2, Des3	DES
Wolf [107]	*not designed*	–	–	–	Analysis	ML/AI		–
Zhang and Lim [110]	RF8	sequential	interactive	instructing	Analysis, Evaluate	general	Des2, Act3	ACT, DES

categorized into *multiple instances* group. We use the online collaborative board platform *Miro* to conduct the thematic analysis.

At the same time, we apply grounded theory to identify the abstract concepts that capture the presentation and navigation of the designed EIs. According to Corbin and Strauss [17], a grounded theory approach generates theories from the data. To construct the theory, many iterations of data collection and theory construction are essential. Specifically, we conduct several iterations of data collection and theory construction from different types of data related to the presentations of the EIs (e.g., format, interface elements, visualisation types, interaction type, UI type). In the *Results* section we will list all the 8 properties that emerged from applying thematic analysis and grounded theory to our dataset, subdivided according to our three research questions.

4 RESULTS

This section presents our findings based on our research questions (Table 1) and a cluster analysis of the current trends in EI research.

4.1 RQ1: How do researchers involve human participants in the design and development of XAI applications?

Based on our analysis, we find that two sub-questions are particularly useful to describe current practices: *at which stage* are human participants involved and *which type* of participants.

In terms of **the stage of involvement**, we analyzed our dataset of publications based on the widely used *ISO human-centered design process framework* [33]. In this standardized framework, after the *Planning* activity, the iterative design process consists of 1) *Analysis*: Understand and specify the context of use (e.g., User group profiles, task models, as-is scenarios, personas, user journey maps), 2) *Specify* the user requirements (e.g., user needs definition, and related forms of understanding users), 3) *Produce* design solutions to meet user requirements, and 4) *Evaluate* the designs against user requirements. Ideally, the development team should involve users and other stakeholders in all activities, especially in 1) and 4). This framework provides us with the first of the eight properties listed in 1: *Activities*. In table 1, we code the identified activities in our surveyed publications, except for Specify activity. Since Specify only indicates the results gathered from the Analysis activities and no specific actions found in this activity, we do not include Specify activity in our data analysis.

Fig. 2 (Left) summarizes the human participants' involvement in different activities (counted by the number of publications). The overwhelming majority (84.9%) of XAI publications involve humans to evaluate the XAI system, while 28.3% in Analysis. The latter indicates that less than a third of the XAI research has conducted some form of user research before setting out to produce XAI solutions. Among all the publications, only 17% involved humans in both Analysis and Evaluation, as the human-centered design process requires. While we acknowledge the differences between research and UX design projects, we believe that the ISO framework, especially at a high level, is useful to shed light on how current XAI research involves human participants. This is especially true because explanations are contextual and user group-specific and understanding users is an essential part of effective HCXAI. This is the second property, *Participant groups*.

A main activity during the Analysis phase is defining user requirements, the results gathered from understanding users in the context of use that include user needs, and problems users have regarding the interactive system. Several publications explicitly gather user requirements for explanations (e.g., [38, 42, 45, 64, 94]). Other publications have

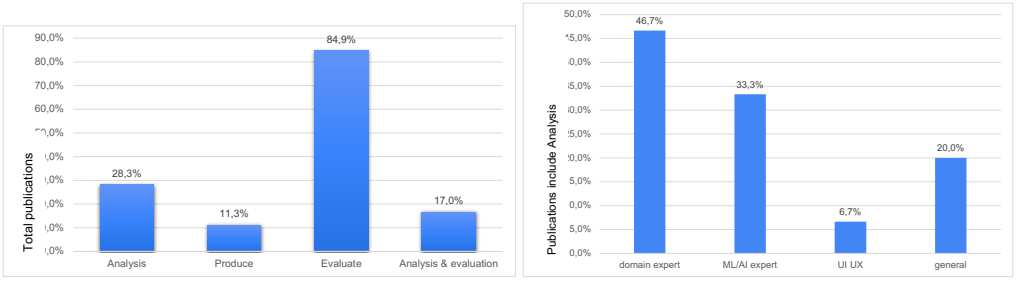

Fig. 2. Left: Distribution percentage of the design activities that human participants are involved in (n=53). Right: Distribution percentage of the five main groups of human participants in the Analysis Activity (n=14).

derived user requirements by understanding the challenges that users face in relation to XAI system (e.g., [24, 57]). We notice that all of these publications attempt to define user requirements (e.g. specific target user groups, user goals, user tasks) at the level of the XAI system in general. None reached the level of specificity of explainable interfaces, which results in a lot of ambiguity when the research team reach the Produce and Evaluation activities.

Among the publications that gather user requirements, 60% of them (17% of the entire surveyed publications) evaluated the system using human participants. Some followed the established user centered design processes (e.g., the ISO framework). They first gather user requirements to drive their design, and then evaluate the prototype based on user requirements (e.g., [16]). In contrast, some publications conducted user requirements gathering and prototype evaluation at the same time (e.g., [38, 71]). The latter case means that the prototype is not evaluated against user requirements.

In addition to defining user requirements, the ISO framework specifies other forms of understanding users in the Analysis activity. In our included publications, understanding users typically takes the form of 1) understanding how users make sense of the machine learning model (e.g., [16, 26, 30, 38, 80, 107, 110]), 2) how they perceive AI (e.g., [60]), or 3) how they perform XAI-related tasks such as comparing different ML models (e.g., [71]). All of these activities focuses on the users' cognitive processes, since the explanations produced by XAI has high information density to all user groups.

In terms of **the type of human participants**, different user groups have different needs when it comes to explanations [58]. Existing surveys have classified the stakeholder groups based on their ML expertise[69] and their purposes of using XAI[58]. Similar to these surveys, we identified 3 stakeholders groups of domain experts, ML/AI experts, and UI/UX experts, and 1 general participant group. Fig. 2 (Right) summarizes different groups involved in the Analysis activity. We find that 46.7% of the surveyed publications involve domain experts such as caregivers [5], lawyers [38], clinical experts [45, 56, 78, 92, 103]. These XAI systems are typically designed to increase trust (e.g., [16, 77, 81]) and transparency [27, 40, 49] for domain experts. Another 33.3% of our surveyed publications use participants with ML/AI expertise in their Analysis activity. This is typically for XAI systems designed for ML experts including tasks such as identifying existing problems [24], ML experts' mental models [25, 30]), and improving the interpretability of the XAI system based on ML experts' insights [31, 40]. The rest of the publications involve UI/UX practitioners and general participants, typically represented by the convenient sample of Amazon Mechanical Turk workers and students (represented as "genernal" in Fig. 2 (Right)). UX/UI practitioners provide insights of user requirements for XAI system [57]. One main use of general participants is to evaluate human-AI collaborating tasks (e.g., [6, 12, 20, 21] and

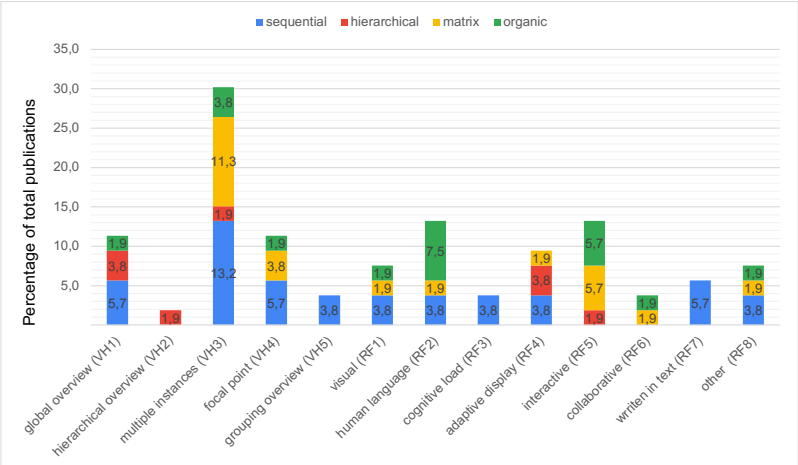

Fig. 3. Design Requirements (including visual hierarchy and required features) correlated with the Selected Information Architecture in the Final EI Designs

XAI system' desiderata (e.g., intepretability [10, 39], trust [8, 12, 37], user satisfaction[8, 10, 37]). These are seemingly generic tasks without requiring specific user expertise.

4.2 RQ2: How are explainable interfaces currently designed?

To answer RQ2, we analyze how the surveyed publications specify design goals and user experience, and what UI characteristics are represented in the final EI design. Compare to existing reviews that have examined the design goals of XAI overall[69, 106], our analysis is much more focused on the EI and how it facilitates the communication of technical explanations to human participants. Overall, 46 out of the total 53 publications (86.8%) presented an explainable interface (EI) as part of an XAI system. 45 publications (84.9%) designed and evaluated an EI, and 40 publications (75.5%) reported design requirements explicitly. Again, we use the term EI broadly to capture the interface between the technical explanations and the users. Existing reviews have analyzed main EI design goals for specific user groups [14, 69]. In this section, we attempt to understand, at a more concrete level, the information flow and the features that researchers decide that their EIs need to have.

4.2.1 Design Requirements: Visual Hierarchy and Required Features. Since the EIs in all surveyed publications use a graphical interface, we use the concept of *visual hierarchy* to understand how researchers seek to organize explanation-related information in the design requirements. Visual hierarchy refers to "the organization of the design elements on the page so that the eye is guided to consume each design element in the order of intended importance" [35]. In our work, we analyze the 31 publications where the researchers explicitly report their design requirements (e.g., explanation with natural human language [21, 27], multiple explanations [38, 40], explanations with highlighting regions of interest [31, 71]). We then use grounded theory [90] to derive five types of visual hierarchy: *global overview* (VH1) ,*hierarchical overview* (VH2), *multiple instances* (VH3), *focal points* (VH4), and *grouping overview* (VH5).

The left side of Fig. 3 shows the distribution of each visual hierarchy type over the number of publications. Note a publication can require multiple information hierarchy. For example, the required visual hierarchy from Sovrano et al. [93] are global overview, multiple instance and grouping overview since they want to have many explanations that

can be grouped and each of them provides global overview to confirm the theory that explanations can be understood by answering multiple answers from different questions. The most common visual hierarchy requirements is for the EI to show multiple instances of explanations at the same time. These 16 publications (30.2%) Multiple instances presents different factors of explanations or illustrates multiple examples of explanations in one view. For example, researchers report that their EI should show multiple explanations with different sets of input data for prediction [71], different pattern characteristics of the data [88], or multiple explanations for comparison [15, 25]. These publications choose multiple instances to help users to compare and contrast different instances and eventually help them build a general understanding of a machine learning model.

The second common type of visual hierarchy requirement is global overview. 11.3% the surveyed publications require explanations to be presented in a all-in-one style where different aspects of the explanations can be presented in the same view. Examples are that explanation should include multifaceted context [40], or be holistic to help users to locate bugs in explanations [49], or provide overviews of different explanations [93].

Another 11.3% of the surveyed publications required focal point as the visual hierarchy of their EI design. They require regions of interest to be highlighted in the EI. An example is that explanations should highlight relevant input data that influence the prediction of the model [92]. In addition, we find 2 publications (3.8%) require what we call grouping overview. In these publications, the researcher specify that the EI needs to present explanations in multiple clusters in the same view. Examples are that explanations should demonstrated in a group [40], or contextual information of explanation should be shown in group [93]. Finally, only 1 publication (1.9%) require hierarchical overview to present explanations in a tree branching structure.An example is that explanations represent different future states that an AI will be in based on the AI's possible sequences of actions [26].

Required features are functionalities that the researchers claim that their final EIs must have. These required features often have a large impact on the design space of the EI. For instance, the requirements of *being interactive* and using *human language* as output will highlight a chatbox interface as a design solution space. From our dataset of surveyed publications, 40 out of 53 publications (75.5%) report explicit required features. 13 publications do not report required features (listed as "*none*" in table 1).

Using the Grounded Theory methods, we identified 8 main types of required features. The right side of Fig. 3 shows the distribution of each required feature. Relatively speaking, "human language," being "interactive," and "adaptive display" are the most common required features. Interactive interfaces, while technically more challenging to develop, are required by researchers to allow users to choose how to explore the EI (e.g., [10, 16, 82]), to provide input to the system (e.g., [37]). Notably, human language is required by the XAI researchers to explain the AI classifications (e.g., [21]) or predictions in human understandable terms (e.g., [16, 27, 43]). Some researchers chose human language so that users can converse with the XAI in the form of human dialog (e.g., [42, 91]), or to make the system sound more intelligent [48]. The "Other" category include the least common features including audio feedback, recurrent explanation, and session recording. "User cognitive load" related features are the least reported. Examples include the requirement to design the EI to reduce the users' cognitive load for visual information [2] and semantic information (e.g., [39]). This is the third property: *required features and visual hierarchy.*

4.2.2 Explainable Interface Design Outcomes: Information architecture, interactivity, and interaction type. After analyzing the reported design requirements of EIs, we examine the final design outcome of the EIs (examples can be seen in Fig. 4). Compared with the design requirements specified *prior to* the final design, we examine how the structure of the information from the system interface is carried out in the final design. Specifically, we classified the EI designs based

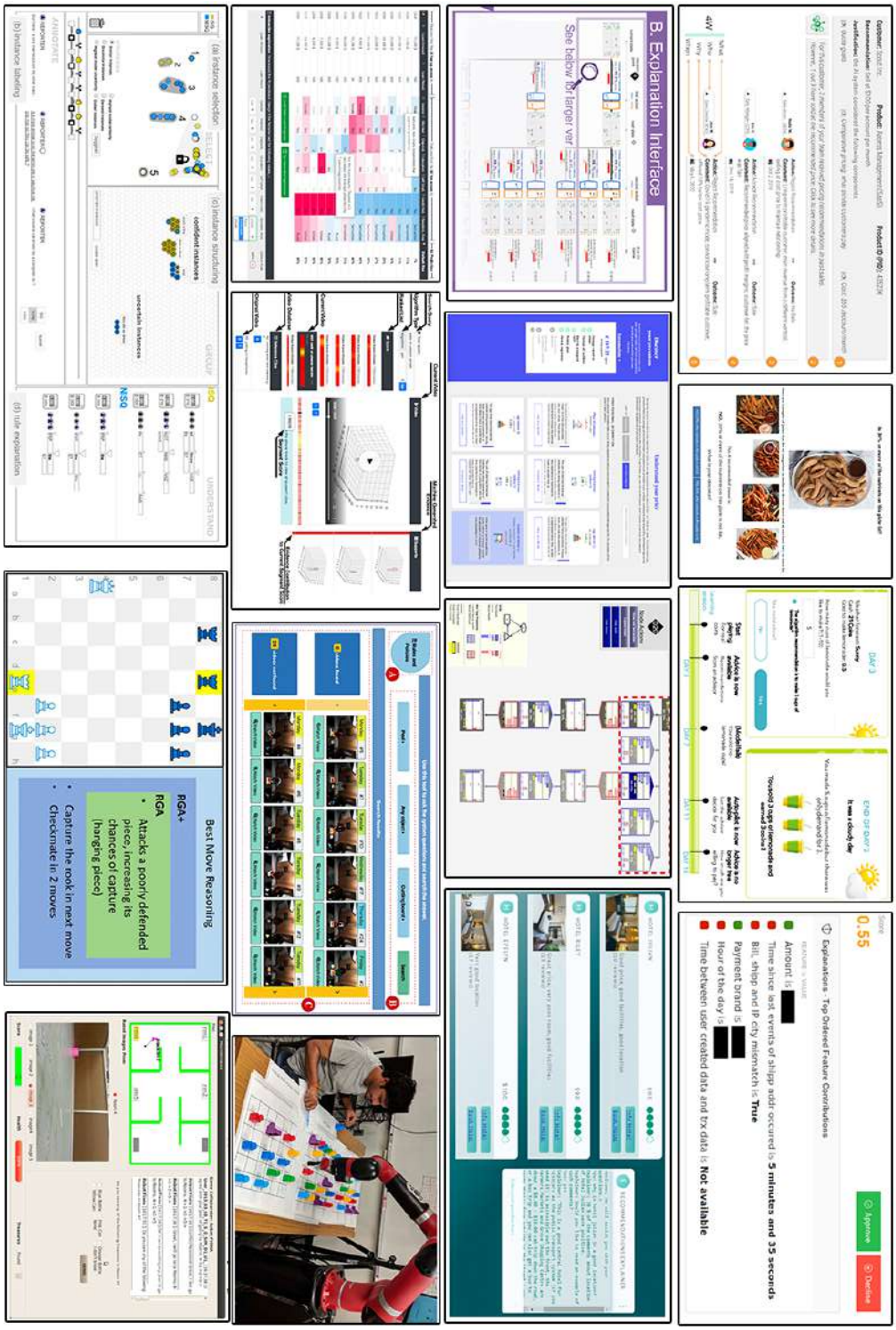

Fig. 4. Explainable Interface Design Examples. Top Row (Left to Right): [8, 11, 30, 46]. Second Row: [10, 26, 42, 49]. Third Row: [15, 50, 75, 97]. Fourth Row: [21, 88, 91]

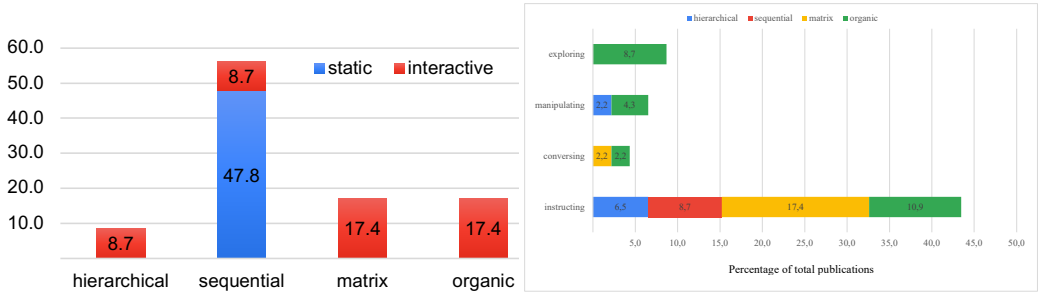

Fig. 5. Left: designed information architecture. Right: Interaction types with designed information architecture

on their *information architecture* and whether they are *interactive* or not. These are the fourth and fifth properties of our analysis.

Information architecture indicates the way the content of a designed EI is organized. We adopt Garret's well-established taxonomy [32], which classifies UI into four primary types of information architecture. We then code all the EIs in our surveyed articles accordingly and our results are summarized in Fig. 5 (Left).

49.1% of total surveyed publications that have the EI final designs included adopt the sequential information architecture, which presents information in a linear flow. In these EIs, users must go through explanations step by step in a pre-defined order. We believe that the main reason is that XAI explanations are typically highly technical. Sequential information architecture allows XAI researchers to control the order as well as the pace of how the explanations are presented to users. For example, Bucinca et al. [11] trained a ML model to predict the amount of fat content given images of food (Fig. 4, Second image Top Row). Their EI explains the prediction by showing to the users each of the food ingredients the ML model recognizes in the input image.

Matrix structure and organic structure are less commonly used in final EI designs. Matrix structure "allows user to move from node to node along two or more dimensions". A good example is Chromik et al. [15]'s interactive EI that shows the data of 16 loan applicants (Fig. 4, First image Third Row). For each applicant, the EI shows the features the ML model uses, its decision for approving/denying the application, and other information (e.g., loan period). Using the spread sheet-like interface, this EI design allows users to sort the different features and even modify the values to see how the ML decision changes. By using the multiple explanations (16 data points) for comparison, the EI encourage users to understand the ML model by contrastive reasoning (comparing between the applicants) and counterfactual reasoning (e.g., what if this applicant had a different gender). Since the matrix structure is suitable to present different pieces of information at once, it is preferred to design explanations that require multiple instances, as seen in Fig. 3. However, with organic structure design, the node of information evolves and depends on the actions that users perform with the system. For example, the chess game that users play with an agent from Das and Chernova [21] demonstrates this type of structure (Fig. 4, Third image Fourth Row). Specifically, since the moves from both the user and agent are completely unpredictable, each move a user makes will result in a different move that the agent will have, and vice versa. This type of interaction provides a free-form of interacting and exploring for users without a predefined structure.

The least common information architecture (7.5%) is hierarchical, which structures information in a tree-like branching structure. This is consistent with our above-mentioned analysis that hierarchical overviews is the least common design requirements. As shown in Fig. 3, the one publication [26] that requires hierarchical overview adopted hierarchical information architecture.

Interaction Types. Finally, Fig. 5 summarizes the distribution of the interaction types [85] of the final EI designs.Rogers et al.'s framework identified four main types of interaction: 1) *Instructing* is when user provide instructions, and commands to a system. Examples are typing commands, choosing options, and pressing buttons. 2) *Conversing* is when user have a two-way communication conversation with a system. For example, user input format can be typing or speaking, and the system response are text or speech. 3) *Manipulating* is when "users interact with objects in a virtual or physical space by manipulating them". Examples are placing, opening, and picking. 4) *Exploring* is when "users move through a virtual environment or a physical space".

Our analysis shows that that Instructing is the predominant interaction type (37.7%). All four information architecture structures are represented in the Instructing type. A main feature here is that users take control of how to approach the explanation, even though it typically takes limited forms such as clicking "Next" in a pre-defined sequence of explanations (e.g., [5, 6, 11, 31, 34, 83])) . More substantive examples are that users can press the button to sort (e.g., [10]), filter (e.g., [45]), or even change the value of certain parameters (e.g., [15]). While conversing seems to be the most "natural" way for users to receive explanations, it is the least common (3.5% of the surveyed publications, and 4.3% of the designed EIs). In both of the two publications that use the Conversing interaction type [42, 91], users interact with the XAI through a chat-bot. This is our sixth property.

4.3 RQ3: How are explainable interfaces currently evaluated?

Evaluation metrics. We investigate which evaluation metrics are currently used to assess EI designs. We collect all the metrics from our surveyed publications, and apply thematic analysis to analyze patterns. Evaluation metric for system's desiderata includes *transparency* (Des1), *effectiveness* (Des2), *efficiency* (Des3), *user trust* (Des4), and *user satisfaction* (Des5). User understanding of the system reflects on their cognitive ability *to understand* (Cog1) and *to learn* (Cog2) how machine learning model makes prediction. Since a user's understanding of the system influences the actions that they perform with the interface, their interaction can indicate their understanding of the system. In return, when users have limitation in interacting with the system, it influences their ability to understand and learn the system. Hence, it is necessary to examine user ability *to control* (Act1), *to respond* (Act2), and to collaborate: *to accomplish task* (Act3) and *to synchronize* (Act4) with the system. This is the seventh property identified.

Metric types. : The evaluation metrics used in our surveyed publications can be classified into system's *desiderata* (DES) (60.4% of surveyed publications), users' ability to perform *cognitive* (COG) tasks (35.8%), and users' ability to take certain *actions* (ACT) (30.2%) such as to learn, control, or respond. This is the eight and last property. It is worth noting that almost all publications evaluate the overall system where XAI is part of; no publication evaluated the EI alone except for only 1 publication from [10] (where they evaluate the design principles of contextualization and allow exploration which are applied on the interface). The results in Fig. 6 are from evaluations of the entire system.

The popular area on the right of Fig. 6 indicates that the majority of publications conduct evaluation for system's desiderata. System's effectiveness is the most common (37.7%) evaluation metric for the XAI system. System's effectiveness is concered with accuracy (e.g., [27, 65]), helpfulness (e.g., [45, 71, 93]), and quality (e.g., [64, 77]).

In terms of evaluating users ability to perform actions, the most common metric is examining how users accomplish tasks (26.4%). Users' ability to accomplish tasks are measured with participants' object performance (e.g., [11, 21]), complete/error rate (e.g., [20, 21, 49, 75, 91]), completion time (e.g., [51, 75, 93]).

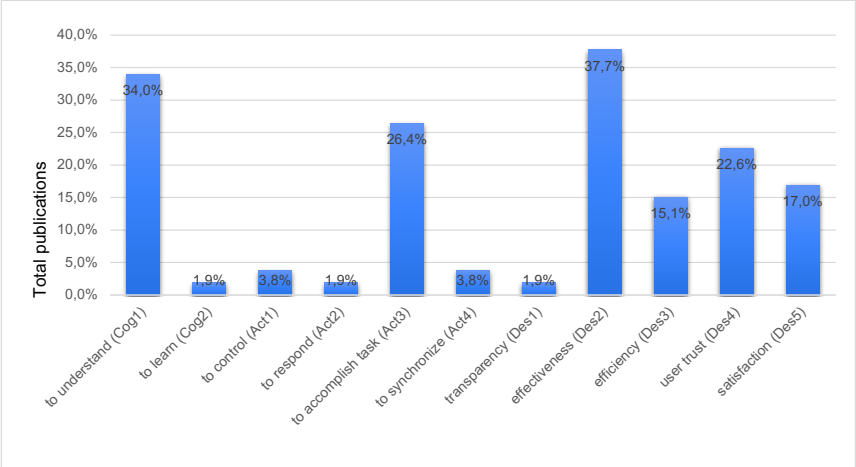

Fig. 6. Specific evaluation metrics (including their types)

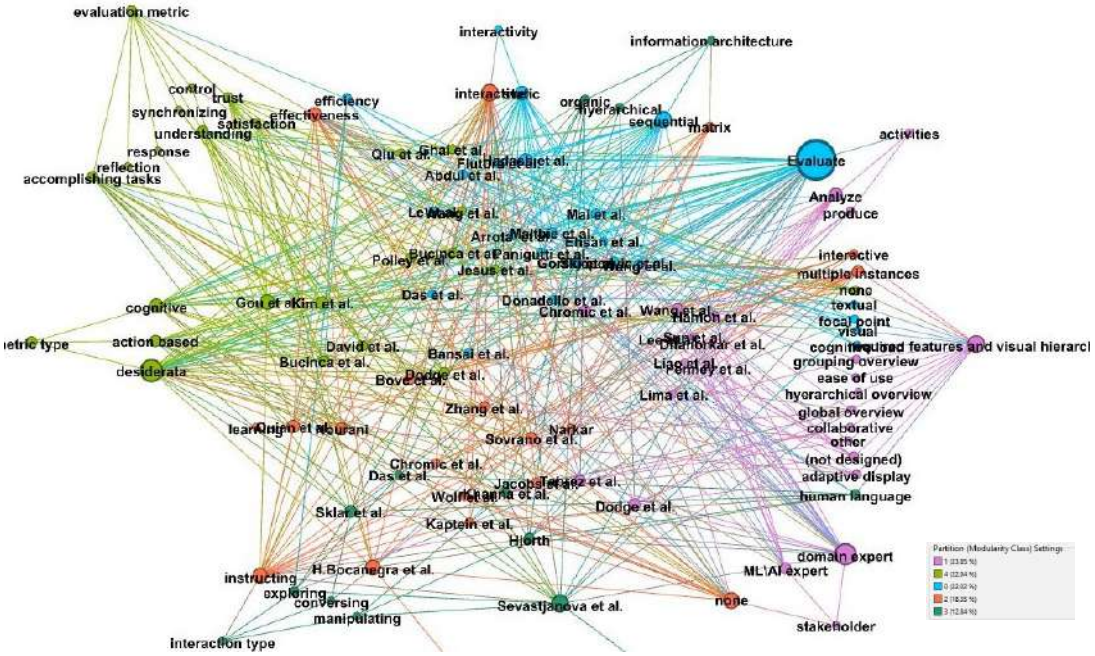

Fig. 7. Overall Cluster Analysis of Literature

While performance task is a goal-driven metric, other action-driven metrics area not often measured. There are a few publications that evaluate users' ability to control (3.8%) (e.g., a sense of control users have [16, 37]), and ability to respond (1.9%).

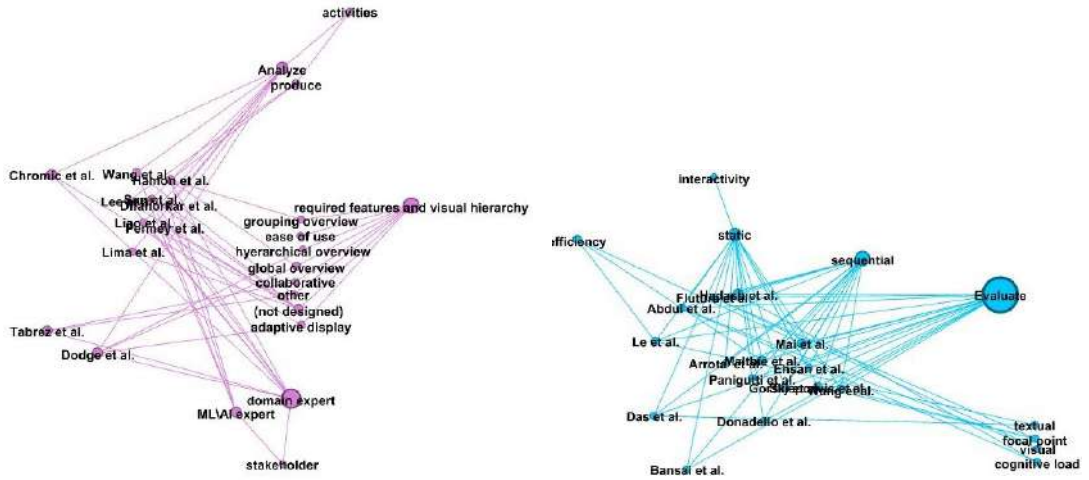

Fig. 8. Left: Detailed View of Cluster 1. Right: Detailed View of Cluster 0

4.4 Cluster analysis of literature

In order to unfold more complex multidimensional relations within the XAI field, we decided to cluster all the surveyed publications (see Fig. 7). The network graph was created in Gephy (https://gephi.org/), an open-source network analysis and visualization package, using the modularity measure. The intent was to identify clusters of publications with a similar configuration of the 8 properties listed in Table 1 (*Participant group, Activities, Required Features and Visual Hierarchy, Interaction Type, Interactivity, Information Architecture, Evaluation Metric, Metric Type*). Modularity is a measure used in Social Network Analysis and is defined as the fraction of the edges that fall within the given groups minus the expected fraction if edges were distributed at random; while communities are defined as having dense connections between the nodes within modules but sparse connections between nodes in different modules. Modularity was preferred to traditional clustering algorithms such as K-means because of three reasons: a) modularity analysis in Gephy allows visualizing the communities detected on a network graph, this feature is highly prized as it facilitates human interpretation and understanding of the nature of the resulting clusters, b) modularity performs well with highly dimensional but sparse data, while both k-means and k-mode struggle with that, and c) modularity is a-parametric and does not require assumptions such as stating a priori the number of desired clusters, as it is necessary with k-means. Gephy implements modularity analysis with the Louvain method. In the network graph shown in Fig. 7, the colors represent the 5 communities (or clusters) identified. Cluster 0 (light blue) captures 22.02% of the publications examined, cluster 1 (pink) covers 23,85%, cluster 2 (orange) includes 18,35%, cluster 3 (dark green) encloses 12,84% and cluster 4 (light green) catches the remaining 22,94%. At the center, all 53 publications are represented, while on the periphery of the graph the 8 properties with their relative categories are used to define the publications. As Fig. 7, most publications involved humans in evaluation. Below, as an example, we will analyze the largest clusters 0 and 1.

4.4.1 Cluster 1. As shown in Fig. 8 (Left) this cluster is defined by *Domain Experts* as stakeholders sometimes paired by *ML\AI experts* as well as *Analysis* and *Produce* activities.

This pattern is indicative of a need to understand how humans make sense of XAI system in the context of games. For example, Dodge et al. [26] attempt to understand human reasoning processes when performing an assessment

of the system's decisions in a real-time strategy game. Also, domain experts' sense-making can confirm whether the applied cognitive theories in XAI system are effective. For example, Information Foraging Theory relates to the understanding of how humans search for information, this theory is adopted in the context of a real-time strategy game to understand how domain experts (players) understand and seek explanations for the actions of the system [80]. Similarly, Tabrez el al. [97] evaluate the cognitive theory of how humans cope with misunderstandings of a system's behavior by introducing rewards and explanations. We assume that since domain experts are already familiar with the studied domain and thus require less training to interact with the system, they require less effort in making sense of the domain. In contrast, if they are unfamiliar with the domain, they might have to put more effort into making sense of the new domain. Hence, the sense-making process for domain experts is less likely to be affected by the unfamiliar domain. We suspect that this could be the reason why domain experts are preferred when evaluating cognitive theories.

This cluster is also defined by a pattern of activities that exclude *Evaluate* but focuses on *Analize* and *Produce*, We hypothesize that domain experts are more involved in the *Produce* activity because the domains under examination are higher stake domains such as clinical and medical. Specifically, intensive care unit (ICU) clinicians are part of the co-design sessions to provide their diagnostic reasoning process and the usage of explanation features as they interact with the interface [103]. Similarly, ICU medical staff were in the loop to validate the decision that the AI system makes on the diagnosis of lung cancer [40]. Also, therapists annotate the exercise dataset from stroke and healthy subjects [56]. Since clinical and medical domains have higher stakes for users, it is essential to include domain experts while building the system.

4.4.2 Cluster 0. As shown in Fig. 8 (Right) this cluster is defined by *Evaluate* activity, *Sequential* information architecture and *Static* interactivity.

Publications that involve participants to evaluate the XAI systems tend to adopt 'black-box' machine learning models and apply agnostic approaches to explain how the model works. Since the main problems of models such as convolution neural networks (e.g., [38]), deep neural networks (e.g., [5]), neural networks (e.g., [31, 55]) is low interpretability, XAI researchers prioritize to evaluate the generated explanation with the interpretability criteria.

Furthermore, since the textual format communicates clearly to users, XAI researchers adopt textual explanations to increase the interpretability level of such black-box models. Specifically, they design user studies with text classifications (e.g., [5, 38]), or adopt Natural Language to increase intuitiveness and understandability of the explanation (e.g., [5, 55]). Textual explanation can also embed semantic features (e.g., [39]) that aid users' sense-making process.

Additionally, processing textual format imposes more cognitive load than visual format, so the content of the explanations should be presented in a sequential order to avoid cognitive overload. When explanations are presented gradually, users are able to understand the explanations better.

5 DISCUSSION

5.1 More attention and new methods are needed to understand user needs for explanations.

Based on our analysis, we find that gathering user requirements before developing an XAI system is not a widespread practice. This finding provides further empirical evidence for why existing XAI research lack usability, practical interpretability, and efficacy for real users [1, 28, 66, 111]. If researchers are not clear about users' needs, their skills, and how they process technical explanations, it is not surprising that real users have difficulties using their XAI systems.

We hypothesize that there are several reasons for this. First, XAI has so far been primarily a technical research field. Its focus has been developing new algorithms that make black-box ML *interpretable*. Under this discipline, XAI systems

are mostly seen as a prototype to showcase and test algorithmic feasibility. However, interpretability only means that the resulting ML models *can* potentially be interpreted. A significant amount of research and HCI design is needed to turn something interpretable into actual explanations that real users find understandable and useful. Second, we believe another reason is that established user research methods are insufficient to gather users' explanation needs. Explanations of AI are highly technical in nature, and there have not been a lot of existing examples for an average user to establish expectations. As a result, most users have problems articulating what their needs are, when it comes to XAI. In turn, even if a research team carries out time-consuming user research tasks, they may not find enough information to extract well-defined user requirements. This methodological challenge is echoed in all aspects of UX design for ML and AI products[109]. We suggest that XAI researchers can use first person methods[63], such as autoethnography[13, 62] and autobiographical design methods[23, 72], to build on their own knowledge. We also suggest more collaboration with HCI designers to provide heuristic evaluations [74] as experts on user interface and user experience.

5.2 Interactivity is a key towards real explanability and actionable understanding.

An increasing amount of recent evidence suggests that interactive explanations can be more user-friendly [57, 67, 103]. However, less than half of the EIs in our surveyed publications are interactive. Among them, many offer limited interactivity such as letting users click through a pre-defined sequence of explanation content. There is also a lack of XAI work on scaffolding, user engagement, and other design methods to structure user interaction. This significantly limits XAI's ability to support more complex interactions around explanations, which may be essential for non-experts. We encourage XAI researchers and UX practitioners design interactive comparing interface to address this needs. A promising direction is to design the interface with matrix information architecture, where it supports interactivity and organize explanations' features in multiple dimensions. Recent work such as [100] offers a promising direction towards exploring new interaction formats of EIs through computer games.

5.3 Evaluation of EI and XAI should reduce confounding factors.

Our findings suggest that while researchers have adopted a wide variety of evaluation metrics, most publications evaluated the performance of the larger system where the EI is a component of. For example, Hjorth [43] designed the XAI interface to help students learn about the policy views and political communications through a machine learning system adopting Natural Language. While XAI interface is an important component to help users communicate with the system, he does not evaluate the interface itself, but evaluate the whole system. This means that there are a lot of confounding factors when XAI researchers try to derive implications from their evaluation results. When the users report overall satisfaction with the entire system, it is difficult to tease out which components contribute to it and in what ways. We encourage future EI research to conduct more targeted evaluation of EIs and their components before evaluating the larger system where XAI is part of.

Related, XAI researchers heavily rely on predefined desiderata (e.g., *trust, causality, transferability, informativeness, and fair and ethical decision making* [61]) as evaluation metrics. While these desiderata offer a general direction, they are often too general to target the specific requirements or to be operationalized as evaluation metrics in a given context of use. For instance, what is considered trustworthy or informative can differ vastly from AI experts to novice end-users. Similarly, how to measure trust is a complicated task in itself. We believe that conducting user research, especially using qualitative methods (e.g., developing user group profiles, task models, and personas) at the early stage, can be a highly effective way to supplement the context-free desiderata.

5.4 Reviewing framework as a generative tool

The work here presented analyzes existing research utilizing eight properties. As seen in figures 2, 3, 5 and 6, the frequency distribution of publications per property is generally quite skewed, for example there is a lot of work that has sequential information structure while very few utilize hierarchical structure, or again the majority of publications examined adopts instructing interaction type while very few implemented a conversational approach. It is worth asking whether maybe there are opportunities delineated by the negative space shown by this analysis, for example there are no instances of EI that implement static interactivity AND hierarchical information architecture, is that because it simply does not make sense or could it be a missed opportunity for future work? Examining even just one of the clusters identified in section 4.4 it is possible to hypothesize EI that do not exist but might improve transparency and interpretability. For example looking at cluster 1, we could imagine an EI that utilizes UX and UI experts rather than domain experts and focuses on evaluation rather than analyses or production.

The analysis here presented could be turned into a generative framework by looking at combinations of properties that are not represented in the dataset.

6 LIMITATIONS

There are several limitations of this paper. First, our systematic search is carried only in the ACM Digital Library database due to the HCI focus on our subject. It is likely that some relevant research is published in other venues, such as NeurIPS, ICML, or arXiv, and therefore is not included in our survey. Since *explainable interface* is an emerging terminology that XAI researcher only started using very recently, all papers must be manually scanned for relevance. A broader search is thus out of scope. We believe that, as one of the first survey on this topic, the number of papers resulted from our systematic search included in this paper is large enough to provide a representative sample of emerging research in explainable interface. Among the six related review articles, 50% of them also only use the ACM Digital Library.

Second, our analysis is based on what the researchers report in their publication. It is possible that, depending on the focus of the paper, researchers may omit certain details about user participation. For instance, if the researchers conducted informal user research to gather user needs, they may not report it in a publication where the focus is on XAI algorithms. Furthermore, there has been growing recognition of *first person methods* [63] that uses HCI researchers' own experience for data collection, as opposed to external users. These first person methods are not considered as part of the process to define user requirements in our analysis.

7 CONCLUSION

Despite its technological breakthroughs, eXplainable Artificial Intelligence (XAI) research has limited success producing the *effective explanations* needed by users. In order to improve XAI systems' usability, practical interpretability, and efficacy for real users, the emerging area of *Explainable Interfaces* (EIs) focuses on the user interface and user experience design aspects of XAI. This paper presents a systematic survey of 53 publications to identify current trends in human-XAI interaction and promising directions for EI design and development. This is among the first systematic survey of EI research.

In conclusion, we conducted a systematic literature review of 53 publications about the design and evaluation of EIs. Compared to existing technical review articles that focuses on *what* to explain, we focuses on *how* current XAI communicate the explanation to human users. In fact, is is only through unraveling the modalities of how EI

can effectively function that we can radically improve future designs. Through three guiding research questions, we examined how participants are involved as well as how EIs are designed and evaluated. Using cluster analysis, we identify current trends in human-XAI interaction and promising directions for EI design and development.

ACKNOWLEDGMENTS

This work is supported by the Danish Novo Nordisk Foundation under Grant Number NNF20OC0066119.

REFERENCES

[1] Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y Lim, and Mohan Kankanhalli. 2018. 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*. 1–18.

[2] Ashraf Abdul, Christian von der Weth, Mohan Kankanhalli, and Brian Y. Lim. 2020. COGAM: Measuring and Moderating Cognitive Load in Machine Learning Model Explanations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–14. https://doi.org/10.1145/3313831.3376615

[3] Amina Adadi and Mohammed Berrada. 2018. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE access* 6 (2018), 52138–52160.

[4] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion* 58 (2020), 82–115.

[5] Luca Arrotta, Gabriele Civitarese, and Claudio Bettini. 2022. DeXAR: Deep Explainable Sensor-Based Activity Recognition in Smart-Home Environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 1 (March 2022), 1–30. https://doi.org/10.1145/3517224

[6] Gagan Bansal, Tongshuang Wu, Joyce Zhou, Raymond Fok, Besmira Nushi, Ece Kamar, Marco Tulio Ribeiro, and Daniel Weld. 2021. Does the Whole Exceed its Parts? The Effect of AI Explanations on Complementary Team Performance. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–16. https://doi.org/10.1145/3411764.3445717

[7] Kathy Baxter, Catherine Courage, and Kelly Caine. 2015. *Understanding your users: a practical guide to user research methods*. Morgan Kaufmann.

[8] Daniel Ben David, Yehezkel S. Resheff, and Talia Tron. 2021. Explainable AI and Adoption of Financial Algorithmic Advisors: An Experimental Study. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, Virtual Event USA, 390–400. https://doi.org/10.1145/3461702.3462565

[9] Alexander Binder, Grégoire Montavon, Sebastian Lapuschkin, Klaus-Robert Müller, and Wojciech Samek. 2016. Layer-Wise Relevance Propagation for Neural Networks with Local Renormalization Layers. In *Artificial Neural Networks and Machine Learning*, Villa A., Masulli P., and Pons Rivero A (Eds.).

[10] Clara Bove, Jonathan Aigrain, Marie-Jeanne Lesot, Charles Tijus, and Marcin Detyniecki. 2022. Contextualization and Exploration of Local Feature Importance Explanations to Improve Understanding and Satisfaction of Non-Expert Users. In *27th International Conference on Intelligent User Interfaces*. ACM, Helsinki Finland, 807–819. https://doi.org/10.1145/3490099.3511139

[11] Zana Buçinca, Phoebe Lin, Krzysztof Z. Gajos, and Elena L. Glassman. 2020. Proxy tasks and subjective measures can be misleading in evaluating explainable AI systems. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*. ACM, Cagliari Italy, 454–464. https://doi.org/10.1145/3377325.3377498

[12] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z. Gajos. 2021. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 1–21. https://doi.org/10.1145/3449287

[13] Marta E Cecchinato, Anna L Cox, and Jon Bird. 2017. Always on (line)? User experience of smartwatches and their role within multi-device ecologies. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 3557–3568.

[14] Michael Chromik and Andreas Butz. 2021. Human-XAI Interaction: A Review and Design Principles for Explanation User Interfaces. In *Human-Computer Interaction – INTERACT 2021 (Lecture Notes in Computer Science)*, Carmelo Ardito, Rosa Lanzilotti, Alessio Malizia, Helen Petrie, Antonio Piccinno, Giuseppe Desolda, and Kori Inkpen (Eds.). Springer International Publishing, Cham, 619–640. https://doi.org/10.1007/978-3-030-85616-8_36

[15] Michael Chromik, Malin Eiband, Felicitas Buchner, Adrian Krüger, and Andreas Butz. 2021. I Think I Get Your Point, AI! The Illusion of Explanatory Depth in Explainable AI. In *26th International Conference on Intelligent User Interfaces*. ACM, College Station TX USA, 307–317. https://doi.org/10.1145/3397481.3450644

[16] Michael Chromik, Florian Fincke, and Andreas Butz. 2020. Mind the (persuasion) gap: contrasting predictions of intelligent DSS with user beliefs to improve interpretability. In *Proceedings of the 12th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*. ACM, Sophia Antipolis France, 1–6. https://doi.org/10.1145/3393672.3398491

[17] Juliet Corbin and Anselm Strauss. 2014. *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage publications.

[18] Marina Danilevsky, Kun Qian, Ranit Aharonov, Yannis Katsis, Ban Kawas, and Prithviraj Sen. 2020. A survey of the state of explainable AI for natural language processing. *arXiv preprint arXiv:2010.00711* (2020).

[19] Arun Das and Paul Rad. 2020. Opportunities and challenges in explainable artificial intelligence (xai): A survey. *arXiv preprint arXiv:2006.11371* (2020).

[20] Devleena Das, Siddhartha Banerjee, and Sonia Chernova. 2021. Explainable AI for Robot Failures: Generating Explanations that Improve User Assistance in Fault Recovery. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Boulder CO USA, 351–360. https://doi.org/10.1145/3434073.3444657

[21] Devleena Das and Sonia Chernova. 2020. Leveraging rationales to improve human task performance. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*. ACM, Cagliari Italy, 510–518. https://doi.org/10.1145/3377325.3377512

[22] Augustin Degas, Mir Riyanul Islam, Christophe Hurter, Shaibal Barua, Hamidur Rahman, Minesh Poudel, Daniele Ruscio, Mobyen Uddin Ahmed, Shahina Begum, Md Aquif Rahman, et al. 2022. A survey on artificial intelligence (ai) and explainable ai in air traffic management: Current trends and development with future research trajectory. *Applied Sciences* 12, 3 (2022), 1295.

[23] Audrey Desjardins and Aubree Ball. 2018. Revealing tensions in autobiographical design in HCI. In *proceedings of the 2018 designing interactive systems conference*. 753–764.

[24] Shipi Dhanorkar, Christine T. Wolf, Kun Qian, Anbang Xu, Lucian Popa, and Yunyao Li. 2021. Who needs to know what, when?: Broadening the Explainable AI (XAI) Design Space by Looking at Explanations Across the AI Lifecycle. In *Designing Interactive Systems Conference 2021*. ACM, Virtual Event USA, 1591–1602. https://doi.org/10.1145/3461778.3462131

[25] Jonathan Dodge, Andrew A. Anderson, Matthew Olson, Rupika Dikkala, and Margaret Burnett. 2022. How Do People Rank Multiple Mutant Agents?. In *27th International Conference on Intelligent User Interfaces*. ACM, Helsinki Finland, 191–211. https://doi.org/10.1145/3490099.3511115

[26] Jonathan Dodge, Roli Khanna, Jed Irvine, Kin-ho Lam, Theresa Mai, Zhengxian Lin, Nicholas Kiddle, Evan Newman, Andrew Anderson, Sai Raja, Caleb Matthews, Christopher Perdriau, Margaret Burnett, and Alan Fern. 2021. After-Action Review for AI (AAR/AI). *ACM Transactions on Interactive Intelligent Systems* 11, 3-4 (Dec. 2021), 1–35. https://doi.org/10.1145/3453173

[27] Ivan Donadello, Mauro Dragoni, and Claudio Eccher. 2020. Explaining reasoning algorithms with persuasiveness: a case study for a behavioural change system. In *Proceedings of the 35th Annual ACM Symposium on Applied Computing*. ACM, Brno Czech Republic, 646–653. https://doi.org/10.1145/3341105.3373910

[28] Finale Doshi-Velez and Been Kim. 2017. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608* (2017).

[29] Filip Karlo Došilović, Mario Brčić, and Nikica Hlupić. 2018. Explainable artificial intelligence: A survey. In *2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO)*. IEEE, 0210–0215.

[30] Upol Ehsan, Q. Vera Liao, Michael Muller, Mark O. Riedl, and Justin D. Weisz. 2021. Expanding Explainability: Towards Social Transparency in AI systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–19. https://doi.org/10.1145/3411764.3445188

[31] Simon Flutura, Andreas Seiderer, Tobias Huber, Katharina Weitz, Ilhan Aslan, Ruben Schlagowski, Elisabeth André, and Joachim Rathmann. 2020. Interactive Machine Learning and Explainability in Mobile Classification of Forest-Aesthetics. In *Proceedings of the 6th EAI International Conference on Smart Objects and Technologies for Social Good*. ACM, Antwerp Belgium, 90–95. https://doi.org/10.1145/3411170.3411225

[32] Jesse James Garret. 2003. The elements of user experience: user-centered design for the web. *Nueva York, NY: AIGA* (2003).

[33] Thomas Geis, Knut Polkehn, Rolf Molich, and Oliver Kluge. 2016. CPUX-UR Curriculum. *UXQB e. V* (2016).

[34] Bhavya Ghai, Q. Vera Liao, Yunfeng Zhang, Rachel Bellamy, and Klaus Mueller. 2021. Explainable Active Learning (XAL): Toward AI Explanations as Interfaces for Machine Teachers. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (Jan. 2021), 1–28. https://doi.org/10.1145/3432934

[35] Kelly Gordon. 2021. Visual Hierarchy in UX: Definition. https://www.nngroup.com/articles/visual-hierarchy-ux-definition/

[36] David Gunning. 2017. Explainable artificial intelligence (xai). *Defense advanced research projects agency (DARPA), nd Web* 2, 2 (2017), 1.

[37] Lijie Guo, Elizabeth M. Daly, Oznur Alkan, Massimiliano Mattetti, Owen Cornec, and Bart Knijnenburg. 2022. Building Trust in Interactive Machine Learning via User Contributed Interpretable Rules. In *27th International Conference on Intelligent User Interfaces*. ACM, Helsinki Finland, 537–548. https://doi.org/10.1145/3490099.3511111

[38] Łukasz Górski and Shashishekar Ramakrishna. 2021. Explainable artificial intelligence, lawyer's perspective. In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law*. ACM, São Paulo Brazil, 60–68. https://doi.org/10.1145/3462757.3466145

[39] Sophia Hadash, Martijn C. Willemsen, Chris Snijders, and Wijnand A. IJsselsteijn. 2022. Improving understandability of feature contributions in model-agnostic explainable AI tools. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–9. https://doi.org/10.1145/3491102.3517650

[40] Ronan Hamon, Henrik Junklewitz, Gianclaudio Malgieri, Paul De Hert, Laurent Beslay, and Ignacio Sanchez. 2021. Impossible Explanations?: Beyond explainable AI in the GDPR from a COVID-19 use case scenario. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Virtual Event Canada, 549–559. https://doi.org/10.1145/3442188.3445917

[41] Jonathan L. Herlocker, Joseph A. Konstan, and John Riedl. 2000. Explaining collaborative filtering recommendations. *Proceedings of the 2000 ACM conference on Computer supported cooperative work - CSCW '00* (2000), 241–250. https://doi.org/10.1145/358916.358995

[42] Diana C. Hernandez-Bocanegra and Jürgen Ziegler. 2021. Conversational review-based explanations for recommender systems: Exploring users' query behavior. In *CUI 2021 - 3rd Conference on Conversational User Interfaces*. ACM, Bilbao (online) Spain, 1–11. https://doi.org/10.1145/3469595.

3469596

[43] Arthur Hjorth. 2021. NaturalLanguageProcesing4All: - A Constructionist NLP tool for Scaffolding Students' Exploration of Text. In *Proceedings of the 17th ACM Conference on International Computing Education Research*. ACM, Virtual Event USA, 347–354. https://doi.org/10.1145/3446871.3469749

[44] Kasper Hornbæk and Antti Oulasvirta. 2017. What is interaction?. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 5040–5052.

[45] Maia Jacobs, Jeffrey He, Melanie F. Pradier, Barbara Lam, Andrew C. Ahn, Thomas H. McCoy, Roy H. Perlis, Finale Doshi-Velez, and Krzysztof Z. Gajos. 2021. Designing AI for Trust and Collaboration in Time-Constrained Medical Decisions: A Sociotechnical Lens. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–14. https://doi.org/10.1145/3411764.3445385

[46] Sérgio Jesus, Catarina Belém, Vladimir Balayan, João Bento, Pedro Saleiro, Pedro Bizarro, and João Gama. 2021. How can I choose an explainer?: An Application-grounded Evaluation of Post-hoc Explanations. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Virtual Event Canada, 805–815. https://doi.org/10.1145/3442188.3445941

[47] José Jiménez-Luna, Francesca Grisoni, and Gisbert Schneider. 2020. Drug discovery with explainable artificial intelligence. *Nature Machine Intelligence* 2, 10 (2020), 573–584.

[48] Frank Kaptein, Bernd Kiefer, Antoine Cully, Oya Celiktutan, Bert Bierman, Rifca Rijgersberg-peters, Joost Broekens, Willeke Van Vught, Michael Van Bekkum, Yiannis Demiris, and Mark A. Neerincx. 2022. A Cloud-based Robot System for Long-term Interaction: Principles, Implementation, Lessons Learned. *ACM Transactions on Human-Robot Interaction* 11, 1 (March 2022), 1–27. https://doi.org/10.1145/3481585

[49] Roli Khanna, Jonathan Dodge, Andrew Anderson, Rupika Dikkala, Jed Irvine, Zeyad Shureih, Kin-Ho Lam, Caleb R. Matthews, Zhengxian Lin, Minsuk Kahng, Alan Fern, and Margaret Burnett. 2022. Finding AI's Faults with AAR/AI: An Empirical Study. *ACM Transactions on Interactive Intelligent Systems* 12, 1 (March 2022), 1–33. https://doi.org/10.1145/3487065

[50] Chris Kim, Xiao Lin, Christopher Collins, Graham W. Taylor, and Mohamed R. Amer. 2021. Learn, Generate, Rank, Explain: A Case Study of Visual Explanation by Generative Machine Learning. *ACM Transactions on Interactive Intelligent Systems* 11, 3-4 (Dec. 2021), 1–34. https://doi.org/10.1145/3465407

[51] Sebin Kim and Jihwan Woo. 2021. Explainable AI framework for the financial rating models: Explaining framework that focuses on the feature influences on the changing classes or rating in various customer models used by the financial institutions.. In *2021 10th International Conference on Computing and Pattern Recognition*. ACM, Shanghai China, 252–255. https://doi.org/10.1145/3497623.3497664

[52] Yubo Kou and Xinning Gui. 2020. Mediating Community-AI Interaction through Situated Explanation: The Case of AI-Led Moderation. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (Oct. 2020), 1–27. https://doi.org/10.1145/3415173

[53] Todd Kulesza, Margaret Burnett, Weng-keen Wong, and Simone Stumpf. 2015. Principles of explanatory debugging to personalize interactive machine learning. In *Proceedings of the 20th international conference on intelligent user interfaces*. 126–137.

[54] Todd Kulesza, Simone Stumpf, Margaret Burnett, Sherry Yang, Irwin Kwan, and Weng Keen Wong. 2013. Too much, too little, or just right? Ways explanations impact end users' mental models. *Proceedings of IEEE Symposium on Visual Languages and Human-Centric Computing, VL/HCC* (2013), 3–10. https://doi.org/10.1109/VLHCC.2013.6645235

[55] Thai Le, Suhang Wang, and Dongwon Lee. 2020. GRACE: Generating Concise and Informative Contrastive Sample to Explain Neural Network Model's Prediction. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, Virtual Event CA USA, 238–248. https://doi.org/10.1145/3394486.3403066

[56] Min Hun Lee, Daniel P. Siewiorek, Asim Smailagic, Alexandre Bernardino, and Sergi Bermúdez i Badia. 2020. An Exploratory Study on Techniques for Quantitative Assessment of Stroke Rehabilitation Exercises. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*. ACM, Genoa Italy, 303–307. https://doi.org/10.1145/3340631.3394872

[57] Q. Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: Informing Design Practices for Explainable AI User Experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–15. https://doi.org/10.1145/3313831.3376590

[58] Q. Vera Liao and Kush R. Varshney. 2022. Human-Centered Explainable AI (XAI): From Algorithms to User Experiences. http://arxiv.org/abs/2110.10790 arXiv:2110.10790 [cs].

[59] Antonios Liapis and Jichen Zhu. 2022. The Need for Explainability in AI-Based Creativity Support Tools. In *Proceedings of the Human Centered AI workshop at NeurIPS 2022*.

[60] Gabriel Lima, Nina Grgić-Hlača, and Meeyoung Cha. 2021. Human Perceptions on Moral Responsibility of AI: A Case Study in AI-Assisted Bail Decision-Making. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–17. https://doi.org/10.1145/3411764.3445260

[61] Zachary C Lipton. 2018. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue* 16, 3 (2018), 31–57.

[62] Andrés Lucero. 2018. Living without a mobile phone: An autoethnography. In *Proceedings of the 2018 Designing Interactive Systems Conference*. 765–776.

[63] Andrés Lucero, Audrey Desjardins, Carman Neustaedter, Kristina Höök, Marc Hassenzahl, and Marta E. Cecchinato. 2019. A sample of one: First-person research methods in HCI. In *DIS 2019 Companion - Companion Publication of the 2019 ACM Designing Interactive Systems Conference*. 385–388. https://doi.org/10.1145/3301019.3319996

[64] Theresa Mai, Roli Khanna, Jonathan Dodge, Jed Irvine, Kin-Ho Lam, Zhengxian Lin, Nicholas Kiddle, Evan Newman, Sai Raja, Caleb Matthews, Christopher Perdriau, Margaret Burnett, and Alan Fern. 2020. Keeping It "Organized and Logical": After-Action Review for AI (AAR/AI). (2020), 12.

[65] Nicholas Maltbie, Nan Niu, Matthew Van Doren, and Reese Johnson. 2021. XAI tools in the public sector: a case study on predicting combined sewer overflows. In *Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. ACM, Athens Greece, 1032–1044. https://doi.org/10.1145/3468264.3468547

[66] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence* 267 (2019), 1–38.

[67] Tim Miller, Piers Howe, and Liz Sonenberg. 2017. Explainable AI: Beware of inmates running the asylum or: How i learnt to stop worrying and love the social and behavioural sciences. In *Proceedings of the IJCAI Workshop on Workshop on Explainable Artificial Intelligence*.

[68] David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G Altman, and PRISMA Group*. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine* 151, 4 (2009), 264–269.

[69] Sina Mohseni, Niloofar Zarei, and Eric D. Ragan. 2021. A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems. *ACM Transactions on Interactive Intelligent Systems* 11, 3-4 (Dec. 2021), 1–45. https://doi.org/10.1145/3387166

[70] Menaka Narayanan, Emily Chen, Jeffrey He, Been Kim, Sam Gershman, and Finale Doshi-Velez. 2018. *How do humans understand explanations from machine learning systems? An evaluation of the human-interpretability of explanation.* Technical Report.

[71] Shweta Narkar, Yunfeng Zhang, Q. Vera Liao, Dakuo Wang, and Justin D. Weisz. 2021. Model LineUpper: Supporting Interactive Model Comparison at Multiple Levels for AutoML. In *26th International Conference on Intelligent User Interfaces*. ACM, College Station TX USA, 170–174. https://doi.org/10.1145/3397481.3450658

[72] Carman Neustaedter and Phoebe Sengers. 2012. Autobiographical design in HCI research: designing and learning through use-it-yourself. In *Proceedings of the Designing Interactive Systems Conference*. 514–523.

[73] Thu Nguyen and Jichen Zhu. 2022. Towards Better User Requirements: How to Involve Human Participants in XAI Research. *arXiv preprint arXiv:2212.03186* (2022).

[74] Jakob Nielsen and Rolf Molich. 1990. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 249–256.

[75] Mahsan Nourani, Chiradeep Roy, Jeremy E Block, Donald R Honeycutt, Tahrima Rahman, Eric Ragan, and Vibhav Gogate. 2021. Anchoring Bias Affects Mental Model Formation and User Reliance in Explainable AI Systems. In *26th International Conference on Intelligent User Interfaces*. ACM, College Station TX USA, 340–350. https://doi.org/10.1145/3397481.3450639

[76] Ingrid Nunes and Dietmar Jannach. 2017. A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction* 27, 3 (2017), 393–444.

[77] Cecilia Panigutti, Andrea Beretta, Fosca Giannotti, and Dino Pedreschi. 2022. Understanding the impact of explanations on advice-taking: a user study for AI-based clinical Decision Support Systems. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–9. https://doi.org/10.1145/3491102.3502104

[78] Cecilia Panigutti, Alan Perotti, and Dino Pedreschi. 2020. Doctor XAI: an ontology-based approach to black-box sequential data classification explanations. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. ACM, Barcelona Spain, 629–639. https://doi.org/10.1145/3351095.3372855

[79] Seyedeh Neelufar Payrovnaziri, Zhaoyi Chen, Pablo Rengifo-Moreno, Tim Miller, Jiang Bian, Jonathan H Chen, Xiuwen Liu, and Zhe He. 2020. Explainable artificial intelligence models using real-world electronic health record data: a systematic scoping review. *Journal of the American Medical Informatics Association* 27, 7 (2020), 1173–1185.

[80] Sean Penney, Jonathan Dodge, Claudia Hilderbrand, Andrew Anderson, Logan Simpson, and Margaret Burnett. 2018. Toward Foraging for Understanding of StarCraft Agents: An Empirical Study. In *23rd International Conference on Intelligent User Interfaces*. ACM, Tokyo Japan, 225–237. https://doi.org/10.1145/3172944.3172946

[81] Sayantan Polley, Rashmi Raju Koparde, Akshaya Bindu Gowri, Maneendra Perera, and Andreas Nuernberger. 2021. Towards Trustworthiness in the Context of Explainable Search. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, Virtual Event Canada, 2580–2584. https://doi.org/10.1145/3404835.3462799

[82] Peizhu Qian. 2022. Evaluating the Role of Interactivity on Improving Transparency in Autonomous Agents. (2022), 9.

[83] Luyu Qiu, Yi Yang, Caleb Chen Cao, Yueyuan Zheng, Hilary Ngai, Janet Hsiao, and Lei Chen. 2022. Generating Perturbation-based Explanations with Robustness to Out-of-Distribution Data. In *Proceedings of the ACM Web Conference 2022*. ACM, Virtual Event, Lyon France, 3594–3605. https://doi.org/10.1145/3485447.3512254

[84] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data minin*.

[85] Yvonne Rogers, Helen Sharp, and Jenny Preece. 2011. Interaction design. *Interaction Design: Beyond Human-Computer Interaction, Wiley* (2011), 1–34.

[86] Thomas Rojat, Raphaël Puget, David Filliat, Javier Del Ser, Rodolphe Gelin, and Natalia Díaz-Rodríguez. 2021. Explainable artificial intelligence (xai) on timeseries data: A survey. *arXiv preprint arXiv:2104.00950* (2021).

[87] James Schaffer, Prasanna Giridhar, Debra Jones, Tobias Höllerer, Tarek Abdelzaher, and John O'donovan. 2015. Getting the message? A study of explanation interfaces for microblog data analysis. In *International Conference on Intelligent User Interfaces, Proceedings IUI*, Vol. 2015-Janua.

345–356. https://doi.org/10.1145/2678025.2701406

[88] Rita Sevastjanova, Wolfgang Jentner, Fabian Sperrle, Rebecca Kehlbeck, Jürgen Bernard, and Mennatallah El-assady. 2021. QuestionComb: A Gamification Approach for the Visual Explanation of Linguistic Phenomena through Interactive Labeling. *ACM Transactions on Interactive Intelligent Systems* 11, 3-4 (Dec. 2021), 1–38. https://doi.org/10.1145/3429448

[89] Avanti Shrikumar, Peyton Greenside, and Anshul Kundaje. 2017. Learning Important Features Through Propagating Activation Differences. In *Proceedings of the 34th International Conference on Machine Learning*.

[90] David Silverman. 2020. *Qualitative research.* sage.

[91] Elizabeth I. Sklar and Mohammad Q. Azhar. 2018. Explanation through Argumentation. In *Proceedings of the 6th International Conference on Human-Agent Interaction (HAI '18)*. Association for Computing Machinery, New York, NY, USA, 277–285. https://doi.org/10.1145/3284432.3284470

[92] Djordje Slijepcevic, Fabian Horst, Sebastian Lapuschkin, Brian Horsak, Anna-Maria Raberger, Andreas Kranzl, Wojciech Samek, Christian Breiteneder, Wolfgang Immanuel Schöllhorn, and Matthias Zeppelzauer. 2022. Explaining Machine Learning Models for Clinical Gait Analysis. *ACM Transactions on Computing for Healthcare* 3, 2 (April 2022), 1–27. https://doi.org/10.1145/3474121

[93] Francesco Sovrano and Fabio Vitali. 2021. From Philosophy to Interfaces: an Explanatory Method and a Tool Inspired by Achinstein's Theory of Explanation. In *26th International Conference on Intelligent User Interfaces*. ACM, College Station TX USA, 81–91. https://doi.org/10.1145/3397481.3450655

[94] Jiao Sun, Q. Vera Liao, Michael Muller, Mayank Agarwal, Stephanie Houde, Kartik Talamadupula, and Justin D. Weisz. 2022. Investigating Explainability of Generative AI for Code through Scenario-based Design. In *27th International Conference on Intelligent User Interfaces*. ACM, Helsinki Finland, 212–228. https://doi.org/10.1145/3490099.3511119

[95] Supriyo, Richard Tomsett, Ramya Raghavendra, Daniel Harborne, Moustafa Alzantot, Federico Cerutti, Mani Srivastava, Alun Preece, Simon Julier, and Raghuveer M Rao. 2017. Interpretability of Deep Learning Models : A Survey of Results. In *Chakraborty, Supriyo, et al. "Interpretability of deep learning models: A survey of results." 2017 IEEE smartworld, ubiquitous intelligence & computing, advanced & trusted computed, scalable computing & communications, cloud & big data computing, Internet.*

[96] Harini Suresh, Steven R Gomez, Kevin K Nam, and Arvind Satyanarayan. 2021. Beyond expertise and roles: A framework to characterize the stakeholders of interpretable machine learning and their needs. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–16.

[97] Aaquib Tabrez, Shivendra Agrawal, and Bradley Hayes. 2019. Explanation-Based Reward Coaching to Improve Human Performance via Reinforcement Learning. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, Daegu, Korea (South), 249–257. https://doi.org/10.1109/HRI.2019.8673104

[98] Nava Tintarev. 2007. Explanations of recommendations. In *Proceedings of the 2007 ACM conference on Recommender systems*. 203–206.

[99] Erico Tjoa and Cuntai Guan. 2020. A survey on explainable artificial intelligence (xai): Toward medical xai. *IEEE transactions on neural networks and learning systems* 32, 11 (2020), 4793–4813.

[100] Jennifer Villareale, Thomas Fox, and Jichen Zhu. 2024. Can Games Be AI Explanations? An Exploratory Study of Simulation Games. In *Proceedings of the Digital Games Research Association (DiGRA) International Conference.*

[101] Jennifer Villareale, Casper Harteveld, and Jichen Zhu. 2022. " I Want To See How Smart This AI Really Is": Player Mental Model Development of an Adversarial AI Player. *Proceedings of the ACM on Human-Computer Interaction* 6, CHI PLAY (2022), 1–26.

[102] Giulia Vilone and Luca Longo. 2020. Explainable artificial intelligence: a systematic review. *arXiv preprint arXiv:2006.00093* (2020).

[103] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y. Lim. 2019. Designing Theory-Driven User-Centric Explainable AI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–15. https://doi.org/10.1145/3290605.3300831

[104] Xinru Wang and Ming Yin. 2021. Are Explanations Helpful? A Comparative Study of the Effects of Explanations in AI-Assisted Decision-Making. In *26th International Conference on Intelligent User Interfaces*. ACM, College Station TX USA, 318–328. https://doi.org/10.1145/3397481.3450650

[105] Yunlong Wang, Priyadarshini Venkatesh, and Brian Y Lim. 2022. Interpretable Directed Diversity: Leveraging Model Explanations for Iterative Crowd Ideation. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–28. https://doi.org/10.1145/3491102.3517551

[106] Oyindamola Williams. 2021. Towards Human-Centred Explainable AI: A Systematic Literature Review. (2021). https://doi.org/10.13140/RG.2.2.27885.92645 Publisher: Unpublished.

[107] Christine T. Wolf. 2019. Explainability scenarios: towards scenario-based XAI design. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. ACM, Marina del Ray California, 252–257. https://doi.org/10.1145/3301275.3302317

[108] Jiachi Xie, Chelsea M Myers, and Jichen Zhu. 2019. Interactive visualizer to facilitate game designers in understanding machine learning. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–6.

[109] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. *Conference on Human Factors in Computing Systems - Proceedings* 4 (2020). https://doi.org/10.1145/3313831.3376301

[110] Wencan Zhang and Brian Y Lim. 2022. Towards Relatable Explainable AI with the Perceptual Process. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–24. https://doi.org/10.1145/3491102.3501826

[111] Jichen Zhu, Antonios Liapis, Sebastian Risi, Rafael Bidarra, and G Michael Youngblood. 2018. Explainable AI for designers: A human-centered perspective on mixed-initiative co-creation. In *2018 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 1–8.