Technological Forecasting & Social Change 186 (2023) 122120

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

journal homepage: www.elsevier.com/locate/techfore

Explainable Artificial Intelligence (XAI) from a user perspective: A synthesis of prior literature and problematizing avenues for future research

AKM Bahalul Haque [a,*], A.K.M. Najmul Islam [a], Patrick Mikalef [b,c]

[a] Department of Software Engineering, LUT University, Finland
[b] Department of Computer Science, Norwegian University of Science and Technology, Norway
[c] Department of Technology Management, SINTEF Digital, Norway

ARTICLE INFO

Keywords:
Explainable AI (XAI)
XAI effects
Trust
Transparency
Understandability
AI Adoption
AI Use

ABSTRACT

The rapid growth and use of artificial intelligence (AI)-based systems have raised concerns regarding explainability. Recent studies have discussed the emerging demand for explainable AI (XAI); however, a systematic review of explainable artificial intelligence from an end user's perspective can provide a comprehensive understanding of the current situation and help close the research gap. The purpose of this study was to perform a systematic literature review of explainable AI from the end user's perspective and to synthesize the findings. To be precise, the objectives were to 1) identify the dimensions of end users' explanation needs; 2) investigate the effect of explanation on end user's perceptions, and 3) identify the research gaps and propose future research agendas for XAI, particularly from end users' perspectives based on current knowledge. The final search query for the Systematic Literature Review (SLR) was conducted on July 2022. Initially, we extracted 1707 journal and conference articles from the Scopus and Web of Science databases. Inclusion and exclusion criteria were then applied, and 58 articles were selected for the SLR. The findings show four dimensions that shape the AI explanation, which are format (explanation representation format), completeness (explanation should contain all required information, including the supplementary information), accuracy (information regarding the accuracy of the explanation), and currency (explanation should contain recent information). Moreover, along with the automatic representation of the explanation, the users can request additional information if needed. We have also described five dimensions of XAI effects: trust, transparency, understandability, usability, and fairness. We investigated current knowledge from selected articles to problematize future research agendas as research questions along with possible research paths. Consequently, a comprehensive framework of XAI and its possible effects on user behavior has been developed.

1. Introduction

Recently, the adoption and use of artificial intelligence (AI)-based applications by various business organizations have been increasing to aid decision-making. For example, the International Data Corporation (IDC) has estimated that the worldwide AI expenditure is supposed to increase to 110 billion US dollars by the end of 2024 (Adadi and Berrada, 2018; IDC, 2018). Because AI has become more prevalent, it has become routine to rely on it to make decisions in our daily lives (Stahl et al., 2021; Mahmud et al., 2022a,b). We use various intelligent systems every day, such as in content and product recommendation (Benbasat and Wang, 2005; Gruetzemacher et al., 2021; Wang et al., 2014; Choi et al., 2012), news websites, social media (Feng et al., 2020), healthcare (Haque et al., 2020), and other public services (Hengstler et al., 2016; Haque et al., 2021; Du and Xie, 2021); however, the working principle of AI systems is unclear as the machine-learning models used in different AI systems do not reveal enough information about the process through which the conclusion is derived (Castelvecchi, 2016). Furthermore, the deep neural network (DNN) models used in advanced AI systems are extraordinarily complex to explain. Only specific people who design the algorithms understand how the system works (Angelov and Soares, 2020). The opacity of AI systems can reduce end users' trust and reliance on using AI-based systems while making critical decisions (Hasan et al., 2021; Baum et al., 2011). To address this problem, researchers and practitioners have called for the requirement to provide explainable Artificial Intelligence (XAI) that allows end users to perceive the

* Corresponding author.
E-mail addresses: bahaul.haque@lut.fi (A.B. Haque), najmul.islam@lut.fi (A.K.M.N. Islam), patrick.mikalef@ntnu.no, patrick.mikalef@sintef.no (P. Mikalef).

https://doi.org/10.1016/j.techfore.2022.122120
Received 28 January 2022; Received in revised form 27 August 2022; Accepted 16 October 2022
Available online 6 November 2022

underlying working principle of the decision-making procedure (Laato et al., 2022; Tiainen, 2021). Understanding the working principles of AI systems is crucial for end users to make effective decisions in different contexts (Scott et al., 1977). For example, in mission-critical use cases, such as healthcare, the decision-making procedure should be understandable for the users (doctors) to rely on the system (Lauritsen et al., 2020).

Furthermore, the General Data Protection Regulation (GDPR) has also emphasized the explainability of AI systems by introducing the "right to explanation" (Goodman and Flaxman, 2017). The regulation includes another policy related to "automated individual decision-making, including profiling," to prevent personal data from being used and processed by automated systems without permission (Malgieri, 2019). In addition, the High-level Expert Group on Artificial Intelligence of the European Commission has also outlined the importance of an explanation to achieve the transparency and reliability of AI in their "Ethics Guidelines for Trustworthy Artificial Intelligence (AI)".[1] Furthermore, governments worldwide are currently adopting automated decision making; for example, the Dutch immigration services are testing automated processes for asylum requests and resident permit applications (Janssen et al., 2020). Such sensitive decision making by government organizations should have explainability for the users as well for those involved in decision making. Therefore, organizations involved in the government should also have XAI as a prerequisite for an automated decision-making system. AI-based decision-making systems can be used for scalable and larger ecosystems; however, the systems need to be incorporated with some principles related to ethics and rights (Fjeld et al., 2020).

Therefore, due to the wide applicability and demand, researchers have investigated XAI across various domains and perspectives. Previously published Systematic Literature Reviews (SLRs) on XAI have focused on the ethical perspective of AI's black box nature (Meske et al., 2022; Wells and Bednarz, 2021), human-centric design patterns for ML-based systems (Chromik and Butz, 2021), personalized explanations of XAI (Schneider and Handali, 2019), behavioral interactions of human and autonomous agents (Anjomshoae et al., 2019), XAI in healthcare domain (Chakrobartty and El-Gayar, 2021; Antoniadi et al., 2021), and AI system communication, design recommendations and tradeoffs of the end user-centric AI (Laato et al., 2022), among others. Despite the plethora of these types of investigations, we have identified two major research gaps concerning end users' explanation needs. First, most prior SLRs (see Table 1) focused on a single domain (e.g., healthcare, transportation, etc.). This limits our understanding of how the end users' explanation needs might vary across different domains. For example, the healthcare professionals' explanation needs for making critical decisions would be significantly different than consumers' decisions regarding their next purchases. Second, most prior SLRs (see Table 1) have been conducted from a technical perspective. To understand the explanation needs of end users, an SLR that reviews studies from the human perspective is needed. However, very few SLR studies (Laato et al., 2022) have been done from the end users' perspectives. Therefore, an SLR conducted across different domains that includes the latest published articles can provide a comprehensive outline of how XAI has advanced in different application domains in recent times. Moreover, a comprehensive study of human-centered XAI can help researchers and practitioners understand how people perceive different types of explanations provided by AI-based systems. The analysis will also provide meticulous insight into the impact of XAI on humans. Hence, we conducted an SLR to critically analyze the previous research on AI users' explanation needs to fulfill the research objectives, which are (1) a synthesis of prior literature on XAI that contains (a) a critical analysis of extant literature to represent current knowledge on XAI in terms of explanation needs and XAI effects and (b) research domains and (2) the development of thematically organized future research avenues.

To address the research objectives, 58 publications were selected by scanning Scopus and Web of Science databases and using rigorous citation chaining techniques. Our SLR has three key findings. First, we found four dimensions of end users' explanation needs: format, completeness, accuracy, and currency. We then linked these dimensions with the five effects of XAI: trust, transparency, understandability, usability, and fairness, which have been discussed in prior literature (Laato et al., 2022) to develop a framework. Finally, we found 10 application domains where XAI research has been conducted. Based on these findings, our paper contributes to the existing XAI literature (Binns et al., 2018; Chazette and Schneider, 2020; Schneider et al., 2021; van der Waa et al., 2020; Laato et al., 2022) by 1) identifying the dimensions of end users' explanation needs and presenting them from information systems research perspective; 2) identifying the outcomes of XAI from the end users' perspectives; 3) identifying research gaps and problematizing future research directions in XAI, particularly from end users' perspectives; and 4) building a framework for XAI research from end users' perspective. Our findings also help practitioners design a more user-friendly and trustworthy XAI system by determining the explanation needs of the end users.

The remainder of the paper is structured as follows. Section 2 outlines the background of XAI and Related Works. Section 3 describes the SLR methodology and literature selection. Section 4 contains the research trend of XAI based on the selected articles, and Section 5 synthesizes previous studies on XAI. This section comprehensively represents the current knowledge of XAI aligned with information systems research. Section 6 critically analyzes the current knowledge to identify future research directions. Section 7 outlines the comprehensive framework of XAI research from the end user's perspective. Section 8 briefly describes the implications of this work, and Section 9 concludes the paper.

2. Background

2.1. Explainable AI and related concepts

Explainable AI, interpretable AI, transparent AI, understandable AI, and responsible AI terminologies are used interchangeably in the literature (Arrieta et al., 2020). XAI has emerged intending to present explanations purveyed to human understanding, trust, and transparency (Gerlings et al., 2021a, 2021b). The relational link connecting the input and the output of an artificial neural network is not observable. Therefore it is necessary to put effort into the explainability and interpretability of the black-box nature of various AI models (Dağlarli, 2020). DARPA, one of the leading research organizations on XAI, explained XAI as an extension of an AI system whose models and decisions can be easily understandable and properly believable by end users (Gunning and Aha, 2019). The understandability and believability of machine-learning models contribute to the interpretability of a machine-learning model for the target audience (Lipton, 2018). Explainability usually indicates how strongly a particular phenomenon can be described so that the audience can effortlessly understand it. Therefore, in XAI, explainability means the AI should be capable of explaining predictions obtained from a model from a more profound methodological point of view to users (Antunes et al., 2012); however, explainable AI can also be defined as: "given an audience, an explainable Artificial Intelligence produces details or reasons to make its functioning clear or easy to understand" (Arrieta et al., 2020).

Interpretability (Lipton, 2018) specifies that the working procedure of the machine models should be made unambiguous and crystal clear to both technical and nontechnical users. Though interpretability and explainability are used interchangeably, there are some basic conceptual differences between them. Explainability means explaining the

[1] on Artificial Intelligence (AI HLEG), H.-L. I. G. (2019). *Ethics Guidelines for Trustworthy AI*. European Commission. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

Table 1
Comparison of previous related systematic review studies.

Study	Purpose	Years included	Source of primary studies
Anjomshoae et al., 2019	Presents a goal-driven literature review of explainable robots and agents to enhance the understanding of the "black box."	All documents are published between the years 2008–2018.	An initial collection of 303 papers were reduced to 62 final selections using seven inclusion criteria. The authors did not mention the types of individual publication. These papers were collected from digital libraries, such as IEEExplore, Science Direct, ACM, and Google Scholar.
Schneider and Handali, 2019	This study provides a structured collection of information that conceptualizes "personalized explanation" and relates the idea to other domains that are intertwined with XAI.	The paper did not mention the publication time of these documents.	They collected research articles and conference papers from the IEEE Xplore, AIS, ACM, and Arxiv databases. Their study did not mention the total number of papers considered.
Antoniadi et al., 2021	Highlighting the indispensability of interpretable AI systems in medical use cases, this study underscores the ethical and fair decision making by AI systems in medical practices. This study claims to provide suggestions to aid future opportunities and tackle foreseeable challenges.	Unknown – 2020 (authors did not specify their starting year as a search criterion).	Using the Google Scholar database, they identified 668 articles based on six combinations of search phrases. Through an intricate elimination and selection process, 33 papers were finally selected. The authors did not specify the publication type of these papers.
Chakrobartty and El-Gayar, 2021	Raising concerns about the un-explainability of AI techniques, especially in the medical sector, this study highlights the methods and practices that emphasize XAI in the medical sector.	This study covers documents published between 2008 and 2020.	Based on eight search keywords, they initially found 66 documents, which were reduced to 22 using several inclusion and exclusion criteria. The authors did not specify the type of publications.
Chromik et al., 2021	To better comprehend the black box, this study presents an argument that advocates that the explanation user interfaces interpretability increases by employing explanation-generating models. This study provides insight into how designers can attune the explanation of AI systems in user interfaces.	Unknown – 2020 (authors did not specify their starting year as a search criteria).	An initial collection of 146 documents was reduced to 91 documents that meticulously matched the research objective.
Gerlings et al., 2021b	This study presents a thoroughgoing discussion on how XAI addresses the black box problem in AI-based applications. By conducting a comprehensive study of recent publications, they attempted to find how XAI contributes to reducing the gap between stakeholders and the black box.	Covers documents from 2016 to 2020.	They collected data from ArXiv, AIS, JSTOR, ACM Digital Library, IEEE Xplore, SAGE, and Science Direct digital libraries. From 221 initial documents, they finally picked 64 documents for their study.
Linardatos et al., 2021	This study highlights the programming implementation in recent studies that contributes to increasing the interpretability of ML models from both theorist and practitioner perspectives.	Not specified.	Not specified.
Wells & Bednarz, 2021	This study accentuates the societal and ethical implications of XAI in the area of reinforcement learning. The study showed limitations, such as lack of user studies, the prevalence of toy-examples, and difficulties providing understandable explanations, in the case of reinforcement learning.	Covers published documents between 2014 and 2020.	Conducting a Boolean search on digital libraries, such as ACM, IEEExplorer, Science Direct, and Springer Link digital libraries, they gathered 520 papers, among which they justify choosing only 25 papers that matched their research interest. • Defense/Military – 1 • Autonomous Vehicles – 2 • Networking 2 • Robotics – 4 • Gridworld – 5 • Games - 16
Laato et al., 2022	The authors identified the high-level objectives of AI communications with end users such as understandability, trustworthiness, transparency, controllability, and fairness. Moreover, the authors provide design recommendations for explanations of AI systems.	Search conducted on October 2020	The search was conducted on both Scopus and Web of Science from XAI from the HCI perspective. 808 unique articles were extracted after removing the duplicates. The final sample size was 25 articles that matched their research objective.
Our study	Our study focuses on (1) a synthesis of prior literature on XAI that contains critical analysis of extant literature to represent current knowledge on XAI in terms of explanation representation, XAI effects, explored research domains and (2) the development of thematically organized future research avenues.	Covers published documents up to July 15, 2022.	Conducted a search on Scopus and Web of Science, which are the two most comprehensive databases for scholarly publications. Based on a wide range of keywords, the search initially revealed 2896 studies both from the Scopus and Web of Science. From them, 58 studies were included as they matched the research objectives and interests. Only end user-centric empirical studies are included, which provides a unique identifier for our study. In contrast to other SLRs (e.g., Laato et al., 2022), we have discussed the explanation representation across different domains, have identified explanation quality dimensions through Wixom and Todd (2005), and have analyzed the effect of XAI across different domains and categorically presented them. Finally, we have provided a synthesized framework by connecting explanation dimensions and XAI effects.

decisions made by machine models in a human-understandable form, and interpretability is the explanation of how or why a model resulted in a particular prediction (Doshi-Velez and Kim, 2017). Transparency is another critical mainstay of XAI, which means being effortlessly viewed through something. In XAI, a model can be considered transparent if it can explain its different steps simplistically to human users (Wachter et al., 2017). Understandability specifies whether the features and attributes of a model are easily recognizable by users without knowing its inner composition. Understandable AI specifies whether the unification of model developers and UI designers can produce a human-centric AI architecture (Arrieta et al., 2020). Finally, responsible AI is a framework from the governance perspective that is comprised of guidelines and policies for AI technologies to ensure integrity, efficiency, and productivity. These policies and guidelines also involve responsible system design, proper monitoring, and awareness (Ghallab, 2019).

2.2. Related works

Practical relevance and research interest in XAI have significantly increased in recent times. We have been able to identify several prior SLRs which focus on various domains. Table 1 represents a comparative analysis of these identified studies. The black-box nature of AI poses ethical concerns and risks since no one can interpret what is going on inside and how the data is being processed (Meske et al., 2022). Therefore, the open development of AI should be closely observed and audited, as the compromises involved may lead to dire consequences (Meske et al., 2022). The explanatory design of the user interface can also contribute to understanding black box AI. Interaction factors, such as transmission, dialogue, control, experience, optimal behavior, tool use, and embodied action, are critical when designing such a system. Four human-centric design patterns for ML-based systems increase the understanding level of a human user through a set of explanation-generating methods (Chromik and Butz, 2021). Naturalness, responsiveness, flexibility, and sensitivity are the four recurring design patterns that are the most frequently used human-centric design patterns (Chromik and Butz, 2021). Personalized explanation enhances the interpretability and understanding of the explainees (Schneider and Handali, 2019); however, there is a substantial research gap regarding collecting personalized and explicit information from the explainees with arguable privacy concerns (Schneider and Handali, 2019).

Research on explainable AI has increased and has primarily focused on policy summarization, human collaboration, visualization, verification, etc. (Wells and Bednarz, 2021); however, research gaps exist in customized algorithms, user testing, and scalability (Wells and Bednarz, 2021). In one systematic review, the explainable nature of the behavioral interaction of agents and robots with human users is discussed (Anjomshoae et al., 2019). The work also summarizes the importance of the explainable nature of intelligent systems for non-expert users. Both technical and non-technical perspectives are important for the XAI domain. One of the seminal scholarly works (discussing the importance of unveiling the black box) comprehensively outlines the need, research challenges, and future research opportunities to provide explainability (Gerlings et al., 2021a, 2021b). Healthcare is a crucial domain for any type of technology use. Similarly, XAI research for the healthcare domain can help doctors make decisions (Chakrobartty and El-Gayar, 2021). XAI in healthcare includes various techniques and methods used for XAI (Chakrobartty and El-Gayar, 2021) and clinical decision making (Antoniadi et al., 2021). Other traditional review studies discuss explainable AI from a technical perspective, such as the interpretability methods of various machine-learning interpretability models (Linardatos et al., 2021). Recently, Laato et al. (2022) identified the high-level objectives of AI communications with end users such as understandability, trustworthiness, transparency, controllability, and fairness. Moreover, they provided design recommendations for explanations of AI systems along with future research directions.

3. Methodology

We have adopted an SLR methodology to summarize the existing studies on XAI. An SLR is a method for locating, assessing, and evaluating relevant research for certain research questions, topics, or phenomena being studied (Kitchenham and Charters, 2007). To analyze prior contributions to AI, the techniques focus on "identifying, evaluating and interpreting all available research relevant to a particular research question, or topic area, or phenomenon of interest" (Kitchenham and Charters, 2007). We identified the search terms and used Boolean operators to generate search strings for searching the Scopus and Web of Science databases. Scopus and Web of Science are among the most comprehensive and recognized databases with reputed scholarly publications. We did not specify any starting year for the search criteria, and therefore the timeline used for the search results would be the day we conducted our last search query, which is July 15th, 2022. We have identified inclusion and inclusion criteria to filter irrelevant studies and to develop the final article list. The search terms are shown in Table 1.

3.1. Literature selection criteria

For the literature selection, we defined a set of well-defined inclusion and exclusion criteria based on the scope of this review work. The inclusion and exclusion criteria are outlined in Table 2.

3.1.1. Search result extraction and analysis

The search terms and the results extracted are provided in Table 3.

From both databases, only conference and journal articles were selected, and the duplicates were removed, leaving 1707 articles. After reading the titles and abstracts, 1190 articles were removed from the list. Full texts of the remaining 517 articles were studied carefully to remove the articles that were not within the scope of our research theme. Furthermore, articles without empirical studies were excluded, which resulted in the final 58 articles. Fig. 1 depicts the screening and selection process.

4. Research trend

Of the 58 studies in our SLR, 13 were journal articles, and 45 were conference articles. Table 4 depicts the publications per year for the selected studies and the number of journal and conference articles. Here, we observed that the number of publications increased from 2018 onwards. This clarifies that explainable AI has been a topic of interest in recent years. Other bibliometric data of the selected articles, such as the number of publications by publishers (Table 5) and the top-five cited articles (according to Scopus), including their author affiliation (Table 6), are presented as well.

5. Synthesis of prior literature

This section provides a critical analysis of the selected research studies and an overview of their findings. This section is divided into (1) Current Knowledge Representation and (2) Research Domains. Table 7 represents the synthesis of prior literature.

5.1. Current knowledge representations

This section represents the current knowledge extracted from the selected articles. The section is divided into three subsections: (1) XAI representation, (2) Effects of Explainable AI, and (3) Explanation Presentation Time.

5.1.1. XAI representation

We have adopted the information quality dimensions proposed by Wixom and Todd (2005) to conceptualize XAI representation dimensions. Wixom and Todd (2005) proposed four information quality

Table 2
Inclusion and exclusion criteria.

Inclusion criteria	Exclusion criteria
1. Conference and journal papers are included 2. Only publications in English language 3. Full-text availability in online databases and repositories 4. Articles with empirical studies are included (end user centric)	1. Review articles, book chapters, magazines, and editorials were excluded 2. Articles published other than in the English language were excluded 3. Full-text was not available on online repositories 4. Duplicate results were removed

dimensions. The dimensions are "format," which defines the user's perception regarding the information presented; "completeness," which represents how adequately all the necessary information is presented; "accuracy," which represents the user's perception of how accurate the information is; and "currency," which represents the user's perception of how up-to-date the information is at the time of access. To understand the quality of the explanation presented by the XAI system, we have aligned our findings with these four dimensions (Wixom and Todd, 2005). These dimensions are described based on the data extracted from the literature.

5.1.1.1. Format. The information representation of XAI systems is primarily either textual, visual, or auditory or in a hybrid mode. Users of different domains have different perceptions of the information representation format. The users of healthcare-related AI systems need explanations in both textual and visual(graphical) formats (Branley-Bell et al., 2020; Cheng et al., 2019; Daudt et al., 2021; Lee and Rich, 2021; Wang et al., 2019; Xie et al., 2019; Rodriguez-Sampaio et al., 2022). For example, if an explanation is presented with appropriate figures, images, and terminologies, the user's understandability increases (Bussone et al., 2015; Cai et al., 2019; Eiband et al., 2019; Hudon et al., 2021). The hybrid explanation format reveals important information about the reasoning involved in a decision-making process (Branley-Bell et al., 2020; Górski and Ramakrishna, 2021). Expert end users of AI-based diagnostic pathology prefer a user-centric design that combines textual and visual explanations (Evans et al., 2022). Users of music recommendations, movie recommendations, drawing tools, and other different types of sports-related systems need explanations in a hybrid format to increase understandability (Cramer et al., 2008; Ehsan et al., 2019; Kouki et al., 2019; Ngo et al., 2020; Oh et al., 2018; Schmidt et al., 2020; Szymanski et al., 2021). The hybrid explanation can include a partial dependence plot as well as documentation. This two-dimensional plot describes how one output is influenced by another input (Szymanski et al., 2021).

Users prefer hybrid explanations that include textual and visual explanations and, in some cases, explanations using indicator lights (Schneider et al., 2021; van der Waa et al., 2020). Explanations in e-commerce (Ehsan et al., 2021; Eslami et al., 2018), education (Cheng et al., 2019; Conati et al., 2021; Mucha et al., 2021; Putnam and Conati, 2019), finance (Binns et al., 2018; Chromik et al., 2021; Cirqueira et al., 2020), law (Liu et al., 2021a, 2021b; Górski and Ramakrishna, 2021),

Table 3
Search terms and results.

Search terms	Databases	Results
"Explainable Artificial Intelligence" OR "Transparent Artificial Intelligence" OR "Interpretable Artificial Intelligence" OR "Understandable Artificial Intelligence" OR "Artificial Intelligence Transparency" OR "Artificial Intelligence Interpretability" OR "Artificial Intelligence Understandability" OR "XAI" OR "Understandable AI" OR "AI Transparency" OR "AI Interpretability" OR "Responsible AI" OR "AI Decision Making" OR "AI trust" OR "AI system use" OR "AI use"	Scopus Web of Science	2669 1345

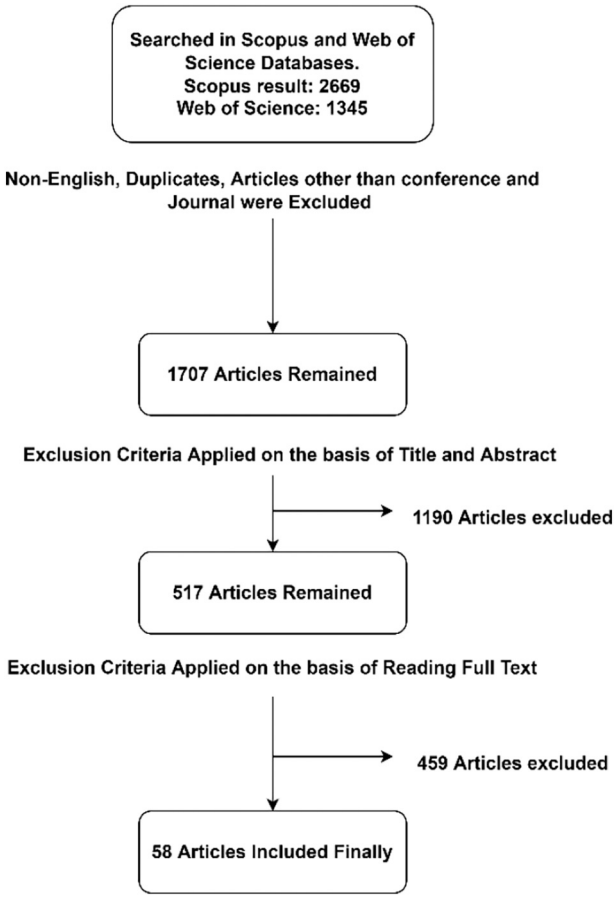

Fig. 1. Article screening and selection process.

Table 4
Number of conference and journal publications by year.

Publication year	Conference publications	Journal publications
2008	0	1
2009	2	0
2010	1	0
2015	1	0
2018	4	0
2019	11	0
2020	10	4
2021	14	4
2022	2	4

Table 5
Number of articles by publisher.

Publishers	No. of publication
Taylor and Francis	1
Springer	12
SAGE	1
IEEE	2
Emerald	1
Elsevier	4
ACM	37

and social networking (Lim and Dey, 2009; Yin et al., 2019) can also be provided in the hybrid mode. For example, the logic behind algorithmic decision-making, working procedures, and product image in e-commerce systems are all discussed in the textual explanation. For law-related systems, counterfactual explanations and logical reasoning behind the decision in laymen's terms are needed (Górski and

Table 6
Top-five cited articles, including their authors and affiliations according to Google scholar (till the date of final submission of this article).

Title	Year	Source	Cited by	Authors with affiliations	Publisher	Type
Designing theory-driven user-centric explainable AI	2019	Conference on Human Factors in Computing Systems - Proceedings	447	Wang, D., School of Computing, National University of Singapore, Singapore, Singapore; Yang, Q., Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA, United States; Abdul, A., School of Computing, National University of Singapore, Singapore, Singapore; Lim, B.Y., School of Computing, National University of Singapore, Singapore, Singapore	ACM	Conference
It's reducing a human being to a percentage"; perceptions of justice in algorithmic decisions	2018	Conference on Human Factors in Computing Systems - Proceedings	353	Binns, R., Dept. of Computer Science, University of Oxford, United Kingdom; Van Kleek, M., Dept. of Computer Science, University of Oxford, United Kingdom; Veale, M., Dept. of Science, Technology, Engineering and Public Policy, University College London, United Kingdom; Lyngs, U., Dept. of Computer Science, University of Oxford, United Kingdom; Zhao, J., Dept. of Computer Science, University of Oxford, United Kingdom; Shadbolt, N., Dept. of Computer Science, University of Oxford, United Kingdom	ACM	Conference
Why and why not explanations improve the intelligibility of context-aware intelligent systems	2009	Conference on Human Factors in Computing Systems - Proceedings	568	Lim, B.Y., Carnegie Mellon University, United States; Dey, A. K., Carnegie Mellon University, 5 United States; Avrahami, D., Intel Research Seattle, United States	ACM	Conference
Assessing demand for intelligibility in context-aware applications	2009	ACM International Conference Proceeding Series	243	Lim, B.Y., Carnegie Mellon University, Pittsburgh, United States; Dey, A.K., Carnegie Mellon University, United States	ACM	Conference
The effects of transparency on trust in and acceptance of a content-based art recommender	2008	User Modeling and User-Adapted Interaction	435	Cramer, H., Human Computer Studies Lab., University of Amsterdam, Amsterdam, Netherlands; Evers, V., Human Computer Studies Lab., University of Amsterdam, Amsterdam, Netherlands; Ramlal, S., Human Computer Studies Lab., University of Amsterdam, Amsterdam, Netherlands; Van Someren, M., Human Computer Studies Lab., University of Amsterdam, Amsterdam, Netherlands; Rutledge, L., Telematica Institute, Enschede, Netherlands, CWI, Amsterdam, Netherlands; Stash, N., Eindhoven University of Technology, Eindhoven, Netherlands, VU University Amsterdam, De Boelelaan Amsterdam, Netherlands; Aroyo, L., Eindhoven University of Technology, Eindhoven, Netherlands, VU University Amsterdam, Amsterdam, Netherlands; Wielinga, B., Human Computer Studies Lab., University of Amsterdam, Amsterdam, Netherlands	Springer	Journal

Ramakrishna, 2021). The explanation format in virtual assistant systems includes voice-based interactions along with textual and visual explanations (Weitz et al., 2019, 2021; Gao et al., 2022). An interactive agent with hybrid (textual and audio-visual) explanations can increase the perception of trust in a system (Weitz et al., 2021). XAI in immigration systems needs both the textual and visual information format because the decision-making requires careful observation of personal details, travel itineraries, and photo matches with travelers (Janssen et al., 2020). In the human resource context, both textual and visual explanations are recommended (Bankins et al., 2022). For criminal justice use cases, the reasoning should include information related to both "why" and "why not" because the counterfactual details help clear any doubt or bias (Dodge et al., 2019). The hybrid explanation format is also required for other context-aware systems (Lim et al., 2009), general decision-making systems (Brennen, 2020; Schrills and Franke, 2020), travel guides (Lim and Dey, 2009), cooking recommendation systems (Broekens et al., 2010), and wearable systems (Danry et al., 2020).

5.1.1.2. Completeness. Completeness in XAI refers to providing the target user with all required information, including on demand supplementary data. For the healthcare domain, the user needs to be presented with patients' demographic information, cardinal symptoms, previous test data, and initial evaluations (Wang et al., 2019; Xie et al., 2019). The visual explanation can include a vivid and concise representation of appropriate diagnosis images, indicators of different properties, bar charts, etc. (Bussone et al., 2015; Cai et al., 2019; Ehsan et al., 2019; Eiband et al., 2018). The textual explanation can include a detailed representation of the decision-making procedure and the algorithms' working principles (Branley-Bell et al., 2020; Bussone et al., 2015; Cai et al., 2019; Daudt et al., 2021; Eiband et al., 2019; Lee and Rich, 2021). In addition, providing users with contextual information and references about the prediction upon request increases users' trust and perception of reliability (Branley-Bell et al., 2020; Daudt et al., 2021; Lee and Rich, 2021; Bove et al., 2021). Contextual information refers to an explanation that is domain-specific or application-specific. Along with explaining the algorithms and machine-learning models, it is also important to include domain-specific contextual information regarding decision making. The contextual information varies across different domains. Therefore, it should be considered by the developers during the design phase of a system. Media and entertainment recommendation systems can explain decision-making by revealing the working procedure of the algorithm, the personal data being used, and visual representation of the recommendation being made (Ehsan et al., 2019; Kouki et al., 2019; Schmidt et al., 2020). For example, the users of music and movie recommendation systems want to see what kind of data has been used for the prediction and the popularity rating of the decision (Kouki et al., 2019; Ngo et al., 2020). In addition, information regarding the movie name, previous ratings, genres, and confidence measurements can be provided as an explanation.

Another example is an online news recommendation system, where the visual explanation includes a two-dimensional partial dependence plot that describes how the output is influenced by the input properties (Szymanski et al., 2021). A textual explanation of XAI can also include product type, price, order details, and other different attributes and features (Ehsan et al., 2021; Eslami et al., 2018; Bankins et al., 2022). The reasons for user agreement and disagreement related to predictions

Table 7
Synthesis of prior literature.

Source	XAI representation	Effects	Explanation presentation time	Research focus
Cramer et al., 2008	Hybrid representation	Accuracy, Trust Transparency	With the recommendation and after the user demands explanation	Media and Entertainment
Branley-Bell et al., 2020	Hybrid representation	Trust Understandability	With the recommendation and after the user demands explanation	Healthcare
Cheng et al., 2019	Hybrid representation	Understandability	After the user demands explanation as supplementary information	Education
Daudt et al., 2021	Hybrid representation	Trust Understandability	Not mentioned explicitely, but analysis shows both with recommendation and after the user demands explation	Healthcare
Lee and Rich, 2021	Hybrid representation	Trust	With the recommendation and after the user demands explanation	Healthcare
Wang et al., 2019	Hybrid representation	Trust Understandability	Not mentioned explicitly	Healthcare
Xie et al., 2019	Hybrid representation	Trust	Not mentioned explicitly	Healthcare
Rodriguez-Sampaio et al., 2022	Hybrid representation	Trust Understandability	With the recommendation and after the user demands explanation	Healthcare
Bussone et al., 2015	Graphical representation	Trust	With the recommendation and after the user demands explanation	Healthcare
Cai et al., 2019	Graphical representation	Transparency	With the recommendation and after the user demands explanation	Healthcare
Eiband et al., 2019	Graphical representation	Transparency	With the recommendation and after the user demands explanation	Recommendation System
Hudon et al., 2021	Hybrid representation	Trust Understandability	Not explicitly mentioned	Media and Entertainment
Górski and Ramakrishna, 2021	Hybrid representation	Understandability Fairness	With the recommendation and after the user demands explanation	Law
Evans et al., 2022	Hybrid representation	Understandability	With the recommendation and after the user demands explanation	Healthcare
Ehsan et al., 2019	Hybrid representation	Understandability	With the recommendation	Media and entertainment
Kouki et al., 2019	Hybrid representation	Trust	With the recommendation	Media and entertainment
Ngo et al., 2020	Hybrid representation	Transparency	With the recommendation	Media and entertainment
Oh et al., 2018	Hybrid representation	Trust Usability	With the recommendation and after the user demands explanation	Media and entertainment
Schmidt et al., 2020	Hybrid representation	Trust	With the recommendation and after the user demands explanation	Media and entertainment
Szymanski et al., 2021	Hybrid representation	Understandability Transparency	With the recommendation and after the user demands explanation	Media and entertainment
Ehsan et al., 2021	Hybrid representation	Trust Understandability	With the recommendation and after the user demands explanation	E-commerce
Eslami et al., 2018	Hybrid representation	Understandability	With the recommendation and after the user demands explanation	E-commerce
Conati et al., 2021	Hybrid representation	Transparency	With the recommendation and after the user demands explanation	Education
Mucha et al., 2021	Hybrid representation	Fairness	With the recommendation and after the user demands explanation	Education
Putnam and Conati, 2019	Hybrid representation	Trust	With the recommendation and after the user demands explanation	Education
Li et al., 2021	Hybrid representation	Transparency	With the recommendation and after the user demands explanation	Human Resource Management
Khosravi et al., 2022	Hybrid representation	Trust	With the recommendation and after the user demands explanation	Education
Binns et al., 2018	Hybrid representation	Understandability Fairness	With the recommendation	Transportation, Finance
Chromik et al., 2021	Hybrid representation	Trust Understandability Usability	With the recommendation and after the user demands explanation	Finance
Cirqueira et al., 2020	Hybrid representation	Trust	With the recommendation and after the user demands explanation	Finance
Liu et al., 2021a	Hybrid representation	Transparency Understandability Fairness	With the recommendation and after the user demands explanation	Legal
Liu et al., 2021b	Hybrid representation	Transparency	With the recommendation and after the user demands explanation	Social Networking
Górski and Ramakrishna, 2021	Hybrid representation	Transparency Understandability Fairness	With the recommendation and after the user demands explanation	Legal
Lim and Dey, 2009	Hybrid representation	Understandability	With the recommendation and after the user demands explanation	Social Networking
Yin et al., 2019	Hybrid representation	Trust	With the recommendation and after the user demands explanation	Social Networking
Weitz et al., 2019		Trust	With the recommendation and after the user demands explanation	Digital Assistant

(*continued on next page*)

Table 7 (*continued*)

Source	XAI representation	Effects	Explanation presentation time	Research focus
	Hybrid representation			
Weitz et al., 2021	Hybrid representation	Trust	With the recommendation and after the user demands explanation	Digital Assistant
Janssen et al., 2020	Hybrid representation	Transparency Fairness	With the recommendation and after the user demands explanation	E-Governance
Bankins et al., 2022	Hybrid representation	Trust	With the recommendation	Human Resource Management
Dodge et al., 2019	Hybrid representation	Trust Fairness	With the recommendation and after the user demands explanation	E-Governance
Lim et al., 2009	Hybrid representation	Understandability	With the recommendation and after the user demands explanation	Recommendation System
Brennen, 2020	Hybrid representation	Trust Usability	With the recommendation and after the user demands explanation	Recommendation System
Schrills and Franke, 2020	Hybrid representation	Trust	With the recommendation and after the user demands explanation	Recommendation System
Broekens et al., 2010	Hybrid representation	Understandability	With the recommendation and after the user demands explanation	Media and Entertainment
Danry et al., 2020	Hybrid representation	Understandability	With the recommendation and after the user demands explanation	Healthcare
Eiband et al., 2018	Graphical representation	Transparency	With the recommendation and after the user demands explanation	Healthcare
Bove et al., 2021	Graphical representation	Trust Understandability	With the recommendation and after the user demands explanation	E-commerce
Chazette and Schneider, 2020	Hybrid representation	Understandability	With the recommendation	Transportation
Schneider et al., 2021	Hybrid representation	Understandability	With the recommendation	Transportation
van der Waa et al., 2020	Hybrid Representation	Understandability Transparency	With the recommendation	Transportation
Park et al., 2021	Hybrid representation	Trust	With the recommendation and after the user demands explanation	Human Resource
Hong et al., 2020	Hybrid representation	Trust	With the recommendation and after the user demands explanation	Social networking
Liao et al., 2020	Hybrid representation	Trust Transparency	With the recommendation and after the user demands explanation	Social networking
Wang and Moulden, 2021	Hybrid representation	Transparency	With the recommendation and after the user demands explanation	Social networking
Dhanorkar et al., 2021	Hybrid representation	Trust	With the recommendation and after the user demands explanation	AI Development
Evans et al., 2022	Hybrid representation	Trust	With the recommendation and after the user demands explanation	Healthcare
Andres et al., 2020	Hybrid representation	Trust	On demand explanation	AI Development
Hind et al., 2020	Hybrid representation	Understandability	With the recommendation and after the user demands explanation	Social networking

in intelligent tutoring systems must be explained to the users as well (Conati et al., 2021; Putnam and Conati, 2019). Therefore, users should be able to request information if the explanations provided do not meet their expectations. Furthermore, other systems, such as grade estimations and university admission decision making, are required to provide students with personal details, academic details, and other required attributes that contribute to decision making (Cheng et al., 2019; Mucha et al., 2021).

Loan application systems, fraud detection, and other banking software are sophisticated decision-making systems. Therefore, explanations for these types of systems should be more detailed and comprise personal details, previous credit history, employment history, and the algorithms' working procedures (Binns et al., 2018; Chromik et al., 2021; Cirqueira et al., 2020). Similarly, for transportation systems, decision making can be explainable using contextual information, confidence measurements, light indicators, and previous decisions in similar situations (Chazette and Schneider, 2020; Schneider et al., 2021; Bove et al., 2021). Flight re-routing systems can provide the reasoning behind choosing specific routes and other supplementary information on demand (Binns et al., 2018). Virtual assistant systems should provide an explanation using the appearance of a virtual agent, such as facial expressions, voice, and gestures. A harmonic combination of explainable AI methods as well as appropriate linguistic representations can make the system trustworthy (Weitz et al., 2019, 2021; Gao et al., 2022). XAI in a human resource management system should explain the working procedure and should display personal information and other attributes both in a textual and visual format (Park et al., 2021). A similar situation is observed in the case of immigration services and criminal justice use cases (Dodge et al., 2019; Janssen et al., 2020). To establish the completeness of the XAI system, the system developers and designers should keep a critical eye on the explanation types and user requirements. The users want "why," "why not," "how," "what if," and "what else" explanations from the systems along with an interactive user interface (Broekens et al., 2010; Conati et al., 2021; Schrills and Franke, 2020). Moreover, developers may consider using different color-coding indicators that can also enhance trust among users (Brennen, 2020). Therefore, they should design and develop an interactive system carefully considering all the requirements of users, device diversity, and regulatory issues to promote the completeness of XAI (Brennen, 2020; Danry et al., 2020; Hong et al., 2020).

5.1.1.3. Accuracy. Users' accuracy perceptions regarding information from XAI systems vary and depend on different factors. Explanations containing personalized prioritization matrices, counterfactual information about specific predictions (Liu et al., 2021a, 2021b), and supplementary information instigate the perception of accuracy and understandability among users (van der Waa et al., 2020; Wang et al.,

2019; Xie et al., 2019). Moreover, this information can help verify decision-making and motivate the user to adopt an AI-based system (Wang et al., 2019). The academic tutoring system shows the confidence value of a decision as an explanation, which helps users accept or ignore a decision (Putnam and Conati, 2019). In addition, for education-related AI tools, explanation accuracy can increase if comparisons are shown between previous and current recommendations, trust scores of different recommendations using different models, etc. (Li et al., 2021; Khosravi et al., 2022). Some explainable AI systems can increase user interaction by providing detailed user instructions (Oh et al., 2018), information of the mental model used (Cramer et al., 2008), collaborative filtering (Ngo et al., 2020), and contextual data (Eiband et al., 2018; Liao et al., 2020; Bove et al., 2021). User involvement in the design process reduces the knowledge gap and promotes accuracy perceptions (Eslami et al., 2018; Ngo et al., 2020; Oh et al., 2018). Users' accuracy perceptions of XAI information are based on an explanation that contains information related to the certainty level of prediction (Bussone et al., 2015; Eiband et al., 2019), algorithmic decision-making procedures (Eiband et al., 2019; Park et al., 2021), claims and evidence (Danry et al., 2020), and information regarding domain expert engagement in the development process (Mucha et al., 2021; Wang and Moulden, 2021).

XAI should produce a human-like explanation and should show the accuracy level of the system to make the system more interpretable and accurate (Janssen et al., 2020; Lim and Dey, 2009; Park et al., 2021). Users' perceptions of the accuracy of the data of the XAI system can be established if the explanation of the algorithmic working procedure is presented sequentially to the users. This sequential flow of actions and information will motivate the user to accept or deny the decision (Broekens et al., 2010; Conati et al., 2021). AI-based law-related decision-making systems can positively affect users' accuracy perceptions by including evidence-based-reasoning sentences, legal rule sentences, and citation sentences (Górski and Ramakrishna, 2021). Explanations including this information act as a reference to the accuracy of the decision made by the system (Górski and Ramakrishna, 2021).

5.1.1.4. Currency. Currency is defined as the user's perception of up-to-date information (Wixom and Todd, 2005); however, for XAI, currency unfolds differently. XAI explains the algorithmic working principle, counterfactual data, supplementary information, and contributing features (Binns et al., 2018; Chromik et al., 2021; Eiband et al., 2019). From the XAI perspective, though the users are presented with an automatic explanation, an on-demand explanation is also available. The on-demand explanation can include the most recent information about any decision (Bussone et al., 2015; Putnam and Conati, 2019; Schrills and Franke, 2020; Wang et al., 2019). Supplementary information regarding the contextual data and the latest and historical references can also be available in XAI systems (Binns et al., 2018; Branley-Bell et al., 2020; Cirqueira et al., 2020; Bove et al., 2021). When designers and developers include the users in the XAI development process, they can acquire up-to-date user requirements (Ngo et al., 2020; Oh et al., 2018).

For fraud detection, loan approval contexts, and other mission-critical systems, it is essential to present the latest information (Binns et al., 2018; Chromik et al., 2021; Cirqueira et al., 2020). Recruitment systems also should use the candidates up to date information for recruitment-related decision-making (Bankins et al., 2022; Li et al., 2021). Financial decision-making such as loan or credit approval, should use the up-to-date financial history of the person (Chromik et al., 2021; Cirqueira et al., 2020). Another use case related to the flight re-routing system offers the latest flight data to users so that travel is flexible and comfortable (Binns et al., 2018). The same goes for media and entertainment recommendation systems, where the users recommend the latest movies and music as part of the process (Kouki et al., 2019). Similarly, instant messaging applications and tour guide systems present the latest explanation data to the user (Lim and Dey, 2009; Yin et al., 2019).

5.1.2. Effects of XAI

Our goal in this paper is to link the XAI representation dimensions with XAI effects. Towards this goal, we adopted the XAI objectives described by Laato et al. (2022) and categorize explainable AI effects into trust, transparency, usability, understandability, and fairness. The effects are briefly explained in the following subsections based on our literature review.

5.1.2.1. Trust. Based on our literature review, we have observed that users' trust is affected by both the stated and the observed accuracy of the machine-learning model. For example, users' trust in a machine-learning model increases or decreases based on the information about stated and observed accuracy (Yin et al., 2019). Prior research studies have shown that providing users with contextual information, historical data, and the proper reference behind decision making enhances trust in the system, particularly in the context of healthcare and finances (Cirqueira et al., 2020; Kouki et al., 2019; Wang et al., 2019; Xie et al., 2019; Dhanorkar et al., 2021; Bove et al., 2021). Furthermore, users' perceptions of bias are reduced if explanations include input value attributes, reference data related to the prediction, and contextual information (Cirqueira et al., 2020; Daudt et al., 2021; Hong et al., 2020; Lee and Rich, 2021; Evans et al., 2022). Moreover, a high confidence level for specific predictions helps users build trust in the system (Bussone et al., 2015; Ehsan et al., 2021).

Explanation styles have a significant impact on users' trust in a system. For example, visual explanations of the input data of the machine-learning model induce a higher level of visibility, understandability, observability, and trust in the system (Schrills and Franke, 2020; Hudon et al., 2021). User trust also varies based on whether they are informed that a human or AI made the decision; however, the variation is mostly observed when the decision is positive. Users tend to trust a system if the decision is positive irrespective of the decision maker (Bankins et al., 2022). The explanation should contain enough detail regarding the prediction and decision-making procedure so that users can feel confident and trust the system. Among various types of visual explanation formats, augmented reality-based explanations and product displays also enhance end users' trust in a system (Rodriguez-Sampaio et al., 2022). Too much information could create cognitive overload and decrease users' understanding and trust (Cramer et al., 2008; Schmidt et al., 2020; Hudon et al., 2021). The explanation should be stakeholder-oriented, such as by designing an interactive user interface to explain to non-technical stakeholders (Andres et al., 2020; Liao et al., 2020). Therefore, increased user interaction by providing adequate instructions and allowing the user to take initiatives would increase reliability and trustworthiness (Dodge et al., 2019; Oh et al., 2018; Putnam and Conati, 2019; Schrills and Franke, 2020).

If a system can simulate human-like expressions using lip sync and body language, it can increase trust (Weitz et al., 2021). For example, virtual assistants' voices, facial expressions, and gestures enhance users' trust. Therefore, a harmonic combination of a human-like facial expression along with an appropriate linguistic representation can have a significant impact on users' trust (Weitz et al., 2019: Gao et al., 2022). From the organizational point of view, employees' trust in any AI-based system is related to effectiveness, job efficiency, data protection, user understanding, and control. Weitz et al. (2019) also observed that though explanations should show the user relevant data along with the attributes, personal data need to be masked for privacy reasons (Wang and Moulden, 2021). Similarly, explanations that include comparisons among different attributes and previous and current recommendations can increase user (students, teachers, and educational researchers) trust in education-related XAI systems (Khosravi et al., 2022). Another study related to human resource management revealed that decreasing the knowledge gap between the user and the system can enhance trust (Chromik et al., 2021). Therefore, the authors also recommended reducing the knowledge gap by collaborating with users during the XAI

Table 8
Future research agenda.

Themes	Issues/topics/research gaps	Challenges and research questions	Proposed approach and research paths
XAI Standardization	Lack of holistic guidelines for XAI development for researchers and practitioners. XAI development process is opaque. Research from a regulatory and compliance perspective is not available. Communication methods and nature among the stakeholders is not defined.	Our review shows no empirical study that provides holistic guidelines or standards for developing an XAI System. GDPR is newly introduced and one of the strictest guidelines. Hence, integrating it into XAI requires rigorous investigation. RQ 1. How can XAI development guidelines be identified?	Alignment of the current software development cycle with XAI development. Different stakeholders can be involved in the XAI software development lifecycle to determine the best practices/guidelines. Identify the stakeholders. Conduct qualitative and quantitative research to identify the stakeholder requirements. Identify different types of stakeholder engagement with the XAI development lifecycle through qualitative and quantitative research. Research across multiple domains can also help portray domain-specific guidelines as well as generic guidelines for XAI development.
		RQ 2. How can we incorporate GDPR as a design requirement of XAI development?	Understand the applicable GDPR articles. Codesign with industry practitioners, end users, and legal experts to investigate the GDPR requirements. Conduct data protection impact assessments to evaluate the GDPR compliance trends. Conduct design science research to identify common guidelines for GDPR compliance.
		RQ 3. How do we integrate the proposed "Ethics guidelines for Trustworthy Artificial Intelligence" by "High-level Expert Group on Artificial Intelligence" from the European Commission?	Investigate and identify the feasibility of integrating ethical guidelines. Co-design with practitioners and experts to outline the suitable ethical guideline requirements.
		RQ 4. How can the XAI stakeholders communicate with others for collaborative development?	Stakeholders should be identified. Codesign with the stakeholders to establish a collaborative development environment. Use iterative evaluations of different types of communication techniques to find the suitable one.
XAI Visualization	Very few theory-guided studies have been conducted. The measurement of information (or explanation) quality dimensions related to XAI are not discussed. The dimensions of information (or explanation) quality are not discussed in detail for the low literacy group of people.	Information (or explanation) quality dimensions are not properly aligned with any current XAI literature. User perception measurements of explanation quality have not been performed. People with relatively low literacy might perceive the explanations differently.	Investigate the information quality dimensions in prior literature and align them with XAI. Conceptualize XAI systems' explanation quality dimensions. Develop or adapt explanation quality dimension measurement scales. Use a theory-guided approach to explain how the explanation quality dimensions affect various outcomes.
		RQ 5. How do we measure the explanation quality dimensions presented by XAI?	
		RQ 6. How does XAI representation differ in the case of relatively low literate people?	Evaluate and understand the need for differences in XAI representation for the low literacy group. Conceptualize different XAI scenarios and present explanations in various formats. Evaluate and measure various representations to find suitable representation techniques for XAI.
XAI Effects	Lack of measurement approach for trust, transparency, understandability, and usability for XAI systems. Longitudinal study has not been performed. Very few studies measure XAI's impact on domain experts, system developers, practitioners, and researchers.	Measuring the impact of different XAI representation formats. XAI effect on low literacy group is not discussed. The effect of AI explanation can have long-term effects; however, they are not outlined.	Design and validate measurement scales.
		RQ 7. How do we measure user trust, transparency, understandability, and usability?	
		RQ 8. How do explanation quality dimensions affect trust, transparency, understandability, and usability?	Investigate the effect of explanation quality dimensions on trust, transparency, understandability, and usability through survey or

(*continued on next page*)

Table 8 (*continued*)

Themes	Issues/topics/research gaps	Challenges and research questions	Proposed approach and research paths
			experiment-based research.
			Behavioral theories can be used to propose suitable research models.
		RQ 9. How do we measure the XAI impact on different stakeholders?	Identify different stakeholders of the XAI ecosystem.
			Codesign with different stakeholders of XAI to understand the effect of explanations.
			Design measurement scale and collect user responses from different stakeholders.
		RQ 10. What is the effect of XAI on low literate people?	Conceptualize case study suitable for people with low literacy.
		RQ 11. What is the longitudinal effect of XAI on end users?	Understand how the effect of XAI changes over time among different stakeholders.
			Investigate user response over time and in different domains to measure the longitudinal effect.

development lifecycle (Chromik et al., 2021; Hong et al., 2020; Park et al., 2021).

5.1.2.2. Transparency. Transparency denotes the concept of revealing the opaque procedure of decision making, allowing the whole procedure to be scrutinized by non-technical/average users if needed (Birkinshaw, 2006; Black, 1997). Therefore, making a process transparent can help to determine the features responsible for decision making, regulating, and controlling the whole process. To promote transparency, different attributes, such as age, gender, income, profession, and other related specifics, can be included in the explanation (Janssen et al., 2020). For autonomous cars, there is a significant tradeoff between explainability and system complexity. If the system needs to be more transparent (explainable), the design becomes complex, and the whole process becomes time-consuming as well (van der Waa et al., 2020).

For movie recommendation systems, content-based collaborative filtering can be adopted to increase transparency. Therefore, item-based recommendations and user-centric recommendations should be distinguishable so that the user is informed about the system and can easily connect the dots between their expectation and the system-provided recommendations (Conati et al., 2021; Ngo et al., 2020; Li et al., 2021; Liu et al., 2021a, 2021b; Dhanorkar et al., 2021). Along with the explanation, contextual information and reference data are also helpful in promoting the transparency of the systems (Liao et al., 2020). A user's involvement in the system development process by facilitating reliable communication between the system development team and the user increases transparency (Cai et al., 2019; Eiband et al., 2018). Therefore, transparency in the decision-making process and user involvement in system development can positively impact system acceptance (Conati et al., 2021; Cramer et al., 2008).

5.1.2.3. Understandability. Experimental research shows that a user's prior knowledge about the system's interactions results in better understandability and trust in the system (Branley-Bell et al., 2020; Cheng et al., 2019; Eslami et al., 2018; Lim et al., 2009; Bove et al., 2021). Moreover, presenting every interaction within the system in a sequential manner helps users understand the working procedure of the system (Broekens et al., 2010). Ehsan et al. (2021) found that social transparency is important in increasing an AI-based system's understandability; however, without background information and proper contextual information, the prediction accuracy (confidence measurement) is nothing but a number (Ehsan et al., 2021; Bove et al., 2021). Explanations with logical reasoning and counterfactual information improve the understandability of the system (Górski and Ramakrishna, 2021). In the case of expert-level end users, counterfactual explanations help to understand the generated explanations, decision making, and factors relevant to the algorithms (Evans et al., 2022). Case-based explanations can increase the understandability of decision making in criminal justice related use cases (Liu et al., 2021a, 2021b). The study by Liu et al. (2021a, 2021b) also showed that if users complete some training before using a system, this can increase the interactive nature of and the familiarity with the explanations (Liu et al., 2021a, 2021b).

Binns et al. (2018) found that a complete view of an explanation with strategic details of each process increases the user's understandability of a system. Here, the complete view denotes that the explanation should be more comprehensive and should include the scope of each event that takes place within the system's boundaries (Binns et al., 2018; Ehsan et al., 2019). Apart from textual explanations, visual explanations of the input data of the machine-learning model promote a higher level of visibility, understandability, observability, and trust in a system (Lim and Dey, 2009; Schrills and Franke, 2020; Daudt et al., 2021). Explanations involving augmented reality can also impact average users' understandability of the system (Rodriguez-Sampaio et al., 2022). In addition to an automated explanation system, an on-demand feedback retrieval system enhances the understandability of autonomous vehicle

decision making (Schneider et al., 2021). Moreover, for wearable systems, auditory feedback of explanations increases a system's understandability (Danry et al., 2020).

In the case of non-technical stakeholders, the information should be clear, concise, and comprehensive so that there is no unnecessary information that might create a cognitive overload (Hudon et al., 2021). Users of AI-based hiring systems require numerical data of the assessment along with the explanation to increase understandability. The attributes should be properly labelled and explained, and the decision should be properly reasoned to increase user understandability (Li et al., 2021). Furthermore, the explanation provision can be on-demand to avoid the monotonous and time-consuming nature of a system (Chazette and Schneider, 2020). Another way to increase understandability is to use fact sheets and mental models for a variety of stakeholders involved in the AI development process (Chromik et al., 2021; Hind et al., 2020). The fact sheet contains all the attributes of data, the prediction mechanism, the working principle, the inherent structure of the model, the training data for machine-learning models, testing protocols, and testing models (Hind et al., 2020). In addition, the developers of XAI must understand the user's mental model before developing the system. Mental model is crucial for any interactive system design since it is based on users' beliefs and perceptions about the external world. Therefore, the developer team must collaborate with end users, domain experts, and other necessary stakeholders to establish dedicated communication (Chromik et al., 2021).

5.1.2.4. Usability. XAI systems can have a positive impact on a system's usability (Oh et al., 2018). According to Chazette and Schneider (2020), for navigation systems, users would like to feel in control of the system because it provides the user with the choice of accepting or rejecting a decision. Furthermore, feedback modalities/features in autonomous vehicles can significantly increase user experiences by making the system more usable and understandable (Chazette and Schneider, 2020).

For the finance and human resource domain, following a particular explanation style is vital because presenting various explanations using a specific explanation style could help the user understand the role of various features and the reasoning behind the prediction, which can improve usability. Similarly, Szymanski et al. (2021) found that in the case of a news article recommendation system, explanations help users assess their own article writing skills and at the same time learn to improve their articles. Furthermore, to increase usability, accessible and interactive interfaces should be designed and developed for non-technical stakeholders (Andres et al., 2020; Brennen, 2020). Involving the stakeholders in the development lifecycle may also increase a system's usability (Chromik et al., 2021).

5.1.2.5. Fairness. The fairness of an intelligent system is dependent on various attributes and values as well as validity. Local explanation, which refers to an explanation of each prediction, enhances system fairness perceptions for the user. Case-based explanations have less impact on fairness criteria, but a global explanation can compensate for this and can enhance user trust (Dodge et al., 2019); however, for a criminal justice use case-based explanation, evidence-based-reasoning, legal rule sentences, and citation sentences have impacts on users' fairness perceptions (Liu et al., 2021a, 2021b; Górski and Ramakrishna, 2021). Social media related health applications and services require explanations with all types of details, logical reasoning, demographic information, and supplementary information to increase fairness perceptions among users (Liu et al., 2021a, 2021b).

For the finance and human resource domains, explanation style is vital to fairness perceptions. As mentioned, the different explanations presented in similar explanation styles could help the user understand the role of various features and the reasoning behind a prediction (Binns et al., 2018). Therefore, the user's ability to differentiate among various reasons will increase, resulting in enhanced fairness perceptions (Binns et al., 2018; Janssen et al., 2020). Furthermore, fairness is perceived more favorably by users when the input influence explanation presented is understood (Binns et al., 2018; Mucha et al., 2021).

5.1.3. Explanation presentation time

Our critical observation of the selected literature shows that explanations are provided to users in two ways. In most cases, the explanation is shown to the user automatically while visualizing the decision itself. In this case, the user wants minimal and adequate information to be provided to avoid cognitive overload (Ehsan et al., 2019; Schmidt et al., 2020; Hudon et al., 2021; Dhanorkar et al., 2021). Therefore, to have more control over an explainable AI system, users also prefer on-demand supplementary and contextual information sharing. Hence, "when the AI should be explainable" revolves around these primary concepts. For different domains, the concept is presented in various ways because the nature of the interaction is not the same across all domains. Some of these scenarios are described in detail.

For medical personnel, both textual and visual explanations and related hints are displayed automatically to doctors after decision making (Branley-Bell et al., 2020; Eiband et al., 2018, 2019; Xie et al., 2019). In addition, supplementary information should be available upon user request for better diagnosis, understandability, and trust (Wang et al., 2019; Xie et al., 2019). The supplementary information can be a combination of a reference to the previous diagnosis or any historical data that might help make an accurate decision (Branley-Bell et al., 2020; Bussone et al., 2015; Daudt et al., 2021; Lee and Rich, 2021). Similarly, various media and entertainment systems require textual and visual explanations immediately with the prediction result (Kouki et al., 2019; Ngo et al., 2020). Music recommendation systems, art recommendation systems, arcade gaming, tour guides, cooking agents, and movie recommendation systems require automatic rational generation along with personalized recommendations (Cramer et al., 2008; Ehsan et al., 2019; Lim and Dey, 2009; Ngo et al., 2020; Schmidt et al., 2020). Another requirement for automatic explanations is to reveal the filtering technique used as well as which data have been considered. For AI-based drawing tools, the user needs information on-demand rather than automatic explanations because users need to lead the task rather than receive suggestions from the system (Oh et al., 2018). One demand explanation is also required for intelligent cooking agents. The users like to lead the task and later like to receive explanations from the system (Broekens et al., 2010).

In the case of the financial domain, users are automatically presented with an explanation regarding decision making. Though the decision making is automatic, the system should show the explanation in different styles to make it more understandable (Binns et al., 2018; Cirqueira et al., 2020). Similar to healthcare decision making, contextual information is also needed upon user request (Chromik et al., 2021; Rodriguez-Sampaio et al., 2022). Autonomous car users require automatic and prompt textual and visual explanations along with the prediction result for quick decision making (Chazette and Schneider, 2020; Schneider et al., 2021). Contextual information about different scenarios should be available upon user request. Moreover, the literature analysis revealed that users of human resource management, e-commerce, and other recommendation systems (Broekens et al., 2010; Conati et al., 2021; Ehsan et al., 2021; Park et al., 2021; Zimmermann et al., 2022) require both on-demand and automatic explanations. Therefore, it is clear that in most cases, users receive explanations automatically along with the prediction result; however, supplementary information is necessary and should be available on-demand in most cases. The supplementary information can be textual, visual, or hybrid because there is no clear infrmation of this requirement in the literature. In addition, job recruiters or recruitment agencies sometimes need to backtrack the decision they made to get help in the next recruitment. Backtracking helps to understand the decision-making process of the system. Therefore, human resource managers and recruiters require mostly on-demand explanations (Li et al., 2021; Daudt et al., 2021).

5.2. Research domains

We have identified 10 domains in which XAI has been used: healthcare, media and entertainment, education, transportation, finance, e-commerce, human resource management, digital assistant, e-governance, and social networking. We comprehensively discuss the use of XAI in these domains in more detail.

5.2.1. Healthcare

Healthcare is one of the most explored research domains in XAI. This domain includes research on clinical decision making, disease diagnosis, and health-related recommendation systems (Branley-Bell et al., 2020; Bussone et al., 2015; Cai et al., 2019; Wang et al., 2019; Rodriguez-Sampaio et al., 2022). The users of AI-based systems in healthcare are primarily doctors with very little technical knowledge. Moreover, they tend to have their own opinion regarding disease detection and clinical decision making (Branley-Bell et al., 2020; Bussone et al., 2015; Lee and Rich, 2021). Therefore, the explanation required by doctors should include sufficient graphical and textual data along with appropriate contextual references (Branley-Bell et al., 2020; Daudt et al., 2021; Xie et al., 2019). In addition, healthcare-based applications, such as fitness apps and nutrition recommendations, should have a communicative and interactive user interface (Eiband et al., 2018, 2019).

5.2.2. Media and entertainment

This research domain consists of music recommendation systems, movie recommendation systems, art recommendation systems for museums and websites, news article recommendation systems, and arcade gaming systems (Kouki et al., 2019; Ngo et al., 2020; Oh et al., 2018; Szymanski et al., 2021). Users of both music and movie recommendations prefer personalized recommendations presented in various explanation styles (Ngo et al., 2020; Schmidt et al., 2020). They require explanations containing the details of the personalized recommendations that include the basic working principle of the system and details regarding the personal information used (Ngo et al., 2020; Schmidt et al., 2020). Users also express their concern regarding the amount of information being presented because too much information can create cognitive overload (Ehsan et al., 2019; Schmidt et al., 2020; Hudon et al., 2021). Therefore, the system should not overwhelm the user with unnecessary and ambiguous information and should provide both textual and visual explanations. In the case of gaming, the user interface should provide prompt hints, and the user interface should be communicative and user-friendly (Cramer et al., 2008; Ehsan et al., 2019; Schmidt et al., 2020).

5.2.3. Education

The education domain includes intelligent tutoring systems, university admission decision making, and grade estimation systems (Cheng et al., 2019; Conati et al., 2021; Mucha et al., 2021; Putnam and Conati, 2019). Investigations on intelligent tutoring systems have revealed that explanations improve the usability of the system (Conati et al., 2021; Putnam and Conati, 2019). In addition, the explanations should provide information about the system's behavior and working procedure as well as the logic behind certain decision-making tasks, such as admission decision making. For admission decision making, users often argue it should be humans who should make the decision, not a machine that simply runs on algorithms (Cheng et al., 2019; Mucha et al., 2021; Khosravi et al., 2022).

5.2.4. Transportation

The transportation domain includes navigation systems, decision support systems for autonomous cars, and flight rerouting systems for the aviation industry (Binns et al., 2018; Chazette and Schneider, 2020; Schneider et al., 2021; van der Waa et al., 2020). For domain experts, case-based explanations are preferable for autonomous car decision support systems (Schneider et al., 2021). Case-based explanations refer to explaining a certain decision in relation to use cases (Schneider et al., 2021). Moreover, the explanations (hints) provided to the user can be visual, textual, light indicators, or a hybrid mode (Schneider et al., 2021; van der Waa et al., 2020). For navigation systems, the user requirements are slightly different because the users require on-demand explanations as well as proper reasoning behind any decision being made. A similar scenario is observed in the case of flight re-routing systems. In both cases, the users wanted to control the flow of suggestions (explanations) provided by the system (Binns et al., 2018).

5.2.5. Finance

Financial use cases of XAI research include insurance, financial fraud detection, and loan applications (Binns et al., 2018; Chromik et al., 2021; Cirqueira et al., 2020). For banking activities, such as insurance claims and loan approvals, explanations regarding specific decision making should be made available to users (Chromik et al., 2021). The operator should be able to see the loan requestor's information, credit history, and other demographic information. Binns et al. (2018) and Cirqueira et al. (2020) also argued that when designing an explainable system, the developers must understand and connect with the user's mental model. An effective XAI system should be able to detect the incorrect mental model and calibrate it accordingly (Binns et al., 2018; Cirqueira et al., 2020).

5.2.6. E-commerce

The authors have investigated the social transparency and design framework that contributes to trust in the decision making of AI-based systems (Ehsan et al., 2021; Eslami et al., 2018) used in e-commerce. A better understanding of artificial intelligence-based systems is required for the promotion of social transparency. Although the effect of transparency of XAI in the long term has not been investigated, it can be utilized as a useful marketing tool (Eslami et al., 2018). Furthermore, because online advertising uses personal data for analytics, a less transparent algorithm may increase privacy issues (Ehsan et al., 2021). Online shopping experiences are better if there is an explainable AI-based system using augmented reality and text. The users have more trust and therefore a better online shopping experience (Zimmermann et al., 2022; Bove et al., 2021).

5.2.7. Human resource management

Employees' attitudes towards accepting artificial intelligence-based decisions in human resource management are influenced by a variety of factors (Binns et al., 2018; Park et al., 2021; Bankins et al., 2022). Employees often believe that the decision making can be biased, manipulative, and an invasion of privacy. Therefore, a psychological burden in accepting AI-based predictions could result. Reducing the knowledge gap by increasing transparency and interpretability can help in understanding decision making (Park et al., 2021). Moreover, collaborating with human users during the design stage can increase the chances of system adoption and can enhance the positive attitude toward the system (Binns et al., 2018; Park et al., 2021). The algorithmic hiring process is becoming increasingly popular. The recruiter's requirements for explainability include providing appropriate reasons behind the decision making, explaining the assessment scores of the candidate, and showing similar recruitments in the organization. A previous recruitment of similar kind can help the recruiters detect any possible bias in the decision-making process. Moreover, if multiple recruiters are using the system and are changing shifts, it is a good idea to provide a summary of previous work each time they log in to the system (Li et al., 2021).

5.2.8. Digital assistants

Previous studies have shown that the more human-like and interactive the system is, the more user trust increases for virtual assistants (Weitz et al., 2019, 2021). Facial expressions, voice, gestures, and verbal comments, especially those related to phonemes, are supportive and

appealing to users. Moreover, end users require linguistic explanations from an XAI system. Hence, an interactive agent with a harmonic combination of explainable AI methods and an appropriate linguistics representation can make a system trustworthy and more user-centered (Gao et al., 2022; Weitz et al., 2021).

5.2.9. E-governance

Empirical analyses have been performed on a criminal justice use case to investigate people's perceptions of the fairness of machine-learning algorithms and to what extent these algorithms need explanations (Dodge et al., 2019; Janssen et al., 2020). To increase understandability, credibility, and trust, the system should explain the algorithm's working procedure, the attributes that contribute to decision making, and the availability of contextual data (Dodge et al., 2019). Similarly, investigations of immigration services use cases reveal that though algorithms can help in decision making, it is not necessary to make all decisions using algorithms (Janssen et al., 2020). One study also revealed that the white box approach (explainable AI approach) can lead to better decision making (Janssen et al., 2020). Therefore, e-governance requires human intervention for critical decision making.

5.2.10. Social networking

Research on the social networking domain has revealed that the participants require both "why" and "why not" explanations for specific system behaviors (Lim and Dey, 2009; Yin et al., 2019; Liu et al., 2021a, 2021b). Therefore, developers can provide user log information, mental model related information, and contextual information on-demand. Moreover, an effective explainable AI system requires human user intervention in the design process through a dedicated communication medium (Yin et al., 2019).

Apart from the application domains of XAI, several studies have discussed the realm of XAI development. Studies related to XAI have been conducted to develop practical guidelines for designers, developers, domain experts, and other related stakeholders (Hind et al., 2020; Hong et al., 2020; Liao et al., 2020; Wang and Moulden, 2021). Hind et al. (2020) designed a question bank as a standard guideline for collecting user requirements for user-centered AI. The guidelines provided in this study can be a vital component in designing a trustworthy, understandable, interactive, and user-centric XAI system (Hind et al., 2020). Developers should also explore the problem space and conceptualize primary and alternative strategies (Hong et al., 2020; Liao et al., 2020). In addition, XAI development requires the active participation of domain experts, product managers, data scientists, auditors, and end users (Wang and Moulden, 2021).

6. Critical analysis of future research agendas

This section focuses on questioning and problematizing future research directions (Alvesson and Sandberg, 2011, 2020). In contrast to the previous section's discussion of the XAI research trend, this section extensively focuses on establishing a critical standpoint of future research directions by analyzing "what" is the current knowledge and "how" it can be improved (Alvesson and Sandberg, 2011, 2020). Therefore, we have reconsidered the current understanding related to XAI's methodological, conceptual, and development issues and investigated the unexplored areas. We have divided the whole observation into three primary thematic categories. The first one considers the standardization practice, the second focuses on representing XAI, and the last considers the overall effect of XAI on humans. Furthermore, rather than simply pointing out the gap that exists in the research findings, we have tried to articulate emerging research questions deduced from the unexplored research areas. We then constructed them in terms of their potential significance to identify specific and feasible research paths. Table 8 provides an overview of the future research directions based on current knowledge.

6.1. Theme 1: XAI standardization

Our analysis reveals that XAI has been used in various domains; however, there is a lack of studies that inform XAI standardization. One of the articles provides the guidelines for UI design for XAI, which both the designers and developers can use if needed (Eiband et al., 2018). Another article proposes a question bank that might be useful for requirement elicitation for explainable AI (Liao et al., 2020); however, these two articles do not offer comprehensive guidelines or standards for developing an explainable AI system. Therefore, the following research questions can be addressed for the XAI standardization theme.

6.1.1. RQ 1. How can XAI development guidelines be developed?

Extensive research on XAI design and development can facilitate

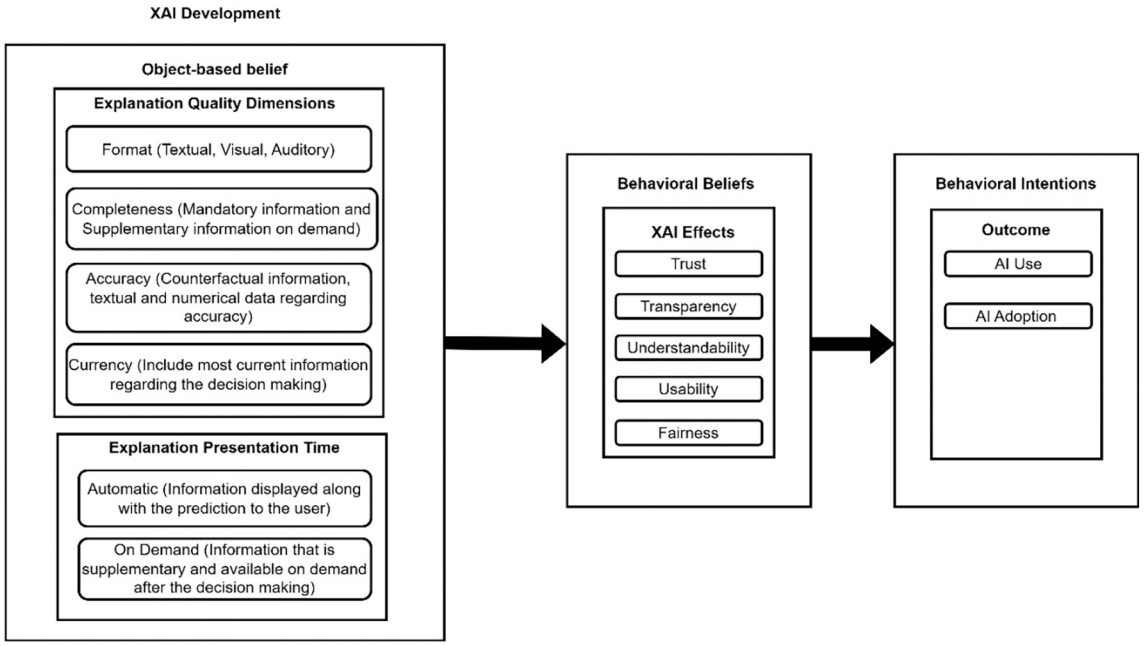

Fig. 2. Synthesized framework for XAI research from a user perspective.

determining standard guidelines and best practices for XAI development. Therefore, an important research direction would be identifying the best practices and guidelines for XAI development. Addressing this research question should include the involvement of all necessary stakeholders in the research. Furthermore, researchers across multiple domains can help create domain-specific guidelines for XAI development. Design science research can also be used to create XAI development guidelines (Hevner and Chatterjee, 2010).

6.1.2. RQ 2. How can we incorporate regulatory and ethical aspects as design requirements of XAI development?

Our observation reveals a lack of empirical studies from a regulatory and compliance perspective. Article 22 of the GDPR discusses "automated individual decision-making, including profiling," to safeguard the data subject's personal information from automatic processing (Malgieri, 2019; European Union, 2018). In addition, Articles 13–15 of the GDPR discuss the data subject's right to know the logic, that is, "meaningful information," regarding the processing of personal data. To be more precise, the data subject has the right to be informed about "meaningful information about the logic involved" if any decision related to the subject is "based solely on automatic, automated processing" (Malgieri, 2019).

To address this research question, researchers must identify the possible GDPR articles related to the XAI system. The requirements of GDPR compliance are mostly related to personal data collection, processing, retention strategy, and destruction. Therefore, co-design work that involves regulators, auditors, privacy offers, and other necessary stakeholders is a useful research direction. Another crucial step is to conduct a data protection impact assessment.

6.1.3. RQ 3. How can the stakeholders communicate with the developer team for XAI development?

We have observed from the review that communication among the developer team and other stakeholders are essential to XAI development; however, there are limited guidelines to initiate and conduct such communication (Meske et al., 2022). Therefore, to address this research question, the researchers can organize co-design workshops with the stakeholders of the XAI ecosystem so that the communication techniques can be identified and evaluated.

6.2. XAI visualization

6.2.1. RQ 4. How do we measure the explanation quality dimensions of XAI?

We have identified explanation quality dimensions in this paper. Previous studies did not empirically measure the explanation quality dimensions. In our review, we also observed that the explanation quality dimensions of AI systems unfold differently than the information quality dimensions. Therefore, researchers can search for the availability of existing measurement scales for format, completeness, accuracy, and currency. If such scales are available, researchers can adapt them to the XAI context. A major adaptation of these scales would be needed, and in fact, researchers may need to develop the scales from scratch by following the standard scale development procedure (Moore and Benbasat, 1991).

6.2.2. RQ 5. How does explanation representation differ in the case of relatively low-literate people?

The low-literacy group tends to have low AI literacy, which makes information representation more challenging. The selected articles used in this work revealed textual, visual, auditory, and hybrid modes of information representation. Different modes are used for different application domains; however, no article investigates neither how to represent the explanations to low-literate people nor how to measure their perceptions of the explanation. Therefore, addressing this research question can help present an explanation suitable for all users. The researchers should evaluate different XAI representations among low-literate people. It is vital to conceptualizing various XAI scenarios that can be presented to them.

6.3. XAI effects

6.3.1. RQ 6. How do we measure the trust, transparency, understandability, and usability of XAI? How do explanation quality dimensions affect trust, transparency, understandability, and usability?

Our observation in this review work revealed that a limited number of studies exist that measure the user perceptions of the transparency, understandability, and usability of an XAI system (Cramer et al., 2008; Daudt et al., 2021; Cheng et al., 2019). Thus, researchers should use existing measurement scales to measure these factors. The identified scales should be adapted to the XAI context. Theory-guided approaches can be used to construct models to investigate how explanation quality affects satisfaction, trust, transparency, understandability, and usability.

6.3.2. RQ 7. How do we measure the XAI impact on different stakeholders?

An AI ecosystem contains different stakeholders, such as designers, domain experts, developers, data scientists, UX engineers, and regulatory bodies (Meske et al., 2022; Laato et al., 2022). For example, domain experts can participate in the XAI development process to identify the feasibility of the explanations. Data scientists can assist the development process by designing more explainable machine-learning models. Similarly, other stakeholders can contribute to XAI development. To understand the impact on different stakeholders, a similar methodological approach can be adopted as we suggested in RQ6.

6.3.3. RQ 8. What is the effect of XAI on low-literate people?

We did not find studies that targeted low-literate people. To address this research question, first, researchers need to identify the low-literate group of people. Experiments can be designed in which AI decision-making and explanations can be presented to collect responses on explanation quality and other important factors. This type of research can also validate the developed scales in RQ4 and RQ6 among low-literate user groups.

6.3.4. RQ 9. What is the longitudinal effect of XAI on various types of end users?

Human-centered XAI can benefit from longitudinal studies because it will help researchers understand the changes in user perceptions overtime at the group level and individually. Most prior research studies on explainable AI are cross-sectional. Researchers can develop relevant research models and test them using longitudinal research design.

7. Synthesized framework for XAI research from users' perspectives

The findings from the current SLR enabled us to construct a comprehensive framework for XAI research from end users' perspectives (Fig. 2). Building on the work of Wixom and Todd (2005), our proposed comprehensive framework suggests that object-based beliefs, such as the explanation quality dimensions (format, completeness, accuracy, and currency) as well as when to explain (automatic and on-demand), impact a number of behavioral beliefs, including trust, transparency, understandability, usability, and fairness. In turn, these behavioral beliefs impact behavioral intention (AI adoption, AI use).

According to Wixom and Todd (2005), object-based beliefs are the characteristics of technology, whereas behavioral beliefs are the anticipated consequences of technology use. Wixom and Todd (2005) suggested that the impacts of object-based beliefs on behavioral beliefs are mediated through the object-based attitude (Eagly and Chaiken, 1993; Fazio and Olson, 2003); however, in a recent empirical study (Islam et al., 2020), it was shown that object-based beliefs can have direct impacts on behavioral beliefs. Therefore, we have proposed direct

relationships between object-based beliefs and behavioral beliefs in our framework. Fig. 2 shows the graphical representation of the framework.

8. Implications

8.1. Theoretical implications

Our SLR findings have five major theoretical contributions. First, from a broad perspective, our study is one of the few studies investigating AI end users' explanation needs. Therefore, our paper contributes to the previously conducted literature reviews (Wells and Bednarz, 2021; Anjomshoae et al., 2019; Gerlings et al., 2021a, 2021b; Laato et al., 2022), particularly by identifying the end users' explanation needs and the impacts.

Second, we adopted Wixom and Todd's (2005) conceptualization of information quality dimensions to conceptualize the explanation quality dimensions of AI systems. Our findings show that explanation quality dimensions are format, completeness, accuracy, and currency. We have also observed that when to explain (automatic and on-demand) is another important factor of XAI. With our findings, we contribute to research conducted to design and govern responsible AI systems (Wearn et al., 2019; Maas, 2018; Peters et al., 2020; Rakova et al., 2021).

Third, we have described the five effects of XAI systems: trust, transparency, understandability, usability, and fairness. Our SLR findings position these factors as the most important effects of XAI. While these factors are described by Laato et al. (2022), our SLR links them with XAI representation dimensions.

Fourth, we have identified three major themes of future research: XAI standardization, XAI visualization, and XAI effects. We have proposed nine possible research questions that future IS researchers can investigate. We have also outlined the possible ways researchers can address these research questions.

Finally, we have proposed a comprehensive framework by connecting explanation-related factors and XAI effects. We further propose that the XAI effects can ultimately influence behaviors, such as AI adoption and use. This framework has implications for researchers. For example, many interesting research models can be developed and tested based on this framework. While the framework is developed using Wixom and Todd's (2005) work, which describes relationships among object-based beliefs, behavioral beliefs, and behavior, hypotheses can be developed from additional theories, such as the IS success model (DeLone and McLean, 1992, 2002), technology acceptance models (Davis, 1989; Davis et al., 1989; Chuttur, 2009), the theory of reasoned actions (Ajzen and Fishbein, 1973; Ajzen and Fishbein, 1980; Fishbein, 1967; Fishbein and Ajzen, 1977; Hale et al., 2002), and the theory of planned behavior (Ajzen, 1985, 1991).

8.2. Practical implications

From a practical standpoint, this SLR can serve as a guideline for designing human-centric AI and measuring its consequences. Because AI is becoming more prevalent in all aspects of life, the findings of this study may drive researchers and enthusiasts to design digital services that are morally sustainable. For example, designers can ensure that their systems provide explanations related to the identified dimensions of explanation quality. Their design should also contain possibilities for both automatic and on-demand explanations. Our findings also outline a need to design XAI systems in various domains, not just for mission-critical systems. Since AI is now being used more than ever in various industrial and corporate decision-making, the findings of this SLR can help understand the employees' behavioral intention to use those systems. As discussed in the literature, various state-of-the-art recruitment systems use data-driven decision-making. In cases like this, the explanation dimensions can help to understand the details of the decision-making process. Hence, the synthesized framework of this SLR can be adopted in various industries and corporate organizations to understand the likelihood of system adoption and use. Therefore, we suggest that system designers consider this need when they design AI-based systems. This also has implications for AI education. We suggest including topics such as explainable AI, responsible AI, and AI governance, among others, as important topics to train AI developers in addition to technical topics.

9. Conclusion

Recently, AI has gained significant momentum, which, if correctly managed, does have the potential to revolutionize various sectors; however, the AI community must overcome the challenge of explainability, an intrinsic hurdle that was not a part of AI-based ecosystems before. This work has comprehensively discussed XAI from the end user's perspective. We have identified the dimensions of explanation quality from existing empirical studies, and we found that the effects of XAI on end users can motivate users to adopt and use AI-based systems. Furthermore, by investigating the selected studies, we have identified crucial future research avenues. Possible directions to address these avenues and a comprehensive framework have also been identified and developed, respectively. Though the widespread application of XAI is yet to be implemented, based on our review, the growing need for XAI is vividly clear. The explanation quality dimensions of XAI outlined in this work are vital to XAI system development because the dimensions can have impacts on trust, understandability, fairness, and transparency.

Our study has three limitations. First, we have considered only the empirical studies on XAI for this review work. Future studies can also consider theoretical papers on XAI.

Second, we used Scopus and Web of Science for the database search. Hence, we might have missed important studies for our work. This limitation can be addressed in the future by conducting searches of other databases.

Third, we have used Wixom and Todd's (2005) information quality dimensions for conceptualizing the explanation quality of AI systems. There are other information quality dimensions proposed by other researchers (Wang and Strong, 1996). Therefore, future studies can use these dimensions to identify additional explanation quality dimensions for AI systems.

CRediT authorship contribution statement

AKM Bahalul Haque: Conceptualization, Methodology, Conducting Primary Search, Data Collection, Writing Original Draft, Analyzing and Addressing the Reviewer's Comments

A.K.M. Najmul Islam: Conceptualization, Reviewing Draft, Editing, Reviewing the Search Result, Critically Analyzing Reviewers' Comments, Supervision

Patrick Mikalef: Conceptualization, Reviewing Draft, Reviewing the Search Result and Data, Critically Analyzing Reviewers' Comments, Supervision

Declaration of competing interest

None.

Acknowledgements

This work was supported by the Slovenian Research Agency (research core funding No. P5-0410).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.techfore.2022.122120.

References

Adadi, A., Berrada, M., 2018. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access 6, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052.

Ajzen, I., 1985. From intentions to actions: a theory of planned behavior. In: Kuhl, J., Beckmann, J. (Eds.), Action Control. SSSP Springer Series in Social Psychology. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-69746-3_2.

Ajzen, I., 1991. The theory of planned behavior. Organ. Behav. Hum. Decis. Process. 50 (2), 179–211.

Ajzen, I., Fishbein, M., 1973. Attitudinal and normative variables as predictors of specific behavior. J. Pers. Soc. Psychol. 27 (1), 41–57. https://doi.org/10.1037/h0034440.

Ajzen, I., Fishbein, M., 1980. Understanding Attitudes and Predicting Social Behavior. Prentice-Hall, Englewood Cliffs, NJ.

Alvesson, M., Sandberg, J., 2011. Generating research questions through problematization. Acad. Manag. Rev. 36 (2), 247–271.

Alvesson, M., Sandberg, J., 2020. The problematizing review: a counterpoint to elsbach and Van Knippenberg's argument for integrative reviews. J. Manag. Stud. 57 (6), 1290–1304.

Andres, J., Wolf, C.T., Cabrero Barros, S., Oduor, E., Nair, R., Kjærum, A., Tharsgaard, A. B., Madsen, B.S., 2020. Scenario-based XAI for humanitarian aid forecasting. In: Conference on Human Factors in Computing Systems - Proceedings, 1–8. https://doi.org/10.1145/3334480.3382903.

Angelov, P., Soares, E., 2020. Towards explainable deep neural networks (xDNN). Neural Netw. 130, 185–194. https://doi.org/10.1016/j.neunet.2020.07.010.

Anjomshoae, S., Calvaresi, D., Najjar, A., Främling, K., 2019. Explainable agents and robots: results from a systematic literature review. In: Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, 2. AAMAS, pp. 1078–1088 (Aamas).

Antoniadi, A.M., Du, Y., Guendouz, Y., Wei, L., Mazo, C., Becker, B.A., Mooney, C., 2021. Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: a systematic review. Appl. Sci. (Switzerland) 11 (11), 5088. https://doi.org/10.3390/app11115088.

Antunes, P., Herskovic, V., Ochoa, S.F., Pino, J.A., 2012. Structuring dimensions for collaborative systems evaluation. ACM Comput. Surv. 44 (2) https://doi.org/10.1145/2089125.2089128.

Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Herrera, F., 2020. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion 58, 82–115.

Bankins, S., Formosa, P., Griep, Y., Richards, D., 2022. AI decision making with dignity? Contrasting workers' justice perceptions of human and AI decision making in a human resource management context. Inf. Syst. Front. 1–19.

Baum, S.D., Goertzel, B., Goertzel, T.G., 2011. How long until human-level AI? Results from an expert assessment. Technol. Forecast. Soc. Chang. 78 (1), 185–195.

Benbasat, I., Wang, W., 2005. Trust in and adoption of online recommendation agents. J. Assoc. Inf. Syst. 6 (3), 4.

Binns, R., Van Kleek, M., Veale, M., Lyngs, U., Zhao, J., Shadbolt, N., 2018. "It's reducing a human being to a percentage"; perceptions of justice in algorithmic decisions. In: Conference on Human Factors in Computing Systems - Proceedings, 2018-April, 1–14. https://doi.org/10.1145/3173574.3173951.

Birkinshaw, P., 2006. Freedom of information and openness: fundamental human rights. Admin. L. Rev. 58, 177.

Black, J., 1997. New institutionalism and naturalism in socio-legal analysis: institutionalist approaches to regulatory decision making. Law Policy 19 (1), 51–93.

Bove, C., Aigrain, J., Lesot, M.J., Tijus, C., Detyniecki, M., 2021. Contextualising local explanations for non-expert users: an XAI pricing interface for insurance. In: IUI Workshops.

Branley-Bell, D., Whitworth, R., Coventry, L., 2020. User trust and understanding of explainable ai: Exploring algorithm visualisations and user biases. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12183 LNCS(Ml), pp. 382–399. https://doi.org/10.1007/978-3-030-49065-2_27.

Brennen, A., 2020. What do people really want when they say they want "explainable AI?" We asked 60 stakeholders. In: Conference on Human Factors in Computing Systems - Proceedings, 1–7. https://doi.org/10.1145/3334480.3383047.

Broekens, J., Harbers, M., Hindriks, K., Van Den Bosch, K., Jonker, C., Meyer, J.J., 2010. Do you get it? User-evaluated explainable BDI agents. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 6251 LNAI, pp. 28–39. https://doi.org/10.1007/978-3-642-16178-0_5.

Bussone, A., Stumpf, S., O'Sullivan, D., 2015. The role of explanations on trust and reliance in clinical decision support systems. In: Proceedings - 2015 IEEE International Conference on Healthcare Informatics, ICHI 2015, pp. 160–169. https://doi.org/10.1109/ICHI.2015.26.

Cai, C.J., Winter, S., Steiner, D., Wilcox, L., Terry, M., 2019. "Hello Ai": Uncovering the onboarding needs of medical practitioners for human–AI collaborative decision-making. In: Proceedings of the ACM on Human-Computer Interaction, 3(CSCW). https://doi.org/10.1145/3359206.

Castelvecchi, D., 2016. Can we open the black box of AI? Nature 538 (7623), 20–23. https://doi.org/10.1038/538020a.

Chakrobartty, S., El-Gayar, O., 2021. Explainable Artificial Intelligence in the Medical Domain: A Systematic Review. AMCIS 2021 Proceedings. https://aisel.aisnet.org/amcis2021/art_intel_sem_tech_intelligent_systems/art_intel_sem_tech_intelligent_systems/1.

Chazette, L., Schneider, K., 2020. Explainability as a non-functional requirement: challenges and recommendations. Requir. Eng. 25 (4), 493–514. https://doi.org/10.1007/s00766-020-00333-1.

Cheng, H.F., Wang, R., Zhang, Z., O'Connell, F., Gray, T., Harper, F.M., Zhu, H., 2019. Explaining decision-making algorithms through UI: strategies to help non-expert stakeholders. In: Conference on Human Factors in Computing Systems - Proceedings, 1–12. https://doi.org/10.1145/3290605.3300789.

Choi, K., Yoo, D., Kim, G., Suh, Y., 2012. A hybrid online-product recommendation system: combining implicit rating-based collaborative filtering and sequential pattern analysis. Electron. Commer. Res. Appl. 11 (4), 309–317. https://doi.org/10.1016/j.elerap.2012.02.004.

Chromik, M., Butz, A., 2021. Human-xai interaction: A review and design principles for explanation user interfaces. In: IFIP Conference on Human-Computer Interaction. Springer, Cham, pp. 619–640.

Chromik, M., Eiband, M., Buchner, F., Krüger, A., Butz, A., 2021. I think I get your point, AI! The illusion of explanatory depth in explainable AI. In: International Conference on Intelligent User Interfaces, Proceedings IUI, pp. 307–317. https://doi.org/10.1145/3397481.3450644.

Chuttur, M.Y., 2009. Overview of the technology acceptance model: origins, developments and future directions. In: Working Papers on Information Systems, 9, pp. 9–37 (37).

Cirqueira, D., Nedbal, D., Helfert, M., Bezbradica, M., 2020. In: Scenario-based Requirements Elicitation for User-centric Explainable AI: A Case in Fraud Detection. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12279 LNCS, pp. 321–341. https://doi.org/10.1007/978-3-030-57321-8_18.

Conati, C., Barral, O., Putnam, V., Rieger, L., 2021. Toward personalized XAI: a case study in intelligent tutoring systems. Artif. Intell. 298, 103503 https://doi.org/10.1016/j.artint.2021.103503.

Cramer, H., Evers, V., Ramlal, S., Van Someren, M., Rutledge, L., Stash, N., Aroyo, L., Wielinga, B., 2008. The effects of transparency on trust in and acceptance of a content-based art recommender. User Model. User-Adap. Inter. 18 (5) https://doi.org/10.1007/s11257-008-9051-3.

Dağlarli, E., 2020. Explainable artificial intelligence (xAI) approaches and deep meta-learning models. Adv. Appl. Deep Learning. https://doi.org/10.5772/intechopen.92172.

Danry, V., Pataranutaporn, P., Mao, Y., Maes, P., 2020. Wearable Reasoner: Towards Enhanced Human Rationality Through A Wearable Device with An Explainable AI Assistant. PervasiveHealth: Pervasive Computing Technologies for Healthcare. https://doi.org/10.1145/3384657.3384799.

Daudt, F., Cinalli, D., Garcia, A.C.B., 2021. In: Research on Explainable Artificial Intelligence Techniques: An User Perspective. Proceedings of the 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design, CSCWD 2021, pp. 144–149. https://doi.org/10.1109/CSCWD49262.2021.9437820.

Davis, F.D., 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q. 13 (3), 319–340. https://doi.org/10.2307/249008.

Davis, Fred D., Bagozzi, Richard P., Warshaw, Paul R., 1989. User acceptance of computer technology: a comparison of two theoretical models. Manag. Sci. 35 (8), 982–1003. https://doi.org/10.1287/mnsc.35.8.982.

DeLone, W.H., McLean, E.R., 1992. Information systems success: the quest for the dependent variable. Inf. Syst. Res. 3 (1), 60–95.

DeLone, W.H., McLean, E.R., 2002. Information systems success revisited. In: Proceedings of the 35th Annual Hawaii International Conference on System Sciences. IEEE, pp. 2966–2976.

Dhanorkar, S., Wolf, C.T., Qian, K., Xu, A., Popa, L., Li, Y., 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, pp. 1591–1602.

Dodge, J., Vera Liao, Q., Zhang, Y., Bellamy, R.K.E., Dugan, C., 2019. Explaining models: An empirical study of how explanations impact fairness judgment. In: International Conference on Intelligent User Interfaces, Proceedings IUI, Part F1476, pp. 275–285. https://doi.org/10.1145/3301275.3302310.

Doshi-Velez, F., Kim, B., 2017. Towards a rigorous science of interpretable machine learning. http://arxiv.org/abs/1702.08608.

Du, S., Xie, C., 2021. Paradoxes of artificial intelligence in consumer markets: ethical challenges and opportunities. J. Bus. Res. 129, 961–974.

Eagly, A.H., Chaiken, S., 1993. The Psychology of Attitudes. Thomson Wadsworth, Belmont, CA.

Ehsan, U., Tambwekar, P., Chan, L., Harrison, B., Riedl, M.O., 2019. Automated rationale generation: A technique for explainable AI and its effects on human perceptions. In: International Conference on Intelligent User Interfaces, Proceedings IUI, Part F1476, pp. 263–274. https://doi.org/10.1145/3301275.3302316.

Ehsan, U., Liao, Q.V., Muller, M., Riedl, M.O., Weisz, J.D., 2021. Expanding explainability: Towards social transparency in ai systems. In: Conference on Human Factors in Computing Systems - Proceedings. https://doi.org/10.1145/3411764.3445188.

Eiband, M., Schneider, H., Bilandzic, M., Fazekas-Con, J., Haug, M., Hussmann, H., 2018. Bringing transparency design into practice. In: International Conference on Intelligent User Interfaces, Proceedings IUI, pp. 211–223. https://doi.org/10.1145/3172944.3172961.

Eiband, M., Buschek, D., Kremer, A., Hussmann, H., 2019. The impact of placebic explanations on trust in intelligent systems. In: Conference on Human Factors in Computing Systems - Proceedings. https://doi.org/10.1145/3290607.3312787.

Eslami, M., Kumaran, S.R.K., Sandvig, C., Karahalios, K., 2018. Communicating algorithmic process in online behavioral advertising. In: Conference on Human

Factors in Computing Systems - Proceedings, 2018-April, 1–13. https://doi.org/10.1145/3173574.3174006.
European Union, 2018. Art. 22 GDPR - Automated individual decision-making, including profiling. General Data Protection Regulation. https://gdpr-info.eu/art-22-gdpr/.
Evans, T., Retzlaff, C.O., Geißler, C., Kargl, M., Plass, M., Müller, H., Holzinger, A., 2022. The explainability paradox: challenges for xAI in digital pathology. Futur. Gener. Comput. Syst. 133, 281–296.
Fazio, R.H., Olson, M.A., 2003. Attitudes: foundation, function and consequences. In: Hogg, M.A., Cooper, J. (Eds.), The Sage Handbook of Social Psychology. Sage, London, UK.
Feng, C., Khan, M., Rahman, A.U., Ahmad, A., 2020. News recommendation systems-accomplishments, challenges future directions. IEEE Access 8, 16702–16725. https://doi.org/10.1109/ACCESS.2020.2967792.
Fishbein, M., 1967. Attitude and the Prediction of Behavior. Readings in attitude Theory and Measurement.
Fishbein, M., Ajzen, I., 1977. Belief, attitude, intention, and behavior: an introduction to theory and research. Philos. Rhetor. 10 (2).
Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., Srikumar, M., 2020. Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI. Berkman Klein Center Research Publication (2020-1).
Gao, M., Liu, X., Xu, A., Akkiraju, R., 2022. In: Arai, K. (Ed.), Intelligent Systems and Applications. IntelliSys 2021. Lecture Notes in Networks and Systems, 296. Springer, Cham.
Gerlings, J., Jensen, M.S., Shollo, A., 2021. Explainable AI, but explainable to whom? http://arxiv.org/abs/2106.05568.
Gerlings, J., Shollo, A., Constantiou, I., 2021. Reviewing the need for explainable artificial intelligence (XAI). In: Proceedings of the Annual Hawaii International Conference on System Sciences, pp. 1284–1293. https://doi.org/10.24251/hicss.2021.156, 2020-Janua.
Ghallab, M., 2019. Responsible AI: requirements and challenges. AI Perspect. 1 (1), 1–7. https://doi.org/10.1186/s42467-019-0003-z.
Goodman, B., Flaxman, S., 2017. European union regulations on algorithmic decision making and a "right to explanation". AI Mag. 38 (3), 50–57. https://doi.org/10.1609/aimag.v38i3.2741.
Górski, Ł., Ramakrishna, S., 2021. Explainable artificial intelligence, lawyer's perspective. In: Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law, pp. 60–68.
Gruetzemacher, R., Dorner, F.E., Bernaola-Alvarez, N., Giattino, C., Manheim, D., 2021. Forecasting AI progress: a research agenda. Technol. Forecast. Soc. Chang. 170, 120909.
Gunning, D., Aha, D.W., 2019. DARPA's explainable artificial intelligence program. AI Mag. 40 (2), 44–58. https://doi.org/10.1609/aimag.v40i2.2850.
Hale, J.L., Householder, B.J., Greene, K.L., 2002. The theory of reasoned action. In: The Persuasion Handbook: Developments in Theory and Practice, 14, pp. 259–286.
Haque, A.K.M.B., Hasan Pranto, T., All Noman, A., Mahmood, A., 2020. Insight about detection, prediction and weather impact of coronavirus (Covid-19) using neural network. Int. J. Artif. Intell. Appl. 11 (4), 67–81. https://doi.org/10.5121/ijaia.2020.11406.
Haque, A.K.M.B., Bhushan, B., Dhiman, G., 2021. Conceptualizing smart city applications: requirements, architecture, security issues, and emerging trends. Expert. Syst. https://doi.org/10.1111/exsy.12753.
Hasan, R., Shams, R., Rahman, M., 2021. Consumer trust and perceived risk for voice-controlled artificial intelligence: the case of Siri. J. Bus. Res. 131, 591–597.
Hengstler, M., Enkel, E., Duelli, S., 2016. Applied artificial intelligence and trust—the case of autonomous vehicles and medical assistance devices. Technol. Forecast. Soc. Chang. 105, 105–120.
Hevner, A., Chatterjee, S., 2010. Design science research in information systems. In: Design research in information systems. Springer, Boston, MA, pp. 9–22.
Hind, M., Houde, S., Martino, J., Mojsilovic, A., Piorkowski, D., Richards, J., Varshney, K.R., 2020. Experiences with improving the transparency of AI models and services. In: Conference on Human Factors in Computing Systems - Proceedings, pp. 1–8. https://doi.org/10.1145/3334480.3383051.
Hong, S.R., Hullman, J., Bertini, E., 2020. Human factors in model interpretability: industry practices, challenges, and needs. In: Proceedings of the ACM on Human-Computer Interaction, 4. CSCW1, pp. 1–26. https://doi.org/10.1145/3392878.
Hudon, A., Demazure, T., Karran, A., Léger, P.M., Sénécal, S., 2021. Explainable artificial intelligence (XAI): how the visualization of AI predictions affects user cognitive load and confidence. In: NeuroIS Retreat. Springer, Cham, pp. 237–246.
IDC, 2018. Worldwide Artificial Intelligence Spending Guide. International Data Corporation. https://www.idc.com/getdoc.jsp?containerId=IDC_P33198.
Islam, A.N., Cenfetelli, R., Benbasat, I., 2020. Organizational buyers' assimilation of B2B platforms: effects of IT-enabled service functionality. J. Strateg. Inf. Syst. 29 (1), 101597.
Janssen, M., Hartog, M., Matheus, R., Yi Ding, A., Kuk, G., 2020. Will algorithms blind People? The effect of explainable AI and decision-makers' experience on AI-supported decision-making in government. Soc. Sci. Comput. Rev. 1–16 https://doi.org/10.1177/0894439320980118.
Khosravi, H., Shum, S.B., Chen, G., Conati, C., Tsai, Y.S., Kay, J., Gašević, D., 2022. Explainable artificial intelligence in education. Comput. Educ. Artif. Intell. 3, 100074.
Kitchenham, B.A., Charters, 2007. In: Guidelines for performing systematic literature reviews in software engineering. Technical Report, Ver. 2.3 EBSE Technical Report, 1. EBSE, pp. 1–54.
Kouki, P., Schaffer, J., Pujara, J., O'Donovan, J., Getoor, L., 2019. Personalized explanations for hybrid recommender systems. In: International Conference on Intelligent User Interfaces, Proceedings IUI, Part F1476, pp. 379–390. https://doi.org/10.1145/3301275.3302306.
Laato, S., Tiainen, M., Islam, A.N., Mäntymäki, M., 2022. How to explain AI systems to end users: a systematic literature review and research agenda. Internet Res. 32 (7), 1–31.
Lauritsen, S.M., Kristensen, M., Olsen, M.V., Larsen, M.S., Lauritsen, K.M., Jørgensen, M. J., Lange, J., Thiesson, B., 2020. Explainable artificial intelligence model to predict acute critical illness from electronic health records. NatureCommunications 11 (1). https://doi.org/10.1038/s41467-020-17431-x.
Lee, M.K., Rich, K., 2021. Who is included in human perceptions of ai?: trust and perceived fairness around healthcare AI and cultural mistrust. In: Conference on Human Factors in Computing Systems - Proceedings. https://doi.org/10.1145/3411764.3445570.
Li, L., Lassiter, T., Oh, J., Lee, M.K., 2021. Algorithmic hiring in practice: recruiter and HR professional's perspectives on AI use in hiring. In: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, pp. 166–176.
Liao, Q.V., Gruen, D., Miller, S., 2020. Questioning the AI: informing design practices for explainable AI user experiences. In: Conference on Human Factors in Computing Systems - Proceedings, 1–15. https://doi.org/10.1145/3313831.3376590.
Lim, B.Y., Dey, A.K., 2009. Assessing demand for intelligibility in context-aware applications. In: ACM International Conference Proceeding Series, 195–204. https://doi.org/10.1145/1620545.1620576.
Lim, B.Y., Dey, A.K., Avrahami, D., 2009. Why and why not explanations improve the intelligibility of context-aware intelligent systems. In: Conference on Human Factors in Computing Systems - Proceedings, 2119–2128. https://doi.org/10.1145/1518701.1519023.
Linardatos, P., Papastefanopoulos, V., Kotsiantis, S., 2021. Explainable AI: a review of machine learning interpretability methods. Entropy 23 (1), 1–45. https://doi.org/10.3390/e23010018.
Lipton, Z.C., 2018. The mythos of model interpretability: in machine learning, the concept of interpretability is both important and slippery. Queue 16 (3). https://doi.org/10.1145/3236386.3241340.
Liu, H., Lai, V., Tan, C., 2021. Understanding the effect of out-of-distribution examples and interactive explanations on human-ai decision making. In: Proceedings of the ACM on Human-Computer Interaction, 5. CSCW2, pp. 1–45.
Liu, R., Gupta, S., Patel, P., 2021. The application of the principles of responsible AI on social media marketing for digital health. Inf. Syst. Front. 1–25.
Maas, M.M., 2018. Regulating for'Normal AI Accidents' Operational Lessons for the Responsible Governance of Artificial Intelligence Deployment. In: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, pp. 223–228.
Mahmud, H., Islam, A.N., Ahmed, S.I., Smolander, K., 2022a. What influences algorithmic decision-making? A systematic literature review on algorithm aversion. Technol. Forecast. Soc. Chang. 175, 121390.
Mahmud, H., Islam, A.K.M.N., Mitra, R.K., Hasan, A.R., 2022b. The Impact of Functional and Psychological Barriers on Algorithm Aversion – An IRT Perspective. In: Papagiannidis, S., Alamanos, E., Gupta, S., Dwivedi, Y.K., Mäntymäki, M., Pappas, I. O. (Eds.), The Role of Digital Technologies in Shaping the Post-Pandemic World. I3E 2022, Lecture Notes in Computer Science, 13454. Springer, Cham.
Malgieri, G., 2019. Automated decision-making in the EU member states: the right to explanation and other "suitable safeguards" in the national legislations. Comput. Law Secur. Rev. 35 (5), 105327 https://doi.org/10.1016/j.clsr.2019.05.002.
Meske, C., Bunde, E., Schneider, J., Gersch, M., 2022. Explainable artificial intelligence: objectives, stakeholders, and future research opportunities. Inf. Syst. Manag. 39 (1), 53–63.
Moore, G.C., Benbasat, I., 1991. Development of an instrument to measure the perceptions of adopting an information technology innovation. Inf. Syst. Res. 2 (3), 192–222.
Mucha, H., Robert, S., Breitschwerdt, R., Fellmann, M., 2021. Interfaces for explanations in human-AI interaction: proposing a design evaluation approach. In: Conference on Human Factors in Computing Systems - Proceedings. https://doi.org/10.1145/3411763.3451759.
Ngo, T., Kunkel, J., Ziegler, J., 2020. In: Exploring Mental Models for Transparent and Controllable Recommender Systems: A Qualitative Study. UMAP 2020 - Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization, pp. 183–191. https://doi.org/10.1145/3340631.3394841.
Oh, C., Song, J., Choi, J., Kim, S., Lee, S., Suh, B., 2018. I lead, you help but only with enough details: Understanding the user experience of co-creation with artificial intelligence. In: Conference on Human Factors in Computing Systems - Proceedings, 2018-April, 1–13. https://doi.org/10.1145/3173574.3174223.
Park, H., Ahn, D., Hosanagar, K., Lee, J., 2021. Human-ai interaction in human resource management: understanding why employees resist algorithmic evaluation at workplaces and how to mitigate burdens. In: Conference on Human Factors in Computing Systems - Proceedings. https://doi.org/10.1145/3411764.3445304.
Peters, D., Vold, K., Robinson, D., Calvo, R.A., 2020. Responsible AI—two frameworks for ethical design practice. IEEE Trans. Technol. Soc. 1 (1), 34–47.
Putnam, V., Conati, C., 2019. Exploring the need for explainable artificial intelligence (XAI) in intelligent tutoring systems (ITS). In: CEUR Workshop Proceedings, p. 2327.
Rakova, B., Yang, J., Cramer, H., Chowdhury, R., 2021. Where responsible AI meets reality: Practitioner perspectives on enablers for shifting organizational practices. In: Proceedings of the ACM on Human-Computer Interaction, 5. CSCW1, pp. 1–23.
Rodriguez-Sampaio, M., Rincón, M., Valladares-Rodríguez, S., Bachiller-Mayoral, M., 2022. Explainable artificial intelligence to detect breast cancer: A qualitative case-based visual interpretability approach. In: International Work-Conference on the Interplay between Natural and Artificial Computation. Springer, Cham, pp. 557–566.

Schmidt, P., Biessmann, F., Teubner, T., 2020. Transparency and trust in artificial intelligence systems. J. Decis. Syst. 29 (4), 260–278. https://doi.org/10.1080/12460125.2020.1819094.
Schneider, Johannes, Handali, Joshua, 2019. Personalized explanation in machine learning: A conceptualization. In: European Conference of Information Systems.
Schneider, T., Ghellal, S., Love, S., Gerlicher, A.R.S., 2021. Increasing the user experience in autonomous driving through different feedback modalities. In: International Conference on Intelligent User Interfaces, Proceedings IUI, 7–10. https://doi.org/10.1145/3397481.3450687.
Schrills, T., Franke, T., 2020. Color for characters - effects of visual explanations of AI on trust and observability. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12217 LNCS, pp. 121–135. https://doi.org/10.1007/978-3-030-50334-5_8.
Scott, A.C., CWJDR, Shortliffe, E.H., 1977. Explanation capabilities of production-based consultation systems. American Journal of Computational Linguistics 1–50. https://aclanthology.org/J77-1006.
Stahl, B.C., Andreou, A., Brey, P., Hatzakis, T., Kirichenko, A., Macnish, K., Wright, D., 2021. Artificial intelligence for human flourishing–beyond principles for machine learning. Journal of Business Research 124, 374–388.
Szymanski, M., Millecamp, M., Verbert, K., 2021. Visual, textual or hybrid: the effect of user expertise on different explanations. In: International Conference on Intelligent User Interfaces, Proceedings IUI, 109–119. https://doi.org/10.1145/3397481.3450662.
Tiainen, M., 2021. To Whom to Explain and What?: Systematic Literature Review on Empirical Studies on Explainable Artificial Intelligence (XAI) (Master Thesis). accessed June 5, 2022. https://www.utupub.fi/handle/10024/151554.
van der Waa, J., Schoonderwoerd, T., van Diggelen, J., Neerincx, M., 2020. Interpretable confidence measures for decision support systems. Int. J. Hum. Comput. Stud. 144 (May), 102493 https://doi.org/10.1016/j.ijhcs.2020.102493.
Wachter, S., Mittelstadt, B., Floridi, L., 2017. Transparent, explainable, and accountable AI for robotics. ScienceRobotics 2 (6). https://doi.org/10.1126/scirobotics.aan6080.
Wang, D., Yang, Q., Abdul, A., Lim, B.Y., States, U., 2019. In: Designing Theory-Driven User-Centric Explainable AI, pp. 1–15.
Wang, J., Moulden, A., 2021. AI trust score: a user-centered approach to building, designing, and measuring the success of intelligent workplace features. In: Conference on Human Factors in Computing Systems - Proceedings. https://doi.org/10.1145/3411763.3443452.
Wang, R.Y., Strong, D.M., 1996. Beyond accuracy: what data quality means to data consumers. J. Manag. Inf. Syst. 12 (4), 5–33.
Wang, Z., Yu, X., Feng, N., Wang, Z., 2014. An improved collaborative movie recommendation system using computational intelligence. J. Vis. Lang. Comput. 25 (6), 667–675. https://doi.org/10.1016/j.jvlc.2014.09.011.
Wearn, O.R., Freeman, R., Jacoby, D.M., 2019. Responsible AI for conservation. Nat. Mach. Intell. 1 (2), 72–73.
Weitz, K., Schiller, D., Schlagowski, R., Huber, T., André, E., 2019. I "do you trust me?": increasing user-trust by integrating virtual agents in explainable AI interaction design. In: IVA 2019 - Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents, pp. 7–9. https://doi.org/10.1145/3308532.3329441.
Weitz, K., Schiller, D., Schlagowski, R., Huber, T., André, E., 2021. "Let me explain!": exploring the potential of virtual agents in explainable AI interaction design. J. Multimodal User Interfaces 15 (2), 87–98. https://doi.org/10.1007/s12193-020-00332-0.
Wells, L., Bednarz, T., 2021. Explainable AI and reinforcement learning—a systematic review of current approaches and trends. Front. Artif. Intell. 4, 550030 https://doi.org/10.3389/frai.2021.550030.
Wixom, B.H., Todd, P.A., 2005. A theoretical integration of user satisfaction and technology acceptance. Inf. Syst. Res. 16 (1), 85–102. https://doi.org/10.1287/ISRE.1050.0042.
Xie, Y., Chen, X.A., Gao, G., 2019. Outlining the design space of explainable intelligent systems for medical diagnosis. In: CEUR Workshop Proceedings, p. 2327.
Yin, M., Vaughan, J.W., Wallach, H., 2019. Understanding the effect of accuracy on trust in machine learning models. In: Conference on Human Factors in Computing Systems - Proceedings, pp. 1–12. https://doi.org/10.1145/3290605.3300509.
Zimmermann, R., Mora, D., Cirqueira, D., Helfert, M., Bezbradica, M., Werth, D., Weitzl, W.J., Riedl, R., Auinger, A., 2022. Enhancing brick-and-mortar store shopping experience with an augmented reality shopping assistant application using personalized recommendations and explainable artificial intelligence. Journal of Research in Interactive Marketing. Vol. ahead-of-print No. ahead-of-print.

AKM Bahalul Haque is a Junior Researcher at the Department of Software Engineering at LUT University. Earlier, he was a lecturer at the Department of Electrical and Computer Engineering, North South University. His works have been accepted and published in international conferences and peer-reviewed journals, including IEEE Access, Expert Systems, Cybernetics and Systems, various International conference proceedings, Tylor and Francis Books, and Springer Book. His research interests include Explainable AI, blockchain, data privacy and protection, and human-computer interaction.

A.K.M. Najmul Islam received the Ph.D. degree in information systems from the University of Turku, Finland, and the M.Sc. (Eng.) degree from the Tampere University of Technology, Finland. He is currently an Adjunct Professor at Tampere University, Finland. He is also an Associate Professor at LUT University, Finland. His research has been published in top outlets, such as European Journal of Information Systems, Information Systems Journal, Journal of Strategic Information Systems, Technological Forecasting and Social Change, Computers in Human Behavior, Internet Research, Computers & Education, Journal of Medical Internet Research, Information Technology & People, Telematics & Informatics, Journal of Retailing and Consumer Research, Communications of the AIS, Journal of Information Systems Education, AIS Transaction on Human-Computer Interaction, and Behaviour & Information Technology.

Patrick Mikalef is a Professor in Data Science and Information Systems at the Department of Computer Science. He has been a Marie Skłodowska-Curie post-doctoral research fellow working on "Competitive Advantage for the Datadriven Enterprise" (CADENT). He received his B.Sc. in Informatics from the Ionian University, his M.Sc. in Business Informatics for Utrecht University, and his Ph.D. in IT Strategy from the Ionian University. His research interests focus on the strategic use of data science and information systems in turbulent environments. He has published work in international conferences and peer reviewed journals, including the European Journal of Information Systems, British Journal of Management, Information and Management, and the European Journal of Operational Research.