IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 18 August 2023, accepted 3 September 2023, date of publication 14 September 2023, date of current version 6 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3315605

RESEARCH ARTICLE

A Conceptual Model Framework for XAI Requirement Elicitation of Application Domain System

MARIA ASLAM, DIANA SEGURA-VELANDIA, AND YEE MEY GOH
Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, LE11 3TU Loughborough, U.K.

Corresponding author: Diana Segura-Velandia (d.segura@lboro.ac.uk)

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/V062042/1.

ABSTRACT The use of data analytics and Machine Learning (ML) branches of AI for predictive and analytic knowledge retrieval has surged significantly in various industries (e.g., health, finance, business, and manufacturing). However, the acceptance of AI has been hindered by opaque models that lack transparency. Explainability in AI (XAI) has gained significant prominence owing to its focus on introducing avenues of accountability in AI. XAI acknowledges the importance of human factors and strives to incorporate them into the design process, recognising that the cognitive effort involved in understanding explanations is a key aspect. Mental Models play a crucial role in the XAI evaluative premise, but their current utility is limited. By intentionally designing explanations that align with users' mental models, their experiences can be significantly enhanced, leading to improved understanding, satisfaction, trust, and performance. This study proposes using Mental Models to elicit explainability requirements and to develop an Ontology-Driven Conceptual Model to facilitate the learning process for a better understanding of explanations.

INDEX TERMS Conceptual model, explainability in AI, mental models, requirements elicitation.

I. INTRODUCTION

Requirements elicitation (RE) is important in systems development. It helps to identify, gather, and define the needs and expectations of the system's stakeholders. There are limited studies in requirements elicitation for explanations, which is deemed conditional to successful use of XAI for key task performances [1]. This is an essential step, as by eliciting requirements from end-users, developers can gain insight into the types of explanations as well as explanative elements that are most useful and meaningful for the users. It requires consideration of paramount factors, such as the domain process, end-users, and constraints brought about by the data and ML tools.

Current research studies in RE for XAI focus on methods that encourage involving end-users in the design and development of XAI systems, such as participatory design and co-creation techniques [2]. Other methods involve the use of scenario-based design [3], [4], question banks [5], and capturing expert knowledge [6] for recognising the explanation needs.

Studies have explored the use of natural language processing techniques to extract requirements from textual data [7]. In addition, there is a growing interest in understanding the impact and value of different explanation formats, such as visual explanations or explanations that use analogies on end-users and how these explanations are used and interpreted [8], [9], [10], [11], [12]. Such knowledge base can significantly highlight end-user preferences for explanation presentation types.

These studies provide significant insight into the importance of RE for XAI and cite this through recommendations of methods, techniques and tools that can support RE process in the development of XAI systems.

However, there is a need for more research on benchmarking the RE process, which considers the cognitive activities that occur when explainability needs emerge.

Requirement Elicitation phenomena is deep-rooted and surfaces more prominently if studied through engagement with cognitive processes that are proactive but less studied.

The associate editor coordinating the review of this manuscript and approving it for publication was Zhaojun Steven Li.

Mental Models (MMs) are seen as cognitive structures that create an internal representation of the world based on personal experiences, perception, and understanding, in an effort to facilitate learning, reasoning, understanding, and decision making [13]. It is not far from reality to compare Mental Models to an interface used by humans to interact with external phenomena and comprehend its reality. The fact that MMs are likely to be imposed on reality [14] amplifies their significance manifold. The process of understanding is key for the construction of MMs and can significantly enhance their quality in terms of stability and accuracy.

Although the significance of MMs and Understanding is undeniable, their current usability solely as evaluative metrics for explanation products at the end of the XAI lifecycle limits their scope. Assessing users' MMs during the RE phase enables a more accurate representation of users' explanation needs. Moreover, emphasising the significance of understanding for the success of the explanatory process aligns it with the goal of facilitating user's understanding of knowledge-enriched explanations.

This study proposes two main research pathways: (i) the application of suitable methodologies to assess MMs for RE for user's explanations in an industrial context and, (ii) developing an advanced knowledge model for assessing MMs, resulting in aligned explanations for understanding, reasoning, and decision-making objectives.

In the context of this study, a Conceptual Model (CM) is posited as a more accurate and comprehensive representation of knowledge that is scientifically accepted or holds intuitive value for users [15]. The design of this model encompasses characteristics, properties, and qualities of explanations presented holistically through an Explanation Ontology. This ontology represents the explanation characteristics as factors that enable the properties and subsequently the qualities of explanations, with cognitive understanding being the focal point. Additionally, it centres around attainment of understanding, which is the ultimate explanation goal.

The main objectives of this study phase are:

i. To emphasise the utilisation of MMs and cognitive understanding as the central focus in Requirement Elicitation (RE) of explanations.
ii. To design and develop a comprehensive CM with a holistic ontology to generate stable and complete explanations that appeal to MMs through their characteristics, properties, and qualities.

These objectives collectively contribute to the creation of an Ontology-Driven Conceptual Model (ODCM) that facilitates Requirements Elicitation for eXplainable Artificial Intelligence (REXAI). By introducing the concept of aligning explanations with users' mental models, this paper presents a new and important approach to enhancing users' comprehension, trust, and acceptance of AI systems. This perspective on incorporating human cognitive requirements into AI design will overcome a fundamental challenge with the lack of transparency and understandability of AI models. The novel ODCM will enable XAI planning by presenting comprehensive options for the end-users to choose from to suit their explanation needs. This is non-trivial because the underpinning explanability needs is challenging because of the abstract nature of explanation concept, and there is no consistency in the literature. To the best of the author's knowledge, there are currently no existing XAI-based Conceptual Models available.

The remaining sections in this paper are organised as follows: Section II provides an in-depth analysis of the significance of Requirements Elicitation in user-centric explanatory process. Section III explores Mental Models, their connection to Understanding and their utility in RE. Section IV presents the theoretical foundations of the Conceptual Model design and the ontology formation process. Section 5 discusses the findings and conclusions drawn from the preceding sections, as well as potential avenues for future research in this field of work.

II. REQUIREMENTS ELICITATION FOR USER-CENTRIC EXPLAINABLE AI

A. RELATED WORK

Research in Requirements Elicitation in explainable AI has gained momentum due to an increase in applications in the field of XAI. It has become evident that explainability needs can be intrinsic and may go unnoticed if the detection methods only focus on basic requirements. As a result, current research areas in RE of XAI concentrate on several key areas, including, user-centric approaches, domain-specific RE, and cognitive modelling.

This paper does not provide an extensive review of the literature in these areas (please see the works of [3], [4], [6], [16], [17], [18], [19], [20], [21], and [22]). However, this paper does reference the methods used in user RE, which are detailed in Table 1 for reference.

User-centric approaches in Requirements Engineering involve various methods such as scenario-based, goal-based, question-bank enquiries, and among others, which aim to integrate users' mindsets into the design process [25].

Scenario-based approaches offer significant support in the design of XAI systems. By employing scenario building techniques, these approaches effectively identify the need for explanations based on domain specificity and inform the design process accordingly. In the context of RE, domain-specific scenarios play a key role in providing illustrations of user interactions within a particular domain process, which allows the development of tailored approaches and solutions that address the unique requirements and constraints of particular application areas. Previous research in domain-specific RE has provided valuable insights into tailoring approaches for various fields, including Software Systems, Information Systems, and Human-Agent Interactive Systems [1], [6].

By embracing user-centric approaches and leveraging domain-specific scenarios, XAI systems can be designed to deliver explanations that align closely with user expectations,

TABLE 1. Some methods that can be used to eliciting requirements in explainable AI systems.

Method	Study Design	References
Scenario-based RE	Scenarios of possible use in system development planning phase for XAI. Use of 'aging-in-place' and 'fraud detection' use-cases.	[3], [4]
Goal-based RE	Evaluation of explanation needs based on users' intended goals.	[23]
Question-bank	Identification of users' XAI needs through use of interviews based on question banks.	[5]
User role based RE	Requirements elicitation of XAI by means of key stakeholder identification and enquiry of their explanability needs.	[24]
Requirements Engineering for XAI	A work-in-progress RE framework based on domain specific knowledge.	[22]

leading to enhanced user experiences and improved system performance.

A pioneer study in scenario-based approach for explanations was conducted by [3]. This study tackled challenges related to uncertainty and domain limitations. Additionally, [4] conducted a study that emphasised the importance of context-awareness in XAI systems. Their use of a fraud-detection case study showcases the relevance of considering the specific context. To enhance the elicitation process, they incorporate insights from the work of [5] by engaging stakeholders through a systematic approach that involves questioning. Another research that has appeared recently tackles RE through the traditional Requirements Engineering route, in which a framework for generic user-centric requirements has been proposed [22].

Question banks play a crucial role in eliciting specific information from users, particularly in the context of XAI systems requirements definition. This is based on a structured approach that enables a comprehensive exploration of various dimensions of explainability related requirements.

A notable example of question bank is presented in the work of [5]. Their question bank focuses on addressing explainability needs in both local and global settings, considering explanations in the form of counterfactuals or example-based approaches. These methods aim at conducting question-oriented sessions with users, leveraging their cognitive intuition to enhance the RE process.

Inspired by prior research, [26] conducts a study to develop RE techniques tailored for Generative AI. By building on existing studies, they aim to incorporate effective methods and strategies for gathering requirements in this domain, addressing its unique challenges appropriately.

While considered an older approach, the use of pre-defined explanation goals as proposed by [23], remains a valuable strategy in the field of REXAI. This study suggests employing the goal-based technique for the evaluation of explanations. Utilising pre-defined goals of explanation offers several advantages including: i) enabling end-users and developers of XAI to recognise constraints that may impact goal completion, ii) assisting XAI developers in selecting appropriate methods of XAI product development that align with the explainability needs of recognised goals, iii) streamlining the XAI process cycle to focus on the identified goals.

While approaches like scenario-based, goal-based, and question bank methods aim to develop a domain-centric focus, their effectiveness in achieving the desired levels of understanding, trust, and satisfaction remains uncertain due to the lack of implementation and validation.

There is an interest in leveraging cognitive models, such as MMs, to better understand how humans perceive, reason, and make decisions. By incorporating these models into the RE process, researchers aim to develop explanations that align with users' cognitive processes, enhancing the effectiveness and usability of AI systems [6], [13], [27], [28]. These models offer promising prospects of capturing and representing the cognitive processes involved in explainability needs in human end-users. However, the actual implementation of these recommendations is currently lacking.

Therefore, it is crucial and timely to address the gap between theory and practice by exploring and implementing methods of requirements enquiry that account for cognitive aspects. This can allow for a more comprehensive understanding of user's MMs, enabling better design and development of XAI systems that truly meet the desired levels of understanding, trust, and satisfaction. This research aims to bridge this gap to realise maximum achievable potential of cognitive-focused approaches in REXAI.

B. COGNITIVE MODELLING

System designers frequently make assumptions about user needs in the absence of prior knowledge of requirements engineering [26]. This approach often excludes users from the decision-making process during system planning, neglecting the importance of their input. Even when users are included, the techniques used may not adequately capture their most critical potential needs. Furthermore, cognitive factors such as understanding, trust, and satisfaction can lead users to misjudge their own learning requirements [29]. Therefore, an effective RE process should prioritise the needs of users' MMs, which are essential for a correct understanding of external phenomena.

MMs continuously evolve and improve as new information becomes available [15]. However, in cases where complex systems are involved, flawed understanding may lead to formulation of a flawed MM. end-users may readily accept such an MM as they may over-estimate their own knowledge depth [6].

Laird, a prominent researcher in MMs, suggests that humans understand the world by constructing simplified working models in their minds during the early stages of discovery. These models are incomplete and based on limited personal perceptions, serving as imitations of the real world and reflecting uncertain knowledge of how the actual counterparts in the real world operate.

C. MENTAL MODEL UTILITY IN XAI

In the XAI context, a MM refers to the cognitive representation or conceptual framework that users develop to understand and make sense of the external world including AI systems. MMs enhance the utility of XAI by enabling demand for tailored explanations, enabling user understanding, fostering transparency and trust in AI systems, supporting reasoning and decision-making and facilitating human-AI interaction.

Explanations in the XAI domain can be evaluated through the measurement of a user's MM [30], [31], [32]. This evaluation involves estimating metrics such as understanding, trust, satisfaction, response time, and others.

In typical AI scenarios, users who only observe the output of a machine learning-based AI system may draw inaccurate or incomplete conclusions due to inherent incompleteness and instability. Augmenting the output with an explanation, clarifying how and why the AI system generated the results, enhances the completeness and stability of the user's MM. This, in turn, facilitates correct decision making and restores trust in the AI system's output [33], as shown in Figure 1, where stability instils confidence in decision making processes.

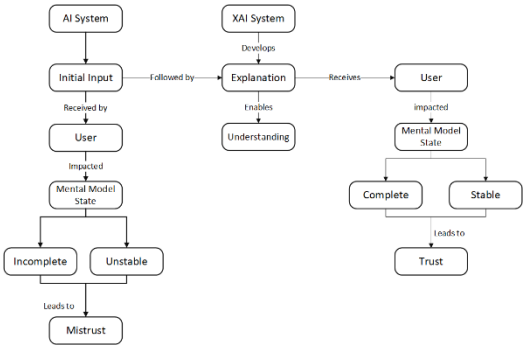


FIGURE 1. Conceptual Model of Explanation process [34].

The abstract nature of understanding a concept can make it challenging for users to express what, how, or how well they understand something, often leading to misguided judgments regarding their level of understanding. To address this issue, a support system can serve as a guide, helping users assess their understanding of an event or a system, along with the underlying knowledge and domain principles involved. Such a support system is also beneficial for assessing MMs, providing a clear overview of factors that can enhance users' understanding of an event or a system. In simpler terms, it enables targeted enquiry into which explanations, at what time in the process, and in what manner can improve the understanding of XAI requirements for both users and designers.

D. SIGNIFICANCE OF UNDERSTANDING

Understanding is paramount to the success of the explainability experience. Explanations are essential for advancing understanding, as they serve as an intellectual goal for explanation process [36]. Given the significant influence of understanding in constructing of accurate MMs, ensuring the degree and correctness of understanding becomes imperative. Users often experience an illusion of understanding, which can be misleading and disrupt the stability of their MM. This, in turn, impacts other end-goal user experiences (EGUE), such as satisfaction, performance, and decision making. Hence, the importance of understanding as an end-goal is evident due to its profound impact on achieving other explainability goals.

While the importance of understanding in achieving satisfaction from explanations is widely acknowledged [21], [35], [36], there is limited research that specifically focuses on improving understanding as a primary end-goal. Despite the recognition of understanding as a critical factor, limited attention has been given to exploring and experimenting with methodologies and approaches aimed at enhancing the level of understanding in the context of explanations. Further investigation and empirical studies are needed to develop specific techniques and strategies that improve understanding as a central goal in explainability. Bridging this research gap will lead to more impactful and satisfying explainable AI experiences, advancing the field.

Significance of understanding and its far-reaching impacts in enhancing mental models is being adopted in studies that deal with technical premise in XAI as well. Understanding of system's logic [37] and explanations aimed at multiple user groups [24] to facilitate understanding at a much larger scale are research themes with growing interest.

E. RE THROUGH COGNITIVE MODELLING

The field of cognitive modelling for RE is evolving. Two main research directions include human-AI interaction during the elicitation process and cognitive support tools to assist users in the RE process.

The first research direction entails examining the cognitive processes through which users develop MMs of AI systems. This involves investigating the factors that contribute to the formation of these MMs, such as the users' prior knowledge, experiences, and interactions with the AI system. Additionally, researchers analyse how AI-generated explanations impact users' mental models, identifying the ways in which these explanations shape users' understanding, perceptions, and expectations. Furthermore, researchers explore how users evaluate these explanations and how their assessments are influenced by their existing mental models. By studying these aspects, a deeper understanding can be gained regarding the intricate interplay between users' mental models, AI-generated explanations, and the evaluation process.

III. CONCEPTUAL MODELLING

In comparison to mental models, conceptual models are accurate and complete representations of scientifically accepted knowledge [15]. If developed with conceptual models, mental models can grow and transform into more accurate forms of

themselves. A Conceptual Modelling approach is a worthy candidate to consider for the modelling of users' mental models. This is a concept that is largely placed in the Social Sciences and Cognitive Psychology realms, but even within that, the advances in understanding are limited. Formally known as "the process of defining certain aspects of the physical and social worlds for enabling understanding and communication", the process of Conceptual Modelling aims to represent the conceptual version of the domain system [38]. The resultant is known as the Conceptual Model (CM).

In the context of Requirements Elicitation for eXplainable Artificial Intelligence (REXAI), a conceptual model represents essential characteristics, properties, and qualities of explanations. However, adopting conceptual modelling techniques for REXAI poses research challenges due to limited literature on the interdisciplinary approaches and advanced modelling techniques essential for their development.

Use of Conceptual Models for RE is seen frequently in the past as well as recent research work. The technique's utility in information systems engineering [38], categorisation during design phase [39], and education science [15] seen.

A. ONTOLOGY-DRIVEN REXAI

The integration of ontologies in XAI requirements engineering is a relevant area of investigation. Ontology-based techniques can be used to elicit, analyse, and represent XAI requirements, aligning them with stakeholder needs while facilitating the design and evaluation of XAI systems.

Ontologies are considered useful tools for representing factual relationships within a specific domain of interest [40]. They aid in various stages including planning [41], development [42], [43], deployment [40], and validation of a benchmark REXAI [44], [45], [46].

Researchers have explored the integration of ontologies in various technical and computational fields [39], including modelling languages [47], data management systems [48], and information systems [45], [49]. The utilisation of ontologies holds promise in creating a conceptual model that enhances users' understanding of the underlying domain. Ontologies are well suited for structuring the contents of conceptual models as they encompass concepts, their semantic relationships, and sequential order of their contents [47].

Ontology-Driven Conceptual Modelling (ODCM) focuses on leveraging ontologies to enhance the development and application of conceptual models in various domains. It provides a structured and systematic approach to capture, organise, and represent domain knowledge in the form of an ontology-based conceptual model.

MMs are easily impacted by simplified models as they are easily understandable [50]. The ODCM provides a simplified model for RE that end-users find simple to understand and look through to choose from the available options of explanations.

Differences exist in theories and functions of ontology for conceptual modelling in way of attributes, properties, and relation representation amongst various studies [51]. The field of ontology for conceptual modelling encompasses various theories and functions, and there are differences among studies regarding the representation of attributes, properties, and relations [51]. Recent perspectives on Bunge's 1977 ontology, criticise its focus on concrete objects and their attributes, suggesting that it lacks consideration of human perception of these objects [51], [52]. This limitation hinders the representation of a human-centric version of a conceptual domain. Consequently, the true essence of conceptual modelling, which aims to represent human understanding of a body of knowledge, may be inadequately captured [53].

By using ontologies, conceptual modelling can benefit in several ways:

i. Ontologies provide a structured framework for modelling semantics, enabling a clear and precise depiction of relationships between entities. The identification of entities and their relationship within a CM can ensure accurate representation of the domain.
ii. Existing theories of ontology can help in the selection of appropriate grammar or semantics for conceptual models, ensuring that the domain is effectively represented. This helps to avoid ambiguous semantics that can obscure the understanding of the domain [45].
iii. Using ontologies in the development of a conceptual model supports a validated and reliable construction process. The application of ontology theories facilitates structuring and classification of different phenomena within a certain domain, aligned with aims of this research.

In the context of XAI, ontologies have been used for identifying design patterns to define explanations [54], defining the relationship between different explanation attributes [55], matching XAI solutions to appropriate explanation types and AI systems [56] and developing guidance for the realisation of requirements elicitation [57].

ODCM for REXAI involves the use of entities that play a role in capturing and representing the conceptual model based on ontological principles. These entities encompass concepts, characteristics, relationships, constraints, hierarchies, and instances, each playing a significant role in the modelling process. These are discussed as follows.

1) ENTITIES

A large body of XAI related literature reports several metrics for measuring explanations. The differences in methods of implementation helps in the identification of how metrics evaluate an explanation product [33], [58] explanation methods [59], and explanative properties of ML models [59]. This research is concerned with the first category.

At the end of an explanation process, an XAI product is revealed, expected to possess certain characteristics that help users achieve explainability goals, resulting in a useful experience with the XAI system [36], [60], [61]. The information conveyed by this product may vary depending on the

context and nature of the underlying event, as dictated by explainability goals. Explanation products can have unique characteristics, properties, qualities, and impacts based on their functions and the methods of measuring each may vary. Some are measured using computational methods referred to as objective metrics [33], [62], [63]; while others rely on subjective methods based on users' feedback [33]. Though various metrics are mentioned in the literature, only a few have been implemented to yield experimental results [64].

This study surveyed various XAI metrics by considering 42 research publications. The search terms used were 'evaluation of explanations in AI' or 'XAI evaluation'. Out of these, 13 studies provided comprehensive lists of methods for measuring metrics [33], [37], dependencies between metrics [21], and a survey of their practical usage [64]. Additionally, other studies focused on experimentally implementing individual or small groups of metrics.

In this study, a focused examination was conducted on 64 distinct terms used for metrics that measure explanations, based on the literature gathered. It should be noted that these metrics do not cover the evaluation of explainability methods (e.g. SHAP, LIME) or the metrics that measure the degree of explainability of ML models (e.g., Neural Networks, Random Forests). While these factors are indeed relevant, the research scope is limited to metrics related to explanation products. The presence of synonyms, antonyms, methodological similarities, or differences among these metrics can pose challenges for researchers when implementing evaluative methods for their XAI systems.

The selected metrics play a valuable role in the development of an ontology for conceptual modelling in XAI. By incorporating these metrics, a comprehensive model of the characteristics, properties, and qualities of explanation can be established for an underlying domain process. Utilising evaluative-metrics-based conceptual models in REXAI serves multiple purposes. Firstly, it enables the assessment of a user's mental model through the recommended conceptual model, benefiting both the end users and XAI designers by ensuring a thorough understanding of all possible explanations that can or cannot be provided to end-users. Secondly, it enhances learning, understanding, and problem-solving within users' mental models, fostering improved trust in the underlying ML models, and embracing the concept of human-in-the-loop early in the REXAI process. Lastly, this conceptual model facilitates the development of domain-specific explanations that are attuned to user needs and domain constraints.

2) DESIGN

In the field of ontology design, defining concepts that exist in a specific time and space is relatively straightforward if their spatially relative nature can be described. However, it becomes more challenging to define concepts that are quantitative, qualitative, hybrid, or abstract in nature. Distinguishing the qualitative nature of a single instance from its quantitative aspects is not a simple task [42]. This research does not attempt to classify these natures, but rather explores both and presents one that best represents the ODCM approach.

The design and construct of ODCM is based on recognising enabling patterns amongst metrics, forming the basis for their classification. Metrics displaying a unitary function, or a collection of simpler functions are classified as Characteristics (Char) of explanations. These explanation characteristics are grouped within Properties (Prop) of explanations based on their behaviours and tendency to have similar goals or outcomes, which are then further grouped under Quality (Qual) of explanations, encompassing a larger scope but similar criteria. The quality of an explanation is crucial for facilitating understanding, which is central to the ODCM and critical for the XAI experience (Section 2.2.2). Understanding, in turn, further facilitates EGUE, including Trust, Satisfaction, and Performance in the preliminary ODCM design.

The classification of metrics into characteristics, properties, and qualities of explanations is based on the following criteria:

Definition: Various definitions for metrics as they appear in literature help identify significant similarities and differences for classification. Differences help disregard less likely definitions, while similarities further strengthen the attributes of each metric.

XAI Function: Each characteristic impacts explanations differently, but there may be similarity in end-user-goals. Characteristics with similar impacts on end-user goals can be grouped together.

Subjective/ Objective: Metrics measured subjectively rely on users' feedbacks, while objective metrics involve computational methods. This criterion determines spatial positioning of metrics in the ontology.

Relation to Other Metrics: The nature of relationship between metrics and their enablement trends is crucial for determining their spatial positioning.

Implementation: Understanding how metrics are implemented helps in studying their construct, function, and impacts. Objective metrics are computationally implemented, making their functions and impacts more concrete, while subjective metrics may vary in different cases.

3) DEVELOPMENT AND MANAGEMENT

Inheritance is a central feature of ontologies [47], and CMs that are based on such relationships are easily representable via an ontological structure.

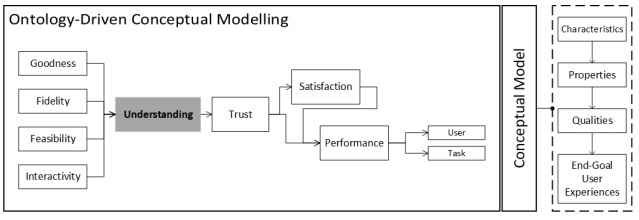


FIGURE 2. Structure of ODCM and resulting CM.

Based on the ontological architecture defined in [40], [47], and [65], the ontology structuring process involves the following steps:

1. Identification of Key terms: It is crucial to identify all key terms and establish a convention for their correct definition. This can include considering their literal meanings or defining them based on their objectives or impacts. These terms are treated as entities in the ODCM. Diagrammatic aids, such as E-R diagrams or relational schemas, are commonly used to enhance the understanding of entity definitions and their relationships with other entities. The goal is to achieve a comprehensive ontology that encompasses all necessary aspects of the CM.

2. Frequent terms: The frequency of metrics appearing in the literature serves as evidence of their significance in XAI evaluation. When redefined as ODCM entities, frequent terms should appear only once. Hence, the ODCM's definition must accurately reflect each entity's core objective in REXAI. Furthermore, for frequently appearing terms, it is essential to observe the nature of their appearances. Metrics in XAI, such as soundness or completeness may have different methods of definition and implementation. While all the differences must be studied, only the most appropriate of such entities should feature in the ontology.

3. Identification of synonyms: In XAI evaluation metrics studies, concepts may appear with different terms. Identifying such concepts is essential but their inclusion in the ontology is debatable. This study proposes their inclusion to provide users with alternative ways of identifying and defining an entity in a more understandable manner during REXAI processes. However, handling synonyms poses challenges. They can confuse designers, developers, and end-users with multiple taxonomic trends for the same concept. Additionally, omitting or deleting synonyms may limit taxonomic diversity since users and systems may have different semantic preferences. Including such terms in the ODCM is challenging but can prove useful.

B. ONTOLOGICAL RELATIONSHIPS

1) GENERIC RELATIONSHIPS

In ontologies,] 'related-to' semantics play a vital role in = defining generic associations as pointed out by [40]. These semantics establish a connection between two entities without specifying the exact nature of the relationship. In our research, the ODCM incorporates subjective entities, including understanding, trust, and satisfaction. To ensure a broad and generic scope of implementation, a certain degree of generalisation amongst entities is necessary.

2) HIERARCHICAL RELATIONSHIPS

In an 'is-a' hierarchy, entities have sub-concepts that act as substitutes every time the entity is invoked. In the case of ODCM, the relationship between *Char -> Prop -> Qual* is generally based on this principle.

3) RELATIONSHIP BETWEEN BASIC TERMS

Once all terms under different classifications are identified, the relationships between them are determined. In the ODCM, the relationship between recognised entities is elaborated in sections 1.1 and 1.2.

4) RELATIONSHIP BETWEEN ONTOLOGIES

Regarding the relationship between ontologies, they can be decomposed into sub-ontologies. Each sub-ontology is built independently and later combined with other sub-ontologies to form a comprehensive ODCM.

C. CONSTRUCT

The ODCM encompasses a comprehensive scope organised into multiple tiers. Each tier has been independently built with its own sub-ontology determining holistic relationships at single-tier-level. These sub-ontologies are combined to form the comprehensive ODCM, as shown in Figure 4.

The sub-ontologies and their key characteristics are outlined below:

1) UNDERSTANDING AND EGUE BLOCK

Understanding serves the central metric in this ontology, with trust, satisfaction, and performance identified as other essential end-goal metrics. While trust extends beyond understanding, it becomes relevant once initial understanding is achieved. These metrics are subjective and measured through qualitative methods. Trust-related metrics such as truthfulness, correctness, and transparency can be quantitatively computed, resulting in a hybrid metric measured through both subjective and objective methods.

2) GOODNESS

The quality of explanation, represented by goodness, aims to enhance the value of explanations. Goodness considers the scope of input data and perturbations of inputs that impact output. Key properties within this premise are broadness, which assesses the scope of the data representing the underlying domain, and perturbation-based properties, which identify input changes and their impact on output.

These properties allow end-users to understand data scope, the impact on delivering the output, and the appropriateness of the ML model being used during RE.

3) FIDELITY

Fidelity metrics evaluate how accurately an explanation represents the underlying domain system, ultimately building user trust in the XAI system. Fidelity quality includes properties and characteristics that foster trust in explanations. Demonstrating that predicted outcomes result from the correct selection of features by the models instills trust. Model-based interpretation, transparency elements, correctness assurance and soundness elements, and confidence-building elements are vital components embedded in explanations.

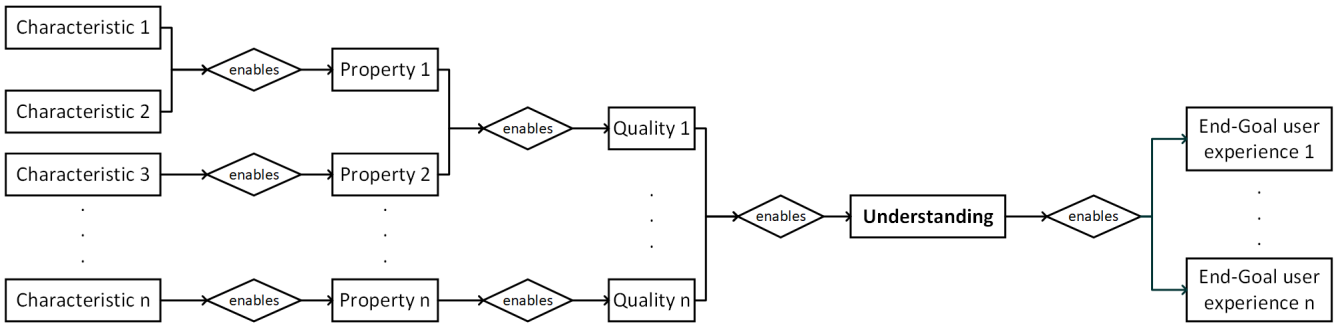


FIGURE 3. E-R diagram.

4) FEASIBILITY

The costs of running digital systems and the feasibility of transitioning towards digitalisation play significant roles in making decisions. Computational costs, especially with big data and complex black boxes can become a financial burden. Runtimes and additional XAI systems can further add to these costs. Having feasibility assessments of the XAI system with generated explanations allows users to weigh the advantages against costs and time constraints. Such explanations also facilitate debugging and selecting more feasible system options.

5) INTERACTIVITY

The sub-ontology on interactivity comprises metrics that enable users' interaction with the HXAI. These metrics define the characteristics mostly used in conversational systems [66], enabling users to adjust explanations to their needs [67]. Interactivity comes into play when users have advanced certainty about desired outcomes, acting as guides [67]. In HXAI, interactive features are used in recommendation systems, where users can correct assumptions [64] or enquire for additional information out of curiosity [22] use interactivity in HXAI. The nature of the relationship that exists in this sub-ontology is one of enablement.

6) CASE-BASED

There are a select number of elements that are recognised to be case-relevant. Their inclusion into the ODCM is necessary as it enhances the ontology with additional REXAI knowledge. End-users can demand details about each characteristic within this sub-ontology to understand their relevance to their specific domain systems. For instance, the chronology characteristic becomes mostly relevant when working with time-series data, where the timing of instances or features holds significance.

IV. DISCUSSION AND CONCLUSION

The field of XAI is characterised by a multitude of metrics, each with its own definitions, implementations methods, and measurement techniques. Despite this diversity, there are similarities in how these metrics are defined and calculated. One valuable clue lies in the measurement approach, which provides insights into the specific explanation elements being measured. Given the vast number of XAI metrics, a hierarchical design becomes necessary to regulate and prioritise what is essential for explanations based on end-users' requirements.

End-users across industries may have varying demands, but they generally seek explanations that help them understand events occurring in the underlying domain system. Hence, understanding is considered an essential end-goal, leaving users satisfied after interacting with the XAI system. Additionally, achieving understanding can trigger other critical goals for the XAI system accomplishment. For this research, trust, satisfaction, and performance were selected as key goals due to their recurrent mention in the literature. The attainment of understanding contributes to increased stability and completeness of mental models, achieved by filling knowledge gaps, answering questions, and instilling trust and satisfaction.

Conceptual models, along with the ODCM design strategy emerge as strong candidates to resolve these issues. By structuring conceptual models with inclusive hierarchies, explanatory design can be critically considered at a fundamental level. The qualities and properties of explanations, combined with unitary explanatory characteristics, enable a simplistic and primitive examination of concepts, breaking down the RE issues for end-users.

The hierarchical nature of the conceptual model provides a clear and concise platform for planning elements of explanations incrementally. It collects all possible options of explanation elements providing end-users with choices. Furthermore, the viewpoint provides a critical analysis of what can and cannot be included, considering data and model constraints. The self-explanatory and simplistic ODCM structure not only encourages end-users' active participation in the RE process but also enhances their understanding of the XAI system and its capabilities.

The ODCM structure of the conceptual model opens up several avenues for future progressive research in the REXAI premise, an extension of XAI planning. By connecting it with ontologies developed by studies mentioned in this paper, particularly in [55], [56], and [68], a comprehensive ontological

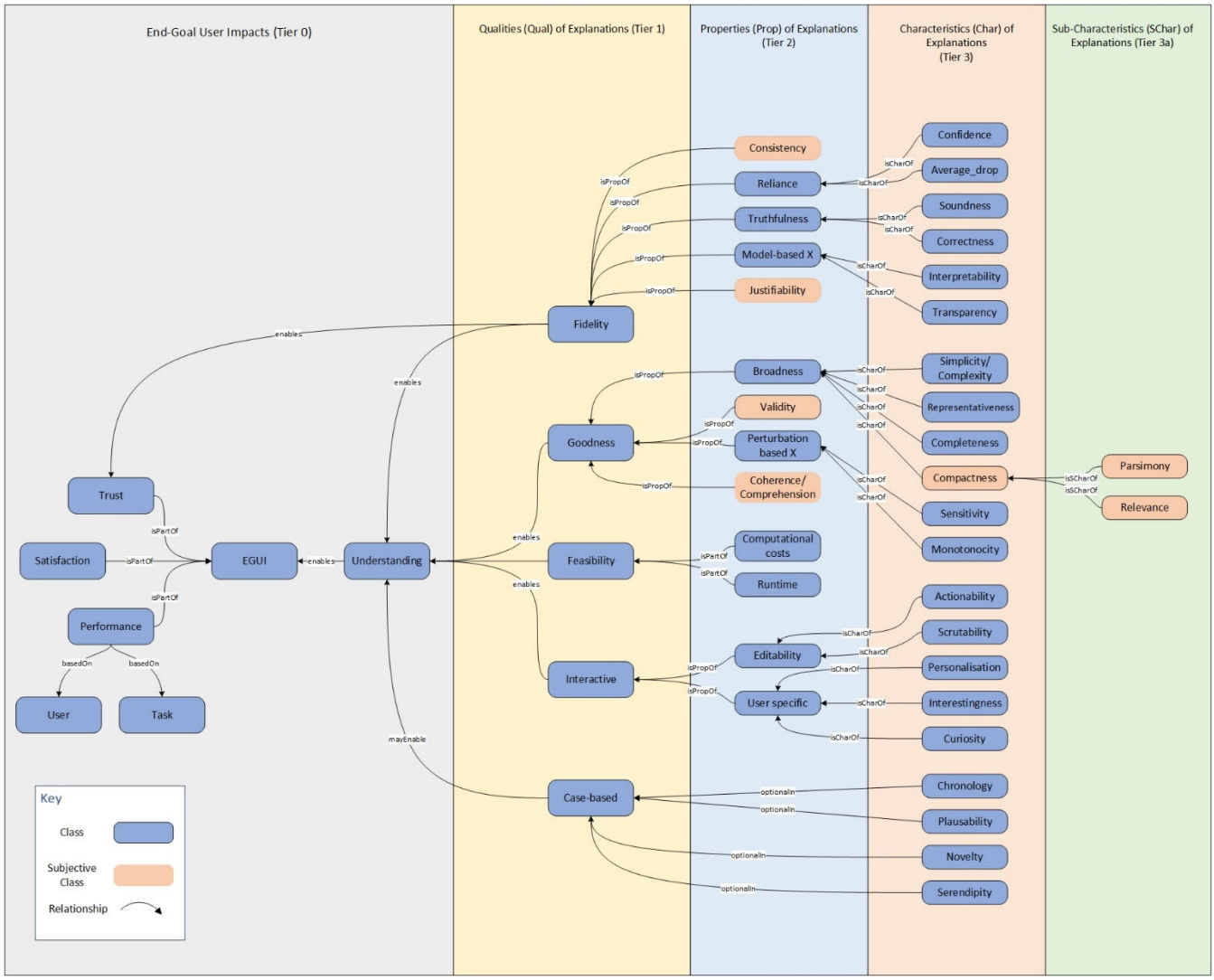


FIGURE 4. **Proposed ODCM for XAI.**

process of explanation lifecycle can be formulated. This includes considerations of feasibilities, parallel operations, co-dependencies, best practices, and monitoring and control elements, all interconnected to form a comprehensive XAI lifecycle.

The Conceptual Model developed as a result of the ODCM caters to the cognitive needs of end-users, enabling intrinsic cognitive activities including learning, reasoning, understanding, and decision-making.

ODCM presents significant potential as a future research subject, particularly concerning requirements elicitation, which exclusively aims at mental model stability and completeness. Furthermore, the elements of the conceptual model, serving as a concept that encompasses a complete scope of available knowledge, sets a benchmark on which explanation planning can be based. Its ontological design allows for flexibility in incorporating additions and upgrades. The agnostic and generic nature of ODCM ensures wide implementation scope, adding value to both current research and future research studies.

The ontology can be extended to include the most suitable explanation formats that align with the characteristics in Tier 3 and 4. An additional benefit of using the ODCM design is its flexibility to complement other RE methods in XAI, such as scenario and goal based, as well as question bank methods. This versatility enables it to serve as a benchmark method for REXAI.

This research further endeavored to develop an ODCM as an XAI system planning tool, extending its application beyond requirements elicitation to other phases of the XAI system development, presentation, and evaluation. An API based on the ODCM, which empowers end-users to envision explanations that can be generated from data is under development. This API will facilitate both end-users and XAI tool developers in accurately capturing REXAI.

The simplistic and self-explanatory design of ODCM makes it inclusive of users with varying digital skill levels, providing them with unrestricted options for designing explanations. The concept of aligning explanations with user mental models proposed in this paper has the potential to enhance users' comprehension and acceptance of AI systems with future growth opportunities to include XAI requirements that may emerge with rapidly evolving practices. Furthermore, its integration into comprehensive ontologies of an end-to-end XAI lifecycle will provide a fundamental and tangible step forward in making AI systems more accessible.

ACKNOWLEDGMENT

The work reported in this paper was undertaken as part of the Made Smarter Innovation: Centre for People-Led Digitalisation, University of Bath, University of Nottingham, and Loughborough University.

REFERENCES

[1] C. Cheligeer, J. Huang, G. Wu, N. Bhuiyan, Y. Xu, and Y. Zeng, "Machine learning in requirements elicitation: A literature review," *Artif. Intell. Eng. Design, Anal. Manuf.*, vol. 36, p. e32, Mar. 2022, doi: 10.1017/s0890060422000166.

[2] U. Ehsan and M. O. Riedl, "Human-centered explainable AI: Towards a reflective sociotechnical approach," in *HCI International 2020—Late Breaking Papers: Multimodality and Intelligence* (Lecture Notes in Computer Science), vol. 12424, C. Stephanidis, M. Kurosu, H. Degen, and L. Reinerman-Jones, Eds. Cham, Switzerland: Springer, 2020, doi: 10.1007/978-3-030-60117-1_33.

[3] C. T. Wolf, "Explainability scenarios: Towards scenario-based XAI design," in *Proc. 24th Int. Conf. Intell. User Interfaces*, Mar. 2019, pp. 252–257, doi: 10.1145/3301275.3302317.

[4] D. Cirqueira, D. Nedbal, M. Helfert, and M. Bezbradica, "Scenario-based requirements elicitation for user-centric explainable AI," in *Machine Learning and Knowledge Extraction* (Lecture Notes in Computer Science), vol. 12279, A. Holzinger, P. Kieseberg, A. Tjoa, and E. Weippl, Eds. Cham, Switzerland: Springer, 2020, doi: 10.1007/978-3-030-57321-8_18.

[5] Q. V. Liao, D. Gruen, and S. Miller, "Questioning the AI: Informing design practices for explainable AI user experiences," in *Proc. CHI Conf. Human Factors Comput. Syst.*, Apr. 2020, pp. 1–15, doi: 10.1145/3313831.3376590.

[6] N. Clewley, L. Dodd, V. Smy, A. Witheridge, and P. Louvieris, "Eliciting expert knowledge to inform training design," in *Proc. 31st Eur. Conf. Cognit. Ergonom.*, Sep. 2019, pp. 138–143, doi: 10.1145/3335082.3335091.

[7] E. Cambria, L. Malandri, F. Mercorio, M. Mezzanzanica, and N. Nobani, "A survey on XAI and natural language explanations," *Inf. Process. Manage.*, vol. 60, no. 1, Jan. 2023, Art. no. 103111, doi: 10.1016/j.ipm.2022.103111.

[8] D. H. Kim, E. Hoque, and M. Agrawala, "Answering questions about charts and generating visual explanations," in *Proc. CHI Conf. Human Factors Comput. Syst.*, Apr. 2020, pp. 1–13, doi: 10.1145/3313831.3376467.

[9] G. Alicioglu and B. Sun, "A survey of visual analytics for explainable artificial intelligence methods," *Comput. Graph.*, vol. 102, pp. 502–520, Feb. 2022, doi: 10.1016/j.cag.2021.09.002.

[10] G. Ras, M. van Gerven, and P. Haselager, "Explanation methods in deep learning: Users, values, concerns and challenges," in *Explainable and Interpretable Models in Computer Vision and Machine Learning* (The Springer Series on Challenges in Machine Learning). Cham, Switzerland: Springer, 2018, doi: 10.1007/978-3-319-98131-4_2.

[11] H. Han, R. Faust, B. Felipe K. Norambuena, R. Prabhu, T. Smith, S. Li, and C. North, "Explainable interactive projections for image data," in *Advances in Visual Computing* (Lecture Notes in Computer Science), vol. 13598. Cham, Switzerland: Springer, 2022.

[12] A. Heimerl, K. Weitz, T. Baur, and E. André, "Unraveling ML models of emotion with NOVA: Multi-level explainable AI for non-experts," *IEEE Trans. Affect. Comput.*, vol. 13, no. 3, pp. 1155–1167, Jul. 2022, doi: 10.1109/TAFFC.2020.3043603.

[13] N. A. Jones, H. Ross, T. Lynam, P. Perez, and A. Leitch, "Mental models: An interdisciplinary synthesis of theory and methods," *Ecology Soc.*, vol. 16, no. 1, pp. 1–14, 2011, doi: 10.5751/ES-03802-160146.

[14] L. Westbrook, "Mental models: A theoretical overview and preliminary study," *J. Inf. Sci.*, vol. 32, no. 6, pp. 563–579, Dec. 2006, doi: 10.1177/0165551506068134.

[15] I. M. Greca and M. A. Moreira, "Mental models, conceptual models, and modelling," *Int. J. Sci. Educ.*, vol. 22, no. 1, pp. 1–11, Jan. 2000, doi: 10.1080/095006900289976.

[16] A. Dahiya and J. Kumar, "Direct user behavior data leads to better user centric thinking than role playing: An experimental study on HCI design thinking," in *Proc. Int. Conf. Human-Comput. Interact.*, vol. 1293. Cham, Switzerland: Springer, 2020, pp. 11–18.

[17] D. Wang, Q. Yang, A. Abdul, and B. Y. Lim, "Designing theory-driven user-centric explainable AI," in *Proc. CHI Conf. Human Factors Comput. Syst.*, May 2019, pp. 1–15, doi: 10.1145/3290605.3300831.

[18] M. Le Guillou, L. Prévot, and B. Berberian, "Bringing together ergonomic concepts and cognitive mechanisms for human-AI agents cooperation," *Int. J. Hum. Comput. Interact.*, vol. 39, no. 9, pp. 1827–1840, 2022, doi: 10.1080/10447318.2022.2129741.

[19] A. Chander and R. Srinivasan, "Evaluating explanations by cognitive value," in *Machine Learning and Knowledge Extraction* (Lecture Notes in Computer Science), vol. 11015, A. Holzinger, P. Kieseberg, A. Tjoa, and E. Weippl, Eds. Cham, Switzerland: Springer, 2018, doi: 10.1007/978-3-319-99740-7_23.

[20] M. Westberg and K. Främling, "Cognitive perspectives on context-based decisions and explanations," 2021, *arXiv:2101.10179*.

[21] J. H.-W. Hsiao, H. H. T. Ngai, L. Qiu, Y. Yang, and C. C. Cao, "Roadmap of designing cognitive metrics for explainable artificial intelligence (XAI)," 2021, *arXiv:2108.01737*.

[22] U.-E. Habiba, J. Bogner, and S. Wagner, "Can requirements engineering support explainable artificial intelligence? Towards a user-centric approach for explainability requirements," in *Proc. IEEE 30th Int. Requirement Eng. Conf. Workshops (REW)*, 2022, pp. 162–165, doi: 10.1109/rew56159.2022.00038.

[23] D. B. Leake, "Goal-based explanation evaluation," *Cogn. Sci.*, vol. 15, no. 4, pp. 509–545, 1991, doi: 10.1016/0364-0213(91)80017-Y.

[24] M. Hall, D. Harborne, R. Tomsett, and V. Galetic, "A systematic method to understand requirements for explainable AI (XAI) systems," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, 2019, pp. 2–3.

[25] D. Buschek, M. Eiband, and H. Hussmann, "How to support users in understanding intelligent systems? An analysis and conceptual framework of user questions considering user mindsets, involvement, and knowledge outcomes," *ACM Trans. Interact. Intell. Syst.*, vol. 12, no. 4, pp. 1–27, Dec. 2022, doi: 10.1145/3519264.

[26] J. Sun, Q. V. Liao, M. Müller, M. Agarwal, S. Houde, K. Talamadupula, and J. D. Weisz, "Investigating explainability of generative AI for code through scenario-based design," in *Proc. 27th Int. Conf. Intell. User Interface*, Mar. 2022, pp. 212–228, doi: 10.1145/3490099.3511119.

[27] M. Anders, M. Obaidi, B. Paech, and K. Schneider, "A study on the mental models of users concerning existing software," in *Requirements Engineering: Foundation for Software Quality* (Lecture Notes in Computer Science), vol. 13216. Cham, Switzerland: Springer, 2022.

[28] N. M. Yusoff and S. S. Salim, "Shared mental model processing in visualization technologies: A review of fundamental concepts and a guide to future research in human-computer interaction," in *Engineering Psychology and Cognitive Ergonomics. Mental Workload, Human Physiology, and Human Energy* (LNAI), vol. 12186. Cham, Switzerland: Springer, 2020.

[29] M. Merry, P. Riddle, and J. Warren, "A mental models approach for defining explainable artificial intelligence," *BMC Med. Informat. Decis. Making*, vol. 21, no. 1, pp. 1–12, Dec. 2021, doi: 10.1186/s12911-021-01703-7.

[30] T. Kulesza, S. Stumpf, M. Burnett, S. Yang, I. Kwan, and W.-K. Wong, "Too much, too little, or just right? Ways explanations impact end users' mental models," in *Proc. IEEE Symp. Vis. Lang. Human Centric Comput.*, Sep. 2013, pp. 3–10, doi: 10.1109/VLHCC.2013.6645235.

[31] M. M. A. De Graaf and B. F. Malle, "How people explain action (and autonomous intelligent systems should too)," in *Proc. AAAI Fall Symp. Ser.*, 2017, pp. 19–26.

[32] T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," *Artif. Intell.*, vol. 267, pp. 1–38, Feb. 2019, doi: 10.1016/j.artint.2018.07.007.

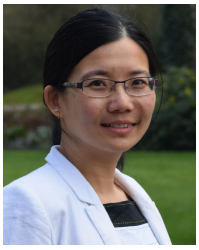
[33] R. R. Hoffman, S. T. Mueller, G. Klein, and J. Litman, "Metrics for explainable AI: Challenges and prospects," 2018, *arXiv:1812.04608*.

[34] S. T. Mueller, R. R. Hoffman, W. Clancey, A. Emrey, and G. Klein, "Explanation in human-AI systems: A literature meta-review," Defense Adv. Res. Projects Agency, USA, Tech. Rep. TA-2_02/19, Feb. 2019.
[35] Z. Zhang, "User interface design based on human-centered explainable AI methods," Aalto Univ., Finland, Tech. Rep., 2022.
[36] M. Langer, D. Oster, T. Speith, H. Hermanns, L. Kästner, E. Schmidt, A. Sesing, and K. Baum, "What do we want from explainable artificial intelligence (XAI)?—A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research," *Artif. Intell.*, vol. 296, Jul. 2021, Art. no. 103473, doi: 10.1016/j.artint.2021.103473.
[37] S. Mohseni, N. Zarei, and E. D. Ragan, "A multidisciplinary survey and framework for design and evaluation of explainable AI systems," *ACM Trans. Interact. Intell. Syst.*, vol. 11, nos. 3–4, pp. 1–45, Dec. 2021, doi: 10.1145/3387166.
[38] V. A. Carvalho, J. P. A. Almeida, C. M. Fonseca, and G. Guizzardi, "Extending the foundations of ontology-based conceptual modeling with a multi-level theory," in *Conceptual Modeling* (Lecture Notes in Computer Science), vol. 9381, P. Johannesson, M. Lee, S. Liddle, A. Opdahl, and Ó. P. López, Eds. Cham, Switzerland: Springer, 2015, doi: 10.1007/978-3-319-25264-3_9.
[39] V. A. Carvalho, J. P. A. Almeida, C. M. Fonseca, and G. Guizzardi, "Multi-level ontology-based conceptual modeling," *Data Knowl. Eng.*, vol. 109, pp. 3–24, May 2017, doi: 10.1016/j.datak.2017.03.002.
[40] V. Sugumaran and V. C. Storey, "Ontologies for conceptual modeling: Their creation, use, and management," *Data Knowl. Eng.*, vol. 42, no. 3, pp. 251–271, 2002, doi: 10.1016/S0169-023X(02)00048-4.
[41] B. A. N. Cenka, H. B. Santoso, and K. Junus, "Personal learning environment toward lifelong learning: An ontology-driven conceptual model," *Interact. Learn. Environ.*, pp. 1–17, Feb. 2022, doi: 10.1080/10494820.2022.2039947.
[42] B. Henderson-Sellers, O. Eriksson, and P. J. ågerfalk, "On the need for identity in ontology-based conceptual modelling," in *Proc. Conf. Res. Pract. Inf. Technol. Ser.*, vol. 165, 2015, pp. 9–20.
[43] G. Guizzardi and V. Zamborlini, "Using a trope-based foundational ontology for bridging different areas of concern in ontology-driven conceptual modeling," *Sci. Comput. Program.*, vol. 96, pp. 417–443, Dec. 2014, doi: 10.1016/j.scico.2014.02.022.
[44] S. T. March and G. N. Allen, "Toward a social ontology for conceptual modeling," *Commun. Assoc. Inf. Syst.*, vol. 34, pp. 1347–1358, Jan. 2014, doi: 10.17705/1cais.03470.
[45] G. Shanks, E. Tansley, and R. Weber, "Using ontology to validate conceptual models," *Commun. ACM*, vol. 46, no. 10, pp. 85–89, Oct. 2003, doi: 10.1145/944217.944244.
[46] T. P. Sales and G. Guizzardi, "Ontological anti-patterns: Empirically uncovered error-prone structures in ontology-driven conceptual models," *Data Knowl. Eng.*, vol. 99, pp. 72–104, Sep. 2015, doi: 10.1016/j.datak.2015.06.004.
[47] M. Erdmann and R. Studer, "Ontologies as conceptual models for XML documents," *Proc. 12th Work. Knowl. Acquis. Model. Manag.*, pp. 1–19, 1999. [Online]. Available: http://ryk-kypc1.narod.ru/ERDMANN.pdf
[48] C. Jayapandian, C. H. Chen, A. Dabir, S. Lhatoo, G. Q. Zhang, and S. S. Sahoo, "Domain ontology as conceptual model for big data management: Application in biomedical informatics," in *Conceptual Modeling* (Lecture Notes in Computer Science), vol. 8824, E. Yu, G. Dobbie, M. Jarke, and S. Purao, Eds. Cham, Switzerland: Springer, 2014, doi: 10.1007/978-3-319-12206-9_12.
[49] R. Weber, "Conceptual modelling and ontology: Possibilities and pitfalls," in *Conceptual Modeling—ER 2002* (Lecture Notes in Computer Science), vol. 2503, S. Spaccapietra, S. T. March, and Y. Kambayashi, Eds. Berlin, Germany: Springer, 2002, doi: 10.1007/3-540-45816-6_1.
[50] C. Kent, M. A. Chaudhry, M. Cukurova, I. Bashir, H. Pickard, C. Jenkins, B. D. Boulay, A. Moeini, and R. Luckin, "Machine learning models and their development process as learning affordances for humans," in *Artificial Intelligence in Education* (Lecture Notes in Computer Science) vol. 12748. Cham, Switzerland: Springer, 2021.
[51] G. Guizzardi, C. Masolo, and S. Borgo, "In defense of a trope-based ontology for conceptual modeling: An example with the foundations of attributes, weak entities and datatypes," in *Conceptual Modeling—ER 2006* (Lecture Notes in Computer Science) vol. 4215, D. W. Embley, A. Olivé, and S. Ram, Eds. Berlin, Germany: Springer, 2006, doi: 10.1007/11901181_10.
[52] G. Allen and S. T. March, "A critical assessment of the Bunge-Wand-Weber ontology for conceptual modeling gove," in *Proc. 16th Annu. Workshop Inf. Technol. Syst. (WITS) Paper*, Jul. 2006, pp. 1–6.
[53] J. P. McCusker, J. Luciano, and D. L. McGuinness, "Towards an ontology for conceptual modeling," in *Proc. Workshop (CEUR)*, vol. 833, May 2014, pp. 191–199.
[54] I. Tiddi, M. d'Aquin, and E. Motta, "An ontology design pattern to define explanations," in *Proc. 8th Int. Conf. Knowl. Capture*, Oct. 2015, pp. 1–9, doi: 10.1145/2815833.2815844.
[55] S. Chari, O. Seneviratne, D. M. Gruen, M. A. Foreman, A. K. Das, and D. L. McGuinness, "Explanation ontology: A model of explanations for user-centered AI," 2020, *arXiv:2010.01479*.
[56] E. Wenink, J. Van Der Waa, and S. Raaijmakers, "Towards FAIR explainable AI: A standardized ontology for mapping XAI solutions to use cases, explanations, and AI systems," *Proc. Int. Conf. PErvasive Technol. Rel. Assistive Environ. (PETRA)*, 2022, pp. 562–568, doi: 10.1145/3529190.3535693.
[57] S. Farfeleder, T. Moser, A. Krall, T. St?alhane, I. Omoronyia, and H. Zojer, "Ontology-driven guidance for requirements elicitation," in *The Semanic Web: Research and Applications* (Lecture Notes in Computer Science), vol. 6643. 2011, pp. 212–226, doi: 10.1007/978-3-642-21064-8_15.
[58] P. Lopes, E. Silva, C. Braga, T. Oliveira, and L. Rosado, "XAI systems evaluation: A review of human and computer-centred methods," *Appl. Sci.*, vol. 2, no. 12, pp. 1–31, 2022, doi: 10.3390/app12199423.
[59] L. Arras, A. Osman, and W. Samek, "CLEVR-XAI: A benchmark dataset for the ground truth evaluation of neural network explanations," *Inf. Fusion*, vol. 81, pp. 14–40, May 2022, doi: 10.1016/j.inffus.2021.11.008.
[60] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, Jun. 2020, doi: 10.1016/j.inffus.2019.12.012.
[61] S. Laato, M. Tiainen, A. K. M. N. Islam, and M. Mäntymäki, "How to explain AI systems to end users: A systematic literature review and research agenda," *Internet Res.*, vol. 32, no. 7, pp. 1–31, Dec. 2022, doi: 10.1108/INTR-08-2021-0600.
[62] J. Zhou, A. H. Gandomi, F. Chen, and A. Holzinger, "Evaluating the quality of machine learning explanations: A survey on methods and metrics," *Electron.*, vol. 10, no. 5, pp. 1–19, Mar. 2021, doi: 10.3390/ELECTRONICS10050593.
[63] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," 2017, *arXiv:1702.08608*.
[64] L. Coroama and A. Groza, "Evaluation metrics in explainable artificial intelligence (XAI)," in *ARTIIS* (Communications in Computer and Information Science). 2022, pp. 401–413, doi: 10.1007/978-3-031-20319-0.
[65] E. Christopoulou, C. Goumopoulos, I. Zaharakis and A. Kameas, "An ontology-based conceptual model for composing context-aware applications," Res. Academic Comput. Technol. Inst., Greece, 2004.
[66] K. Sokol and P. Flach, "Explainability fact sheets," in *Proc. Conf. Fairness, Accountability, Transparency*, 2020, pp. 56–67, doi: 10.1145/3351095.3372870.
[67] J. M. Rožanec, B. Fortuna, and D. Mladenič, "Knowledge graph-based rich and confidentiality preserving explainable artificial intelligence (XAI)," *Inf. Fusion*, vol. 81, pp. 91–102, May 2022, doi: 10.1016/j.inffus.2021.11.015.
[68] C. Panigutti, A. Perotti, and D. Pedreschi, "Doctor XAI An ontology-based approach to black-box sequential data classification explanations," *Proc. Conf. Fairness, Accountability, Transpar.*, 2020, pp. 629–639, doi: 10.1145/3351095.3372855.

MARIA ASLAM received the B.Sc. degree in computer science and the M.Sc. degree in big data science and technology from the University of Bradford, U.K., in 2012 and 2019, respectively. She is currently pursuing the Ph.D. degree with Loughborough University, U.K., with a focus on explainability in artificial intelligence and its utility and implementation in manufacturing industry.

DIANA SEGURA-VELANDIA received the bachelor's degree in chemical engineering from the University of Los Andes, Bogotá, Colombia, and the Ph.D. degree in artificial intelligence in product formulation from Loughborough University, U.K.

With a rich academic background, she has harnessed her expertise to apply cutting-edge artificial intelligence techniques across diverse domains, including electronics manufacturing, healthcare, and automotive industries. Her notable achievements include devising innovative AI methods for real-time monitoring, defect prediction, and design optimization in electronics manufacturing. Her interdisciplinary work has extended to the healthcare sector, where she has contributed to the development of AI-driven solutions for improved patient care. In the automotive domain, her focus on asset traceability and tracking has led to advancements in supply chain management and efficiency. She is currently a Lecturer of ICT for manufacturing with the Department of Mechanical, Manufacturing, and Electrical Engineering, Loughborough University, she continues to drive progress in her field. Her research interests include span artificial intelligence, decision-making, advanced materials, sensor technologies, and dynamic modeling and simulation.



YEE MEY GOH received the B.Eng. degree (Hons.) in mechanical engineering from Universiti Tenaga Nasional, Malaysia, in 2001, and the Ph.D. degree from the University of Bristol, U.K., in 2005.

She founded the Digital Automation Systems Design Laboratory, Intelligent Automation Centre, focusing on modeling and understanding human factors and knowledge that will enable inclusive and responsible development of industrial technologies, such as robotics and AI. She is currently a Reader of transdisciplinary digital manufacturing with Loughborough University. She is the Co-Director of the Made Smarter Innovation Centre on People-Led Digitalisation, working with the University of Bath and the University of Nottingham and a range of industries to improve technology adoption through a people-led approach to digitalization.

Dr. Goh is a fellow of the Higher Education Academy (FHEA) and a member of the Design Society (MDS) and the International Society of Transdisciplinary Engineering (ISTE). She serves on the Committee of the Consortium of UK University Manufacturing and Engineering (COMEH).