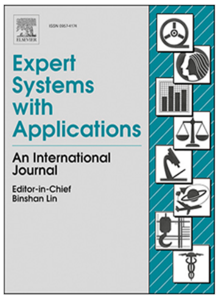
A practical exploration of the convergence of case-based reasoning and explainable artificial intelligence.

PRADEEP, P., CARO-MARTÍNEZ, M. and WIJEKOON, A.

2024

Journal Pre-proof



A practical exploration of the convergence of Case-Based Reasoning and Explainable Artificial Intelligence

Preeja Pradeep, Marta Caro-Martínez, Anjana Wijekoon

PII: S0957-4174(24)01600-2
DOI: https://doi.org/10.1016/j.eswa.2024.124733
Reference: ESWA 124733

To appear in: *Expert Systems With Applications*

Received date : 11 March 2024
Revised date : 14 June 2024
Accepted date : 5 July 2024

Please cite this article as: P. Pradeep, M. Caro-Martínez and A. Wijekoon, A practical exploration of the convergence of Case-Based Reasoning and Explainable Artificial Intelligence. *Expert Systems With Applications* (2024), doi: https://doi.org/10.1016/j.eswa.2024.124733.

A Practical Exploration of the Convergence of Case-Based Reasoning and Explainable Artificial Intelligence

Preeja Pradeep [a], Marta Caro-Martínez [b,*], Anjana Wijekoon [c]

[a] *Insight Centre for Data Analytics, School of Computer Science and IT, University College Cork, Ireland*
[b] *Facultad de Informatica Universidad Complutense de Madrid Spain*
[c] *School of Computing Robert Gordon University Aberdeen Scotland*

Abstract

As Artificial Intelligence (AI) systems become increasingly complex, ensuring their decisions are transparent and understandable to users has become paramount. This paper explores the integration of Case-Based Reasoning (CBR) with Explainable Artificial Intelligence (XAI) through a real-world example, which presents an innovative CBR-driven XAI platform. This study investigates how CBR, a method that solves new problems based on the solutions of similar past problems, can be harnessed to enhance the explainability of AI systems. Though the literature has few works on the synergy between CBR and XAI, exploring the principles for developing a CBR-driven XAI platform is necessary. This exploration outlines the key features and functionalities, examines the alignment of CBR principles with XAI goals to make AI reasoning more transparent to users, and discusses methodological strategies for integrating CBR into XAI frameworks. Through a case study of our CBR-driven XAI platform, iSee: Intelligent Sharing of Explanation Experience, we demonstrate the practical application of these principles, highlighting the enhancement of system transparency and user trust. The platform elucidates the decision-making processes of AI models and adapts to provide explanations tailored to diverse user needs. Our findings emphasize the importance of interdisciplinary approaches in AI research and the significant

*Corresponding Author
Email addresses: ppradeep@ucc.ie (Preeja Pradeep), martcaro@ucm.es (Marta Caro-Martínez), a.wijekoon1@rgu.ac.uk (Anjana Wijekoon)

role CBR can play in advancing the goals of XAI.

Keywords: Case-Based Reasoning, CBR-driven XAI, Explainable Artificial Intelligence, Human-understandable Explanations, Trustworthy AI

1. Introduction

The prevalence and swift progress of Artificial Intelligence (AI) [1] technology has brought to light the issue of system opacity, where the intricate workings of AI models often remain unclear to users. This lack of transparency creates obstacles to user trust and acceptance. It raises significant ethical and accountability concerns, particularly in critical areas of decision-making, including healthcare, finance, and law enforcement. To address this challenge of making AI decisions comprehensible, Explainable Artificial Intelligence (XAI) [2] has emerged as a vital area of research. However, developing effective XAI solutions faces several limitations due to the intricate and diverse nature of AI algorithms and the distinct requirements of various stakeholders. Among the methodologies explored, Case-Based Reasoning (CBR) [3] emerges as a compelling approach, which uses intuitive, example-based reasoning mechanisms to enhance the explainability of AI systems.

The convergence of CBR and XAI offers a promising pathway to address the above-mentioned challenges. CBR, with its foundation in analogical reasoning from past cases, inherently possesses qualities of transparency and interpretability. It provides a natural mechanism for explanation by drawing parallels between current problems and previously encountered cases, thus offering insights into the reasoning process in an intuitive manner. However, while CBR's potential to enhance explainability is recognized, its integration into the broader landscape of XAI requires careful consideration of methodological approaches, technological infrastructures, and user-centric design principles.

This paper comprehensively explores why and how a convergence of CBR and XAI is necessary and beneficial for advancing the field of AI towards more interpretable, trustworthy, and user-friendly systems. Our literature review found a few surveys on CBR for XAI. Keane et al. [4] advocate using CBR as a transparent counterpart to opaque AI systems, namely Artificial Neural Networks (ANN), enhancing interpretability within the XAI domain. Weber et al. [5] further this concept with knowledge-based XAI, which merges domain expertise with AI, leveraging CBR principles for more precise AI decision explanations through supervised classification. This method treats AI inputs and outputs as case problems and solutions, enriching explanations with domain-specific insights. Schoenborn et al. [6] underscore the importance of explanations in decision-making, presenting Explainable CBR (XCBR) as CBR systems that generate explanations. Unlike traditional case-based methods, XCBR offers a structured taxonomy for researchers aiming to create and apply explanations. Gates et al. [7] highlight the critical need for evaluating CBR explanations to enhance the understanding of intelligent systems. They propose evaluation strategies, survey XCBR systems, and define dimensions for categorizing CBR explanation components, suggesting future research avenues and community efforts to improve XCBR evaluations.

Our paper advances beyond previous studies by offering a comprehensive view that connects theoretical insights with real-world implementations, emphasizing the integration of CBR and XAI to meet the growing need for interpretable AI systems. [rev] As a consequence, we solve the following gap in the literature: we establish a guide when implementing CBR & XAI models, providing actual implementations of real-world use cases as examples. We explore how CBR techniques can be seamlessly integrated into XAI to clarify AI decision-making for users of all expertise levels, identifying key features or functionalities essential for this synergy. Unique to our study is the inclusion of a case study on a CBR-driven XAI platform we developed, showcasing how CBR enhances transparency and trust in AI systems. This examination highlights CBR's practical benefits in XAI and provides practitioners and researchers with valuable insights for designing and implementing effective CBR-driven XAI systems. [rev] Therefore, the main contribution of this paper is twofold. Following the format used by other authors previously [8] we specify our contribution in a list format:

1. We offer an in-depth literature analysis, focusing on the theoretical process for implementing interpretable AI systems considering XAI and CBR synergies. The outcome of this study is a guide to building such systems for AI developers.
2. We illustrate the development of a CBR-driven platform, a real use case implemented through the steps from the theoretical process described in point (1). Therefore, we detail how AI practitioners can use our guide through a practical example.

The guidelines proposed here to implement CBR and XAI synergies in the context of transparent AI model development might be used by XAI developers in four main situations. First, when XAI designers want to use CBR as a methodology to implement a post-hoc explainer for a black-box AI model, or when they want to use CBR to drive XAI procedures. Our study and exemplification can guide the design of these systems when designers are not familiar with CBR and/or XAI. Second, when XAI designers want to explain a CBR-based system. Although CBR is an inherently transparent methodology, our study might help designers consider different CBR features to take into account when creating explanations. Therefore, they might explore which steps or resources from the CBR methodology could be more useful to show to users as an explanation in a specific situation. Third, when designers need to implement one or more steps in the CBR circle applying them to XAI, but they do not know all the possibilities within that step. This might happen, for example, if designers are not experts on ontologies and semantic similarity metrics, but they want or need to leverage the advantages of generating explanations. The same might happen with the rest of the CBR resources and steps. Fourth, when designers need or want to analyze an example of how to use our guidelines or specific CBR procedures to implement their own XAI solutions. For instance, if designers do not know how to use semantic knowledge in their system and want to determine which semantic information is necessary, they can observe different options to apply in our guidelines. Moreover, we describe a platform developed as an XAI & CBR synergy: our iSee platform is an example of the different problems that XAI designers can encounter when developing explanation experiences, and how they can be solved using CBR. We expected this description to be illustrative for different XAI practitioners when developing their explanations experiences.

The outline of this article is as follows. Section 2 explores how CBR's inherent principles converge with XAI goals to enhance AI explainability. Moreover, we explore the practical methodologies by which CBR can be interwoven with XAI to demystify AI decisions. This includes a detailed discussion on the methodological integrations and adaptations necessary for CBR techniques to enhance the clarity and relevance of explanations provided by AI systems, filling a gap in the current discourse that often separates theoretical potential from practical application. Afterwards, Section 3 delineates the ethical implications and bias mitigation in the XAI domain. In Section 4, we provide a real-world CBR-driven XAI platform to enhance our understanding of the interplay between CBR and XAI, as well as the design and implementation of the platform. Finally, closing remarks are presented in

Section 5.

2. Converging CBR and XAI for Explainability

The convergence of CBR within XAI frameworks offers a structured approach to generating explanations. This involves identifying past cases that parallel the current decision context and providing a narrative or reasoning trail that users can follow and understand. For instance, in a medical diagnosis AI system, a CBR-driven explanation could detail how the AI's recommendation matches or diverges from previous diagnoses under similar patient conditions, thereby grounding the AI's decision in concrete, understandable examples. Explanations must be tailored to suit various contexts and specific requirements, considering the users' objectives, the breadth of the explanation needed, or the data at hand. Selecting the optimal explanation strategy for specific AI applications and users presents a challenge. This issue can be addressed by creating a unified platform that enables AI developers to identify and implement the most compelling explanation technique for particular scenarios.

Achieving explainability transcends technical challenges, representing a complex effort that caters to the diverse requirements of different stakeholders [9]. These stakeholders include data scientists, domain experts, developers, regulators, and end-users, contributing varied perspectives, objectives, and challenges. The main goals of XAI, including informativeness, transferability, accessibility, fairness, confidence, interactivity, and causality, underscore the alignment of these goals with the unique information requirements of various stakeholders, as outlined in the literature [10]. To make AI decisions understandable to users, XAI systems must tailor explanations to match the interpretability of various data types, such as tables, text, time-series, and images, which differ in their ease of human understanding [11]. Chari et al. [12] highlight the necessity of employing diverse explainability types, such as *case-based*, *contrastive*, *counterfactual*, and *trace-based*, to cater to the varying requirements of users. Explanation scope delineates the extent of interpretation, ranging from *global*, which encompasses the whole model, to *local*, concentrating on the reasoning for individual predictions [13, 14]. Moreover, explanation methods underline key text elements influencing outcomes, providing insights into the model's logic. The effectiveness of these methods depends on their alignment with stakeholder needs, ensuring insights improve user comprehension and meet expectations [13]. XAI methods include *ante-hoc*, where models are inherently interpretable, offering direct insights into workings, and *post-hoc*, balancing predictiveness with interpretability by revealing decisions without detailing internal mechanisms [15].

We will examine CBR's unique ability to provide intuitive explanations with other XAI methods, highlighting its capability to offer transparent, easily understandable analogies rather than dissect model architecture or quantify feature importance. Post-hoc methods, namely model-agnostic and model-specific methods [10, 15] primarily focused on interpreting or explaining the decisions of machine learning models, either without regard to (agnostic) or with specific consideration of (specific) the model's internal mechanisms. In contrast, CBR uses actual instances from the case base for analogical reasoning, presenting complete scenarios that closely match the current problem. Feature importance methods [16] offer insights into the contribution of individual features but might fall short of delivering a comprehensive view of the decision-making process. Meanwhile, CBR provides explanations through analogies, showcasing complete cases similar to the current situation. This method allows users to fully understand the logic behind decisions, making the explanations more accessible and relevant. Furthermore, counterfactual explanations [17] show how small input changes can alter decision outcomes. While crafting meaningful counterfactuals in complex domains is challenging, CBR incorporates counterfactual reasoning by showcasing how minor variations in similar cases affect outcomes. Counterfactual discovery often employs CBR, focusing on optimization-based or example-based algorithms using CBR techniques for Nearest Unlike Neighbors (NUNs) retrieval and adaptation for actionable decision changes [18, 19]. This approach provides insights into decision boundaries without artificial scenarios, positioning CBR as a vital component of XAI for bridging the gap between intricate AI operations and user understanding. CBR offers example-based explanations [20], also called factual or instance-based explanations, which show users past cases solved with the predicted AI model solution. This resonates well with users by mirroring the human tendency to learn from previous experiences. It is a highly effective method in XAI for its clarity and natural fit with human cognitive processes [21].

Implementing a CBR-driven XAI system involves several challenges, including the selection and adaptation of past cases to fit the current problem context accurately. There is also the need to maintain a comprehensive, up-to-date case base reflecting the diversity and complexity of real-world problems. Moreover, the effectiveness of CBR-based explanations depends on the system's ability to select relevant and understandable cases for the intended audience, necessitating careful consideration of user needs and preferences in the design of the explanation generation process. Our literature review identified fundamental elements for converging CBR and XAI: structured case representation, domain knowledge integration, experience-based reasoning, similarity-based retrieval, adaptation and learning, case base maintenance, and an iterative and interactive process. Table 1 outlines these essential CBR principles and their relevance to XAI, which will be discussed in the following sections. These principles correspond to the identified functionalities and demonstrate their significance for XAI by illustrating how each principle enhances AI systems' transparency, comprehensibility, and interpretability.

2.1. Case Structuring and Domain Knowledge for Enhancing Explainability and Semantic Interpretation

In CBR, knowledge is stored as cases containing a problem, solution, and outcome, collected in a case base [29]. Case representation describes the organization and storage method within the system. Machine-readable ontologies are a popular approach that facilitates creating user-specific explanations by clarifying complex data relationships and enhancing comprehension [12]. Utilizing ontologies in case description improves interoperability and fosters collaboration across CBR methods, enriching explanations with diverse knowledge sources. Semantic knowledge for case representation enhances AI interpretability and explanation simplicity, enabling natural language and visually clear, example-based explanations, such as loan approvals based on similar past cases. This semantic richness enhances the clarity and significance of explanations for XAI systems, helping users grasp both the rationale and the specifics of AI decisions. For example, Tiddi et al. [30] introduced an Ontology Design Pattern (ODP) for diverse explanatory concepts, while the Food Explanation Ontology (FEO) [31] formalized domain-specific answers for AI-driven food recommendations. Another approach [32] simplified explanation integration in recommender systems, addressing user expectations and knowledge with a new conceptual model, RecOnto, guiding effective explanation development. Chari et al. [12] proposed an explanation ontology for user-centered AI design, addressing questions including "How, Why, Why-not, What-if, and How-to."

Incorporating domain knowledge [5] into XAI systems enhances transparency and comprehension, particularly in finance or healthcare sectors,

Table 1: CBR Principles and Their Suitability for XAI

Functionality	CBR Principle	Relevance to XAI
Structured Representation of Cases [22].	Cases in CBR are structured with a problem description and its corresponding solution, which may also include annotations about the case's context or rationale.	Structured representation enables the system to explicitly map decisions to prior instances, enabling users to easily trace back to similar cases to comprehend the rationale behind a particular decision.
Domain Knowledge Integration [23].	Integrated with domain-specific knowledge, such as ontologies or rules, to enhance its reasoning capabilities.	Utilizing domain knowledge can provide comprehensive and context-aware explanations, offering users a deeper understanding of the decision-making process.
Experience-Based Reasoning [24].	"Similar problems share similar solutions," which relies on past experiences (cases) to address new and comparable problems.	The reasoning process is transparent, as it relies on concrete past instances, enabling it to provide clear and relatable explanations for its decisions by referring to past cases.
Similarity-Based Retrieval [25].	Utilize the similarity between the current problem and past cases to retrieve relevant cases.	The similarity metrics and criteria provide an objective framework for case retrieval, thereby enhancing transparency in the decision-making process.
Adaptation and Learning [26].	Adapt solutions from past cases that may not perfectly fit the current problem, and they also learn by storing new experiences.	Adaptation process can be transparent by showcasing how past solutions are modified for the current context, and with continuous learning, the system's knowledge remains updated and relevant.
Case Base Maintenance [27].	Undertake regular reviews and updates by eliminating obsolete cases and adding new and valuable experiences	Ensure that the explanations remain relevant and accurate over time, bolstering the system's credibility.
Iterative and Interactive Process [28]	Utilizes an iterative approach, wherein the system engages in a dialogue with users to refine problem descriptions or validate solutions.	Interactivity allows users to be involved in the reasoning process, enhancing trust and providing opportunities for real-time clarification.

making AI decisions more relevant and understandable. This method utilizes expert insights and industry-specific data for case adaptation, fostering trust and informed decision-making. It also supports post-hoc verification, ensuring the scientific accuracy of AI recommendations and improving user collaboration by clarifying AI reasoning in real-world contexts [33]. Domain knowledge and ontologies refine explanations, boosting accuracy, trust, and user satisfaction [34]. Doctor XAI [35] demonstrated integrating domain knowledge into ontologies improves explanations, particularly temporal data. Similarly, ontologies aid in clarifying global post-hoc explanations in decision trees [35, 36]. Studies by Islam et al. [37] shown the application of domain knowledge in finance and cybersecurity to improve black-box model explainability, achieving competitive performance with enhanced explanations.

2.2. Experience-based Reasoning and Similarity Retrieval for Explanation Generation

The CBR cycle [38], leveraging historical data and past experiences, iteratively refines problem-solving with each new case, applying its practical methodology across fields, including medical diagnosis, legal reasoning, and more. The first phase in the CBR cycle is Retrieval, where the system compares a new instance with all stored cases in the case base using similarity metrics [39], focusing on features or criteria specific to the domain. The system retrieves cases most similar to the instance, offering explanations based on problems closely related historically. For example, medical diagnosis finds patients with similar histories and symptoms, emphasizing the importance of accurate problem definition and appropriate similarity metrics for effective case retrieval. The choice of similarity metric is pivotal, influencing outcomes and solution quality, emphasizing careful consideration of similarity metrics to enhance example-based explanations and the application of CBR in XAI. We will explore how experience-based reasoning and similarity-based retrieval generate meaningful explanations, as outlined in Table 1. Experience-based reasoning [40] utilizes past case knowledge to address new challenges, drawing from similar past solutions and their outcomes for guidance. Similarly, similarity-based retrieval [41] searches for analogous cases using defined metrics, aiding in identifying practical solutions for comparable situations.

Cunningham [42] categorizes CBR similarity metrics into four groups: *direct similarity mechanisms* use feature vectors for straightforward comparisons, such as Overlap metrics and Euclidean distances; *transformation-based measures*, including Edit distances and Tree Edit Distance (TED), assess the effort to change one case into another, highlighting case differences.; *information-theoretic measures* analyze raw case data, bypassing feature vectors; and *machine learning-based metrics* apply Machine Learning (ML) techniques to define similarities, necessitating explanations for the ML reasoning processes. Similarity metrics fall into three categories: local, global, and quasi-local [43]. *Local* metrics assess similarity based on a single attribute, offering transparency but limited breadth in evaluation. Meanwhile, considering all attributes, *global* metrics provide comprehensive case comparisons at the cost of reduced transparency [44]. Finally, *quasi-local* metrics strike a balance by evaluating a selected subset of attributes. The choice between these metrics depends on the trade-off between accuracy and explanation transparency in seeking optimal case-based explanations. Moreover, similarity metrics leverage structural or semantic knowledge to assess case similarities [45]. *Structural knowledge* employs data structures, namely graphs, trees, or Behaviour Trees (BTs) for case solutions [46]. In contrast, *semantic knowledge* uses ontologies to describe cases by concepts and properties within a domain [47]. Integrating both approaches enriches similarity assessments, which will be illustrated in Section 4 using BTs for solution representation and ontologies for explainer semantics.

During the retrieval phase, the *MAC/FAC approach* is often employed [48]. This approach involves a two-step filtering process, which begins by identifying cases from the case base that match the query's problem features [49]. While this step is straightforward, the subsequent similarity assessment is critical for retrieval and may lack transparency regarding feature contribution to case ranking [49]. Selecting suitable similarity metrics and providing visual explanations can help users understand the basis of case selection, enhancing the transparency of CBR predictions [41].

2.3. Adaptation and Learning for Customized AI Explanations

We will delve into how CBR systems customize solutions through adaptation and learning, highlighted in Table 1, crucial for tailoring responses to user-specific needs. During the second phase in the CBR cycle, i.e., Reuse or Adaptation [38], the CBR system modifies solutions from past cases to fit the new problem based on the adaptability indicated by case similarity in the retrieve phase. This process enables the generation of personalized explanations, such as updating a medical treatment plan for a new patient's unique situation. The system needs to deliver solutions that are precise and explained in an accessible manner to users. Adaptation techniques for CBR systems are classified into transformational and generative. *Transformational adaptation* involves altering the structure of a solution to fit a new problem, requiring significant modifications [26]. *Generative adaptation*, on the other hand, rethinks the solution-creation process for new scenarios, often building parts of the solution from scratch without relying solely on past cases, addressing complex problems, or bridging solution gaps [26, 50]. *Constructive* *adaptation* [51], a subset of generative adaptation, combines elements from similar cases to create a new solution, such as combining cultural activities from different vacations to recommend a unique urban cultural tour itinerary. The literature explores two methods for learning adaptation knowledge in case reuse: weighted majority voting and case difference heuristic (CDH) [52]. Incorporating ontology into case descriptions, as discussed in Section 2.1, streamlines retrieval and adaptation, ensuring solutions closely match queries with minimal adjustments. This method enhances solution tailoring and elucidates the decision-making process for users, boosting transparency and comprehension.

The revision stage [38], the third phase of the CBR cycle, involves assessing and potentially modifying the solution applied to a new problem to refine its effectiveness. This continuous evaluation and adjustment stage enhances the CBR system's performance and accuracy. An example includes monitoring a patient's response to a tailored treatment plan and adjusting it based on their feedback. Adaptation, particularly during the CBR cycle's revision phase, evaluates and adjusts solutions based on feedback to ensure alignment with user queries. This phase scrutinizes the suitability of solutions and iteratively refines them, integrating new knowledge into the system for enhanced future problem-solving. Successful adjustments result in updated cases stored in the case base, continuously enriching the system's knowledge base and capabilities [53].

2.4. Case Base Maintenance for Enhanced Explanation Quality

This section explores the importance of case base maintenance (CBM) in the CBR retain phase, as outlined in Table 1. Retain phase [38] updates the case base by saving newly solved cases, including problem descriptions, adapted solutions, and relevant details for future reference. For example, new patient cases, including their treatments and outcomes, are recorded to aid future medical diagnoses. CBM is crucial for the accuracy and reliability of CBR integrated with XAI systems, ensuring the case base remains current and effective in providing clear explanations [27, 54]. CBM involves updating the case base to reflect new knowledge, deleting obsolete or redundant cases, merging cases to enhance reasoning, and correcting inconsistencies to maintain or improve system efficiency and explanation quality [27, 55]. Strategies for CBM include optimizing case representation and pruning unnecessary cases, resulting in a streamlined case base that facilitates faster retrieval and sustains problem-solving competence [56].

The literature outlines various algorithms for CBM, including the k-Nearest Neighbors (k-NN) [57] classifier for identifying and removing redundant or noisy cases and instance reduction algorithms for optimizing the case base by clustering or instance-based learning [58]. Effectiveness is evaluated using ML techniques, including Hold-Out and Cross-Validation, to compare performance metrics of the original and updated case bases [55]. The quality of explanations depends on the accuracy and diversity of cases, necessitating regular updates to remove outdated information, thereby maintaining the relevance and trustworthiness of explanations [59]. Transparent documentation of changes and rationale for case updates ensures users understand the maintenance process while reducing redundancy improves efficiency and consistency of explanations [60].

The structure of the case base significantly influences explainability, with a well-organized case base facilitating the precise tracing of solutions and enhancing user trust through reliable explanations. Effective case base maintenance ensures the system remains dynamic, improving by adding new cases and expanding its adaptability and learning (see Section 2.3) capabilities. Personalizing the case base for specific user needs or domains further enhances explanations, improving user satisfaction. CBM is crucial for maintaining high standards of explainability and fostering trust in AI applications.

2.5. Iterative and Interactive Process for Enhancing User-Centric Explanations

This section explores the critical role of iterative and interactive learning, mentioned in Table 1, emphasizing the necessity of engaging user interaction, effective feedback mechanisms, and continuous learning to uphold a user-focused approach [28, 61, 62]. Personalizing explanations [28, 63] by allowing users to influence their depth and scope significantly enhances the transparency and relevance of predictive systems. This process adapts explanations to meet individual user needs, making complex AI decisions more understandable and actionable. For example, personalization enables users to receive specific guidance on improving their financial profiles, offering concrete, actionable insights rather than generic feedback in a credit-scoring XAI system. This tailored interaction fosters a more engaging user experience by directly addressing user queries with personalized information.

Crafting user-friendly interfaces that align with user preferences and simplify system interaction is crucial. These interfaces must adapt to user feedback and behavior, enhancing intuitiveness and user focus. Incorporating simple feedback tools, including ratings and comments, facilitates user engagement and input [64]. For example, while collaborative filtering uses user ratings for personalized suggestions, it is often opaque. A novel approach combines CBR and Formal Concept Analysis for transparent explanations in recommendation systems, utilizing user interactions to identify and explain item recommendations effectively [65]. This method improves explanation transparency and gathers user feedback to refine and trust the explanation process.

Causal explanations in XAI emerge from dialogues tailored to specific "why" questions, emphasizing accuracy and relevance [66]. Such conversations, incorporating text and visuals, allow explanatory agents to address multiple aspects of AI decision-making, making explanations more intuitive and user-centric [67, 68]. For instance, using the Locally Interpretable Model-Agnostic Explainer (LIME) algorithm for image explanations involves segmenting images to show how changes affect AI outputs, thus personalizing the explanation process. This interactive and visual approach enhances AI transparency, fosters user engagement by allowing them to influence the explanation process, and supports a deeper understanding of AI decisions, aligning with fairness and accountability objectives.

Surveys and feedback forms are essential for gathering user input on a system's functionality and explanations, with studies by Smith et al. [62] revealing the importance of providing feedback opportunities alongside explanations to enhance user satisfaction and system improvement. Tailoring feedback mechanisms to user profiles and leveraging implicit feedback through user interactions enhance system relevance and engagement. Techniques, namely graph-based approach [69], use user data for personalized explanations, while feedback-driven updates refine the case base, ensuring its effectiveness [70, 71]. Adapting XAI algorithms based on feedback [72] and employing iterative refinement methods, namely IREX [73], allow for continuous system accuracy and user understanding improvement.

3. Ethical Implications and Bias Mitigation

Addressing biases in CBR systems is essential for fairness and accuracy, as biases from underrepresented cases in the case base can skew decisions. To mitigate biases, diversifying the case base to reflect a wide array of situations and conducting regular audits to adjust retrieval algorithms are crucial steps. Adjusting similarity metrics or integrating fairness constraints ensures more equitable case selection [74, 75]. Research [37, 76] highlights the importance of identifying and addressing biases, suggesting XAI techniques for visual bias evaluation and fairness reporting. Developing fair-by-design ML models [77] and employing algorithms, including CERTIFAI [78] to assess model robustness and fairness can lead to less biased, understandable explanations, enhancing the fairness and transparency of AI systems.

Historical data biases can impair AI model effectiveness, necessitating a comprehensive approach to bias mitigation involving data scrutiny, fairness audits, and collaboration with domain experts [14]. Transparency in AI decision-making enhances interpretability, aiding bias identification and trust-building, despite explanations not assuring system trustworthiness [79]. Trustworthy AI development strategies include quantitative metrics for explanation quality and human evaluation methods to ensure reliability before practical application [15]. Furthermore, the trade-off between transparency and privacy in model explanations requires careful management to prevent information leakage, as analyzed in research on backpropagation and perturbation-based explanations [80].

Privacy and data protection are critical in CBR and XAI systems due to extensive datasets, including sensitive information [10]. Balancing data use for AI performance with privacy rights poses a significant challenge [14], necessitating GDPR compliance via anonymization and encryption. Clear guidelines for AI decision accountability are essential, with XAI enhancing this by enabling decision contestation and action modification for better future outcomes [81]. Testing AI algorithms for policy compliance without revealing proprietary details is crucial, as is allowing external verification to ensure objectives are met and providing explanations for discrepancies [82]. Collaborating with experts across fields can offer deep insights into addressing potential biases. Incorporating a 'human-in-the-loop' approach [2] ensures a balance between AI's capabilities and human oversight, allowing users to report biases and impacts, which are crucial for identifying and correcting system biases. Furthermore, ethical training for AI developers and awareness among users is critical to embedding ethical considerations into AI's responsible development and use [10, 83].

4. iSee: A CBR-driven XAI Platform

The surge of interest in XAI has led to an extensive array of methods for explaining AI decisions, known as explainers, which cater to diverse contexts and needs, such as user objectives, the scope of explanation, or data type [32]. While various explanation methods benefit research and industry, selecting the optimal approach for specific AI applications and user requirements presents a significant challenge [49]. The *iSee: Intelligent Sharing of Explanation Experience* [1] project aims to address this issue by creating a unified platform that enables AI developers to select and implement the most compelling explanation strategy for particular AI scenarios. "Explanation strategies" refer to the diverse methods and techniques developed to interpret ML models and elucidate their predictions, recommendations, and diagnoses [84]. These strategies are designed to cater to the varying needs of different stakeholders, such as technological experts, domain experts, and impacted individuals or subjects [9], who may have distinct backgrounds, skills, and objectives. As these strategies evolve, they equip practitioners with the knowledge to select the most appropriate methods for explaining AI behavior in various contexts. The platform incorporates tools for evaluating the efficacy of explanation strategies in specific contexts and utilizes detailed knowledge structures. These structures help compare scenarios, understand contextual differences, and effectively adapt explanations to meet varying user needs.

iSee is a CBR-driven XAI recommender that aims to enhance the explainability of AI systems by incorporating greater abstraction in the explanation process. The iSee consortium comprises researchers who advocate using the CBR paradigm to capture the knowledge and experience gained from successful adaptation of explainability in AI systems. iSee leverages these experiences to assist AI systems in building explainability that adheres to regulations, such as the EU's 'right to explanation' [2]. The key terms in iSee are as follows: (i) Explanation is an artifact created to enhance a user's comprehension of a system's decision or output, (ii) Explainer (or explanation algorithm) is the algorithmic element within an XAI system tasked with explaining the system's AI component, (iii) Explanation Experience refers to the interaction between the explainee and the XAI system, (iv) Explanation intents refers to the explainee's motives and reasons behind needing explanations, (v) Explanation strategy is understood as the combination of explainers and other workflow components to generate an explanation expe-

[1] https://isee4xai.com/
[2] https://gdpr.eu/

rience that offers different explanations according to the explanation intents.

To the best of our knowledge, iSee stands out from other XAI platforms by providing explanation experiences tailored to individual users and prioritizing their needs. Existing XAI libraries, such as Alibi [85], Dalex [86], and Xplique [87], provide a wide range of explainer tools primarily for developers. However, the iSee platform enables users to customize and apply explainers on-demand and uses CBR to recommend optimal explanation strategies tailored to individual user profiles and case details. iSee offers a comprehensive explainer catalog for diverse data types and includes unique explainers developed by iSee team, such as DiSCERN [88], PertCF [89], IREX [90], and specialized time-series explainer, namely CBRFox [91].

4.1. iSee CBR methodology

The iSee platform [92], rooted in CBR methodology, empowers AI designers to capture and share intricate "explanation experiences" with peers facing similar explanation needs. These experiences leverage various XAI methods to understand the system, tailored to users' requirements thoroughly. Aimed at creating an open catalog of such experiences, it supports the adaptation and customization of explanations to meet diverse needs across trustworthy AI applications. The CBR methodology enables the transfer of solutions from past explanation experiences by customizing them to fit new scenarios. Additionally, users can personalize these solutions based on their preferences. The integration of this process within the CBR cycle involves several steps, as shown in Figure 1. Building on the case representation and domain knowledge integration aspects outlined in Section 2.1, we developed an ontology named iSeeOnto [93]. This ontology, formulated through literature review and analysis of real-world applications, outlines the essential concepts for delineating an explanation experience, which will be elaborated in Section 4.2. Relevant features of best practice explanation experiences are then gathered from different use cases, stored in a case base, and retrieved based on ontology-based weighted similarity. Based on experience-based reasoning and similarity retrieval discussion in Section 2.2, iSee assesses similarities between query cases and the case base for accurate retrieval utilizing cloodCBR [94]. Further details on this cloud-based CBR system are provided in Section 4.3. The best matching case solution or an explanation strategy is represented using BT, as mentioned in Section 2.2. The adaptation and learning function, explored in Section 2.3, leads to developing the iSee reuse strategy, designed to customize solutions for specific query cases, as elaborated in Section 4.4.

Utilizing insights from Section 2.5, the explanation strategy is integrated within a chatbot interface to facilitate interactive feedback loops with end-users, enabling tailored explanation delivery. Feedback collected during this interaction informs further strategy refinement during the revision phase, and successful strategies are archived in the case base for future application, as discussed in Section 2.4. The processes of revision and retention are further elaborated in Section 4.5, ensuring a dynamic and evolving approach to explanation generation within CBR-driven XAI systems.

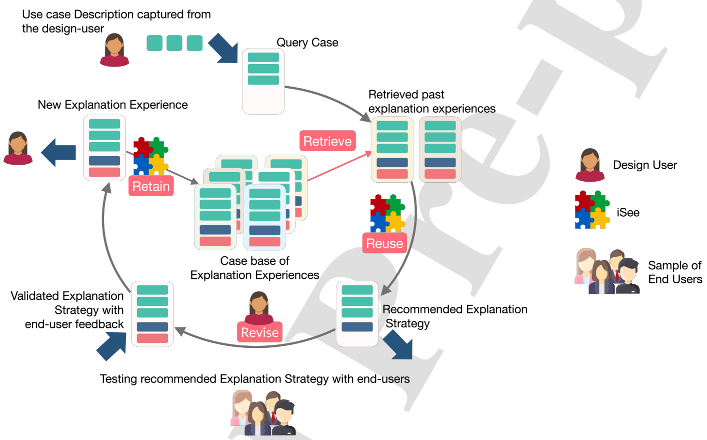


Figure 1: iSee CBR methodology showing the four phases

4.2. Formalisation of the Explanation Experiences

iSeeOnto [93] is an ontology designed explicitly for user-centered XAI, focusing on capturing explanation experiences. It facilitates the characterization of these experiences as cases, each comprising a description, a solution, and a result, which the iSee CBR engine can then re-purpose. The development of iSeeOnto employs the NeON methodology [95] for constructing ontology networks, guiding the definition of the Ontology Requirements Specification Document (ORSD) [96] to outline the purpose, intended end-users, and requirements to be met by iSeeOnto. Various explanation intents and the types of explanations provided in [12] were helpful in the iSeeOnto conceptualization and formalization stages.

A case outlines an instance of providing an explanation, detailing how a particular explanation strategy fulfills specific explanatory needs. Consequently, the ontology delineates the case *Description* as three primary concepts that characterize such an explanation experience: (1) the specific *AI model* requiring explanation, (2) the necessary features of the Explainer tool to elucidate the AI model, and (3) the *User* along with their knowledge levels for understanding the explanation. The *Solution*, i.e., the explanation strategy to be applied in this explanation experience case, explains the AI model while fulfilling the user's explanation intents. In iSee, we use BTs to formalize the solutions [92], as mentioned in Section 2.2. A BT solution contains the questions that reflect user intents, the explainers that will be executed to address said intents, and other structures required to model the relationships between them. The formalization of the BTs for representing explanation strategies is also included in iSeeOnto as *Solution* and BT concepts and all the concepts that define a BT. The *Result* concerns the collective evaluations made by end-users about the explanation obtained from executing the solution. Our previous work [92] provides a comprehensive description of how we modeled explanation strategies using BT.

iSeeOnto concepts, namely *Explainer*, *AI model*, and *User*, include more sub-concepts necessary to apply the CBR retrieval and reuse steps to get similarities between cases. The case description consists of attributes such as AI Task, AI Method, Dataset Type, Portability, Scope, Target, Presentation, Concurrentness, Intent, TechnicalFacilities, AIKnowledge, DomainKnowledge, and User Questions [92]. The explainer has an explainability technique that returns an *Explanation Type* with a specific *Presentation* format, processes a *DatasetType*, has different attributes about explainability (*Concurrentness*, *Portability*, *Scope*), has a *Computational Complexity*, applies to different AI methods and AI tasks, is implemented by a *Framework*, needs or not *Training data* and has a *Model Access Type*. An AI model, which is trained using a *DatasetType*, solves an *AI Task*, which has an *AI goal*, using an *AI method*. The user, who has an *Intent* and *TechnicalFacilities* (that can handle an *Explanation Modality*), asks a *Question*, which has a *Target*. The user also possesses *Knowledge*: *AIKnowledge*, and *DomainKnowledge*. More details of iSeeOnto [93] can be found here [3].

[3] `https://w3id.org/iSeeOnto/explanationexperience`

4.3. Retrieval of Explanation Experiences

The iSee case retrieval system is built on the CloodCBR framework and is seamlessly integrated with the iSee Dashboard. Clood is a cloud-based CBR framework based on a microservices architecture that facilitates the design and deployment of CBR applications of various sizes [97]. It supports semantic similarity metrics for local similarity, such as similarity table, word embedding-based similarity, and ontology-based similarity measures. The use of these similarity metrics helps to reduce retrieval overhead. A more detailed description of CloodCBR can be found in [97].

The retrieval process is triggered by a query characterized solely by its description, and the goal is to uncover the solution corresponding to this query. When creating a query case, retrieval is done by selecting a part of the knowledge that the design user provides. Specific attributes and an explanation strategy are chosen to represent the case. The ontology components are used to consider particular characteristics or properties of a case, which are then matched using similarity metrics, namely Wu & Palmer, Query Intersection, and Exact Match [97, 92]. Every case attribute is assigned a local similarity metric, which is then used to compute the global similarity as the average of local similarities between the attributes of the query case and a case from the case base. The weights or importance values of the local similarity vary depending on the attribute type and similarity metrics. If there is a case with query attributes filled by iSee ontology classes or individuals, the case retrieval task is to search for explanation strategies from the nearest neighbors. The system allows users to retrieve top 'k' cases, explore recommended explanation strategies, and manually refine the selected strategy. The iSee case base comprises 18 seed cases from 50 peer-reviewed papers that describe how to implement XAI within AI systems with user evaluation [92]. The case base will continue to grow by adding new seed cases from the literature and retaining new experiences with more complex strategies created within the iSee CBR platform.

4.4. Reuse of Explanation Experiences

In iSee, the reuse step requires adapting the proposed solutions from the retrieval step discussed in Section 4.3. These solutions are represented as BTs, with leaves indicating the explainers to apply and composite nodes showing how they will execute [92]. While retrieval can provide several solutions that satisfy AI model needs and user requirements, they may not always suit the use case. The only applicability requirement considered during retrieval is that the AI model and explainer process the same data type. To ensure that the BT is suitable for the use case, it is necessary to verify other explainer properties, such as the AI task and AI method that the explainer can explain, as well as the implementation framework used to implement the explainer. The user adapts retrieval solutions to change unsuitable explainers with iSee's reuse functionalities. These functionalities enable the reuse of explainers and subtrees in the BT during the iSee platform's reuse step.

The iSee platform facilitates a tool called the Explanation Experiences Editor (iSeeE3) [98], which allows users to create new BTs from scratch or edit them manually. It also offers iSee Explanation Library, which provides a list of explainers that combine over 50 from various XAI libraries [99]. The repository of explainers can be found on GitHub [4]. The iSee platform provides two main functionalities for its reuse step: explainers adaptation and BTs adaptation.

- **Explainers adaptation**. Users can adapt the non-applicable explainers in the BT and get a list of similar explainers applicable to the case from the iSee explainer library. The similarity of these explainers is calculated using semantic similarity metrics based on the semantic knowledge that describes an explainer, according to iSeeOnto. Users can specify the properties of the new explainer they require by filling out a form, and they will receive a list of explainers that are similar to their query and fulfill their requirements. Figure 2 illustrates the explainer reuse functionality facilitated by iSeeE3. The "Search substitute explainers" button shows the set of similar and applicable explainers to the one selected that the user might replace. The "Substitute explainer with criteria" button will allow users to access this functionality through a form to specify the explainer properties needed in the list of recommendations.
- **BTs adaptation**. Users can replace the whole BT with another applicable one. They will receive a list of similar BTs based on edit distances from the case base solutions whose explainers are relevant to the use case. iSee searches for similar BTs, utilizing an adapted Levenshtein edit distance to calculate the transformation cost between BTs through insertions, deletions, and substitutions. In iSee, this metric is tailored

[4] `https://github.com/isee4xai/iSeeExplainerLibrary`

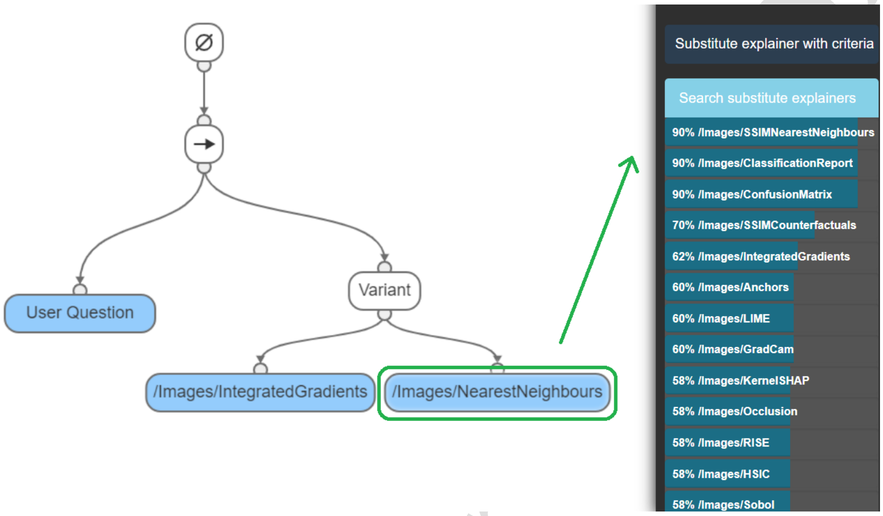


Figure 2: Explainers adaptation functionality: iSeeE3 advocates substituting explainers with similar solutions. In the example, we want to replace the Nearest Neighbors explainer with another similar one applicable to that use case. Node marked '→' signifies sequence node, and 'Variant' is a composite node.

for graph structures, comparing node and adjacency lists to assess differences between BTs. The iSee platform also enables users to specify the properties of the explainers they want in the new BT by filling out a form. Figure 3 shows how we can carry out this type of adaptation in iSee.

Users can perform both functionalities manually or automatically. They can select the new element from the recommended list of explainers or BTs in the manual version. Nevertheless, the automatic version allows users to replace the non-applicable explainers or the whole BT by clicking a button, and the iSee platform will recommend the most similar and applicable option.

In CloodCBR, failure-driven transformational case reuse [100] is introduced to boost problem-solving efficiency by re-purposing solutions from previous cases. This approach involves refining less-than-ideal solutions by incorporating elements from the most closely related cases in sparse databases. It consists in identifying failures, formulating new solutions tailored to address those failures, and testing them against the intended explanation goals.

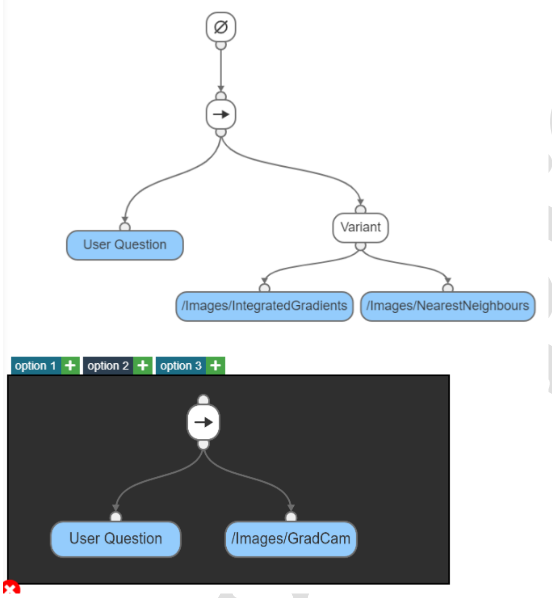


Figure 3: BTs adaptation functionality: iSeeE3 allows users to find new similar BTs to replace the current solution. Here, the user can substitute a BT with two explainers at the top with a simpler one with a different explainer at the bottom.

Through empirical testing, we compared different solution reuse approaches and verified the efficacy of the revamped solutions. This method holds promise for enhancing the utilization of relevant cases and pinpointing suitable adaptation methods for novel issues, particularly in databases with few cases and complex solutions.

4.5. Revision and Retention of Explanation Experience

In the revision phase, the iSee platform revises the explanation experience, allowing design users to modify the retrieved strategy to meet additional requirements manually. These revisions are based on the feedback from the end-users who evaluated the explanation strategy using the chatbot [92]. The chatbot uses the execution workflow defined by the BT and begins by asking users questions about their profile and intention before delivering the relevant explanations. After an interactive session, the system collects user feedback on the evaluation questionnaire and structures them into a User Evaluation Result [92]. When numerous user interactions have been accumulated, the design user reviews the User Evaluation Results. Negative feedback, indicating unmet explanation needs or dissatisfaction with the provided explanations, prompts the design user to refine the explanation strategy via revision. Once positive feedback confirms end-user satisfaction, the case can be retained in the case base as a successful explanation experience, ready for future recommendations.

Retention allows us to add fully developed use cases, which progressed through all stages of the iSee CBR, within the case base for future use. These complete cases are crucial for responding to new queries on the iSee platform, as they encompass problem, solution, and result components, ensuring that recommended strategies are proven and effective. Without the result component, there is a risk of suggesting unverified or failed strategies. Therefore, complete cases are vital for embodying the full spectrum of explanatory experience and defining 'success.' To reflect state-of-the-art, the iSee platform also needs to continuously incorporate new seed cases contributed by XAI practitioners, including innovative explanatory algorithms and strategies. However, these contributions may need more associated outcomes, indicating their utility in user satisfaction. We have conducted a case study on Radiology Fracture Detection (RFD) in our prior research [92], which demonstrated the functionality and workings of the iSee platform in great detail.

[rev]

4.6. Example Use Case: Radiograph Classification for Fracture Detection

We demonstrate the use of the iSee platform in the real-world using an example radiograph fracture detection AI system provided by one of the industry partners. Their AI system is implemented using ConvNet-based architecture for binary classification of fractures in radiographs. The stakeholder explanation needs stems from the need to improve the quality of their product for end-users (i.e. clinicians and radiologists) and to increase regulatory compliance with relevant governance bodies. Accordingly, the design user described two user groups: 1) clinicians who are using the AI system for decision support; and 2) managers who are looking to evaluate the compliance, risk, and regulatory requirements.

The design user leverages iSee's *retrieve*, *reuse*, and *revision* functionalities to create a complete Explanation Experience case and *retains* in the case base. Firstly, a description and implementation of a ConvNet model for binary classification of black and white radiography images are entered into the iSee Cockpit. Details of a clinician persona are then entered, alongside corresponding intents in transparency and performance, thus completing the explanation experience case description. The iSee platform retrieval functionality uses the case description to query the iSee case base and *retrieve* a set of candidate cases containing historical best practices of explanation strategies. The design user *reuses* the recommended solution and performs adaptation to obtain a personalized strategy either using explainer or BT adaptation. Finally, when revisions are complete, the case now contains a solution component. This proposed solution is then evaluated with target end users (e.g., a group of clinicians) to formulate the outcome. This workflow formulates a complete case, which can subsequently be *retained* in the case base to inform future practice.

4.7. Ethical Considerations in iSee

The iSee platform is susceptible to confidentiality breaches, a common challenge for complex explainability frameworks and multifunctional recommender systems. These breaches could occur at strategic levels, such as data protection risks from 'intent' understanding measures, use-case levels through explainer models using training data from other AI systems, or individual user levels by utilizing personal data for explanations. Legal requirements might necessitate avoiding non-anonymized data in sensitive areas, namely healthcare, yet data breach risks persist across scenarios. Identifying these ethical dimensions is crucial for the iSee system to enable effective threat modeling and enhance system security. Security in iSee emphasizes the importance of monitoring abnormal behavior by explainers to protect against cyber threats. Proactive and reactive measures will strengthen security, including user feedback and a suspicious behavior report system. Ensuring data confidentiality while maintaining the iSee platform's effectiveness and providing informative explanations is challenging, as there is a risk of inadvertently disclosing personal data via model explanations. Accountability is another measure that involves defining the responsibility of users interacting with iSee, particularly design users acting as organizational representatives, ensuring the correct interpretation of data and results. An ethical agreement will govern iSee component use to ensure compliance, smooth adoption, and alignment with real-world applications and policy.

5. Conclusions

This paper demonstrates the effective convergence of CBR with XAI by deploying a novel CBR-driven XAI platform, iSee. Our exploration of this integration reveals the significant potential of CBR to enhance explainability, transparency, and user trust in AI systems [rev] [101]. By leveraging the strengths of CBR — its ability to provide relatable, intuitive explanations based on historical cases [rev] [102] — we have outlined a comprehensive framework for developing XAI systems that are not only effective but also user-centric. This framework involves factors such as case representation, domain knowledge integration, experience-based reasoning, similarity retrieval, adaptation, and user feedback, which emphasize the essential role of CBR in developing adaptable and efficient XAI systems that offer human-comprehensible explanations. The case study of the iSee platform serves as a concrete example of how CBR can be seamlessly melded with XAI to address complex decision-making processes, ensuring that AI systems remain understandable to a diverse array of users [rev] [92]. Selecting the optimal explainer from a vast array is challenging, requiring consideration of various factors. The iSee platform addresses this by employing CBR and leveraging user feedback. This paper also discusses how iSee provides AI designers with tools to access and leverage past explanation experiences, aiming to establish itself as a premier resource for fostering trust in AI. This research paves the way for future advancements, suggesting a promising direction for achieving more transparent and trustworthy AI systems within industry and academia. As the demand for XAI continues to grow, the synergy between CBR and XAI highlighted in this study underscores the critical role of interdisciplinary research in bridging the gap between advanced technological capabilities and human-centric computing needs.

Acknowledgment

This research is funded by the iSee project, a CHIST-ERA project financed by the European Union, which received funding for Ireland from the Irish Research Council under grant number CHIST-ERA-2019-iSee (with support from Science Foundation Ireland under Grant number 12/RC/2289-P2 at Insight the SFI Research Centre for Data Analytics at UCC, which is co-funded under the European Regional Development Fund), for Spain from the MCIN/AEI and European Union “Next Generation EU/PRTR” under grant number PCI2020-120720-2 and the UK from EPSRC under grant number EP/V061755/1.

References

[1] Y. K. Dwivedi, L. Hughes, E. Ismagilova, G. Aarts, C. Coombs, T. Crick, Y. Duan, R. Dwivedi, J. Edwards, A. Eirug, et al., Artificial intelligence (ai): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy, International Journal of Information Management 57 (2021) 101994.

[2] V. Hassija, V. Chamola, A. Mahapatra, A. Singal, D. Goel, K. Huang, S. Scardapane, I. Spinelli, M. Mahmud, A. Hussain, Interpreting black-box models: a review on explainable artificial intelligence, Cognitive Computation 16 (1) (2024) 45–74.

[3] B. López, Case-based reasoning: a concise introduction, Springer Nature, 2022.

[4] M. T. Keane, E. M. Kenny, How case-based reasoning explains neural networks: A theoretical analysis of xai using post-hoc explanation-by-example from a survey of ann-cbr twin-systems, in: Case-Based Reasoning Research and Development: 27th International Conference, ICCBR 2019, Otzenhausen, Germany, September 8–12, 2019, Proceedings 27, Springer, 2019, pp. 155–171.

[5] R. Weber, M. Shrestha, A. J. Johs, Knowledge-based xai through cbr: There is more to explanations than models can tell, arXiv preprint arXiv:2108.10363 (2021).

[6] J. M. Schoenborn, R. O. Weber, D. W. Aha, J. Cassens, K.-D. Althoff, Explainable case-based reasoning: A survey, in: AAAI-21 Workshop Proceedings, 2021.

[7] L. Gates, D. Leake, Evaluating cbr explanation capabilities: Survey and next steps, in: ICCBR Workshops, 2021, pp. 40–51.

[8] N. E. Zamri, M. A. Mansor, M. S. M. Kasihmuddin, S. S. Sidik, A. Alway, N. A. Romli, Y. Guo, S. Z. M. Jamaludin, A modified reverse-based analysis logic mining model with weighted random 2 satisfiability

logic in discrete hopfield neural network and multi-objective training of modified niched genetic algorithm, Expert Systems with Applications 240 (2024) 122307.

[9] F. Cabitza, A. Campagner, G. Malgieri, C. Natali, D. Schneeberger, K. Stoeger, A. Holzinger, Quod erat demonstrandum?-towards a typology of the concept of explanation for the design of explainable ai, Expert Systems with Applications 213 (2023) 118888.

[10] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, et al., Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai, Information fusion 58 (2020) 82–115.

[11] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, D. Pedreschi, A survey of methods for explaining black box models, ACM computing surveys (CSUR) 51 (5) (2018) 1–42.

[12] S. Chari, O. Seneviratne, D. M. Gruen, M. A. Foreman, A. K. Das, D. L. McGuinness, Explanation ontology: a model of explanations for user-centered ai, in: International Semantic Web Conference, Springer, 2020, pp. 228–243.

[13] M. Langer, D. Oster, T. Speith, H. Hermanns, L. Kästner, E. Schmidt, A. Sesing, K. Baum, What do we want from explainable artificial intelligence (xai)?–a stakeholder perspective on xai and a conceptual model guiding interdisciplinary xai research, Artificial Intelligence 296 (2021) 103473.

[14] S. Mohseni, N. Zarei, E. D. Ragan, A multidisciplinary survey and framework for design and evaluation of explainable ai systems, ACM Transactions on Interactive Intelligent Systems (TiiS) 11 (3-4) (2021) 1–45.

[15] A. F. Markus, J. A. Kors, P. R. Rijnbeek, The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies, Journal of biomedical informatics 113 (2021) 103655.

[16] M. Saarela, S. Jauhiainen, Comparison of feature importance measures as explanations for classification models, SN Applied Sciences 3 (2021) 1–12.

[17] I. Stepin, J. M. Alonso, A. Catala, M. Pereira-Fariña, A survey of contrastive and counterfactual explanation generation methods for explainable artificial intelligence, IEEE Access 9 (2021) 11974–12001.

[18] A. Wijekoon, N. Wiratunga, I. Nkisi-Orji, C. Palihawadana, D. Corsar, K. Martin, How close is too close? the role of feature attributions in discovering counterfactual explanations, in: International Conference on Case-Based Reasoning, Springer, 2022, pp. 33–47.

[19] E. Delaney, Case-based explanation for black-box time series and image models with applications in smart agriculture, ICCBR Doctoral Consortium 2022 1613 (2022) 0073.

[20] D. Leake, D. Mcsherry, Introduction to the special issue on explanation in case-based reasoning, The Artificial Intelligence Review 24 (2) (2005) 103.

[21] J. van der Waa, E. Nieuwburg, A. Cremers, M. Neerincx, Evaluating xai: A comparison of rule-based and example-based explanations, Artificial Intelligence 291 (2021) 103404.

[22] S. H. El-Sappagh, M. Elmogy, Case based reasoning: Case representation methodologies, International Journal of Advanced Computer Science and Applications 6 (11) (2015) 192–208.

[23] R. Bergmann, G. Pews, W. Wilke, Explanation-based similarity: A unifying approach for integrating domain knowledge into case-based reasoning for diagnosis and planning tasks, in: Topics in Case-Based Reasoning: First European Workshop, EWCBR-93 Kaiserslautern, Germany, November 1–5, 1993 Selected Papers 1, Springer, 1994, pp. 182–196.

[24] A. J. Cañas, D. B. Leake, A. G. Maguitman, Combining concept mapping with cbr: Towards experience-based support for knowledge modeling., in: FLAIRS Conference, Citeseer, 2001, pp. 286–290.

[25] R. L. De Mantaras, D. McSherry, D. Bridge, D. Leake, B. Smyth, S. Craw, B. Faltings, M. L. Maher, M. T COX, K. Forbus, et al., Retrieval, reuse, revision and retention in case-based reasoning, The Knowledge Engineering Review 20 (3) (2005) 215–240.

[26] W. Wilke, R. Bergmann, Techniques and knowledge used for adaptation during case-based problem solving, in: Tasks and Methods in Applied Artificial Intelligence: 11th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems IEA-98-AIE Benicàssim, Castellón, Spain, June 1–4, 1998 Proceedings, Volume II 11, Springer, 1998, pp. 497–506.

[27] B. Chebel-Morello, M. K. Haouchine, N. Zerhouni, Case-based maintenance: Structuring and incrementing the case base, Knowledge-Based Systems 88 (2015) 165–183.

[28] K. Sokol, P. Flach, One explanation does not fit all: The promise of interactive explanations for machine learning transparency, KI-Künstliche Intelligenz 34 (2) (2020) 235–250.

[29] I. Watson, F. Marir, Case-based reasoning: A review, The knowledge engineering review 9 (4) (1994) 327–354.

[30] I. Tiddi, M. d'Aquin, E. Motta, An ontology design pattern to define explanations, in: Proceedings of the 8th International Conference on Knowledge Capture, 2015, pp. 1–8.

[31] I. Padhiar, O. Seneviratne, S. Chari, D. Gruen, D. L. McGuinness, Semantic modeling for food recommendation explanations, in: 2021 IEEE 37th International Conference on Data Engineering Workshops (ICDEW), IEEE, 2021, pp. 13–19.

[32] M. Caro-Martínez, G. Jiménez-Díaz, J. A. Recio-García, Conceptual modeling of explainable recommender systems: an ontological formalization to guide their design and development, Journal of Artificial Intelligence Research 71 (2021) 557–589.

[33] R. Roscher, B. Bohn, M. F. Duarte, J. Garcke, Explainable machine learning for scientific insights and discoveries, Ieee Access 8 (2020) 42200–42216.

[34] D. Spoladore, M. Sacco, A. Trombetta, A review of domain ontologies for disability representation, Expert Systems with Applications (2023) 120467.

[35] C. Panigutti, A. Perotti, D. Pedreschi, Doctor xai: an ontology-based approach to black-box sequential data classification explanations, in: Proceedings of the 2020 conference on fairness, accountability, and transparency, 2020, pp. 629–639.

[36] R. Confalonieri, T. Weyde, T. R. Besold, F. M. del Prado Martín, Using ontologies to enhance human understandability of global post-hoc explanations of black-box models, Artificial Intelligence 296 (2021) 103471.

[37] S. R. Islam, W. Eberle, S. K. Ghafoor, M. Ahmed, Explainable artificial intelligence approaches: A survey, arXiv preprint arXiv:2101.09429 (2021).

[38] D. M. R. LOPEZ, D. MCSHERRY, D. BRIDGE, D. LEAKE, B. SMYTH, S. CRAW, B. FALTINGS, M. L. MAHER, M. T. COX, K. FORBUS, et al., Retrieval, reuse, revision and retention in case-based reasoning, The Knowledge Engineering Review 20 (3) (2005) 215–240.

[39] G. Finnie, Z. Sun, Similarity and metrics in case-based reasoning, International journal of intelligent systems 17 (3) (2002) 273–287.

[40] D. Wang, Q. Yang, A. Abdul, B. Y. Lim, Designing theory-driven user-centric explainable ai, in: Proceedings of the 2019 CHI conference on human factors in computing systems, 2019, pp. 1–15.

[41] P. Marín-Veites, K. Bach, Explaining cbr systems through retrieval and similarity measure visualizations: A case study, in: International Conference on Case-Based Reasoning, Springer, 2022, pp. 111–124.

[42] P. Cunningham, A taxonomy of similarity mechanisms for case-based reasoning, IEEE Transactions on Knowledge and Data Engineering 21 (11) (2008) 1532–1543.

[43] L. Lü, T. Zhou, Link prediction in complex networks: A survey, Physica A: statistical mechanics and its applications 390 (6) (2011) 1150–1170.

[44] M. Caro-Martínez, G. Jiménez-Díaz, J. A. Recio-Garcia, A graph-based approach for minimising the knowledge requirement of explainable recommender systems, Knowledge and Information Systems (2023) 1–31.

[45] A. Günay, P. Yolum, Structural and semantic similarity metrics for web service matchmaking, in: International Conference on Electronic Commerce and Web Technologies, Springer, 2007, pp. 129–138.

[46] H. Xu, Y. Wei, Y. Cai, B. Xing, Knowledge graph and cbr-based approach for automated analysis of bridge operational accidents: Case representation and retrieval, PLoS one 18 (11) (2023) e0294130.

[47] A. Abou Assali, D. Lenne, B. Debray, Case retrieval in ontology-based cbr systems, in: KI 2009: Advances in Artificial Intelligence: 32nd Annual German Conference on AI, Paderborn, Germany, September 15-18, 2009. Proceedings 32, Springer, 2009, pp. 564–571.

[48] K. D. Forbus, D. Gentner, K. Law, Mac/fac: A model of similarity-based retrieval, Cognitive science 19 (2) (1995) 141–205.

[49] J. M. Darias, M. Caro-Martínez, B. Díaz-Agudo, J. A. Recio-Garcia, Using case-based reasoning for capturing expert knowledge on explanation methods, in: International Conference on Case-Based Reasoning, Springer, 2022, pp. 3–17.

[50] B. Smyth, M. T. Keane, Using adaptation knowledge to retrieve and adapt design cases, Knowledge-based systems 9 (2) (1996) 127–135.

[51] E. Plaza, J.-L. Arcos, Constructive adaptation, in: European Conference on Case-Based Reasoning, Springer, 2002, pp. 306–320.

[52] W. Wilke, I. Vollrath, K.-D. Althoff, R. Bergmann, A framework for learning adaptation knowledge based on knowledge light approaches, in: Proceedings of the fifth german workshop on case-based reasoning, 1997, pp. 235–242.

[53] F. Fdez-Riverola, J. M. Corchado, J. M. Torres, An automated hybrid cbr system for forecasting, in: European Conference on Case-Based Reasoning, Springer, 2002, pp. 519–533.

[54] B. Smyth, Case-base maintenance, in: International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, Springer, 1998, pp. 507–516.

[55] E. Lupiani, J. M. Juarez, J. Palma, Evaluating case-base maintenance algorithms, Knowledge-Based Systems 67 (2014) 180–194.

[56] L. Cummins, D. Bridge, Maintenance by a committee of experts: The mace approach to case-base maintenance, in: International Conference on Case-Based Reasoning, Springer, 2009, pp. 120–134.

[57] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE transactions on information theory 13 (1) (1967) 21–27.

[58] B.-R. Dai, S.-M. Hsu, An instance selection algorithm based on reverse nearest neighbor, in: Advances in Knowledge Discovery and Data Mining: 15th Pacific-Asia Conference, PAKDD 2011, Shenzhen, China, May 24-27, 2011, Proceedings, Part I 15, Springer, 2011, pp. 1–12.

[59] K. Göbel, C. Niessen, S. Seufert, U. Schmid, Explanatory machine learning for justified trust in human-ai collaboration: Experiments on file deletion recommendations, Frontiers in Artificial Intelligence 5 (2022) 919534.

[60] E. C. Tsang, X. Wang, An approach to case-based maintenance: Selecting representative cases, International Journal of Pattern Recognition and Artificial Intelligence 19 (01) (2005) 79–89.

[61] T. Kulesza, S. Stumpf, M. Burnett, I. Kwan, Tell me more? the effects of mental model soundness on personalizing an intelligent agent, in: Proceedings of the sigchi conference on human factors in computing systems, 2012, pp. 1–10.

[62] A. Smith-Renner, R. Fan, M. Birchfield, T. Wu, J. Boyd-Graber, D. S. Weld, L. Findlater, No explainability without accountability: An empirical study of explanations and feedback in interactive ml, in: Proceedings of the 2020 chi conference on human factors in computing systems, 2020, pp. 1–13.

[63] J. Schneider, J. Handali, Personalized explanation in machine learning: A conceptualization, arXiv preprint arXiv:1901.00770 (2019).

[64] G. Adomavicius, A. Tuzhilin, Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions, IEEE transactions on knowledge and data engineering 17 (6) (2005) 734–749.

[65] J. L. Jorro-Aragoneses, M. Caro-Martínez, B. Díaz-Agudo, J. A. Recio-García, A user-centric evaluation to generate case-based explanations using formal concept analysis, in: International Conference on Case-Based Reasoning, Springer, 2020, pp. 195–210.

[66] D. J. Hilton, Conversational processes and causal explanation., Psychological Bulletin 107 (1) (1990) 65.

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

[68] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen, et al., Guidelines for human-ai interaction, in: Proceedings of the 2019 chi conference on human factors in computing systems, 2019, pp. 1–13.

[69] M. Caro-Martinez, J. A. Recio-Garcia, G. Jimenez-Diaz, An algorithm independent case-based explanation approach for recommender systems using interaction graphs, in: Case-Based Reasoning Research and Development: 27th International Conference, ICCBR 2019, Otzenhausen, Germany, September 8–12, 2019, Proceedings 27, Springer, 2019, pp. 17–32.

[70] M. Nick, Reducing the case acquisition and maintenance bottleneck with user-feedback-driven case base maintenance., in: FLAIRS Conference, 2006, pp. 376–382.

[71] Q. V. Liao, M. Pribić, J. Han, S. Miller, D. Sow, Question-driven design process for explainable ai user experiences, arXiv preprint arXiv:2104.03483 (2021).

[72] Y. Ramon, T. Vermeire, O. Toubia, D. Martens, T. Evgeniou, Understanding consumer preferences for explanations generated by xai algorithms, arXiv preprint arXiv:2107.02624 (2021).

[73] C. E. Sosa-Espadas, M. G. Orozco-del Castillo, N. Cuevas-Cuevas, J. A. Recio-Garcia, Irex: Iterative refinement and explanation of classification models for tabular datasets, SoftwareX 23 (2023) 101420.

[74] A. M. Saghiri, S. M. Vahidipour, M. R. Jabbarpour, M. Sookhak, A. Forestiero, A survey of artificial intelligence challenges: Analyzing the definitions, relationships, and evolutions, Applied sciences 12 (8) (2022) 4054.

[75] A. Alam, Developing a curriculum for ethical and responsible ai: A university course on safety, fairness, privacy, and ethics to prepare next generation of ai professionals, in: Intelligent Communication Technologies and Virtual Mobile Networks, Springer, 2023, pp. 879–894.

[76] G. Ras, M. van Gerven, P. Haselager, Explanation methods in deep learning: Users, values, concerns and challenges, Explainable and interpretable models in computer vision and machine learning (2018) 19–36.

[77] E. Soares, P. Angelov, Fair-by-design explainable models for prediction of recidivism, arXiv preprint arXiv:1910.02043 (2019).

[78] S. Sharma, J. Henderson, J. Ghosh, Certifai: Counterfactual explanations for robustness, transparency, interpretability, and fairness of artificial intelligence models, arXiv preprint arXiv:1905.07857 (2019).

[79] The Royal Society, Explainable AI: The Basics, https://royalsociety.org/-/media/policy/projects/explainable-ai/AI-and-interpretability-policy-briefing.pdf, accessed: [11 February 2024] (2019).

[80] R. Shokri, M. Strobel, Y. Zick, On the privacy risks of model explanations, in: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, 2021, pp. 231–241.

[81] W. Saeed, C. Omlin, Explainable ai (xai): A systematic meta-survey of current challenges and future opportunities, Knowledge-Based Systems 263 (2023) 110273.

[82] S. Chakraborty, R. Tomsett, R. Raghavendra, D. Harborne, M. Alzantot, F. Cerutti, M. Srivastava, A. Preece, S. Julier, R. M. Rao, et al.,

Interpretability of deep learning models: A survey of results, in: 2017 IEEE smartworld, ubiquitous intelligence & computing, advanced & trusted computed, scalable computing & communications, cloud & big data computing, Internet of people and smart city innovation (smartworld/SCALCOM/UIC/ATC/CBDcom/IOP/SCI), IEEE, 2017, pp. 1–6.

[83] A. Pant, R. Hoda, S. V. Spiegler, C. Tantithamthavorn, B. Turhan, Ethics in the age of ai: An analysis of ai practitioners' awareness and challenges, ACM Transactions on Software Engineering and Methodology (2023).

[84] A. Wijekoon, N. Wiratunga, C. Palihawadana, I. Nkisi-Orji, D. Corsar, K. Martin, isee: Intelligent sharing of explanation experience by users for users, in: Companion Proceedings of the 28th International Conference on Intelligent User Interfaces, 2023, pp. 79–82.

[85] J. Klaise, A. Van Looveren, G. Vacanti, A. Coca, Alibi explain: Algorithms for explaining machine learning models, The Journal of Machine Learning Research 22 (1) (2021) 8194–8200.

[86] H. Baniecki, W. Kretowicz, P. PiÄ, J. WiĹ, et al., Dalex: responsible machine learning with interactive explainability and fairness in python, Journal of Machine Learning Research 22 (214) (2021) 1–7.

[87] T. Fel, L. Hervier, D. Vigouroux, A. Poche, J. Plakoo, R. Cadene, M. Chalvidal, J. Colin, T. Boissin, L. Béthune, et al., Xplique: A deep learning explainability toolbox, arXiv preprint arXiv:2206.04394 (2022).

[88] A. Wijekoon, N. Wiratunga, A user-centred evaluation of discern: Discovering counterfactuals for code vulnerability detection and correction, Knowledge-Based Systems 278 (2023) 110830.

[89] B. Bayrak, K. Bach, Pertcf: A perturbation-based counterfactual generation approach, in: International Conference on Innovative Techniques and Applications of Artificial Intelligence, Springer, 2023, pp. 174–187.

[90] C. E. Sosa-Espadas, M. Cetina-Aguilar, J. A. Soladrero, J. M. Darias, E. E. Brito-Borges, N. L. Cuevas-Cuevas, M. G. Orozco-del Castillo,

Irex: A reusable process for the iterative refinement and explanation of classification models., in: ICCBR Workshops, 2022, pp. 241–245.

[91] M. F. Valdez-Ávila, C. Bermejo-Sabbagh, B. Diaz-Agudo, M. G. Orozco-del Castillo, J. A. Recio-Garcia, Cbr-fox: A case-based explanation method for time series forecasting models, in: International Conference on Case-Based Reasoning, Springer, 2023, pp. 200–214.

[92] A. Wijekoon, N. Wiratunga, K. Martin, D. Corsar, I. Nkisi-Orji, C. Palihawadana, D. Bridge, P. Pradeep, B. D. Agudo, M. Caro-Martínez, Cbr driven interactive explainable ai, in: International Conference on Case-Based Reasoning, Springer, 2023, pp. 169–184.

[93] M. Caro-Martínez, A. Wijekoon, J. A. Recio-García, D. Corsar, I. Nkisi-Orji, Conceptual modelling of explanation experiences through the iseeonto ontology., in: CEUR Workshop Proceedings, Vol. 3389, CEUR Workshop Proceedings, 2023.

[94] I. Nkisi-Orji, N. Wiratunga, C. Palihawadana, J. A. Recio-García, D. Corsar, Clood cbr: Towards microservices oriented case-based reasoning, in: International Conference on Case-Based Reasoning, Springer, 2020, pp. 129–143.

[95] M. C. Suárez-Figueroa, A. Gómez-Pérez, M. Fernández-López, The neon methodology for ontology engineering, in: Ontology engineering in a networked world, Springer, 2011, pp. 9–34.

[96] M. C. Suárez-Figueroa, A. Gómez-Pérez, Ontology requirements specification, in: Ontology Engineering in a Networked World, Springer, 2011, pp. 93–106.

[97] I. Nkisi-Orji, C. Palihawadana, N. Wiratunga, D. Corsar, A. Wijekoon, Adapting semantic similarity methods for case-based reasoning in the cloud, in: International Conference on Case-Based Reasoning, Springer, 2022, pp. 125–139.

[98] M. Caro-Martinez, J. M. Darias, B. Diaz-Agudo, J. A. Recio-Garcia, iseee3—the explanation experiences editor, SoftwareX 21 (2023) 101311.

[99] J. M. Darias, B. Díaz-Agudo, J. A. Recio-Garcia, A systematic review on model-agnostic xai libraries., in: ICCBR Workshops, 2021, pp. 28–39.

[100] I. Nkisi-Orji, C. Palihawadana, N. Wiratunga, A. Wijekoon, D. Corsar, Failure-driven transformational case reuse of explanation strategies in cloodcbr, in: International Conference on Case-Based Reasoning, Springer, 2023, pp. 279–293.

[101] F. Sørmo, J. Cassens, A. Aamodt, Explanation in case-based reasoning–perspectives and goals, Artificial Intelligence Review 24 (2005) 109–143.

[102] T. R. Roth-Berghofer, Explanations and case-based reasoning: Foundational issues, in: European Conference on Case-Based Reasoning, Springer, 2004, pp. 389–403.

Author list (emails and ORCIDs)

Preeja Pradeep	ppradeep@ucc.ie	https://orcid.org/0000-0001-6708-4964
Marta Caro-Martínez	martcaro@ucm.es	https://orcid.org/0000-0001-5239-8207
Anjana Wijekoon	a.wijekoon1@rgu.ac.uk	https://orcid.org/0000-0003-3848-3100

Highlights

- Investigate CBR's role in enhancing AI explainability.
- Principles for developing CBR-XAI platforms are highlighted.
- Novel integration of CBR with XAI is showcased through a CBR-driven XAI platform.

Conceptualization

Preeja Pradeep

Marta Caro-Martínez

Anjana Wijekoon

Methodology

Preeja Pradeep

Marta Caro-Martínez

Anjana Wijekoon

Software

Preeja Pradeep

Marta Caro-Martínez

Anjana Wijekoon

Validation

Preeja Pradeep

Marta Caro-Martínez

Anjana Wijekoon

Formal analysis

Preeja Pradeep

Marta Caro-Martínez

Anjana Wijekoon

Investigation

Preeja Pradeep

Marta Caro-Martínez

Anjana Wijekoon

Resources

Preeja Pradeep

Marta Caro-Martínez

Anjana Wijekoon

Data Curation

Preeja Pradeep

Marta Caro-Martínez

Anjana Wijekoon

Writing - Original Draft

Preeja Pradeep

Marta Caro-Martínez

Anjana Wijekoon

Writing - Review & Editing

Preeja Pradeep

Marta Caro-Martínez

Anjana Wijekoon

Visualization

Preeja Pradeep

Marta Caro-Martínez

Anjana Wijekoon

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: