iSee: intelligence sharing of explanation experience of users for users.

WIJEKOON, A., WIRATUNGA, N., PALIHAWADANA, C., NKISI-ORJI, I., CORSAR, D. and MARTIN, K.

2023

iSee: Intelligent Sharing of Explanation Experience by Users for Users

ANJANA WIJEKOON, NIRMALIE WIRATUNGA, CHAMATH PALIHAWADANA, IKECKUKWU NKISI-ORJI, DAVID CORSAR, and KYLE MARTIN, Robert Gordon University, Scotland

The right to obtain an explanation of the decision reached by an Artificial Intelligence (AI) model is now an EU regulation. Different stakeholders of an AI system (e.g. managers, developers, auditors, etc.) may have different background knowledge, competencies and goals, thus requiring different kinds of interpretations and explanations. Fortunately, there is a growing armoury of tools to interpret ML models and explain their predictions, recommendations and diagnoses which we will refer to collectively as explanation strategies. As these explanation strategies mature, practitioners will gain experience that helps them know which strategies to deploy in different circumstances. What is lacking, and is addressed by iSee, is capturing, sharing and re-using explanation strategies based on past positive experiences. The goal of the iSee platform is to improve every user's experience of AI, by harnessing experiences and best practices in Explainable AI.

CCS Concepts: • **Human-centered computing** → *Natural language interfaces*; • **Computing methodologies** → Knowledge representation and reasoning.

ACM Reference Format:
Anjana Wijekoon, Nirmalie Wiratunga, Chamath Palihawadana, Ikeckukwu Nkisi-Orji, David Corsar, and Kyle Martin. 2023. iSee: Intelligent Sharing of Explanation Experience by Users for Users. In *28th International Conference on Intelligent User Interfaces (IUI '23 Companion), March 27–31, 2023, Sydney, NSW, Australia.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3581754.3584137

1 INTRODUCTION

Studies on interpretable Artificial Intelligence (AI) and regulatory guidelines have highlighted the need for providing explanations for AI decisions as AI techniques have become progressively opaque [4, 5]. Many industries that have incorporated AI into their processes are now facing the challenge of satisfying these guidelines and many do not have the resources to conduct the necessary research and development to implement explanation techniques that meet the needs of their end users. iSee addresses this "cold start" challenge by proposing to reuse explanation experiences. An "explanation experience" refers to an applied use-case of an AI system where an explanation strategy of one or more explanation techniques have been implemented to address end-user explanation needs and the outcome has been formally evaluated. The proposed iSee platform makes use of these past experiences to recommend appropriate explanation strategies for new use cases that are looking to implement explainable AI within its system.

The iSee platform brings together eXplainable Artificial Intelligence (XAI) and Human Computer Interaction (HCI) researchers, AI system designers, and end-users to create explanation experiences. The Case-based Reasoning life-cycle, detailed in [1], forms the back-end engine that retrieves, reuses, revises and retains explanation experiences as cases. Figure 1 depicts the iSee CBR life-cycle and Figure 2 depicts the case structure. The main focus of this demonstration is to showcase the tools provided by the iSee platform for different stakeholders to create explanation experiences and facilitate the CBR processes.

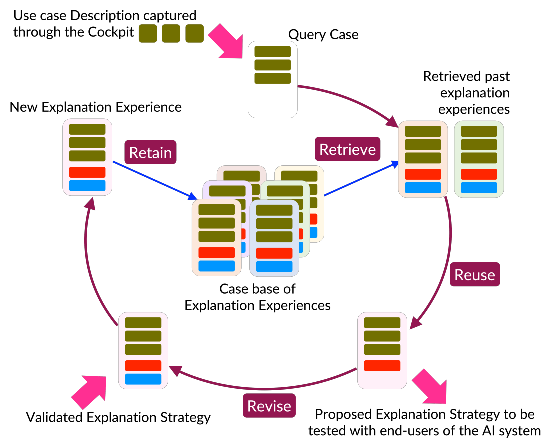

Fig. 1. iSee System

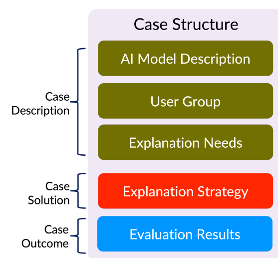

Fig. 2. An explanation experience case

In iSee we refer to AI system developers as design users and AI system consumers as end-users. Design users use the iSee platform to describe their AI system, and use the iSee Case-based Reasoning (CBR) recommender to retrieve and re-use explanation strategies. CBR recommender makes evidence-based recommendations where it retrieves past cases of similar XAI systems that received positive feedback from the end-users. Once a use case is assigned an explanation strategy, end-users of the AI System will use the iSee platform to interact with the explanation strategy and create explanation experiences. Similar to previous work from [7, 9] and [14] iSee views the explanation experience to be interactive and conversational which is an iterative process of the following steps: 1) the end-user expressing their explanation needs; 2) the chatbot presenting explanations by executing the strategy recommended to satisfy expressed needs; and 3) the end-user evaluating their experience.

The iSee platform also offers tools for both XAI/HCI practitioners to contribute tools and promote open science and reusability. Researchers who develop and empirically evaluate XAI methods can contribute their algorithms to be re-used by industry use cases as explanation strategies. For this, iSee offers tools to standardise the description of XAI methods using a set of semantic attributes. This standardised description facilitates linking of XAI implementations with the iSee platform for generating explanations for a specific use case. All components are formalised by the iSeeOnto ontology [1], which facilitates the interoperability of the many heterogeneous components that underpin the iSee platform.

[1] https://www.w3id.org/iSeeOnto/explanationexperience

The iSee project is associated with five use cases from four industry partners [2]. We followed a user-participatory methodology for the design and implementation of the iSee platform which was driven by iterative feedback from design users of these use cases. iSee is an ongoing project and the platform is currently being evaluated by the end users of the five use cases.The rest of the paper describes the tools offered in the iSee platform for each stakeholder and their implementation details.

2 ISEE TOOLS FOR XAI/HCI PRACTITIONERS

iSee implements tools to bring together three types of artefacts from XAI/HCI Practitioners: XAI methods, XAI evaluation metrics and complete use cases. Artefacts generated by each tool are depicted in Figure 3.

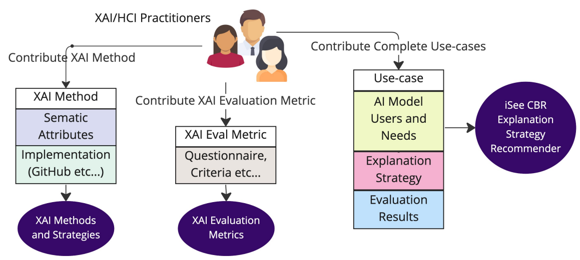

Fig. 3. iSee tools for XAI/HCI practitioners

2.1 Contribute XAI methods

iSee provides tools for XAI Researchers to integrate existing XAI methods with the iSee platform. First, an XAI method is added to the XAI methods repository maintained on GitHub [3] and next semantic attributes of the XAI method are added to the iSee platform. The semantic attributes can be found here [4] as formalised in iSeeOnto. The goal of this tool is to create a formalised repository of the XAI method for reuse. iSee platform currently contains eighteen XAI methods added from literature including common methods like LIME [10], SHAP [8], Anchors [11], ALE [2] and GradCam [13].

2.2 Contribute XAI evaluation metrics

Evaluation and feedback are key components of creating explanation experiences. In literature, evaluation of explanations comprises of a questionnaire with multiple questions aimed at evaluating a specific dimension such as goodness, satisfaction, change in user's mental model and trust [6, 16]. The iSee platform facilitates the creation and sharing of evaluation metrics for reuse. Currently, iSee contains four metrics empirically validated in literature: Hoffman Explanation Goodness Checklist (7 questions); Hoffman Explanation Satisfaction Scale (8 questions); Hoffman Trust Scale (7 questions) [6]; and Cahour-Forzy Trust Scale [3] (4 questions).

2.3 Contribute complete use cases

XAI and HCI practitioners can also contribute complete use cases to the iSee platform. A complete use case consists of a description of the AI model, target user group(s), their explanation needs, the explanation strategy that was developed to

[2] https://isee4xai.com/usecases/
[3] https://github.com/isee4xai/ExplainerLibraries
[4] https://www.w3id.org/iSeeOnto/explainer

provide explanations and end-user feedback received. The explanation strategy can consist of one or more XAI methods that are already added to the platform. The use case should also describe the metric used to evaluate the explanation strategy with the target user group(s) to demonstrate its suitability and applicability to the use case. A use case that has demonstrated a positive experience with its target users will be added to the case base of the CBR recommender such that its explanation strategy can be recommended to a new case that is similar to it in the AI model and target user group(s). Although there is an abundance of use cases found in literature they are often incomplete due to the lack of empirical evidence from users. Currently, we have extracted 7 complete use cases from literature to include in the iSee CBR recommender [5]. The lack of complete cases is one of the main limitations of the iSee recommender and the expectation is as the platform grows, complete use cases with empirical evidence are created within the iSee platform and will be retained in the case base for future recommendations.

3 ISEE TOOLS FOR DESIGN USERS

A design user is an AI system developer and/or manager who is looking to implement explainability into their AI system. This can be motivated by either complying with regulations or simply improving end-user trust and experience of the system. Use case creation is depicted in Figure 4. Given the AI model details and target user group(s) of the case, the iSee platform recommends an explanation strategy that has created positive explanation experiences in the past for similar cases. The design user has the opportunity to modify the use case and get a different recommendation or edit the recommended explanation strategy itself as needed. The design user can also modify the explanation strategy as implementation requirements, end-user needs change as indicated in end-user feedback.

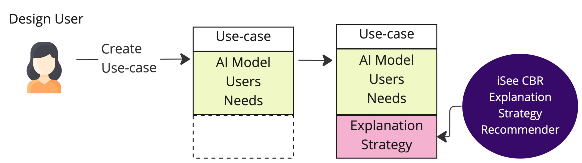

Fig. 4. iSee tools for the design user

4 ISEE TOOLS FOR END USERS

Once a use case has received an explanation strategy from the CBR recommender, it is ready to be tested by the end users. End users will access the conversational interface *iSeeChatbot* where they are able to express their explanation needs, receive explanations and evaluate their experience. Users can modify their explanation needs to receive multiple explanations (multi-shot) if the explanation strategy supports more than one XAI method. The *iSeeChatbot* is driven by a Behaviour Tree that models the dialogue pathways [15]. The use case created by the design user instantiates the dialogue model, which dynamically adapts to the persona and the explanation need indicated by the end user during the conversation to select the appropriate explanation strategy. The XAI methods contributed by researchers are used to execute XAI methods in the selected strategy and XAI evaluation metrics from researchers are used to evaluate the experience. The iSeeChatbot is intended as a testing tool where the design user can select a sample of end users to test a recommended strategy. Iterative feedback from testing can be used to identify new explanation needs and modify the explanation strategy accordingly.

[5]https://github.com/isee4xai/iSeeCases

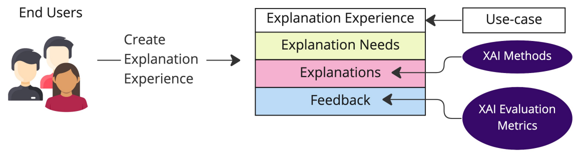

Fig. 5. iSee tools for the end user

5 SYSTEM DESIGN AND IMPLEMENTATION

iSee followed a participatory design approach [12] where co-design and co-creation activities informed the current system design. In the initial stage, user-centred co-design activities were carried out to gather information for the case description. The gathered data were analysed to understand the concerns and implications around data sharing for each industry use case. Based on this analysis, a case description was created that effectively captures the necessary information for recommending and implementing explanation strategies while protecting confidential data. The following phase involved iterative feedback from the use cases, focusing on how to capture the AI model, target users, and their explanation needs.

The iSee platform has a cloud-based architecture; technologies and inter-dependencies of components are depicted in Figure 6. Access to the iSee platform can be requested from the iSee website and accessed via https://cockpit-dev.isee4xai.com/. Some screenshots of the iSee platform tools are captured in Figure 7.

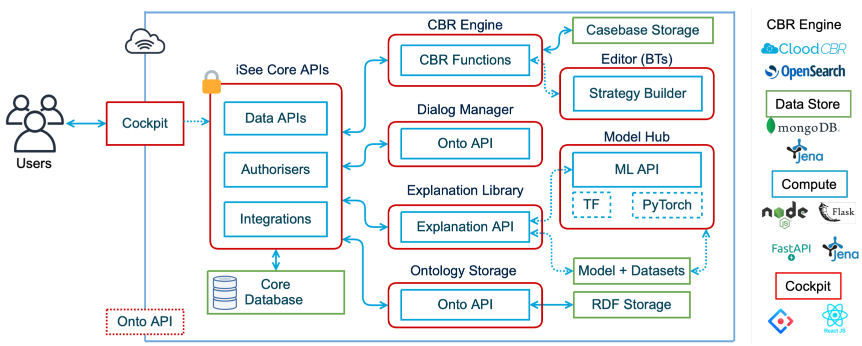

Fig. 6. iSee platform architecture

6 CONCLUSION

This paper presents a demonstration of the iSee platform. iSee is an ongoing project aimed at improving the user experience of AI by harnessing experiences and best practices in Explainable AI. To this end, iSee brings together research and developments from XAI/HCI practitioners and experience from industry partners to promote XAI reuse in the real world. This paper presented a high-level overview of tools available for XAI/HCI practitioners, AI system developers/managers and AI decision consumers. Currently, the iSee platform is tested by end users from four industry partners. iSee is planned to be released as an open platform for promoting the reuse of XAI methodologies and best practices in 2024.

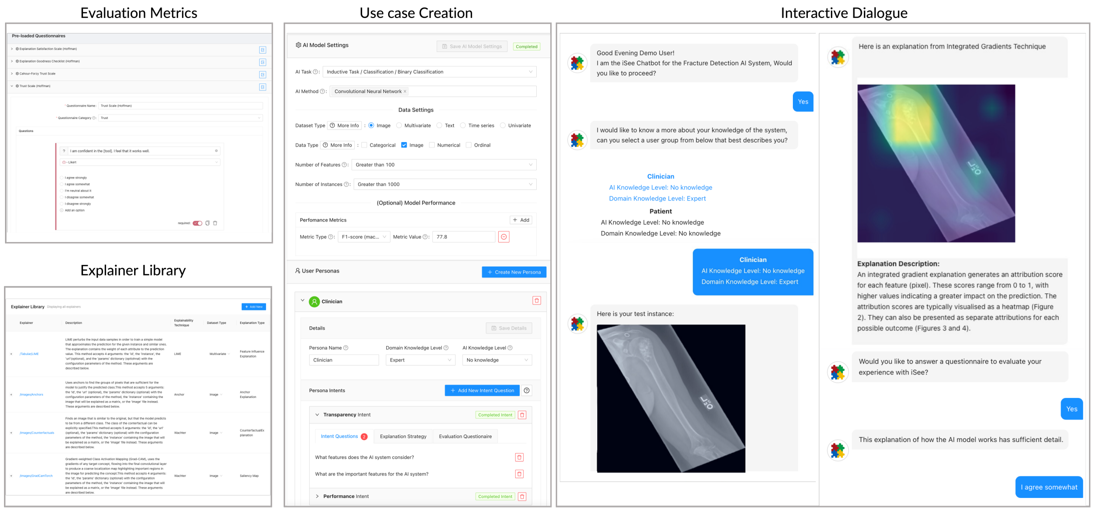

Fig. 7. iSee platform screenshots

ACKNOWLEDGMENTS

iSee is an EU CHIST-ERA project (https://www.chistera.eu/projects/isee) which received funding for the UK from EPSRC under the grant number EP/V061755/1.

REFERENCES

[1] Agnar Aamodt and Enric Plaza. 1994. Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI communications* 7, 1 (1994), 39–59.

[2] Daniel W Apley and Jingyu Zhu. 2020. Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82, 4 (2020), 1059–1086.

[3] Béatrice Cahour and Jean-François Forzy. 2009. Does projection into use improve trust and exploration? An example with a cruise control system. *Safety science* 47, 9 (2009), 1260–1270.

[4] Martin Ebers. 2020. Regulating Explainable AI in the European Union. An Overview of the Current Legal Framework (s). *An Overview of the Current Legal Framework (s)(August 9, 2021). Liane Colonna/Stanley Greenstein (eds.), Nordic Yearbook of Law and Informatics* (2020).

[5] David Gunning and David Aha. 2019. DARPA's explainable artificial intelligence (XAI) program. *AI magazine* 40, 2 (2019), 44–58.

[6] Robert R Hoffman, Shane T Mueller, Gary Klein, and Jordan Litman. 2018. Metrics for explainable AI: Challenges and prospects. *arXiv preprint arXiv:1812.04608* (2018).

[7] Sophie F Jentzsch, Sviatlana Höhn, and Nico Hochgeschwender. 2019. Conversational interfaces for explainable AI: a human-centred approach. In *Explainable, Transparent Autonomous Agents and Multi-Agent Systems: First International Workshop, EXTRAAMAS 2019, Montreal, QC, Canada, May 13–14, 2019, Revised Selected Papers 1*. Springer, 77–92.

[8] Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems* 30 (2017).

[9] Prashan Madumal, Tim Miller, Liz Sonenberg, and Frank Vetere. 2019. A Grounded Interaction Protocol for Explainable Artificial Intelligence. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*. 1033–1041.

[10] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 1135–1144.

[11] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2018. Anchors: High-precision model-agnostic explanations. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 32.

[12] Douglas Schuler and Aki Namioka. 1993. *Participatory design: Principles and practices*. CRC Press.

[13] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*. 618–626.

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

[15] Anjana Wijekoon, David Corsar, and Nirmalie Wiratunga. 2022. Behaviour Trees for Conversational Explanation Experiences. *arXiv preprint arXiv:2211.06402* (2022).

[16] Jeffrey C Zemla, Steven Sloman, Christos Bechlivanidis, and David A Lagnado. 2017. Evaluating everyday explanations. *Psychonomic bulletin & review* 24, 5 (2017), 1488–1500.

Received 09 January 2023