

revAlsor Suggestions

revAlsor Response:

```
▼ {
  ▼ "Theme" : {
    "abstract" :
      "The article is well-written and provides a clear overview of the proposed method for evaluating human-AI decision-making systems. However, there are a few areas where improvements can be made to enhance readability and emphasize key points:

      1. Enhance References:
         - Provide references for the specific AI models used in the system (e.g., revenue prediction, analogues analysis, hypothesis testing).
         - Include references for the concepts of trust, interpretability, and explainability in the context of human-AI decision-making systems.

      2. Address Theme Violations:
         - Clarify the specific challenges and limitations of evaluating human-AI decision-making systems.
         - Discuss potential risks or ethical considerations associated with combining multiple AI models in decision-making.

      3. Include Missing Related Works:
         - Provide a brief overview of existing methods or approaches for evaluating human-AI decision-making systems.
         - Discuss any previous research on retrospective methods or end-user walkthroughs in the context of HCI and AI.

      4. Fill in Missing Information:
         - Explain the specific steps involved in the retrospective end-user walkthrough method.
         - Provide details on the financial decision-making context and how the system combines visual analytics and AI models.

      5. Clarify Explanations:
         - Clearly explain the purpose and benefits of the retrospective end-user walkthrough method.
         - Elaborate on the higher order cognitive processes and the role of AI model outputs in human-AI decision-making.

      6. Add Limitations Section:
         - Discuss any limitations or potential biases in the qualitative assessment with the 29 participants.
         - Address any limitations of the proposed method and potential challenges in its application.

      7. Any Other Suggestions:
         - Consider providing examples or case studies to illustrate the application of the method and its outcomes.
         - Emphasize the implications of the method for HCI practitioners and the potential impact on the design of human-AI interaction systems.

      Overall, the article provides a solid foundation for evaluating human-AI decision-making systems. By addressing the suggested improvements, the article can further enhance its readability and emphasize the key points of the proposed method."
```

"introduction" :
"1 INTRODUCTION

Human-AI decision-making can be defined as the paradigm in which human decision-makers are supported by one or more Artificial Intelligence (AI) models providing recommendations/predictions for a certain task [17]. In recent years, with the advances in AI and its pervasive use in multiple domains (e.g., healthcare, finance, justice, and recruiting), explainable AI (XAI) has become fundamental for humans to assess and make decisions based on AI models, especially for legal reasons [14, 35]. XAI can be defined as the field that aims to make the output of AI systems more understandable to humans [1]. According to DARPA (Defense Advanced Research Projects Agency), XAI aims at producing more explainable models, while maintaining a high level of prediction accuracy and enabling users to understand, appropriately trust, and effectively manage human AI partnership [1]. The explanation provided in XAI considers two main characteristics: complexity and domain. The complexity characteristic refers to matching the content (explanation) with the content consumer (user). The domain characteristic highlights the need to tailor the explanation to the domain in which the AI system is being used [14]. According to Wang et al. [34], explanations should satisfy three properties: (1) Improve understanding of the AI model; (2) Help identify model uncertainty; (3) Support calibrated trust in the model. These works highlight the strong interconnectedness between XAI and trust when AI models are used in decision-making systems.

As new requirements for Human-AI decision-making systems emerge (XAI, trust, uncertainty recognition, and planning future interactions), new methods and tools are needed to help HCI practitioners understand how to bridge the gap between humans and AI decision-making systems to realize the best of this human-AI partnership. However, in contrast to traditional HCI approaches, there is a lack of methods to specifically assess the interplay between different AI models [17], especially when recency bias may influence how people deal with recommendations [8, 11].

Decision-making is known to be influenced by cognitive processes and affective processes [11, 31]. Two of the most commonly used techniques that support gaining insights about these processes are Thinking-Aloud Protocol and Cognitive Walkthrough. Thinking-Aloud Protocol is a technique where participants are encouraged to verbalize their thoughts related to performing tasks while interacting with a user interface (UI) [19]. This technique is commonly employed in formative user studies, supporting the understanding of the rationale and task-solving strategies used by participants [18]. Moreover, it has two important variations: Concurrent Thinking-Aloud (CTA), where participants verbalize their thoughts while performing a task, and Retrospective Thinking-Aloud (RTA), where participants verbalize their thoughts while reviewing/replaying their tasks. However, performing CTA in certain tasks can be intrusive, impact the overall user experience, and influence results [5, 18, 22]. On the other hand, post-test questionnaires may fail to capture nuances of the overall interaction [5, 18, 22]. This complexity motivates the incorporation of the "walkthrough" to bridge this gap, using the system as an artifact to reduce cognitive load as an external representation [5]. Cognitive Walkthrough (CW) is a review method in which an HCI expert simulates performing tasks to provide insight into how users may interact with the UI [18]. In specific decision-making domains, subject matter experts (SMEs) play a key role. Hence, the method proposed in this work aims at incorporating this aspect by combining RTA with an SME-performed walkthrough in a debriefing phase. According to [18], debriefing provides a valuable opportunity for sharing thoughts that did not fit into the session, offering an opportunity for participants to share opinions and correct misunderstandings.

1.1 Contribution Statement

This work is part of a long-term project on human-AI decision-making that explores how SMEs combine multiple AI models in different domains including data science, discovery of new materials, and scientific workflows. While the project considers multiple domains, the user study and the system described in this paper relate to the finance domain. This work proposes a method to help HCI practitioners gain a better understanding of how people combine multiple AI model outputs, in a debriefing phase, given that verbalizing all the rationale during the interaction with multiple AI models can be overwhelming for users. The main contribution of this work is in providing a method, called Retrospective End-User Walkthrough, for screening the level of understanding and composition of AI model outputs, including why certain AI models were ignored by users. We explore the role of AI model outputs in decision-making, XAI, trust in those outputs, and how SMEs plan future use and interactions with the system after their first use of the system. The method was evaluated with 29 participants; four people participated in a pilot to validate the protocol and identify main usability issues; 25 participants (n=25) explored the insights derived by the proposed method.

The remainder of this paper is structured as follows: section 2 presents related work, section 3 details how the approach was evaluated, section 4 details the results, section 5 discusses results obtained so far, and section 6 presents some reflections regarding the suitability of the proposed method and our future research directions."

"conclusion" :

"The conclusion section is clear and provides a summary of the key findings and implications of the study. However, there are a few suggestions to enhance readability and emphasize key points:

1. Enhance References:
 - Provide a full citation for Hutchins [15].
 - Consider adding references to support the discussion on recency bias, confirmation bias, trust via transitivity, and trust by projected similarity.
2. Address Theme Violations:
 - The conclusion does not violate any specific themes.
3. Include Missing Related Works:
 - Consider mentioning any related works or studies that have explored similar topics or methods for evaluating human-AI decision-making systems.
4. Fill in Missing Information:
 - Provide a brief explanation of what RTA and CW stand for to ensure clarity for readers who may not be familiar with these terms.
5. Clarify Explanations:
 - Clarify the term "distributed cognition process" and explain how the system itself serves as an artifact for this process.
6. Add Limitations Section (if necessary):
 - Consider discussing any limitations or potential challenges of the proposed method, such as sample size limitations or potential biases in participant selection.
7. Any Other Suggestions:
 - Consider rephrasing the sentence "Future work is planned to increase the number of participants and to implement requirements identified for bringing AI model outputs to task flows the users commonly perform in a seamless and integrated way" for clarity and to specify the exact requirements that will be implemented.

Overall, the conclusion section effectively summarizes the main contributions of the study and highlights the implications for HCI practitioners and AI experts. With the suggested improvements, the section will be more informative and engaging for readers."

}
"Grammar" : {

"abstract" :

"Clarity and Engagement Suggestions:

1. Replace "towards a specific goal are proposed every day" with "for achieving specific goals are proposed regularly" in the sentence: "Evaluating human-AI decision-making systems is an emerging challenge as new ways of combining multiple AI models towards a specific goal are proposed every day."
2. Replace "amongst others" with "among other factors" in the sentence: "As humans interact with AI in decision-making systems, multiple factors may be present in a task including trust, interpretability, and explainability, amongst others."
3. Replace "so that they may gain an understanding" with "to gain an understanding" in the sentence: "The method consists of employing a retrospective end-user walkthrough with the objective of providing support to HCI practitioners so that they may gain an understanding of the higher order cognitive processes in place and the role that AI model outputs play in human-AI decision-making."
4. Replace "in the context of financial decision-making" with "for financial decision-making" in the sentence: "The system combines visual analytics, three AI models for revenue prediction, AI-supported analogues analysis, and hypothesis testing using external news and natural language processing to provide multiple means for comparing companies. Beyond traditional results on tasks and usability problems, outcomes presented suggest that the method is promising in highlighting why AI models are ignored, used, or trusted, and how future interactions are planned."
5. Replace "such as where AI model outputs could be placed to increase adoption of AI-supported approaches" with "including the optimal placement of AI model outputs to enhance the adoption of AI-supported approaches" in the sentence: "It also provides a means for eliciting information architecture requirements such as where AI model outputs could be placed to increase adoption of AI-supported approaches."
6. Replace "HCI practitioners researching human-AI interaction can benefit by adding this step to user studies in a debriefing stage" with "HCI practitioners studying human-AI interaction can enhance their research by incorporating this step into user studies during the debriefing stage" in the sentence: "We suggest that HCI practitioners researching human-AI interaction can benefit by adding this step to user studies in a debriefing stage as a retrospective Thinking-Aloud protocol would be applied, but with emphasis on revisiting tasks and understanding why participants ignored or connected predictions while performing a task."

"introduction" :

"Clarity and Engagement Suggestions:

1. Replace "especially for legal reasons" with "particularly for legal compliance" in the sentence: "[...] especially for legal reasons [14, 35]."
2. Replace "the explanation provided in XAI considers" with "XAI explanations consider" in the sentence: "The explanation provided in XAI considers two main characteristics: complexity and domain."
3. Replace "when AI models are used in decision-making systems" with "in decision-making systems that utilize AI models" in the sentence: "These works highlight the strong interconnectedness between XAI and trust when AI models are used in decision-making systems."
4. Replace "to realize the best of this human-AI partnership" with "to optimize the human-AI partnership" in the sentence: "new methods and tools are needed to help HCI practitioners understand how to bridge the gap between humans and AI decision-making systems to realize the best of this human-AI partnership."
5. Replace "influence how people deal with recommendations" with "impact individuals' response to recommendations" in the sentence: "However, in contrast to traditional HCI approaches, there is a lack of methods to specifically assess the interplay between different AI models [17], especially when recency bias may influence how people deal with recommendations [8, 11]."
6. Replace "Two of the most commonly used techniques" with "Two commonly used techniques" in the sentence: "Two of the most commonly used techniques that support gaining insights about these processes are Thinking-Aloud Protocol and Cognitive Walkthrough."
7. Replace "This technique is commonly employed" with "This technique is frequently used" in the sentence: "This technique is commonly employed in formative user studies, supporting the understanding of the rationale and task solving strategies used by participants [18]."
8. Replace "This complexity motivates the incorporation of the 'walkthrough'" with "To address this complexity, the 'walkthrough' is incorporated" in the sentence: "This complexity motivates the incorporation of the 'walkthrough' to bridge this gap, using the system as an artifact to reduce cognitive load as an external representation [5]."
9. Replace "Hence, the method proposed in this work aims at incorporating this aspect" with "Therefore, the proposed method in this work aims to incorporate this aspect" in the sentence: "Hence, the method proposed in this work aims at incorporating this aspect by combining RTA with an SME-performed walkthrough in a debriefing phase."
10. Replace "This work is part of a long-term project" with "This work is a part of an ongoing long-term project" in the sentence: "This work is part of a long-term project on human-AI decision-making that explores how SMEs combine multiple AI models in different domains including data science, discovery of new materials, and scientific workflows."



"conclusion" : "Clarity and Engagement Suggestions:

1. Split the long paragraph into smaller, more focused paragraphs to improve readability and organization.
2. Replace "RTA" with "Retrospective Task Analysis" to provide clarity on the acronym."

}

"Cohesion" :

"The coherence between the abstract, introduction, and conclusion of the article is generally good. The abstract provides an overview of the proposed method for evaluating human-AI decision-making systems, while the introduction provides background information on the importance of explainable AI and the need for methods to assess the interplay between different AI models. The conclusion summarizes the key findings of the study and discusses the implications for HCI practitioners and AI experts.

However, there are a few areas where the coherence could be improved:

1. In the abstract, it is mentioned that the method was qualitatively assessed with 29 participants, but this information is not mentioned in the introduction or conclusion. It would be helpful to include this information in the introduction to provide context for the study and in the conclusion to reinforce the validity of the findings.
2. The introduction discusses the importance of explainable AI (XAI) and its connection to trust in decision-making systems, but this connection is not explicitly mentioned in the abstract or conclusion. To improve coherence, the abstract and conclusion could mention the role of XAI in building trust and how the proposed method contributes to understanding the trustworthiness of AI model outputs.
3. The conclusion mentions the identification of recency bias, confirmation bias, trust via transitivity, and trust by projected similarity, but these effects are not discussed in the abstract or introduction. To improve coherence, the abstract and introduction could briefly mention these effects as potential insights gained from the proposed method.

To improve cohesiveness between these sections, the writer can consider the following suggestions:

1. Ensure that key information mentioned in the abstract is also mentioned in the introduction and conclusion to provide consistency and reinforce the main findings of the study.
2. Make explicit connections between the concepts discussed in the abstract, introduction, and conclusion. For example, explicitly mention the connection between XAI and trust in the abstract and conclusion, and briefly mention the identified effects in the introduction.
3. Provide a brief summary of the main findings in the conclusion to reinforce the key insights gained from the study.

By implementing these suggestions, the writer can enhance the coherence between the abstract, introduction, and conclusion, making the overall article more cohesive and understandable for readers."

}

Submit another text