Conceptual Models as a Basis for a Framework for Exploring Mental Models of Co-Creative AI

Jeba Rezwana[1,*], Mary Lou Maher[2]

[1]Towson University, MD, USA
[2]University of North Carolina at Charlotte, NC, USA

Abstract
Recent generative AI tools have moved human-AI co-creativity into the forefront of mainstream culture. Yet designing co-creative AI systems that effectively respond to human partners' values, preferences, and goals poses a significant challenge. Gaining insights into users' mental models is essential for the development of human-centered co-creative AI. This paper introduces a conceptual model of co-creative AI as a framework for exploring and analyzing users' mental models of co-creative AI. Our framework guides the design of tools for investigating mental models to align co-creative AI with users' needs and values.

Keywords
Co-Creative AI, Mental Models, Conceptual Models, Framework, Human-AI Co-Creation

1. Introduction

The availability of popular generative AI tools for creative domains like ChatGPT [1], DALL.E [2], and Github Copilot [3] has created widespread public interest, placing human-AI co-creativity into the mainstream culture. However, designing effective co-creative AI systems that effectively respond to human partners' values, preferences, and goals poses a significant challenge. In the dynamic creative process, diverse strategies and reasoning are essential, requiring adaptable AI agents to accommodate evolving human ideas. To achieve cognitive convergence in a co-creation, Fuller and Magerko propose understanding users' mental models [4].

A mental model is an individual's understanding of how something functions, allowing them to explain, predict, and act within systems [5]. Mental models are subjective constructs shaped by an individual's beliefs, values and experiences [6]. In the realm of human-centered co-creative AI, understanding users' mental models is crucial for aligning AI with their values and needs, fostering motivation for AI use. The effectiveness of co-creative AI is influenced by users and their social and cultural contexts. Staggers and Norcio [7] suggested that researchers must be aware of users' mental models to make human-centered designs.

The literature on users' mental models of co-creative AI is notably scarce, leaving several important questions unanswered. For instance, what are the constructs of conceptual models of co-creative AI? Conceptual models are representations of a target system developed purposefully by experts, unlike mental models, which are developed quickly and often unconsciously [6, 8].

Joint Proceedings of the ACM IUI Workshops 2024, March 18-21, 2024, Greenville, South Carolina, USA
*Corresponding author.
jrezwana@towson.edu (J. Rezwana); m.maher@uncc.edu (M. L. Maher)
0000-0003-1824-249X (J. Rezwana); 0000-0002-4150-0322 (M. L. Maher)

Conceptual models of co-creative AI enable exploring the diverse mental models held by users. Gaining insights into users' mental models enables the development of human-centered co-creative AI systems that better align with users' needs and values.

In this paper, we propose a conceptual model of co-creative AI leveraging the existing literature, which can be used as a framework to explore and analyze mental models of co-creative AI. The growing field of human-AI co-creativity would benefit from studying users' mental models of co-creative AI as they would help inform the design of human-centered co-creative AI.

2. Related Works

2.1. Mental Model Theory

Mental models are cognitive constructs developed through real-world experiences that allow humans to understand how a system functions [9]. These models allow people to understand, explain and predict phenomena and act accordingly [5, 10]. Therefore, users with well-developed mental models should be able to produce more accurate results when interacting with a system. Staggers and Norcio [7] suggested that designers and researchers must be aware of users' mental models. Norman [6] highlights that these models are subjective and prioritize usefulness over accuracy [6]. The contents of mental models can be concepts, relationships between concepts or events and associated procedures [11]. Humans construct mental models by drawing on analogies or metaphors of past represented objects or interactions [7, 12]. Young proposes eight types of mental models, including analogy as a form of mental model [13].

A conceptual model is a purposefully constructed accurate representation of the target system, coherent with scientifically accepted knowledge and typically developed systematically by experts [14, 6]. Human mental models are black boxes and will never be completely transparent [9]. Therefore, a solid conceptual model of a system is necessary as a tool to investigate an individual's mental model of that system [6]. Conceptual models are simplified representations of the target system [15].

2.2. Mental Models of AI

There have been a few studies on mental models of AI based on deep neural networks. Gero et al. developed a conceptual model of an AI system in a game setting and explored how users develop their mental models, uncovering peoples' preconceived notions affecting their mental models [8]. Tullio [16] investigated how users build mental models of an intelligent agent predicting an office worker's availability. Kulesza et al. [11, 17] examined mental models of an intelligent music recommender system, using surveys to identify participants' mental models and found that a 15-minute tutorial significantly improved the soundness of their mental models.

Various works in HCI studied users' mental models of AI systems, though very few studied mental models of co-creative AI. Llano et al. [18] asserted that equipping co-creative AI with users' mental models not only would enable better coordination but would also provide a valuable resource for co-creative AI to explain, justify, and defend their contributions. The idea of mental models as a key aspect of the design of real-time co-creative systems has been highlighted previously [19]. Davis et al. asserted that users' mental models help co-creative AI effectively structure, organize, interpret, and act on sensory data in real-time, which is critical for meaningful co-creation [20].

3. Characterization of a Conceptual Model of Co-Creative AI: A Framework

In this section, we propose a conceptual model as a framework for exploring users' mental models of co-creative AI by drawing on an existing foundation of research.

First, it is important to clarify the definition of conceptual models as it has been used interchangeably with the term mental models in the literature despite being distinct concepts. According to Norman [6], mental models encompass four distinct aspects: the target system (t), which is the actual system a user uses; the *conceptual model* (C(t)) of the target system, which provides a representation of the system developed by experts; the *mental model* (M(t)) of the target system, which users create in their head through the interaction with the target system; and lastly, the *scientist's conceptualization* of the mental model (C(M(t))). Given the inherent complexity of co-creative AI systems, aligning the conceptual model with the mental model can pose challenges. Gero et al. argued that a precise description of the neural network architecture and training procedure does not represent an appropriate conceptual model of an AI [8]. Rather, conceptual models are simplified representations of the target system which can be as simple as an analogy [15]. For instance, an analogy between Rutherford's atom and the solar system can be considered a conceptual model. Thus, our goal is to develop a simplified conceptual model that captures the essence of co-creative AI while making it useful for investigating mental models.

Human-AI co-creativity involves humans and AI collaborating on a creative process as they generate artifacts or ideas. By definition, co-creative AI systems involve a computational creativity (generative AI model) component, an interaction/collaboration component and a utility component. Our conceptual model draws inspiration from Kantosalo et al.'s research, which introduces three categories of metrics for evaluating co-creative systems: novelty (creative divergence), interaction and value [21]. These categories of metrics, as asserted by Kantosalo et al., are components to discern core behaviors of co-creative systems which is the basic purpose of conceptual models [21]. For our framework, we adapt the categories from their research and propose three main constructs for conceptual models of co-creative AI: the *Creativity Model*, the *Interaction Model* and the *Utility Model* (Figure 1).

We conducted a literature review to identify the key components of each construct, as characterizing conceptual models requires alignment with scientifically accepted knowledge [6, 8]. We used Google Scholar as our search database to conduct the literature review. To identify the key components of each construct, we used keywords based on the three constructs of our conceptual model. The keywords we used are: "mental models in co-creativity," "conceptual models in co-creativity," "computational creativity model," "interaction in co-creativity," "usefulness of co-creative AI," "utility of AI," and "co-creativity model." We considered documents published from 1995 until 2023. We did not include papers that are not in English, papers that by title or abstract are outside the scope of the research, and papers that do not involve co-creativity. We included

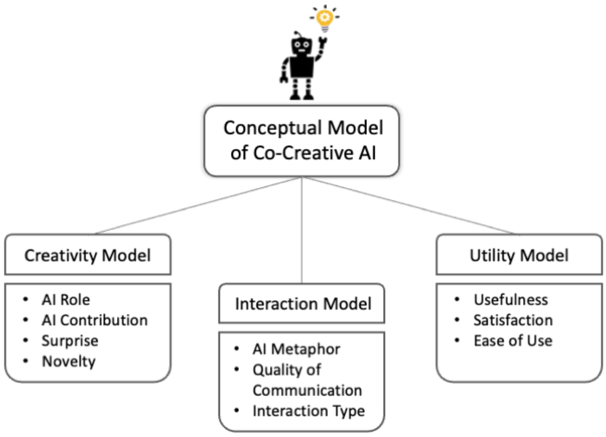

Figure 1: Constructs of the proposed conceptual model of co-creative AI.

papers describing the components of computational creativity, interaction in co-creativity and utility of co-creation.

Figure 1 shows our conceptual model, including the three main constructs and the key components of each construct. We refer to the specific publications that contributed to defining the key components of each construct of the framework in the sections below. In these sections, the first paragraph defines the construct and the core components of the construct, followed by specific questions that are presented to explore each of those components. The last paragraph references the relevant publications that provided the basis for defining the components of each construct.

3.1. Creativity model

The creativity model encompasses the computational creativity aspect of co-creative AI, focusing on creative content generation by AI and the roles of AI in the creative process. This model represents how a co-creative AI generates content, the capabilities of the AI and the extent of surprise, value, and novelty inherent in the generated content. To effectively understand or present the creativity model of co-creative AI, we suggest exploring the following components: *AI roles*, *AI contribution*, *surprise*, and *novelty*. The following questions could be utilized to explore these components.

- What can the co-creative AI actually do? (*AI role*)
- How does the co-creative AI contribute to the creative process? (*AI Contribution*)
- How surprising is the contribution of the co-creative AI? (*Surprise*)
- How novel is the contribution of the co-creative AI? (*Novelty*)

These questions can be augmented in alignment with specific research or design objectives, as the questions presented serve as a customizable template adaptable to research needs.

The following bodies of research provide the basis for the questions we propose to explore in understanding the creativity model. Colton et al. emphasized that computational creativity theory should highlight AI roles and AI contributions of co-creative AI as key components for explaining the creative process [22]. Kantosalo et al. further assert that the computational creativity model represents how a co-creative AI generates creative content and how it contributes to the creative process [23]. According to Boden, a creativity model comprises surprise, novelty, and value components [24]. Additionally, Grace et al. characterize creative outcomes with computational models of surprise, novelty, and value [25, 26].

3.2. Interaction model

The interaction model represents how the AI interacts and collaborates with humans. It includes the metaphorical representation of the type of collaboration, the communication quality between the collaborators and the characterization of the human-AI interaction. When investigating users' understanding of the interaction model of a co-creative AI, we suggest the following components to be explored: *AI metaphor*, *quality of communication* between the collaborator and *interaction type*. We also offer the following questions to be asked to explore these components.

- What is the appropriate metaphor for the co-creative AI? (*AI metaphor in co-creation*)
- What is the quality of the communication between humans and co-creative AI? (*Quality of Communication*)
- Is the interaction of the co-creative AI collaborative or tool-like? (*Interaction type*)

If needed, the above questions can be adapted and extended to the particular context and research objectives. The following research helped us formulate these questions.

The advantages of employing AI metaphors in interaction design have been extensively discussed within design research [27, 28, 29, 30, 31]. Research shows that metaphors impact the perception of collaboration [32]. About the characterization of human-AI interaction type, Kantosalo and Toivonen assert that contrary to how co-creative AI agents are often viewed in the literature, research in computational creativity aims to develop AI agents that are equal collaborators in the creative process [23]. Additionally, understanding the quality of communication between humans and AI is important for capturing the interaction dynamics [33].

3.3. Utility Model

The utility model encompasses the system's *usefulness*, *ease of use* and overall *satisfaction* when interacting with a co-creative AI. When conceptualizing the utility model through the lens of users, the following questions need to be explored:

- How useful is the co-creative AI? (*Usefulness*)
- How satisfactory is the co-creative AI? (*Satisfaction*)
- How easy is it to use the co-creative AI? (*Ease of Use*)

Similar to the other two constructs, the above questions can be expanded upon considering the context, domain and research objective. Below we discuss the research that inspired us in formulating these questions.

To identify the main components of the utility model, we adapted the technology acceptance model (TAM), a model to understand user acceptance of technology, which is an indicator of utility [34]. The two main constructs in TAM are perceived usefulness and ease of use. Satisfaction is a major usability variable [35] and is frequently used in the literature to measure a system's usability or user experience [36, 37].

4. Conclusions and Future Work

In this paper, we develop and describe a conceptual model of co-creative AI as a framework for exploring mental models of co-creative AI. Developing conceptual models of AI proves challenging due to the absence of a direct one-to-one correspondence between the AI model and its behavior, as emphasized by Gero et al. [8]. This complexity intensifies when tackling conceptual models of co-creative AI, where the intricate dynamics of open-ended collaboration with humans add an additional layer of challenge. We developed a conceptual model of co-creative AI, including three main components: *creativity model*, *interaction model* and *utility model*.

The literature on users' mental models of co-creative AI is notably scarce, leaving several important questions unanswered. For instance, what types of mental models of co-creative AI do users possess? Our conceptual model of co-creative AI facilitates the investigation and representation of key components in users' mental models. Researchers can employ surveys, interviews, and other methods based on the framework to explore and analyze users' models across diverse contexts and user groups. Understanding users' mental models facilitates the development of value-sensitive and personalized AI. Furthermore, while developing explainable co-creative AI, the constructs of the conceptual model can be used for conveying contextual information regarding each construct to users, eventually suggesting an accurate model of the system, as highlighted in recent research [38, 39]. This framework can also be used to develop tools that help users understand complex co-creative AI in different domains. The conceptual model also raises questions regarding decision-making within various modules of co-creative AI. For instance, how should we approach AI roles, contributions, metaphors, and interaction design? Further research is necessary to explore these questions.

References

[1] ChatGPT: Optimizing Language Models for Dialogue — openai.com, Online. URL: https://openai.com/blog/chatgpt/.
[2] DALL·E 2 — openai.com, Online. URL: https://openai.com/dall-e-2/.
[3] GitHub Copilot · Your AI pair programmer — github.com, Online. URL: https://github.com/features/copilot.
[4] D. Fuller, B. Magerko, Shared mental models in improvisational performance, in: Proceedings of the intelligent narrative technologies III workshop, 2010, pp. 1–6.
[5] P. N. Johnson-Laird, Mental models: Towards a cognitive science of language, inference, and consciousness, 6, Harvard University Press, 1983.

[6] D. A. Norman, Some observations on mental models, in: Mental models, Psychology Press, 2014, pp. 15–22.

[7] N. Staggers, A. F. Norcio, Mental models: concepts for human-computer interaction research, International Journal of Man-machine studies 38 (1993) 587–605.

[8] K. I. Gero, Z. Ashktorab, C. Dugan, Q. Pan, J. Johnson, W. Geyer, M. Ruiz, S. Miller, D. R. Millen, M. Campbell, et al., Mental models of ai agents in a cooperative game setting, in: Proceedings of the 2020 chi conference on human factors in computing systems, 2020, pp. 1–12.

[9] W. B. Rouse, N. M. Morris, On looking into the black box: Prospects and limits in the search for mental models., Psychological bulletin 100 (1986) 349.

[10] J. Rasmussen, On the structure of knowledge-a morphology of metal models in a man-machine system context, Technical Report, RISOE NATIONAL LAB ROSKILDE (DENMARK), 1979.

[11] 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.

[12] A. Collins, D. Gentner, How people construct mental models, Cultural models in language and thought 243 (1987) 243–265.

[13] R. M. Young, Surrogates and mappings: Two kinds of conceptual models for interactive devices, in: Mental models, Psychology Press, 2014, pp. 43–60.

[14] I. M. Greca, M. A. Moreira, The kinds of mental representations–models, propositions and images–used by college physics students regarding the concept of field, International Journal of Science Education 19 (1997) 711–724.

[15] I. M. Greca, M. A. Moreira, Mental models, conceptual models, and modelling, International journal of science education 22 (2000) 1–11.

[16] J. Tullio, A. K. Dey, J. Chalecki, J. Fogarty, How it works: a field study of non-technical users interacting with an intelligent system, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2007, pp. 31–40.

[17] T. Kulesza, S. Stumpf, M. Burnett, S. Yang, I. Kwan, W.-K. Wong, Too much, too little, or just right? ways explanations impact end users' mental models, in: 2013 IEEE Symposium on visual languages and human centric computing, IEEE, 2013, pp. 3–10.

[18] M. T. Llano, M. d'Inverno, M. Yee-King, J. McCormack, A. Ilsar, A. Pease, S. Colton, Explainable computational creativity, arXiv preprint arXiv:2205.05682 (2022).

[19] J. McCormack, P. Hutchings, T. Gifford, M. Yee-King, M. T. Llano, M. d'Inverno, Design considerations for real-time collaboration with creative artificial intelligence, Organised Sound 25 (2020) 41–52.

[20] N. Davis, C.-P. Hsiao, Y. Popova, B. Magerko, An enactive model of creativity for computational collaboration and co-creation, in: Creativity in the Digital Age, Springer, 2015, pp. 109–133.

[21] A. Kantosalo, P. T. Ravikumar, K. Grace, T. Takala, Modalities, styles and strategies: An interaction framework for human-computer co-creativity., in: ICCC, 2020, pp. 57–64.

[22] S. Colton, J. W. Charnley, A. Pease, Computational creativity theory: The face and idea descriptive models., in: ICCC, Mexico City, 2011, pp. 90–95.

[23] A. Kantosalo, H. Toivonen, Modes for creative human-computer collaboration: Alternating

and task-divided co-creativity, in: Proceedings of the seventh international conference on computational creativity, 2016, pp. 77–84.
[24] M. A. Boden, The creative mind: Myths and mechanisms, Psychology Press, 2004.
[25] K. Grace, M. L. Maher, D. Fisher, K. Brady, Data-intensive evaluation of design creativity using novelty, value, and surprise, International Journal of Design Creativity and Innovation 3 (2015) 125–147.
[26] K. Grace, M. L. Maher, N. Davis, O. Eltayeby, Surprise walks: Encouraging users towards novel concepts with sequential suggestions, in: Proceedings of the 9th International Conference on Computational Creativity (ICCC 2018). Association of Computational Creativity, 2018.
[27] D. Saffer, The role of metaphor in interaction design, Information Architecture Summit 6 (2005).
[28] A. F. Blackwell, The reification of metaphor as a design tool, ACM Transactions on Computer-Human Interaction (TOCHI) 13 (2006) 490–530.
[29] A. N. Antle, M. Droumeva, G. Corness, Playing with the sound maker: do embodied metaphors help children learn?, in: Proceedings of the 7th international conference on Interaction design and children, 2008, pp. 178–185.
[30] C. Faulkner, The essence of human-computer interaction, Prentice-Hall, Inc., 1998.
[31] J. T. Hackos, J. Redish, User and task analysis for interface design, volume 1, Wiley New York, 1998.
[32] P. Khadpe, R. Krishna, L. Fei-Fei, J. T. Hancock, M. S. Bernstein, Conceptual metaphors impact perceptions of human-ai collaboration, Proceedings of the ACM on Human-Computer Interaction 4 (2020) 1–26.
[33] J. Rezwana, M. L. Maher, Designing creative ai partners with cofi: A framework for modeling interaction in human-ai co-creative systems, ACM Transactions on Computer-Human Interaction (2022).
[34] F. D. Davis, A technology acceptance model for empirically testing new end-user information systems: Theory and results, Ph.D. thesis, Massachusetts Institute of Technology, 1985.
[35] E. Frøkjær, M. Hertzum, K. Hornbæk, Measuring usability: are effectiveness, efficiency, and satisfaction really correlated?, in: Proceedings of the SIGCHI conference on Human Factors in Computing Systems, 2000, pp. 345–352.
[36] N. Hill, G. Roche, R. Allen, Customer satisfaction: the customer experience through the customer's eyes, The Leadership Factor, 2007.
[37] P. �Klaus, S. Maklan, Towards a better measure of customer experience, International journal of market research 55 (2013) 227–246.
[38] K. I. Gero, T. Long, L. B. Chilton, Social dynamics of ai support in creative writing, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 2023, pp. 1–15.
[39] J. D. Weisz, M. Muller, J. He, S. Houde, Toward general design principles for generative ai applications, arXiv preprint arXiv:2301.05578 (2023).