Customizing the Balance between User and System Agency in Human-AI Co-Creative Processes

Caterina Moruzzi[1], Solange Margarido[2]
[1]Design Informatics, School of Design, The University of Edinburgh, UK
[2]CISUC, Department of Informatics Engineering, University of Coimbra, Portugal
cmoruzzi@ed.ac.uk, solange@dei.uc.pt

Abstract

The computational creativity community has long been engaging with the challenge of achieving a balance between unobtrusiveness and interjection in the interaction dynamics between computational systems and human users. This challenge has recently become more urgent due to the pervasive adoption and use of Generative AI tools in creative practices. This paper contributes toward understanding the trade-off between user and system agency, building on a user-centered co-creativity framework that identifies key dimensions responsible for the modulation of agency between users and AI in co-creative processes. We consider how the inclusion of these dimensions in co-creativity tools can help us achieve core values in human-AI interaction, such as alignment, co-adaptation, and explainability. We close the paper by suggesting a way to tailor co-creative experiences, describing the profiles of different AI personae that users might choose to interact with.

Introduction

The recent widespread integration of Generative AI (GenAI) models for creative content generation into creative processes — e.g., text-to-image generation models such as DALL-E 2 (Ramesh et al. 2022), Midjourney (midjourney.com/), Adobe Firefly (adobe.com/uk/products/firefly.html), or Stable Diffusion (stablediffusionweb.com) — is accentuating the need to determine the optimal balance between user-driven actions and AI assistance (Epstein et al. 2023). The latter is a long-standing discussion in the fields of Human-Computer Interaction (HCI) — or, to be more precise, in the field of Human-Centered-AI (HCAI) (Xu and Dainoff 2023) — and computational creativity (CC) (Lugrin, Pelachaud, and Traum 2022; Maes and Kozierok 1993). This examination is essential for maintaining the enhanced efficiency gained through automating repetitive tasks while also upholding users' sense of agency, control, and responsibility (Heer 2019). This inquiry becomes especially relevant in the realm of creative endeavors where personal agency and freedom of expression hold paramount importance.

Specifically, there is a need for understanding the trade-off between providing users with customization options that increase the controllability of the system they are interacting with on one side, and reducing the complexity of the interface to improve user experience and efficiency on the other (Cavalcante Siebert et al. 2023; Liu, Huang, and Holopainen 2023; Löbbers, Barthet, and Fazekas 2023; Shi et al. 2024).

In this paper, we build on our recent work (Moruzzi and Margarido 2024), aimed at addressing the challenge of striking a balance between user and system agency (Lawton et al. 2023) by proposing a user-centered model of H-AI co-creativity (UCCC). This model lists key dimensions of user-system interaction that co-creativity tools should allow their users to modulate in order to adjust the degree of control they have on the creative process.

We map the dimensions included in the UCCC model to key desirable values in H-AI interaction, i.e., alignment, co-adaptation, and explainability. This allows us to reflect on how the dynamics of co-creativity between users and AI systems can contribute to a wider understanding of the steps we should take to progress toward a meaningful relationship between human users and AI. We move on to expand on the suggestion we made in our previous work to implement the UCCC model into a prototype of a customization tool: the CoCreatAI app (Moruzzi and Margarido 2024). We show how, within this prototype, users are able to choose the desired level of customization of the interface and, accordingly, have the option of selecting which profile of AI co-creation partner to collaborate with.

This paper aims to contribute to an ongoing discussion in the fields of HCI and CC on the necessity of understanding how to reach an optimal equilibrium between automation and agency in H-AI interactions. The description of different AI profiles that users might choose to interact with is provided in line with the acknowledged necessity to tailor the co-creation environment to the diverse needs of individual users engaged in creative processes (Lin et al. 2023) and aims at improving the user experience and addressing the inherent variability in human preferences for collaboration.

In the next section, we provide an overview of how the topic of agency has been addressed by the HCI and CC community. We then move on to summarize the key dimensions of the model outlined in our previous work (Moruzzi and Margarido 2024) and map the dimensions to three key H-AI interaction values. We close the paper with a description of the AI profiles that the users can choose from to adapt the co-creative process to their own creative preferences and approaches.

User and System Agency

Due to the rapid increase of accessibility and use of GenAI tools and their deployment in social practices such as creative processes (Oppenlaender et al. 2023), the interest in the topic of agency in AI has considerably grown in the last few months. Yet, addressing the topic of (human or artificial) agency is far from being an easy task (Moruzzi 2022). Assigning agentive capabilities to a system can be interpreted in various ways. It may imply recognizing the system as autonomous and in control of its own actions, considering it responsible and accountable for its actions (Voiklis et al. 2016), or acknowledging the system as the author of the latter, especially in unique tasks like crafting a novel or composing music, leading to the attribution of intellectual property to the system. In addition, how we understand agency can change according to whether it is attributed to human or artificial agents. Intentionality, control, and causal powers are aspects of agency that are usually attributed to human agents (Chambon, Sidarus, and Haggard 2014; Legaspi, He, and Toyoizumi 2019; Moreno and Etxeberria 2005), while artificial agents have been deemed capable of being adaptive, autonomous, at least in part, and interactive (Franklin and Graesser 1996; Guzman and Lewis 2020; Russell and Norvig 2020).

Investigating the level of agency displayed by technological systems in one or more of these forms may give us insights into whether and how agency attribution influences the perception of technology as either a tool or a co-creator in human-machine collaborative processes. The growing sophistication of GenAI models used in creative applications is gradually tilting the balance of agency towards the AI system's side of the spectrum. Consequently, there is an urgent need to explore how adjusting agency levels can influence users' experiences and perceptions during collaborative creative processes involving technological tools (Davis 2013; Evans, Robbins, and Bryson 2023).

The UCCC model (Moruzzi and Margarido 2024) takes inspiration from works in the HCI and CC literature, which suggest various frameworks to elucidate the factors shaping co-creative processes between users and technology. Lubart's classification of computer roles in HCI as nanny, penpal, coach, or colleague is one of the most famous among them (Lubart 2005), and we draw inspiration from his classification – as well as from frameworks of the role of computational systems in co-creative interaction proposed in the literature (Guzdial et al. 2019; Kantosalo et al. 2020; Rezwana and Maher 2023) – in listing the AI profiles further on in the paper.

The UCCC framework presents desirable dimensions, each containing a varied number of parameters, which users can act upon to customize the creative experience to their own needs and style in real-time and in a synchronous way during co-creative processes with an AI system. The UCCC model is agnostic, i.e., it does not refer to any specific co-creative field of application but is instead conceived based on H-AI co-creative processes in general.

The model sits within the paradigm of mixed-initiative co-creativity. This paradigm, introduced in the work of Yannakakis, Liapis, and Alexopoulos (2014), is at the intersection between traditional creativity support tools like computer-aided design (CAD), which support the user in an unobtrusive way, and computational creativity, where artificial agents typically exert more control over the process in an autonomous or semi-autonomous manner (Lawton, Grace, and Ibarrola 2023). Mixed-initiative co-creativity assumes that both the human and the artificial agent involved in the creative process proactively contribute to the process, sharing the agency and control with the partner (Dellermann et al. 2019).

User-centered Co-creativity Model

In this section, we summarize the structure of the UCCC framework, with its dimensions and parameters. See our previous work for a full description (Moruzzi and Margarido 2024).

Interaction Guidance & Response

This category refers to the (**User Expertise** and **Model Operation Insights**) and to the (**Session Reflection**) stages of the H-AI interaction.

User Expertise

- **Expertise level**
  - Beginner
  - Intermediate
  - Advanced

Through the "User Expertise" dimensions, before commencing the co-creative process, users can indicate their level of expertise, allowing the AI to tailor interactions accordingly. In some of the dimensions of the UCCC framework, the user can select the option 'According to my level of expertise', thus leaving to the AI system the role of adapting the interaction based on the user's familiarity with the practice in question.

Model Operation Insights

- **Model interpretation**
  - Yes
    - Always
    - On demand
  - No
- **Tutorial**
  - Yes
  - No

The "Model Operation Insights" dimension responds to the need, from an HCI perspective, for users to access the model operations, to understand the complexity of the system they are interacting with and the parameters that are responsible for the actions that the system takes during the co-creative interaction (Shi et al. 2024). Users can opt to access information regarding model operations, utilizing, for example, visualization methods like clustering or dimensionality reduction (Chatzimparmpas et al. 2020; Llano et al. 2020). If users choose 'Yes', they can specify whether they prefer this information to be displayed constantly (e.g., on a dedicated section of the interface) or only upon request.

Session Reflection

User

Perception of control

- Low ◆—●—◆ High

Enjoyment

- Low ◆—●—◆ High

Interaction

- Difficult ◆—●—◆ Easy

AI

User's performance on xyz

- Overall
- Compared to previous sessions

The "Session Reflection" dimension refers to the post-interaction phase, where users reflect on the concluded co-creative session while the AI offers feedback to the user. Users can rate the perceived level of control, enjoyment, and difficulty during the interaction on a discrete scale. The AI system can use the responses provided by the user to adjust its behavior in subsequent interactions. Additionally, the AI can provide users with feedback on various aspects of their performance, according to the specific co-creative process, e.g., the novelty and originality of the output, the speed and accuracy of execution, etc.

Interaction Configuration

Interaction Interface

Customization level

- Limited
- Moderate
- Extensive

Storage preference

- Session-based
- Profile-based

Control panel layout

- Minimalist
- Moderate
- Comprehensive
- According to my level of expertise

The Interaction Configuration category refers to formal aspects of the interaction between the user and the AI. It includes the dimensions **Interaction Interface**, **Interaction Modality**, and **Feedback & Adaptation**.

The "Interaction Interface" dimension comprises three parameters. The 'Customization level' empowers users to determine their degree of control over customizing other dimensions within the Interaction Configuration and Interaction Dynamics categories. Opting for limited customization restricts users to a predefined subset of parameters, simplifying the selection process. Users may prefer this option to avoid feeling overwhelmed by numerous choices or due to time constraints. Conversely, extensive customization may appeal to experienced users seeking control over each available parameter.

The 'Storage preference' parameter pertains to whether users prefer for their settings and feedback to be limited to the current session (session-based) or saved for future interactions with the AI system instead (profile-based).

Additionally, the 'Control panel layout' parameter determines the complexity of the interface design. Users may opt for a minimalist or comprehensive layout based on their preferences, similar to the rationale behind the 'Customization level' parameter. Alternatively, users can allow the AI to automatically select a layout based on the level of expertise they indicated at the start of the co-creative interaction.

Interaction Modality

User → AI

Communication channel

- Text-based
- Voice-based
- Gesture-based
- Touch-based
- Behavioural

AI → User

- **Communication channel**
  - Text-based
  - Voice-based
  - Visual
  - Haptic

The "Interaction Modality" dimension allows users to select their preferred modality of communication with the AI. The success of text-to-image generation has resulted in disproportionate attention toward delivering prompts to the AI model through natural language over other modalities. Some GenAI models are moving in the direction of enabling users to interact with the models in modalities other than natural language, but this is an area that necessitates further research (Lin et al. 2023; Shi et al. 2024). This dimension lists some of the possible communication channels that users and AI systems can use to communicate during the co-creative process. It is doubtless that some of these modalities are difficult to implement (even if there are ongoing studies, for example, in the field of brain-computer interfaces, which might seem to be one of the hardest communication channels to achieve (de la Torre-Ortiz et al. 2020; Kavasidis et al. 2017), and more than one communication modality can be involved in a single interaction process. Nevertheless, this list, however incomplete, has the aim of encouraging to consider the possible interaction modalities applicable to co-creative AI systems, beyond natural language.

Feedback and Adaptation

User → AI

- **Feedback type**
  - Direct
  - Indirect

AI → User

- **Explanation detail**
  - Full
  - Moderate
  - Least
  - According to my level of expertise

The "Feedback & Adaptation" dimension is structured based on the communication direction: from the user to the AI and from the AI to the user. User feedback to the AI can be direct or indirect. Indirect feedback stems from user choices during the interaction, such as accepting or disregarding AI suggestions (Maes and Kozierok 1993). Direct feedback occurs through the communication channel specified in the "Interaction Modality" dimension. Conversely, the AI can offer users varying levels of explanation detail, ranging from 'Full' to 'Least'. The possibility for the user to understand what consequences their actions during a co-creative process produce, is crucial to a constructive interaction and a fulfilling user experience (El-Assady and Moruzzi 2022; Shi et al. 2024). On the other hand, the feedback that the user expresses to the AI model plays a relevant role in the adaptation of the model to the user's expectations and preferences. The user can choose to leave to the AI the role of adapting the explanation provided according to their level of expertise. In this case, it is important that the AI gives explanations in a language that the user is familiar with. In the 'Values in H-AI Interaction' section, we will come back to this specific parameter of interaction, highlighting its relevance for the development of a mutual theory of mind between the user and the AI (Larsson, Font, and Alvarez 2022).

Interaction Dynamics

The Interaction Dynamics category includes the dimensions that are responsible for the adaptation of the level of control that the user can have on the process: **Task Management**, **AI Interaction Approach**, and **Task Completion Authority**.

Task Management

- **Task assignment**
  - Unified task
  - Divided task
- **Processing approach**
  - Turn-taking
    - **Intervention pace**
      - Slow
      - Normal
      - Fast
  - Parallel

Within the "Task Management' dimension, the parameter 'Task assignment' delineates whether the user and the AI collaborate on the same or different tasks during co-creative interaction (Kantosalo and Toivonen 2016). Task execution may follow a turn-taking model (Thomaz and Chao 2011) or proceed concurrently. In parallel task execution, both the user and the AI independently contribute to specific aspects or components of the overarching task simultaneously. In turn-taking scenarios, users can select the intervention pace, ranging from ‘Slow’ to ‘Fast’.

AI Interaction Approach

Engagement
- Suggesting
- Taking initiative

Contribution style
- Pleasing
- Provoking

Content control
- Overwriting
- Not overwriting

The “AI Interaction Approach” dimension empowers users to influence how the AI behaves during interactive co-creative processes. Initially, users can determine whether they want the AI to take the initiative or merely suggest options. In the latter scenario, the AI offers suggestions, which users can accept or disregard (Larsson 2022). Furthermore, users can select the style in which the AI contributes to the process. Opting for a ‘Pleasing’ contribution style entails the AI facilitating users in their tasks, whereas a ‘Provoking’ AI may diverge from user expectations by suggesting alternative directions or creating different content (Kantosalo and Toivonen 2016; Rezwana and Maher 2023). The level of agency assumed by the AI in the interaction can be adjusted by users through the ‘Content control’ parameter. If users opt against overwriting, they retain final decision-making authority, allowing them to accept or manually override AI suggestions.

Task Completion Authority

Final decision
- User
- AI
- User and AI

The “Task Completion Authority” dimension pertains to who determines when the task is finished. Users can opt to retain full control over the submission decision, delegate it entirely to the AI, or share the decision-making responsibility with the AI. In the latter scenario, task submission occurs only when both co-creating partners reach a mutual agreement.

Values in H-AI Interaction

Careful consideration of agency dimensions in co-creative processes, and of how their modulation can afford different kinds of interactions between users and technology, is crucial not only for improving the users’ experience but also for promoting a more informed study on the influence that various levels of agency distribution have on other pressing issues in H-AI interaction, such as the attribution of responsibility, authorship, and trust. In this section, we consider how the dimensions outlined in the UCCC model can help us move towards some key desirable values in H-AI interaction, namely Alignment, Co-adaptation, and Explainability (see Figure 1). These values are essential for the partners who participate in collaborative processes to gradually develop trust in each other (Kästner et al. 2021; Schoeffer, Machowski, and Kuehl 2021), facilitating collaboration and a seamless blending of ideas (Davis 2013).

Alignment. AI alignment is a field of study concerned with the investigation of “whether AI technologies align with human interests and values and function in a safe and ethical manner” (Hagendorff and Fabi 2022). In an interaction state, the participating agents manage to effectively communicate with each other by adapting their knowledge representation models – or mental models – to the interlocutor (Johnson-Laird 1989). In mixed-initiative interactions between human users and AI, for the interaction to result in a correct updating of the mental models, users need a representation of the AI, and the AI needs a representation of the user (Kulesza et al. 2012).

In the context of H-AI interaction in co-creative processes, we can understand alignment as the harmonization of preferences, strategies, and decision-making between human users and AI, which can be achieved only if the participants in the interaction share a common mental model (McCormack et al. 2020). Alignment involves ensuring that the AI system’s outputs are in line with the intentions and expectations of the user, fostering a synergistic and effective co-creative environment. Mental models can vary in richness according to how much an agent understands about the phenomena and the context in which they are moving (Hindennach et al. 2023). This is why it is important to model the AI system’s behavior based on the user’s literacy and expertise. The feedback and explanations that users and AI can offer to each other during the interaction make sure that each co-creating participant in the process forms a rich and accurate mental model about their interlocutors, to ensure effective interactions (Wang and Goel 2022).

Co-adaptation. If successful, alignment between participants in a process results in the co-adaptation of the interlocutors’ knowledge representation models and a mutual achievement of the respective goals and needs (Van Zoelen, Van Den Bosch, and Neerincx 2021). The “Feedback & Adaptation” dimension in the UCCC model is designed to facilitate a co-adaptation between users and the AI. By acting on the parameters of these dimensions, users can select the type of feedback through which they communicate their agreement or disagreement with the action the AI has just executed. In this way, the AI can adjust the next moves to conform to the user’s expectations and preferences. The direct or indirect feedback provided to the AI during the co-creation process can also con-

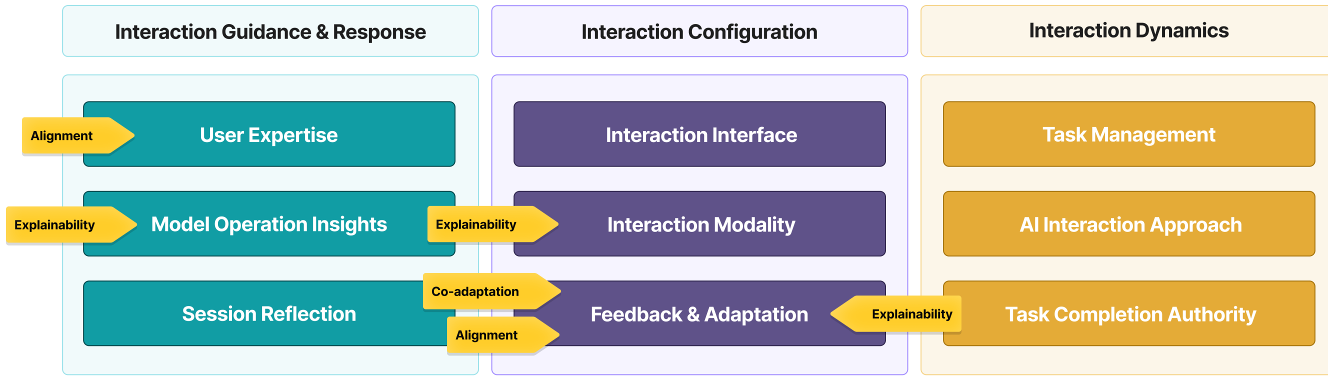

Figure 1: *H-AI Interaction Values*. The figure shows the correspondence between desirable values in H-AI interaction, i.e. Alignment, Co-adaptation, and Explainability, and the dimensions of the UCCC framework that can contribute to achieving each of them.

tribute to the alignment between the AI and the creative approach and strategies of the user(Rezwana and Maher 2023; Yamagami et al. 2023). Current co-creative approaches, as is the case in the game design domain, still mainly rely on indirect feedback, which may not be enough to promote successful co-adaption. The importance of establishing means for an ongoing dialog between the user and the AI is one of the aspects that the UCCC framework aims to highlight through the aforementioned dimension.

Conversely, the explanations provided by the AI assist users in comprehending the steps leading to a specific move or output. Users can then contemplate whether to modify their approach to encourage the AI to replicate an outcome if they are satisfied with it or avoid repeating certain actions if the result is unsatisfactory. The user's satisfaction with the AI-generated output can also increase thanks to a richer understanding of the mechanisms behind the steps taken by the AI system (Larsson, Font, and Alvarez 2022).

Explainability. The "Feedback & Adaptation" dimension is also significant for the third desirable value in H-AI interaction: Explainability. As a field of research, explainable AI (XAI) aims at producing scrutable AI models that maintain at the same time a high level of performance and accuracy (Ehsan et al. 2021). Research shows how explanations are useful for building trust between the AI and the user (Zhu et al. 2018; Ahn et al. 2021). In the context of the UCCC framework, having access to the "Model Operation Insights" and control over the AI's level and modality of explanation can empower the user to be more aware of the distribution of agency between themselves and the model.

The processes of XAI encompass an understanding of the AI's decisions and behavior, the diagnosis of the AI's performance and applicability, and a refinement of the AI models for the given users, tasks, and data (El-Assady and Moruzzi 2022). Explanations can be communicated through different media, such as natural language explanations, visualizations, interactive interfaces, local explanations, approximations, or case-based explanations (Lipton 2001). The "Interaction Modality" is, therefore, another crucial aspect of explainability: the choice of the channel used for mutual communication between the user and the AI participating in a co-creative process can determine the success in fostering a shared understanding between agents participating in collaborative processes (Llano et al. 2020).

AI Personae

In our previous work (Moruzzi and Margarido 2024), we proposed a way to implement the UCCC framework into a prototype of a customization tool for co-creative processes, the CoCreatAI app. We expand here on one of its functionalities, namely the possibility for the user to select the preferred level of customization and, if appropriate, one among six different AI personae to collaborate with.

When selecting a *Moderate* level of customization, the user of the app is offered the option to choose their preferred profile for the AI they will engage with during the process. We here expand on each one of the six AI personae (see Table 1).

AI as Assistant. The attribution of the role of an assistant to AI stems from the former conventional view of computational systems as merely support tools. This AI persona closely matches Lubart's nanny role (2005) or the support role described by Maher (2012). The HCI and CC communities now acknowledge that computers can take on different roles with varying degrees of initiative, autonomy, and authority in the creative process. Nevertheless, there are several instances in which the user and the specific creative context may benefit from a more passive approach from the AI. In such cases, the computer agent's mission can be summed up in one simple question: "How can I help?". Its focus should be on assisting the user to accomplish their tasks more quickly and efficiently. As such, a user with this preference can expect the system to have a suggesting engagement and pleasing contribution style, in order to help the user fulfill their goals without being intrusive. Evidently, an AI assistant is not supposed to overrule the user's intentions and, therefore, should not overwrite user content. The AI co-creating is also usually working on different tasks from the user, as it tends to take on the repetitive and burdensome tasks that the user wants to take off their plate. Since the role of AI should be to assist the user without raising any

	AI Persona					
	Assistant	**Inpirational Source**	**Idea Challenger**	**Colleague**	**Quality Control**	**Director**
Control panel layout	Comprehensive	Moderate	Moderate	Moderate	Moderate	Minimalist
Comm. channel (H→AI)	Behavioural	Behavioural	Text-based	Text-based	Behavioural	Behavioural
Comm. channel (AI→H)	Text-based	Text-based	Text-based	Text-based	Text-based	Text-based
Kind of feedback (H→AI)	Indirect	Indirect	Direct	Direct	Indirect	Indirect
Explanation detail (AI→H)	Least	Least	Full	Moderate	Moderate	Least
Task assignment	Divided	Unified	Unified	Unified	Unified	Unified
Processing approach	Turn-taking	Turn-taking	Turn-taking	Parallel	Turn-taking	Turn-taking
Intervention pace	Normal	Normal	Normal	N/A	Normal	Normal
Engagement	Suggest	Suggest	Take initiative	Take initiative	Take initiative	Take initiative
Contribution style	Pleasing	Provoking	Provoking	Pleasing	Pleasing	Provoking
Content control	Not ovr.	Not ovr.	Not ovr.	Not ovr.	Overwriting	Overwriting
Submission decision	User	User	User	Both	Both	AI

Table 1: *Table of AI Personae*. The table shows the parameters from the UCCC framework dimensions which would be automatically selected in the customization app when the user selects one of the six AI personae we discuss in the paper as their co-creation partner.

major questions or issues, indirect feedback from the user to the AI and the least amount of explanation detail of the AI's actions seems the most appropriate for this scenario – the system compliantly does what it is asked to do. The user should have the final decision.

AI as Inspirational Source. Sometimes, human users go through periods of creative blocks in which they find it difficult to come up with new ideas or move forward with a project that involves creative execution. In these cases, users may want some source of inspiration, but without giving up their control and sense of authorship over the creative artifact and what the final product will be. In this scenario, AI takes a more provoking stance in order to inspire the user in new directions, but it does so in a suggestive way, with the user either drawing on its suggestions or not. As with the AI as Assistant persona, the computational agent should be unobtrusive, not overwriting user content and providing minimal detail in its explanations, as its suggestions are merely intended to surprise the user and possibly trigger their creative thinking. The system never intends to impose its ideas on the human co-creator. In this case, the user and the AI are engaged in the same task and the user takes the final decision.

AI as Idea Challenger. A human user who chooses AI as Idea Challenger is actively looking to change their line of thinking by working with a co-creator who challenges their usual creative patterns and preconceptions about what the "right" approach is. The human user still wants to have more control over the creative process but is open to being questioned and challenged about the choices they make, as long as it is duly justified. In this sense, the agents work on the same task, and the AI can take some initiative and contribute directly to the creative artifact in a provoking style. Despite taking initiative, the system cannot overwrite user creations. It is expected that the system offers a high detail of explanation, in order to justify the alternative directions it presents and why the user should avoid more obvious choices. The user can also give direct feedback to the computational agent to allow for a debate of ideas and to be able to express more clearly whether or not they agree with the AI's alternatives. Towards the end of the creative process, the decision of when and whether to conclude it still rests with the human user.

AI as Colleague. The role of AI as Colleague stands in the optimal middle ground of the mixed-initiative co-creativity paradigm. By picking this AI persona, the human agent should expect a balanced partnership and a very close and strong collaboration between them and the computational agent (Margarido et al. 2024). Our AI as Colleague persona closely relates to Lubart's colleague role (2005). In this scenario, it is essential that the computational agent takes initiative and is able to contribute to the creative process autonomously. Here the co-creators work on the same task in parallel, meaning they don't have defined shifts and can contribute simultaneously and interrupt each other if they deem it necessary. The AI leans more towards a pleasing stance, as the goal is to converge to a result that satisfies both parties so that both the system and the human adapt to each other and make compromises in order to reach a common ground. Likewise, the AI avoids overwriting the human partner's actions, instead fostering a discussion in which both can agree to make changes to the creative artifact. Since the aim is to combine the ideas of one another in a seamless manner that pleases both, it is crucial to ensure bi-directional communication between the co-creative agents. Thus, the human user must be able to give direct feedback to the system. The system, on the other hand, generally provides explanations in moderate detail, so that it is sufficient to allow it to explain its process and contributions, but not excessively detailed to the point where it can slow down the pace and fluidity of the dialogue. To successfully finalize the co-creative task, both the user and the AI must agree on its completion.

AI as Quality Control. There are creative contexts where ensuring the quality and feasibility of artifacts is very important. Thus, the human co-creator may need the AI to perform quality control throughout the process so that certain requirements are fulfilled. This may involve, for example, ensuring the playability of a map and compliance with the game rules when working on level design. In creative writing, this can involve ensuring coherence of the writing tone and with the particular theme and setting of the story, as well as performing reviews of grammar, syntax, spelling, and clarity of the text. When working on a house project, there are several guidelines for ideal configurations and measurement ranges to ensure the functionality and practicality of the layout. The AI as Quality Control takes the initiative to correct, direct, improve, or block actions by the human designer, potentially overwriting their content. Despite having the control to overwrite user actions, the AI takes on a pleasing role in order to meet the user's intentions as much as possible without compromising the quality of the artifacts. In this co-creative setting, the human user only gives indirect feedback to the AI and the AI provides explanations in moderate detail, particularly if it rejects or changes a large portion of the user's actions. The task is only completed with the approval of both the human agent and the artificial agent.

AI as Director. The AI as Director persona is the option where the AI is given the most control over the creative process. This persona can be equated to the manager role of Guzdial et al. (2019). In this case, the human user is willing to be guided by the artificial agent, either because they are inexperienced with the creative context and want to learn from the AI, or because they have great confidence in the system's capabilities, or simply because they are curious to see which paths the system will take them. As such, the system takes the lead in the process, providing directions for the human co-creator to follow along. The AI is not concerned with following the user's ideas and intentions. On the contrary, it contributes to the creative process in a provocative style, steering the addition of content to meet its own needs. The system can overwrite the additions made by the human user if it considers that they do not meet its expectations. Since the human user follows along with the AI and adapts their decision-making based on its guidance, the AI offers simple explanations and a minimalist control panel layout for the user to simply participate without much room for interjections. The AI is in charge of the submission decision.

Limitations and Future Research

The six AI personae we presented in the previous section serve as baseline examples of co-creative AI profiles with predefined dimensions of the UCCC framework to facilitate the user's choice of customization. We admit that other AI personae may exist in addition to these, and may even be configured and added to the catalog by the user.

A limitation of the CoCreatAI app in its prototype form that we acknowledge is the technical challenge of implementing it due to the number and complexity of the dimensions of the UCCC framework. To address this limitation, in future research, we plan to work on a proof of concept of the CoCreatAI app, which will reflect the principles and dimensions delineated in the UCCC framework and will serve as a functioning system enabling user interaction, offering a concrete demonstration of the co-creative process. Through this proof of concept, we will be able to understand what key dimensions should be retained in the framework and which dimensions might be abandoned. We will iteratively develop the application through user feedback, addressing technical complexities as they emerge and improving the design of the app based on this feedback. In addition, considering the swift advancements in technology within the co-creative AI domain, we anticipate the feasibility of developing tools in the near future that approximate the customization level outlined by the UCCC framework.

A further challenge that may emerge from the implementation of this customization tool is how to ensure that users who are too busy to be willing to make thoughtful customization choices have a good experience of the tool. We envision two possible approaches to address this challenge: i) exploring ways to present the framework dimensions and the AI profiles so that users with different kinds of expertise and previous experience with AI tools can make an informed choice with minimum effort required, and ii) carrying out preliminary studies of implicit user behavior to understand which of the AI personae would best serve different user types. In this latter case, the AI personae would be part of an internal representation of the system, without being surfaced in the interface of the tool, and the tool will adapt to different users on the basis of limited explicit questioning at the beginning of the interaction.

Conclusions

In this paper, we expanded on our previous work (Moruzzi and Margarido 2024), discussing how the dimensions included in their UCCC framework allow us to reflect on key values of H-AI interactions, such as Alignment, Co-adaptation, and Explainability. We, then, described the characteristics of six AI personae that the users can choose to interact with through the CoCreatAI app, a customization device based on the dimensions included in the UCCC model, which can be applied to various co-creative tools to enhance the user experience and allow for a fine-grained modulation of the balance between user and system agency. With this work, we aim to contribute to the ongoing discussion within the CC community on how to achieve the right balance between user and system agency in co-creative processes.

Acknowledgments

We would like to thank our anonymous reviewers for their throughtful insights and comments. This work was supported by the UK EPSRC Next Stage Digital Economy Centre in the Decentralised Digital Economy (DECaDE) (EP/T022485/1), by the RKEI Award 2023-2024 of the University of Edinburgh, and by the FCT – Foundation for Science and Technology, I.P./MCTES through national funds (PIDDAC), within the scope of CISUC R&D Unit – UIDB/00326/2020 or project code UIDP/00326/2020. The second author is funded by Foundation for Science and Technology (FCT), Portugal, under the grant 2021.07047.BD.

References

[Ahn et al. 2021] Ahn, D.; Almaatouq, A.; Gulabani, M.; and Hosanagar, K. 2021. Will we trust what we don't understand? impact of model interpretability and outcome feedback on trust in ai. *arXiv preprint arXiv:2111.08222*.

[Cavalcante Siebert et al. 2023] Cavalcante Siebert, L.; Lupetti, M. L.; Aizenberg, E.; Beckers, N.; Zgonnikov, A.; Veluwenkamp, H.; Abbink, D.; Giaccardi, E.; Houben, G.-J.; Jonker, C. M.; et al. 2023. Meaningful human control: actionable properties for ai system development. *AI and Ethics* 3(1):241–255.

[Chambon, Sidarus, and Haggard 2014] Chambon, V.; Sidarus, N.; and Haggard, P. 2014. From action intentions to action effects: how does the sense of agency come about? *Frontiers in human neuroscience* 8:320.

[Chatzimparmpas et al. 2020] Chatzimparmpas, A.; Martins, R. M.; Jusufi, I.; and Kerren, A. 2020. A survey of surveys on the use of visualization for interpreting machine learning models. *Information Visualization* 19(3):207–233.

[Davis 2013] Davis, N. 2013. Human-computer co-creativity: Blending human and computational creativity. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 9, 9–12.

[de la Torre-Ortiz et al. 2020] de la Torre-Ortiz, C.; Spapé, M. M.; Kangassalo, L.; and Ruotsalo, T. 2020. Brain relevance feedback for interactive image generation. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, UIST '20, 1060–1070. New York, NY, USA: Association for Computing Machinery.

[Dellermann et al. 2019] Dellermann, D.; Ebel, P.; Söllner, M.; and Leimeister, J. M. 2019. Hybrid intelligence. *Business & Information Systems Engineering* 61(5):637–643.

[Ehsan et al. 2021] Ehsan, U.; Passi, S.; Liao, Q. V.; Chan, L.; Lee, I.; Muller, M.; Riedl, M. O.; et al. 2021. The who in explainable ai: how ai background shapes perceptions of ai explanations. *arXiv:2107.13509*.

[El-Assady and Moruzzi 2022] El-Assady, M., and Moruzzi, C. 2022. Which biases and reasoning pitfalls do explanations trigger? decomposing communication processes in human–ai interaction. *IEEE Computer Graphics and Applications* 42(6):11–23.

[Epstein et al. 2023] Epstein, Z.; Hertzmann, A.; of Human Creativity, I.; Akten, M.; Farid, H.; Fjeld, J.; Frank, M. R.; Groh, M.; Herman, L.; Leach, N.; et al. 2023. Art and the science of generative ai. *Science* 380(6650):1110–1111.

[Evans, Robbins, and Bryson 2023] Evans, K. D.; Robbins, S. A.; and Bryson, J. J. 2023. Do we collaborate with what we design? *Topics in Cognitive Science*.

[Franklin and Graesser 1996] Franklin, S., and Graesser, A. 1996. Is it an agent, or just a program?: A taxonomy for autonomous agents. In *International workshop on agent theories, architectures, and languages*, 21–35. Springer.

[Guzdial et al. 2019] Guzdial, M.; Liao, N.; Chen, J.; Chen, S.-Y.; Shah, S.; Shah, V.; Reno, J.; Smith, G.; and Riedl, M. O. 2019. Friend, collaborator, student, manager: How design of an ai-driven game level editor affects creators. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, 1–13.

[Guzman and Lewis 2020] Guzman, A. L., and Lewis, S. C. 2020. Artificial intelligence and communication: A human–machine communication research agenda. *New Media & Society* 22(1):70–86.

[Hagendorff and Fabi 2022] Hagendorff, T., and Fabi, S. 2022. Methodological reflections for ai alignment research using human feedback. *arXiv preprint arXiv:2301.06859*.

[Heer 2019] Heer, J. 2019. Agency plus automation: Designing artificial intelligence into interactive systems. *Proceedings of the National Academy of Sciences* 116(6):1844–1850.

[Hindennach et al. 2023] Hindennach, S.; Shi, L.; Miletić, F.; and Bulling, A. 2023. Mindful explanations: Prevalence and impact of mind attribution in xai research. *arXiv preprint arXiv:2312.12119*.

[Johnson-Laird 1989] Johnson-Laird, P. N. 1989. Mental models.

[Kantosalo and Toivonen 2016] Kantosalo, A., and Toivonen, H. 2016. Modes for creative human-computer collaboration: Alternating and task-divided co-creativity. In *Proceedings of the seventh international conference on computational creativity*, 77–84.

[Kantosalo et al. 2020] Kantosalo, A.; Ravikumar, P. T.; Grace, K.; and Takala, T. 2020. Modalities, styles and strategies: An interaction framework for human-computer co-creativity. In *ICCC*, 57–64.

[Kästner et al. 2021] Kästner, L.; Langer, M.; Lazar, V.; Schomäcker, A.; Speith, T.; and Sterz, S. 2021. On the relation of trust and explainability: Why to engineer for trustworthiness. In *2021 IEEE 29th International Requirements Engineering Conference Workshops (REW)*, 169–175. IEEE.

[Kavasidis et al. 2017] Kavasidis, I.; Palazzo, S.; Spampinato, C.; Giordano, D.; and Shah, M. 2017. Brain2image: Converting brain signals into images. In *Proceedings of the 25th ACM International Conference on Multimedia*, MM '17, 1809–1817. New York, NY, USA: Association for Computing Machinery.

[Kulesza et al. 2012] Kulesza, T.; Stumpf, S.; Burnett, M.; and Kwan, I. 2012. 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*, 1–10.

[Larsson, Font, and Alvarez 2022] Larsson, T.; Font, J.; and Alvarez, A. 2022. Towards ai as a creative colleague in game level design. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 18, 137–145.

[Larsson 2022] Larsson, T. 2022. Ai as a creative colleague in game level design.

[Lawton et al. 2023] Lawton, T.; Ibarrola, F. J.; Ventura, D.; and Grace, K. 2023. Drawing with reframer: Emergence and control in co-creative ai. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, 264–277.

[Lawton, Grace, and Ibarrola 2023] Lawton, T.; Grace, K.; and Ibarrola, F. J. 2023. When is a tool a tool? user perceptions of system agency in human–ai co-creative drawing. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, 1978–1996.

[Legaspi, He, and Toyoizumi 2019] Legaspi, R.; He, Z.; and Toyoizumi, T. 2019. Synthetic agency: sense of agency in artificial intelligence. *Current Opinion in Behavioral Sciences* 29:84–90.

[Lin et al. 2023] Lin, Z.; Ehsan, U.; Agarwal, R.; Dani, S.; Vashishth, V.; and Riedl, M. 2023. Beyond prompts: Exploring the design space of mixed-initiative co-creativity systems. *arXiv preprint arXiv:2305.07465*.

[Lipton 2001] Lipton, P. 2001. What good is an explanation? In *Explanation: Theoretical approaches and applications*. Springer. 43–59.

[Liu, Huang, and Holopainen 2023] Liu, H. X.; Huang, Y.; and Holopainen, J. 2023. How to use generative ai as a design material for future human-computer (co-) creation?

[Llano et al. 2020] Llano, M. T.; d'Inverno, M.; Yee-King, M.; McCormack, J.; Ilsar, A.; Pease, A.; and Colton, S. 2020. Explainable computational creativity. In *ICCC*, 334–341.

[Löbbers, Barthet, and Fazekas 2023] Löbbers, S.; Barthet, M.; and Fazekas, G. 2023. Ai as mediator between composers, sound designers, and creative media producers. *arXiv preprint arXiv:2303.01457*.

[Lubart 2005] Lubart, T. 2005. How can computers be partners in the creative process: classification and commentary on the special issue. *International Journal of Human-Computer Studies* 63(4-5):365–369.

[Lugrin, Pelachaud, and Traum 2022] Lugrin, B.; Pelachaud, C.; and Traum, D. 2022. *The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application*. ACM.

[Maes and Kozierok 1993] Maes, P., and Kozierok, R. 1993. Learning interface agents. In *AAAI*, volume 93, 459–465.

[Maher 2012] Maher, M. L. 2012. Computational and collective creativity: Who's being creative? In *ICCC*, 67–71.

[Margarido et al. 2024] Margarido, S.; Roque, L.; Machado, P.; and Martins, P. 2024. Boosting mixed-initiative co-creativity in game design: A tutorial. *arXiv preprint arXiv:2401.05999*.

[McCormack et al. 2020] McCormack, J.; Hutchings, P.; Gifford, T.; Yee-King, M.; Llano, M. T.; and D'inverno, M. 2020. Design considerations for real-time collaboration with creative artificial intelligence. *Organised Sound* 25(1):41–52.

[Moreno and Etxeberria 2005] Moreno, A., and Etxeberria, A. 2005. Agency in natural and artificial systems. *Artificial Life* 11(1-2):161–175.

[Moruzzi and Margarido 2024] Moruzzi, C., and Margarido, S. 2024. A user-centered framework for human-ai co-creativity. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24), May 11–16, 2024*. ACM, New York, NY, USA.

[Moruzzi 2022] Moruzzi, C. 2022. Creative agents: Rethinking agency and creativity in human and artificial systems. *Journal of Aesthetics and Phenomenology* 9(2):245–268.

[Oppenlaender et al. 2023] Oppenlaender, J.; Silvennoinen, J.; Paananen, V.; and Visuri, A. 2023. Perceptions and realities of text-to-image generation. *arXiv preprint arXiv:2306.08363*.

[Ramesh et al. 2022] Ramesh, A.; Dhariwal, P.; Nichol, A.; Chu, C.; and Chen, M. 2022. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125* 1(2):3.

[Rezwana and Maher 2023] Rezwana, J., and Maher, M. L. 2023. Designing creative ai partners with cofi: A framework for modeling interaction in human-ai co-creative systems. *ACM Transactions on Computer-Human Interaction* 30(5):1–28.

[Russell and Norvig 2020] Russell, S., and Norvig, P. 2020. *Artificial Intelligence: A Modern Approach*. USA: Prentice Hall Press, 4th edition.

[Schoeffer, Machowski, and Kuehl 2021] Schoeffer, J.; Machowski, Y.; and Kuehl, N. 2021. A study on fairness and trust perceptions in automated decision making. *arXiv preprint arXiv:2103.04757*.

[Shi et al. 2024] Shi, J.; Jain, R.; Doh, H.; Suzuki, R.; and Ramani, K. 2024. An hci-centric survey and taxonomy of human-generative-ai interactions.

[Thomaz and Chao 2011] Thomaz, A. L., and Chao, C. 2011. Turn-taking based on information flow for fluent human-robot interaction. *AI Magazine* 32(4):53–63.

[Van Zoelen, Van Den Bosch, and Neerincx 2021] Van Zoelen, E. M.; Van Den Bosch, K.; and Neerincx, M. 2021. Becoming team members: Identifying interaction patterns of mutual adaptation for human-robot co-learning. *Frontiers in Robotics and AI* 8:692811.

[Voiklis et al. 2016] Voiklis, J.; Kim, B.; Cusimano, C.; and Malle, B. F. 2016. Moral judgments of human vs. robot agents. In *2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN)*, 775–780. IEEE.

[Wang and Goel 2022] Wang, Q., and Goel, A. K. 2022. Mutual theory of mind for human-ai communication. *arXiv preprint arXiv:2210.03842*.

[Xu and Dainoff 2023] Xu, W., and Dainoff, M. 2023. Enabling human-centered ai: A new junction and shared journey between ai and hci communities. *Interactions* 30(1):42–47.

[Yamagami et al. 2023] Yamagami, M.; Madduri, M.; Chasnov, B. J.; Chou, A. H.; Peterson, L. N.; and Burden, S. A. 2023. Co-adaptation improves performance in a dynamic human-machine interface. *bioRxiv* 2023–07.

[Yannakakis, Liapis, and Alexopoulos 2014] Yannakakis, G. N.; Liapis, A.; and Alexopoulos, C. 2014. Mixed-initiative co-creativity.

[Zhu et al. 2018] Zhu, J.; Liapis, A.; Risi, S.; Bidarra, R.; and Youngblood, G. M. 2018. Explainable ai for designers: A human-centered perspective on mixed-initiative co-creation. In *2018 IEEE conference on computational intelligence and games (CIG)*, 1–8. IEEE.