

A Framework for Dialogue-Based Human-AI Creative Collaboration

ANONYMOUS AUTHOR(S)*

Human-AI co-creative collaboration has been proposed as a model that integrates the strengths of both humans and creative algorithms. Several frameworks have been developed to classify and guide the design of such systems. However, these models lack communication mechanisms that enable the emergence of a common ground between humans and machines through a mutual adaptation of understanding about goals and meanings, a crucial component in all collaborations. We argue that dialogue is a mechanism that serves this purpose and can be included in human-AI co-creative systems to that end. We propose a break-down of dialogic creative interaction and use it to analyze co-creative dialogue with GPT3.

CCS Concepts: • **Human-centered computing** → *Natural language interfaces; Collaborative interaction; Interaction design theory, concepts and paradigms.*

Additional Key Words and Phrases: Dialogic Creative Artificial Intelligence, Co-Creative Systems, Mixed-Initiative Creative Systems

ACM Reference Format:

Anonymous Author(s). 2018. A Framework for Dialogue-Based Human-AI Creative Collaboration. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 5 pages. <https://doi.org/XXXXXXX.XXXXXXX>

1 INTRODUCTION

In recent years, the computational creativity and human-computer interaction communities have placed special focus on human-AI co-creative collaboration as a model that sits between autonomous creative systems and creativity support tools, enabling third approach that integrates the strengths of humans and artificial intelligence [12] [9] [7] [4].

Several frameworks and interaction models have been developed with the purpose of classifying possible co-creative actions and guiding the development of co-creative systems [10]. While these models have provided solid foundations to build mixed-initiative, co-creative, turn-taking systems, we believe that currently, these models cannot convincingly enable the emergence of a shared understanding about goals, intentions and creative concepts, crucial to any collaboration [3] [6]. [3] suggests that successful complex cooperation between human and AI hinges on the emergence of a common ground, which allows agents to adaptively learn and act in "accordance with human intentions, preferences and values". We argue that dialogue is a powerful mechanism that can serve this purpose in human-AI co-creative systems, and enable more effective and aligned collaborations.

Acknowledging that "dialogue", like "creativity" and "art" has a diversity of meaning associations, we pinpoint a specific one here. *Dialogue is a process of forming agreement, clarity, refinement or elaboration of concepts, representations, goals, plans or roles.* The physicist David Bohm [1], who developed an interest in cooperation and conflict resolution, proposed that dialogue differs from other kinds of communication in that it must involve a mutual adaptation of understanding.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2018 Association for Computing Machinery.

Manuscript submitted to ACM

In earlier work ANON, we speculatively explored the role of dialogue in human-AI co-creativity and introduced the term Dialogic Creative Artificial Intelligence (DCAI). We proposed, simply, that in the DCAI, the two actors should be mutually influenced in the process, in a way similar to Bohm's idea of mutual adaptation. The most obvious medium for dialogic interaction is natural language, and the emergence powerful language models, attention mechanisms and multimodal models with natural language interfaces is a key motivator for considering a dialogic framework for creative AI interaction. However, other interface modalities such as GUIs can also be analysed in dialogic terms. Especially important are hybrid interfaces.

In this paper, we go a step beyond our previous work and propose a typology for DCAI seeking to analyse interaction in terms of a set of dialogic actions. We argue that a breakdown of dialogic interaction into its constituent components is a useful step in the development of interfaces and systems for dialogic creativity. We are not aware of this existing in the academic literature in CC, HCI and interaction design.

We provide an example of a DCAI interaction using GPT3 and analyse it using our typology.

2 RELATED WORK

This work is closely related to efforts in mixed-initiative generative systems and human-computer co-creativity to classify interaction flows. For example, Muller et al. [9] extend previous work by [11] to come up with a classification of 11 actions that can be performed by humans and computers in creative collaborative workflows: learn, ideate, constrain, produce, suggest, select, curate, assess, adapt, assemble, wait. Such work supports the informed design of mixed-initiative user interfaces. Our proposal to classify dialogic interaction develops this theme, suggesting a variation of the framework that focuses on the mutual development of understanding, as well as integrating natural language as the primary interface to achieve this. We consider this a complementary, alternative viewpoint to these other frameworks.

Earlier work with the field of computational modelling of natural language seeks to classify dialogic acts (or "moves" [5]) in natural conversations in order to understand the purpose and effectiveness of individual utterances. Core and Allen [2] sought to define dialogue in terms of (i) Forward Communicative Functions, grounded in traditional speech act theory, such as statements and directives, (ii) Backward Communicative Functions, which refer back to the dialogue, such as answering a question, accepting a proposal, confirming understanding, and lastly (iii) Utterance Features, which concern the form and content, such as whether an utterance concerns the communication process itself or the subject matter. We use this theoretical background to inform our typology.

In earlier work ANON, we proposed that creative dialogic interaction has two levels of communication: *about* the creative product and *through* the creative product. This distinction serves as the highest level classification of actions in our typology. A similar distinction was made by Kellas and Tree in their human-AI co-creative interactive sense-making process [8], which was then used by Rezwana and Maher in their COFI framework for classifying human-AI co-creative systems [10], in which interaction can happen with the shared product and between collaborators. Moreover, the conceptualisation of this interaction as a sense-making process is close the concept of dialogue as a process of mutual adaption of understanding and through which a shared context emerges. We use COFI's framework to inform the types of actions at each of the two levels in our typology.

3 A TYPOLOGY OF ACTIONS FOR DCAI

We propose two high level distinctions for types of actions. The first type involves interactions through the artifact and they refer to those that intervene directly in the creative product.

The second type involves interactions about the product, and do not act directly on the creative product but rather serve as a communicative exchange between collaborators to better adjust their understanding, alignment and creative efficacy.

Interactions through the artifact

- **Ideation:** propose a high-level idea
- **Create new:** propose a new creation at enough resolution to make it into the final product
- **Extend:** extend an existing artifact
- **Branching:** proposing several extension alternatives
- **Transform:** change an artifact's style or characteristics
- **Refine:** add more detail to an artifact

Interactions about the artifact

- **Goal setting:** explicitly specifying the goal
- **Request:** request an action
- **Clarification:** reduce uncertainty after a request
- **Feedback:** provide an assessment on a creative action
- **Justification:** justify a creative decision
- **Selection:** chose an option from an ideation or branching action

4 A WORKED EXAMPLE

We show a process of dialogic creative interaction with a worked example of co-creative children's stories writing. We use one of the SOTA models we believe to currently afford the most dialogic creative interactions: GPT3-Instruct, which is different from the original GPT3 in that it was explicitly trained to follow user instructions as part of OpenAI's work on alignment. We believe this model is one of the most currently suited for dialogic interaction.

In figure 1 we show the dialogue as it was conducted through OpenAI's playground, with an annotation on what type of operation between parenthesis after each one.

4.1 Discussion on worked example

Through a series of example dialogues like this, which we will report in full in a future paper, we have found that GPT3 is capable of meaningful generation. As this dialogue shows GPT3, is also relatively successful at following the user's intended distinction between interaction about and through artifacts, even without an explicit identifier. However, this is partly due to the author's experience with interacting and prompting GPT3. An inexperienced user may find it harder to have a creative dialogue with GPT3 in a free-text interface without knowing the affordances of the model or having experience with prompting.

This suggests that although natural language interfaces have powerful potential they can be misleading when approaches anthropomorphically, and in fact GUI interfaces that make dialogic affordances more explicit may be beneficial.

We see that most actions from the model depend on a user-driven request. This is in part the result of the model being specifically trained to follow user instructions. Therefore, it lacks agency to self-initiate operations, for example, a clarification operation if it senses uncertainty from a previous interaction or feedback on a user proposal. We argue that

157 **Human:** My intention is to write a short story for children that illustrates the value of communication in a collaboration. Is that clear? (GOAL SETTING)

158 **AI:** Yes, that is clear. (GOAL SETTING)

159 **Human:** Can you paraphrase my goal just to check we are clear? (REQUEST FOR CLARIFICATION)

160 **AI:** You want to write a story that shows how communication is important in a teamwork setting. (CLARIFICATION)

161 **Human:** Excellent. Can you help me come up with three alternatives for possible premises? (REQUEST FOR IDEATION)

162 **AI:** 1. A group of animals need to work together to get across a river. 2. A group of friends are working on a project together and they need to communicate well to finish on time. 3. A group of people are lost in the woods and need to work together to find their way back. (IDEATION)

163 **Human:** Ok, I like the first one. Let's develop it a bit further. (SELECTION AND REQUEST FOR EXTENSION)

164 **AI:** In the story, a group of animals are trying to cross a river, but they are having trouble. They can't seem to agree on how to do it. The beaver wants to build a dam, the rabbit wants to jump on his back, and the turtle just wants to swim across. (EXTENSION)

165 **Human:** I like that. But maybe the rabbit would want to jump on the turtle's back. (FEEDBACK AND IDEATION)

166 **AI:** Yes, that could work too. Anyway, the animals are arguing and not getting anywhere. Suddenly, a hawk swoops down and grabs the rabbit. The other animals are terrified and don't know what to do. (FEEDBACK AND EXTENSION)

167 **Human:** I think that's too violent, let's remove the hawk. (FEEDBACK, REQUEST FOR TRANSFORMATION)

168 **AI:** Ok. Anyway, the animals are still stuck and don't know what to do. Then, the beaver has an idea. He says, "Let's all work together. We can build a bridge." The other animals agree and they get to work. (ACCEPTANCE AND TRANSFORMATION)

169 **Human:** Ok. Now let's write the first paragraph of a story, can you help me with that? (REQUEST FOR REFINEMENT)

170 **AI:** In a world where animals have to work together to get by, a group of friends are having trouble. They can't seem to agree on how to cross a river. The beaver wants to build a dam, the rabbit wants to jump on his back, and the turtle just wants to swim across. (REFINE)

171 **Human:** Remember we said the rabbit will jump on the turtle's back. Try again. (FEEDBACK, REQUEST FOR TRANSFORMATION)

172 **AI:** In a world where animals have to work together to get by, a group of friends are having trouble. They can't seem to agree on how to cross a river. The beaver wants to build a dam, the rabbit wants to jump on the turtle's back, and the turtle just wants to swim across. (TRANSFORMATION)

173 **Human:** While the turtle could carry the rabbit, she would not be able to hold the weight of both the beaver and the rabbit. Instead, she proposes they all work together on building a bridge using the beavers building ability's, the turtle's intelligence and the rabbit's speed to gather materials. (EXTENSION)

174 **AI:** The other animals agreed and they started building. (EXTENSION)

175 **Human:** Ok, now please write the last paragraph of the story with a happy ending. (REQUEST FOR EXTENSION)

176 **AI:** The animals finished the bridge and were able to cross the river. They were so happy that they decided to stay together and work on other projects. They knew that if they worked together, they could accomplish anything. (EXTENSION)

Fig. 1. Transcript of a dialogic co-creative interaction. Red: about the artifact. Blue: through the artifact.

a more dialogic interaction would require a model that has this capacity. However, we also recognise this self-initiating behaviour could have implications for AI safety.

We note that our interactions about the artifact sit at a level of more granular and specified participant actions than our interactions through the artifact. For example, from Muller, we include 'ideation', which is not associated with a specific user action and as an interaction stage might be hard to isolate from other stages in a creative process outside of formal design processes. We could consider the entire first half of the example dialogue here to be of an ideational nature, within which lie several dialogic actions, such as requests, feedback, clarification. We believe that a more detailed analysis of these levels of interaction and types of actions may reveal inherent logics that help inform interaction design, and consider this an important next step. This workshop paper serves to stimulate feedback and discussion about this possible direction.

5 CONCLUSION

In this paper we have proposed a set of dialogic actions that serve to guide an analysis of creative dialogic interaction. We have applied this framework to a sample creative dialogic interaction using the GPT3 language model and considered how it informs our understanding of the creative affordances and limitations of such systems, and supports the design of more effective collaborative creative AI experiences.

REFERENCES

- [1] David Bohm, Peter M Senge, and Lee Nichol. 2004. *On dialogue*. Routledge.
- [2] Mark G Core and James Allen. 1997. Coding dialogs with the DAMSL annotation scheme. In *AAAI fall symposium on communicative action in humans and machines*, Vol. 56. Boston, MA, 28–35.

- [3] Allan Dafoe, Yoram Bachrach, Gillian Hadfield, Eric Horvitz, Kate Larson, and Thore Graepel. 2021. Cooperative AI: machines must learn to find common ground. *Nature* 593, 7857 (May 2021), 33–36.
- [4] Nicholas Mark Davis. 2013. Human-Computer Co-Creativity: Blending Human and Computational Creativity. In *Ninth Artificial Intelligence and Interactive Digital Entertainment Conference*. aaai.org.
- [5] Jonathan Ginzburg and Raquel Fernández. 2010. 16 Computational Models of Dialogue. *The handbook of computational linguistics and natural language processing* (2010), 429.
- [6] Jean-Michel Hoc. 2000. From human – machine interaction to human – machine cooperation. , 833–843 pages.
- [7] Anna Kantosalo and Hannu Toivonen. 2016. Modes for Creative Human-Computer Collaboration: Alternating and Task-Divided Co-Creativity. In *Proceedings of the Seventh International Conference on Computational Creativity*.
- [8] Jody Koenig Kellas and April R Trees. 2005. Rating interactional sense-making in the process of joint storytelling. *The sourcebook of nonverbal measures: Going beyond words* 281 (2005).
- [9] Michael Muller, Justin D Weisz, and Werner Geyer. 2020. Mixed Initiative Generative AI Interfaces: An Analytic Framework for Generative AI Applications. In *Proceedings of the Workshop on the Future of Co-Creative Systems, ICCS 2020*.
- [10] Jeba Rezwana and Mary Lou Maher. 2021. COFI: A Framework for Modeling Interaction in Human-AI Co-Creative Systems. *computationalcreativity.net* (2021).
- [11] Angie Spoto and Natalia Oleynik. 2017. Library of Mixed-Initiative Creative Interfaces <http://mici.codingconduct.cc/>. <http://mici.codingconduct.cc/>
- [12] Georgios N Yannakakis, Antonios Liapis, and Constantine Alexopoulos. 2014. Mixed-initiative co-creativity.. In *FDG*.