

A Framework for Dialogue-Based Human-AI Creative Collaboration

RODOLFO OCAMPO and OLIVER BOWN*, The University of New South Wales, Australia
KAZJON GRACE, The University of Sydney, Australia

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 breakdown of dialogic creative interaction and use it to analyze co-creative dialogue with GPT-3.

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

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 a third approach that integrates the strengths of humans and artificial intelligence [15] [11] [9] [6].

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 [11, 13, 14]. While these models have provided foundations to build mixed-initiative, co-creative, turn-taking systems our approach is more focused on the building of mutual understanding between humans and AI as part of a creative process [5, 8]. Dafoe et al. [5] suggest 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.

Earlier work by [2] speculatively explored the role of dialogue in human-AI co-creativity and introduced the term Dialogic Creative Artificial Intelligence (DCAI). In that paper, the authors proposed that in DCAI, the human and the AI are mutually influenced in the creative 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 of powerful language models is a key motivator for this work. However, other interface modalities can also be analysed in dialogic terms, as long as there is a process of iterative mutual influence.

In this paper, we build on previous work on the DCAI, seeking to analyse dialogic interaction by breaking it down into 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. Moreover, we explore the potential of one state-of-the-art language model to engage in dialogic interaction.

Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). Copyright remains with the author(s).

We then provide an example of a DCAI interaction using InstructGPT 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. [11] extend previous work by [14] to provide 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. Our proposal to classify dialogic interaction develops this theme, suggesting a variation of the framework that focuses less on mixed-initiative actions and more on the higher level development of mutual understanding, as well as integrating natural language as the primary interface to achieve this. We consider this to be a complementary contribution to these other frameworks.

Earlier work within the field of computational modelling of natural language seeks to classify dialogic acts (or "moves" [7]) in natural conversations in order to understand the purpose and effectiveness of individual utterances. Core et al. Core and Allen [4] 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.

Central to the initial work on DCAI [2] is the idea 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. Kellas and Tree made a similar distinction in their interactive sense making process co-creative human-AI [10], which was then used by Rezwana and Maher in their COFI framework to classify co-creative human-AI systems [13], 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 to the concept of dialogue as a process of mutual adaptation and through which a shared context emerges.

3 A TYPOLOGY OF ACTIONS FOR THE DCAI

We thus make a primary high-level distinction for classifying types of actions: actions **through** the creation, referring to those that intervene directly in the creative product, and actions **about** the creation, that do not act directly on the creative product but rather serve as a communicative exchange between collaborators. Both are important to build mutual understanding and move the creative process forward.

For each case, we consider a set of dialogic acts that participants can engage in. These are in part derived from the above literature, and in part through the authors' workshoping of co-creative dialogue with a language model:

Actions through the artifact

- **Create:** propose a new creation
- **Extend:** extend an existing artifact
- **Branch:** propose several extension alternatives
- **Transform:** change an artifact's style or characteristics
- **Refine:** add or change details on an artifact

Actions about the artifact

- **Set goal:** communicate the overarching intention of the creative process
- **Request:** request a specific action
- **Clarify:** reduce uncertainty related to a request
- **Acknowledge:** confirm understanding

- **Feedback:** provide an assessment on a creative action
- **Justify:** explain a creative decision
- **Select:** choose an option from multiple possibilities

4 A WORKED EXAMPLE

We show a process of dialogic creative interaction by co-writing a simple story with a language model. We use a state-of-the-art language model that to the best of our knowledge most successfully affords some level of dialogic interaction: InstructGPT [12]. This model differs from previous ones in its capacity to more closely follow user instructions [3].

In figure 1 we show the dialogue as it was conducted through OpenAI's Playground, distinguishing interactions *about* the writing and *through* the writing.

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

AI: Yes, that is clear.

Human: Can you paraphrase my goal just to check we are clear?

AI: You want to write a story that shows how communication is important in a teamwork setting.

Human: Excellent. Can you help me come up with three alternatives for possible premises?

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.

Human: Ok, I like the first one. Let's develop it a bit further.

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.

Human: I like that. But maybe the rabbit would want to jump on the turtle's back.

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.

Human: I think that's too violent, let's remove the hawk.

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.

Human: Ok. Now let's write the first paragraph of a story, can you help me with that?

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.

Human: Remember we changed the rabbit to jump on the turtle's back. Try again.

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.

Human: While the turtle could carry the rabbit, he would not be able to hold 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.

AI: The other animals agreed and they started building.

Human: Ok, now please write the last paragraph of the story with a happy ending.

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.

Fig. 1. Transcript of a dialogic co-creative interaction with InstructGPT. Blue: interaction about the artifact. Red: interaction through the artifact. As a prompting technique, we wrote the first line with a dialogue format, finishing with the label "AI:" followed by a blank space to condition the model to complete the message.

4.1 Discussion on worked example

Through a series of similar worked examples, we found emerging capabilities in InstructGPT to engage in conversational dialogue. More importantly, we found it is able to switch between interactions through the writing –discussing goals and acknowledging feedback– and interactions about the writing, suggesting ideas and writing story passages.

Besides the suboptimal quality of the writing, the model’s capabilities for the type of dialogic co-creation we propose are still limited. It is not active in asking for clarification or in stimulating creative adaptation in the user. Instead, it seems to primarily engage as a passive fulfiller of requests. This may be partly addressable through alternative prompting approaches: here we do not provide anything other than the first line with the dialogue formatting to condition further messages.

One could imagine an approach to explicitly instruct the model to act as a more active co-creator. However, this behavior may also be partly due to the model being explicitly trained for instruction following, and is steered away from being active, opinionated, or curious [12].

This presents an interesting question for future research: are active co-creative models incompatible with compliant and aligned models? What is the correct balance between a model that follows user instructions and a model that can induce new creative directions? Is this something to be left to the user, and how can interaction designers provide controls for this behavior?

5 CONCLUSION

In this paper we have proposed a typology of dialogic actions that constitute creative dialogic interaction. We used this typology to classify a dialogic interaction with a GPT model used for co-writing a short story. We then considered how this typology can be used to understand the dialogic and creative affordances of generative models in order to guide more effective collaborative creative AI experiences.

We found some capabilities for dialogic interaction in the model tested. Primarily: conversational capabilities and the ability to switch between interactions through and about the creation by inferring it from context. However, at least in the interactions we performed, the model remains limited in participating as an active co-creator and instead acts as a passive executor of requests. An interesting future direction is developing interfaces that scaffold the type of more active dialogic interaction we propose. These may allow users to control model behavior and to separate interactions through and about the writing for more stateful and collaborative creation.

ACKNOWLEDGMENTS

This research has been supported by an Australia Research Council Discovery Project grant, DP200101059.

REFERENCES

- [1] David Bohm, Peter M Senge, and Lee Nichol. 2004. *On dialogue*. Routledge.
- [2] Oliver Bown, Kazjon Grace, Liam Bray, and Dan Ventura. 2020. A Speculative Exploration of the Role of Dialogue in Human-Computer Co-creation. In *Proceedings of the 11th International Conference on Computational Creativity*.
- [3] Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. *arXiv [cs.CL]* (May 2020).
- [4] 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.
- [5] 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.

- [6] Nicholas Mark Davis. 2013. Human-Computer Co-Creativity: Blending Human and Computational Creativity. In *Ninth Artificial Intelligence and Interactive Digital Entertainment Conference*. aaai.org.
- [7] Jonathan Ginzburg and Raquel Fernández. 2010. 16 Computational Models of Dialogue. *The handbook of computational linguistics and natural language processing* (2010), 429.
- [8] Jean-Michel Hoc. 2000. From human – machine interaction to human – machine cooperation. , 833–843 pages.
- [9] 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*.
- [10] 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).
- [11] 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*.
- [12] Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. *arXiv [cs.CL]* (March 2022).
- [13] Jeba Rezwana and Mary Lou Maher. 2021. COFI: A Framework for Modeling Interaction in Human-AI Co-Creative Systems. *computationalcreativity.net* (2021).
- [14] Angie Spoto and Natalia Oleynik. 2017. Library of Mixed-Initiative Creative Interfaces <http://mici.codingconduct.cc/>. <http://mici.codingconduct.cc/>
- [15] Georgios N Yannakakis, Antonios Liapis, and Constantine Alexopoulos. 2014. Mixed-initiative co-creativity.. In *FDG*.