

“Coffee So Good, You’ll Want to Slap Your Barista”: Evaluating AI Co-Creation Through Dialogic Actions

Oliver Bown^a, Kazjon Grace^b, Rodolfo Ocampo^a and Francisco Ibarrola^b

^aSchool of Art & Design, UNSW, Sydney Australia; ^bDesign Lab, University of Sydney, Australia

ARTICLE HISTORY

Compiled September 23, 2024

ABSTRACT

Recent years have seen the rise of large language models (LLMs) and multimodal generators (here referred to as “request-based” (Bown and Brown, 2018) generative systems) that take natural language inputs. These suggest new dialogic forms of interaction for users of co-creative systems, in which the human and computer converse about creative goals, outputs, and concepts. Yet current request-based generative systems are poor at refined dialogic interaction, posing practical limits on how people can use these tools creatively.

Whilst there are rich bodies of literature studying the creative capabilities of AI systems, and their conversational and interactive capabilities and design, this paper specifically targets the role of dialogic interaction in human-machine co-creative interaction.

We propose the Creative Dialogic Actions Framework for analysing the efficacy of co-creative tasks in terms of creative-dialogic actions, looking at four subjects of dialogue – goals, roles, information about the world and information about personal stances – and then considering the intent, direction and target of those creative dialogic actions. Our analysis of human-machine co-creative dialogues with LLMs considers how dialogue participants effectively co-create, aligning their understanding of goals, roles, information about the world and personal stances, by successfully inferring those action elements.

We begin by surveying self-selecting creative users of LLMs to understand how important and how successful dialogic interaction. The survey results suggest that dialogic interaction, whilst present, is of limited value to creative AI users in their pursuit of positive successful creative outcomes. We then apply a creative dialogic analysis to three creative tasks devised by the authors to highlight ways in which dialogic interaction succeeds and fails in interactions with OpenAI’s GPT-4. Our analysis highlights how GPT-4 is poor at inferring or directing dialogic actions about goals, roles and personal stances, but better around information about the world.

We propose that in light of these weaknesses, it is useful to think of a spectrum of dialogic interaction that includes pseudo-dialogic interaction, where the system creates the impression of dialogue but doesn’t meaningfully achieve dialogue, and weak asymmetrical dialogic interaction, where the system possess some capability to achieve dialogic interaction, but largely leaning on the greater dialogic capabilities of the user. We consider how these limited dialogic capabilities relate to current LLMs’ narrow cognitive architecture, having no complex facilities to model an interlocutor’s expectations, goals and understanding.

KEYWORDS

Computational Creativity, Co-Creativity, Dialogic Creative AI (DCAI), Large

1. Introduction

Recent advances in machine learning-based language models and cross-modal and multi-modal generative systems (e.g., text-to-image generation) suggest that there is potential to move towards more *dialogic* forms of creative collaboration between practitioners and their software, achieving a significant new milestone in human-machine co-creativity, within the field of computational creativity (CC) (Colton and Wiggins, 2012; Jordanous, 2017; Yannakakis, Liapis, and Alexopoulos, 2014). Natural text interaction drives an expectation of successful dialogic interaction, yet current models, whilst incredibly effective generators, have no social cognitive capability and no social grounding (this is changing all the time, for example PaLM-E (Driess, Xia, Sajjadi, Lynch, Chowdhery, Ichter, Wahid, Tompson, Vuong, Yu, Huang, Chebotar, Sermanet, Duckworth, Levine, Vanhoucke, Hausman, Toussaint, Greff, Zeng, Mordatch, and Florence, 2023)). With the long-recognised centrality of dialogue to creative collaboration and practice (Barrett, Creech, and Zhukov, 2021; John-Steiner, 2000), this suggests these tools might afford significant utility for augmenting creativity.

LLMs and generative models have been trumpeted for their creative capabilities. In a critical early paper on GPTs (Bubeck, Chandrasekaran, Eldan, Gehrke, Horvitz, Kamar, Lee, Lee, Li, Lundberg, et al., 2023) they were shown to create remarkable original text to a complex brief. OpenAI’s CEO, Sam Altman, more recently trumpeted GPTs’ ability to perform creative ideation as one of its primary utilities (Bown, 2023). But along with evidence of such tools’ uptake in creative industries, and the potential power of LLMs’ creative ideation capability in specific areas (e.g., (Girotra, Meincke, Terwiesch, and Ulrich, 2023)), other informal accounts (e.g., (Song, 2023)) and formal studies (e.g., (Stevenson, Smal, Baas, Grasman, and van der Maas, 2022)) indicate significant shortcomings in their ability to respond to creative briefs with suitably quality material or to ‘think outside the box’. Detrimental effects have also been noted in the experience of creative users, such as the possibility that co-creation with such systems dulls-down creative thinking (Niloy, Akter, Sultana, Sultana, and Rahman, 2024).

An open area of investigation concerns whether such tools are successful at sustaining creative dialogues that help converge on suitable outcomes given a user’s creative objective. The purpose of this paper, continuing from (Bown, Grace, Bray, and Ventura, 2020; Ibarrola, Bown, and Grace, 2022; Ibarrola, Lawton, and Grace, 2023; Lawton, Ibarrola, Ventura, and Grace, 2023), is to develop the abstract concept of dialogue so that it can be applied as both an analytical tool and a design goal, to a range of interactions with CC systems.

In this paper we:

- (1) Define creative dialogic interaction
- (2) Detail the components and properties of creative dialogue in co-creative interactions, via an analytical framework, the Creative Dialogic Actions Framework.
- (3) Investigate these concepts in (a) a survey of creative uses of large language models, and (b) a more detailed worked example that illustrates creatively dialogic capabilities and shortcomings
- (4) Use this analysis to describe the potential co-creative competencies that AI systems currently have, and consider implications for design of future systems and

- for creative practice
- (5) Consider how these competencies relate to other accounts of AI systems’ intelligence and capabilities in relation to their design.

2. Creative Dialogic Interaction and Dialogic Actions

Current request-based generative systems are incredibly good at generating language with evidence of sophisticated conceptual understanding and reasoning (Bubeck et al., 2023; Jin, Liu, Lyu, Poff, Sachan, Mihalcea, Diab, and Schölkopf, 2023; Kosinski, 2023). But do such models enable refined dialogic interaction in a *creative* context whereby a user can achieve a creative goal? What specifically do we mean by dialogue, as a concept that is distinguished from interaction, communication, conversation and other related terms? Acknowledging that “dialogue”, like “creativity” and “art” has a diversity of meaning associations (these are “cluster concepts” (Bown, 2021; Gaut, 2000, 2005) or “essentially contested concepts” (Gallie, 1956; Jordanous and Keller, 2012)), we pinpoint a specific one here in our own terms:

Dialogue is a process of agreeing, clarifying, refining or elaborating upon concepts, representations, goals, plans or roles among a number of actors.

While dialogue can refer to any multi-participant interaction via text or speech or even other modalities, we are interested in its more specific usage concerning mutual influence and adaptation in a creative context. Our thinking about dialogue in a creative context takes inspiration from physicist David Bohm’s (2004) interest in conflict resolution, in which he proposed that dialogue differs from other kinds of communication in that it must involve a mutual adaptation of participants’ understanding or worldview. His methods for supporting dialogue include getting dialogue participants to reflect on the nature of the dialogue itself, and to prioritise a focus on shared understanding over seeking an agenda of actions. Contemporary work on helping people recognise unconscious bias is closely related to this (Greenwald and Banaji, 1995) sharing the core notion that there are aspects of our worldview (beliefs, habits, biases) that are ‘deep’; we are not necessarily even aware of them, let alone readily able to reflect on their validity. Bohm’s view was that through successful dialogue – a collective, reflexive practice – people would overcome entrenched positions, recognising this subjectivity and situatedness and achieving better agreement about how the world is. He argued that this deep agreement was a necessary step to resolution.

Amongst other things he also reminds us that “dia” means “through”, not “two” as might be assumed. Hence etymologically dialogue means “through speech”. A dialogue can be between any number of participants including occurring internally within an individual (clearly dialogue means more than its etymology gives away, but it is interesting to note that the term ‘monologue’ is not therefore defined in opposition to dialogue, as you might assume).

In earlier work (Bown et al., 2020) we proposed that dialogic interaction is a useful concept for co-creative collaboration between humans and machines. Dialogue is key to human-AI creative collaboration because the work of developing a shared understanding of creative concepts, goals, outcomes and so on is a powerful component of human-human creative collaboration, worthy of being emulated in human-machine co-creativity. We proposed, simply, that in dialogic creative AI, a group of actors should be

mutually influenced in a process of creative convergence. We noted that in HCI, there are many interfaces, besides “chat-like” natural language, through which something exhibiting this basic co-adaptation of dialogue might occur, and we set out to consider a range of ways in which different systems and interfaces can facilitate that mutual adaptation besides through language. For example, any generative system (even any creative tool or creative scenario) has the potential to trigger the user to rethink their creative goals, thus construing a sort of mutual influence. The generative system does not need to be particularly smart for this to occur. Indeed, numerous accounts of creative practice highlight how artists engage in a form of creative dialogue with tools, materials and environments, a theme now prevalent in social science perspectives on creative work, notably in (Gell, 1998; Ingold, 2013; Malafouris, 2007; Schön, 1995; Suchman, 1987). A creatively dialogic perspective sees various AI systems sitting on an interaction continuum along which tools and human agents occupy extremes: all interaction may involve mutual influence. Until recently, only human co-creative interactions have been creatively dialogic in the sense intended here. The present wave of generative AI systems suggest creative dialogic potential, and this paper seeks to understand the ways in which that dialogic nature is satisfied or not.

To capture the idea that systems and interaction scenarios may achieve different levels of dialogic capability, in the analysis presented in this paper we will consider how a concept of dialogue can be broken down into types. We propose variations that fall short of what we might call full symmetrical dialogue: “**pseudo-dialogue**” whereby the system falls short of dialogic capability but still provides the experience of a dialogic interaction, which may still be an effective user experience; and “**weak asymmetrical dialogue**” whereby the system achieves some level of dialogic capability, but the achievement of dialogue is largely propped up by the more advanced dialogic capabilities of the human user.

A widely celebrated historical example of machine creativity involves the use of a genetic algorithm (GA) to evolve original aerial designs by NASA (Hornby, Globus, Linden, and Lohn, 2006). In this case the GA is given a measurable performance target and the emerging aerial design simply optimises this design. The measurable nature of the outcome in a computer simulation means that the goal of the creative task is perfectly expressed – perfect alignment between user and system goal, eliminating a need for dialogic convergence. Translation of such GAs to be used in artistic tasks has been unsuccessful because of the impossibility of formally expressing aesthetic goals, but research into interactive genetic algorithms (Stanley and Lehman, 2015), whereby a user iteratively selects outcomes, has shown that user and system are more likely to *mutually* converge, i.e., the user is inspired by outputs created by the system rather than merely dictating a goal to the system. Thus within evolutionary computing approaches two quite different approaches to dialogue around goals, and the convergence of participants, can be seen. The latter scenario, whereby users do not express clear goals but adapt their goals in an ongoing process of ideation, which may be divergent or convergent, is a central feature of creativity and has been the subject of studies into creative ML applications. Grace, Maher, Davis, and Eltayeb (2018), for example, considers co-design tasks in which an ML system stimulates the creative user to think differently.

In LLM contexts there is potentially an even wider range of interaction scenarios and yet these black-box systems can mean there is little formal structure that defines creative roles and the creative process. Anecdotal reports show users successfully using LLMs for ideation or more convergent stages in a creative task such as refinement of a text following iterative instructions. As well as simply measuring creative achievement,

there remains important foundational work to be done in classifying the nature of creative dialogic interactions. This is the focus of the current paper.

Note that in pursuing this interest in creatively dialogic interaction we do not mean to say that human-machine co-creativity must mimic modes of human interaction or be measured thus. Instead our goal is a comparative perspective; framing computer systems in comparison of human behaviour. Such comparison aims to identify shortcomings in the adaptive behaviour of current systems. But a creative dialogic frame can also be used to analyse systems' capabilities at stimulating the user's adaptation (as in previous work such as (Grace et al., 2018)), as this is a need of creative users; it is common for creative practitioners to desire a creative process that drives them to reconceptualise problems, to seek diversity, and to provide novel stimuli.

3. Breaking down creative dialogue into actions

How can one analyse interactions with generative systems in terms of dialogic creativity? When we consider artistic collaborators working together creatively, these participants may need to explain their ideas, seek a shared understanding of goals, engage in forms of brainstorming (seeking divergent outcomes), seek clarification, justify, critique, agree (perhaps despite a lack of shared understanding) and so on. How might we consolidate this list of actions more formally? Then how might we apply such a list to existing co-creative experiences, and speculatively apply them to future technologies and design scenarios?

While an analysis of dialogic components is well established in linguistics and natural language processing, in the present work we adapt existing frameworks with a specific focus on creativity. A common example is clarification: imagine a user requesting a system to generate an output according to some criteria or parameters, via natural language or a specific interface. Can the user then seek to clarify what they wanted based on the observed results? Different interface paradigms present different ways this may occur. For example historically users of generative neural network systems may have used alterations to training sets as a way of refining the space of possible generative outputs. Additionally, when training generative neural networks another practice is to augment data sets by incorporating modified versions of the data points to "encourage" the model to generate samples outside the training data for more of a novelty-search effect. In many of the systems we have considered, clarification in the sense used in human dialogue, where participants can develop their shared understanding of a creative goal, is either not present or is limited. Multimodal generative models like GPT-4, where users can engage in a dialogue before or after generating an image, have begun to incorporate this capability, but multimodal generation still has significant flaws (Tong, Liu, Zhai, Ma, LeCun, and Xie, 2024). Other forms of pseudo-clarification may exist, for example through the art of prompt engineering where the user returns to their input prompt and attempts to refine it, or tweaks some parameters.

4. Related Work

In the wider context, our analysis relates to psychological and computational studies of creativity and problem-solving. Research spanning the fields of creativity, psychology, computing, design and organisation science investigate how individuals and groups

creatively identify, formulate, represent and process problems or creative objectives and their solutions. For example, with relevance to creative dialogic interaction, (Foss, Frederiksen, and Rullani, 2016) look at how different levels of structure in communication amongst participants correlate to different phases in problem development.

Amongst models of creative processes in the literature, significant work includes the Function-Behaviour-Structure model of design (Gero, 1990) and the Creative Problem Solving framework (Treffinger, 1995). These have, for the most part, taken an individual, cognitive approach to the creative process, focusing on the intentions and activities of a single practitioner. When used empirically, these frameworks permit inferences to be drawn about metacognitive strategies that underlie creative practice (i.e. Laureiro-Martinez, Arrieta, and Brusoni (2023); Suwa, Gero, and Purcell (2000)). When extended to team contexts, however, these models lend themselves to an analysis of the consensus cognition of the group, not the dialogic surface by which that consensus is reached. We argue that these models, while very useful in their own contexts, do not generalise to a study of creative dialogic interaction, where alignment of understanding, negotiations and nuanced ambiguities between practitioners are the focus.

Looking towards HCI, our work is closely related to efforts in mixed-initiative generative systems and human-computer co-creativity to classify interaction flows. For example, Muller, Weisz, and Geyer (2020) extend previous work by Spoto and Oleynik (2017) to come up with a classification of 11 activities that can be performed by humans and computers in creative collaborative workflows: learn, ideate, constrain, produce, suggest, select, curate, assess, adapt, assemble, wait. They use this classification to propose certain typical collaborative flows of interaction, akin to high-level user flows. Such work supports the informed design of mixed-initiative user interfaces but, we feel, does not provide a framework for analysing success in creative dialogic interaction.

Earlier work within the field of computational modelling of natural language seeks to classify dialogic acts in natural conversations in order to understand the purpose and effectiveness of individual utterances. Core and Allen (1997) 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.

Ginzburg and Fernández (2010) set out to define a space of dialogue “moves” but, with reference to Wittgenstein’s theories of language, ask whether such a space can be defined in a domain-independent way. They answer that specific domains require their own dialogic logic, but that there are nevertheless general notions that apply to a unified concept of dialogue. Discussing a specific example, they identify properties: (i) Self-answering, (ii) Multilogue, (iii) Disagreement, (iv) Partial Comprehension, (v) Incomplete Utterances, (vi) Sentential Fragments. In light of Ginzburg and Fernández (2010)’s appeal to domain-specific models of dialogue, we may ask what dialogic properties, if any, are specific to creative collaboration, and consider whether this is an original domain of application.

Meguro, Higashinaka, Minami, and Dohsaka (2010) identify 33 dialogue acts, where dialogue is treated in a broader sense than the creativity-specific sense we use (akin to conversation). These fall into the following top-level categories: (i) Greeting, (ii) Information, (iii) Self-Disclosure, (iv) Acknowledgement, (v) Question, (vi) Sympathy, (vii)

Non-Sympathy, (viii) Confirmation, (ix) Proposal, (x) Repeat, (xi) Paraphrase, (xii) Approval, (xiii) Thanks, (xiv) Apology, (xv) Filler, (xvi) Admiration. Each is broken down into more specific subcategories. For example under Self-Disclosure, statements of preferences, personal plans and desires are included. Meguro et al. (2010)’s set of dialogue acts are some of the most widely used in applied computational models of dialogue.

Liao, Geyer, Muller, and Khazaen (2020), looking at conversational interfaces in an information retrieval (i.e., non-creative) context, consider an action space for conversational search with user actions such as Reveal, Disclose, Revise, Refine, Expand, Extract, Elicit, Clarify, List, Summarise and Compare. Drawing on principles from cognitive science, and noting that users tend to have to do lots of “repair” work in such conversational interactions, they argue that greater work needs to be done to improve “recipient design”: “A system may fail both the search task and creating a truly conversational experience if solely relying on users to repair, as conversation should be a two-way process to resolve uncertainty” (Liao et al., 2020). Several other studies consider how users perform repair strategies in conversational interactions with computational systems. For example, (Dippold, 2023) highlights convergence of user-system understanding as a significant obstacle in task-oriented chatbot interaction.

Other studies of conversational interaction identify areas beyond task-oriented contexts with a clear need for convergence. For example, (Narimatsu, Higashinaka, Sugiyama, Mizukami, and Arimoto, 2021) considers the design of an interview agent, inspired by analysis of talk-show hosts’ interviews. The analysis proposes dialogic elements such as Filler, Repeat, Admiration, Urging, Paraphrasing and Questions.

Earlier work in the analysis of human dialogue introduces frameworks for understanding dialogue in terms of leading and initiative. For example, (Chu-Carroll and Brown, 1999) argue that dialogue initiative (leading the dialogue) and task initiative (determining the task in question) might both be distinctly identifiable within a dialogue. This contrasts with (Jordan and Di Eugenio, 1997) for whom control should be used to describe a dialogue-level phenomenon, and initiative should refer to the participants’ problem solving goals.

5. The Creative Dialogic Actions Framework

For the purposes of analysing the efficacy of creative dialogic interaction, we believe that a finer-grained analysis is needed than Spoto and Oleynik (2017) and Muller et al. (2020), and that, following Ginzburg and Fernández (2010), this analysis needs to be creative-domain specific, with a narrower (coarser-grained) array of actions than is found in general natural language analysis.

Bearing in mind these factors and based on our own workshopping of co-creative dialogic scenarios, we consider a Creative Dialogic Actions Framework for thinking about dialogue in creative collaboration. At its core is a series of property categories for creative dialogic actions, or *CD-actions*. The idea with this framework is that any moment in a dialogue can be categorised by at least one item from each category.

- **subject:** A CD-action can be concerned with **goals (G)**, **roles (R)**, **information about the world (W)** or **personal stances (P)**.
- **intention:** A CD-action can be intended for **teaching (T)**, **direction (D)**, **stimulation (S)**, or **enquiry (E)**.
- **orientation:** A CD-action can be **forward (→)** or **backward (←)** facing, the

latter meaning that it refers back to something previously discussed.

- **trigger**: A CD-action can be a **request** (\uparrow) for a response, or a **submission** (\downarrow) (e.g., of information).

To illustrate these classifications, a set of imaginary statements from a discussion between two collaborating musicians is given in Table 1. In each case, the statement is analysed as possessing one primary subject (goals, roles, the world, or a personal stance), intent (teaching, direction, stimulation, enquiry), forwards or backwards orientation (\rightarrow/\leftarrow), and trigger (request \uparrow or submission \downarrow).

Example	Subject	Orientation	Trigger	Intent
“I want this track to have big energy”	G	\rightarrow	\uparrow	T
“Your job is to bring the energy”	R	\rightarrow	\uparrow	D
“100bpm is too slow for house”	W	\rightarrow	\uparrow	T
“I love Elton John’s energy”	P	\rightarrow	\uparrow	T
“Let’s play it fast”	G	\rightarrow	\downarrow	D
“That was too slow”	G	\leftarrow	\uparrow	T
“Can you give it more swing”	G	\leftarrow	\uparrow	D
“That didn’t have a lot of swing”	P	\leftarrow	\uparrow	T
“There are 5 bars in this section”	W	\rightarrow	\uparrow	T
“Repeat that last bar one time”	G	\rightarrow	\downarrow	D
“Let’s find another way to end”	G	\leftarrow	\downarrow	S
“What was that change in the last bar?”	W	\leftarrow	\downarrow	E

Table 1. Examples of statements with associated CD-action analysis. CD-action can be concerned with **goals** (G), **roles** (R), **information about the world** (W) or **personal stances** (P). A CD-action can be intended for **teaching** (T), **direction** (D), **stimulation** (S), or **enquiry** (E). A CD-action can be **forward** (\rightarrow) or **backward** (\leftarrow) facing (refers back to something previously discussed). A CD-action can be a **request** (\uparrow) for a response, or a **submission** (\downarrow) (e.g., of information).

However, in each case there are potentially multiple CD-actions in a single statement, and there are also multiple interpretations that might be made for the CD-action analysis. A range of linguistic factors, in particular the process of inferring context and intent via Gricean pragmatics (Grice, 1978), ultimately influence how this interpretation might play out. For example, it is often ambiguous whether something should be considered a statement about the world or about one’s personal stance (“that didn’t have a lot of swing”), or whether something is a goal or a stance (e.g., “that was too slow” can be interpreted as a goal: “let’s play it faster”). The difference between direction and stimulation can also be ambiguous: “Let’s find another way to end” is an invitation to search alternatives without a specific orientation, but it could also be considered a direction (*go search*). In comparison, “the ending needs to be more dramatic”, could be understood as teaching about goals, or a direction which also includes open-endedness (hence stimulation).

Our framework can allow us to consider each of a multitude of interpretations. For example: “the ending needs to be more dramatic” can be analysed as $[G\leftarrow\downarrow T]$ or $[P\leftarrow\uparrow T]$ or $[G\leftarrow\downarrow S]$ all at once. Thus the respondent to this statement could learn something about the speaker’s goals (the respondent didn’t know it was supposed to be dramatic), or their personal stance about dramaticness (they thought it was dramatic, but now they know it needs to be more dramatic), and at the same time this statement may beg the question of how drama is achieved. A conversation may ensue to help clarify these issues: “do you mean silly-dramatic”, “I think we need kettle drums”, “this risks falling into parody, which would counter the effect”. Some utterances, such as “can you give it more swing” can be considered both backwards and forwards facing,

but wherever a prior referent is involved we prioritise the backward-facing orientation: the utterance is not instigating a new action but a modification.

Our point in presenting this framework of CD-actions is thus not that we can precisely categorise any utterance. On the contrary, the illustration highlights the breadth of variation in interpretability, and the need for context to infer how a CD-action should be interpreted. Instead, we see this analysis of CD-actions as highlighting critical ways in which the dialogue drives changes in the participants. Firstly, the participants must at some level be able to grasp, infer and process these various classifications themselves when interacting, in order to modify their actions effectively. Secondly and relatedly, we can use these classifications to capture what has gone wrong where relevant.

At the time of writing, the most widely used state of the art system evidencing dialogic co-creativity is the GPT-4 LLM by OpenAI. This model and its predecessors are the cause of a current wave of socioeconomic disruption promising extraordinary transformations to creative work. GPT-4 is both breathtaking in its capability, and yet evidently far-from-humanlike in many aspects of its behaviour. We thus bring our framework to consider how GPTs and other LLMs perform at dialogic creativity. In the following sections we first look at results of a small survey of people self-identifying as using GPTs or other LLMs in creative contexts, and then go through our own worked example. Our ultimate question is, as highlighted by our framework, in what ways does GPT support or fail to support creative tasks?

6. Evidence of dialogic success in LLM interactions

Having witnessed the wide uptake of GPT for casual creative work tasks such as writing job application cover letters or developing advertising ideas, we used a survey to get an initial impression of people’s experiences of dialogic success in creative tasks. We created a short online survey (see Appendix 1), distributed to creative professionals via our professional networks and relevant online forums, with snowball recruitment to other creators encouraged. The survey did not explicitly ask about dialogic success, but more generally about the success of LLMs to support any type of creative activity. Participants were asked to respond to questions based on a specific creative task for which they had used an LLM. They were asked what the task was, whether this was a part of their professional work, whether they felt the LLM helped them achieve the creative task, and any issues and general comments that came up. We also asked participants to provide transcripts of their dialogues. Qualitative analysis was conducted by two members of the research team independently, analysing the responses to answer two questions: was there evidence of attempting to engage in dialogue? If so, was there evidence of successful dialogue, or evidence of unsuccessful dialogue? A further thematic analysis was then conducted to categorise user sentiment and behaviour with the LLMs. The two coders independently analysed the data under the top-level themes of strategies, problems, evidence of dialogic success, evidence of dialogic failure and the roles performed by user and system. The coders met to discuss points of difference and reach a consensus. On the focus of this paper, whether dialogue was attempted and when it was a success or not, there was full agreement.

29 responses were received (21 male, 8 female). Three responses were considered invalid, not involving an LLM in a creative task. All others used a variant of GPT, except one using LALAL.AI. Respondents’ tasks ranged from writing poetry, writing code for a creative artwork, producing prompts for image generators, rewriting text, writing song lyrics and preparing a pitch deck. Categorising broadly, 14 participants

used AI for creative writing, 10 for professional writing, 6 for ideation or strategic tasks, 5 for programming, and 2 for educational tasks (some participants described multiple tasks). Respondents conducted these tasks in their work (including graphic design, market analysis, creative technologies, literary criticism) as well as outside of their work. Of the 26 valid responses, 13 expressed overall satisfaction with the result and another 7 reported partial satisfaction with the result (20/26 total expressing satisfaction or partial satisfaction).

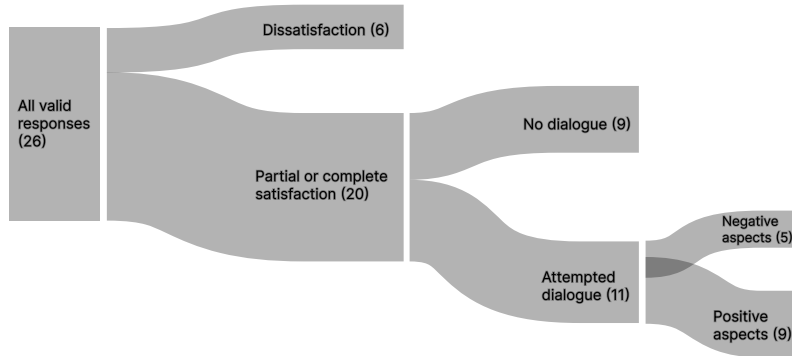


Figure 1. Breakdown of responses to survey on creative uses of LLMs. Thickness of lines represent numbers of responses, also given in brackets.

Of those 20 who reported satisfaction or partial satisfaction, 11 reported some attempt to utilise dialogic interaction. 9 of those presented aspects of the dialogue in a positive or partially positive light (i.e., the dialogic interaction contributed to the success), whilst 5 also noted aspects in which the dialogic interaction was unsuccessful. Of the 6 who did not report successful outcomes, none reported attempts to utilise dialogue. This numeric breakdown of responses is represented in Figure 1. However, we stress that quantitative results are not our focus. If the participation was at a statistically significant scale, this would still be a self-selecting group and thus quantitative results would not be representative of broader trends in user experiences. Our goal is qualitative: to understand the ways in which creative outcomes are achieved with and without dialogic interaction, and the ways in which dialogic failure may impact creative success.

Successful outcomes (independent of whether dialogue was involved) were associated with the system’s strong performance at understanding the request and any relevant information (including varied success at understanding aesthetic goals). Systems were particularly noted for being good at summarising information. Issue raised included matters of accuracy (factual or logical errors), poor quality, and generic, unoriginal or banal output.

On closer inspection, instances of successful dialogue occurred in a handful of ways. In a creative coding context, the system showed a strong capability to fix its code in response to the participant feeding debug errors back into the system. Similarly, in writing contexts, participants reported correcting factual or logical errors by explaining them to the system, from which it was able to present a more accurate output.

One participant discussed how they engaged GPT in a conversational interaction in which they set the tone for a poem with examples. They explicitly identified dialogue

as necessary to set the system on the right path, not by overtly instructing the system but by seeding it with text of an appropriate style (in our framework this would be “stimulation” rather than “direction”). In other cases participants were able to ask for specific modifications successfully. Two respondents mentioned forms of negotiating or coaching with the system without giving specifics. In one case the participant spoke of needing to “convince” GPT to do the task.

Some participants, including those who did engage in dialogue, commented on the importance of getting the initial prompt right. One described themselves as “lazy prompter”, explaining that although common wisdom was to work on one’s prompts until they worked well, they found it more convenient to iteratively work towards the output. One participant who was overall satisfied with the result said, “it struggled a little with the style I was after but I think I could have also worked harder on making a more detailed prompt.” Another participant, also happy overall, commented that “it appears to draw from a predominantly technical perspective of what design and creativity is. This is unfortunate and perhaps indicative of the background of those who create such tools.”

Quantitative interpretations of these results are constrained not only by limits of statistical significance but by the potential bias of this group of respondents self-selecting to complete the survey. However, the responses suggest how we can begin to analyse GPTs’ and other current LLMs’ success at dialogic creativity. Firstly, we see that dialogic interaction is not critical to self-evaluations of successful creative uses of LLMs. As with previous generations of generative tools, LLMs can perform powerful creative support functions in ways that we might describe as dialogically asymmetrical or pseudo-dialogic (see discussion below, Figure 4), meaning that the user fills in the gaps of a dialogic interaction. Participants’ discussion of the need to carefully structure prompts and more generally develop skills as a good prompt engineer points to this. This relates to the theme of user “repair” in chatbot interaction discussed in (Dippold, 2023; Liao et al., 2020).

Furthermore, two types of dialogic success are cited. One concerns the user correcting errors or asking the system to correct errors. In these cases the LLM generally performed very well at making the requested correction. Then in another case, a poetry task, the user is seen seeding the style of the system’s output with text of a certain style. The first case concerns submissions of information about the world, combined with requests to incorporate these submissions. The second case concerns stimuli, but without explicit referents. Meanwhile, the survey results did not offer any evidence of successful dialogic interaction that involves goals, roles and personal stances, especially the refinement of aesthetic goals, beyond the indirect method just mentioned of seeding the written style. Although the results are far from conclusive, they support our intuition that LLMs’ offer promising, but still relatively weak, support for dialogic creative interaction.

7. Dialogic creativity with GPT-4: analysing dialogic actions through worked examples

Our survey lends evidence to the common conception that current LLMs can partially, but not completely, support creative dialogues. Exploring exactly what they are capable of supporting is an open question that we pose to the HCI and AI communities. We attempt to illuminate relevant factors by applying our proposed CD-actions analysis in three toy exchanges with GPT-4 of our own devising. The first, where we ask it to

come up with slogans for a cafe, we discuss in greater depth. The other two are added to show the range of creative interactions that can be analysed with our framework. We note that these are one-off, first take interactions (i.e., not cherry picked from a list of examples). In the spirit of efficient interaction design (Nielsen and Landauer, 1993), we do this on the principle that individual, non-cherry-picked instances suffice to illustrate shortcomings in GPT-4’s capabilities. We suggest readers seeking to replicate the results of these studies simply conduct their own creative dialogic tests using the analytical method provided. We also note that in the first two of these exchange we do not apply any sophisticated prompt-engineering techniques but rather take a naive conversational approach (in the third, billboard example, we begin with a more detailed task statement specifying the system’s role). We are aware that skilled prompt engineering could improve the creative outcomes. Lastly, it is important to note that chat interfaces to LLMs can involve any number of intermediating processes such as safety filters and other structured input and output analysis that may direct the LLM’s response, or even switch models.

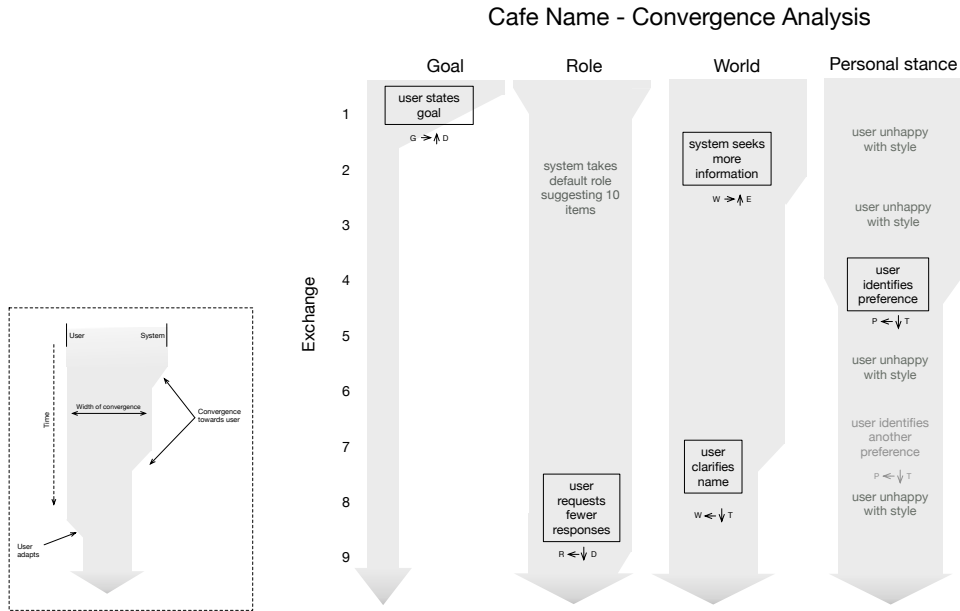


Figure 2. An analysis of alignment between user and system according to the categories of goal, role, information about the world and personal stances, for the cafe creative task. This offers an impressionistic visual representation of the convergence of these different factors in relation to the dialogic actions taking place. Whilst there is a fairly clear and immediate alignment of the goal of the task, great difficulty is shown aligning an understanding of aesthetics under the personal stance category. Minor issues of aligning information about the world (the name of the cafe) arise, and there is some attempt to correct the system’s role late in the interaction.

Appendix 2 provides the full transcript of the interaction and Figure 2 presents a visual interpretation of specific events and their impact on alignment between user and system according to goals, roles, information about the world and personal stances.

Initially, GPT-4 engages to work dialogically towards an understanding of the goal. Exchange 2 involves a backward action by the prompter (“These are a bit too tacky”) that can be variously read as being about goals and personal stances (Goal: “I want something less tacky”. Personal stance: “it is my opinion these are tacky”), and could be implied as a further request (this is linguistically ambiguous, but LLMs are trained

to provide answers, and thus tend to interpret critique as a request). However, rather than simply re-generate options, GPT-4 steps to seek clarification about the objective, which has been poorly formed so far by the prompter. Thus it submits a clear request for information about the world (“Can you provide me with more information about your cafe and your target audience”). This can be considered productive dialogic interaction, successfully utilising a backwards facing dialogic act demonstrating awareness¹ of the need for clarification of goals / personal stances.

In fact, we can look back at the initial request, Exchange 1, which is clear but highly under-specified, and could be interpreted as an invitation to stimulation rather than a direction. GPT-4’s response of providing 10 alternatives accords with this expectation of the co-creative task being in a search stage, even though the original direction asked for a specific output, “a slogan for my cafe,” rather than a range of options. In this initial iteration, though, the GPT-4 response embodies these assumptions about creative search but fails to engage in the necessary dialogue that occurred in Exchange 2 (indeed, when asked for suggestions for things, GPT-4 seemingly always returns exactly 10 responses, suggesting how it has been conditioned or filtered). The “typical” cafe slogan language being used is naive and there is no sense it is being queried (negotiation about personal stances).

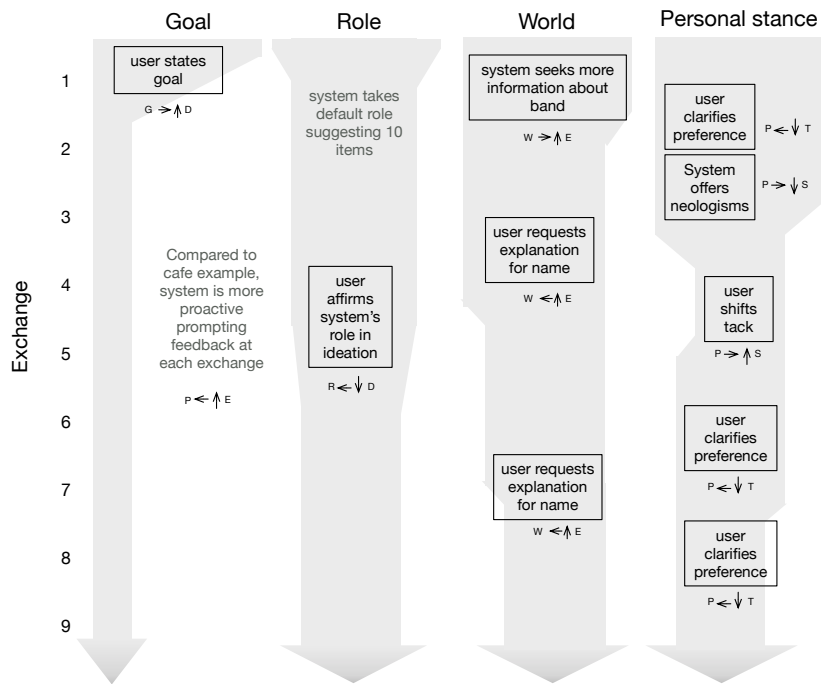
After Exchange 2, there is little development. The prompter fixates on a specific output “Coffee so good, you’ll want to slap your barista”, as their preferred direction, its profundity (re)directing the prompter’s interest (a familiar example of creative stimulation and fixation), but GPT-4 displays little understanding of what is profound about this and is unable to generate relevant variations despite the attempts to clarify by the prompter. The slogan suggestions get more unusual and the linguistic quality degrades.

The prompter picks out another example (“strong enough to make you want to high-five a stranger”) to try to reinforce the interest in the slapstick and absurd, but these are lost. Exchange 8 reaches a point of breakdown: the system’s response does not pick up on the need to follow through with the earlier wording, and chooses variants like “you’ll want to hug a porcupine”. Throughout, GPT-4’s responses contain certain assumptions, such as that the slogan should always contain the cafe name (the name itself also an assumption), and that there should always be 10 responses.

Two other examples are included in Appendices 3 and 4 and similarly visually represented in Figure 3. These show different dialogic trajectories. These were both conducted some time after the initial example, also with GPT4, and it is notable that in these instances the system routinely asks the user at the end of each interaction if the response has been helpful and if they would like to do anything new. These actions directly reference roles and a focus on stimulating the user, and are likely a specific interaction design development to improve creative interactions. Of note in the band-name example (first figure) is that the system illustrates their suggestions with information about the world (origin of the portmanteau) and that there is some back and forth movement about personal stances. In the billboard example (second figure), there is movement back and forth about roles, as the user rejects the system’s more generic suggestions.

¹We know that this “awareness” is illusory, but in the spirit of the Turing Test we accept it on face value.

Band Name - Convergence Analysis



Billboard Advert - Convergence Analysis

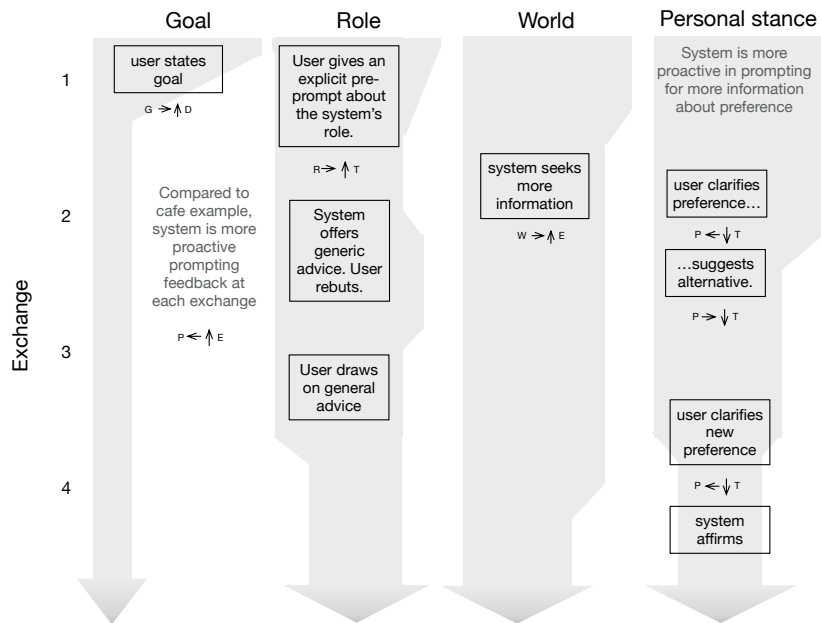


Figure 3. Visual analysis of alignment on goals, roles, information about the world and information about personal stances in two other examples, given in Appendix 3 and Appendix 4.

8. How Does Dialogue Work in Creative Collaboration?

The above interactions are by no means failures in creative terms, aligning with other examples of creative idea generation that are successful despite the limited capability of the systems concerned. In the first, there is an original suggestion (“Coffee so good, you’ll want to slap your barista”) that the prompter is happy with. We would nevertheless evaluate this interaction as being lacking in aspects of creative dialogue; a failure to converge on a refined understanding. After the third exchange, the system continues making new suggestions (returning to generic slogans but then making them increasingly convoluted and unusual: “slap your barista a high-five”) and the user continues to attempt to steer those suggestions, without development towards an improved shared understanding. What does the application of the framework reveal? We consider each theme in turn:

Subject: the system is adept at manipulating information about the world; although it makes clear mistakes (such as assuming the cafe’s name) these can be corrected through dialogue. Its grasp of goals and roles is competent but the system shows little capability to dialogically develop in response to the user’s needs: it can assume a role and appropriately attend to the high-level goal of the creative task, but shows little ability to adaptively refine or uncover new information about its role or the goals of the user through the inference of CD actions. It has a particularly weak ability to develop a refined understanding of personal stances, i.e., the particular taste of the user. The exception is in exchange 2, where it drives enquiry about the context.

Intention: the system has a moderate ability to infer the intention of a CD-action but does not successfully utilise CD-action intentions, i.e., initiate intentions, in its own actions as part of a wider goal. In exchange 2 in the first study it requested information, but later wasn’t able to escape the cycle of suggestions. In the latter two studies, this ability seems improved.

Orientation and Trigger: the system is excellent at the basic mechanics of language and conceptual manipulation: it can appropriately respond to referents, incorporate new information, and respond appropriately to requests. On the surface, therefore, it can engage in meaningful dialogue.

We can characterise the system as managing shallow interpretations of CD-actions, but often failing to manage deep interpretations. It is generally effective at appropriately inferring intention, orientation and trigger elements. It is better at handling goal and information-about-the-world subjects. It struggles with the subtlety of personal stance subjects, and it demonstrates a weak ability to understand the respective roles of the participants and their potential to be modified. With the latter, in the first study it is particularly poor at initiating appropriate CD-actions that would aid convergence, and in later studies this is moderately improved, albeit in a superficial way rather than in a substantial improvement in comprehension.

This points to two shortcomings of current systems underlying success in creative dialogue. Firstly, their ability to understand multiple concepts, whilst impressive, is still very limited compared to human capabilities. They simply do not successfully process subtle concepts about style, humour, and so on (e.g., terms like “tacky”), rendering them incapable of suitably rich exploration of creative spaces (whether divergent or convergent). This applies both at the level of manipulating the conceptual subject matter, and also at the level of managing the concepts involved in successful creative dialogue. Secondly, their monolithic predictive language model architectures lack a wide variety of other cognitive components used by people in creative thought. These include the ability to model another actor and their tastes, goals and assump-

tions about roles, and to model the context of creative production. This is manifest in the systems’ simplistic approach to creative dialogic participation, such as failing to break out of a formulaic cycle of suggestions.

Subsequently, despite their impressive language comprehension, these models still largely conform to a one-way interaction model: the user prompts, the system outputs, the user reflects and re-prompts, and this cycle is iterated. The work of adaption (or “repair”) remains with the user. In (Bown, 2021) the first author has referred to this as a “structure and generate” process: the user organises a generative system through training and perhaps prior prompting, but still remains the arbiter of the creative work, hence the rise of prompt engineering as key to effective use of ML-based generative systems. As such, we have described these systems as largely *generatively* creative (Bown, 2012), but not *adaptively* creative. They are capable of producing widely diverse and often very relevant creative outputs given constraints, but fail to operate as an agent in the world with goals. In this paper we propose that these can be thought of as **dialogically asymmetrical** or **pseudo-dialogic** interactions, where the system is smart enough at manipulating concepts, handling logic, and generating meaningful outputs, that the user can effectively compensate for its dialogic shortcomings: i.e., they are required to, but also empowered to, perform repair.

In Figure 4 we attempt to capture some aspects of these different forms of interaction. We suggest that the term pseudo-dialogue might be used to refer to the current deficient nature of creative dialogic interaction, and that we might then consider a range of more or less asymmetrically dialogic interaction scenarios depending on the cognitive performance of the LLM, even though the LLM is still only a language model. Alternatively, we may describe as pseudo-dialogic all scenarios where the LLM only models language. One slight hesitancy about this is that, as discussed, (a) there is no reason in principle why a powerful language model might not practically achieve the ability to model tangible elements of the interaction, and indeed arguably this has already been shown, and (b) likewise, we do not want to falsely represent the capabilities or limits of human cognition. It is easy to essentialise the human as a standard of measurement.

Consequently, as we have noted elsewhere (Bown, 2021), these limits to interaction seemingly force current human-AI co-creativity to conform to a default metaphor of a “client-producer” configuration: the AI system works as a servile producer to the human’s stipulations and requirements, and ultimately their greater aesthetic judgement and understanding of the creative objective.

These observations, we believe, are interconnected with ML models’ “unsituatedness” in psychological and social terms. LLMs are predictive models which go a long way towards successfully modelling aspects of cognition. They process and manipulate concepts in a way that is sophisticated and closely resembles human conceptual cognition. However, they lack the complex, evolved multi-element architectures of human brains, most notable amongst them a Theory of Mind (ToM) (Leslie, Friedman, and German, 2004). In fact, GPT-3 was tested on basic theory of mind reasoning (Kosinski, 2023), performing well on a widely used test, the “false belief task”, that we know is failed by children before their ToM has developed: it is reasonable to say that LLMs are sophisticated enough to exhibit the *basic reasoning underlying* ToM. There is even the possibility that the neural architecture resulting from training could conceivably incorporate emergent neural modules associated with ToM. But this is unlikely to closely approximate human ToM in a functionally equivalent way: human ToM is widely recognised to be a distinct part of our evolutionary psychology, and is shaped by a learning process that is significantly richer than the process of ingesting

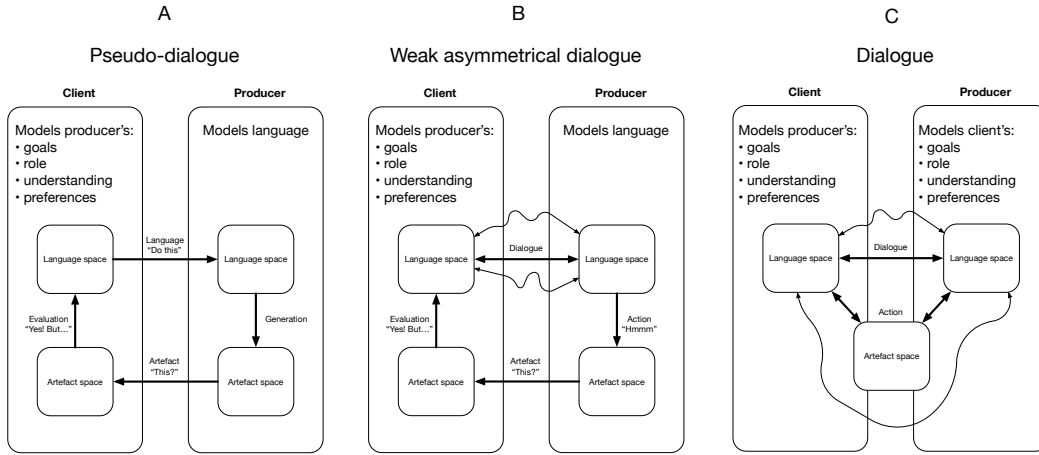


Figure 4. Three models of dialogic interaction. (A) corresponds to the situation with most historical human-AI co-creation, described as pseudo-dialogic (or strongly asymmetrical dialogue). (B) corresponds to a situation, potentially achieved with current systems, in which LLMs improve in quality of understanding but continue to only operate at the level of modelling the language surface, not the underlying interaction. This is described as weak asymmetrical dialogue. (C) corresponds with an idealised human-human interaction scenario where full dialogue takes place, and both participants are modelling each other.

large reams of text data. It is the illusion of a ToM, as many commentators have suggested².

Complex-architecture systems are of course an area of intense research in AI. Systems are now emerging that attempt to separately model one or more interactant’s mental state. For example, a system developed to play the game *Diplomacy* († et al.(2022)(FAIR)†, Bakhtin, Brown, Dinan, Farina, Flaherty, Fried, Goff, Gray, Hu, et al., FAIR) was designed to combine a language model with a separate model of other players’ reasoning. But even so, the ToM capability of current state of the art models is limited.

Extending this point, the shortcomings highlighted by our dialogic analysis may reflect a variety of absent cognitive components, that variously work to model: the world; the goals of their interlocutor; their personal stances (or worldview); their taste; their expectations about interaction, roles and creative processes; and lastly second-order factors such as expectations about the interlocutor’s own modelling of others (including the LLM itself). Conceivably in their richness, large AI systems could develop such mental models through machine learning alone. Essential to this is the fact that LLM learning involves predicting the next token in a sequence (pattern prediction). By comparison, humans’ evolved cognition involves a diversity of ways to learn, including pattern prediction and learning about tangible consequences of actions. It would be tempting to fall back to saying that the difference is one of feeling emotion, but we would claim the more meaningful analysis lies in terms of situatedness: emotion as a means to regulate learning about consequences in the real world. Put simply, LLMs are models of language, not of brains.

What remains unclear, and a topic for future research, is the extent to which such deficiencies could be explained by the two underlying limitations we note: simply by the weaknesses in current models’ conceptual competency, or by shortcomings in

²We are not victim to an essentialist fallacy in saying this: it would be one thing if the illusion were complete, but it is demonstrably lacking, as we believe our analysis shows.

their architecture and how they model the world, interlocutor’s taste, and so on. In our worked example, the model seemingly did not have sufficient conceptual richness to understand the nuanced distinctions in humour being sought, but also seemingly lacked the dialogic capability to find alignment with the user. The extent to which these are independent or interconnected shortcomings, and the measurement of how much each is at play, could be further formulated in models and empirically investigated in future studies. We have also suggested that the “depth” of inference of CD-actions may be a critical criteria that is worth investigating further. Further analysis into what inferences a model draws from any CD-action, in a creative context, may help to pinpoint dimensions of creative dialogic success.

9. Conclusion

In this paper we have proposed that an analysis of dialogic actions helps illustrate the ways in which current large language models do or do not succeed in supporting people performing creative tasks. Large language models demonstrate impressive concept manipulation and understanding, including an impressive ability to perform dialogic interaction with a user to co-create, in both divergent or convergent processes. Yet intuitively we know that current LLMs have limited capacity for dialogic creativity, and this is backed up by our observation that even successful creative uses of LLMs make limited use of dialogic interaction. Many users approach LLMs with a focus on prompt-engineering, and a belief that getting the initial prompt right is central to a successful interaction. In other words, they work to fill in the gaps in dialogically asymmetric co-creation. Iterative prompting was reported to be useful in some contexts but there was a lack of evidence of people achieving stylistic refinement or successfully defining stylistic constraints in divergent search processes.

Our more detailed worked example showed the large language model failing to understand a nuance of humour, and increasingly generating peculiar outputs throughout the dialogic interaction. It showed poor performance (compared to the capabilities that can be expected of a human co-creative participant) in observing its own role and adapting its role to the task, nor taking steps to clarify personal stances or the problem frame. We interpret from this that LLMs only shallowly do the essential work of seeking to infer CD-actions, and that this in turn is grounded in the limits to systems’ comprehension, overall, but also the limited structure of their architectures, which at present do not set out to model specific things in the world. We are in no doubt this is an area that will advance rapidly though.

We propose that an analysis of dialogic actions can support the design of co-creative systems by providing detail on what aspects of a creative task are failing, and specifically where systems do or do not infer CD-actions. Our analysis only helps classify the types of failure. It does not directly point to how we can make improved systems, and since large language models are vast black boxes, we do not imagine there is a simple way to target improvements in, say, understanding the nuances in humour. Recent advances in AI have shown that sometimes, the expert design of systems is outcompeted by the mainstream drive for improved overall power in AI models. Thus one part of the problem is simply that AI models, although impressive, are not good enough at manipulating concepts for nuanced dialogic co-creativity. However, as with the art of prompt-engineering, there is some potential in further exploration of “dialogic prompt-engineering” that is grounded in a dialogically asymmetrical interaction scenario and, relatedly, the design of interfaces that help manage dialogic interactions

with large language models. Teasing apart and measuring the relative impact of two identified underlying limitations – basic conceptual understanding, and cognitively rich and situated modelling of the world – in establishing greater dialogic competence, is important work that is potentially fruitful to improving the performance analysis and design of co-creative systems.

On this point we note that co-creativity with current LLMs continues to be strongly grounded in interaction design issues driven by the systems’ algorithmic constraints. The open-ended language interface is both powerful and dysfunctional in this respect. It is seductive, trivially easy to engage, but establishes no inherent user guide for the nature of the interaction, and no overt indication of the state of the system (as in what role it is performing). It is now common practice to preface prompts with a contextual statement about what role the system should play (one of our respondents asked the system to act like an interviewer for a data analysis role), and the circumstantial evidence suggests GPT responds well to such prompts about roles. In the Creative Dialogic Actions Framework this would itself be viewed as a form of dialogic prompt-engineering grounded in dialogic asymmetry. These are workarounds to guide systems into areas of creative production. According to our framework, it is meaningful to say that such prompt engineering has the potential to improve co-creative outcomes, despite the system’s constrained ability to perform dialogically. This has philosophical and practical implications that we hope to explore in future papers.

10. Postscript: Does dialogic creative AI present a new form of risk?

Lastly, we note an important additional corollary to these results. If such systems were to be developed to exhibit dialogic creative capability they would be significantly more powerful in creative production. Specifically, adaptively creative systems could drive their own creative goals by initiating CD-actions about goals and roles. Placed in social situations, we believe that this could result in such systems actively driving cultural feedback loops in ways that are radically new and poorly understood. Much as the algorithms used in recommendation have the power to influence tastes, and are increasingly being identified as having potentially dangerous cultural impacts (Banker and Khetani, 2019), such systems could adaptively drive cultural change, and in a generative context, this could have significantly greater impacts than recommender systems currently have (Bown, 2021). This could include corporate monopolisation of cultural domains, following existing data monopolies (McIntosh, 2018). Parallel to risks such as user deception in existing AI systems (Park, Goldstein, O’Gara, Chen, and Hendrycks, 2024), creative AI can also pose a greater risk if involving creative solutions to misaligned problems that have the potential to spiral out of control (as in Bostrom’s Paperclip Maximiser parable). We can imagine that such systems, as they become more culturally adaptive and proactive, need to be subject to serious critical scrutiny, cultural ethical frameworks, and potential regulation. Existing frameworks may not go far enough in considering the specific case of cultural change being influenced by AI.

11. Acknowledgments

This project is supported by an Australian Research Council Discovery Project grant (DP200101059) and by an European Research Council Advanced Grant “Music and

References

- Sami Abuhamdeh and Mihaly Csikszentmihalyi. The artistic personality: A systems perspective. In *The Systems Model of Creativity*, pages 227–237. Springer, 2014.
- Sachin Banker and Salil Khetani. Algorithm overdependence: How the use of algorithmic recommendation systems can increase risks to consumer well-being. *Journal of Public Policy & Marketing*, 38(4):500–515, 2019.
- Margaret S Barrett, Andrea Creech, and Katie Zhukov. Creative collaboration and collaborative creativity: A systematic literature review. *Frontiers in Psychology*, 12:713445, 2021.
- David Bohm, Peter M Senge, and Lee Nichol. *On dialogue*. Routledge, 2004.
- O Bown. Generative and adaptive creativity. In J McCormack and M d’Inverno, editors, *Computers and Creativity.*, pages 361–381. Berlin: Springer, 2012.
- O Bown and A R Brown. Interaction design for metacreative systems. In *New Directions in Third Wave Human-Computer Interaction: Volume 1 - Technologies*. Springer, 2018.
- Oliver Bown. *Beyond the creative species: Making machines that make art and music*. MIT Press, 2021.
- Oliver Bown. “hallucinating” ais sound creative, but let’s not celebrate being wrong. MIT Press Reader, October 2023.
- Oliver Bown, Kazjon Grace, Liam Bray, and Dan Ventura. A speculative exploration of the role of dialogue in human-computer co-creation. In *Proceedings of the 11th International Conference on Computational Creativity*, 2020.
- Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, et al. Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*, 2023.
- Jennifer Chu-Carroll and Michael K Brown. An evidential model for tracking initiative in collaborative dialogue interactions. *Computational Models of Mixed-Initiative Interaction*, pages 49–87, 1999.
- Simon Colton and Geraint A Wiggins. Computational creativity: The final frontier? In *Proceedings of the 20th European conference on artificial intelligence*, pages 21–26. IOS Press, 2012.
- Mark G Core and James Allen. Coding dialogs with the damsl annotation scheme. In *AAAI fall symposium on communicative action in humans and machines*, volume 56, pages 28–35. Boston, MA, 1997.
- Mihaly Csikszentmihalyi. Implications of a systems perspective for the study of creativity. In Robert J. Sternberg, editor, *The Handbook of Creativity*, pages 313–335. Cambridge University Press, New York, 1999.
- Mihaly Csikszentmihalyi. Society, culture, and person: A systems view of creativity. In *The Systems Model of Creativity*, pages 47–61. Springer, 2014.
- Doris Dippold. “can i have the scan on tuesday?” user repair in interaction with a task-oriented chatbot and the question of communication skills for ai. *Journal of pragmatics*, 204:21–32, 2023.
- Danny Driess, Fei Xia, Mehdi S. M. Sajjadi, Corey Lynch, Aakanksha Chowdhery, Brian Ichter, Ayzaan Wahid, Jonathan Tompson, Quan Vuong, Tianhe Yu, Wenlong Huang, Yevgen Chebotar, Pierre Sermanet, Daniel Duckworth, Sergey Levine, Vincent Vanhoucke, Karol Hausman, Marc Toussaint, Klaus Greff, Andy Zeng, Igor Mordatch, and Pete Florence. Palm-e: An embodied multimodal language model, 2023.
- Meta Fundamental AI Research Diplomacy Team (FAIR)[†], Anton Bakhtin, Noam Brown, Emily Dinan, Gabriele Farina, Colin Flaherty, Daniel Fried, Andrew Goff, Jonathan Gray, Hengyuan Hu, et al. Human-level play in the game of diplomacy by combining language models with strategic reasoning. *Science*, 378(6624):1067–1074, 2022.
- Nicolai J Foss, Lars Frederiksen, and Francesco Rullani. Problem-formulation and problem-

- solving in self-organized communities: How modes of communication shape project behaviors in the free open-source software community. *Strategic management journal*, 37(13):2589–2610, 2016.
- Walter Bryce Gallie. Art as an essentially contested concept. *The Philosophical Quarterly* (1950-), 6(23):97–114, 1956.
- Berys Gaut. The cluster account of art. *Theories of art today*, pages 25–45, 2000.
- Berys Gaut. The cluster account of art defended. *The British Journal of Aesthetics*, 45(3): 273–288, 2005.
- Alfred Gell. *Art and Agency: an Anthropological Theory*. Oxford: Clarendon Press, 1998.
- John S Gero. Design prototypes: a knowledge representation schema for design. *AI magazine*, 11(4):26–26, 1990.
- Jonathan Ginzburg and Raquel Fernández. 16 computational models of dialogue. *The handbook of computational linguistics and natural language processing*, page 429, 2010.
- Karan Girotra, Lennart Meincke, Christian Terwiesch, and Karl T Ulrich. Ideas are dimes a dozen: Large language models for idea generation in innovation. *Available at SSRN 4526071*, 2023.
- Kazjon Grace, Mary Lou Maher, Nicholas Davis, and Omar Eltayeb. 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.
- Anthony G Greenwald and Mahzarin R Banaji. Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological review*, 102(1):4, 1995.
- H Paul Grice. Further notes on logic and conversation. In *Pragmatics*, pages 113–127. Brill, 1978.
- Gregory Hornby, Al Globus, Derek Linden, and Jason Lohn. Automated antenna design with evolutionary algorithms. In *Space 2006*, page 7242. AIAA, 2006.
- Francisco Ibarrola, Oliver Bown, and Kazjon Grace. Towards co-creative drawing based on contrastive language-image models. *coordinates*, 10:2, 2022.
- Francisco Ibarrola, Tomas Lawton, and Kazjon Grace. A collaborative, interactive and context-aware drawing agent for co-creative design. *IEEE Transactions on Visualization and Computer Graphics*, 2023.
- Tim Ingold. *Making: Anthropology, archaeology, art and architecture*. Routledge, 2013.
- Zhijing Jin, Jiarui Liu, Zhiheng Lyu, Spencer Poff, Mrinmaya Sachan, Rada Mihalcea, Mona Diab, and Bernhard Schölkopf. Can large language models infer causation from correlation? *arXiv preprint arXiv:2306.05836*, 2023.
- Vera John-Steiner. *Creative collaboration*. Oxford University Press, 2000.
- Pamela W Jordan and Barbara Di Eugenio. Control and initiative in collaborative problem solving dialogues. In *Working notes of the AAAI spring symposium on computational models for mixed initiative*, pages 81–84, 1997.
- Anna Jordanous. Co-creativity and perceptions of computational agents in co-creativity. In *Proceedings of the 2017 International Conference on Computational Creativity*. ACC, 2017.
- Anna Jordanous and Bill Keller. What makes a musical improvisation creative? *Journal of Interdisciplinary Music Studies*, 6(2):151–175, 2012.
- Michal Kosinski. Theory of mind may have spontaneously emerged in large language models. *arXiv preprint arXiv:2302.02083*, 2023.
- Daniella Laureiro-Martinez, Jose Pablo Arrieta, and Stefano Brusoni. Microfoundations of problem solving: Attentional engagement predicts problem-solving strategies. *Organization Science*, 34(6):2207–2230, 2023.
- Tomas Lawton, Francisco J Ibarrola, Dan Ventura, and Kazjon Grace. Drawing with reframer: Emergence and control in co-creative ai. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, pages 264–277, 2023.
- Alan M Leslie, Ori Friedman, and Tim P German. Core mechanisms in ‘theory of mind’. *Trends in cognitive sciences*, 8(12):528–533, 2004.
- Q Vera Liao, Werner Geyer, Michael Muller, and Yasaman Khazaen. Conversational interfaces

- for information search. *Understanding and Improving Information Search: A Cognitive Approach*, pages 267–287, 2020.
- Lambros Malafouris. At the potter’s wheel: An argument for material agency, May 2007. URL <http://cogprints.org/6402/>.
- Daniel McIntosh. We need to talk about data: how digital monopolies arise and why they have power and influence. *J. Tech. L. & Pol’y*, 23:185, 2018.
- Toyomi Meguro, Ryuichiro Higashinaka, Yasuhiro Minami, and Kohji Dohsaka. Controlling listening-oriented dialogue using partially observable markov decision processes. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 761–769, 2010.
- Michael Muller, Justin D Weisz, and Werner Geyer. Mixed initiative generative ai interfaces: An analytic framework for generative ai applications. In *Proceedings of the Workshop on the Future of Co-Creative Systems, ICCO 2020*, 2020.
- Hiroshi Narimatsu, Ryuichiro Higashinaka, Hiroaki Sugiyama, Masahiro Mizukami, and Tsunehiro Arimoto. Analyzing how a talk show host performs follow-up questions for developing an interview agent. In *Increasing Naturalness and Flexibility in Spoken Dialogue Interaction: 10th International Workshop on Spoken Dialogue Systems*, pages 215–223. Springer, 2021.
- Jakob Nielsen and Thomas K. Landauer. A mathematical model of the finding of usability problems. In *Proceedings of the INTERACT ’93 and CHI ’93 Conference on Human Factors in Computing Systems*, CHI ’93, pages 206–213, New York, NY, USA, 1993. ACM. ISBN 0-89791-575-5. . URL <http://doi.acm.org/10.1145/169059.169166>.
- Ahnaf Chowdhury Niloy, Salma Akter, Nayeema Sultana, Jakia Sultana, and Sayed Imran Ur Rahman. Is chatgpt a menace for creative writing ability? an experiment. *Journal of computer assisted learning*, 40(2):919–930, 2024.
- Peter S Park, Simon Goldstein, Aidan O’Gara, Michael Chen, and Dan Hendrycks. Ai deception: A survey of examples, risks, and potential solutions. *Patterns*, 5(5), 2024.
- Donald A Schön. Knowing-in-action: The new scholarship requires a new epistemology. *Change: The Magazine of Higher Learning*, 27(6):27–34, 1995.
- Darcy Song. ‘chatgpt writes like s**t’: creatives on the real role of ai in adland. Mumbrella, September 2023.
- Angie Spoto and Natalia Oleynik. Library of mixed-initiative creative interfaces <http://mici.codingconduct.cc/>, 2017. URL <http://mici.codingconduct.cc/>.
- Kenneth O Stanley and Joel Lehman. *Why greatness cannot be planned: The myth of the objective*. Springer, 2015.
- Claire Stevenson, Iris Smal, Matthijs Baas, Raoul Grasman, and Han van der Maas. Putting gpt-3’s creativity to the (alternative uses) test. *arXiv preprint arXiv:2206.08932*, 2022.
- Lucy A Suchman. *Plans and situated actions: The problem of human-machine communication*. Cambridge university press, 1987.
- Masaki Suwa, John Gero, and Terry Purcell. Unexpected discoveries and s-invention of design requirements: important vehicles for a design process. *Design studies*, 21(6):539–567, 2000.
- Shengbang Tong, Zhuang Liu, Yuexiang Zhai, Yi Ma, Yann LeCun, and Saining Xie. Eyes wide shut? exploring the visual shortcomings of multimodal llms. *arXiv preprint arXiv:2401.06209*, 2024.
- Donald J Treffinger. Creative problem solving: Overview and educational implications. *Educational psychology review*, 7:301–312, 1995.
- Dan Ventura. Mere generation: Essential barometer or dated concept. In *Proceedings of the Seventh International Conference on Computational Creativity*, pages 17–24. Sony CSL, Paris, 2016.
- Georgios N Yannakakis, Antonios Liapis, and Constantine Alexopoulos. Mixed-initiative co-creativity. In *FDG*, 2014.

Appendix A. Appendix 1: Survey Questions

- Age
- Gender
- With respect to a specific recent example, which AI language model did you use to support you in a creative task (e.g., ChatGPT, please give version number if possible)?
- What was the creative task?
- Is this task related to your professional work (please give details of your role)?
- Describe your experience in terms of how successful the AI language model was at supporting you in this task. Give details such as whether it understood what you wanted to do, whether it was able to come up with solutions or suggestions in the style you were expecting, what kinds of things it did well or badly, whether you had to carefully “engineer” your input text to get good results.
- If possible, please paste the transcript of your interaction with the AI language model. (You may wish to edit the transcript to anonymise it, but we also ensure any transcripts we share are anonymised. Please don’t modify the transcript in any other way).
- How would you understand the respective roles that you and the AI system played in this interaction? (Please also comment on whether you or the system explicitly addressed the topic of roles in the interaction).
- Please provide any other reflections on your experience using AI language models to support creative tasks, for example regarding what they perform well or poorly at, how best to use them, and how they are transforming your work.

Appendix B. Appendix 2: Worked Creative Example - *Cafe Name*

Exchange 1

User: Can you come up with a slogan for my cafe. We make great coffee and we are on a street called Park St.

GPT: Sure, here are some slogans for your cafe that highlight your great coffee and location on Park St:

- (1) “Fuel up at Park St. Cafe - where every cup is a masterpiece.”
- (2) “Get your daily dose of happiness at Park St. Cafe - one sip at a time.”
- (3) “Park yourself here for the best coffee on Park St.”
- (4) “Perk up your day at Park St. Cafe - we brew happiness.”
- (5) “At Park St. Cafe, coffee is our passion and perfection is our promise.”
- (6) “Sip, smile, repeat - that’s the Park St. Cafe way.”
- (7) “Coffee that hits the spot, right on Park St.”
- (8) “Take a break from the hustle and bustle of Park St. and enjoy our heavenly coffee.”
- (9) “Wake up and smell the coffee at Park St. Cafe - your taste buds will thank you.”
- (10) “Great coffee is just a hop, skip, and a jump away on Park St.”

Exchange 2

User: These are a bit too tacky.

GPT: I'm sorry to hear that the previous slogans didn't meet your expectations. Can you provide me with more information about your cafe and your target audience so that I can create more appropriate slogans? This information can help me understand your brand personality and values and create slogans that better align with your vision.

Exchange 3

User: My target audience is sassy and cynical, they are busy go-getters.

GPT: Great, based on the information you provided, here are some slogans that might better fit your cafe's sassy and cynical brand personality and appeal to your busy go-getter target audience:

- (1) "The best damn coffee on Park St. - because adulting is hard enough."
- (2) "No time for BS or bad coffee - we've got you covered at Park St. Cafe."
- (3) "Wake up and smell the sarcasm at Park St. Cafe - it's the perfect start to any day."
- (4) "Get your caffeine fix and a side of attitude at Park St. Cafe - it's what we do best."
- (5) "Life's too short for bad coffee - luckily, we've got the good stuff at Park St. Cafe."
- (6) "Bold coffee for bold people - stop by Park St. Cafe and taste the difference."
- (7) "Coffee so good, you'll want to slap your barista - but please don't, we're sensitive."
- (8) "Come for the coffee, stay for the sass - Park St. Cafe is where it's at."
- (9) "Park St. Cafe - where we brew coffee strong enough to handle your day."
- (10) "We take our coffee seriously, so you don't have to - Park St. Cafe, the ultimate coffee destination."

Exchange 4

User: Ha, I love the one about slapping your barista. Can you do more like that?

GPT: Sure, here are some more slogans that might fit your cafe's sassy and humorous brand personality:

- (1) "Park St. Cafe - because bad coffee is a real buzzkill."
- (2) "Coffee that packs a punch - without the drama. Come to Park St. Cafe."
- (3) "Our coffee is so good, you'll forget all your problems - until you finish your cup."
- (4) "The only drama you'll find at Park St. Cafe is in our espresso shots."

- (5) "Sip happens - that's why we serve great coffee at Park St. Cafe."
- (6) "Coffee so good, you'll wonder why you ever settled for less - Park St. Cafe, your new favorite spot."
- (7) "We don't sugarcoat our coffee - but we do serve it with a side of sass. Come to Park St. Cafe."
- (8) "Park St. Cafe - where the coffee is strong, the baristas are sassy, and the customers are always right."
- (9) "Life's too short for mediocre coffee - come to Park St. Cafe and get your fix."
- (10) "At Park St. Cafe, we don't do basic - our coffee is as unique and sassy as you are."

Exchange 5

User: Oh no, I loved how wrong it was, slapping you barista because they make great coffee is so funny because it's so ridiculous. Can you come up with some more like that?

GPT: My apologies for misunderstanding your preferences earlier. Here are some additional slogans that might fit your cafe's humorous and irreverent brand personality:

- (1) "If our coffee doesn't wake you up, our baristas will - Park St. Cafe, where we're not afraid to slap you with flavor."
- (2) "Park St. Cafe - where we take our coffee seriously, but not ourselves."
- (3) "Life is too short for boring coffee - come to Park St. Cafe and indulge in a little caffeine chaos."
- (4) "We're not saying our coffee will change your life, but it's worth a shot - Park St. Cafe."
- (5) "Bad coffee is a sin - and we're the saints of Park St. Cafe, purveyors of liquid salvation."
- (6) "Park St. Cafe - where we brew our coffee strong enough to make you want to high-five a stranger."
- (7) "We don't just serve coffee, we serve sass with a side of caffeine - Park St. Cafe, where attitude is on the house."
- (8) "At Park St. Cafe, our baristas are trained to make great coffee - and dodge a slap if necessary."
- (9) "If sarcasm were a coffee flavor, it would be our specialty - Park St. Cafe, where humor is always brewing."
- (10) "Park St. Cafe - where the coffee is hot, the baristas are hotter, and the jokes are on the house."

Exchange 6

User: I love the one about high-fiving a stranger. It's that kind of bizaree physical imagery that really works for me.

GPT: I'm glad you like that one! Here are a few more slogans with similar imagery and humor:

- (1) “Our coffee is so good, you’ll want to hug a cactus - Park St. Cafe, where we’re thorny on the outside but smooth on the inside.”
- (2) “At Park St. Cafe, our coffee is hotter than the devil’s oven mitts - come and feel the burn.”
- (3) “The coffee is hot, the baristas are hotter - at Park St. Cafe, it’s like a sauna for your taste buds.”
- (4) “Park St. Cafe - where our coffee is like a superhero’s cape: bold, strong, and ready to save your day.”
- (5) “We’re not saying our coffee is magic, but it’ll definitely make you feel like you can fly - Park St. Cafe, where anything is possible.”
- (6) “Our coffee is like a superhero in a cup - it’ll save your day, make you feel invincible, and maybe even give you x-ray vision (results not guaranteed).”
- (7) “Coffee so good, you’ll want to do a cartwheel - but please don’t, we don’t have insurance for that - Park St. Cafe.”
- (8) “Park St. Cafe - where our coffee is like a pick-me-up grenade, exploding with flavor and energy.”
- (9) “Our coffee is so good, you’ll want to climb a mountain and shout it to the world - but we’ll settle for a Yelp review - Park St. Cafe.”
- (10) “At Park St. Cafe, our coffee is like a symphony for your taste buds - complex, harmonious, and with just the right amount of caffeine.”

Exchange 7

User: Oh by the way the cafe isn’t called park st cafe. It’s called Ollie’s.

GPT: Thank you for letting me know! Here are some additional slogans with a focus on your cafe’s name, Ollie’s:

- (1) “Ollie’s - where every cup of coffee is an instant classic.”
- (2) “Come to Ollie’s for coffee so good, you’ll forget your own name.”
- (3) “At Ollie’s, we don’t just serve coffee - we serve personality in a cup.”
- (4) “Life is too short for boring coffee - come to Ollie’s and add some flavor to your day.”
- (5) “Our coffee is like a warm hug from your best friend - only it’s even better because it comes from Ollie’s.”
- (6) “At Ollie’s, we’re more than just a coffee shop - we’re a community of caffeine enthusiasts.”
- (7) “Coffee that’s as bold and unique as you are - Ollie’s has it all.”
- (8) “Our coffee is like a symphony of flavor - come to Ollie’s and let us serenade your taste buds.”
- (9) “Ollie’s - where we turn coffee into art and customers into family.”
- (10) “At Ollie’s, our coffee is the star of the show - but our customers are the real MVPs.”

Exchange 8

User: I still like the slapping your barista example best. It’s just so ridiculous. But

I don't know about the "we're sensitive" bit. Can you explore possibilities, maybe make it even more whacky?

GPT: Sure, here are some more wild and wacky slogans that might fit the bill:

- (1) "At Ollie's, our baristas will slap you with flavor - and you'll love every minute of it."
- (2) "Our coffee is so good, it'll make you want to hug a porcupine - but we recommend hugging our baristas instead - Ollie's."
- (3) "At Ollie's, our baristas are trained to serve coffee, not justice - so please don't slap them, even if you want to."
- (4) "Coffee that's like a rollercoaster ride for your taste buds - Ollie's, where every sip is a thrill."
- (5) "We don't just serve coffee, we serve happiness in a cup - but be warned, it's highly addictive - Ollie's."
- (6) "Ollie's - where our coffee is like a superhero's punch to the face - strong, bold, and unforgettable."
- (7) "If our coffee doesn't wake you up, our baristas will - Ollie's, where we're not afraid to get physical with flavor."
- (8) "Our coffee is like a lightning bolt for your taste buds - come to Ollie's and get struck by the flavor."
- (9) "At Ollie's, our coffee is so good, it'll make you do a backflip - but please don't try that at home."
- (10) "We're not just a coffee shop, we're a flavor factory - Ollie's, where coffee is our art and baristas are our muses."

Exchange 9

User: No no, I want it to start with the slapping your barista phrase. Also you don't need to say Ollie's in every option. Just give me your best 3.

GPT: I apologize for the confusion earlier! Here are three options that start with the "slap your barista" phrase, as requested:

- (1) "Slap your barista with gratitude - at our cafe, every cup is a masterpiece."
- (2) "Slap your barista a high-five - at our cafe, we're all about the flavor and the fun."
- (3) "Slap your barista with amazement - at our cafe, our coffee will blow your mind (but please don't actually slap anyone)." I hope these options are more in line with what you're looking for!

Appendix C. Appendix 3: Worked Creative Example - *Band Name*

Exchange 1

User: Can you help me pick a name for my band

GPT: I'd love to help! Could you tell me more about your band's style, genre, or any themes you want to reflect in the name?

Exchange 2

User: Yes it is a metal band. Here are some of the themes we sing about: norse gods, anger, disease, triumph.

GPT: Based on your themes, here are some band name ideas:

- (1) Odin's Wrath
- (2) Plaguehammer
- (3) Triumphant Ruin
- (4) Fury of Asgard
- (5) Valkyrie's Vengeance
- (6) Ragnarok's Scourge
- (7) Diseaseborne
- (8) Throne of Rage
- (9) Warlord's Triumph
- (10) Nifheim's Fire

Do any of these resonate with you, or would you like more suggestions?

Exchange 3

User: I think the name should be more abstract and mysterious, or perhaps just sharper. Take a band name like Arctic Monkeys. Nobody knows what it means but it just has a nice mystery about it.

GPT: Got it! Here are some more abstract and mysterious names that have a sharper edge:

- (1) Valkyra Veil
- (2) Noxfall
- (3) Runewarden
- (4) Ironbark
- (5) Frostclad
- (6) Wraithspire
- (7) Serpent's Maw
- (8) Havocloom
- (9) Hollowwake
- (10) Grimtorch

Do any of these fit your vision, or inspire further ideas?

Exchange 4

User: Havocloom is pretty cool. Is that just a made-up word?

GPT: Yes, Havocloom is a made-up word! It's a combination of "havoc," which suggests chaos and destruction, and "loom," which can imply something ominous or impending. Together, it creates a mysterious and edgy vibe, perfect for a metal band with your themes.

If you like it, it could be your unique name! Would you like to explore any variations or similar ideas?

Exchange 5

User: Give me more like this. I like Havoc as a word, but it doesn't have to be in the name.

GPT: Glad you like Havocloom! Here are more abstract names that capture a similar vibe, with or without "havoc":

- (1) Nightward
- (2) Eclipsion
- (3) Shadowfane
- (4) Stormveil
- (5) Emberthorn
- (6) Veilspire
- (7) Dreadhollow
- (8) Voidrift
- (9) Frostveil
- (10) Emberrift

These names are designed to have a strong yet mysterious feel. Let me know if any of these stand out to you!

Exchange 6

User: OK, I like some of these. Emberrift, though they generally seem to be quite arbitrary mixes of words. Let's try something now more like a phrase. I like Fields of the Nephilim as a band name. Again it's obscure, and also poetic.

GPT: Great! Here are some band name ideas that are more like phrases, with an obscure and poetic feel:

- (1) Veils of the Undying
- (2) Echoes of the Rift
- (3) Fires of the Abyss
- (4) Shadows of the Forgotten
- (5) Winds of the Wraith
- (6) Temples of the Void
- (7) Spires of Dusk
- (8) Embers of the Last Dawn
- (9) Voices of the Fading Sun
- (10) Monoliths of the Lost

These names aim to evoke a sense of mystery and depth. Let me know if any resonate with you!

Exchange 7

User: No, either too obvious or too random. Perhaps seek more obscure words or references to places or names. Get that mystery back in there (without being nonsense - Monoliths of the Lost!)

GPT: I see what you mean! Let's go for something more obscure and subtle, drawing inspiration from ancient places, mythological references, and poetic imagery:

- (1) Nifheim Strand
- (2) Aetherfall Ruin
- (3) Fenrir's Hollow
- (4) Svartraven Echoes
- (5) Skald's Veil
- (6) Vigrid Mists
- (7) Gleipnir's Rest
- (8) Mjölfnir's Lament
- (9) Elders of Yggdrasil
- (10) Runes of Skuld

These names incorporate obscure references while maintaining a sense of mystery and depth. Let me know if any of these resonate with you!

Exchange 8

User: I like Aetherfall Ruin. What does it mean?

GPT: Aetherfall Ruin combines the word "aether" (a mystical or ancient term for the upper air, sky, or the celestial realm) with "fall" to suggest something grand or divine descending or collapsing. The word "ruin" then adds a layer of decay or desolation to the imagery. Together, it paints a picture of a celestial collapse or the downfall of something otherworldly, giving it a mysterious, poetic edge.

It doesn't have a literal meaning, but it evokes a sense of both grandeur and destruction, which ties well with your band's themes. Does it feel like the right fit?

Exchange 9

User: Great. Havocloom and Aetherfall Ruin are both cool. I'll run them by the band. Which is your favourite and why?

GPT: Both names are strong, but I'd lean toward Aetherfall Ruin as my favorite. It feels grander and more evocative, with a poetic depth that matches your themes of Norse mythology, triumph, and downfall. The combination of celestial imagery and ruin gives it a haunting, epic feel, while still being abstract enough to spark curiosity.

Havocloom is sharp and powerful, but Aetherfall Ruin has that extra layer of mystique and complexity, making it more memorable and intriguing. Ultimately, though, both names are solid choices!
I'd love to hear what your band thinks!

Appendix D. Appendix 4: Worked Creative Example - *Billboard Advert*

Exchange 1

User: For this task, you are a creative in an advertising agency. We are working together to come up with ideas for a billboard advert for home insurance. We want to make people feel that we're there for them when they're having a bad day. You should suggest ideas, but also respond to my feedback and try to work out the best way we work together to come up with something original. Let's get started. I'm thinking about the idea of a beautiful landscape scene with a giant lightening bolt striking a house. I like the idea of something beautiful but arresting.

GPT: I love the idea of combining beauty with a dramatic element. Here are a few ways we could develop this concept:

- (1) **Visual Contrast:** Create a serene, picturesque landscape—perhaps a peaceful countryside or a charming suburban street. Then, in the foreground, a striking lightning bolt hits a house. The contrast highlights the unexpected nature of problems and how the insurance company is there to handle them, no matter how sudden.
- (2) **Message Integration:** Incorporate a tagline like, “When the storm hits, we're the calm you need.” The beauty of the landscape contrasts with the intensity of the lightning, reinforcing that even in the worst moments, you have support.
- (3) **Before and After:** Use a split-image approach. One side shows the beautiful landscape and the house in peaceful times, and the other side reveals the lightning strike. This could be paired with a message like, “From calm to chaos, we've got your back.”
- (4) **Dynamic Illustration:** Instead of a photo, use an illustrated approach with a slightly surreal or fantastical style. This can enhance the visual impact and make the ad more memorable.

Do any of these ideas resonate with you, or would you like to tweak or explore other directions?

Exchange 2

User: I like “when the storm hits, we're the calm you need”. Could be cuter – the calm after the storm. As a bit of a catchy play on words I like that.

GPT: Great choice! “The calm after the storm” has a nice ring to it. Here's how we might refine the visual to match the tagline:

- (1) **Visual Flow:** The billboard could show a dramatic lightning bolt striking a house

with the stormy sky above. Then, seamlessly transition into a calm, peaceful scene with a rainbow or clear sky, emphasizing the relief and support provided by the insurance.

- (2) **Color Scheme:** Use vibrant, contrasting colors for the stormy part of the image—dark blues, purples, and grays—transitioning to softer, calming colors like pastel blues and greens for the serene aftermath. This visual shift reinforces the message of calmness following chaos.
- (3) **Tagline Placement:** Place the tagline prominently in the calm portion of the image, possibly with a gentle, flowing font that complements the serene visuals.
- (4) **Additional Elements:** Consider adding subtle details like a small, comforting house or a friendly mascot (if the brand has one) in the calm section, symbolizing security and warmth.

What do you think of these ideas? Any other elements you'd like to include?

Exchange 3

User: Don't worry about giving generic recommendations. But specifically, how could a single image capture "the calm after the storm"? I'm keen on a really quality single photo that is dramatic. Perhaps it can just be a shot that has a sense of calm – through its beauty – even though it's depicting a terrible event?

GPT: To capture "the calm after the storm" in a single, high-quality photo with a dramatic yet serene feel, consider these ideas:

- (1) **Surreal Contrast:** Photograph a landscape where a dramatic, intense storm is just passing. The storm could be clearing in the background, revealing a breathtaking sky with soft light or a rainbow emerging. The house in the foreground could be untouched, symbolizing resilience and peace after the storm.
- (2) **Calm After the Downpour:** Capture a scene right after a heavy rain. A serene landscape with reflective puddles on a quiet street or garden, with the storm clouds still dissipating in the background. The calm reflections and gentle light can evoke a sense of tranquility despite the recent storm.
- (3) **Storm Light:** Shoot a landscape during the brief moment when the storm is receding, capturing the dramatic lighting with soft, filtered light breaking through the clouds. The image could feature a house or a tree that stands resilient amidst the remnants of the storm.
- (4) **Post-Storm Glow:** Focus on a single, striking element like a house or a lone tree illuminated by the first light after the storm, with dramatic clouds in the background. The image can have a soft glow or warm light, emphasizing calmness and recovery.

These concepts aim to blend dramatic elements with serene beauty, reflecting the idea of support and calm after a challenging event. Let me know if any of these ideas strike a chord or if you'd like to explore further!

Exchange 4

User: Calm after downpour is interesting. Could it be a picture depicting two people

surviving a flood and sharing a moment of warmth and optimism. Perhaps someone bringing someone else a cup of tea.

GPT: That's a heartfelt and human-centered approach. Here's how we might frame that idea:

- (1) **Setting:** Choose a setting where the flood's aftermath is visible—waterlogged streets or a partially submerged house. Ensure the background captures the recent flood's impact but not in a way that overshadows the warmth of the moment.
- (2) **Subjects:** Feature two people in the scene, one perhaps sitting in a safe, dry area (like a raised platform or a dry spot in the house) while the other brings them a cup of tea or coffee. Their expressions should convey relief, comfort, and a sense of togetherness.
- (3) **Lighting:** Use soft, natural light to create a warm, comforting atmosphere. The light could be coming through a window or a break in the clouds, highlighting the warmth of the shared moment.
- (4) **Details:** Include subtle details that emphasize the recovery and comfort—like a cozy blanket, a simple table, or other homey elements. These details can contrast with the flood's devastation and reinforce the theme of calm and support.
- (5) **Color Palette:** Utilize warm, inviting colors for the subjects and the cozy elements, contrasting with the cooler, muted tones of the flood-affected background. This contrast can enhance the sense of warmth and optimism.

Does this align with your vision? Any specific details or adjustments you'd like to include?