

# Mimetic Poet

Jon McCormack, Elliott Wilson, Nina Rajcic and Maria Teresa Llano

SensiLab

Monash University

Caulfield East, Victoria, AU

Jon.McCormack@monash.edu Elliott.Wilson@monash.edu Nina.Rajcic@monash.edu Teresa.Llano@monash.edu

## Abstract

This paper presents the design and initial assessment of a novel device that uses generative AI to facilitate creative ideation, inspiration, and reflective thought. Inspired by magnetic poetry, which was originally designed to help overcome writer’s block, the device allows participants to compose short poetic texts from a limited vocabulary by physically placing words on the device’s surface. Upon composing the text, the system employs a large language model (LLM) to generate a response, displayed on an e-ink screen. We explored various strategies for internally sequencing prompts to foster creative thinking, including analogy, allegorical interpretations, and ideation. We installed the device in our research laboratory for two weeks and held a focus group at the conclusion to evaluate the design. The design choice to limit interactions with the LLM to poetic text, coupled with the tactile experience of assembling the poem, fostered a deeper and more enjoyable engagement with the LLM compared to traditional chatbot or screen-based interactions. This approach gives users the opportunity to reflect on the AI-generated responses in a manner conducive to creative thought.

## Introduction

The advent of large generative AI models, such as transformers, heralds a new era in language-based interaction with machines. However, despite much potential, the majority of interactions with generative AI systems, such as Large Language Models (LLMs) is via simple “chat” interfaces and the form of that interaction is commonly in a question and answer format. This form of interaction reinforces the notion that the machine interlocutor is an “intelligent” but subservient entity, task-focused, willing to please and able to answer questions put to it, but unable to initiate its own dialogue because it lacks intention (McCormack 2024). Moreover, any dialogue-based interaction via screens does not consider implicit context, something that is particularly important in creative applications and in forming aesthetic judgements (Leder and Nadal 2014).

In this paper we propose an alternative way to engage with LLMs, building a novel form of interface designed to support ideation and creative thinking, either for individuals or groups. Called the *Mimetic Poet*, the device augments traditional magnetic poetry, using it as the mechanism for communicating with an LLM in an interactive dialogue over extended periods of time (days or weeks). Through the use of

constraints and physical interactions, the device promotes a slower form of interaction with LLMs, allowing space for contemplation and thought beyond conventional dialogue-based screen interfaces.

In the sections that follow we first introduce some background rationale for building this device and for the mode of interaction it uses. This rationale draws together theories of intention in writing and in design as a basis for interacting with a language-based AI model, which somewhat paradoxically, has no authorial intention. This turns out to be an advantage as it enables us to place both human and AI on a more equal footing in terms of intention.

Next we describe the system design and its technical operation in detail, including both the physical design and internal prompt chaining with the LLM (OpenAI’s GPT-4 API). To evaluate our system we installed it in our research lab for a period of two weeks and asked members of the lab to use it whenever they felt like it. At the conclusion of this period we held a focus group to draw some preliminary findings on the system’s efficacy as both an interface and as a way of promoting ideation and creative thought.

Lastly we reflect on the current limitations of the device and briefly discuss possibilities for improvement, along with some more general observations on people’s perception of current generative AI models.

## Background and Related Work

### The Intentional Fallacy

There’s a general agreement in literary theory that the meaning of a text does not reside exclusively with the author. Known as the “intentional fallacy” (Wimsatt and Beardsley 1946), it accounts for the idea that we may not know what the author of a text intended, and cannot ask them if, for example, they are no longer living. Additionally, the interpretation of meaning in a text is not exclusively determined by the author, where unintended meaning may arise due to historical or cultural context, or the author might deliberately lie or try to obscure the meaning. The philosopher Don Ihde sees parallels in technological design, describing what he calls the “designer fallacy” in thinking that “a designer can design into a technology, its purposes and uses” (Ihde 2008). Ihde is a leading proponent of *postphenomenology* (Ihde 1993), a methodological tool widely used in human-

computer interaction (HCI) design to analyse the relationships between humans and technology, in particular how technology mediates our view of the world and our actions in it. Postphenomenology sits in contrast to more traditional approaches to interface design, such as “user-centered design” (Norman and Draper 1986), which generally limits design considerations to the direct interactions between human and machine.

This dualism of intentional fallacy in text and designer fallacy in interface design is something we weave together in this work, which explores relationships people might have with the emerging new technologies of generative artificial intelligence. We explore design possibilities beyond conventional forms of interaction with LLMs, such as “chatbot” interfaces popularised by systems such as OpenAI’s ChatGPT, instead designing speculative possibilities for what human-AI relationships *could be* (Rajcic and McCormack 2023). The value of this form of design extends beyond functionality or aesthetics. By challenging assumptions and conceptions about the role that objects and technology play in our lives, this form of speculative design—as process and way of thinking—can serve as a means of stimulating different ways of “speculat[ing] about possible futures; and as a catalyst for change.” (Dunne and Raby 2013, p.33).

If we accept that the technological designer’s intention does not fully determine how people may use their design, nor its broader socio-cultural effects, then we are free to consider the possibilities a design might facilitate rather than mandate (e.g. rather than asking “*what is its function or purpose?*” we consider “*what might it make possible?*”). In speculative design we design possible futures to better understand and evaluate the implications of new technologies before they become embedded in society; to enact change rather than conforming to the status quo (Dunne and Raby 2013). This becomes particularly important with the rapid, and largely unregulated, rise of generative AI and its ethical, social and cultural impact on human creativity (Rajcic, Llano, and McCormack 2024).

The intentional fallacy also finds new meaning in generative AI systems, in particular LLMs which can generate coherent text with apparent intention and meaning. However, an LLM does not have a discernible authorial intent, rather each output is a statistical accumulation of individual authorial intent, from those whose work was included in the training data<sup>1</sup>. Moreover, the models are prone to “hallucinations” where they produce factually incorrect or semantically misleading output. If the intention doesn’t come from the LLM, and what the model produces may be a hallucination, then there may be situations where the reader is freer to make their own interpretation. This includes scenarios where reliance on factual or explanatory information is not essential, such as divergent creative thinking, the use of analogy or metaphor in activities such as brainstorming or creative ideation.

---

<sup>1</sup>The exception being in cases where LLMs can be coaxed into verbatim repeating of their training data (Carlini et al. 2021; Nasr et al. 2023).

## Creativity Support Tools

The study of creativity has been enriched by the development of theories such as divergent thinking (McCrae 1987), which emphasizes the generation of multiple solutions to a given problem. This concept, introduced by Guilford in the 1950s, underscores the importance of thinking in varied and unique directions as a hallmark of creativity. Additionally, the theory of intrinsic motivation (Hennessey and Amabile 1998) highlights the role of self-motivation in fostering creativity, suggesting that creative endeavors are most fruitful when driven by genuine interest and satisfaction in the work itself.

On the other hand, Creativity Support Tools (CSTs) are generally designed to aid individuals in furthering their creative faculties by providing resources, inspiration, and technological scaffolding to navigate the creative process. These tools range from simple brainstorming tools that facilitate idea generation through divergent thinking, such as Brian Eno’s *Oblique Strategies* (Harford 2016, Ch.1), and towards more complex systems that integrate technology into the process. However, ways of understanding how generative AI systems might support creativity via creative strategies is relatively underdeveloped (Koch et al. 2019; 2020; Verheijden and Funk 2023).

In the evolving landscape of computational creativity, generative text has been explored through various developments of algorithmic sophistication. Early research explored the modeling of emotion within algorithmically generated poetry (Misztal and Indurkha 2014), the constraining of poetic structure of generated poems (Colton, Goodwin, and Veale 2012), and poetry as responding to input images (Loller-Andersen and Gambäck 2018).

As technology advanced, research shifted towards interactive and collaborative systems (Oliveira et al. 2019). Research also looked into the development of generative poetry as a Creativity Support Tool (CST) for creative writers (Booten and Gero 2021), as well as for creativity in general (Kantosalo et al. 2014). This evolution reflects a growing interest in not only automating the generation of poetic content but also in developing systems that can engage with human users in a more meaningful, collaborative manner.

Researchers have furthermore explored the embedding of generative poetry into physical interfaces (Rajcic and McCormack 2020; 2023; Stiles 2022) introducing an innovative dimension to the interaction between humans and computational creative systems. Such approaches not only challenge the traditional boundaries of poetic expression but also invite users to engage with poetry in more tangible and immersive ways, increasing accessibility of generative poetry, by allowing engagement without the need for conventional screen interfaces.

## Found Poetry

Magnetic Poetry, a literary phenomenon characterized by the arrangement of individual words on magnetic surfaces to create poetry, emerged as a modern iteration of the found poetry technique. The genesis of Magnetic Poetry can be traced back to 1993, when songwriter Dave Kapell, seeking

to overcome writer’s block, developed the concept by scattering words on his refrigerator, allowing for spontaneous and serendipitous writing. This innovation offered a tangible medium for creative writing while making it fun and accessible to a wider audience without the prerequisite of formal literary training.

The technique of “found” poetry has a rich historical lineage extending into the early 20th century. Found poetry re-purposes existing texts, extracting and rearranging words and phrases to form new meanings and poetic expressions (Barda and Barda 2019). This method challenges traditional notions of authorship and creativity, suggesting that art can emerge from the re-contextualization of pre-existing material. The Dadaists of the 1920s, notably Tristan Tzara with his “cut-up” technique, and later the Beat Generation poets in the 1950s, such as Brion Gysin and William S. Burroughs (Adema 2017), significantly contributed to the development and popularization of found poetry. They explored the potential of random and aleatory processes in literary creation, laying the groundwork for contemporary practices such as Magnetic Poetry.

The effect of constraints on creative thinking is illustrated in the practices of found and Magnetic Poetry. By imposing limitations on the selection and arrangement of words, these poetic forms paradoxically liberate the creative process. This phenomenon underscores the role of constraints not as creative barriers, but as catalysts that stimulate creative thinking and problem-solving. Constraints serve a crucial role in the creative process by focusing attention and reducing the overwhelming possibilities that can lead to creative block (Medeiros, Partlow, and Mumford 2014). Working within a fixed lexicon, participants are naturally led to forge unexpected connections between words and ideas.

## System Design

In this section we discuss the system design and overall concept of the *Mimetic Poet*, beginning with the design rationale for using magnetic poetry as an interface.

### A Poetic Interface

To speculatively explore possibilities for human-AI interaction we opted to use poetics as the means of communication between human and machine. By using poetry or poetic language, we mask or obscure the authorial intention of the human and AI authors, placing both on a more equal footing. While poetry and poetic text are the primary communication mechanism for the system, its overall goal is to support creative ideation and divergent thinking rather than being an AI poet. As such, responses are more often *poetic* rather than *poetry*, for example taking the form of aphorisms, metaphor or allegorical narrative rather than literally being poems.

The use of magnetic poetry lowers the “barrier of entry” for people who may not be used to creating poetic text and may not consider themselves poets or even authors in any traditional sense. As both an “interface” and form of expression it has a number of advantages, which we classify into three categories:

**constraints:** magnetic poetry uses a limited vocabulary of possible words from which to compose the poem. This helps to reduce decision paralysis, even picking up words at random can be used to begin. The constrained size of available space helps keep poems brief and to the point.

**usability:** the playful, engaging and physical manipulation of words is simple, moving or editing of words inspires thought on meaning and context.

**physicality:** physical words allow for serendipitous creativity, e.g. seeing interesting combinations of words in close physical proximity around the device (Fig. 1) can inspire a poem’s creation. Being in a shared physical space (rather than on a personal device such as a phone or laptop screen) encourages co-creation in groups of people, who can collaborate on a poem’s construction.

### Slow technology

Standard “chat” style interfaces with LLMs take the form of a continuous dialogue where the person asks a question then receives an instantaneous response. While such dialogues are suited to interactions such as question and answer, or specific goal or task directed activities (“summarise the following report”, “give me a recipe with these ingredients”, etc.), they reinforce concepts of immediacy and divisibility in task-focused language.

In contrast, the use of a physical interface and its affordances emphasises a different approach to using the technology and how to engage with it. Having to manually assemble a poem from a fixed set of words as physical objects requires a distinct mode of thought over standard conversation or text messaging. Assembling the poem and waiting for a response allows for the human participant to consider how an AI might interpret their intention from the poem’s text. The use of an e-ink display means the machine’s responses will always be visible on screen (even when the power is turned off), giving a stronger sense of permanence and significance over traditional chat-based interfaces.

This aligns with the philosophy behind the slow technology approach (Odom et al. 2012; Hallnäs and Redström 2001), which advocates for more reflective human-machine interactions. This counters the fast-pacing, immediate nature of contemporary technology use, as exemplified by the question-response mode of the chat interface described above. By adopting this “slow technology” approach, we implicitly require the human participant to slow down in their interactions with the AI, leaving more space for contemplation, reflection, and evaluation of the exchanges, elevating their significance and prominence both physically and temporally. Our hypothesis is that, together, these features will encourage and support more nuanced understanding of an AI’s role and capabilities, while at the same time assisting with human creative thinking.

In addition to expressing oneself to the AI through magnetic poetry, constraining the language model to reply in a poetic way helps to circumvent problems of both intention and hallucination – there is no “wrong” way to write a poem and issues of truth or factual accuracy are less relevant than in more didactic or information-based tasks.



Figure 1: The Mimetic Poet machine, showing key elements of the device

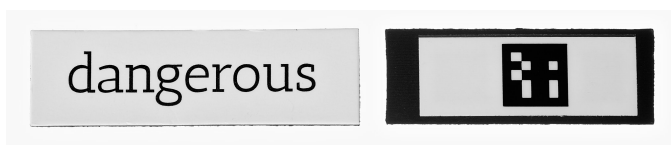


Figure 2: A magnetic poetry word (left) and the fiducial marker attached on the reverse side (right)

### The Device

The *Mimetic Poet* is designed as a stand-alone device that can sit easily in a studio, workplace or home (Fig. 1). It consists of a flat surface (the *slate*) upon which you can assemble your poem to be read by the system. A small drawer underneath the slate is used to store individual words used to create the poem. We found that placing the words around the device facilitated easy composition of poems. The words themselves were sourced from a standard magnetic poetry kit<sup>2</sup>. On the rear side of each word is a small fiducial marker (Fig. 2) that a machine vision system can recognise when the word is placed on the slate (Romero-Ramirez, Muñoz-Salinas, and Medina-Carnicer 2018). We use this marker recognition as the mechanism for communication between human and machine.

Inside the device, a camera and mirror system is used to detect the words placed on the slate. A Raspberry Pi 5 computer performs image processing and fiducial marker detection to identify the words placed on the slate in the sensing area (Figure 4). We use a camera that is sensitive to infrared (IR) light and apply a visible light filter to the lens, minimising any adverse effects of external visible light on the marker

<sup>2</sup>The standard magnetic poetry kit has several hundred words, we used a subset of 175 words in the initial version of the system.

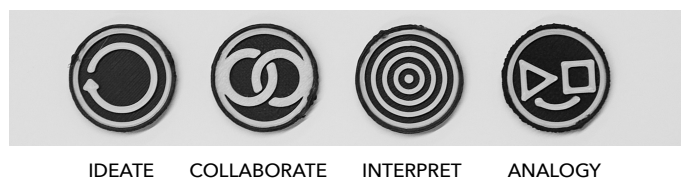


Figure 3: Four special markers that determine the poet's mode when placed on the slate.

recognition and allowing the device to be used irrespective of external lighting conditions. An array of IR LEDs inside the device provides direct illumination of any objects placed on the slate.

As words can be placed in an arbitrary location we developed an algorithm that captures the position and orientation of each word then uses a series of heuristics to determine how best to translate new lines and ensure correct word ordering. The aim was to match, as closely as possible, how a person would “read” the sequence of words placed on the slate. This was essential as we found participants often used interesting physical arrangements of the words as part of a poem's visual aesthetic.

To compute the correct ordering of placed words we developed a novel sorting algorithm that uses ray-tracing (Whitted 1980) to determine the best sequencing of placed markers. The marker detection system provides the positions of the centre ( $m_c$ ) and the four corners of each detected marker. The marker with the highest vertical ( $y$ ) position is first selected and removed from the list of unsorted detected markers. The top-left and bottom-left corners of the selected marker are subtracted from each other to form a vector representing the left edge of the marker. Next, we compute the tangent vector to the marker edge vector,  $\vec{v}_t$ , which is used

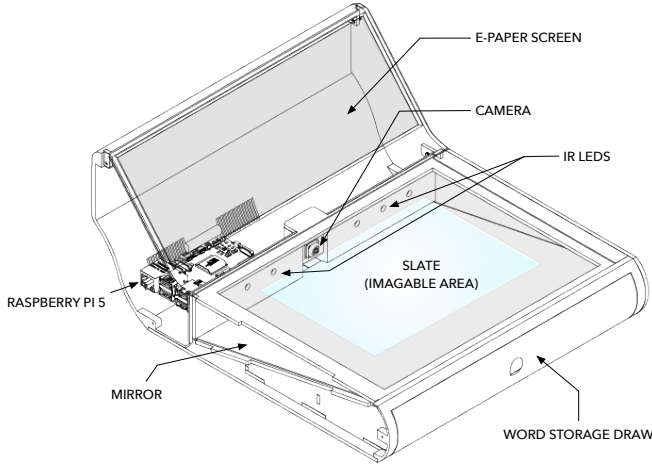


Figure 4: Schematic diagram of the Mimetic Poet Slate

to create a line,  $l$  with start and end points  $k$  units to the left and right of the maker respectively.

$$\begin{aligned} start &= k\vec{v}_t + m_c \\ end &= -k\vec{v}_t + m_c \\ l &= end - start \end{aligned} \quad (1)$$

As the units used by the system are measured in pixels, we set  $k = 1000$ . This line is then intersection tested (Eberly 2001) across each remaining unsorted marker, using a circle to represent the bounds of the marker.

$$p_c = start + \left( (m_c - start) \cdot \frac{l^2}{|l|^2} \right) \quad (2)$$

Markers that collide with the line and the initial marker that was used to create it are removed from the unsorted list and moved to a new list that represents a single line of tiles:

```

if  $|p_c - m_c| < tileHeight$  then
  move  $m$  to line list
end if

```

This list is then sorted by the distance of each marker from the start point of the line. This ensures that even if the highest marker wasn't the left-most tile in the line that the order is correct. It also means that the sorting algorithm works with significantly skewed/diagonal lines of tiles and even upside down tile will sort in a meaningful way (i.e. they will be read right to left because the start and end points of the line will have flipped). After all the markers intersecting the line have been processed the algorithm is repeated until the unsorted marker list is empty.

If no word is moved for a few seconds, the text of the poem, including new lines, punctuation, etc. is converted to a string and the string sent to the AI subsystem (described in detail below). In addition to placing words on the slate, we designed four special markers (Fig. 3) that allow participants to select a different mode or personality for the AI.

We initially experimented with the idea that the markers should indicate the kind of response the participant is looking for from their poem:

1. **interpret**: the system attempts to interpret the input poem, based on emotional tone or content, and provide a "reading" of the participants mental or emotional state.
2. **collaborate**: the system tries to collaborate with the participant on poetry generation, generating a variant on the text supplied using the same words as are available to the participant.
3. **ideate**: the system uses the input poem to help with ideation, responding with an idea or strategy that builds upon similarities between the poetic concept presented and ideas that the author may be interested in.
4. **analogy**: the system constructs an analogous text based on concepts from a different discipline.

In early testing with this scheme, we observed that while participants liked the ability to control and direct the system, the responses were often wordy or too cryptic for people to see obvious relationships between poem and response. Moreover, differentiating the modes, both visually in terms of the marker icon and conceptually in terms of the response, proved difficult. To address these issues, we developed a prompt chaining scheme (Wu et al. 2022) to better control the LLM's responses to the input poems.

### Prompting the LLM

The design of prompts used internally plays a pivotal role in steering the direction and quality of the outputs generated by the system. We implemented a set of prompt chains for the preliminary testing of the system. To this end, we utilised LangChain (Chase 2022), a python package which facilitates prompt design and chaining with interchangeable LLM services. For this study, we used OpenAI's GPT4, accessed via the Python API.

The prompts were written with respect to the four modes (Figure 3) and are shown in Table 1. For each mode, we first constructed a prompt that exemplified the desired outcome of the mode. The second stage of the prompt chain was to summarise the first response, placing a restriction on the LLMs output. Incorporating a second stage in the prompt chain, where the system's output is condensed or reinterpreted, introduces an essential feedback mechanism. This step is not only a constraint, but filters the initial response, promoting clarity, brevity, and a different presentation of the original output. This was thought to generate unusual, poetic, and particularly cryptic responses. This configuration serves as an example from which participants of the study (detailed below) based their initial impressions, and suggested modifications or improvements to the prompts.

### Study

To better understand how the efficacy of the *Mimetic Poet* as both a human-AI interface and system for supporting creative thinking, we undertook a preliminary study within our research laboratory. We placed the device in a communal



Mode	Prompt 1	Prompt 2
Interpret	I just wrote the following text: {poem}. Speculate on what I'm feeling when writing this. Please keep the interpretation short (2-3 sentences).	Summarise this: {response} in only 5-15 words.
Collaborate	Select words from the following text: {poem} to form a question that the text seems to be asking or addressing. Then, use other words from the text to answer it (2-3 sentences).	Summarise this: {response} in only 5-15 words.
Ideate	The user just input the following text: {poem} Try and develop a creative idea or strategy that builds upon similarities between these words/concepts presented. Please keep your response short (2-3 sentences).	Reword your answer here: {response} in only 5-15 words.
Analogy	Reframe this the following text with reference to a different discipline: {poem}	Repeat the following: {response} except obscure it further.

Table 1: The prompt chaining for each mode, with ‘poem’ variable being the input poem by the user, and ‘response’ being the LLM output resulting from the first prompt.

area of the workplace where people often gathered. Participants ( $N = 14$ ) were instructed on the basic purpose of the device (a machine to support creativity that uses magnetic poetry to communicate with an AI) and encouraged to use it as they saw fit. The device was installed over a two week period and during that time participants generated 413 individual poems (an average of 29 poems/participant). No remuneration was provided for participation and the project was approved by our university Ethics committee.

The most popular mode was “collaborate” (59%), followed by “ideate” (19%), “interpret” (14%) and “analogy” (8%). Participants used only 111 of the 175 available words, with the AI responses generating 6,833 unique words, a forty-fold increase. Interestingly, the most popular human-selected words (*human, dead, deception, memory, machine, bad, filth, heaven, delicious, eat*) were almost identical to those selected by the AI (*human, memory, community, machine, flower, heaven, filth, wonder, empty, nature*), indicating that the responses often used words from the input text.

At the conclusion of the study period, we held an in-person focus group to hear participants’ thoughts and experience of using the device. The focus group consisted of semi-structured interviews and open discussion in the same area that the *Mimetic Poet* was installed, allowing participants to directly demonstrate their use of the device and to refer to specific aspects of its design by indicating or showing. The focus group was attended by participants ( $N = 6$ ). All comments were audio recorded and transcripts produced, along with researcher’s notes and still photography.

## Methodology

We collated the transcripts, notes and internal machine logs and undertook a basic analysis using two methods. Firstly, individual researchers analysed the text and notes and drew out specific themes of importance from participants responses. The research team then came together to compare findings and, following discussion, individual findings were merged, resulting in four identifiable themes: (i) the inter-



Figure 5: A participant exploring the *Mimetic Poet* during the focus group session.

face itself and physicality of interaction; (ii) usability; (iii) Participant’s overall perceptions of the AI; (iv) The AI’s responses to participant’s poems. We discuss findings under each of the themes below.

## Themes

“I have really enjoyed being invited to engage in something playful. It is a pleasant break from work that puts me in a good mood and perhaps relieves work related ‘stuckness’ or frustration. I think being playful is important for creative people.”

— Focus Group Participant

**Interface/Physicality** Participants of our focus group found the interface “playful” and “like a game”, with the

magnet poetry component of the system intuitive and fun to engage with. One participant remarked that the device was “hard not to interact with”. A majority of participants commented on the physical design being “retro” and reminiscent of old or obsolete technologies. This led to a good deal of curiosity and interest. Other participants also enjoyed the fact that it “free[s] your mind from the urgency of working with a screen”. The choice to embed the technology into a physical device was enjoyed by all participants, and reminded them of a “board game” rather than carrying the usual connotations of interacting with AI.

With respect to the limited vocabulary available, participants found this both enjoyable and frustrating; “that’s the beauty of fridge poetry”. However, with reference to the machine’s responses, one participant noted it annoying that the system has, in contrast, an unlimited vocabulary: “all I have is simple nouns and verbs, and it comes with high philosophical concepts”. Some participants enjoyed the process of scanning through and picking up the individual word magnets as an “inspirational” activity. Yet, other participants commented on their preference to be able to write freely to the system. For example, open writing interface would be “giving more space for the participant to explore . . . rather than putting some random words that someone else picked”. While the constraint made it easier for some to engage, there is a sense that some participants struggled to communicate with the AI; that their individuality was lost in the process.

**Usability** All participants found the device easy and intuitive to use and were impressed with the reliability of the word recognition. However some commented on the delay between finishing the poem and getting a response being too long. This expectation for immediate responses contrasts the slow-technology approach adopted in the *Mimetic Poet*, and while participants found this to be an issue in their initial interactions, they also commented that this would not be a problem for sustained use “I would definitely have something like this if I’m working continuously on a writing piece, if I need to ignite some ideas”, where it could be seen more “as a companion”, in contrast of it being “a one encounter” type of interaction. Upon further discussion participants suggested wanting more immediate feedback from the device in terms of what it is doing “it takes time to figure out that you need to wait and that there could be some glitches in between”.

Overall, participants identified a personal or individual use as their preferred mode of interaction “I feel like it could be a private moment, and then you can share it with some other people”, “just for me, for my reflective practice”, “I would love to have one of these at home to just play around and see what happens”. However, they also highlighted the benefits of more social or group use, “it could even be fun to use it in a session like this when we’re trying to ideate things together”, and even one participant commented how they got inspired on their own work seeing the response another participant got from the device “but it wasn’t something that I did. It was something that . . . was doing before”.

Finally, participants also mentioned the ability to customise their device as something they would find useful

“as time goes by, it does have more information about you . . . it would introduce dynamic”. Additionally, although they found it fun to use, some of the participants mentioned getting tired of the interactions over a longer period. Customisation as the device learns from human interactions might help alleviate this issue.

**AI Perception** Participants pointed out that the responses from the AI have a very similar positive tone, which was a “limiting factor” for creative expression, “if you are limited to just that spectrum of the emotion . . . then you are never going to really explore things properly.” As tech companies try to make LLMs safe for public use, they have adopted a dominant positive tone in the models responses. In a sustained interaction, this may lead to a gap between someone’s intentions and the device’s ability to meet those intentions.

We acknowledge that controls for this must be in place, but for creative expression a balance may be required. On this, participants mentioned they would like the ability to change the “personality” of the device “I think when it displays personality it is engaging and enjoyable”, citing examples when the responses seemed to have “some kind of hidden sentiment, like sarcasm or double meaning” as more engaging. Although participants did not report this as a factor to enhance creative thinking, they repeatedly refer to this type of interaction as playful and enjoyable, something that would set them in a better mood, which previous studies have shown to enhance creative thinking (Vosburg 1998; Baas, De Dreu, and Nijstad 2008).

**AI Response** In addition to wanting responses with different tones or as the participants put it, “personalities”, they also mentioned that the more poetic-like responses were the most interesting as they differed from the usual descriptive LLM style. On this, most of the participants agreed that the *Analogy* mode took them closer to the kind of preferred response, with one of them saying “is the most proactive answer . . . it feels the closest to poetry”.

An interesting insight was also on how much detail to give about the nature of a response, with some participants stating that they would like to understand “the intent of the prompt” as this would help them “understand what’s happening behind it”. These participants believed this could lead to the ability to “tune [the device] to your own particular needs”. However, other participants stated that they would prefer not knowing since they could assign their own interpretation to the response “me putting myself in, could dictate what I’m going to get out [which] is the whole purpose”.

## Discussion

As the system designers, we found the focus group feedback very helpful in considering how the current design is perceived by its users, and ways we can improve it. However, as a small study on an initial prototype, our work has a number of limitations. Only a subset (6) of participants took part in the focus group, so we weren’t able to canvas all participant’s opinions of the device. However, the large number of interactions logged over the two week evaluation period suggests that the *Mimetic Poet* received a lot of at-

Mode	Input Poem	Output Text
Interpret	brain problem see over here each bad judge secret life insidious their obscene picture is already across from a good few	You seem to be grappling with feelings of frustration and anxiety, possibly due to feeling misunderstood or judged.
Collaborate	hate delicious body beautiful anxious heart	Delicious hate, body beautiful, Anxious heart, artfully dutiful.
Ideate	do promise a radiant world forest see animal their same cry beyond science slow broken heaven	Post-apocalyptic mobile game-animals restore nature, create new world
Analogy	shine promise water thought until flower already soft	Dewy water softly caresses the budding flower, refreshing the mind's hardened soil

Table 2: Example text generated by the system for each mode, based from participants input poems

tention and use – an average of 41 new poems were created each day. Secondly, participants were members of our laboratory and were all creative practitioners or researchers highly familiar with technology and the creative use of Artificial Intelligence. This is not necessarily a disadvantage, as all participants had previously attempted to use tools like ChatGPT for creative ideation or to support creative thinking. The majority found such tools ill-suited to the task and saw the *Mimetic Poet* as a playful alternative.

Although at the heart of our design is the idea of encouraging reflection following the principles of slow technology, the study showed us that this will represent a challenge as some participants longed for the immediacy of current interactions. Research in the design of these types of technologies, which are intended to “surround us and therefore is part of our activities for long periods of time” (Hallnäs and Redström 2001, pp 161), have emphasised the need of persistent use in order to develop such intended relationship (Mazé and Redström 2005). For this, a study with a more sustained period of use is needed.

Lastly, while a number of participants reported finding using the tool stimulating and interesting, we did not find that the device, in its current form, was directly responsible for solving practical creative problems.

Despite these limitations, we feel that the *Mimetic Poet* has significant potential, which we hope to realise in the next iteration, currently under development. The feedback from our study has suggested several design changes that would improve engagement and quality of responses. These include:

**Interface:** additional feedback while composing a poem was seen as favourable amongst several participants; adding markers that represent different AI “personalities” that allow participants to customise the responses as an alternative to the current mode markers (Fig. 3).

**AI responses:** we currently use a simple prompt chaining technique (Table 1) to prompt GPT-4. Further development of more advanced prompt chains is needed to increase the quality and suitability of the AI responses. Experimenting with other LLMs, in particular ones that can be easily fine-tuned (such as Mistral7b) would help overcome the overly positive and didactic responses participants often received from the current system.

**AI perception:** currently our zero-shot prompting does not allow the LLM to make use of past interactions, leading to a more transactional interaction from participants. Adding the ability to personalise the model and incorporate past interactions into the prompt chain would allow the LLM to also reference a participant’s history of interactions, enriching the experience over the long term.

## Conclusions

In this paper we have presented the *Mimetic Poet*, a novel device that uses magnetic poetry as the means of communication with a LLM for the purposes of encouraging creative thinking and ideation. We designed and built the device, then evaluated it with a group of participants over a two week period, canvassing views of user experience in a focus group session. Our findings showed that some encounters with the device helped participants in creative thinking and ideation, and that the interface was a desirable alternative to traditional chat-based interfaces, with our interface preferred as an inspirational device.

We view these alternative human-AI interaction methods as a means to catalyse wider conversations about AI’s role in human creativity, presenting new ways for us to engage with artificial systems. By examining logs of human input and AI responses, it’s clear that humans are currently the more creative of the two participants. . . for now.



## References

- Adema, J. 2017. Cut-up. In *Keywords in Remix Studies*. Routledge. 104–114.
- Baas, M.; De Dreu, C. K.; and Nijstad, B. A. 2008. A meta-analysis of 25 years of mood-creativity research: Hedonic tone, activation, or regulatory focus? *Psychological bulletin* 134(6):779.
- Barda, J., and Barda, J. 2019. Techniques of assemblage. *Experimentation and the Lyric in Contemporary French Poetry* 95–167.
- Booten, K., and Gero, K. I. 2021. Poetry machines: Eliciting designs for interactive writing tools from poets. In *Creativity and Cognition*, 1–5.
- Carlini, N.; Tramer, F.; Wallace, E.; Jagielski, M.; Herbert-Voss, A.; Lee, K.; Roberts, A.; Brown, T.; Song, D.; Erlingsson, U.; and others. 2021. Extracting training data from large language models. In *30th USENIX security symposium (USENIX security 21)*, 2633–2650.
- Chase, H. 2022. *LangChain*. <https://github.com/langchain-ai/langchain>.
- Colton, S.; Goodwin, J.; and Veale, T. 2012. Full-face poetry generation. In *ICCC*, 95–102.
- Dunne, A., and Raby, F. 2013. *Speculative everything: Design, fiction, and social dreaming*. Cambridge, MA and London England: MIT Press.
- Eberly, D. H. 2001. *3D game engine design: a practical approach to real-time computer graphics*. San Francisco, Calif.; London: Morgan Kaufmann.
- Hallnäs, L., and Redström, J. 2001. Slow technology—designing for reflection. *Personal and ubiquitous computing* 5:201–212.
- Harford, T. 2016. *Messy: How to be creative and resilient in a tidy-minded world*. London, UK: Little Brown and Company.
- Hennessey, B. A., and Amabile, T. M. 1998. Reality, intrinsic motivation, and creativity. *American Psychologist* 53(6):674–675.
- Ihde, D. 1993. *Philosophy of technology: An introduction*. New York, NY, USA: Paragon House Publishers.
- Ihde, D. 2008. The designer fallacy and technological imagination. In Vermaas, P. E.; Kroes, P.; Light, A.; and Moore, S. A., eds., *Philosophy and design*. Springer. 51–59.
- Kantosalo, A.; Toivanen, J. M.; Xiao, P.; and Toivonen, H. 2014. From isolation to involvement: Adapting machine creativity software to support human-computer co-creation. In *ICCC*, 1–7.
- Koch, J.; Lucero, A.; Hegemann, L.; and Oulasvirta, A. 2019. May AI?: Design ideation with cooperative contextual bandits. In Dey, A., and Zhao, S., eds., *Proceeding of ACM SIGCHI 2019*, Paper No. 633. New York, NY: ACM / ACM SIGCHI.
- Koch, J.; Taffin, N.; Beaudouin-Lafon, M.; Laine, M.; Lucero, A.; and Mackay, W. E. 2020. ImageSense: An intelligent collaborative ideation tool to support diverse human-computer partnerships. *Proc. ACM Hum.-Comput. Interact.* 4(CSCW1).
- Leder, H., and Nadal, M. 2014. Ten years of a model of aesthetic appreciation and aesthetic judgments: The aesthetic episode – Developments and challenges in empirical aesthetics. *British Journal of Psychology* 105:443–464.
- Loller-Andersen, M., and Gambäck, B. 2018. Deep learning-based poetry generation given visual input. In *ICCC*, 240–247.
- Mazé, R., and Redström, J. 2005. Form and the computational object. *Digital creativity* 16(1):7–18.
- McCormack, J. 2024. Autonomy, intention, performativity: Navigating the AI divide. In Trillo, R. A., and Poliks, M., eds., *Choreomata: Performance and performativity after AI*. Boca Raton, FL, USA: CRC Press. 240–257.
- McCrae, R. R. 1987. Creativity, divergent thinking, and openness to experience. *Journal of personality and social psychology* 52(6):1258.
- Medeiros, K. E.; Partlow, P. J.; and Mumford, M. D. 2014. Not too much, not too little: The influence of constraints on creative problem solving. *Psychology of Aesthetics, Creativity, and the Arts* 8(2):198.
- Misztal, J., and Indurkha, B. 2014. Poetry generation system with an emotional personality. In *ICCC*, 72–81.
- Nasr, M.; Carlini, N.; Hayase, J.; Jagielski, M.; Cooper, A. F.; Ippolito, D.; Choquette-Choo, C. A.; Wallace, E.; Tramèr, F.; and Lee, K. 2023. Scalable extraction of training data from (production) language models. *arXiv pre-print*.
- Norman, D. A., and Draper, S. W. 1986. *User centered system design: new perspectives on human-computer interaction*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Odom, W.; Banks, R.; Durrant, A.; Kirk, D.; and Pierce, J. 2012. Slow technology: critical reflection and future directions. In *Proceedings of the Designing Interactive Systems Conference*, 816–817.
- Oliveira, H. G.; Mendes, T.; Boavida, A.; Nakamura, A.; and Ackerman, M. 2019. Co-poetryme: interactive poetry generation. *Cognitive Systems Research* 54:199–216.
- Rajcic, N., and McCormack, J. 2020. Mirror ritual: An affective interface for emotional self-reflection. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, 1–13.
- Rajcic, N., and McCormack, J. 2023. Message ritual: A posthuman account of living with lamp. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–16.
- Rajcic, N.; Llano, M. T.; and McCormack, J. 2024. Towards a diffractive analysis of prompt-based generative AI. *arXiv preprint arXiv:2403.01783*.
- Romero-Ramirez, F. J.; Muñoz-Salinas, R.; and Medina-Carnicer, R. 2018. Speeded up detection of squared fiducial markers. *Image and Vision Computing* 76:38–47.
- Stiles, S. 2022. *TECHNELEGY*. Eyewear Publishing.
- Verheijden, M. P., and Funk, M. 2023. Collaborative diffusion: Boosting designerly co-creation with generative AI.

In *Extended abstracts of the 2023 CHI conference on human factors in computing systems*, CHI EA '23. New York, NY, USA: Association for Computing Machinery.

Vosburg, S. K. 1998. The effects of positive and negative mood on divergent-thinking performance. *Creativity research journal* 11(2):165–172.

Whitted, T. 1980. An improved illumination model for shaded display. *Communications of the ACM* 23(6):343–349.

Wimsatt, W. K., and Beardsley, M. C. 1946. The intentional fallacy. *The Sewanee Review* 54(3):468–488.

Wu, T.; Jiang, E.; Donsbach, A.; Gray, J.; Molina, A.; Terry, M.; and Cai, C. J. 2022. PromptChainer: Chaining large language model prompts through visual programming. In *Extended abstracts of the 2022 CHI conference on human factors in computing systems*, CHI EA '22. New York, NY, USA: Association for Computing Machinery.