Pareiduo: A Human–LLM Dialogue in Sand

Jiao Zhao
Department of Urban Studies and Planning
Massachusetts Institute of Technology
Cambridge, MA 02139
jiaoz17@mit.edu

Qian Xiang
Graduate School of Design
Harvard University
Cambridge, MA 02139
xianq734@gmail.com

Abstract

We introduce Pareiduo, an interactive device that facilitates drawing games between humans and AI. Drawings are a fundamental human artifact and a new way for AI to understand abstract representations. Pareiduo encourages players to sketch in sand on a glowing surface with feedback and instructions driven by LLMs. Two players take on complementary roles: Shadow (handling sand formations) and Light (dealing with voids), working together to help the AI identify their creations. The system combines visual intelligence, image processing, and embedded electronics to offer a screen-free, multisensory interface that promotes face-to-face interaction. Pareiduo reimagines human-AI perceptual differences not as mistakes but as reflections of the variety within abstract representation. The project explores the creative space where human imagination and machine perception meet in an open-ended conversation, sparking reflections on representation, standardization, and teamwork in the digital age.

1 Introduction

Interpreting patterns and creating abstract representations are fundamental aspects of human cognition[1, 2]. Across cultures and histories, abstraction has served not only as a way to make sense of the world, but also as a means of building connection and solidarity, allowing people to share perception, emotion, and meaning through symbolic form[3]. This process finds an analogue in computation. Computation itself can be understood as a system of representation, an attempt to formalize and manipulate patterns of knowledge. Abstract representations like drawings extend beyond efficiency to speak to personal desires and individuality within collective contexts. Sand drawing has deep roots across diverse cultures, serving as a means of collective interpretation and social engagement. In contemporary settings, sand play has evolved into various forms, such as sand painting as performative art and sand tray play for children's education and recreation. As sand is easily manipulated, coninupulsy formable sand drawing has the potential an inviting medium that lowers the barrier to participation, inspiring the naive, playful side of drawing as a game rather than as an accomplishment obliged to create something extraordinary.

Reading abstract representations differs between large language models and humans, creating interesting discrepancies that reflect partial knowledge about how we recognize things. We present Pareiduo, an interactive game device that explores these discrepancies and similarities of perception through sand drawing(Figure 1). The game begins with role assignment of two players as either the Shadow or Light player. The Shadow player works with sand formations to create recognizable patterns, while the Light player focuses on creating meaning through the empty spaces between formations. Each player receives a secret object to draw. Players take turns within timed rounds, working on the same surface through opposite approaches. Their goal is to collectively create a pattern that allows the AI to recognize both secret objects. Throughout the game, the AI provides feedback about what it perceives in the evolving sand patterns. As rounds progress, players can decide when to end

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

Figure 1: Photo of the designed device

the game to reveal final results. The AI then discloses its recognition of each player's object and provides explanations of its recognition process and judgment reasoning. Pareiduo creates a slow, conversational experience that explores not only the interpretation gaps between humans and AI, but also how humans can collaborate to maximize both positive and negative space in creative dialogue.

This project makes three key contributions. First, it introduces sand drawing as a novel modality for object detection by large language models. Second, it develops a multimodal physical interface enabling collaborative drawing and natural interaction without screens, encouraging traditional face-to-face engagement. Third, it probes interpretative divergence between human and machine perception, reframing this divergence not as inaccuracy or inefficiency but as something that reflects the diversity and richness of abstract representations. This perspective is particularly important in ongoing discussions around standardization, where computation is often seen as diminishing individuality and diversity.

2 Related Work

Object detection from abstract representations, such as sketches, has been extensively studied in both computer science academia and industry. Sketch-A-Net's CNN[4] architecture achieved better accuracy than humans in sketch recognition. The Sketchy Database[5]provides a benchmark for sketch-based image retrieval and supports studies of cross-domain learning between human-drawn abstractions and photographs. Google's Quick, Draw![6] collected more than 50 million sketches through an online game, interestingly connecting scientific research with collective play. It also demonstrated how line drawings can train neural networks to recognize simplified objects. Trained AI systems like Pixelor [7] and Sketch-RNN [6] create sketches with optimal strokes for recognition. Many other methodologies and recent models have been developed, such as Sketchy Scene[8] and SketchyCOCO[9]. These investigations often frame human-machine comparisons through performance metrics like accuracy and precision. Yet beyond accuracy lies a fertile space of misalignment and imaginative possibility.

More creative precedent work explores interpretation rather than detection, acknowledging the discrepancies embedded in the diversity of human experience, while revealing computational perspectives on collaboration. Artist Tom White's work on perception engines uses Vision Transformers to create artwork recognizable by vision AI. Research into co-drawing workflows with professional artists reveals that artists prefer co-creative AI over didactic AI[10]. SketchPartner was developed to work with designers, interpreting sketches and generating new ones to inspire design exploration[11]. SketchQA works in the opposite direction, generating sequence data to guide human guessers and using data from vision question answering to train the model[12]. Although these experiments demonstrate interesting approaches, none explore diverse materials beyond line-based sketching and none establish collaboration between multiple humans and computers that preserves the social character of game play.

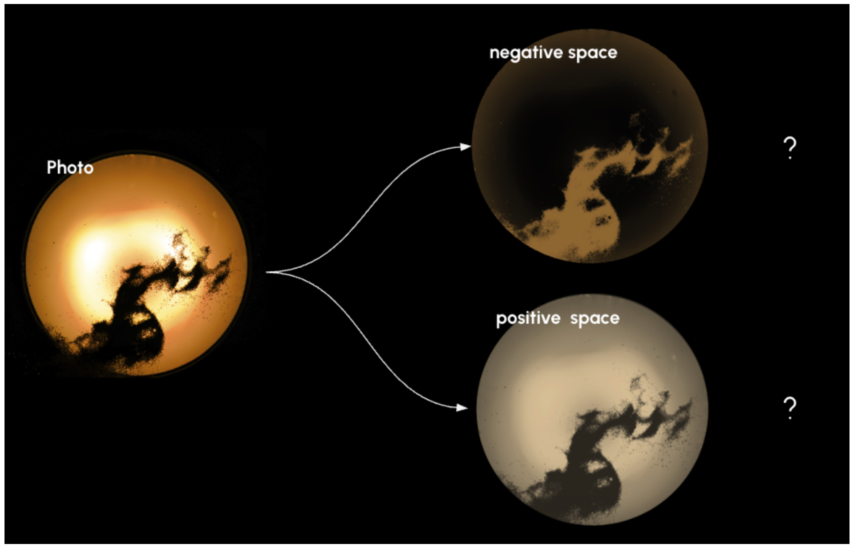

Figure 2: Figure-ground perception

3 Process design

3.1 Concepts: accumulated discrepancy and spatial ambiguity

The pareiduo draws on two conceptual sources. The first is the telephone game, where messages degrade as they pass through multiple participants. We transpose this logic of accumulated discrepancy into human-AI interaction, asking: Can miscommunication itself become a playful mechanic when humans collaborate with large language models? The second influence comes from figure-ground perception exercises: visual puzzles that exploit positive and negative space. Certain drawing games challenge players to identify objects through either the material marks or the residual voids between them. Pareiduo builds on this duality by assigning one player to work with sand formations as positive space and the other to work with the gaps between them as negative space(Figure 2), the interaction and collaboration between two players will center around this strategy as they work on the same surface to draw.

3.2 Game Design: multimodal interaction as structured dialogue

The game operates through a four-phase state machine(Figure 3) that structures multiple rounds of interaction as dialogue rather than isolated turns. Players first receive audio prompts with simple observational questions generated dynamically by the LLM, for example. Who has darker eyes? or "Who is wearing fewer colors?" These prompts avoid prescriptive categories and encourage spontaneous, face-to-face negotiation, assigning roles as either Shadow (working with sand) or Light (working with voids). Each player then leans toward their designated side of the device, where embedded apertures deliver a whispered audio riddle generated by the LLM. This riddle conceals a target object that the player must help the AI recognize. By lowering the audio volume, we ensure that objectives remain private, creating asymmetric knowledge between collaborators. Players then alternate every 15 seconds, shaping the sand surface according to their role. Light intensity functions as the temporal signal, brightness indicates the start of a turn and gradually dims to mark its end, prompting players to yield control without verbal instruction. After every four rounds, the LLM provides audio feedback describing what it perceives in the current configuration and offering guidance for refinement. When players believe their pattern successfully communicates both objects, they press a guess-tell button. The LLM then evaluates the attempt, announces recognition outcomes, and explains its interpretive process before resetting for the next session. This structure creates a slow, iterative conversation between human intention and machine perception, where discrepancies accumulate not as failures but as sites of collaborative negotiation.

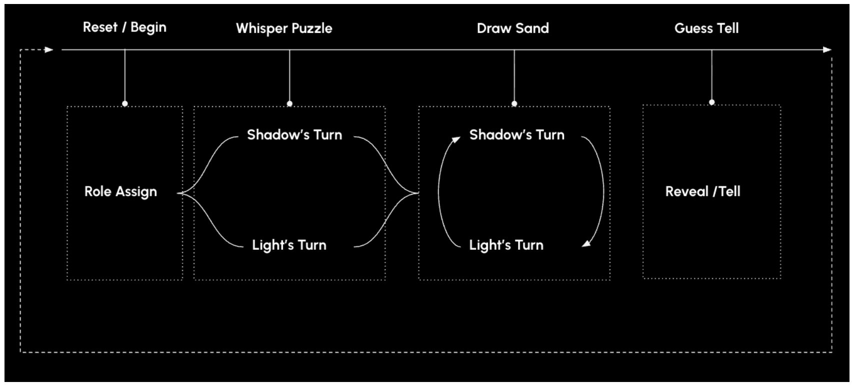

Figure 3: Gameplay Architecture

4 Technical design

To implement the process described above, we developed two integrated systems: a physical interface that supports sand manipulation, audio feedback, and timing cues, and a software architecture that connects image processing with LLM-based prompting for role assignment, object selection, visual interpretation, and feedback generation.

4.1 Physical system: minimal aesthetics and embedded sensing

The device is an illuminated customized sand table(Figure 4). It consists of 3D-printed components, rolled aluminum cones, acrylic surfaces, LED backlighting, an embedded camera, and speakers. Arduino-controlled electronics manage lighting transitions, button inputs, and data transfer between physical and computational layers. The entire housing is 3D-printed with aluminum cones positioned to evoke machine agency; all audio feedback emanates from these apertures, reinforcing the sense that players are playing with a non-human observer. The cone height positions the device as a low table, facilitating comfortable reach and collaborative gestures. Dark exterior surfaces ensure that light gradients remain the primary temporal signal during gameplay. The drawing surface features four textured wing surfaces that reference Zen garden aesthetics. Four buttons are embedded between the wings: one initiates gameplay, one triggers final evaluation, and two deliver whispered objectives to individual players. At the center, outer rings are used for a sand reservoir, and a transparent acrylic sheet serves as the drawing surface. All wiring and electronics are concealed between dual-shell layers, with larger speakers integrated into the wing structures. This minimal design strategy directs attention to the illuminated drawing surface—the site of actual interaction—while receding the technological infrastructure into the background.

4.2 Software design and LLMs integration

The software employs a modular architecture comprising a React-based web application, a pre-processing pipeline for image enhancement through color thresholding, and API integration for vision-based object recognition and text-to-speech synthesis. Our prompt design strategy balances functional constraints with conversational tone across four interaction phases. For role assignment, we instruct the LLM to generate quick, observable prompts based on physical characteristics, producing instructions like "Look at each other's eyes under the light. The person whose eyes reflect more light is the Light player; the other is the Shadow player." This grounds role distribution in immediate, embodied comparison. For object concealment, we prompt the LLM to generate "concise, one-sentence descriptions of objects without naming them directly," producing riddles such as "It has a long handle and bristles at one end, sweeping away what the wind leaves behind." During drawing phases, prompts supply preprocessed images and the player's target object, instructing the LLM to identify two plausible interpretations and provide modification steps in "encouraging, playful" language. The LLM returns structured JSON with actionable guidance: "Smooth the top edge into a gentle curve. Add a small triangle on the right side for a tail." For final evaluation, we request "natural, conversational analysis" with a 1-5 rating and reasoning, producing feedback like "Your

Figure 4: Technical components

fish looks more like a leaf caught in the current, still graceful, but missing its tail fins. Rating: 3/5." By prioritizing tone and constraint articulation in prompt design, we shape the LLM's role as a conversational mediator rather than an authoritative judge.

5 Discussion and future work

By detailing our Pareiduo approach, we contribute to provoke philosophical thinking about representation in the computational age and comparisons between human and computer perception, while giving users a gaming, relaxing, playful experience. In this section we reflect on how we designed the interactive process, share feedback from initial user testing, and discuss future work.

5.1 Design insights: perceptual alignment and collaborative game

Through review sessions of our artwork, we observed several groups of people playing with our device and examining the results. We found that whether there is convergence or divergence between human perception and the LLM's recognition does not significantly influence the experience. This supports our hypothesis that the artwork is not about accuracy, and that divergence exists between humans as well, as we observed from bystander comments. However, in extended play sessions where two players sit down and engage with our device for a sustained period, we found that continuous divergence generates frustration. This finding suggests the need to implement chain-of-thought prompting to encourage the LLM to articulate the reasoning behind its judgments and to increase the proportion of instructions over time to enhance user experience. Our design encourages collaboration between two players and the LLM. The goal is to create a pattern where the black sand can be recognized as the object assigned to the Shadow player, and the remaining surface can be recognized as the object assigned to the Light player, both by the LLM. We challenge the conventional notion that drawing-and-guessing games should be rapid competitions. Users enjoyed our collaborative framework, and the two players communicated throughout the game. We position human-computer interaction within the broader context of human-human interactions that enhance relationships and communication.

5.2 Limitations: more research needed for diverse material modalities

We explored several vision-language models, including LLaMA 3.2-Vision, Qwen2-VL, GPT-4o, and Claude 3.5 Sonnet, ultimately selecting GPT-4o for its performance in recognizing objects from sand drawings. We processed the sand medium into images through a camera and applied image processing to mask irrelevant areas and enhance model performance. While the performance met our expectations, we call for more research on abstract representations—not line drawings or standard visual images, but sketchy drawings in different materialities. Past artworks may already present rich databases for exploration in this direction.

6 Conclusion

Pareiduo situates human-AI interaction to fundamental aspects of human experience: touch, play, and interpret abstract representations. The work deliberately engages two core human capacities. First, pareidolia, our tendency to see familiar forms in ambiguous stimuli, represents a uniquely human way of making sense of the world through visual memory and cultural association. Second, the tactile joy of manipulating materials connects us to childhood play and embodied learning, allowing participants to simply enjoy touching the sand without having to create something useful.

The discrepancies between human perception and LLMs' perception become central to understanding how we interpret abstract patterns in reference to our visual memory of objects in the world. By introducing LLMs into sand drawing, Pareiduo recontextualizes an embodied, communal form of expression within a computationally mediated setting, preserving its physicality while expanding its interpretive possibilities. Differences between a participant's intended form and the AI's perception are not treated as failures but as generative events, prompting players to adapt strategies, reconsider their representations, and explore alternate interpretations. This distinguishes itself from research that focuses on improving LLM performance and accuracy, instead highlighting the value of ambiguity and openness in collaborative creation.

7 Authors

Jiao Zhao is a Master's candidate in City Planning at the Massachusetts Institute of Technology. Her interest includes computational design, data visualization, and interactive installations.

Qian Xiang is a Master's candidate in Design Studies at the Harvard Graduate School of Design. Her practice integrates tangible media, visual design, and product design.

References

[1] Agustin Fuentes. Human niche, human behaviour, human nature. *Interface Focus*, 7(5):20160136, August 2017. doi: 10.1098/rsfs.2016.0136. URL `https://royalsocietypublishing.org/doi/10.1098/rsfs.2016.0136`. Publisher: Royal Society.

[2] Mark P. Mattson. Superior pattern processing is the essence of the evolved human brain. *Frontiers in Neuroscience*, 8:265, August 2014. ISSN 1662-4548. doi: 10.3389/fnins.2014.00265. URL `https://pmc.ncbi.nlm.nih.gov/articles/PMC4141622/`.

[3] Tim Ingold. *Lines : a brief history*. Routledge, London ;, 2007. ISBN 0415424267.

[4] Qian Yu, Yongxin Yang, Yi-Zhe Song, Tao Xiang, and Timothy Hospedales. Sketch-a-Net that Beats Humans, July 2015. URL `http://arxiv.org/abs/1501.07873`. arXiv:1501.07873 [cs].

[5] Patsorn Sangkloy, Nathan Burnell, Cusuh Ham, and James Hays. The sketchy database: learning to retrieve badly drawn bunnies. *ACM Transactions on Graphics*, 35(4):1–12, July 2016. ISSN 0730-0301, 1557-7368. doi: 10.1145/2897824.2925954. URL `https://dl.acm.org/doi/10.1145/2897824.2925954`.

[6] David Ha and Douglas Eck. A Neural Representation of Sketch Drawings, May 2017. URL `http://arxiv.org/abs/1704.03477`. arXiv:1704.03477 [cs].

[7] Ayan Kumar Bhunia, Ayan Das, Umar Riaz Muhammad, Yongxin Yang, Timothy M. Hospedales, Tao Xiang, Yulia Gryaditskaya, and Yi-Zhe Song. Pixelor: a competitive sketching AI agent. so you think you can sketch? *ACM Trans. Graph.*, 39(6):166:1–166:15, November 2020. ISSN 0730-0301. doi: 10.1145/3414685.3417840. URL `https://dl.acm.org/doi/10.1145/3414685.3417840`.

[8] Changqing Zou, Qian Yu, Ruofei Du, Haoran Mo, Yi-Zhe Song, Tao Xiang, Chengying Gao, Baoquan Chen, and Hao Zhang. SketchyScene: Richly-Annotated Scene Sketches, August 2018. URL http://arxiv.org/abs/1808.02473. arXiv:1808.02473 [cs] version: 1.

[9] Chengying Gao, Qi Liu, Qi Xu, Limin Wang, Jianzhuang Liu, and Changqing Zou. Sketchy-COCO: Image Generation from Freehand Scene Sketches, April 2020. URL http://arxiv.org/abs/2003.02683. arXiv:2003.02683 [cs].

[10] Chipp Jansen and Elizabeth Sklar. Exploring Co-creative Drawing Workflows. *Frontiers in Robotics and AI*, 8, May 2021. ISSN 2296-9144. doi: 10.3389/frobt.2021.577770. URL https://www.frontiersin.org/journals/robotics-and-ai/articles/10.3389/frobt.2021.577770/full. Publisher: Frontiers.

[11] Pegah Karimi, Jeba Rezwana, Safat Siddiqui, Mary Lou Maher, and Nasrin Dehbozorgi. Creative sketching partner: an analysis of human-AI co-creativity. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*, pages 221–230, Cagliari Italy, March 2020. ACM. ISBN 978-1-4503-7118-6. doi: 10.1145/3377325.3377522. URL https://dl.acm.org/doi/10.1145/3377325.3377522.

[12] Ravi Kiran Sarvadevabhatla, Shiv Surya, Trisha Mittal, and Venkatesh Babu Radhakrishnan. Game of Sketches: Deep Recurrent Models of Pictionary-style Word Guessing, January 2018. URL http://arxiv.org/abs/1801.09356. arXiv:1801.09356 [cs].