From Mark-Making to Meaning-Making: Opportunities and Challenges for a Multimodal Chatbot to Facilitate Therapeutic Art Activity

Le Lin
City University of Hong Kong
Hong Kong SAR, China
lelin3-c@my.cityu.edu.hk

Zihao Zhu
City University of Hong Kong
Hong Kong SAR, China
zihaozhu9-c@my.cityu.edu.hk

Rainbow Tin Hung Ho
The University of Hong Kong
Hong Kong SAR, China
tinho@hku.hk

Jing Liao*
City University of Hong Kong
Hong Kong SAR, China
jingliao@cityu.edu.hk

Yuhan Luo*
City University of Hong Kong
Hong Kong SAR, China
yuhanluo@cityu.edu.hk

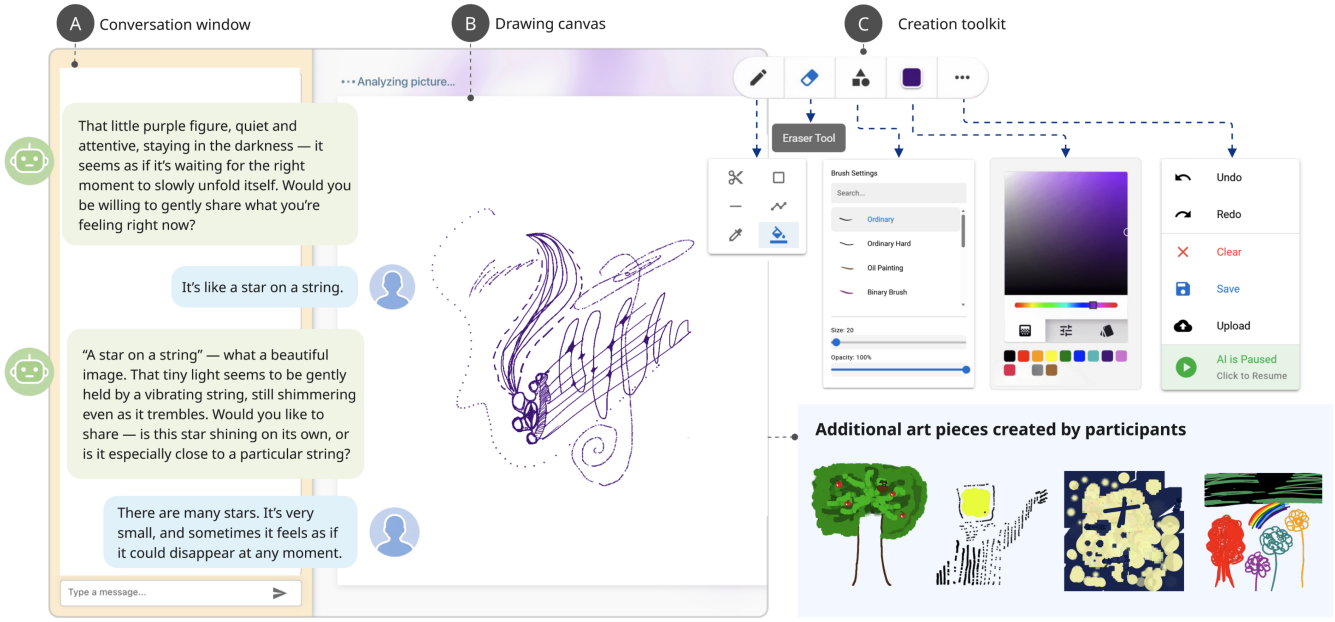

Figure 1: The overview of our AI-mediated therapeutic art activity platform.

Abstract

Therapeutic art thrives on the synergy between creative visual production and reflective dialogue. While traditional digital tools often fail to bridge these two modalities, the emergence of Multimodal Large Language Models (MLLMs) offers a new frontier for AI-mediated support. This paper introduces an ongoing project exploring an AI platform designed to interpret and respond to both the visual nuances of a user's drawing and their accompanying narrative. Reflecting on a preliminary evaluation with five domain experts, we move beyond technical feasibility to discuss several critical open questions: How can AI navigate the subjective ambiguity of art? How does AI mediation alter the therapeutic alliance? How can researchers and developers draw boundaries of machine empathy and the ethical implications of automating sensitive expressive interventions?

Keywords

Therapeutic art, Image understanding, Chatbot customization, Human-AI collaboration, Multimodel agent

1 Introduction and Backgrounds

Therapeutic art activities, such as expressive drawing and painting, are vital tools for emotional processing and stress reduction, significantly improving psychosocial health [4, 20]. Unlike purely verbal communication, these activities leverage visual metaphors to help individuals externalize complex internal thoughts that are

*Corresponding author.

often difficult to articulate [20]. However, the transition from simple mark-making to meaningful self-discovery typically requires a skilled practitioner to guide the reflective process [1, 4]. While the importance of human therapists is undeniable, access to such specialized support remains a significant barrier for many [19].

Recent advancements in multimodal large language models (MLLMs) have shifted digital support from basic text-based chatbots to sophisticated visual interaction [11, 14, 22], offering a pathway to broaden the accessibility of therapeutic art and reduce the social stigma associated with traditional therapy [9]. Existing research has explored AI as a co-creator in digital painting [3] or as a guide for therapist-customized homework [13]. However, a critical gap remains: these systems often lack the ability to link dialogue directly to the evolving visual metaphors on the canvas. This prevents the synchronized multimodal reflection—the interplay between seeing and talking—that is essential to the therapeutic process [13].

While some systems have integrated multimodal understanding, they are rarely designed for therapeutic contexts [7, 10, 17]. Even recent platforms tailored for specific groups, such as trauma-impacted youth, often rely on pre-defined assets that can hinder spontaneous artistic expression [21]. To address this, we developed a web-based, free-form drawing system integrated with a MLLM. The system analyzes visual creations in real time, facilitating a responsive dialogue grounded in the user's immediate visual expression.

In a preliminary qualitative evaluation with five domain experts (full description of the study procedure and findings can be found in [12]), participants recognized the chatbot's potential to foster therapeutic engagement through its real-time visual comprehension. However, the study also surfaced several tensions, including whether the "simulated empathy" displayed by MLLM to mirror language patterns without a shared lived experience can sufficiently sustain a therapeutic alliance, and how digital interaction address the loss of physical touch, the tactile resistance of traditional media, and the sensory grounding that is often central to somatic art therapy. Based on the experts' feedback, we synthesized four imperative directions for future research:

- 1) Identifying appropriate entryways for AI-led inquiry and establishing robust protocols to manage sensitive emotional content and clinical risk.
- 2) Navigating the subtle tension between Depth—determining how far to probe a specific visual metaphor and knowing when to stop to avoid over-interpretation—and Width—directing the user's attention toward untouched or previously unaware elements on the canvas to broaden their perspective.
- 3) Exploring the transition from a purely conversational assistant to a collaborative partner capable of direct, shared visual engagement on the canvas.

2 System Design & Expert Evaluation

To explore the area for MLLM to facilitate therapeutic art activities, we developed an MLLM-powered web system serving as a design probe. The detailed design rationale and system archiecture can be found in [12]. Below, we describe our system design and expert evaluation procedure.

2.1 System Design

The system consists of a web-based frontend and a multi-module backend. The frontend provides an interactive drawing canvas with a conversational sidebar for simultaneous creative and verbal expression (see Figure 1). The backend is powered by the *Qwen3-VL-Plus* [18], featuring two core modules: (1) an **Image Understanding Module** that conducts real-time analysis of users' drawing content, for which we designed a structured prompt that instructs the model to generate three levels of interpretations of the user's artwork; and (2) a **Conversation Module** that integrates the conversational history, current drawing, and analysis results from the image understanding module as input, and synthesizes these contextual inputs to enable the chatbot to generate context-aware, therapeutically appropriate responses. To this end, we designed task-specific prompts and in-context examples tailored to therapeutic art contexts, which are engineered to encourage reflective expression, elicit open-ended and drawing-focused questions, and encourage further creation.

2.2 Expert Evaluation

We conducted individual semi-structured evaluations with five participants: one certificated art therapist, three healing arts practitioners, and one graduate student in expressive art therapy. Participants included one male and four female, and their ages ranged from 23 to 32 (M = 28.6, SD = 3.26). The evaluation involved participants introducing their professional backgrounds and practice workflow, freely exploring the system for therapeutic art activities, sharing their thoughts and perspectives on AI-mediated art therapy. We briefly report the findings in the following sections. The full description can be found in [12].

2.2.1 Positive Outlook. Participants in general shared a positive outlook regarding the role of MLLM AI in therapeutic art activities: *"AI can bring art to everyone's life, fostering their perceptual and aesthetic awareness for wellbeing"* (P1). They recognized the chatbot's potential to serve as a *"low-stakes entry point"* for emotional expression through art-making (P1, P3–P5), praising its **visual comprehension** (P3–P5), and the ability to stay open, neutral, and encourage self-expression. For instance, P3 noted that the chatbot's *"precise analysis of the visual creation"*, which *"deepened the conversation in a meaningful way"*. P4 found that questions asked by the chatbot were grounded in the visual details, such as color and shape of each elements, making the interaction feel personalized.

2.2.2 Tensions. On the other hand, participants all highlighted the importance for mitigating potential risks such as inadvertently triggering severe emotional disturbance. While AI can understand the image content, P1 and P3 pointed out that neither can it truly "understand" how people feel nor "observe" their creation process (e.g., facial expression, body movement). Furthermore, the conversation lacked the depth required for self-discovery (P2), and the digital experience often misses the tactile and sensory richness in physical creation (P1, P5). In light of these challenges, they raised several areas for future improvement.

Entryways & Risk Management. All participants suggested a screening process to exclude individuals, for whom unguided deep-diving into visual metaphors could lead to emotional dysregulation. P4 suggested that the system should monitor for signs of therapeutic

rupture—instances where the user becomes frustrated, angry, or feels misunderstood by the AI: *"If the user had rejected the AI's suggestions, or if the system can detect signs of anger or dissatisfaction in their emotions, it may be the right moment to introduce human interventions."*

Bespoke Alignment Between User Profile and AI's Therapeutic Style. Participants emphasized that a "one-size-fits-all" AI persona cannot meet the diverse user needs and proposed two dimensions of personalization. First, art therapy is grounded in a wide range of theoretical frameworks, and the user should have an option to choose which style they prefer (P4). P5 shared a related example from the user's perspective: *"If I were a user and a dancer, I would prefer to choose a therapist who understands dancing. Otherwise, they might not fully grasp my experiences."*

Balancing Therapeutic Depth and Width. All participants hoped the chatbot could ask more details about the visual elements they created on the Canvas. Particularly, P1 and P5 mentioned that the questions should move from the big picture to smaller pieces, which can help lower the user's defenses, enabling a continuous exploration of their thoughts. Beyond deep reflection, P2 highlighted the necessity of "*conversational width*." This could give users more space for imagination rather than limiting their thoughts in one box, and for people experience acute mental health distress, this can *"avoid overstepping into potentially triggering territory"* (P2).

Enriching Visual Interactivity. During the debriefing interviews, participants expressed interest in enhancing the art-making process through contextual UI interactions and human–AI co-creation. For example, P5 suggested that the AI chatbot could engage with specific visual elements by highlighting or masking regions directly on the canvas for discussion, rather than putting them in pure language description, as visual communication is an important part of drawing-based art therapy. P2 added, *"Compared to text, I personally tend to respond more naturally to visual elements."*

3 Discussions & Future Research Directions

Our expert evaluation suggests that MLLMs can act as a "reflective mirror" that faciliate self-expression and art creation, yet several fundamental gaps remain between algorithmic processing and therapeutic presence. We identify four critical areas for future interdisciplinary research.

3.1 Safe Entryways

A key challenge in identifying appropriate "entryways" is the transition from passive observer to active inquirer. This is because AI-led inquiry risks opening emotional wounds that it may not have the capacity to close [21]. Thus, future research needs to explore ways for understanding visual and linguistic markers of individual emotional flooding—a state where the user's nervous system is overwhelmed or mental system is vulnerable [6]. Because emotional expression is deeply idiosyncratic, identifying these markers requires an idiographic, longitudinal approach rather than a universal classification model. For an MLLM, this necessitates establishing a personalized affective baseline by tracking a user's unique "visual dialect" over time. As suggested by Santosa et al. [16], a deviation from a user's common expression pattern—such as a sudden shift in stroke density, spatial fragmentation, or use of color relative to their interaction history—serves as a more reliable indicator of emotion fluctuation than any generalized metric.

3.2 Navigating the Trajectory of Conversation

The central dilemma in guiding the creative process lies in managing the trajectory of inquiry. Our study identify a fundamental tension between the depth and width of the exploration, which requires the AI to possess the ability to navigate the user's creative flow. While probing a single visual metaphor can lead to cathartic breakthroughs, it carries the significant risk of over-interpretation or "hallucinating" meaning [2]. If an MLLM imposes a standard symbolic interpretation (e.g., assuming a red circle must represent anger) before the user has arrived at that conclusion, it can cause narrative foreclosure, where the user's organic process of self-discovery is silenced by the AI's suggestion. Future work must investigate non-directive prompting strategies that facilitate scaffolding [5], ensuring the AI only moves as deep as the user's current zone of proximal development allows.

3.3 Adapting to Individual Profiles

A limitation of current MLLMs is their tendency toward a monolithic therapeutic style, usually characterized by a generic, supportive tone. However, as experts highlighted, there are many distinct streams of art therapy practice regarding the media use (e.g., water color versus wet clay), focus of inquiry (e.g., process versus product), and role of the art creation (e.g., extension of the self verus the third party in the space), etc [2, 15, 20]. As such, the efficacy of an intervention is contingent upon its alignment with the individual's clinical goals, cultural background, and the evolving therapeutic relationship [8]. Future research should explore how AI can dynamically shift between these established theoretical frameworks to match a user's situational needs.

3.4 Toward Collaborative Visual Agency

The current state of AI-mediated art therapy is primarily conversational, where the AI "sees" but does not "touch." Our study points to a desire for a more collaborative interaction paradigm. For the future research, we aim to explore the transition from a conversational assistant to a partner capable of shared visual engagement. This includes investigating co-creative acts where the AI and user manipulate the canvas together. Such interaction could help bridge the materiality gap by fostering a sense of relational connection and shared agency that is currently missing from purely dialogue-based systems.

References

[1] Sherry L Beaumont. 2012. Art therapy approaches for identity problems during adolescence. *Canadian Art Therapy Association Journal* 25, 1 (2012). doi:10.1080/08322473.2012.11415557

[2] Tessa Dalley. 2008. *Art as therapy: An introduction to the use of art as a therapeutic technique.* Routledge. doi:10.4324/9780203134825

[3] Xuejun Du, Pengcheng An, Justin Leung, April Li, Linda E Chapman, and Jian Zhao. 2024. DeepThInk: Designing and probing human-AI co-creation in digital art therapy. *International journal of human-computer studies* 181 (2024), 103–139. doi:10.1016/j.ijhcs.2023.103139

[4] Holly Feen-Calligan. 1996. Art therapy as a profession: Implications for the education and training of art therapists. *Art Therapy* 13, 3 (1996). doi:10.1080/07421656.1996.10759216

[5] Francesca Fusco, Caroline Pitt, Michele Newman, Jason Yip, and Nikita Soni. 2026. Towards Understanding Children's Collaborative Interaction Patterns in

Child-AI Co-creative Interfaces. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*. 1–20. doi:10.1145/3772318.3791831

[6] Gärd Holmqvist, Åsa Roxberg, Ingrid Larsson, and Cristina Lundqvist-Persson. 2017. What art therapists consider to be patient's inner change and how it may appear during art therapy. *The arts in psychotherapy* 56 (2017), 45–52. doi:10.1016/j.aip.2017.07.005

[7] Zeyuan Huang, Cangjun Gao, Yaxian Shan, Haoxiang Hu, Qingkun Li, Xiaoming Deng, Cuixia Ma, Yu-Kun Lai, Yong-Jin Liu, Feng Tian, et al. 2025. SketchGPT: A sketch-based multimodal interface for application-agnostic LLM interaction. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. 1–18. doi:10.1145/3746059.3747598

[8] Amanda Lazar, Jessica L Feuston, Caroline Edasis, and Anne Marie Piper. 2018. Making as expression: Informing design with people with complex communication needs through art therapy. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–16. doi:10.1145/3173574.3173925

[9] Yi-Chieh Lee, Yichao Cui, Jack Jamieson, Wayne Fu, and Naomi Yamashita. 2023. Exploring effects of chatbot-based social contact on reducing mental illness stigma. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–16. doi:10.1145/3544548.3581384

[10] Jiaye Leng, Hui Ye, Pengfei Xu, Miu-Ling Lam, and Hongbo Fu. 2025. GenFO-Drawing: Supporting Creative Found Object Drawing With Generative AI. *IEEE Transactions on Visualization and Computer Graphics* (2025). doi:10.1109/TVCG.2025.3626754

[11] Chaoyu Li, Sid Padmanabhuni, Maryam S Cheema, Hasti Seifi, and Pooyan Fazli. 2025. Videoa11y: Method and dataset for accessible video description. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–29. doi:10.1145/3706598.3714096

[12] Le Lin, Zihao Zhu, Jing Liao, and Yuhan Luo. 2026. Exploring a Multimodal Chatbot as a Facilitator in Therapeutic Art Activity.. In *Proceedings of the 2026 ACM Designing Interactive Systems Conference*. https://doi.org/10.1145/3802974.3809425

[13] Di Liu, Jingwen Bai, Zhuoyi Zhang, Yilin Zhang, Zhenhao Zhang, Jian Zhao, and Pengcheng An. 2025. TherAIssist: Assisting Art Therapy Homework and Client-Practitioner Collaboration through Human-AI Interaction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, Volume 9, Issue 3* (2025). doi:10.1145/374949

[14] Franklin Mingzhe Li, Michael Xieyang Liu, Cynthia L Bennett, and Shaun K Kane. 2026. ADCanvas: Accessible and Conversational Audio Description Authoring for Blind and Low Vision Creators. (2026).

[15] Judith Aron Rubin. 2016. *Approaches to art therapy: Theory and technique*. Routledge. doi:10.4324/9781315716015

[16] Shannon Sie Santosa, Qian Wan, Junnan Yu, and Yuhan Luo. 2026. EmoFlow: From Tracking to Sense-Making of Emotions Through Creative Drawing. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*.

[17] Weiyan Shi, Sunaya Upadhyay, Geraldine Quek, and Kenny Tsu Wei Choo. 2026. TalkSketch: Multimodal Generative AI for Real-Time Sketch Ideation with Speech. In *International Workshop on Creative AI for Live Interactive Performances*. Springer, 83–97. doi:10.1007/978-3-032-16893-1_6

[18] Qwen Team. 2025. Qwen3-vl technical report. *arXiv preprint arXiv:2511.21631* (2025). doi:10.48550/arXiv.2511.21631

[19] Lesley Uttley, Alison Scope, Matt Stevenson, Andrew Rawdin, Elizabeth Taylor Buck, Anthea Sutton, John Stevens, Eva Kaltenthaler, Kim Dent-Brown, and Chris Wood. 2015. Systematic review and economic modelling of the clinical effectiveness and cost-effectiveness of art therapy among people with non-psychotic mental health disorders. *Health Technology Assessment (Winchester, England)* 19, 18 (2015), 1. doi:10.3310/hta19180

[20] Asnat Weinfeld-Yehoudayan, Johanna Czamanski-Cohen, Miri Cohen, and Karen L Weihs. 2024. A theoretical model of emotional processing in visual artmaking and art therapy. *The Arts in Psychotherapy* 90 (2024), 102196. doi:10.1016/j.aip.2024.102196

[21] Christine Wu, Ila Krishna Kumar, and Rosalind Picard. 2025. Navigating Emotions Through Art: Recommendations for Designing Art-Therapy Based Chatbots for Trauma-Impacted Youth: Recommendations for Designing Art-Therapy Based Chatbots for Trauma-Impacted Youth. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–7. doi:10.1145/3706599.3719869

[22] Chanjin Zheng, Zengyi Yu, Yilin Jiang, Mingzi Zhang, Xunuo Lu, Jing Jin, and Liteng Gao. 2025. ArtMentor: AI-Assisted Evaluation of Artworks to Explore Multimodal Large Language Models Capabilities. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–18. doi:10.1145/3706598.3713274