Interactive Generative Art for Enhancing Emotional Expression in Digital Mental Health

Jerome M. Dawson
Department of Computer Science, University of North Texas, Denton, TX, USA.
jeromedawson338@unt.edu

Gerald Koskinen
Department of Computer Science, University of Houston, Houston, TX, USA.
koskinen194@uh.edu

Qianbo Sun
School of Computing, Clemson University, Clemson, SC, USA.
qianbo.sun@clemson.edu

Abstract

Interactive generative art occupies a promising yet underexamined position between expressive arts interventions and computational mental health support. This paper develops a system-level account of how interactive generative art can enhance emotional expression in digital mental health. Rather than treating generative models as isolated technical artifacts, the analysis focuses on architectural composition, interaction infrastructure, deployment pressures, governance constraints, robustness requirements, fairness challenges, and long-term sustainability. The discussion integrates affective computing, human-computer interaction, generative modeling, and clinical implementation research to identify structural trade-offs that shape expressive benefits and risks. A central argument is that emotional expression is not produced solely by an image or sound generator; it emerges from the orchestration of perception, interpretation, feedback loops, clinical framing, and sociotechnical trust. Interactive generative art can externalize difficult emotional states, scaffold reflective awareness, and support therapeutic dialogue, but only when systems are designed with attention to latency, interpretability, safety, and inclusive representation. The paper further examines regulatory tensions around automated mental health tools, the need for robust failure management, and conditions for sustainable deployment across diverse care settings. By reframing generative art as a system-level intervention rather than a feature-level enhancement, the paper offers a structural lens for researchers, clinicians, and policy actors seeking to align computational creativity with the ethical demands of mental health care.

Keywords

interactive generative art; digital mental health; affective computing; human-AI co-creation; socio-technical systems; governance.

1. Introduction

Digital mental health interventions have expanded rapidly as a response to persistent gaps in access, cost, and stigma that characterize traditional mental health services [1]. Many existing tools emphasize structured self-reporting, psychoeducation, or text-based cognitive behavioral exercises, which can be effective but often struggle to engage users whose emotional states are difficult to verbalize. Affective computing has long sought to bridge this gap by enabling computational systems to recognize, interpret, and respond to human emotional processes [2]. At the same time, advances in personal sensing and machine learning have created new possibilities for inferring psychological states from behavioral and physiological signals [3]. In parallel, generative image-to-image translation and related techniques have altered the expressive capacity of interactive media, allowing users to transform visual and auditory content in ways that feel immediate, participatory, and personally meaningful [4]. These developments invite a broader system-level examination of how interactive generative art might support emotional expression not simply as a creative novelty but as a carefully embedded component of digital mental health infrastructure.

The value of generative art for mental health cannot be reduced to model accuracy or aesthetic quality. Emotional expression in therapeutic contexts depends on trust, safety, interpretation, and the user's sense of agency. Interactive generative systems can translate bodily movement, physiological arousal, vocal tone, or manual drawing into evolving visual or musical forms, thereby creating external representations of internal states. This process may facilitate emotional disclosure, reflection, and clinical conversation. However, such benefits depend on infrastructure decisions that are frequently overlooked in proof-of-concept studies. These include interaction latency, content safety, data governance, model bias, clinical oversight, and long-term maintainability. A system-level perspective is therefore essential. This paper examines interactive generative art as a sociotechnical intervention, emphasizing structural trade-offs across architecture, deployment, fairness, robustness, and policy.

2. System-Level Foundations of Interactive Generative Art in Mental Health

Interactive generative art systems for mental health typically combine multiple functional layers. These include an input layer that captures user behavior or biosignals, a generative layer that synthesizes aesthetic content, an output layer that displays or performs the content, and a reflective layer that supports interpretation. In clinical or wellness settings, a further governance layer mediates access, documentation, and escalation. Generative adversarial networks have been foundational in enabling controllable synthesis across domains, while conditional generation allows user inputs to shape outputs in semantically meaningful ways [5]. More recent text-conditional systems have expanded the range of expressive mappings, though they also introduce new questions about prompt sensitivity and content control [6]. From a system standpoint, the central challenge is not whether a model can generate compelling art, but whether the entire interaction loop can remain responsive, interpretable, and safe under the emotional variability of real users.

A defining structural feature of this domain is the coupling between aesthetic openness and clinical predictability. Generative systems are valued because they can produce surprising, evocative, and nonliteral forms that soften the constraints of verbal expression. Yet mental health settings often require some degree of predictability to avoid triggering harmful content or reinforcing maladaptive interpretations. This tension cannot be resolved simply by constraining model outputs after generation. It must be addressed through layered design choices, including shared control between user and system, curated generative spaces, real-time moderation, and context-aware fallback strategies. The expressive value of generative art is thus a property of the entire human-machine system rather than an attribute of the generative model alone.

3. Architectural Design and Interaction Infrastructure

The architecture of interactive generative art systems must balance expressive richness with usability under emotional distress. Research on design aesthetics indicates that visual quality influences perceived usability and emotional response, but aesthetic appeal alone does not guarantee meaningful engagement [7]. Interaction design must support mental models that users can grasp even when attention is fragmented or affect is intense [8]. Many mobile mental health applications have been criticized for low retention and weak integration into daily life, despite acceptable surface design [9]. Effective generative art systems therefore require more than elegant interfaces. They require coherent interaction grammars, clear feedback, and graceful recovery from user confusion. Early conversational systems such as ELIZA demonstrated that even simple interaction loops can generate therapeutic presence when users project meaning into system responses [10]. Interactive generative art extends this principle into visual, auditory, and embodied domains, where ambiguity is often a resource rather than a defect.

Architecturally, the system should separate the generative engine from the interpretive layer and the clinical wrapper. This separation enables modular updates, differential risk assessment, and context-specific configuration. For example, an abstract visual generative module may be reused across clinical and wellness contexts, while the interpretive prompts, safety filters, and escalation pathways differ. Latency is another infrastructure concern. Emotional expression often unfolds in real time through gesture, voice, or physiological change. If generative output lags behind user action, the felt sense of agency diminishes and the therapeutic quality of the interaction degrades. Edge processing, adaptive resolution, and predictive pre-rendering can reduce perceived latency, but they must be balanced against privacy and device accessibility. A system designed for high-end virtual reality may not function equitably on low-cost smartphones, making architectural modularity a fairness issue as well as an engineering concern.

4. Generative Models and Emotional Expression Mechanisms

Generative models support emotional expression through transformation, metaphor, and co-creation. A user may begin with an abstract mark, a sound, or a movement trace, and the system may elaborate it into a complex visual texture or harmonic progression. This process can externalize emotion without requiring explicit verbal labeling. It also creates a shared artifact that can be observed, revised, and discussed. Human-AI interaction guidelines emphasize the importance of clarity about system capability, feedback timing, and user control [11]. These guidelines are especially relevant when generative models are invited into emotionally sensitive territory. Users should understand that the system is not interpreting their mental state with clinical certainty, but rather offering aesthetic amplification that may support reflection. A recent system-level study of AI-assisted human-pet artistic musical co-creation illustrates how co-creation can be structured as a wellness ritual rather than as a diagnostic procedure [12]. The emphasis on partnership, play, and aesthetic participation suggests that generative systems may be most effective when they avoid clinical authority and instead support expressive agency.

The interpretive mechanism matters as much as the generative model itself. A user may find emotional meaning in a produced image or sound, but that meaning is shaped by prompts, labels, and the surrounding therapeutic framing. Systems can support reflection by offering open questions rather than fixed emotional classifications. They can also allow users to modify generative outputs, thereby reinforcing a sense of authorship. This participatory loop distinguishes generative art from passive content delivery. It also introduces design challenges because the system must track user preferences, infer likely safety risks, and maintain coherent interaction across sessions. Personalization can increase engagement, but excessive adaptation may narrow the expressive space and inadvertently reinforce maladaptive patterns. Structural trade-offs therefore arise between personalization and creative novelty, between reflective open-endedness and clinical guidance, and between user autonomy and system responsibility.

5. Deployment Contexts and Clinical Integration

Deployment of interactive generative art in mental health settings requires careful alignment with clinical workflows and regulatory expectations. In formal care contexts, clinicians may need to understand why a particular output appeared, what input produced it, and how it relates to a patient's emotional state. Automated systems that make consequential recommendations or influence clinical interpretation can raise expectations of explanation, even when existing data protection frameworks do not always guarantee a right to explanation for every automated decision [13]. Although generative art systems are typically framed as expressive tools rather than diagnostic devices, their integration into therapy can blur this boundary. A clinician may ask a patient to interpret a generated image, and that interpretation may then inform treatment planning. Under these conditions, the system becomes part of a clinical sense-making process, and its opacity may become ethically significant.

Contextual deployment also demands attention to the distinction between wellness use and clinical use. Wellness applications may operate with lower regulatory oversight but still reach vulnerable users. Clinical integration, by contrast, may require validation studies, adverse event reporting, professional training, and interoperability with electronic health records. Generative art systems intended for clinical use must be designed to produce auditable interaction logs, support clinician review, and include escalation pathways when users express distress. At the same time, excessive logging can undermine the sense of psychological safety that expressive art requires. This trade-off cannot be eliminated by technical means alone. It requires governance policies that clarify what data are recorded, who may access them, and how they may be used. The architecture should support contextual data minimization, allowing richer recording only when clinically justified and consented.

6. Robustness, Safety, and Failure Management

Robustness in interactive generative art involves more than model uptime or software reliability. It includes the capacity to handle emotionally volatile inputs without producing harmful, destabilizing, or coercive outputs. Generative systems can fail in subtle ways, such as misinterpreting a movement, producing semantically incongruent imagery, or reinforcing a user's negative self-perception through unintended aesthetic associations. The ethics of algorithms literature highlights that algorithmic systems embed moral assumptions through data selection, optimization targets, and interaction design [14]. In mental health contexts, these assumptions can influence what counts as a valid emotional expression, whose aesthetic preferences are normalized, and how distress is framed. Robustness must therefore be assessed at the level of the sociotechnical system, including human moderators, fallback interactions, and feedback loops that allow users to flag harmful content.

Failure management should distinguish between technical failures, semantic failures, and emotional failures. A technical failure may involve model latency, output corruption, or platform incompatibility. A semantic failure may involve a generated image that is technically valid but emotionally confusing. An emotional failure may occur when the system responds in a way that dismisses, pathologizes, or exploits user vulnerability. Each mode requires different safeguards. Technical failures can be addressed through redundancy and graceful degradation. Semantic failures require domain-specific content constraints and user controls. Emotional failures demand relational design, such as offering pauses, human contact options, and nonjudgmental recovery prompts. A system that only logs errors may miss the most consequential harms. Robustness frameworks should therefore include structured review of interaction transcripts, escalation metrics, and qualitative feedback from users and clinicians.

7. Fairness, Bias, and Inclusive Design

Fairness in interactive generative art is shaped by representational diversity, accessibility, and the differential distribution of expressive affordances. Generative models trained on biased visual or musical datasets may reproduce narrow aesthetic norms, underrepresent certain cultural traditions, or systematically distort particular bodies and emotions. Intersectional accuracy disparities in automated perception systems demonstrate that fairness problems are not hypothetical but measurable and often severe [15]. A generative art system that misreads movement or facial expression for some demographic groups may exclude them from the intended expressive benefits. More broadly, bias in machine learning systems arises from data collection, feature selection, model objectives, and evaluation practices [16]. Large generative models also absorb social stereotypes from training corpora, requiring careful auditing and mitigation when deployed in mental health [17].

Inclusive design must go beyond demographic balancing in training data. Interaction modalities should accommodate users with different motor, sensory, and cognitive capacities. Visual generative systems may need audio descriptions or haptic equivalents. Voice-driven systems may need text or gesture alternatives. Language models used in prompts should avoid assuming fluency in a dominant language. Moreover, fairness requires attention to the cultural specificity of emotional expression. Aesthetic forms that seem neutral, such as color palettes, musical modes, or abstract shapes, can carry culturally specific meanings. Systems that impose a single expressive vocabulary risk alienating users whose emotional lives are organized differently. Participatory design with diverse user communities is therefore not an optional add-on but a structural requirement for fairness. This also implies that fairness evaluation should include measures of expressive agency, cultural fit, and perceived safety, not only statistical parity in model outputs.

8. Governance, Privacy, and Policy Implications

Governance of interactive generative art in mental health must address privacy, clinical accountability, intellectual property, and evolving artificial intelligence regulation. Generated artifacts may reveal sensitive emotional states, and interaction histories may be more revealing than the final images or sounds. Privacy protection requires clear data governance structures that limit retention, prevent unauthorized secondary use, and support user deletion. Generative systems can also produce content that raises copyright or authorship questions, though these concerns are secondary to the more immediate risk of emotional harm. High-level artificial intelligence principles emphasize beneficence, non-maleficence, autonomy, justice, and explicability, but such principles require translation into specific system requirements [18]. In mental health, the principle of non-maleficence may demand pre-deployment review of trigger risks, while autonomy may require meaningful consent processes that go beyond generic terms of service.

Policy actors face a difficult balance between encouraging therapeutic innovation and preventing unregulated mental health claims. Interactive generative art systems may be marketed as wellness tools even when they function as adjuncts to therapy. Regulatory definitions should focus on the function of the system rather than the technology used. If a system interprets emotional expression, records clinically relevant patterns, or recommends actions, it may require higher oversight. Interoperability with health systems raises additional questions about medical device classification and data protection. Governance should also support independent auditing of generative content, safety filters, and bias mitigation. Without external review, developers may lack incentives to document known limitations. Transparent documentation, incident reporting, and post-market surveillance can help align system behavior with public health values.

9. Sustainability and Long-Term System Evolution

Sustainability requires architectural and organizational strategies that allow interactive generative art systems to evolve without accumulating harmful debt. Short-term deployments often rely on cloud-based generative models with high computational costs, which may be unsustainable in low-resource mental health settings. Virtual and augmented reality interventions have shown therapeutic promise, but their infrastructure demands remain unevenly distributed [19]. Emerging evidence on artificial intelligence in mental health suggests that realization depends on translational infrastructure, clinical validation, and ongoing safety monitoring rather than breakthrough model performance alone [20]. Positive computing perspectives also emphasize that wellbeing technologies should be designed for long-term human flourishing rather than short-term engagement metrics [21].

Long-term evolution should be guided by modular architecture, open standards, and federated governance. Modular separation between generative engines, safety layers, interaction templates, and clinical reporting allows components to be replaced as models improve. Open standards for interaction logs and safety annotations can support cross-platform learning while preserving privacy. Engagement-oriented mechanisms such as gamified feedback can sustain motivation, but they must not overwhelm the reflective aims of emotional expression [22]. The system should be able to scale down for low-bandwidth environments and scale up for immersive therapeutic settings. Sustainability also depends on workforce capacity. Clinicians and support staff need training to interpret generative artifacts, respond to user distress, and recognize when the system is not appropriate. A sustainable sociotechnical system therefore includes documentation, education, maintenance routines, and community accountability, not only software updates.

10. Conclusion

Interactive generative art offers a distinct pathway for enhancing emotional expression in digital mental health by transforming internal states into shareable aesthetic forms. Its value, however, is not guaranteed by generative capability. The structural argument advanced in this paper is that expressive benefit depends on how generative models are embedded within broader interaction, clinical, and governance systems. Architecture must support low-latency participatory loops, modular safety controls, and equitable access across devices and sensory modalities. Deployment must clarify the boundary between wellness and clinical use while preserving auditable pathways for clinician review. Robustness must account for technical, semantic, and emotional failure modes. Fairness must be evaluated through inclusive design and cultural fit, not only aggregate model metrics. Governance must translate high-level principles into practical safeguards for privacy, consent, and accountability. Finally, sustainability requires modular evolution, infrastructure equity, and long-term organizational commitment. These system-level considerations define the conditions under which interactive generative art can responsibly enhance emotional expression, not as a technological substitution for human care, but as an expressive infrastructure that supports reflection, dialogue, and therapeutic presence.

References

1. Torous, J., Myrick, K. J., Rauseo-Ricupero, N., & Firth, J. (2020). Digital mental health and COVID-19: Using technology today to accelerate the curve on access and quality tomorrow. JMIR Mental Health, 7(3), e18848. https://doi.org/10.2196/18848

2. Picard, R. W. (1997). Affective computing. MIT Press.

3. Mohr, D. C., Zhang, M., & Schueller, S. M. (2017). Personal sensing: Understanding mental health using ubiquitous sensors and machine learning. Annual Review of Clinical Psychology, 13, 23-47. https://doi.org/10.1146/annurev-clinpsy-032816-044949

4. Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1125-1134). IEEE.

5. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139-144. https://doi.org/10.1145/3422622

6. Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with CLIP latents. arXiv preprint arXiv:2204.06125.

7. Sonderegger, A., & Sauer, J. (2010). The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. Applied Ergonomics, 41(3), 403-410. https://doi.org/10.1016/j.apergo.2009.09.002

8. Norman, D. A. (2013). The design of everyday things (Revised ed.). Basic Books.

9. Bakker, D., Kazantzis, N., Rickwood, D., & Rickard, N. (2016). Mental health smartphone apps: Review and evidence-based recommendations for future developments. JMIR Mental Health, 3(1), e7. https://doi.org/10.2196/mental.4984

10. Weizenbaum, J. (1966). ELIZA—A computer program for the study of natural language communication between man and machine. Communications of the ACM, 9(1), 36-45. https://doi.org/10.1145/365153.365168

11. Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1-13). ACM. https://doi.org/10.1145/3290605.3300233

12. Wang, Z., Ma, L., Jin, Y., Feng, Y., Pan, X., Ji, S., & Zhang, K. (2025, August). AI-assisted human-pet artistic musical co-creation for wellness therapy. In Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (pp. 10216-10224).

13. Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. International Data Privacy Law, 7(2), 76-99. https://doi.org/10.1093/idpl/ipx005

14. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 2053951716679679. https://doi.org/10.1177/2053951716679679

15. Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of Machine Learning Research, 81, 77-91.

16. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), 1-35. https://doi.org/10.1145/3457607

17. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610-623). ACM. https://doi.org/10.1145/3442188.3445922

18. Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1). https://doi.org/10.1162/99608f92.8cd550d1

19. Riva, G., Baños, R. M., Botella, C., Mantovani, F., & Gaggioli, A. (2016). Transforming experience: The potential of augmented reality and virtual reality for enhancing personal and clinical change. Frontiers in Psychiatry, 7, 164. https://doi.org/10.3389/fpsyt.2016.00164

20. Koutsouleris, N., Hauser, T. U., Skvortsova, V., & De Choudhury, M. (2022). From promise to practice: Towards the realisation of AI-informed mental health care. The Lancet Digital Health, 4(11), e829-e840. https://doi.org/10.1016/S2589-7500(22)00153-4

21. Calvo, R. A., & Peters, D. (2014). Positive computing: Technology for wellbeing and human potential. MIT Press.

22. Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining gamification. In Proceedings of the 15th International Academic MindTrek Conference (pp. 9-15). ACM. https://doi.org/10.1145/2181037.2181040