

# Multimodal Activity Traces for Creative Repurposing

**KEXUE FU**, Studio for Narrative Spaces, City University of Hong Kong, China

**JINGFEI HUANG**, Harvard University, USA

**LONG LING**, Tongji University, China

**TOBY JIA-JUN LI**, University of Notre Dame, USA

**RAY LC**, Studio for Narrative Spaces, City University of Hong Kong, China

CCS Concepts: • **Human-centered computing** → **Interaction paradigms**.

Additional Key Words and Phrases: Multimodality, Creativity Support, Storytelling

## ACM Reference Format:

Kexue Fu, Jingfei Huang, Long Ling, Toby Jia-Jun Li, and RAY LC. 2026. Multimodal Activity Traces for Creative Repurposing. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*, April 13–17, 2026, Barcelona, Spain. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

## 1 Introduction

Creative work has always been multimodal. Kandinsky used music to empower his paintings, while Richard Strauss wrote symphonic poems from on narrative descriptions reminiscent of paintings. We think about stories in the form of visual scenes, and understand visual environments in the form of textual descriptions in a novel. In the age of generative AI, this multimodal nature of creative work is instantiated in the form of text-to-image[2, 17], and image-to-text tools[9]. Creators ideate with language models, generate images in separate tools, sketch in yet another application, and archive fragments across cloud folders and chat histories [5, 10–12]. Yet these modalities remain siloed across disconnected platforms, preventing creators from fluidly combining and referencing across modalities in ways that would naturally spark cross-modal inspiration [3, 4, 16]. Compounding this fragmentation, the multimodal artifacts produced during AI-assisted creative sessions — images, sketches, generated text, and their coupled transformations — are rarely preserved in a form that supports future reuse. Instead, they dissolve into chat histories and session logs, invisible to both the creator and the researcher.

This loss matters for creative practice. Recent work suggests that versioning in creative domains is not primarily about precise rollback or linear progress, but about psychological safety, reuse, and long-term recombination [6]. Creative practitioners treat prior artifacts as a palette — a reservoir of fragments, variations, and partial ideas that can be revisited, repurposed, and transformed over time [13]. When multimodal artifacts are preserved and made accessible,

---

Authors' Contact Information: **Kexue Fu**, Studio for Narrative Spaces, and City University of Hong Kong, School of Creative Media, Hong Kong, SAR, China, [kexuefu2-c@my.cityu.edu.hk](mailto:kexuefu2-c@my.cityu.edu.hk); **Jingfei Huang**, Harvard University, Graduate School of Design & School of Engineering and Applied Science, Cambridge, Massachusetts, USA, [jingfeihuang@mde.harvard.edu](mailto:jingfeihuang@mde.harvard.edu); **Long Ling**, Tongji University, College of Design and Innovation, Shanghai, China, [lucyling0224@gmail.com](mailto:lucyling0224@gmail.com); **Toby Jia-Jun Li**, University of Notre Dame, Department of Computer Science and Engineering, Notre Dame, Indiana, USA, [toby.j.li@nd.edu](mailto:toby.j.li@nd.edu); **RAY LC**, [ray.lc@cityu.edu.hk](mailto:ray.lc@cityu.edu.hk), Studio for Narrative Spaces, and City University of Hong Kong, Hong Kong, SAR, China.



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

© 2026 Copyright held by the owner/author(s).

Manuscript submitted to ACM

Manuscript submitted to ACM

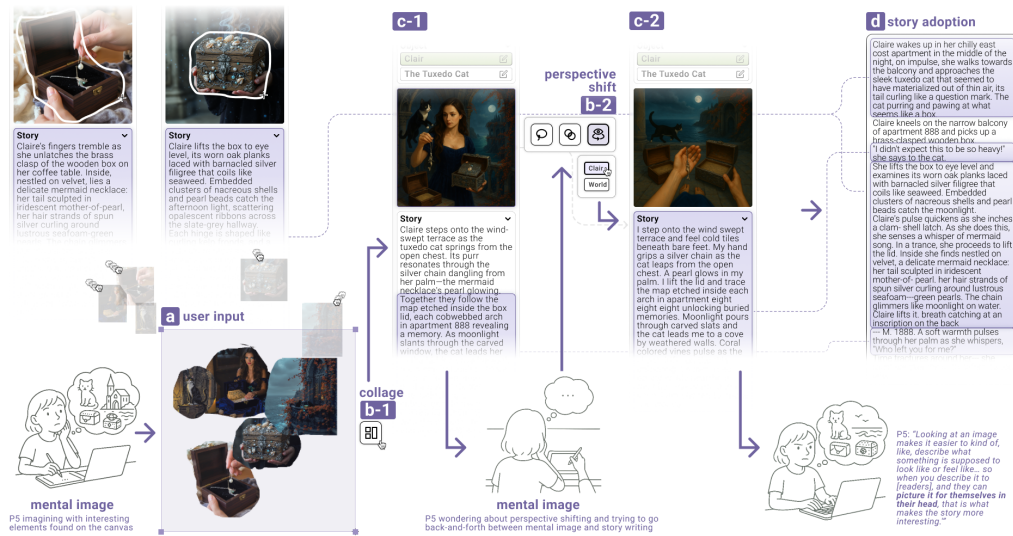


Fig. 1. The canvas serves as a persistent representational space where creative iterations across text and image are recorded as multimodal traces. Rather than disappearing into linear histories, these artifacts remain available for reuse, enabling creators to branch into new exploratory trajectories.

they serve as cognitive scaffolds: a previously generated image can re-anchor a drifting narrative; a discarded sketch can re-emerge as the seed of a new scene; a stylistic variant rejected in one context can inspire a tonal shift in another. In this sense, the creative process is not a linear progression from idea to output, but a recursive act of accumulation and recomposition — one that depends critically on the availability of prior material as active creative resources.

This invisibility is a problem not only for practitioners but for the creativity research community. Creative activity traces have proven to be a powerful lens for understanding how people ideate, iterate, and make sense of emerging work [6]. However, existing trace analysis approaches — including embedding-based linkographs [8, 14], clustering-based design space modeling [1], and time-series visualizations of co-creation trajectories — have largely been developed for single-modality artifacts, such as code, text drafts, or ordered sequences of design actions [13]. What counts as a trace when creativity unfolds across modalities? How do we model creative activity that is distributed, relational, and cross-representational rather than linear and modality-specific? These questions do not yet have answers. We argue that the rise of multimodal AI-assisted creativity demands a corresponding expansion of the creative activity trace (CAT) analytical toolkit — one that can represent, compare, and interpret traces that span tightly coupled modalities.

## 2 Example Project

To better ground our argument in a concrete system, we present Vistoria [7]<sup>1</sup>, a multimodal co-editing system for fictional story writing that externalizes the creative process as a persistent, recomposable record of image–text artifacts. Imagine a system that treats both textual writing and visual manipulation as part of the same paradigm of creative story-writing. Configuring images of a story’s flow is just as powerful and equivalent as Writing the story in text editing form.

<sup>1</sup>Demo video: <https://www.youtube.com/watch?v=JWaj2VJG5IU>.

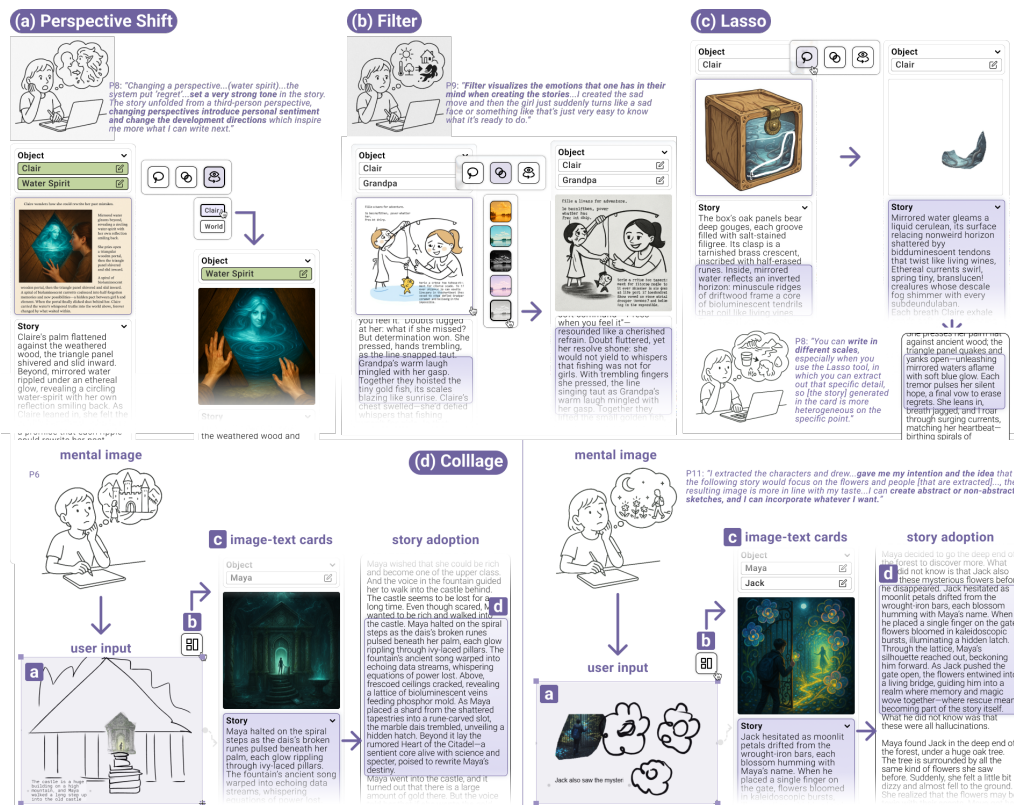


Fig. 2. How the Vistoria system’s multimodal text–image co-editing features are designed to externalize and trace the creative process, enabling creators to reuse existing artifacts to generate new ones, while integrating these artifacts into the final creation.

## 2.1 Vistoria: a multimodal system to support text-image co-editing

Vistoria supports multimodal trace preservation and reuse through two mechanisms. First, it transforms ephemeral multimodal inputs – sketches, textual annotations, and reference images – into persistent *image–text cards*, each pairing a generated visual with a corresponding narrative segment. These cards are not discarded after use; they accumulate on the canvas as a living record of the creative process, remaining available for future recombination (Figure 1). Second, Vistoria introduces *instrumental operations* – Lasso, Collage, Perspective Shift, and Filter – that apply uniformly across modalities and make simultaneous changes to text and images at the same time (Figure 2). A Lasso selection extracts regions from an image or fragments of text to generate a new, focused card; Collage merges elements across cards to construct new scenes; Perspective Shift simultaneously reframes the visual viewpoint and regenerates the narrative voice; Filter jointly adjusts image style and prose tone. Each operation produces a new artifact while preserving the source materials, creating a branching, recomposable trace of multimodal creative decisions.

## 2.2 Related Findings

We present findings from a user study with 12 fiction writers investigating how presenting creative traces multimodally and enabling the reuse of creative artifacts shape creators’ perceptions and creation behaviors.

*Cross-Modal Interplay as a Driver of Richer Narrative Detail.* Participants repeatedly oscillated between text – valued for its openness and imaginative potential – and images, which grounded abstract ideas in concrete visual form. This interplay depended critically on both modalities remaining simultaneously visible as persistent traces: when text and images co-existed on the canvas, participants could reference images while writing, directly adopt or rewrite system-generated text, and allow inspiration to emerge from juxtaposition. AI-generated images further functioned not merely as illustrations but as provocations – their randomness pushed narrative development in unanticipated directions. These benefits depend on structurally linked multimodal content, sustained recording of creative artifacts across modalities, and the capacity for ongoing cross-modal comparison.

*Visible, Operable Traces Support Agency and Ownership.* In the text-only baseline, participants occupied a largely passive role – editing content they did not initiate. With Vistoria, the persistence of multimodal traces on the canvas enabled a fundamental role shift: participants could see their creative history, select from it, and transform it, moving from passive editors to active curators of their own accumulated materials. This produced significantly higher expressiveness, immersion, and collaboration (all  $p < .05$ ), alongside a strong sense of ownership – one participant described the final story as emerging entirely from "manipulating and combining self-generated ideas." Ownership, here, is a trace effect: it arises from the ability to see and operationally control the accumulated record of one's own creative decisions.

### 3 Discussion and Future Work

*Trace-Aware Personalization* Multimodal traces, if appropriately structured, could enable systems to move beyond generic, model-driven suggestions toward support that is genuinely personalized – grounded in who the creator is and how they have worked before. Rather than responding to each prompt in isolation, a trace-aware system could reason over a creator's accumulated history: recognizing which visual styles were consistently rejected, identifying which character or scene combinations sparked the most generative branching, or noticing when a current direction closely echoes a previously abandoned one. At moments of creative stagnation, such a system could surface forgotten fragments as prompts for recombination; at moments of convergence, it could flag unexplored directions that the trace suggests the creator has historically found productive. The creative trace, in this vision, becomes not merely a record of what was made, but an active resource for shaping what comes next.

*Trace Persistence as a Design Affordance.* Version control research has shown that creative practitioners use histories not as audit trails but as palettes – sources of material for reuse, recombination, and confidence [15]. Vistoria operationalizes this insight at the level of multimodal artifacts: image-text cards persist on the canvas, remaining available as seeds for future creative moves. Future work should examine what properties a multimodal trace must have to remain generatively useful over time – and how systems can surface latent connections between spatially or temporally distant fragments without overwhelming the creator.

*Toward Longitudinal and Cross-Domain Trace Studies.* The current study examined short, bounded sessions. Real creative practice accumulates traces across weeks and projects. Longitudinal studies are needed to understand how multimodal traces age – which fragments remain generative, which become obsolete, and how creators develop strategies for navigating growing trace archives much like heuristics learning. Beyond story writing, the analytical questions raised here apply broadly: music composition, interaction design, and visual art all generate multimodal creative traces that the current framework hope to generalize to in the future.

## References

- [1] Alberto Alvarez, Jose Font, and Julian Togelius. 2022. Toward Designer Modeling Through Design Style Clustering. *IEEE Transactions on Games* 14, 4 (2022), 676–686. doi:10.1109/TG.2022.3143800
- [2] Yun Chen, Yiwei Wang, Antoni B. Chan, Jixing Li, and Ray LC. 2025. Once More with (the Right) Feeling: How Historical Fiction Writing Processes of Character Design, Plot Outline, and Context Checking Are Affected by Co-Writing with ChatGPT. In *HCI in Business, Government and Organizations*, Keng Leng Siau and Fiona Fui-Hoon Nah (Eds.). Springer Nature Switzerland, Cham, 79–101. doi:10.1007/978-3-031-92823-9\_7
- [3] John Joon Young Chung and Max Kreminski. 2024. Patchview: LLM-powered Worldbuilding with Generative Dust and Magnet Visualization. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) (UIST '24). Association for Computing Machinery, New York, NY, USA, Article 77, 19 pages. doi:10.1145/3654777.3676352
- [4] John Joon Young Chung, Melissa Roemmele, and Max Kreminski. 2025. Toyteller: AI-powered Visual Storytelling Through Toy-Playing with Character Symbols. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (CHI '25). Association for Computing Machinery, New York, NY, USA, Article 331, 23 pages. doi:10.1145/3706598.3713435
- [5] Hai Dang, Frederik Brudy, George Fitzmaurice, and Fraser Anderson. 2023. WorldSmith: Iterative and Expressive Prompting for World Building with a Generative AI. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) (UIST '23). Association for Computing Machinery, New York, NY, USA, Article 63, 17 pages. doi:10.1145/3586183.3606772
- [6] Nicholas Davis, Chih-Pin Hsiao, Kunwar Yashraj Singh, Brenda Lin, and Brian Magerko. 2017. Creative Sense-Making: Quantifying Interaction Dynamics in Co-Creation. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition* (Singapore, Singapore) (CC '17). Association for Computing Machinery, New York, NY, USA, 356–366. doi:10.1145/3059454.3059478
- [7] Kexue Fu, Jingfei Huang, Long Ling, Sumin Hong, Yihang Zuo, Ray LC, and Toby Jia jun Li. 2026. Vistoria: A Multimodal System to Support Fictional Story Writing through Instrumental Text-Image Co-Editing. arXiv:2509.13646 [cs.HC] <https://arxiv.org/abs/2509.13646>
- [8] Gabriela Goldschmidt. 2014. *Linkography: unfolding the design process*. Mit Press.
- [9] Yuanning Han, Ziyi Qiu, Jiale Cheng, and RAY LC. 2024. When Teams Embrace AI: Human Collaboration Strategies in Generative Prompting in a Creative Design Task. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (CHI '24). Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3613904.3642133
- [10] Chinmay Kulkarni, Stefania Druga, Minsuk Chang, Alex Fiannaca, Carrie Cai, and Michael Terry. 2023. A Word is Worth a Thousand Pictures: Prompts as AI Design Material. arXiv:2303.12647 [cs.HC] <https://arxiv.org/abs/2303.12647>
- [11] Long Ling, Xinyi Chen, Ruoyu Wen, Toby Jia-Jun Li, and RAY LC. 2024. Sketchar: Supporting Character Design and Illustration Prototyping Using Generative AI. *Proc. ACM Hum.-Comput. Interact.* 8, CHI PLAY (Oct. 2024), 337:1–337:28. doi:10.1145/3677102
- [12] Xiaohan Peng, Janin Koch, and Wendy E. Mackay. 2024. DesignPrompt: Using Multimodal Interaction for Design Exploration with Generative AI. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) (DIS '24). Association for Computing Machinery, New York, NY, USA, 804–818. doi:10.1145/3643834.3661588
- [13] Xinyu Shi, Yinghou Wang, Ryan Rossi, and Jian Zhao. 2025. Brickify: Enabling Expressive Design Intent Specification through Direct Manipulation on Design Tokens. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (CHI '25). Association for Computing Machinery, New York, NY, USA, Article 424, 20 pages. doi:10.1145/3706598.3714087
- [14] Amy Smith, Barrett R Anderson, Jasmine Tan Otto, Isaac Karth, Yuqian Sun, John Joon Young Chung, Melissa Roemmele, and Max Kreminski. 2025. Fuzzy Linkography: Automatic Graphical Summarization of Creative Activity Traces. In *Proceedings of the 2025 Conference on Creativity and Cognition* (CC '25). Association for Computing Machinery, New York, NY, USA, 637–650. doi:10.1145/3698061.3726915
- [15] Sarah Serman, Molly Jane Nicholas, and Eric Paulos. 2022. Towards Creative Version Control. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 336 (Nov. 2022), 25 pages. doi:10.1145/3555756
- [16] Zihan Yan, Chunxu Yang, Qihao Liang, and Xiang 'Anthony' Chen. 2023. XCreation: A Graph-based Crossmodal Generative Creativity Support Tool. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) (UIST '23). Association for Computing Machinery, New York, NY, USA, Article 48, 15 pages. doi:10.1145/3586183.3606826
- [17] Daijin Yang, Yanpeng Zhou, Zhiyuan Zhang, Toby Jia-Jun Li, and RAY LC. 2022. AI as an Active Writer: Interaction strategies with generated text in human-AI collaborative fiction writing. In *Joint Proceedings of the IUI 2022 Workshops: APEX-UI, HAI-GEN, HEALTHI, HUMANIZE, TExSS, SOCIALIZE*. CEUR-WS Team, 56–65. [https://scholars.cityu.edu.hk/en/publications/publication\(d901f5a2-0600-422f-b588-db5a59871961\).html](https://scholars.cityu.edu.hk/en/publications/publication(d901f5a2-0600-422f-b588-db5a59871961).html)