The World as a Canvas: Materializing Imagination through Prompting the Physical World

HyunA Seo
Seoul National University
Seoul, Republic of Korea
hyunas1996@gmail.com

Nels Numan
Google
Seattle, Washington, USA
nelsn@google.com

Hye-Young Jo
University of Colorado Boulder
Boulder, Colorado, USA
hye-young.jo@colorado.edu

Jaewook Lee
Paul G. Allen School of Computer Science & Engineering
University of Washington
Seattle, Washington, USA
jaewook4@cs.washington.edu

Jon E. Froehlich
Paul G. Allen School of Computer Science & Engineering
University of Washington
Seattle, Washington, USA
jonf@cs.uw.edu

Ruofei Du
Google XR
San Francisco, California, USA
ruofei@google.com

Ryo Suzuki
University of Colorado Boulder
Boulder, Colorado, USA
ryo.suzuki@colorado.edu

Youngki Lee
Computer Science and Engineering
Seoul National University
Seoul, Republic of Korea
youngkilee@snu.ac.kr

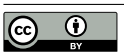
Abstract

This workshop brings together researchers and practitioners in HCI, eXtended Reality (XR), and AI to explore how generative AI can lower barriers to creating XR experiences. Generative AI can now turn rough ideas into images, videos, and apps within minutes, yet creating XR experiences remains demanding. XR creations must be grounded in physical space, previewed before they exist, steered through multimodal user inputs, co-created across heterogeneous devices, inspected for AI errors that unfold in interaction, and shared beyond a single headset. Through keynotes, posters and demos, and hands-on prototyping activities, we will turn these challenges into design considerations and a shared research agenda. We will publish the outcomes on the workshop website to seed a growing community advancing generative AI-powered XR for everyday creativity: **worldascanvas.github.io**

Keywords

Extended Reality, Generative AI, Creativity Support Tools, XR Authoring

ACM Reference Format:
HyunA Seo, Nels Numan, Hye-Young Jo, Jaewook Lee, Jon E. Froehlich, Ruofei Du, Ryo Suzuki, and Youngki Lee. 2026. The World as a Canvas: Materializing Imagination through Prompting the Physical World. In *The 39th Annual ACM Symposium on User Interface Software and Technology (UIST Adjunct '26), November 02–05, 2026, Detroit, MI, USA*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3830397.3842699



This work is licensed under a Creative Commons Attribution 4.0 International License.
UIST Adjunct '26, Detroit, MI, USA
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2855-6/26/11
https://doi.org/10.1145/3830397.3842699

1 Introduction

Humans share a deep desire to create, an act that makes us feel alive [21, 38]. Over the past decades, digital technologies have expanded the ways people create by lowering the cost of failure and speeding up iteration [9, 31]. Generative AI is making creation even more accessible. People can now turn rough ideas into various forms of media, from images to videos and even 3D-printable objects [26, 27, 32], within minutes and without mastering complex software [18, 43].

Meanwhile, as XR devices and spatial computing mature, creation is moving beyond flat screens into the real world. Users engage in embodied interaction with what they make [6, 13, 29], and their creations blend playfully into everyday environments [3, 12, 15, 30, 33]. This physical grounding makes authoring situated: virtual characters and personalized spaces are created in situ to fit the user's surroundings [4, 15, 16, 36, 37, 42], and entire environments are generated on demand and continuously reshaped as the user interacts with them [1, 5, 24, 25]. The embodied and shared nature of XR also makes authoring multimodal and collective. Remote collaborators co-create across heterogeneous devices and differing mental imagery [11, 23, 41], and creators build by sketching, speaking, and gesturing [8, 19, 28, 34, 40]. Across these practices, generative AI acts not just as a content generator but as a partner throughout the authoring process (e.g., supporting ideation, mediating collaboration, and adapting to user context) [7, 11, 14, 20].

Yet many challenges still remain in authoring XR experiences [22, 35]. XR authoring requires creators to coordinate spatial, temporal, and interactive elements at once (e.g., position, motion, and behavior), and even a minor jarring element can break immersion. Since XR content blends with physical space, creators must mentally simulate how a creation will look and behave among physical objects. Moreover, with multiple generative models taking on different roles

in a single pipeline, an early error propagates downstream and becomes hard to correct. Though XR captures rich signals of intent (e.g., gesture, gaze, and facial expression) [10, 17], how creators can effectively express intent through these signals and how AI should interpret them remain open questions. Even after a creation is complete, sharing it, a key reward of creating, remains cumbersome in XR, and once remixed across media, its provenance and ownership become hard to trace.

Our workshop aims to identify shared interests and emerging challenges at the intersection of generative AI in XR authoring, and to build a collective vision for how this integration should move forward. UIST is the ideal venue, since it has built many of the interaction techniques, toolkits, and devices that XR creation stands on, and gathers researchers in HCI, XR, and generative AI alongside practitioners in education, industrial design, and remote collaboration. We will consolidate the discussions into a shared research agenda, published on the workshop website as a starting point for continued community discussion and collaboration.

2 Scope

We will explore the following topics during the workshop:

In-situ Creation within Physical Space. In-situ XR creation must account for the user's physical surroundings, from coarse properties (e.g., lighting, scale, and object categories) down to finer ones such as material and weight. How can generative models produce content and behaviors that stay physically plausible? What limits how finely creators can author such physical interactions: sensing, representations, or generative pipelines?

Previsualization and Mental Imagery. At the outset, creators picture the result and mentally project it onto their space, yet people differ greatly in how vividly they can do this. How much should tools previsualize to support imagination without constraining or anchoring creators, and in what form? Beyond showing possible outcomes, how can previews help creators develop and explore their own mental imagery?

Expressing and Interpreting Creative Intent. In XR, creators speak with more than words; gesture, gaze, facial expression, and movement all carry meaning. What creative intent do people convey through these signals, and how should generative AI interpret it? How can creators and AI negotiate ambiguous intent and refine a creation through iteration?

Co-Creation across Hardware Heterogeneity. Collaborators rarely hold identical devices; one may wear a headset while another joins from a tablet or phone. How can generative AI adapt content and interactions across devices to maintain coherent co-creation when input, display, and compute vary widely? What roles suit each device, and how should work be handed off between them?

Evaluating Generated Content. Generative AI inevitably makes mistakes, and in XR such errors are especially difficult to inspect and repair, as they may surface only through interaction, across viewpoints, and over time. How can interfaces help creators detect such errors and evaluate the quality of generated content? At which stage of a generative pipeline should errors be detected and corrected, and how can this process be made transparent to creators?

Cross-Media Creation, Transformation, and Provenance. Generative AI also lets content cross media boundaries: existing images, videos, stories, and objects can become interactive XR experiences, while XR creations can be translated into videos or physical artifacts. What should be preserved, transformed, or newly generated in such translations? And as content undergoes repeated transformation and remixing, how can its provenance, attribution, and ownership remain traceable?

3 Workshop Plan

Our workshop is a full-day, in-person event for 25–35 participants, designed around hands-on interaction and community building. A dedicated website will announce the theme, application process, and schedule, and host the outcomes after the event.

3.1 Workshop Activities

Opening and Speed Introductions (45 minutes). Organizers welcome participants and outline the day's goals. Participants introduce themselves one-on-one for two minutes and then rotate to a new partner, guided by icebreaker prompts and by self-introduction slides collected in advance.

Keynotes (90 minutes). We invite two speakers, one from academia and one from industry, to share their perspectives on generative AI for XR authoring. Each gives a 30-minute talk followed by a 15-minute Q&A. The aim is that participants enter the day's activities with a shared picture of what these systems can already do and where they still fail.

Poster and Demo Session (75 minutes). Posters and demos selected through the call for participation are set up as stations around the room, where participants try the systems firsthand and talk with the presenters. Each participant carries a short printed booklet, a passport, to note the challenges they encounter and the ideas each system sparks. These notes become the material participants draw on in the afternoon.

Group Discussion (30 minutes). Organizers place one challenge bucket drawn from § 2 at each table, and participants join the table whose topic they find most compelling. Drawing on the keynotes, posters, demos, and their passport notes, each team narrows its bucket into one concrete problem to address in the prototyping session that follows.

Low-Fidelity Prototyping (45 minutes). Each team develops a low-fidelity prototype that includes the interfaces, tools, and experiences needed to address its chosen problem. For rapid prototyping, we provide disposable clear face shields as stand-ins for optical see-through displays, inspired by prior work on low-cost XR study setups [2, 39]. Participants draw visual cues directly on the shields and use sketches and sticky notes to represent other modalities (e.g., haptics and audio). Through this process, participants surface design considerations for the shared research agenda.

Closing (45 minutes). Each team gives a five-minute presentation of its prototype. Organizers then summarize the day, outline the post-workshop plan, and open a community channel to continue the conversation.

3.2 Post-Workshop Activities

After the event, we will publish a report on the workshop website summarizing the discussed challenges, prototypes, and design considerations, illustrated with photographs. This report will serve as a basis for further discussion, feedback, and collaboration in the community.

4 Call for Participation

We invite researchers, designers, and practitioners working on XR authoring, generative AI, and creativity support, as well as those applying these technologies in domains such as education, industrial design, and remote collaboration. Posters and interactive demos, whether polished or early-stage, are especially welcome. Demo presenters should briefly describe their setup and equipment needs so we can coordinate with the venue.

Acknowledgments

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. RS-2024-00463802), JST PRESTO Grant Number JPMJPR23I5, and NSF CAREER Award IIS-2543251.

References

[1] Naisha Agarwal, Judith Amores, and Andrew D. Wilson. 2026. Storycaster: An AI System for Immersive Room-based Storytelling. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. doi:10.1145/3772318.3791305

[2] Karan Ahuja, Chris Harrison, Mayank Goel, and Robert Xiao. 2019. MeCap: Whole-Body Digitization for Low-Cost VR/AR Headsets. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology (UIST '19)*. doi:10.1145/3332165.3347889

[3] Ronald T. Azuma. 1997. A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments* 6, 4 (1997), 355–385. doi:10.1162/pres.1997.6.4.355

[4] Fernanda De La Torre, Cathy Mengying Fang, Han Huang, Andrzej Banburski-Fahey, Judith Amores Fernandez, and Jaron Lanier. 2024. LLMR: Real-time Prompting of Interactive Worlds using Large Language Models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. doi:10.1145/3613904.3642579

[5] Fernanda M. De La Torre, Javier Hernandez, Andrew D. Wilson, and Judith Amores Fernandez. 2025. Sonora: Human-AI Co-Creation of 3D Audio Worlds and its Impact on Anxiety and Cognitive Load. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. doi:10.1145/3706598.3713316

[6] Paul Dourish. 2001. *Where the Action Is: The Foundations of Embodied Interaction*. MIT Press, Cambridge, MA.

[7] Ruofei Du, Benjamin Hersh, David Li, Xun Qian, Nels Numan, Zhongyi Zhou, Yanhe Chen, Xingyue Chen, Jiahao Ren, Robert Timothy Bettridge, Faraz Faruqi, Xiang Anthony Chen, Steve Toh, and David Kim. 2026. Vibe Coding XR: Accelerating AI + XR Prototyping with XR Blocks and Gemini. arXiv:2603.24591 [cs.HC]

[8] Runlin Duan, Yuzhao Chen, Yichen Hu, and Karthik Ramani. 2026. JustShape: Exploring Co-Speech Gestures for Multimodal LLM-Powered 3D Parametric Modeling. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. doi:10.1145/3772318.3790641

[9] Jonas Frich, Lindsay MacDonald Vermeulen, Christian Remy, Michael Mose Biskjaer, and Peter Dalsgaard. 2019. Mapping the Landscape of Creativity Support Tools in HCI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Article 389, 18 pages. doi:10.1145/3290605.3300619

[10] Christian Holz and Andrew D. Wilson. 2011. Data Miming: Inferring Spatial Object Descriptions from Human Gesture. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. 811–820. doi:10.1145/1978942.1979060

[11] Erzhen Hu, Mingyi Li, Jungtaek Hong, Xun Qian, Alex Olwal, David Kim, Seongkook Heo, and Ruofei Du. 2025. Thing2Reality: Enabling Spontaneous Creation of 3D Objects from 2D Content using Generative AI in XR Meetings. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25)*. doi:10.1145/3746059.3747621

[12] Keiichi Ihara, DaeHo Lee, Manato Abe, Hye-Young Jo, and Ryo Suzuki. 2026. CinemaWorld: Generative Augmented Reality with LLMs and 3D Scene Generation for Movie Augmentation. arXiv:2603.08060 [cs.HC]

[13] Scott R. Klemmer, Björn Hartmann, and Leila Takayama. 2006. How Bodies Matter: Five Themes for Interaction Design. In *Proceedings of the 6th Conference on Designing Interactive Systems (DIS '06)*. 140–149. doi:10.1145/1142405.1142429

[14] Geonsun Lee, Min Xia, Nels Numan, Xun Qian, David Li, Yanhe Chen, Achin Kulshrestha, Ishan Chatterjee, Yinda Zhang, Dinesh Manocha, David Kim, and Ruofei Du. 2025. Sensible Agent: A Framework for Unobtrusive Interaction with Proactive AR Agents. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25)*. doi:10.1145/3746059.3747748

[15] Jaewook Lee, Filippo Aleotti, Diego Mazala, Guillermo Garcia-Hernando, Sara Vicente, Oliver James Johnston, Isabel Kraus-Liang, Jakub Powierza, Donghoon Shin, Jon E. Froehlich, Gabriel Brostow, and Jessica Van Brummelen. 2025. ImaginateAR: AI-Assisted In-Situ Authoring in Augmented Reality. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25)*. doi:10.1145/3746059.3747635

[16] Jingyu Lee, Dongho Han, Jinmyeong Kim, and Youngki Lee. 2026. SpatialBlend: Relational Reconstruction for Cross-Space Presence in Heterogeneous Spaces. In *Proceedings of the 39th Annual ACM Symposium on User Interface Software and Technology (UIST '26)*. doi:10.1145/3830398.3830690

[17] Jaewook Lee, Jun Wang, Elizabeth Brown, Liam Chu, Sebastian S. Rodriguez, and Jon E. Froehlich. 2024. GazePointAR: A Context-Aware Multimodal Voice Assistant for Pronoun Disambiguation in Wearable Augmented Reality. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. doi:10.1145/3613904.3642230

[18] David Li, Nels Numan, Xun Qian, Yanhe Chen, Zhongyi Zhou, Evgenii Alekseev, Geonsun Lee, Alex Cooper, Min Xia, Scott Chung, Jeremy Nelson, Xiuxiu Yuan, Jolica Dias, Tim Bettridge, Benjamin Hersh, Michelle Huynh, Konrad Piascik, Ricardo Cabello, David Kim, and Ruofei Du. 2025. XR Blocks: Accelerating Human-centered AI + XR Innovation. arXiv:2509.25504 [cs.HC]

[19] Jian Liao, Adnan Karim, Shivesh Singh Jadon, Rubaiat Habib Kazi, and Ryo Suzuki. 2022. RealityTalk: Real-Time Speech-Driven Augmented Presentation for AR Live Storytelling. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology (UIST '22)*. doi:10.1145/3526113.3545702

[20] Nicolai Marquardt, Asta Roseway, Hugo Romat, Payod Panda, Michel Pahud, Gonzalo Ramos, Steven Drucker, Andrew D. Wilson, Ken Hinckley, and Nathalie Henry Riche. 2025. ImaginationVellum: Generative-AI Ideation Canvas with Spatial Prompts, Generative Strokes, and Ideation History. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25)*. doi:10.1145/3746059.3747631

[21] Abraham H. Maslow. 1943. A Theory of Human Motivation. *Psychological Review* 50, 4 (1943), 370–396. doi:10.1037/h0054346

[22] Michael Nebeling and Maximilian Speicher. 2018. The Trouble with Augmented Reality/Virtual Reality Authoring Tools. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 333–337.

[23] Nels Numan, Gabriel Brostow, Suhyun Park, Simon Julier, Anthony Steed, and Jessica Van Brummelen. 2025. CoCreatAR: Enhancing Authoring of Outdoor Augmented Reality Experiences Through Asymmetric Collaboration. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. doi:10.1145/3706598.3714274

[24] Nels Numan, Shwetha Rajaram, Balasaravanan Thoravi Kumaravel, Nicolai Marquardt, and Andrew D. Wilson. 2024. SpaceBlender: Creating Context-Rich Collaborative Spaces Through Generative 3D Scene Blending. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24)*. doi:10.1145/3654777.3676361

[25] Shwetha Rajaram, Nels Numan, Balasaravanan Thoravi Kumaravel, Nicolai Marquardt, and Andrew D. Wilson. 2024. BlendScape: Enabling End-User Customization of Video-Conferencing Environments through Generative AI. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24)*. doi:10.1145/3654777.3676326

[26] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical Text-Conditional Image Generation with CLIP Latents. arXiv:2204.06125 [cs.CV]

[27] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 10684–10695.

[28] Karl Toby Rosenberg, Rubaiat Habib Kazi, Li-Yi Wei, Haijun Xia, and Ken Perlin. 2024. DrawTalking: Building Interactive Worlds by Sketching and Speaking. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24)*. doi:10.1145/3654777.3676334

[29] HyunA Seo, Youngki Lee, Rajesh Balan, and Thivya Kandappu. 2025. EmoShortcuts: Emotionally Expressive Body Augmentation for Social Mixed Reality Avatars. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*. 1–16.

[30] HyunA Seo, Juheon Yi, Rajesh Balan, and Youngki Lee. 2024. Gradualreality: Enhancing physical object interaction in virtual reality via interaction state-aware blending. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–14.

[31] Ben Shneiderman. 2007. Creativity support tools: accelerating discovery and innovation. *Commun. ACM* 50, 12 (2007), 20–32. doi:10.1145/1323688.1323689

[32] Uriel Singer, Adam Polyak, Thomas Hayes, Xi Yin, Jie An, Songyang Zhang, Qiyuan Hu, Harry Yang, Oron Ashual, Oran Gafni, Devi Parikh, Sonal Gupta, and Yaniv Taigman. 2023. Make-A-Video: Text-to-Video Generation without Text-Video Data. In *The Eleventh International Conference on Learning Representations (ICLR)*.

[33] Mel Slater. 2009. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, 1535 (2009), 3549–3557. doi:10.1098/rstb.2009.0138

[34] Ryo Suzuki, Rubaiat Habib Kazi, Li-Yi Wei, Stephen DiVerdi, Wilmot Li, and Daniel Leithinger. 2020. RealitySketch: Embedding Responsive Graphics and Visualizations in AR through Dynamic Sketching. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (UIST '20)*. doi:10.1145/3379337.3415892

[35] Yiliu Tang, Mengke Wu, Jason Situ, and Yun Huang. 2026. Empowered XR through Generative AI: Balancing Superpowers and Risks. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. doi:10.1145/3772318.3791307

[36] Cyrus Vachha, Yixiao Kang, Zach Dive, Ashwat Chidambaram, Anik Gupta, Eunice Jun, and Björn Hartmann. 2025. Dreamcrafter: Immersive Editing of 3D Radiance Fields Through Flexible, Generative Inputs and Outputs. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. doi:10.1145/3706598.3714312

[37] Suibi Che-Chuan Weng, Shih-Yu Ma, Sawyer Reinig, Pritalee Kadam, Ada Yi Zhao, Amy Banić, Ryo Suzuki, and Ellen Yi-Luen Do. 2026. Editing Reality: Designing In-Situ Co-Creation with Generative AI in Mixed Reality. In *Proceedings of the 2026 ACM Designing Interactive Systems Conference (DIS '26)*. doi:10.1145/3800645.3813087

[38] Donald W. Winnicott. 1971. *Playing and Reality*. Tavistock Publications, London.

[39] Julia Woodward, Feben Alemu, Natalia E. López Adames, Lisa Anthony, Jason C. Yip, and Jaime Ruiz. 2022. "It Would Be Cool to Get Stampeded by Dinosaurs": Analyzing Children's Conceptual Model of AR Headsets Through Co-Design. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. doi:10.1145/3491102.3501979

[40] Yichen Andy Yu, Wanru Li, Qiaoran Wang, Jymon Ross, Gavin Johnson, Mandy Lui, and Qiao Jin. 2026. SpatialPrompt: XR-Based Spatial Intent Expression as Executable Constraints for AI Generative 3D Design. In *Companion Publication of the 2026 ACM Designing Interactive Systems Conference (DIS '26 Companion)*. doi:10.1145/3802974.3809475

[41] Adam Zeman, Michaela Dewar, and Sergio Della Sala. 2015. Lives without imagery – Congenital aphantasia. *Cortex* 73 (2015), 378–380. doi:10.1016/j.cortex.2015.05.019

[42] Lei Zhang, Jin Pan, Jacob Gettig, Steve Oney, and Anhong Guo. 2024. VRCopilot: Authoring 3D Layouts with Generative AI Models in VR. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24)*. doi:10.1145/3654777.3676451

[43] Zhongyi Zhou, Jing Jin, Vrushank Phadnis, Xiuxiu Yuan, Jun Jiang, Xun Qian, Kristen Wright, Mark Sherwood, Jason Mayes, Jingtao Zhou, Yiyi Huang, Zheng Xu, Yinda Zhang, Johnny Lee, Alex Olwal, David Kim, Ram Iyengar, Na Li, and Ruofei Du. 2025. InstructPipe: Generating Visual Blocks Pipelines with Human Instructions and LLMs. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. doi:10.1145/3706598.3713905