XR and AI: AI-Enabled Virtual, Augmented, and Mixed Reality

Ryo Suzuki
University of Calgary
Calgary, Canada
ryo.suzuki@ucalgary.ca

Mar Gonzalez-Franco
Google
Seattle, USA
margon@google.com

Misha Sra
UC Santa Barbara
Santa Barbara, USA
sra@cs.ucsb.edu

David Lindlbauer
Carnegie Mellon University
Pittsburgh, USA
davidlindlbauer@cmu.edu

ABSTRACT

This workshop aims to unite experts and practitioners in XR and AI to envision the future of AI-enabled virtual, augmented, and mixed reality experiences. Our expansive discussion includes a variety of key topics: Generative XR, Large Language Models (LLMs) for XR, Adaptive and Context-Aware XR, Explainable AI for XR, and harnessing AI to enhance and prototype XR experiences. We aim to identify the opportunities and challenges of how recent advances of AI could bring new XR experiences, which cannot be done before, with a keen focus on the seamless blending of our digital and physical worlds.

CCS CONCEPTS

• **Human-centered computing → Mixed / augmented reality**.

KEYWORDS

Augmented Reality; Mixed Reality; Virtual Reality; Generative AI; Large Language Models; Computer Vision; Machine Learning; Human-AI Interaction

1 INTRODUCTION

The advent of generative AI, large language models, and new foundation models in computer vision have fostered numerous technological innovations in AI research and its applications. However, interface design and interaction with AI technologies is predominantly confined to graphical user interfaces (GUIs) on computers or smartphones. We believe that the next step in Human-AI Interaction should be oriented towards "real-world" AI interfaces, where the AI technology naturally and seamlessly integrates into our everyday life by blending bits and atoms. By blending AI with extended reality (XR), we can enable human-AI interactions that are more physical, spatial, and context-aware, beyond the limitations of current GUI-based interfaces. However, despite immense potential, the integration of XR and AI has barely been investigated yet.We believe it is a critical future direction in HCI research and there is need to build, connect, and grow a community around these shared domains to collectively push the boundary of AI-XR research.

The goal of this XR and AI workshop (https://xr-and-ai.github.io) is to bring XR and AI researchers together to gather insights, facilitate collaboration, and shape the future of real-world-oriented human-AI interaction research. In particular, we aim to develop a strategic roadmap for future directions, which will, for instance, define key research questions, explore new application domains, and identify the challenges and opportunities in AI-XR research. But foremost, this workshop will bring together a research community that will collaborate on this focus in the future. Finally, we plan to establish the groundwork for a reference paper to inform future research and to initiate a proposal for follow-up activities, like workshops and seminars, to continue growing the momentum towards these research efforts.

2 PERSPECTIVES OF INTEREST

This workshop welcomes HCI researchers and practitioners in XR (VR/AR/MR), AI, machine learning, and computational interaction domains to share diverse perspectives and expertise. There are several domains that are not fully explored yet in the literature of XR and AI [5]. We plan to discuss the topics that include but are not limited to the following areas:

Generative XR. We explore the integration of generative AI for XR applications. For instance, by leveraging the recent advances in image generation such as GAN and Stable Diffusion, there is a vast potential to transform various aspects of XR interactions, such as content creation [9] 3D and avatar animation [1], and immersive environment design [3]. Generative XR is not limited to image generation, but can be integrated with generative AI for text, motion, video, and 3D object as well. In particular, we are interested in understanding how generative AI can leverage the unique physical and spatial aspects of XR applications, such as 3D scene understanding and spatial and tangible interactions. What are the challenges and possibilities in such integration?

LLMs for XR. LLMs for XR focus specifically on the exploration of large language models to augment mixed reality experiences.

Languages are an important part of XR experiences, including in-person communication [7, 10], physical document reading [6], and language learning. How can we leverage the LLMs to naturally augment such interactions, instead of explicitly typing questions on the screen? Beyond text-based interaction, what kind of new applications are possible by integrating the recent advances in multi-modal LLMs to XR environment?

Adaptive and Context-Aware XR*.* Adaptive and context-aware XR focuses on the development of XR interfaces that seamlessly blend into the user's everyday environment [4, 8]. Current research investigates how computational methods from optimization and machine learning can be leveraged to create XR interfaces that understand and adapt to the context of users and their environment.

Explainable AI for XR*.* Explainable AI for XR addresses the growing need for user-friendly interfaces that help users understand, customize, and interact with mixed reality applications [14]. It involves creating prototyping tools aimed at making XR experiences more accessible and understandable. The challenge is how to enable non-technical users to author adaptive behaviors and how conventional interface prototyping methods can be adapted for the development of adaptive XR.

Prototyping XR with AI*.* As XR interfaces become more accessible to end users, lowering the barrier to design and develop XR applications becomes increasingly critical [12]. Current tools primarily focus on developing applications and experiences for static contexts. Therefore, a major challenge lies in enabling non-technical users to author adaptive behaviors [2]. This is particularly crucial in the development of tangible interfaces, which could unlock significant potential for immersive XR experiences [11, 13].

Other XR x AI Topics*.* We also welcome any possible XR and AI themes, such as intelligent interaction techniques, AI-enabled accessibility in XR, new evaluation methodologies, and real-world-oriented human-AI interaction.

3 WORKSHOP FORMAT AND PLANS

Introductions and Lightning Talks*.* The organizers will kick off the workshop with introductions and each participant's lighting talks highlighting their related research experience. To start community building, we will also facilitate pre-workshop engagement to cultivate and propose an initial set of themes. We also invite a couple of speakers who will deliver presentations to provoke further discussions.

Theme Organization and Discussion*.* Participants will collectively extrapolate themes from the presentations and shared readings. The discussed themes will be combined with the initial set of topics developed prior to the workshop to serve as guiding directions for the first round of discussions. Once a preliminary list of discussion topics has been defined, each topic will be assigned a 'table.' During the session, participants will rotate between tables to engage in focused discussions of topics of their choice. One participant at each table will be designated as the discussion mediator, whose responsibilities will involve guiding and documenting the discussion. The second session on theme organization will begin

with lightning talks by the discussion mediators summarizing earlier conversations. Participants will then collectively revisit the discussion topics, reorganizing accordingly based on the results of the first session and expert perspectives. With the refined list of topics, the remainder of the session will follow the same format as the first round of discussions.

Defining Future Challenges and Research Directions*.* In this final discussion session, we will begin by regrouping and refining the list of discussion topics based on the results of the discussion sessions. The final workshop session will focus on summarizing the workshop findings and defining next steps. First, the organizers will provide a recap of the workshop activities, including the defined themes from the pre-workshop activities and a summary of the morning and afternoon discussion outcomes. The floor will then be opened for participants to contribute their reflections on the workshop discussion. A final discussion will be held around potential future directions, such as follow-up workshops and publications.

4 ORGANIZERS

Ryo Suzuki is an Assistant Professor in the Department of Computer Science at the University of Calgary. His research interests focus on augmented reality and tangible user interfaces. His research aims to transform everyday environments into dynamic physical and spatial media with the power of AI and AR. His website is https://ryosuzuki.org/.

Mar Gonzalez-Franco is a Research Manager at Google, where she leads Blended Interaction Research and Devices (BIRD) group. She has pioneering research in the field of extended reality in various domains such as human perception of virtual avatar, haptic interfaces for VR, multi-device interactions for XR, and ML-enabled novel experiences. Her website is https://margonzalezfranco.github.io/.

Misha Sra is an Assistant Professor in the Department of Computer Science at the University of California, Santa Barbara, where she leads the Human-AI Integration Lab. Her research interest lies at the intersection of XR and Human-AI Interaction, focusing specifically on interfaces that augment human physical capabilities, AI-infused XR interfaces, and human-AI creative collaboration. Her website is https://sites.cs.ucsb.edu/~sra/.

David Lindlbauer is an Assistant Professor at the Human-Computer Interaction Institute at Carnegie Mellon University where he leads the Augmented Perception Lab. His research focuses on creating and studying enabling technologies and computational approaches for adaptive user interfaces to increase the usability of AR and VR interfaces, with applications in casual interaction, productivity, health and robotics. His website is https://davidlindlbauer.com/

REFERENCES

[1] Karan Ahuja, Eyal Ofek, Mar Gonzalez-Franco, Christian Holz, and Andrew D Wilson. 2021. Coolmoves: User motion accentuation in virtual reality. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2 (2021), 1–23.
[2] Ruofei Du, Na Li, Jing Jin, Michelle Carney, Scott Miles, Maria Kleiner, Xiuxiu Yuan, Yinda Zhang, Anuva Kulkarni, Xingyu Liu, et al. 2023. Rapsai: Accelerating Machine Learning Prototyping of Multimedia Applications through Visual Programming. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–23.

[3] Ran Gal, Lior Shapira, Eyal Ofek, and Pushmeet Kohli. 2014. FLARE: Fast layout for augmented reality applications. In *2014 IEEE international symposium on mixed and augmented reality (ISMAR)*. IEEE, 207–212.

[4] Jens Grubert, Tobias Langlotz, Stefanie Zollmann, and Holger Regenbrecht. 2017. Towards Pervasive Augmented Reality: Context-Awareness in Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics* 23, 6 (2017), 1706–1724. https://doi.org/10.1109/TVCG.2016.2543720

[5] Teresa Hirzle, Florian Müller, Fiona Draxler, Martin Schmitz, Pascal Knierim, and Kasper Hornbæk. 2023. When XR and AI Meet-A Scoping Review on Extended Reality and Artificial Intelligence. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–45.

[6] Zhen Li, Michelle Annett, Ken Hinckley, Karan Singh, and Daniel Wigdor. 2019. Holodoc: Enabling mixed reality workspaces that harness physical and digital content. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–14.

[7] 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*. 1–12.

[8] David Lindlbauer, Anna Maria Feit, and Otmar Hilliges. 2019. Context-aware online adaptation of mixed reality interfaces. In *Proceedings of the 32nd annual ACM symposium on user interface software and technology*. 147–160.

[9] Vivian Liu, Jo Vermeulen, George Fitzmaurice, and Justin Matejka. 2022. 3DALL-E: Integrating text-to-image AI in 3D design workflows. *arXiv preprint arXiv:2210.11603* (2022).

[10] Xingyu" Bruce" Liu, Vladimir Kirilyuk, Xiuxiu Yuan, Alex Olwal, Peggy Chi, Xiang" Anthony" Chen, and Ruofei Du. 2023. Visual Captions: Augmenting Verbal Communication With On-the-Fly Visuals. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–20.

[11] Kyzyl Monteiro, Ritik Vatsal, Neil Chulpongsatorn, Aman Parnami, and Ryo Suzuki. 2023. Teachable Reality: Prototyping Tangible Augmented Reality with Everyday Objects by Leveraging Interactive Machine Teaching. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.

[12] Michael Nebeling. 2022. XR tools and where they are taking us: characterizing the evolving research on augmented, virtual, and mixed reality prototyping and development tools. *XRDS: Crossroads, The ACM Magazine for Students* 29, 1 (2022), 32–38.

[13] 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*. 166–181.

[14] Xuhai Xu, Anna Yu, Tanya R Jonker, Kashyap Todi, Feiyu Lu, Xun Qian, João Marcelo Evangelista Belo, Tianyi Wang, Michelle Li, Aran Mun, et al. 2023. XAIR: A Framework of Explainable AI in Augmented Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–30.