Research Statement - Ryo Suzuki

Augment Human Thought and Creativity with the Power of AR and AI

The advent of generative AI and large language models have fostered numerous technological innovations in AI research. However, today's AI interfaces are predominantly limited to 2D computer screens, leaving current AI systems unable to engage with and respond to our physical environment. My research goal is to shift this paradigm towards a **real-world-oriented human-AI interaction** by integrating augmented reality (AR) and AI interfaces. Through this, I aim to seamlessly weave AI into the fabric of our everyday lives, moving beyond the constraints of current screen-based interactions. I believe this integration of AR and AI will significantly amplify our cognitive and creative abilities by enabling us to think and collaborate with **real objects** in the **real world**, rather than just virtual objects on screens. This approach empowers us to utilize the entire range of human capabilities, tapping into the rich, holistic modes of thinking that humans have developed since prehistoric times.

Towards this goal, my research has developed AI-enabled AR interfaces [8] by pushing the boundaries of AR x AI research in the following key areas: 1) *AI-mediated augmented communication:* enhancing human communication like conversation and presentation with the power of AR and AI, 2) *AI-powered content creation for AR:* leveraging machine learning-driven authoring for dynamic and interactive AR content creation, 3) *ML-enabled tangible AR interaction:* integrating AR with interactive machine teaching to make tangible interactions ubiquitous and adaptive, and 4) *Blended virtual-physical interaction with AR and robotics:* further blending virtual and physical worlds through the integration of visual augmentation and physical actuation. Building upon these themes, I also outline my future research directions that focus on incorporating generative AI, large language models, and explainable AI into AR interfaces.

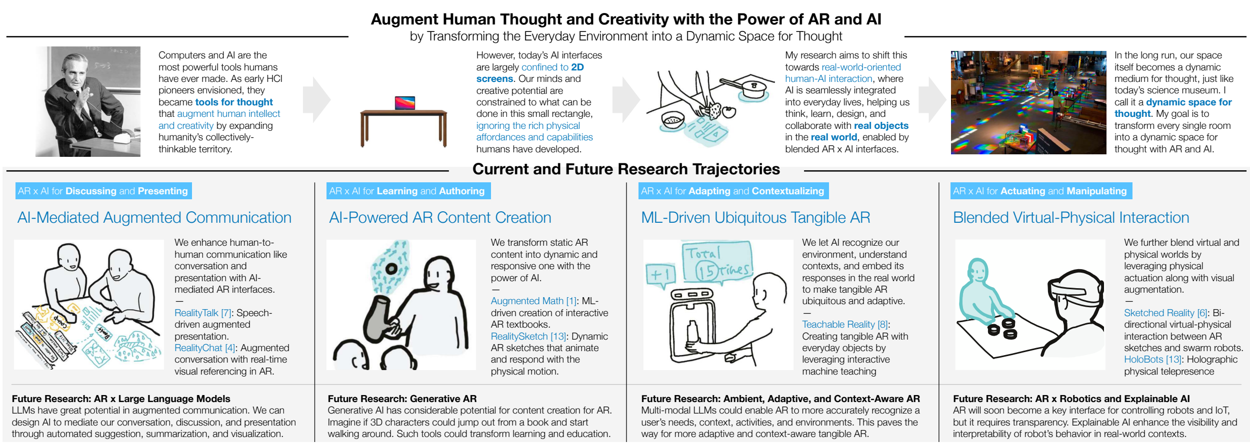

Figure 1: Overview of my vision, motivation, and current research trajectories.

1. AI-Mediated Augmented Communication

The integration of AR and AI has tremendous potential to enhance our everyday communication. Historically, spoken languages have remained fundamental methods of human-to-human communication, yet this medium of communication has not changed over thousands of years. In this research, I aim to augment spoken languages through embedded visual information and AI-mediated interactions (Figure 2).

RealityTalk [6] demonstrates this AI-mediated augmented communication within a presentation context (Figure 3). In this project, we explore a new interaction technique with transcribed spoken language, allowing presenters not just to visualize their spoken words but to interact with them in real time. This includes using embedded visuals and animations to make the storytelling more engaging and dynamic. By using this system, users can craft live, augmented presentations enriched with speech-driven interactions. Similarly, RealityChat [4] explores the concept of augmented conversation, which aims to support

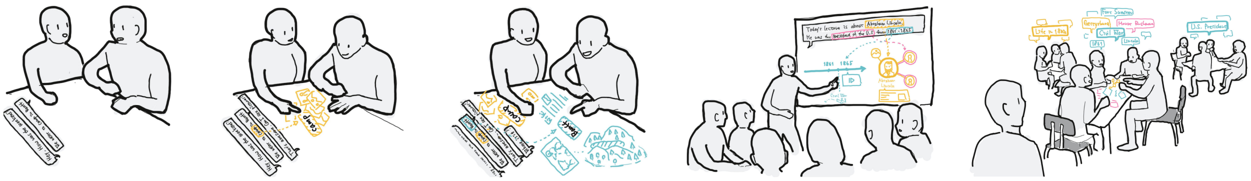

Figure 2: By leveraging AI-mediated AR interfaces, we can augment human-to-human communication, which includes conversation, presentation, discussion, and collaboration.

co-located in-person conversations via embedded on-the-fly visual referencing (Figure 3). Our AR interface provides relevant visual references in real-time, based on keywords extracted from the spoken conversation. By embedding these visual references in AR around the conversation partner, augmented conversation can reduce distraction and friction, allowing users to maintain eye contact and supporting more natural social interactions. Such interaction capability will enhance in-person social interaction by augmenting the human's natural conversation through AR.

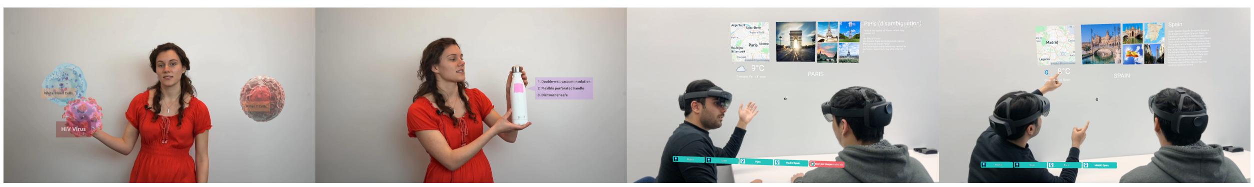

Figure 3: RealityTalk [6] (Left) and RealityChat [4] (Right) are examples of AI-mediated augmented presentation and conversation

2. AI-Enabled Dynamic and Interactive Content Creation for AR

Augmented reality has tremendous potential for educational purposes. AR-based textbooks, for instance, enable children to engage with learning material in an immersive and interactive manner, which is not possible with traditional static textbooks. However, the creation of AR content today is largely limited to a manual process, resulting in much of the AR content being *static* and merely floating in mid-air. This is because developing *interactive* AR content involves considerable programming effort and time. To tackle this challenge, I have developed AI-powered on-demand and automatic content creation for AR.

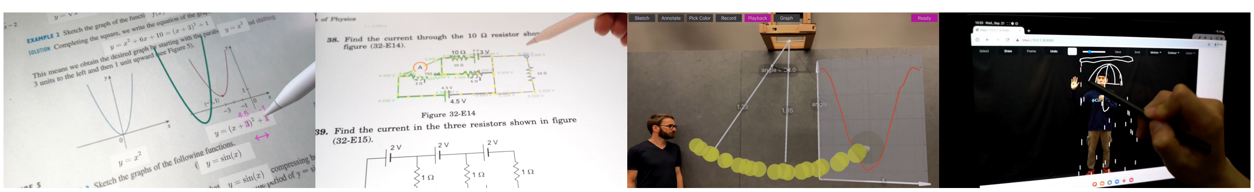

Figure 4: Augmented Math [1] (Left), Augmented Physics [2] (Middle Left), RealitySketch [12] (Middle Right), and RealityCanvas [17] (Right) are examples of AI-enabled interactive content creation in AR.

Augmented Math [1] and Augmented Physics [2] demonstrate the concept of AI-generated AR textbooks by extracting and augmenting content on static paper (Figure 4). Instead of manually creating interactive content, we introduce a machine learning-based approach that transforms static math and physics textbooks into dynamic, interactive AR tutorials overlaid directly onto the pages. Our system begins by scanning a static document, employing optical character recognition and computer vision to identify values, symbols, and graphs. These elements are then converted into interactive visuals which are seamlessly integrated with the original document using AR. As a result, our approach enables even non-technical users like teachers and students to convert their textbooks into interactive, customizable, and explorable explanations in real-time.

RealitySketch [12] and RealityCanvas [17] also bring a novel dimension to AR dynamic content creation through AI-powered sketching. The unique aspect of these works is that the user can sketch *animated* and *responsive* graphics embedded in the real world with object tracking (Figure 4). Sketched elements drawn in these tools are embedded and bound to physical objects so that they can dynamically move with the corresponding physical motion. We demonstrate a rich application space of such animated AR sketching, such as augmented storytelling, physics experiments, sports training, and classroom teaching.

3. Ubiquitous Tangible AR with ML-Enabled Approach

Seamless integration of virtual and physical elements is a key to immersive AR experiences, yet enabling such blended interactions remains a significant challenge due to the complexity and intricacy of the implementation. Currently, the integration of AR with tangible objects predominantly relies on either marker-based tracking or embedded physical sensors. These methods, however, lack flexibility and practicality, which is why tangible user interfaces, despite their potential, are rare to find outside of research labs.

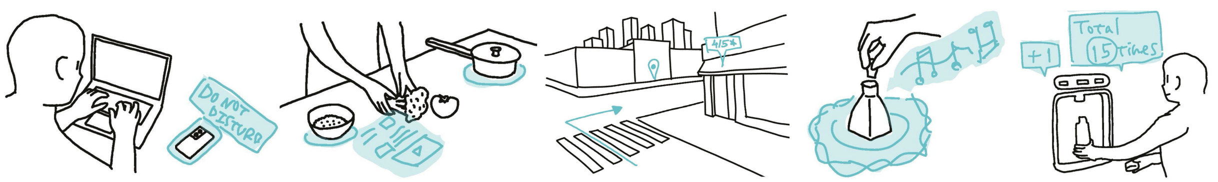

Figure 5: By leveraging AI-powered interfaces, we make tangible AR pervasive and adaptive, weaving them into the fabric of everyday life.

Techable Reality [7] aims to address this challenge by combining AR with interactive machine teaching, a novel machine learning and computer vision technique that utilizes user-generated data for custom classification (Figure 6). This approach empowers users to easily create their own tangible and gestural interactions in real-time. This functionality facilitates the rapid prototyping of functional tangible AR applications without the need for complex setup and tedious programming. Our new approach makes tangible AR interfaces more ubiquitous and adaptive, enabling users to create context-aware interfaces for various objects and environments (Figure 5).

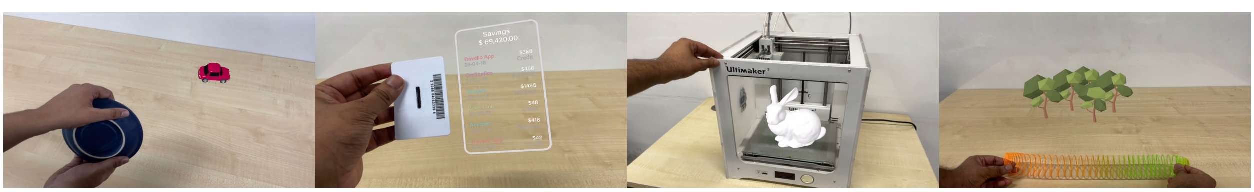

Figure 6: Teachable Reality [7] is an example of ML-enabled ubiquitous tangible AR interfaces.

4. Blended Virtual-Physical Interaction with AR and Robotics

Currently, augmented reality primarily focuses on *visual* augmentation, while the *physical* aspects of our environment remain static and non-programmable. Looking beyond just visual augmentation, I envision a future where AR can be seamlessly incorporated into dynamic physical environments by leveraging robotics and IoT devices. This integration of AR with actuated environments will significantly enrich the interaction between the virtual and physical worlds, delivering both expressive visual and tangible feedback for a more immersive experience.

Sketched Reality [5] explores this blended virtual-physical interaction with AR and swarm robots (Figure 7). This project combines AR sketching with actuated tangible interfaces for a two-way interactive experience. For example, virtually sketched elements like ropes and walls can interact with, move, and even constrain physical robots, blurring the boundary between virtual and physical worlds. Building on this concept, HoloBots [3] takes it a step further by introducing holographic physical telepresence, enabling holographic users to physically interact with remote objects as if they were present in the same location (Figure 7). The integration of AR and robotics opens up a broad spectrum of design possibilities, and I have mapped out these future opportunities in AR and Robotics Survey paper [10].

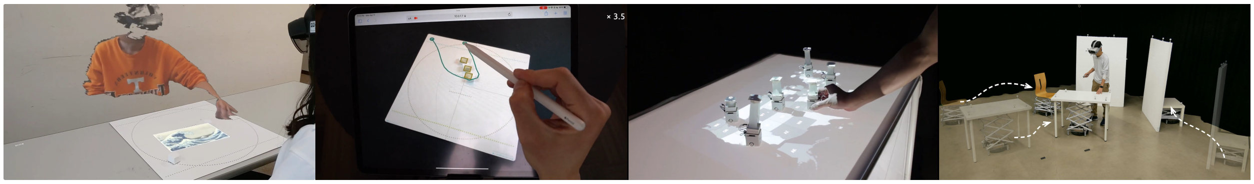

Figure 7: HoloBots [3] (Left), Sketched Reality [5] (Middle Left), ShapeBots [16] (Middle Right), RoomShift [9] (Right) are examples of blended virtual-physical interactions enabled by AR and robotics.

My past research has further explored actuated environments facilitated by swarm robotics. Projects like ShapeBots[16], Reactile[11], and FluxMarker[15] showcase such interfaces using tabletop swarm robots (Figure 7). Expanding beyond this scale, I have also investigated large-scale swarm robotics with projects like RoomShift[9] and LifTiles [13], which feature robots capable of autonomously reconfiguring room layouts. These robots can dynamically adapt physical spaces, for instance, rearranging chairs and desks dynamically (Figure 7). One of the key advantages of dynamic physical interfaces is their ability to deliver rich haptic and tactile feedback. HapticBots [14] demonstrates this by dynamically forming and modifying physical shapes for haptic interactions in VR. This approach offers versatility across a range of applications, including education, remote collaboration, and medical training.

Research Agenda

LLMs for AR: Leveraging Large Language Models for AI-Mediated Augmented Communication
With the advent of ChatGPT, I see significant opportunities to incorporate large language models into AR. We have started experimenting with LLMs for AR document enhancement for augmented reading, but with the recent advent of multi-modal LLMs, there is even greater potential for augmented communication. I am also interested in conducting long-term studies to evaluate system effectiveness by leveraging its flexibility and generalizability.

Generative AR: Leveraging Generative AI for In-situ and On-Demand AR Content Creation
Generative AI has significant potential in AR content creation. For example, integrating generative AI tools like Stable Diffusion or ControlNet with AR sketching could enable the immediate creation of AR scenes that seamlessly merge with the real environment. Additionally, incorporating 3D content generation could transform the way to create augmented textbooks for educational experiences. For my future work, I aim to investigate how generative AI can be used to generate immersive AR applications beyond 2D interfaces.

Adaptive Tangible AR: Designing Adaptive AR Interfaces for Contextual Physical Space and Activity
I also plan to explore AR assistants that adapt to the user's context and environment. Building upon my previous work, I intend to develop context-aware assistants combining computer vision and large language models to achieve this goal. By harnessing object recognition and scene understanding capabilities in AR and AI, this system would be capable of autonomously recognizing a user's actions and the surrounding information to provide adaptive and contextually relevant responses. Such interfaces will be seamlessly integrated into the user's physical environments, enhancing their everyday lives.

Explainable AR: Explainable AI and Robotics with AR-Enhanced Visualization
Finally, integrating explainable AI with AR is an important research direction. This addresses the increasing need for privacy and accessibility in AR. In addition, such explainable AR interfaces could become important to understanding robot behaviors, as the transparency of robots' decisions becomes essential. Future research should focus on merging explainable AI and robotics within AR, enhancing the visibility and interpretability of AI in real-world contexts.

References

[1] Neil Chulpongsatorn, Mille Skovhus Lunding, Nishan Soni, and **Ryo Suzuki**. Augmented math: Authoring ar-based explorable explanations by augmenting static math textbooks. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pages 1–16, 2023.

[2] Aditya Gunturu, Yi Wen, Jarin Thundathil, Mashiat Mustaq, Rubaiat Habib Kazi, and **Ryo Suzuki**. Augmented physics: An ml-integrated authoring tool to augment static physics textbooks with overlaid interactive simulations. *arXiv preprint*, 2023.

[3] Keiichi Ihara, Mehrad Faridan, Ayumi Ichikawa, Ikkaku Kawaguchi, and **Ryo Suzuki**. Holobots: Augmenting holographic telepresence with mobile robots for tangible remote collaboration in mixed reality. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pages 1–12, 2023.

[4] Shivesh Jadon, Mehrad Faridan, Edward Mah, Rajan Vaish, Wesley Willett, and **Ryo Suzuki**. Realitychat: Augmented conversation with embedded speech-driven on-the-fly referencing in ar. *arXiv preprint*, 2023.

[5] Hiroki Kaimoto, Kyzyl Monteiro, Mehrad Faridan, Jiatong Li, Samin Farajian, Yasuaki Kakehi, Ken Nakagaki, and **Ryo Suzuki**. Sketched reality: Sketching bi-directional interactions between virtual and physical worlds with ar and actuated tangible ui. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, pages 1–12, 2022.

[6] Jian Liao, Adnan Karim, Shivesh Singh Jadon, Rubaiat Habib Kazi, and **Ryo Suzuki**. 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*, pages 1–12, 2022.

[7] Kyzyl Monteiro, Ritik Vatsal, Neil Chulpongsatorn, Aman Parnami, and **Ryo Suzuki**. 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*, pages 1–15, 2023.

[8] **Ryo Suzuki**, Mar Gonzalez-Franco, Misha Sra, and David Lindlbauer. Xr and ai: Ai-enabled virtual, augmented, and mixed reality. In *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pages 1–3, 2023.

[9] **Ryo Suzuki**, Hooman Hedayati, Clement Zheng, James L Bohn, Daniel Szafir, Ellen Yi-Luen Do, Mark D Gross, and Daniel Leithinger. Roomshift: Room-scale dynamic haptics for vr with furniture-moving swarm robots. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–11, 2020.

[10] **Ryo Suzuki**, Adnan Karim, Tian Xia, Hooman Hedayati, and Nicolai Marquardt. Augmented reality and robotics: A survey and taxonomy for ar-enhanced human-robot interaction and robotic interfaces. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–33, 2022.

[11] **Ryo Suzuki**, Jun Kato, Mark D Gross, and Tom Yeh. Reactile: Programming swarm user interfaces through direct physical manipulation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2018.

[12] **Ryo Suzuki**, Rubaiat Habib Kazi, Li-Yi Wei, Stephen DiVerdi, Wilmot Li, and Daniel Leithinger. 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*, pages 166–181, 2020.

[13] **Ryo Suzuki**, Ryosuke Nakayama, Dan Liu, Yasuaki Kakehi, Mark D Gross, and Daniel Leithinger. Lifttiles: constructive building blocks for prototyping room-scale shape-changing interfaces. In *Proceedings of the fourteenth international conference on tangible, embedded, and embodied interaction*, pages 143–151, 2020.

[14] **Ryo Suzuki**, Eyal Ofek, Mike Sinclair, Daniel Leithinger, and Mar Gonzalez-Franco. Hapticbots: Distributed encountered-type haptics for vr with multiple shape-changing mobile robots. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, pages 1269–1281, 2021.

[15] **Ryo Suzuki**, Abigale Stangl, Mark D Gross, and Tom Yeh. Fluxmarker: Enhancing tactile graphics with dynamic tactile markers. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 190–199, 2017.

[16] **Ryo Suzuki**, Clement Zheng, Yasuaki Kakehi, Tom Yeh, Ellen Yi-Luen Do, Mark D Gross, and Daniel Leithinger. Shapebots: Shape-changing swarm robots. In *Proceedings of the 32nd annual ACM symposium on user interface software and technology*, pages 493–505, 2019.

[17] Zhijie Xia, Kyzyl Monteiro, Kevin Van, and **Ryo Suzuki**. Realitycanvas: Augmented reality sketching for embedded and responsive scribble animation effects. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pages 1–14, 2023.