Blankscope: Real-time Bridging of Human Perception with AI Imagination

Chiun Lee*
Graduate School of Design
Harvard University
Cambridge, MA 02138
chiun_lee@gsd.harvard.edu

Qingyun Liu*
Graduate School of Design
Harvard University
Cambridge, MA 02138
qingyun_liu@gsd.harvard.edu

Krystal Montgomery†
Massachussetts Institute of Technology
Cambridge, MA 02139
mkrystal@mit.edu

Abstract

Imagine seeing the world through a different lens. **Blankscope** is an interactive artwork that uses generative AI to reflect on our symbiotic relationship with machines and how they can shape our perception of reality. The device offers a unique duality: a live, unfiltered view of the world alongside an AI-reimagined version of the same scene. By using physical controls, the user becomes a collaborator, dynamically shifting the AI's perspective to explore different cultural and temporal lenses. This process intentionally blurs the line between what is real and what is imagined, prompting users to consider the nature of shared authorship with non-human entities and the new forms of creative responsibility that will emerge in the age of AI.

Figure 1: Blankscope dual lens interface with real-time AI generation

*Equal contribution.

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

1 Introduction

In recent years, the proliferation of generative AI has raised new questions about authorship and perception. As machine-generated content becomes increasingly indistinguishable from reality, how do we discern what is 'reality'? How do we define historical narratives when pasts and futures can be synthetically rendered? **Blankscope** explores these questions through a dual-lens device that merges tangible interaction with reality and real-time generative imagery.

This project investigates the boundaries between physical and synthetic perception by enabling users to see both the captured present and a generated alternative world. One lens displays live video input received from a webcam, and the other renders a real-time AI-generated image sequence. The AI model is controlled by the user's adjustments of cultural and temporal parameters. Through tangible rotary encoders, users actively participate in shaping what they see, shifting between generated historical reconstructions and speculative futures among different cultural contexts.

Positioned at the intersection of human-AI interaction, mixed reality, and tangible perception, this work proposes a form of embodied co-creation, where meaning emerges not only from the machine but from the tension and fusion between human perception and machine-generated reality.

2 Methodology

To facilitate a beginner-friendly experience, we focused on providing simple, yet impactful, controls that directly influence the AI's output and, consequently, the user's perception.

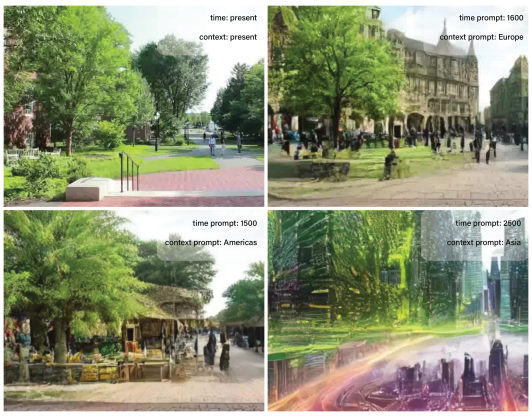

Figure 2: Different realities based on time and context prompts

Figure 3: Blankscope dual realities overlaying and merging into one

Key System Components: Our system comprises several interconnected components working in concert to deliver the immersive experience. A comfortable Binocular Enclosure houses a High-Resolution Camera capturing the real-time scene and two microdisplays. Rotary Encoders provide user input for time-based and style-based AI transformations, wirelessly controlled via a Xiao ESP32S3. A Dedicated Local Computer provides the primary processing power for complex generative AI tasks, with OpenAI API integration translating encoder inputs into detailed textual prompts for the AI model. The AI Image Generation Model, heavily based on radames/Real-Time-Latent-Consistency-Model, processes the live camera feed and textual prompts to generate altered visual outputs. Finally, the Display System utilizes two 0.7-inch OLED microdisplays (1920x1080).

Technical Implementation: **Blankscope** integrates hardware prototyping, real-time machine learning inference, and prompt-driven content generation into a compact dual-lens device. The system comprises three primary subsystems: physical interaction, cloud-based AI processing, and real-time visual output.

Input: Tangible Encoding of Time and Context. User input is mediated through two rotary encoders connected to a Xiao ESP32-S3 microcontroller. Encoder A encodes temporal displacement using

sequential 'a' and 'd' characters, each representing ±100 years from a 2025 baseline. Encoder B maps rotational position to one of five global cultural contexts (Asia, Africa, Europe, Americas, Oceania), divided across 360 degrees. This input stream is transmitted via serial communication to a Raspberry Pi 5, which formats the codes and sends them through a Cloudflare Tunnel to a remote NVIDIA GPU-enabled server for processing.

Prompt Generation and AI Rendering. The remote server interprets incoming signals through a Python-based prompt chain. Time and culture inputs are parsed into a specific year, a cultural label with historical context, and a cinematic, photo-realistic prompt via GPT-4o mini API. This prompt is then passed to the Real-Time-Latent-Consistency-Model (RTLCM) to generate temporally and culturally conditioned visuals. The pipeline ensures consistent material and visual language, reflecting real-world and speculative hybrid aesthetics.

Output: Dual-Lens Mixed Reality Display. Final visual output is routed back to the Raspberry Pi and displayed through two Sony ECX336A micro OLEDs with custom VR optics. The left lens shows the webcam's live input (reality), while the right displays the AI-generated video. A fusion mode allows the two videos to overlap within the viewer's perceptual space, with adjustable overlay ratio and interpupillary alignment.

3 Exploring the Theme of Humanity

The intricate relationship between perception, media, and technology has long been a focal point in human-device interaction research. Recent advancements in Large Language Models (LLMs) and AI are driving new forms of human-machine co-creation, with a growing number of projects using lens-based form factors to explore speculative imagination through immersive and re-contextualized digital experiences. However, few current interfaces allow users to co-create alternative realities through embodied, real-time interaction. Our system addresses this gap by creating a novel paradigm for human-AI co-creation in physical spaces, drawing inspiration from "Large Language Objects" to invite users to engage with reality as a fluid, co-generated narrative [1].

Blankscope serves as an experiential platform to engage with questions about AI perception and human vision. The collaboration in **Blankscope** is a real-time, embodied dialogue, where the AI's strength lies in its ability to rapidly generate vast, data-driven visual interpretations, and the human's unique strength is in synthesizing these disparate visual inputs, applying emotional context, and performing the final act of creative synthesis within their own mind. The artwork emerges from this powerful synergy, where the machine provides the material and the human mind gives it meaning.

The work also considers the new human rituals, responsibilities, or roles that will emerge as more and more visual AI tools become part of our devices. The act of using **Blankscope** becomes a new ritual of contemplative co-creation. The user's role shifts from a passive author to an active "navigator" of shared realities, a position that comes with the responsibility to critically evaluate the AI's output, to understand its inherent biases and stereotypes, and to consciously decide how to interpret and integrate its vision with the present moment. This is a new form of human responsibility—to be a discerning partner to a non-human collaborator. The device makes shared authorship tangible: the user initiates the creative act with their input from the physical rotary encoders; the AI responds with its generative interpretation; and the user's mind completes the artistic process.

The work prompts us to consider what principles and values we imbue in artificial intelligence systems, and what are the perils of anthropomorphizing non-human systems. By presenting the AI's visions as imaginative interpretations rather than precise recreations, **Blankscope** reveals the inherent biases and "fragmented cultural narratives" embedded in the model's training data. The user's active role in navigating and synthesizing these visions serves as a critical check against the peril of anthropomorphizing the AI. It reminds the user that they are collaborating with a powerful tool that interprets, rather than a conscious partner that understands, highlighting the importance of human ethical and emotional wisdom as a counterbalance to the machine's capabilities.

Author Bios

Chiun Lee is an interdisciplinary designer at Harvard University. His work focuses on human-machine interaction, AI perception, and spatial design. With a background across human-computer/AI

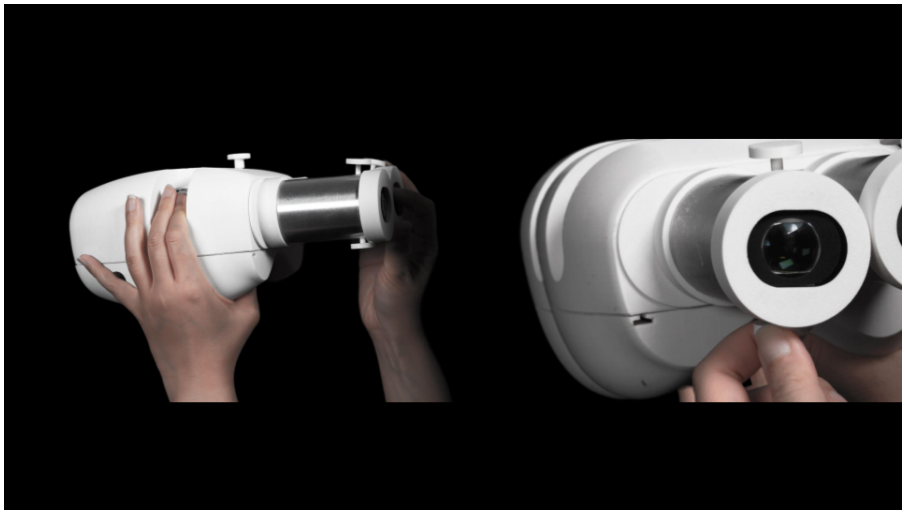

Figure 4: blankscope interactions

interaction and architecture, he explores how technology and design can create dynamic, engaging experiences that invite curiosity, encourage participation, and reframe our relationship with space.

Qingyun Liu is a creative technologist at Harvard University and a researcher at the MIT Media Lab and Microsoft Research Asia. Her work focuses on the intersection of human perception, AI and embodied interaction. With a background across human-computer/AI interaction and urban design, she investigates how AI and machines can understand human senses, memory, perception and surrounding environment, aiming to augment human and connection across realities.

References

[1] M. Coelho and J.-B. Labrune. Large language objects: The design of physical ai and generative experiences. *Interactions*, 31(4):43–48, 2024.