Title	Indra : exploring human agency and creative collaboration with embodied AI robotic systems
Sub Title	
Author	中川, 真里奈(Nakagawa, Marina) 南澤, 孝太(Minamizawa, Kouta)
Publisher	慶應義塾大学大学院メディアデザイン研究科
Publication year	2025
Jtitle	
JaLC DOI	
Abstract	
Notes	修士学位論文. 2025年度メディアデザイン学 第1173号
Genre	Thesis or Dissertation
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO40001001-00002025-1173

Master's Thesis
Academic Year 2025

Indra: Exploring Human Agency and Creative Collaboration with Embodied AI Robotic Systems

Keio University
Graduate School of Media Design

Marina Nakagawa

A Master's Thesis
submitted to Keio University Graduate School of Media Design
in partial fulfillment of the requirements for the degree of
Master of Media Design

Marina Nakagawa

Master's Thesis Advisory Committee:

Kouta Minamizawa	(Main Research Supervisor)
Masa Inakage	(Sub Research Supervisor)

Master's Thesis Review Committee:

Kouta Minamizawa	(Chair)
Masa Inakage	(Co-Reviewer)
Shutaro Takeda	(Co-Reviewer)

Abstract of Master's Thesis of Academic Year 2025

Indra: Exploring Human Agency and Creative Collaboration with Embodied AI Robotic Systems

Category: Design

Summary

We present INDRA, an Interactive Deep-Dreaming Robotic Artist designed to support embodied, real-time collaboration between humans and machines in traditional painting. INDRA operates through a turn-based co-creation framework, where the artist and the robot take alternating turns contributing to a shared canvas. Rather than automating creativity, INDRA enhances it acting as a responsive partner that adapts to the user's input through speech recognition, computer vision, and generative image synthesis. Visually embodied as a cybernetic Maneki-neko the Japanese beckoning cat INDRA is designed to be playful and approachable, challenging impersonal perceptions of AI. The system encourages improvisation and shared authorship, positioning the machine not as a tool, but as a collaborator. By blending emerging technology with traditional arts, INDRA promotes a more inclusive, human-centered vision of AI in creative practice.

Keywords:

robotics, embodied ai, painting, sense of agency, creativity, collaboration

Keio University Graduate School of Media Design

Marina Nakagawa

Contents

List of Figures

List of Tables

Acknowledgements

I would like to express my deepest gratitude to my family for their unwavering support throughout this journey. To my father, mother, and older sister—thank you for your patience, encouragement, and belief in me. Your presence, whether near or far, gave me the strength to continue this work even during the most challenging times. This project would not have been possible without your love and support.

I am also profoundly thankful to Abbi, Carmen, Daniel, Daphne, Denise, Erika, Kinga, Potato Chip, Priya, Sohei, and Sofie. Each of you, in your own way, has offered invaluable support, through thoughtful conversation, mentorship, creative collaboration, emotional grounding, and generous companionship. Your belief in me and my work has been a source of motivation, perspective, and strength throughout this process.

Chapter 1
Introduction

1.1. Background

As artificial intelligence continues to advance, its integration with physical systems—commonly referred to as embodied AI—has opened new opportunities for augmenting human skill through real-time, responsive collaboration. While much of contemporary research in AI focuses on autonomous performance, there is a growing need to explore how AI can function as a partner to human users in domains requiring dexterity, creativity, and agency.

Human skill is traditionally situated in tasks involving complex sensorimotor coordination and high-level decision-making—domains that resist full automation due to their reliance on tacit knowledge and nuanced intent. Fields such as artistic creation and dynamic object manipulation (e.g., juggling) exemplify this challenge, as they require adaptability, timing, and an understanding of context that current AI systems often lack in isolation.

To address this gap, this thesis investigates the design and implementation of embodied AI systems that support, rather than replace, human skill expression. Specifically, it introduces and examines two complementary systems:

- INDRA (Interactive Deep-Dreaming Robotic Artist): A collaborative robotic painting system designed to preserve human artistic agency during co-creation with a physical AI partner.
- ASURA: A telepresence platform enabling a human user to control multiple robotic arms to juggle a hockey puck in real-time, assisted by AI to manage the complexity of multi-limbed coordination.

Both systems aim to reveal how physical embodiment, AI assistance, and real-time feedback can be integrated to create tools that extend human capability in meaningful and interpretable ways. This research draws on multidisciplinary foundations in robotics, human-computer interaction (HCI), cognitive science, and creative practice. It contributes to a growing discourse around human-centered AI, teleoperation, and interactive skill-sharing systems, and offers engineering design insights for building systems that enhance the experience of embodied skill through AI collaboration.

1.2. Motivations

Traditional human skills, whether in the fine arts or juggling (dexterous physical manipulation), represent deeply embodied practices rooted in cultural identity, expertise, and self-expression. As artificial intelligence (AI) becomes increasingly integrated into these domains, there is growing concern about how it may displace rather than support human agency, particularly when engaging traditional artists [1]. In creative fields, AI-generated imagery has raised concerns about authorship, control, and creative ownership, while in robotics and automation, telepresence systems risk reducing users to mere operators of opaque machine intelligence. These shifts threaten to erode meaningful human participation and may contribute to psychological phenomena such as Creative Displacement Anxiety (CDA) [2], as well as broader public skepticism toward AI technologies.

The motivation behind is not just to augment human skill with artificial capabilities, but to build systems that listen, respond, and ultimately empower.

The idea for Indra didn't start with a technical challenge but with a blank canvas.

As someone who struggles with anxiety, especially around creative performance, I often found myself frozen in front of a canvas, uncertain of where or how to begin. A mostly empty space can feel more threatening than inspiring. I was afraid of making the "wrong" mark, of ruining something before it had even started. This experience of artistic paralysis, a kind of creative block fueled by perfectionism and fear, was something I wanted to find a solution for, utilizing new technology. I designed Indra not just as a tool, but as a partner that could help initiate the painting process and could make the first move when I couldn't. When Indra makes a brushstroke, it breaks the silence of the surface—suddenly, the canvas is no longer empty and I don't have to face it alone. That small shift made it easier to respond, to take a turn, and to keep the creative cycle going.

In parallel, I've always been deeply connected to digital creation. I use creative coding to generate images and experiences, often building visual systems from lines of code. But despite this background, I never found an easy or satisfying way to integrate that kind of digital expressiveness into physical painting until I started building Indra. It became a way to bring these two creative worlds together: the unpredictability and presence of real paint with the generative, iterative logic of code. These personal motivations evolved into broader design goals. If having a co-creative partner could reduce my own anxiety and unblock my engagement with a canvas, perhaps it could support others, too, especially those who face barriers to creative expression, whether physical, cognitive, or emotional. From this emerged Indra's role not just as a painting collaborator, but as an accessible, embodied AI system designed to encourage participation, reduce intimidation, and center the human artist.

By embedding generative AI within a real-time, turn-based painting loop, Indra explores how agency can be shared without being diminished. Unlike autonomous tools that produce finished outputs, Indra foregrounds the human in the creative cycle. Each interaction is responsive, conversational, and incomplete until the artist responds. This aligns with a growing body of research in HCI and AI ethics that argues for systems that preserve meaningful human control, foster trust, and support co-adaptive behavior between humans and machines.

Indra examines how *human agency* evolves when integrating physical generative AI into a traditional art workflow. However, the language and design surrounding AI often imply an entity with its own intent, potentially undermining artists' sense of ownership and originality [3,4]. Our goal is to *improve perceptions of generative AI among traditional artists* while preserving their *artistic agency*, the ability to recognize themselves as the primary author of an artwork.

The rise of AI-generated art has significant psychological and social implications for traditional artists. The increasing reliance has led to Creative Displacement Anxiety (CDA), a psychological response to AI-generated art. CDA highlights how AI's ability to mimic or surpass human creativity leads to identity loss, economic anxiety, and reduced artistic motivation among creatives. It also causes skepticism, alienation, and choice paralysis among consumers, emphasizing the need for ethical AI integration to mitigate its negative effects [2]. Socially, the online art community might experience a divide between proponents of AI art and those who reject its use due to misalignment of core values, suggesting the design of AI art technologies should be human value-centered [5]. Applying generative AI to traditional painting requires considering the effects on the existing painting community. Art Workers Japan found from a survey of creative field workers (25,560 responses) that 59% of creatives worry about losing work opportunities to AI, and 92% feared their work would be scrapped as art theft [6]. This highlights the concerns regarding AI potentially replacing human artists or misappropriating their creations. Indra doesn't use AI to enhance productivity or output, but to reduce the psychological pressure that often surrounds artistic creation. Especially for those who experience anxiety, creative block, or uncertainty in front of a blank canvas, it acts as a responsive collaborator that breaks the silence and lowers the threshold to begin. It creates space for participation, encouraging users to see themselves as the primary agents of the work, offering a vision of how embodied AI can support human creativity.

As robotics systems become increasingly integrated into human workflows, the question of how humans can meaningfully control and coordinate complex embodied systems is gaining urgency. Tasks that require rapid, bimanual, or high-precision motor control (surgical assistance, manufacturing, teleoperated maintenance) demand a delicate balance between machine intelligence and human intent. However, many existing systems treat the human either as a supervisor (issuing high-level commands) or a low-level actuator (manually guiding each joint), with little room for fluid co-regulation or adaptive skill sharing.

As a explanatory step in response to this gap and toward development of Indra, we developed Asura, a real-time, AI-assisted telepresence system that allows a user to control two robotic arms to perform a puck-juggling task on a hockey table. This task represents a form of dynamic object manipulation, requiring the coordinated control of multiple effectors with tight timing and feedback loops. The details are described in Appendix.

Central to Asura's design is the principle of meaningful human control: the idea that AI systems should be transparent, supportive, and correctable, ensuring that the user remains the primary agent in task execution [7, 8]. Rather than automating the juggling outright, Asura provides assistive inference—predicting trajectories, stabilizing errors, and adjusting force output while preserving the user's control over timing and direction. This approach aligns with models of collaborative control in robotics, where human and machine inputs are blended in real time according to task context and user behavior [9, 10].

By exploring a non-utilitarian but skill-intensive scenario like robotic juggling, Asura shifts the narrative from performance optimization to expressive physical control, positioning telepresence not as a form of remote labor but as a medium for augmented skill embodiment. The project contributes to a growing discourse in HCI and robotics that emphasizes co-adaptive systems, where human and AI operate as partners, dynamically modulating influence and responsibility [11, 12].

Ultimately, Asura offers a platform to study how shared control frameworks and embodied AI can support human agency, not only in traditional robotic domains but in domains where rhythm, responsiveness, and fluency define expertise. Although Asura could constitute an independent research project separate from Indra, it shares overlapping research themes and technical skill developments. Therefore, we describe Asura accordingly, referring Appendix.

1.3. Research Question

1. Can INDRA extend human skill through real-time physical interaction?
    - Does embodiment (robot arms, physical presence) affect a user's sense of control, authorship, and collaboration compared to screen-based AI systems?
    - How do users perceive the boundary between their intention and the AI's action in a shared physical workspace?
2. Can INDRA support human agency in creative tasks?
    - What kinds of interactions reduce fear of replacement and foster a sense of cooperation?

- Can AI systems be made more legible or transparent to maintain users' sense of authorship and reduce perceptions of loss or theft of agency?

These questions are not just about how AI works, but how it feels to work with it. When a machine shares our canvas or our tools, when it moves beside us and not behind a screen, the dynamic changes. We start to ask: Who's leading? Who's responding? What parts of this outcome are mine? This thesis focuses on those moments—of uncertainty, of negotiation, of shared creation—to better understand what makes AI feel like a partner rather than an intruder. It's from this place that the following hypothesis takes shape.

1.3.1 Hypothesis

When embodied AI systems such as Indra and Asura are designed with human-centered interaction principles, including transparency, real-time responsiveness, and clearly shared control, users are expected to report a stronger sense of agency, less anxiety related to performance, and a greater willingness to see AI as a collaborative partner. These outcomes are likely to correspond with increased trust, perceived authorship, and sustained engagement.

This hypothesis treats agency not as a fixed setting or a simple measure of control, but as something relational. It shifts depending on context, shaped by timing, responsiveness, and the capacity of the system to listen. Agency can emerge in quiet moments, when a robot pauses, follows, or responds with care. Whether it begins with a painted line or a well-timed assist, the AI's role is not to take over, but to offer space for human expression.

From this perspective, the goal is not simply to show that AI improves performance. It is to ask whether it can help people begin. Can it hold open the canvas long enough for someone to overcome hesitation? Can it meet hesitation not with command, but with invitation?

By exploring these questions through systems like Indra and Asura, this research attempts to capture something quieter than performance, a tool in the act of creation. A reminder that collaboration, when designed with care, doesn't reduce human control. It reminds us that we were in control all along.

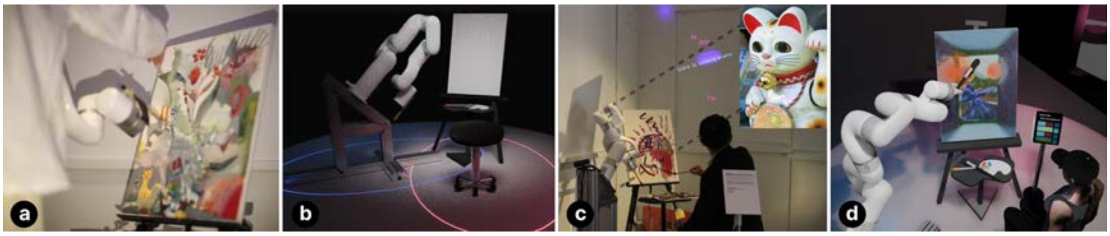

Figure 1.1 Teaser of Indra System, a painting robot proposed in this thesis.

1.4. Thesis Structure

This thesis is organized into seven chapters, each building toward a comprehensive understanding of how embodied AI systems influence human agency in creative collaboration.

- **Chapter 1** introduces the motivation for the study, defines key terms such as agency, embodiment, and co-creation, and presents the research questions and hypotheses that guide the investigation.
- **Chapter 2** reviews relevant literature on agency in human-computer interaction, including foundational theories and contemporary debates. It also situates the work within existing studies on collaborative robotics, human-centered AI, and generative creativity. Special attention is given to recent systems like Google DeepMind's Veo, which exemplify emerging strategies for preserving user authorship in generative AI tools.
- **Chapter 3** outlines the technical design of the AI system developed for this research—*INDRA* an embodied robotic painting collaborator. This chapter discusses their architecture, interface mechanisms, and the conceptual rationale for each system's design choices.
- **Chapter 4** provides results of multiple validation tests of deployed protoype of Indra from various demo and exhibition opportunities, allowing for a range of initial observation crucial to the final user study.
- **Chapter 5** presents the methodology for the user study, including participant recruitment, evaluation metrics, study flow, and data collection instruments. Reports the output data results of the user study.

- **Chapter 6** reports the overall findings of the user study, drawing on both quantitative data and qualitative feedback to examine differences in perceived agency and authorship. Discusses these findings in relation to the broader literature, highlighting how embodiment, interaction modality, and task context shape the dynamics of human-AI collaboration.

- **Chapter 7** concludes the thesis by summarizing key contributions, identifying limitations, and proposing directions for future research. It also reflects on the ethical and inclusive design implications of creating AI systems for creative practice.

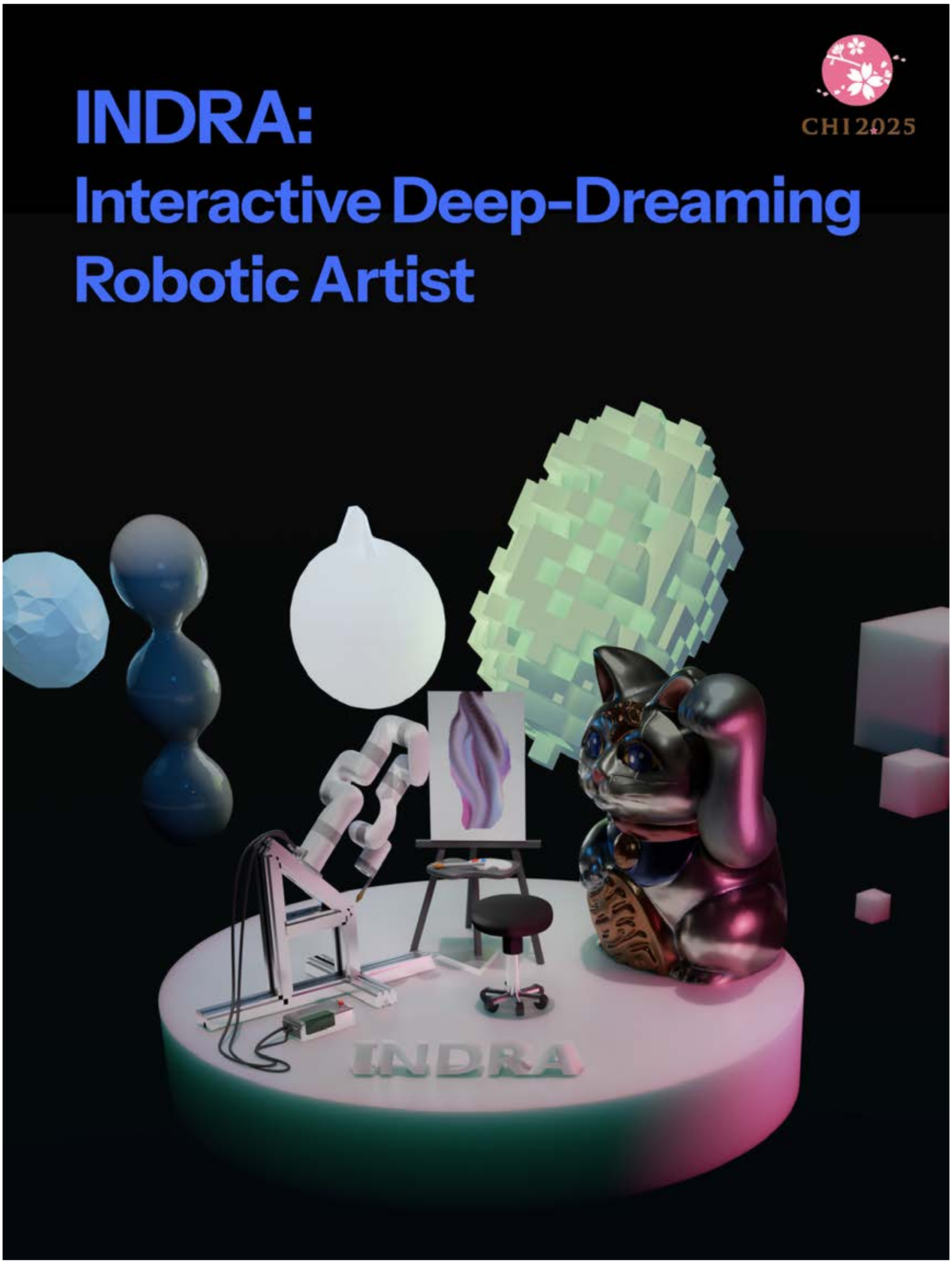

Figure 1.2 Indra's Poster for CHI'25 Interactivity exhibition.

Chapter 2
Literature Review

2.1. Agency in Human-Computer Interaction

In HCI, agency refers to the user's perception of initiating and controlling actions within an interaction system. When interacting with AI, users may experience a shift in agency due to the system's autonomy and decision-making capabilities.

According to frameworks developed by Limerick et al. [13] and Shin and Park [14], agency is influenced by:

- The transparency of the AI's intent
- The predictability and responsiveness of its actions
- The user's ability to override or guide those actions

Indra and Asura introduce complex embodied interactions where the AI visibly acts in the world. This visibility intensifies the importance of agency: users must not only feel involved but also understand their influence on the AI's behavior. If users cannot clearly see how their input affects the shared outcome (e.g., brush strokes, juggling paths), agency may be diminished.

- Interfaces that clarify intent and control sharing (e.g., turn-taking in INDRA)
- Real-time feedback loops (e.g., predictive assistance in ASURA)
- Mechanisms that keep the user's intent at the center of the output

The growing presence of AI in interactive and embodied systems has intensified scholarly debate around the preservation of human agency—broadly defined as the

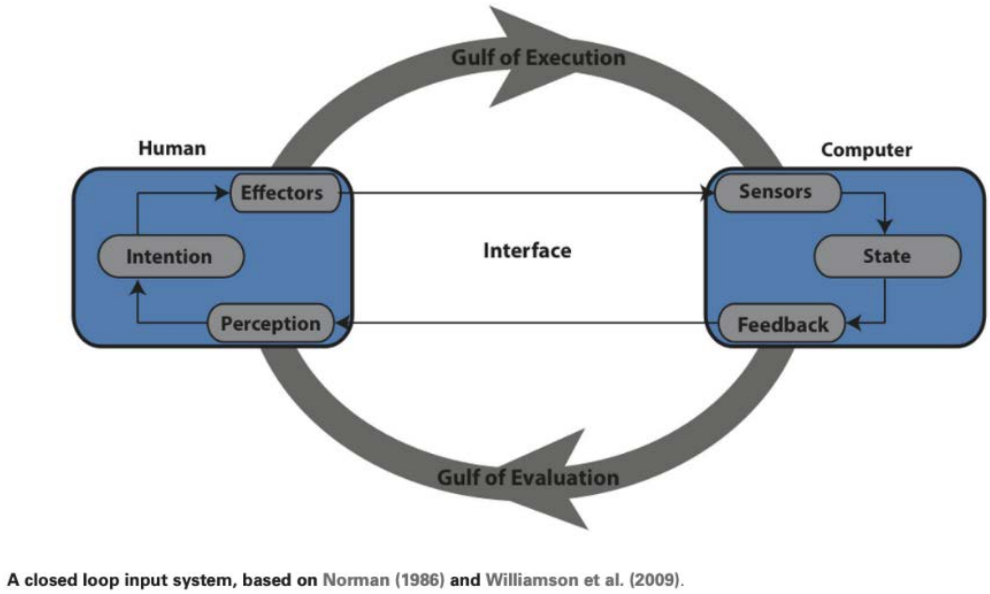

Figure 2.1 Closed loop system of human-computer interaction showing the cycle of execution and evaluation, key for Human Agency.

capacity for individuals to act intentionally and to perceive themselves as the origin of their actions [13, 14]. This discourse is particularly urgent in systems where AI does not merely offer passive automation but becomes an active, embodied agent that is capable of shaping physical, creative, or cognitive outcomes.

A notable contribution to this debate is the Pew Research Center's 2023 study, The Future of Human Agency, which surveyed over 500 experts, including computer scientists, ethicists, and futurists, on the trajectory of AI and human control [15]. Respondents were nearly evenly divided: 56% anticipated that by 2035, most AI systems would not be designed to prioritize human agency in decision-making, while 44% believed systems would be sufficiently human-centered to support user autonomy. Experts voiced particular concern about the rise of "black box" AI systems, which obscure operational logic and thus limit users' ability to interpret, intervene, or feel ownership over outcomes [16, 17].

These concerns intersect directly with research on Human-Centered Design (HCD) in AI, which emphasizes transparency, interpretability, and shared control as prerequisites for preserving user autonomy [1, 18]. In embodied contexts, where AI systems have a physical presence and interact in real time—as in the cases of Indra and Asura—the threat to agency becomes not merely cognitive but sensorimotor and experiential. That is, when AI operates within a shared physical environment, its actions can overwrite or dominate human contributions unless carefully mediated through co-adaptive or collaborative control strategies.

This tension is evident in the design of collaborative robotics systems, which aim to balance human initiative with machine assistance [9, 11]. Here, agency is negotiated dynamically: users may delegate some motoric or computational load to the AI, but effective systems provide mechanisms for intervention, intent shaping, and context-sensitive support [10, 12]. In embodied AI robotic systems, the user remains the central creative or manipulative agent, but the AI subtly scaffolds timing, form, or coordination in service of skill expression.

Crucially, this reframing of AI not as a replacement but as a co-regulator of human activity responds to broader calls within the HCI community to design human-AI systems that center agency rather than automation [19]. Such approaches resonate with emerging views on value-sensitive design, where moral values such as autonomy, dignity, and accountability are embedded in the structure of interaction systems [1].

This emphasis on user-centered design is not limited to embodied systems. The development of generative AI has introduced new paradigms in creative production, particularly within domains like video generation. Google DeepMind's Veo, a recent advancement in text-to-video synthesis, exemplifies this trajectory by foregrounding mechanisms that give users granular influence over the generative process. Veo enables creators to specify detailed aspects of a scene, including camera movement, lighting, and visual style, through structured text input. This approach prioritizes interpretability and modular control, allowing users to shape outputs in ways that reflect their own intentions and stylistic choices [20]. Instead of fully automating the creative process, Veo is structured to remain pliable and directive-sensitive, supporting workflows where the AI acts as a responsive toolset rather than a sole generator. Additionally, Veo integrates with Flow, an interface designed to simplify the assembly of scenes and integration of narrative components. This system facilitates a layered approach to video-making in which users retain authorship over structural and expressive decisions while drawing on the efficiency of AI-assisted tools [21]. These developments reflect a broader movement in human-computer interaction that shifts away from monolithic automation toward co-configurability. AI systems are increasingly designed to maintain human framing and interpretive input throughout the interaction process. In this context, Veo illustrates how even non-embodied systems can be structured to reinforce user authorship, transparency, and fine-grained control within complex creative tasks.

In sum, the current literature foregrounds agency as a central metric for evaluating embodied AI systems. When systems like Indra and Asura are designed according to HCD principles and collaborative control paradigms, they not only augment human skill but reinforce the user's identity as an active, autonomous contributor to the interaction. This aligns with the broader sociotechnical imperative: that emerging technologies must not only "reveal the magic" through technical novelty, but do so in a manner that respects and amplifies the user's agency.

2.1.1 Meaningful Levels of Human Control

A growing body of scholarship has emphasized the importance of maintaining meaningful levels of human control in AI-driven systems, particularly those that operate in dynamic, embodied environments [7]. The term refers to a design paradigm wherein users retain the ability to influence, supervise, or override AI behaviors in ways that are intuitive, context-aware, and aligned with their goals. Unlike superficial control where human input is symbolic or disconnected, meaningful control ensures that the user's actions directly affect system behavior in perceptible and consequential ways. In embodied AI systems, this becomes especially relevant: when physical machines execute actions in the world, the stakes of autonomous behavior increase, and the margin for misalignment between human intent and machine execution narrows.

For embodied AI robotic systems like Indra, meaningful control is not only a safety or usability requirement, but also a core mechanism for preserving perceived agency and task ownership. In Indra, this is achieved through turn-taking structures and gesture-responsive generation, allowing artists to guide the flow of a painting while interpreting AI interventions as collaborative rather than autonomous. In Asura, the user's control over dual robotic arms is augmented by AI support that is assistive, not substitutive, ensuring that skill expression remains centered on the user's input. This echoes broader recommendations in the HCI and AI ethics communities, which argue for embedding meaningful control to promote accountability, trust, and co-adaptive learning [8].

2.2. Robotics in Art and Human Skill

Robots have long fascinated artists and technologists alike, straddling automate and animate roles in independent art-making and embodied skill enactment. This section surveys major historical milestones and recent systems that integrate robotics with artistry or dexterous gesture-based practices.

Historical Context and Early Systems

One of the earliest examples, Edward Ihnatowicz's *The Senster* (1970), was a robotic sculpture that responded to human presence through mirrored arm gestures and sound sensing, a precursor to embodied art interactivity. Around the same time, Harold Cohen's "AARON" (1973–) explored generative visual creativity via rule-based AI; while non-robotic, it laid foundational work for algorithmic painting and later influenced physically embodied painting systems [22].

Notable Robotic Painters

e-David (2009, University of Konstanz) The e-David robot paints by iteratively refining a canvas through visual feedback, adding strokes until convergence with a reference image. Unlike AARON's abstraction, e-David emphasizes realist approximation, a blend of algorithmic perception and material sensitivity [23].

Ai-Da (2019) Ai-Da is a humanoid robotic artist equipped with camera-based vision, a language model, and a hydraulic drawing arm. She produces drawings and paintings in public exhibitions (e.g., British Library, Venice Biennale), sparking debate on authorship and authenticity in machine-created art [24].

Pindar Van Arman's Systems (2005–) Van Arman's robotic painters (e.g., CloudPainter) focus on portraiture and abstraction through real-time co-creation.

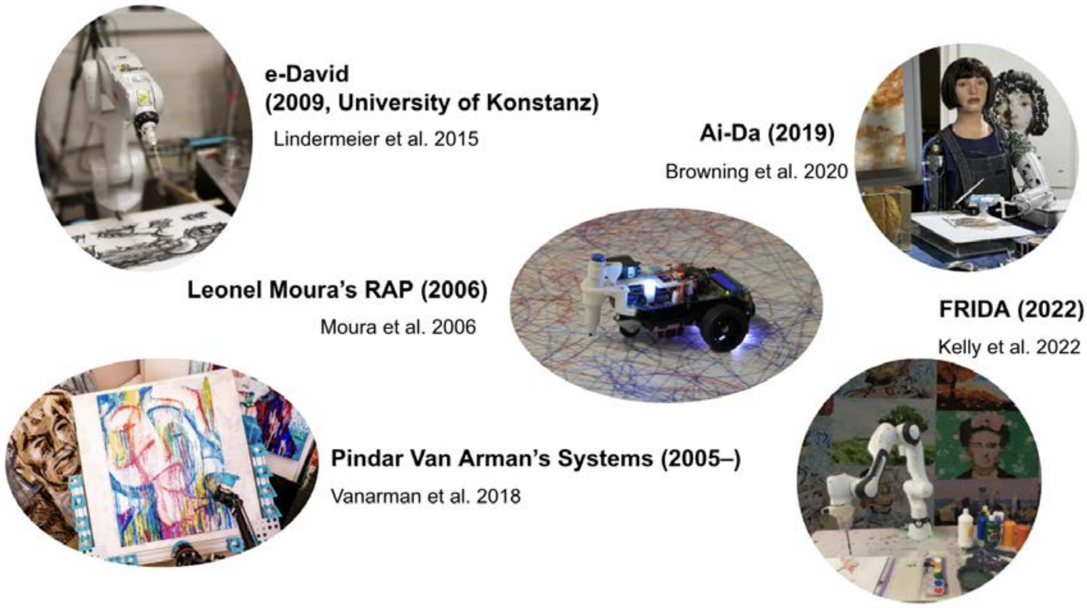

Figure 2.2 Portraits of Notable Robotic Painters.

His work has received awards from RobotArt competitions and explores increasing robot autonomy while preserving human aesthetic input [25].

Leonel Moura's RAP (2006) The Robotic Action Painter uses stigmergic algorithms to decide where to place marks and when to stop. It has been exhibited widely and represents an early effort in creative autonomy for embodied robots [26].

FRIDA (2022) Developed by Carnegie Mellon University, FRIDA leverages diffusion models and real2sim2real pipelines to translate digital artistic prompts into physically rendered brushstrokes [27].

Dexterous and Juggling Robots

Robotic dexterity also manifests in physical, skill-intensive tasks like juggling, which mirror Asura's domain:

- Chiba University's juggling robot demonstrates two-ball tosses using high-speed cameras and control loops at 500 fps [28].

- Hiroshima University's robot performs flower-stick juggling via hybrid feedback control [29].
- OMRON's contact-juggling system simulates human continuous touch-based manipulation [30].

2.3. Generative Renaissance

Generative AI art tools like DALL·E [31], MidJourney, and Stable Diffusion [32] generate images from text prompts, while robotic painting systems like Ai-Da and AICAN [33, 34] demonstrate physical outputs by autonomous agents. However, these systems often bypass the co-creative process and reduce human input to a prompt. CoFrida [35] moves closer to interactivity but remains largely simulation-based, lacking embodied, real-time participation. This echoes a growing interest in "embodied AI", where a system perceives and acts in its environment through a robotic body, allowing adaptive and tangible exchange [36, 37]. Embodied collaboration can foster more intuitive engagement than purely screen-based interactions [3, 37].

This project views AI art as part of a longer historical arc of creative tools evolving alongside human skill. Historically, new media have challenged the role of the hand in art. While tools were used to extend human capability, they often met resistance when perceived to displace skill or presence. This "historical resistance" reappears in current debates around AI art, as creatives face fears of replacement and loss of authorship [3, 4].

Embodied knowledge, rooted in gesture, repetition, and material intuition, underlies traditional art practices. Systems like Indra and Asura build on this lineage by inviting physical interaction rather than replacing it. They explore how computational tools can scaffold rather than supplant embodied skill.

Technological disruption is not new. The advent of photography, synthetic pigments, or mechanical reproduction all caused cultural upheaval, yet also expanded what art could be. Today's AI tools echo these shifts, introducing new questions about authorship, labor, and control.

Art as presence emphasizes slowness, trace, and internal expression—the human mark left behind. Systems like Indra preserve this by structuring AI involve-

By 2035, technologies *will not* be designed to allow humans to easily be in control of most tech-aided decision-making that is relevant to their lives

56% of the experts canvassed agreed with the statement that by 2035 smart machines, bots and systems WILL NOT be designed to allow humans to easily be in control of most tech-aided decision-making relevant to their lives. Among the most common themes they expressed:

- **Powerful interests have little incentive to honor human agency:** The dominant digital-intelligence tools and platforms the public depends upon are operated or influenced by powerful elites – both capitalist and authoritarian – that have little incentive to design them to allow individuals to exert more control over their tech-abetted daily activities. One result of this could be a broadening of the digital divide.
- **Humans value convenience and will continue to allow black-box systems to make decisions for them:** People already allow invisible algorithms to influence and even sometimes "decide" many if not most aspects of their daily lives – that won't change. In addition, when they have been given an opportunity to exercise some control over their tech tools and activities, most have *not* opted to do so.
- **AI technology's scope, complexity, cost and rapid evolution are just too confusing and overwhelming to enable users to assert agency:** They are designed for centralized control, not personalized control. It is not easy to allow the kind of customization that would hand essential decision-making power to individuals. And these systems can be too opaque even to their creators to allow for individual interventions.

Source: Nonscientific canvassing of select experts conducted June 29–Aug. 8, 2022.
"The Future of Human Agency"

PEW RESEARCH CENTER and ELON UNIVERSITY'S IMAGINING THE INTERNET CENTER, 2023

Figure 2.3 Survey data from pew research on the future of human agency.

ment around turn-taking, making the co-creative process legible and intentional.

Finally, **generative tension** refers to the unresolved boundary between human and machine intent. Rather than resolve this tension, Indra embraces it—inviting moments of surprise, misalignment, and negotiation as part of the shared creative process.

2.4. Attitudes towards Robots and AI

Attitudes toward artificial intelligence and robotics are a critical factor influencing the success of human-AI interaction systems. These attitudes affect not only user acceptance and trust, but also interaction outcomes, compliance, and perceived agency. As embodied AI becomes increasingly embedded in daily tasks, from creative co-production to physical manipulation, understanding how people respond to these systems has become a central concern in HCI and robotics research.

2.4.1 Transparency and Trust

Transparency in AI systems significantly impacts user trust. Stange et al. [38] found that while transparency can enhance trust, excessive detail may overwhelm users, leading to decreased trust. Similarly, Ali et al. [39] argue that the effectiveness of transparency depends on the presentation modality and task context, with poorly timed or overly technical explanations potentially undermining trust.

2.4.2 Individual Differences

Several studies have applied validated scales to examine how personality traits and social context influence AI perception. For example, research has found that individual differences, including prior experience, openness to new technologies, and cultural background, significantly mediate user attitudes toward robots and AI [40, 41]. Afroogh et al. [42] further observed that users' generalized trust in technology is predictive of how they interpret errors and adapt their expectations of AI collaborators, suggesting that user traits play a key role in sustaining trust across different interaction scenarios.

Importantly, these attitudes are not fixed. They evolve over time and are shaped by the quality of interaction. Lee et al. [43] showed that even initially skeptical users can develop positive views of robotic collaborators if the interaction is designed to support predictability, feedback, and perceived intent alignment.

2.4.3 Case Studies on Attitudes Toward Robots and AI

Understanding how users form attitudes toward AI and robots (particularly in embodied, interactive contexts) is essential for informing the design of robotic systems. Recent empirical work in human-computer interaction has deepened our understanding of how users form attitudes toward artificial intelligence and robots, namely in contexts involving real-time interaction, creativity, or physical co-presence.

User trust in robots has been shown to develop over time and can be influenced by the **social expressiveness of the system**. In a longitudinal deployment of a robotic barista, users engaged more positively when the robot exhibited extroverted and socially responsive behaviors, regardless of their own personality traits [44]. This suggests that perceived sociability in embodied agents can serve as a compensatory mechanism, modulating acceptance even in routine or task-driven scenarios.

The design of robot personality also has implications for trust calibration. For example, when robot behavior deviated from users' expectations based on social cues such as mismatching personality traits, trust in the system deteriorated, particularly during cooperative tasks [45]. This underscores the importance of aligning behavioral signaling with interaction context to maintain credibility and rapport.

Beyond individual encounters, user attitudes toward AI systems are influenced by their **perceived moral behavior** and reputational associations. Manoli et al. [46] demonstrate a phenomenon of **moral spillover**, wherein unethical behavior by one AI agent leads to broader distrust of unrelated AI systems. Notably, this effect was more pronounced than when similar actions were committed by human agents, highlighting the fragility of public trust in AI collectives.

Trust also proves vulnerable in contexts where errors occur. Malle et al. [47] found that users were more likely to assign blame to autonomous systems than to human collaborators when harm occurred, especially when the system's decision-making process was not transparent. These findings reinforce the need for embodied AI systems to integrate explainability mechanisms that clarify system actions in real time, an issue particularly salient for ASURA, where shared manipulation involves high-speed decision coordination.

Cultural values further shape how authority is perceived in AI systems. Sambasivan et al. [48] explored AI interactions in India and found users expressed deference and even gratitude toward AI advisors, despite negative outcomes. Rather than viewing errors as design flaws, many participants internalized them as user failures. Such culturally embedded responses complicate assumptions about universality in AI trust design and demand more context-sensitive approaches to interface and error communication. Similarly, Zhou et al. [49] revealed that public perceptions of conversational AI are mediated by prior experience, cultural background, and expectations around competence and creativity.

Attitudes toward AI also evolve as users interact with system errors. In a controlled investigation of trust development and repair, recent work has shown that initial trust is easily lost following an error, and common mitigation strategies such as apologies or disclaimers often fail to restore it [50]. This finding challenges assumptions about resilience in human-AI rapport and suggests that error handling must be both functional and emotionally coherent.

Among children, perceptions of robot reliability also shape trust formation. In structured experiments, young users were able to distinguish between trustworthy and untrustworthy robots based on prior exposure, indicating that interaction history plays a key role in shaping early AI social cognition [51]. For educational or creative applications of embodied AI, such as collaborative drawing or teaching tools, this has implications for long-term interaction design and developmental appropriateness.

More broadly, recent HCI work has sought to formalize frameworks for trust in AI. Emerging models emphasize the need for systems to be interpretable, consistent, and adjustable, qualities that foster calibrated trust rather than blind acceptance or disengagement [52]. These frameworks are pertinent to Asura, where human and machine control must dynamically co-adapt.

Finally, studies of AI-assisted decision-making in collaborative environments have shown that user trust depends strongly on alignment between AI reasoning and user expectations. When users are unable to understand or anticipate the AI's rationale, trust deteriorates, even when the system performs reliably [53]. This supports the inclusion of real-time feedback and intent modeling to ensure transparency in high-agency tasks.

Collectively, these studies demonstrate that trust and attitude formation are shaped not only by system performance, but are deeply influenced by cultural norms, social signaling, system transparency, and prior experience. These findings underscore the importance of designing embodied AI with attention to long-term user engagement, shared control, and responsiveness to human cognitive and social expectations.

This supports broader calls in HCI for human-centered design approaches that do not simply aim for compliance or efficiency, but instead foster positive affect, ownership, and agency in human-AI partnerships.

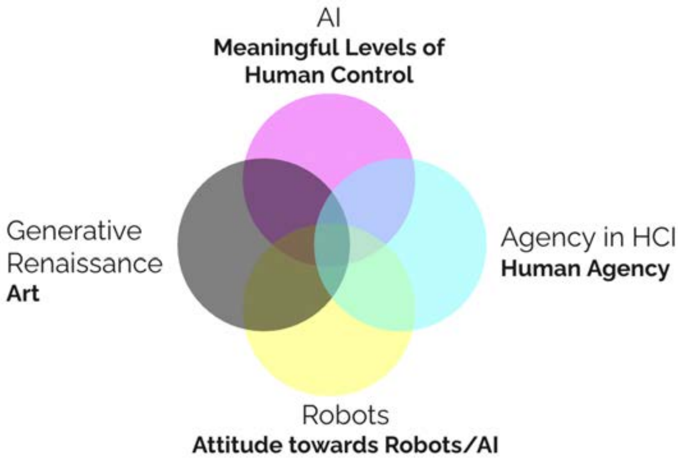

Figure 2.4 Research as an intersection of 4 concepts.

2.5. Position of this work

In summary, current AI-driven art tools and robotic painting systems face several key limitations: they are often confined to digital formats, lack real-time responsiveness, and offer minimal turn-based collaboration.

To address these gaps, we propose Indra, a real-time, user-aware AI agent that merges speech recognition, computer vision, and diffusion-based image synthesis into a single interactive and adaptive system. Our design builds on insights from layered AI architectures [54, 55], which decompose complex tasks into modular levels of perception and control. Early research on subsumption [54] and layered agents [55, 56] illustrated how real-time reactivity could be achieved by distributing responsibilities across multiple system layers. Recent developments in these principles integrate high-level generative models, allowing for rapid adaptation to user interactions and changing environments [57]. Indra leverages this layered design to combine diffusion models with ambient user inputs—spoken prompts, on-canvas sketches, and object recognition. A layered AI approach is particularly effective for *co-creative* painting, as each generative process addresses a distinct aspect of user interaction (visual, linguistic, or gestural), and collectively produces an evolving shared artwork [58]. Unlike purely digital AI art applications, Indra physically interacts with a shared canvas, creating a tangible and evolving collaboration. This tactile engagement introduces a layer of complexity and intimacy, prompting artists to explore the boundaries between control and creative partnership.

2.6. Summary

This chapter surveyed key literature on agency in human-computer interaction, particularly in relation to embodied AI systems that support creative collaboration. Foundational frameworks emphasize transparency, responsiveness, and shared control as essential to preserving user agency, especially when AI systems act visibly in the world. Research on human-centered and value-sensitive design underlines the ethical imperative to prioritize human intention and interpretability in these interactions.

Case studies of systems like Indra and Asura demonstrate how turn-taking, gesture-responsive input, and real-time adaptation can reinforce human authorship and reduce feelings of replacement or loss of control. These approaches contrast with more autonomous generative systems, yet recent developments such as Google DeepMind's Veo and Flow show that non-embodied tools can also be structured to preserve user direction and narrative ownership through thoughtful interface design.

In parallel, a growing body of work examines how trust, cultural norms, prior experience, and perceived system transparency shape user attitudes toward AI and robots. These findings suggest that collaboration quality, not just technical performance, determines whether users view AI as a supportive partner or an intrusive agent. Systems that scaffold user control without diminishing expression are better positioned to foster long-term engagement and trust.

Taken together, these studies form the conceptual foundation for this thesis. They clarify the stakes of agency in embodied AI design, highlight the limitations of current generative tools, and motivate the creation of these systems that seek to center users not only as initiators, but as sustained co-creators in physical and creative tasks.

Chapter 3
Concept and System Design

3.1. Vision

We live in a paradoxical present where human technological ingenuity has outpaced moral imagination. Artificial intelligence and robotics that were once symbols of a radiant future now elicit apprehension, alienation, and even fear as narratives of erasure, of lost work, lost control, and ultimately, a loss of self. As we move into the post-anthropocene, a period defined not by the end of the human but by the decentralization of human exceptionalism, new forms of coexistence with technology must be envisioned, not merely engineered.

This work is situated within what can be described as a post-anthropocene humanism, a reimagined humanism that does not assert dominance over the machine or nature, but instead seeks relationality, co-agency, and mutual transformation. As Braidotti notes, the posthuman is not the end of humanity, but an expansion of what counts as human experience, entangled with the nonhuman, the machinic,

Figure 3.1 Painting using the Parallel Avatar System.

Figure 3.2 First portrait I painted through controlling a robot arm.

and the planetary [59]. It is in this space of entanglement that I locate my work. My vision is to design AI-robotic systems that extend the self through embodied interaction, and to confront the emotional and conceptual rift that now defines public discourse on technological futures. Rather than designing systems that replace human creativity or dexterity, these robotic collaborators reveal the nuanced gradients of human-machine entanglement. Indra, in its painterly patience, engages users in co-creative dialogue. Asura, in contrast, amplifies bodily control, allowing a single user to embody four robotic arms, a choreography that destabilizes traditional notions of the self while offering a new frame for skill, coordination, and agency.

The public is not yet fluent in such relational modes. Surveys from 2025 show persistent unease: 62% of Americans express discomfort with AI's role in creative fields [60], while in Japan, where the cultural intimacy with robots is often romanticized, recent surveys suggest that trust diminishes sharply when robots perform tasks perceived as "human" or "expressive," such as drawing, storytelling, or caregiving [61, 62]. This discomfort is not just technical; it is philosophical. Bruno Latour cautions that "we have never been modern" [63], reminding us that our separation from the nonhuman was always a convenient fiction. Post-Anthropocene humanism dissolves this boundary and demands new tools of sensemaking that

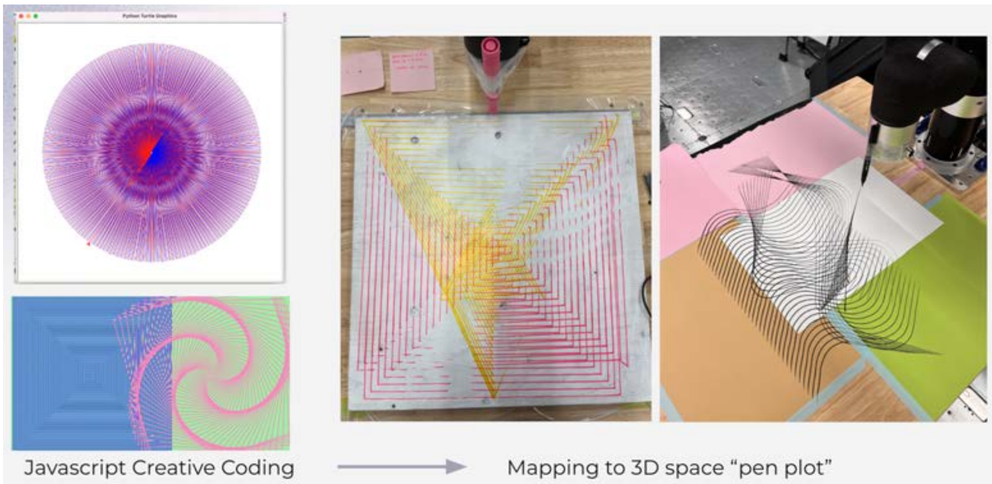

Figure 3.3 My previous creative code drawings of geometric patterns from math functions. Robotics enable bringing it to physical by mapping it essentially like a "pen plotter", which was the first step to getting INDRA to paint.

are speculative, experiential, and poetic.

My systems aim to function as such tools. They do not soothe public fear with polished automation, but instead, make visible the flux of shared authorship. They foreground the choreography between flesh and code, gesture and algorithm, asking: What happens to human agency when AI acts not for us, but with us, or through us? As Rosi Braidotti writes, "What is needed... is a new way of combining critique and creativity, theory and practice, to achieve a generative ethics for our technologically mediated lives" [64]. This thesis is a step in that direction.

3.2. Concept Design

This project began as a response to a shared sense of unease. When image generators like DALL·E were first released, artists around me expressed disconnection and frustration. They felt as though their creative identities were being pushed aside by systems that didn't understand the care, history, and emotion behind making art. Rather than arguing against the technology, I began wondering if we could shape it differently, toward connection rather than displacement.

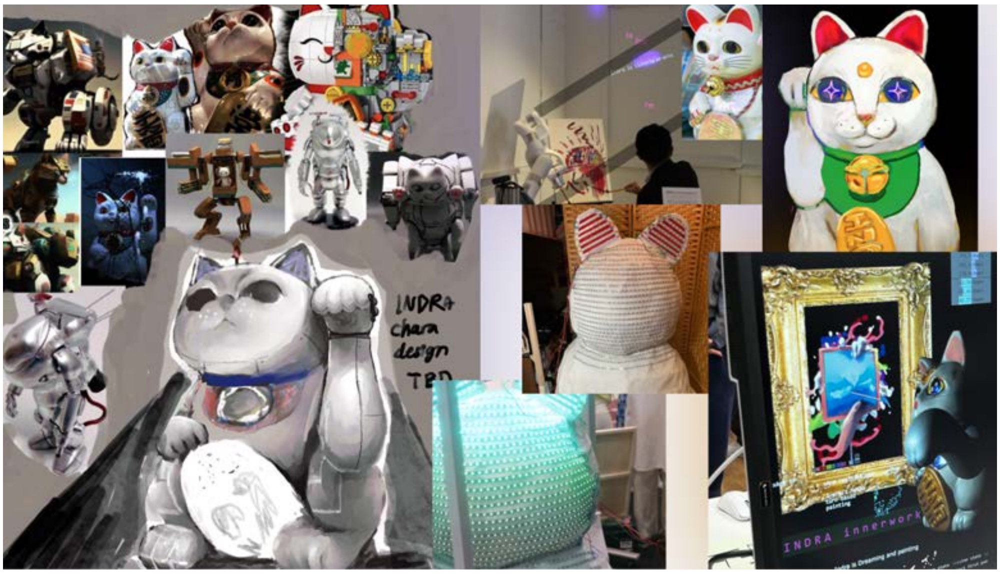

Figure 3.4 Indra's Cybernetic Avatar character mockup of a mecha maneki neko cat.

The first prototype was a simple digital canvas interface where users could draw with a mouse, then click a button to see the system respond with a generative pattern created through creative coding. It was a small gesture, but something about the exchange felt meaningful. There was a sense of control of timing, of mutual influence, and this reminded me of how the creative coding community treats code not just as a utility, but as an expressive medium. This small experiment became the seed for a much larger idea.

I envisioned a way to bring this digital co-creation into the physical world. What if a robotic system could collaborate with a person on a real canvas—not to replicate human creativity, but to participate in it? From this question, Indra was conceptualized: a painting robot designed to share authorship with its human partner through turn-taking and mutual response. Rather than trying to predict or complete a user's intentions, Indra waits, listens, and responds in its own visual language. Together, the human and the robot build a painting, layer by layer, in an evolving conversation of gestures. To make this feel more legible and emotionally resonant, I gave Indra a cybernetic avatar that appears on a nearby screen. This avatar doesn't pretend to be human, but it does help frame the robot's movements in familiar, embodied terms. People often speak about AI as if it were a being. Rather than resist this tendency, I treated it as a design material, creating a presence that can be interpreted and engaged with. The avatar allows the machine to communicate intention, pacing, and rhythm through its posture and gaze, making the act of collaboration feel more natural and alive.

While Indra explores creative expression, Asura was designed to support embodied skill. Where Indra co-creates on a canvas, Asura co-performs in motion and asks whether intelligent systems could extend, rather than replace, our physical abilities. It assists in subtle ways. When a puck comes close to a gripper, the robot automatically catches it. If a user is about to miss, the system steps in with soft assistance just enough to recover without taking full control. Visual cues support timing and orientation, letting the user stay focused on rhythm and flow. The experience is playful, sometimes challenging, and often surprising. Asura was built to show that AI can enhance physical coordination in ways that still leave the user in charge.

At their core, both systems are grounded in the same belief: that creative and physical skill are worth preserving, expanding, and celebrating. Indra offers a path for shared visual authorship, where the AI participates in the poetic language of mark-making. Asura reveals how bodies and machines can move together in collaborative action. Together, they suggest that intelligence, when shaped with care and imagination, can be a companion to human expression—rather than its competitor.

3.3. INDRA: Interactive Deep-Dreaming Robotic Artist

3.3.1 System Overview

INDRA (INteractive Deep-Dreaming Robotic Artist) is a collaborative AI art system that co-creates paintings with a human user. The system is designed around a turn-based interaction model, where a robotic arm and a human participant take sequential turns painting on a shared canvas. Indra processes visual input from the evolving artwork and interprets verbal or gestural cues to inform its next

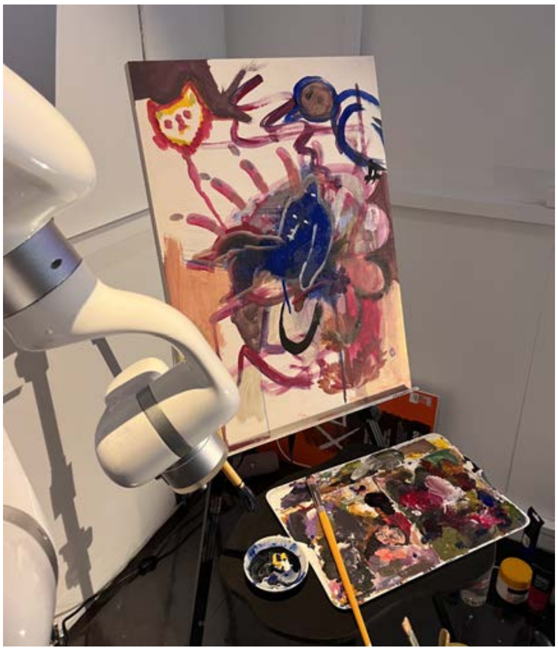

Figure 3.5 .

move. This physical and temporal layering creates a shared visual narrative that is incrementally shaped by both agents.

Architecture: Indra employs a layered control architecture inspired by subsumption design principles, in which low-level sensorimotor feedback operates continuously beneath higher-level modules responsible for decision-making, visual analysis, and stroke generation. The system integrates a UFactory xArm7 robotic arm, which is controlled via a real-time interface built in TouchDesigner. TouchDesigner serves as the primary environment for managing multimodal input and output, including canvas state monitoring, projection feedback, and user interaction channels. For visual analysis, Indra leverages pre-trained generative models, such as Stable Diffusion and ControlNet derivatives, to interpret the developing composition and assess semantic layout. In parallel, a secondary language module parses spoken input to extract narrative or stylistic intent, informing the AI's next action during its turn in the co-creation process.

Interface Mechanisms: Users can interact with Indra in three primary ways: physically painting, drawing on a digital tablet, or speaking short prompts. Indra translates these signals into brushstroke behaviors, selecting position, color, and

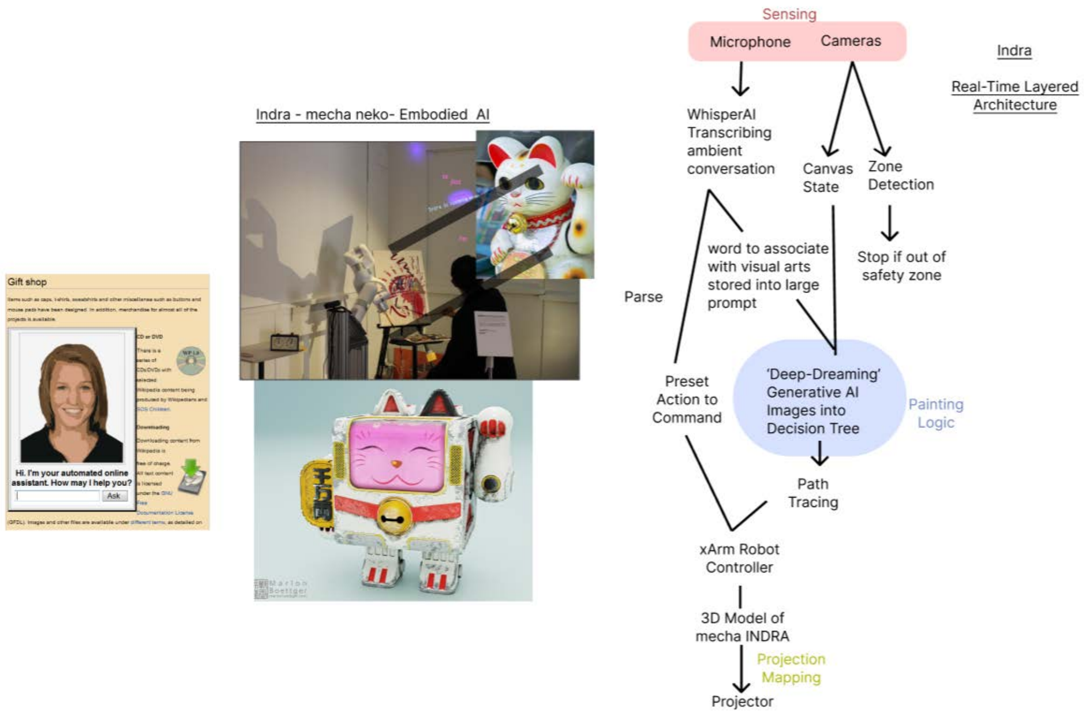

Figure 3.6 Diagram of avatar basis for Indra and initial architecture used in the Encounters Exhibition.

Figure 3.7 xArm 7, a 6 degree of freedom robotic arm by uFactory.

Figure 3.8 Screen captures from Indra Siggraph Asia 2024 supplementary video depicting turn-based interaction.

gesture style. A display next to the canvas provides a visual overlay of the AI's "thoughts," rendering transparency into its decision-making.

Design Rationale: The system was deliberately designed to preserve human authorship by limiting the AI's autonomy and giving users visible control over the progression of the painting. The use of turn-taking structures and analog input reinforces the user's role as primary author, while the AI acts more as a narrative and stylistic amplifier.

3.3.2 User Interaction

Indra offers a versatile and inclusive user experience, dynamically adapting its behavior through sensor data and real-time user inputs. This includes visual feedback, voice commands, and physical interactions with the canvas. The tablet interface provides intuitive controls, featuring module buttons and a drawing window for an additional control layer, enabling artists to directly take command of the robot arm. With its segmented design, Indra accommodates a wide range of users, enhancing the creative process for trained and novice artists. For experienced artists, Indra provides dynamic feedback and real-time composition analy-

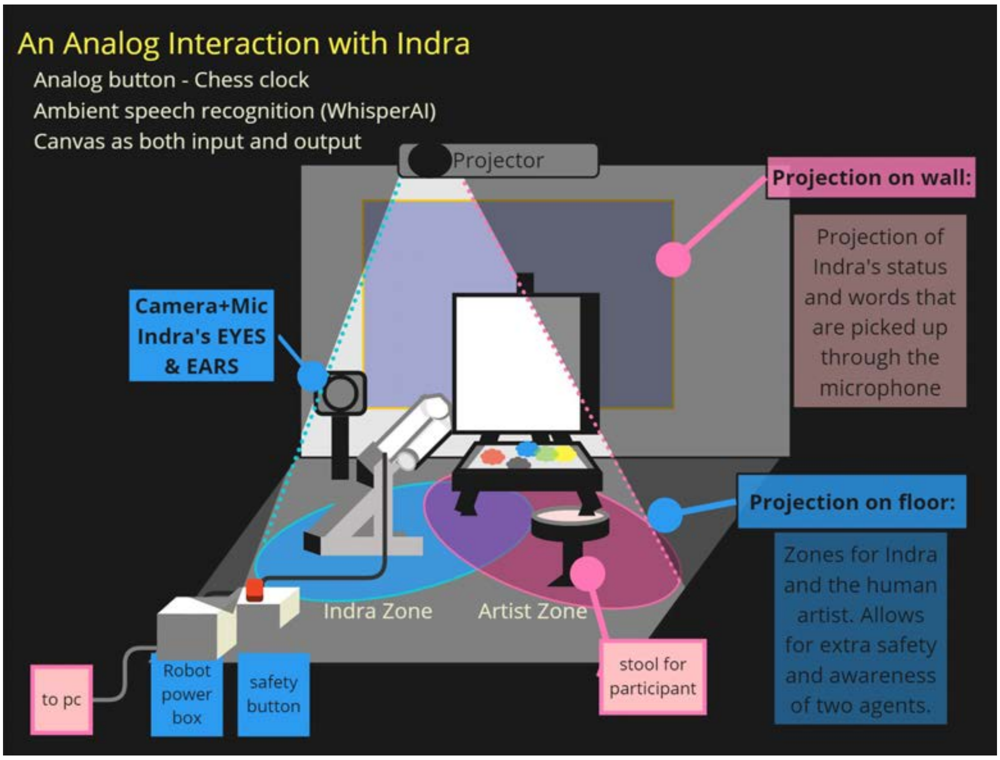

Figure 3.9 Diagram showing in-depth system layout of spacial design for Encounters Exhibition of Indra.

Figure 3.10 Zones from the robot arm and the users rolling stool, with overlap where the canvas stands, representing two creative agents merging.

sis, introducing elements that inspire innovation and expand artistic possibilities. Digital artists with limited traditional painting experience can bridge the gap to conventional media, using familiar digital tools to create physical works. For children, Indra transforms painting into a playful, educational experience, teaching concepts like color mixing and composition through interactive collaboration. Meanwhile, for artists with physical disabilities, Indra enables accessibility by translating digital tablet strokes into physical canvas movements, extending their creative reach and overcoming physical limitations. By customizing its features to suit a variety of user needs, Indra effortlessly integrates into diverse creative workflows, fostering a unique and engaging artistic journey.

3.4. Technical Implementation

The Indra system is designed as a real-time, multimodal robotic painting platform, integrating vision, audio, and user control into a responsive AI-augmented co-creative experience. This section outlines the core technical design of the system, emphasizing the dataflow, subsystem integration, and real-time control strategies used to support embodied AI-human co-creation. Below, each core subsystem is discussed in technical detail.

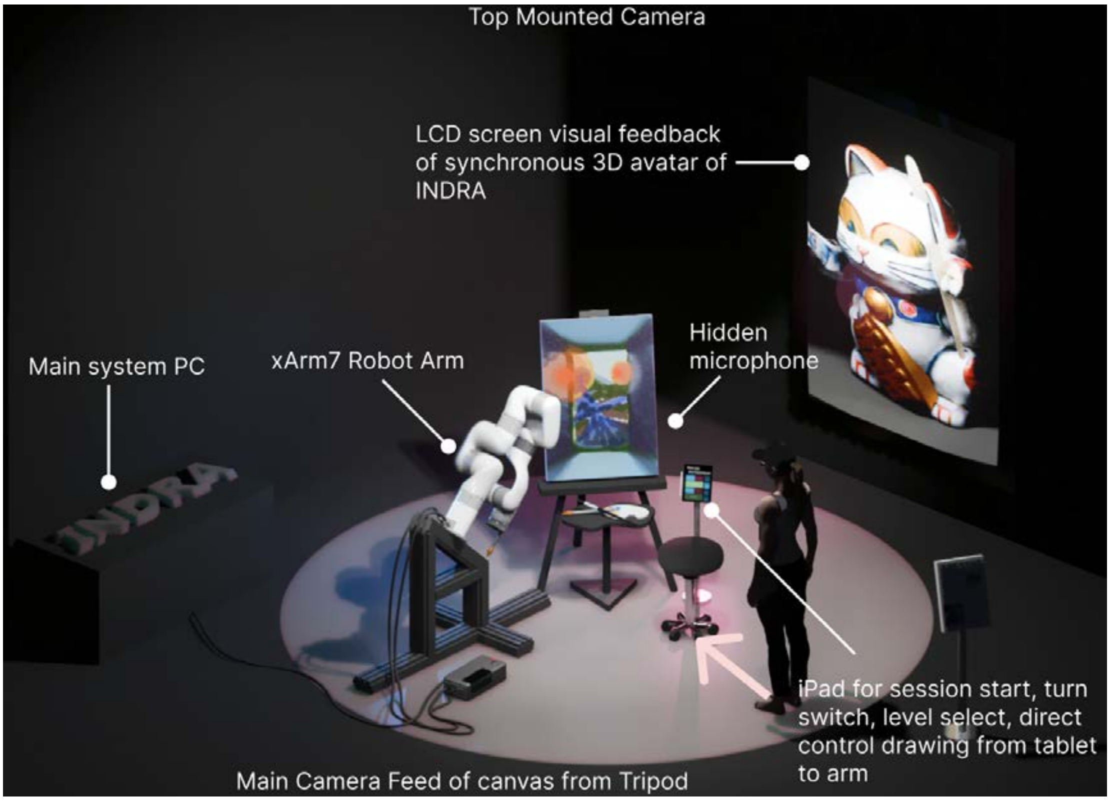

Figure 3.11 Labeled diagram of a rendering of Indra layout for CHI Interactivity.

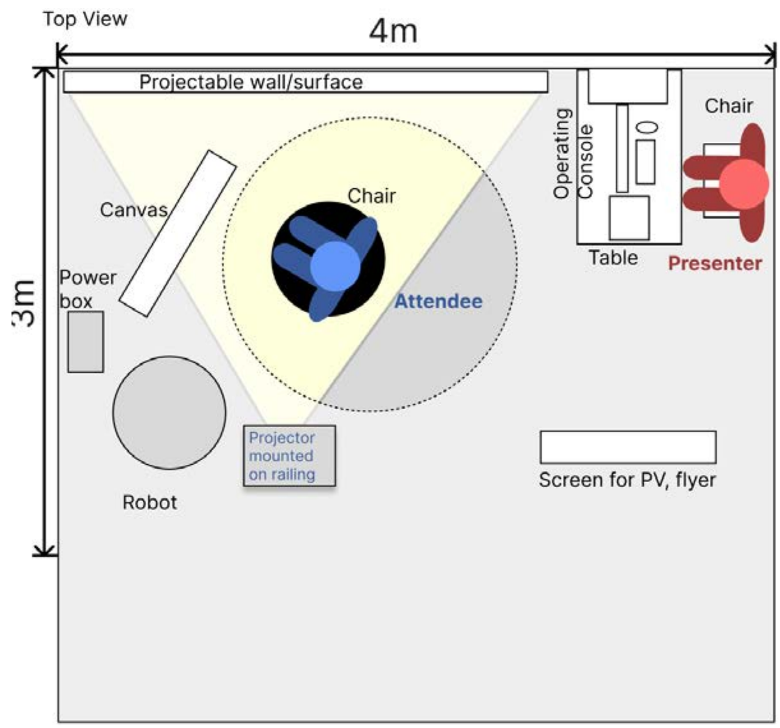

Figure 3.12 Top view layout of Indra, used during 3-day Encounters exhibition.

Figure 3.13 The xArm is mounted on a custom aluminum frame stand which tilts the arm forward 45 degrees on the x-axis.

The ideal requirements are a 3 by 4 meter space for safety and accessibility reasons. Setup takes 3 hours and requires 2 100V power supplies, with a projectable wall or surface available. If a ceiling mounting rail is supplied, the floor can also be used for projection. Venue lighting can be bright, but semi-darkness is preferred. A desktop PC runs main system while one laptop PC controls visual output and additionally the audio module to offload the main PC processing. The addition of the iPad in future iterations of Indra adds one Macbook laptop to enable wireless tablet drawing functionality.

3.4.1 Robot Arm Control via xArm Python SDK

At the heart of Indra's embodied interactivity is the xArm 6, a 6-degree-of-freedom robotic arm produced by UFactory. The robot is controlled programmatically via the xArm Python SDK, a Python-based software development kit provided by UFactory—to directly send commands to the robot. The xArm SDK provides both high-level and low-level control APIs for motion, gripper control, and safety states, and supports multiple modes including position, velocity, and servo-based control. In this project, TouchDesigner acts as the main control environment, serving as the real-time visual programming platform to integrate AI modules, scripts, sensor inputs, and robotic commands. TouchDesigner is connected to

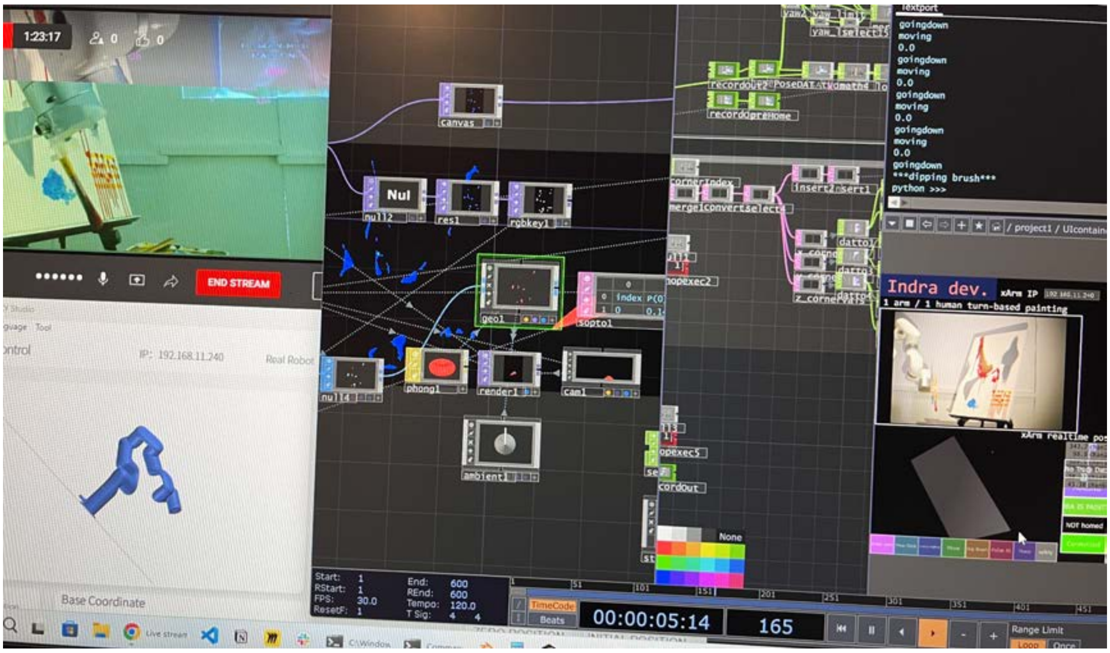

Figure 3.14 Snapshot of the TouchDesigner network and live feed during the Encounters exhibition.

the xArm using a custom "xArm Controller" network that translates planned motion paths into servo commands via the SDK. This enables sub-second level responsiveness and smooth actuation, which would be infeasible through standard robotic middleware such as ROS (Robot Operating System) due to its latency and non-realtime nature in this context.

Unlike middleware frameworks such as ROS (Robot Operating System), which emphasize modularity and abstraction but introduce latency, the xArm SDK allows for direct, low-latency command execution over TCP/IP sockets. This direct control is essential for real-time painting tasks, where micro-delays would result in stroke jitter, positional error, or overshooting.

A dedicated TouchDesigner component called the *xArm Controller* handles this communication. Within this network, user-defined Python scripts send joint space or Cartesian space commands at 30–110 Hz, depending on servo mode performance and system load. This module implements our custom path planning routine, which converts 2D drawing coordinates into mapped 3D trajectories within the robot's workspace.

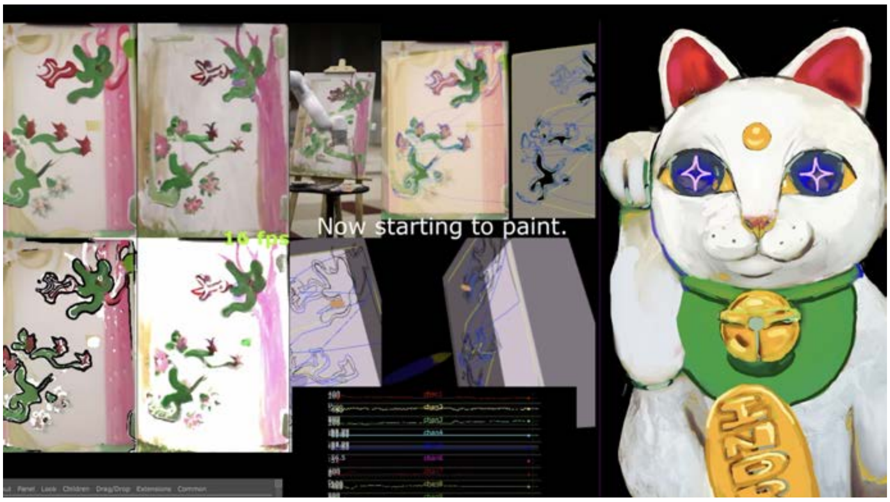

Figure 3.15 .

3.4.2 Custom Path Planning and Coordinate Mapping

Path planning in Indra is built from a minimal but effective algorithmic stack. Rather than relying on conventional path planning libraries like MoveIt or OpenRAVE, we developed a novel drawing-based trajectory planner within TouchDesigner. To render a 2D image into robotic painting strokes, the system first maps the 2D canvas plane into 3D coordinate space relative to the robot's base.

Input data (from user-drawn strokes or generated images) is first processed through the *TOPtoSOP* operator in TouchDesigner. This operator converts a raster image into vectorized contour data. The image is leveled into binary form using luminance thresholds and adaptive contrast filtering. The contours are then smoothed using a Chaikin-like curve simplification and simplified further to reduce redundant vertices. The resulting path is a 2D curve set, which is then mapped to a 3D surface. The canvas is treated as a planar surface in world coordinates, calibrated in advance using fiducial markers and manual pose estimation. The z-height for brush pressure is generated during path tracing and is calculated from the normal vector of the canvas to simulate lifting and pressing strokes, while stroke order is handled by a top-down "printer-like" heuristic for sequential efficiency and clarity.

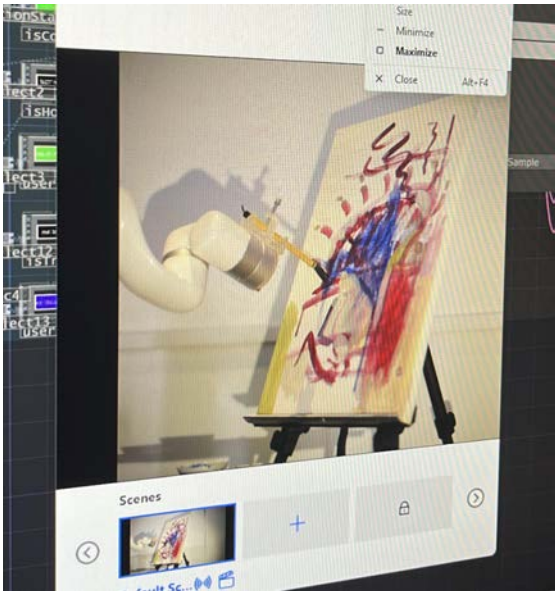

Figure 3.16 Live camera feed from the Canon EOS Webcam Utility during the Encounters Exhibition.

3.4.3 Vision Subsystem: Camera Input, Object Detection, and Stable Diffusion

Indra uses a Canon EOS 5D Mark III DSLR camera, connected via Canon's EOS Webcam Utility, to capture a high-resolution video stream of the canvas from a distance of 3 meters. The video is streamed into TouchDesigner for processing.

Two main vision-based AI models are used:

1. **YOLOv8 (Ultralytics):** Integrated into TouchDesigner using a Python-processed custom operator, *YOLOv8* is a computer vision model that performs object detection and classification on the camera feed in real-time. It identifies general objects (e.g., "tree", "house", "cat") within the canvas state and passes them to the language prompt generator. The words serve as semantic anchors, giving INDRA a basis for interpreting user intent and constructing narrative or compositional responses. Only the top N (N=5 by confidence score) noun objects are retained per frame. The detected objects are not corrected or filtered if a misclassification occurs, the system treats it as a meaningful prompt. This aligns with the project's design philoso-

Figure 3.17 Participant in deep focus during interaction with Indra.

phy, which embraces ambiguity and error as generative artistic possibilities, rather than noise to be eliminated.

2. **StreamDiffusion:** It's core, *Stable Diffusion* is a generative AI latent diffusion model. StreamDiffusion is a custom adaptation of Stable Diffusion for real-time use, implemented via a Python server that communicates with TouchDesigner. StreamDiffusion accepts textual prompts from both YOLO output and transcribed audio, generating imagined "next states" of the painting. The image is treated as a conceptual planning output, not a direct rendering.

To maintain performance, both YOLO and StreamDiffusion are run only once per turn to reduce GPU/CPU load, as opposed to continuous inference.

3.4.4 Audio Input and Natural Language Integration

Ambient audio is captured using a cardioid microphone mounted near the canvas. Audio is piped into the OpenAI Whisper ASR system, transcribed into timestamped natural language, and sent into TouchDesigner.

No speaker identification is performed. The system is intentionally designed

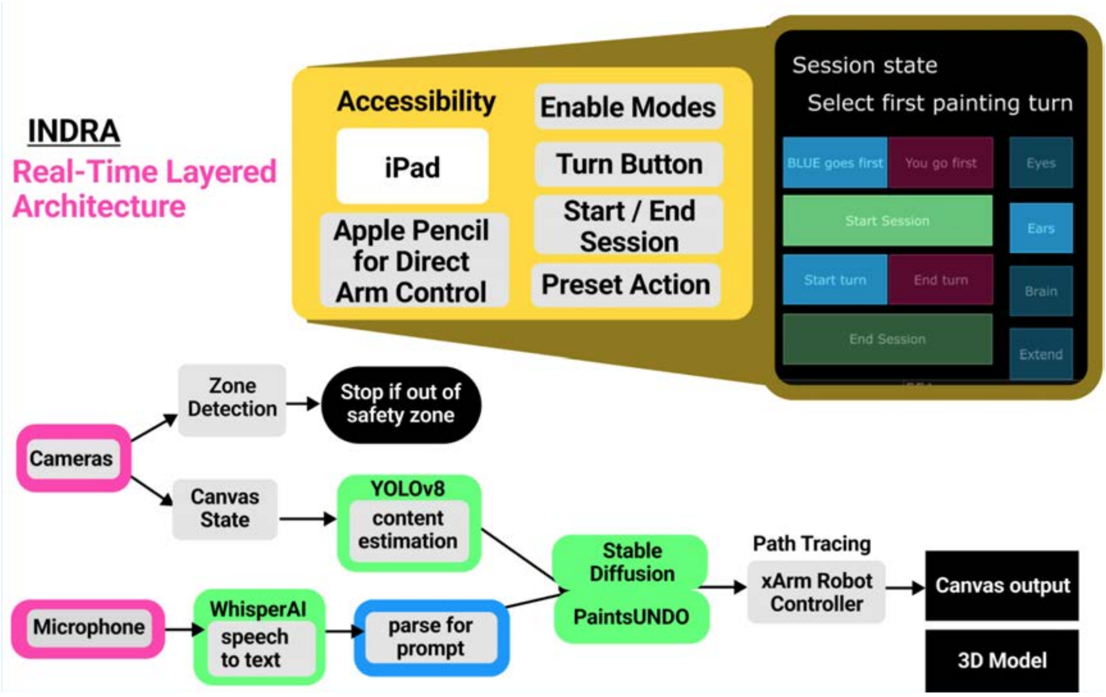

Figure 3.18 Flow diagram of Indra's real-time layered architecture and accessibility interface.

to treat all voices as valid contributors, simulating a shared studio environment. Transcribed words are parsed using part-of-speech tagging, and visual or artistic terms (e.g., "bright", "mountain", "add more contrast") are weighted higher in prompt generation. This enables open-ended collaborative input in a studio environment, where multiple viewers may make verbal contributions.

The transcription is passed into the same semantic processing pipeline as the object detection results, allowing the system to combine visual and verbal cues into a unified interpretation.

3.4.5 Multimodal Prompt Fusion

YOLOv8 outputs and Whisper transcriptions are combined using a context-weighted prompt assembler. The system constructs a single prompt string that summarizes both sources, favoring visual descriptors and nouns. This string becomes the input for StreamDiffusion, which generates a visual hypothesis of the next potential

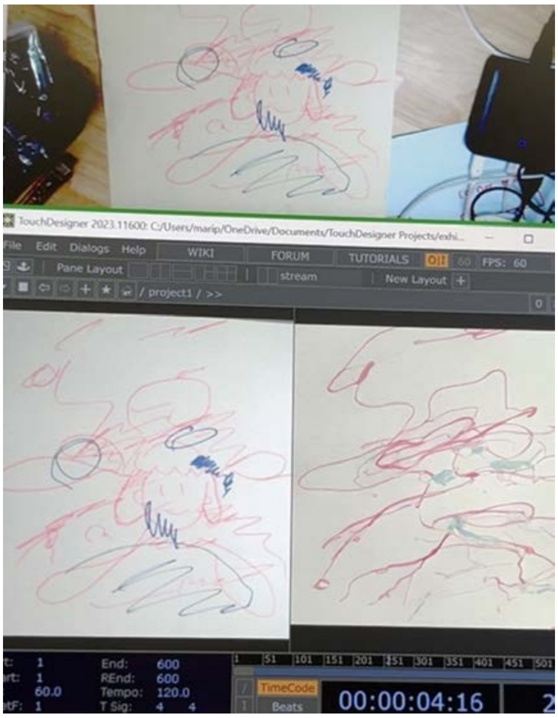

Figure 3.19 Preview of extracted image from Stable Diffusion done on scribble drawing.

painting direction. The prompt string might resemble: *"A surreal landscape with mountains and soft brush textures, inspired by the word 'dreamlike' and a detected object 'tree'."* This fusion ensures that the system's output is grounded in both perception and conversation, forming a bridge between human language and visual intent. Next, we generate an imagined "next state" of the canvas.

3.4.6 Generative Vision: StreamDiffusion

To synthesize its next painting step, Indra uses StreamDiffusion, with prompts set to include the transcribed speech and YOLOv8 detections.

The result is a 512x512 pixel image reflecting what Indra "thinks" should come next based on the human-AI dialogue. This generated image is not painted directly. Instead, it undergoes a series of transformations to make it paintable:

1. **Image Binarization and Stroke Extraction:** The image is converted

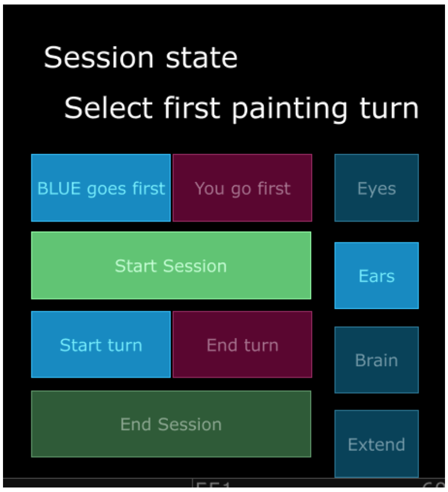

Figure 3.20 Initial UI of session state selector. User can select if they want to take the first turn.

into black and white, and processed using TouchDesigner's TOPtoSOP node, which extracts vector outlines from the image.

2. **Path Simplification and Smoothing:** These outlines are smoothed to remove visual artifacts and reduce complexity. Stroke paths are sorted in a heuristic order (top to bottom, left to right) similar to a printer to reduce arm travel time.

3. **Motion Planning:** The simplified paths are converted into 3D coordinates relative to the canvas. These coordinates are sent to the xArm controller for real-time execution.

3.4.7 Interaction Model and Turn-Based Control

Indra's core HCI logic is structured around turn-taking. A physical "Turn" button on the tablet interface lets the user end their current action and pass control to Indra. Conversely, Indra automatically returns control to the user once it completes a drawing task. This model creates a rhythm of interaction rather than continuous co-activation.

The rhythm of interaction is entirely user-driven:

Figure 3.21 Evolution of the canvas from turn taking between user and Indra.

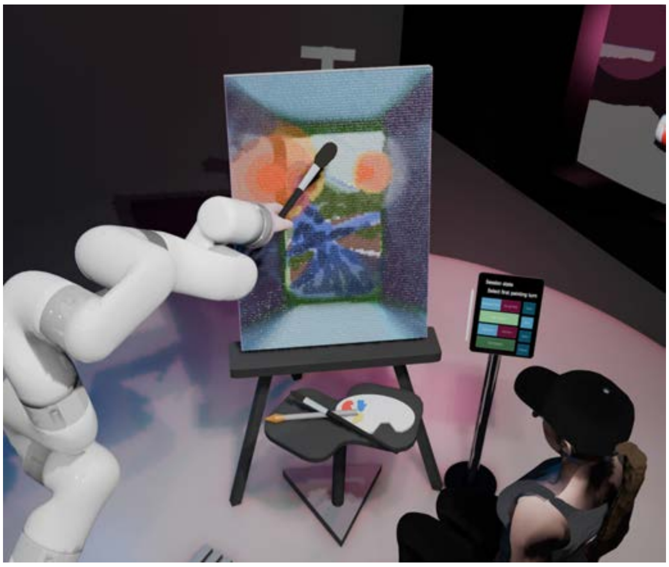

Figure 3.22 View from close-up of how the user is able to share the canvas space with Indra in a safe proximity, and the placement of the tablet.

1. Users can interrupt Indra at any time.
2. Users can take many turns in rapid succession or choose to let Indra operate more autonomously.
3. A full session is capped at 15 minutes for practical demonstration purposes, but within that window, control boundaries are fluid.

This turn model is flexible where users can take control at any time and interrupt Indra's actions. This level of override capability supports meaningful human control and creative pacing. All of Indra's input modalities—vision, audio, and

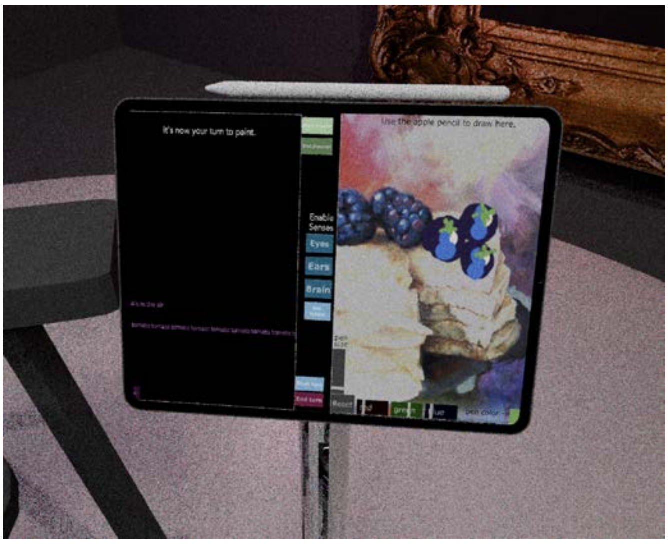

Figure 3.23 iPad interface used for tablet drawing, module activation, and turn switching.

tablet—are synchronized within TouchDesigner using shared data operators and a logic controller that governs when each module activates.The tablet UI also contains toggles to activate or deactivate each modality:

- **Brain:** Enables/disables StreamDiffusion and YOLO.
- **Eyes:** Activates/deactivates the camera feed.
- **Ears:** Toggles microphone input and Whisper.
- **Arm:** Switches the robot to accept direct drawing input from the tablet.

Tablet drawings are processed similarly to AI-generated images—converted via TOPtoSOP and path-planned in 3D space. The digital canvas (iPad) enables users to draw strokes digitally using an Apple Pencil, which can then be sent to Indra for physical replication. This allows users with physical limitations to co-create without touching the canvas directly.

3.4.8 Real-Time Responsiveness and Latency Management

The system maintains a 30–60 FPS update cycle in TouchDesigner. The xArm servo control is run asynchronously at 30–110 Hz. The system introduces a small delay buffer (150–200ms) in path execution to smooth motion and avoid stutter. This delay is visually justified via the avatar's behavior and visible planning lines. Internal collision detection and spatial boundary constraints are coded using joint limit checks and planar workspace masks. A jolt detection routine monitors accelerometer data and frame skips, activating a safety stop button if thresholds are exceeded.

3.4.9 Failure Modes and System Robustness

When an input modality fails (e.g., no microphone access or Whisper crash), its corresponding UI toggle is disabled. Users are encouraged to re-enable it. Indra does not retry or self-correct failure but allows graceful degradation. For example, if no speech is detected, only vision is used for the prompt. If no camera feed is available, the system waits for direct drawing input. This fault-tolerant, modular fallback strategy emphasizes agency over automation.

3.5. Design Iteration

The following subsections document four key deployment contexts Indro prototype was exhibited and tested. Each setting offered distinct participant demographics, spatial constraints, and experiential framing, allowing for diverse insights into system performance and user perception.

3.5.1 Encounters Exhibition

Early prototypes of Indra featured no digital interface between the robot and the user. Turn-taking was mediated by a physical chess clock, which was monitored via webcam and computer vision to detect when it was Indra's turn to act. This fully analog setup emphasized co-presence and temporal pacing but revealed significant limitations in interaction flexibility, particularly during public exhibitions. At the

Figure 3.24 Participant at Encounters exhibition looking to the wall after taking a turn painting with Indra to see the ambient speech that was audibly picked up by the system and projected onto the wall.

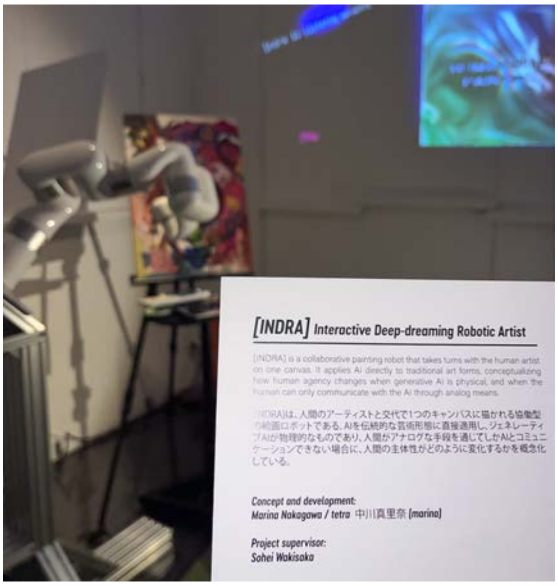

Figure 3.25 Description panel of the first public version of Indra at Encounters.

Figure 3.26 Touchdesigner system (left), me setting up Indra (middle), and a user painting simultaneously at Encounters.

Encounters exhibition, participants often struggled to understand the analog logic of the system.

3.5.2 Open Feedback Reflections

Participants were given the opportunity to leave open-ended comments after their session with Indra. Many chose to reflect on the nature of collaboration, authorship, and emotional response to the interaction. A recurring theme was the sense of shared creation between human and machine. One participant remarked, "Really interesting experience, made me think of co-authorship and human–non-human collaboration." Another wrote:

> "I felt it was a collaboration. The robot felt like a single individual, like a human."

The question of authorship emerged repeatedly. Several users described the interaction as improvisational and open-ended, with one participant noting:

> "I had the sense that it was augmenting my symbolic marks (circles, lines, etc.) but it never turned into something specific. That might be the essence of co-creation with a non-human."

Another commented on the emotional resonance and spontaneity of the exchange:

> "Because I believe creation is born through relationships, I was excited to see what the arm's lines, shapes, and colors would make me feel, and how I would respond. Painting one image together, improvisationally,

Figure 3.27 Painting session at Encounters exhibition where projected words that the system picks up from the users utterances are shown on the wall as text.

> made me wonder whether the process itself is what truly matters. What does it mean for a painting to be finished?"

Participants also described the experience as *emotionally evocative and strangely introspective*, with phrases like:

> "Very strange feeling where I try to find myself.",
> "Being seen.",
> "Nice dramatic music—the robot feels more powerful than me."

Others reflected on the newness of the experience, with one participant stating:

> "I thought it was interesting that the AI draws in response to my movement and the surrounding sounds."

Some saw INDRA as an *expressive agent*:

> "Indra is a fascinating creator, she has as much agency and creativity as any of us. I hope to see her art journey grow and evolve."

Overall, these qualitative responses indicate that participants not only engaged with Indra as a tool but often regarded it as a collaborator, interpreter, or even performer, suggesting the system succeeded in fostering a sense of mutual engagement and reflective co-creation.

3.5.3 Miraikan Open Lab

During the Miraikan Open Lab, accessibility limitations became increasingly apparent. Children and a participant using a wheelchair expressed interest in engaging with Indra but faced challenges due to the physical layout of the system. The canvas height and positioning made it difficult for those with limited reach or mobility to comfortably participate.

These encounters directly informed subsequent design modifications. Observing a wheelchair user express fear and hesitation around the robot highlighted the need for more accessible interaction modes. In response, a tablet-based drawing interface was introduced, allowing users to contribute digitally and have their strokes physically rendered by INDRA. This feature enables participation without

Figure 3.28 Indra paintings made from each day of the Encounters exhibition.

needing to be in close proximity to the robotic arm, expanding access for users with different physical abilities.

Additionally, when children engaged with INDRA, many referred to it as a character or being, asking what “she” looked like. These interactions inspired a more visual representation of Indra’s presence, leading to the development of an animated facial interface. By giving Indra a more expressive identity, the system became more approachable, especially for younger participants, and reinforced the sense of collaboration with a co-creative partner.

3.5.4 KMD Forum

These experiences directly informed the development of a more inclusive and adaptable interface. At the KMD Forum, we introduced Indra Innerwork, a digital-only version of the system that excluded the robotic arm. In this iteration, users drew on an iPad and observed the AI’s contributions rendered on a shared digital canvas in real time, simulating a co-creative exchange in a more accessible format.

Figure 3.29 A child drawing with Innerwork:Indra on an iPad, showing a rose illlustration

3.5.5 Summary

This chapter traces the evolving journey of Indra, an embodied AI collaborator that blurs the boundaries between human creativity and machine agency. From its humble beginnings as a fully analog turn-taking experiment to a refined, accessible digital interface, Indra represents a vision of artistic partnership that embraces diversity, inclusion, and shared authorship.

Each deployment, whether in public exhibitions or international conferences, has revealed new ways people engage with AI as a creative partner. Challenges found in early prototypes led to innovations that not only improved usability but also helped users feel more in control and connected with the system.

The upcoming proof of concept study builds on these rich experiences. Through this work, Indra hopes to inspire a future where human and machine creativity work together, with technology enhancing rather than replacing the unique human spark of imagination and agency.

Chapter 4
Proof of Concept

This section outlines the planned validation study focusing solely on Indra to assess its effectiveness in supporting human agency, creative authorship, and trust during real-time embodied human-AI collaboration. While Asura was developed earlier as a technical prototype to investigate control paradigms, only Indra was used in the final user-facing study. Insights from previous Indra and Asura tests informed the design of this study, which targets a refined evaluation of Indra's interaction modalities. The primary goal is to understand how system design elements such as transparency, responsiveness, and turn-taking influence users' perceived control, authorship, and acceptance of AI as a collaborative partner.

4.1. Purpose

To evaluate how users perceived agency, authorship, and collaboration with Indra, the installation incorporated a mixed-method user study. Participants engaged in the co-creation session, and followed with a post-questionnaire. Additional qualitative data was gathered through short interviews and observational field notes, focusing on users' emotional responses, interpretations of AI behavior, and overall impressions of working with a robotic creative partner.

CHI Interactivity served as a critical field deployment for assessing Indra's usability, legibility, and inclusiveness in a high-traffic, international context. The diversity of participants—ranging from professional designers and HCI researchers to children and general attendees—allowed for rich variation in perspectives. The event also marked the first full integration of Indra's accessible interface design, including tablet-based drawing and visual feedback systems that accommodated users with different levels of artistic skill and physical ability. These adaptations

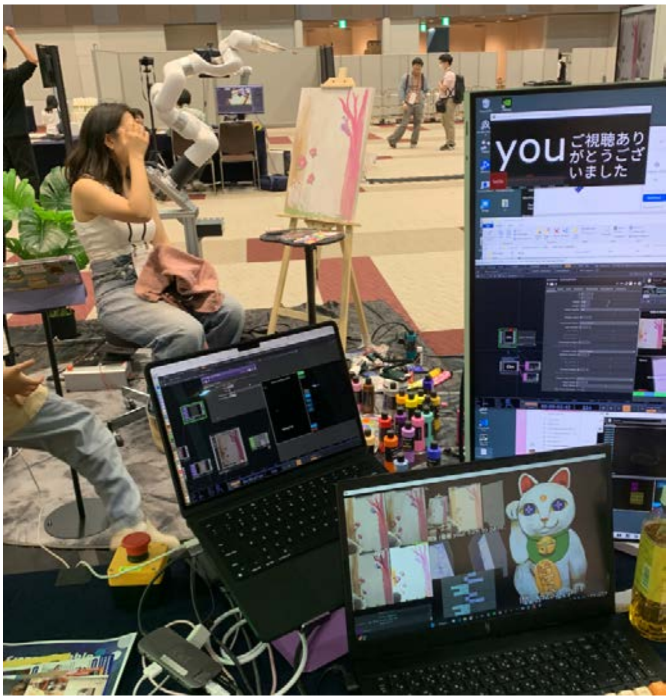

Figure 4.1 Three computer setup running Indra system at CHI.

directly supported the thesis's broader goal of designing collaborative AI systems that prioritize transparency, adaptability, and human agency.

4.2. Content

At the 2025 ACM CHI conference on Human Factors in Computing Systems, Indra was presented in the Interactivity session as a live, embodied co-creation demonstration during a three-day public exhibition held in Yokohama, Japan. The installation featured a dedicated painting station composed of a canvas, a uFactory xArm7 robotic arm, a large LCD display with a synchronized 3D avatar, and a tablet interface. Visitors were invited to engage in 5–7 minute sessions of collaborative painting with Indra, alternating turns with the robotic system using a touchscreen-based control panel.

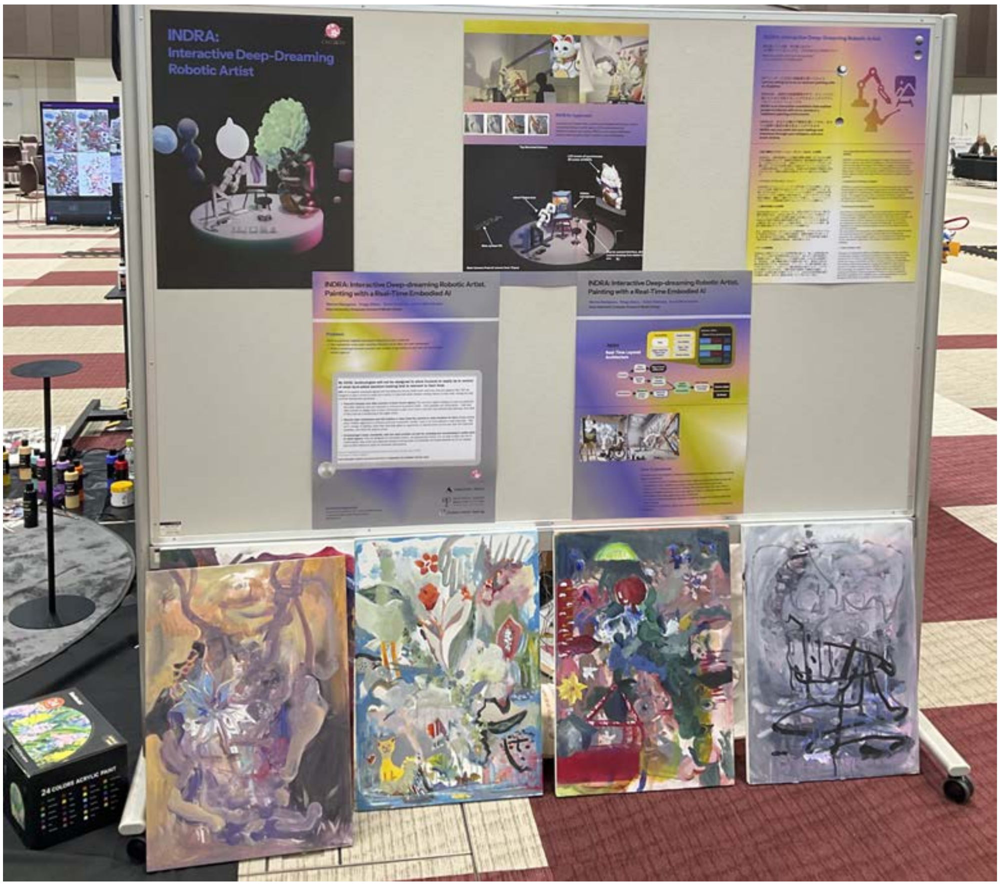

Figure 4.2 Facade of Indra demo booth showing posters and completed paintings.

4.2.1 Ethical Considerations

All participants will be provided with informed consent forms that outline the nature and purpose of the study, their right to withdraw at any time, and how their data will be anonymized and stored. Special accommodations will be made for participants with physical or cognitive disabilities, ensuring equitable access to each system. No personal identifiers will be collected, and all interview data will be pseudonymized. The study complies with institutional ethics protocols and will undergo formal review prior to commencement.

4.3. Methodology

The study involves eight participants aged 20s to 60s, balanced between those with formal artistic backgrounds and non-artists. Participants will interact with INDRA in sessions designed to capture diverse perspectives on co-creative agency.

The study flow includes:

1. **Indra session:** 15-minute collaborative painting interaction.
2. **Questionnaire:** Demographics, experience with AI or art, baseline AI attitudes (ATTARI), Perceived agency (SoAS), trust (GAToRS), and user feedback.
3. **Semi-structured interview:** Reflections on agency, system behavior, emotional responses, and collaboration.

4.3.1 Participants

There are 8 respondents with an age range spanning from their 20s to 60s. Select participants are invited to respond after painting with Indra at the CHI Interactivity exhibition hall after interacting with the system for at least 15 minutes. The aim is to capture a range of perspectives from both experienced artists—such as painters, illustrators, or digital creatives—and non-artists, including those from unrelated disciplines or with no prior creative background. No prior experience with robotics or AI systems will be required, and accessibility needs will be accommodated as necessary.

Items

	Wording	Facet	Valence
1	*AI* will make this world a better place.	*Cognitive*	Positive
2	I have strong negative emotions about *AI*.	*Affective*	Negative (reverse-coded)
3	I want to use technologies that rely on *AI*.	*Behavioral*	Positive
4	*AI* has more disadvantages than advantages.	*Cognitive*	Negative (reverse-coded)
5	I look forward to future *AI* developments.	*Affective*	Positive
6	*AI* offers solutions to many world problems.	*Cognitive*	Positive
7	I prefer technologies that do not feature *AI*.	*Behavioral*	Negative (reverse-coded)
8	I am afraid of *AI*.	*Affective*	Negative (reverse-coded)
9	I would rather choose a technology with *AI* than one without it.	*Behavioral*	Positive
10	*AI* creates problems rather than solving them.	*Cognitive*	Negative (reverse-coded)
11	When I think about *AI*, I have mostly positive feelings.	*Affective*	Positive
12	I would rather avoid technologies that are based on *AI*.	*Behavioral*	Negative (reverse-coded)

Figure 4.3 12 Question items from the ATTARI-12, Attitudes towards AI questionnaire.

4.3.2 Data Collection & Analysis

Quantitative data will be collected through the following validated instruments:

- **Sense of Agency Scale (SoAS)** to measure perceived control and authorship
- **ATTARI-12** to evaluate attitudes toward AI in collaborative contexts
- **GAToRS** to assess trust in robotic systems

Qualitative data will be obtained through observational notes, online feedback form, and post-session interviews. These interviews will be dynamic and open-ended, focusing on user interpretations of system behavior, perceived authorship, emotional reactions, and reflections on co-creation.

Post-Session Questionnaire

After the session, participants complete an online post-experience questionnaire designed to capture system-specific perceptions.

1. Demographic data (age, gender, occupation, artistic background), prior experience with AI or robotics.

2. A short version of the ATTARI-12 scale to measure initial affective and cognitive responses to AI technologies in creative contexts.

3. The Sense of Agency Scale (SoAS), adapted for HCI contexts, measuring the extent to which participants felt in control of the co-creative process.

4. The GAToRS scale, assessing trust in the robotic system's behavior and responsiveness.

Interview

After completing their session with Indra and the accompanying questionnaires, participants were invited to provide open-ended reflections in a free-response section of the post-questionnaire. Additionally, informal conversations were conducted on-site to capture further qualitative impressions. These discussions were not audio-recorded due to the fast-paced and public nature of the exhibition environment, but researcher notes were taken immediately afterward.

Prompts centered on perceived agency, emotional experience, and interaction with the system. Example questions included:

- "How much of the outcome do you feel reflects your own creativity?"
- "Did the system feel like a tool, a partner, or something else?"
- "Were there moments of surprise or confusion?"
- "How would you describe the collaboration process?"
- "How do you imagine such systems might be used in the future?"

4.4. Results

During the three-day CHI Interactivity exhibition, participants created a series of collaborative paintings with Indra. Across the works, several common visual elements appeared. Many paintings featured organic imagery such as flowers, trees, eyes, and animal-like forms. These figures were often surrounded by expressive abstraction, including layered brushstrokes, varied textures, and bold color

Figure 4.4 Three paintings made between Indra and participants, one for each day of the CHI exhibition.

combinations. The compositions show a balance between intentional forms and spontaneous mark-making, suggesting that participants were responding to both their own ideas and Indra's contributions. Instead of polished final pieces, the paintings reflect an ongoing process of experimentation and shared authorship. This aligns with the system's goal of supporting improvisational, co-creative art-making rather than producing refined outputs.

Analysis of post-questionnaire data revealed subtle trends across age and gender. Sense of agency (SoAS) scores increased slightly with age, suggesting older participants felt more in control during co-creation with Indra. Female participants reported higher agency than males, while males showed slightly higher trust in AI interpretability (ATTARI). General attitudes toward robots (GAToRS) remained consistently positive across all groups, indicating broad receptivity to embodied AI systems regardless of demographic differences.

The General Attitudes Toward Robots (GAToRS) responses reveal a nuanced perception of AI and robotic systems among participants. Overall, participants expressed strong support for the use of robots in assistive and complementary roles. Statements such as "Robots are a good thing for society, because they help people" and "Robots are necessary because they can do jobs that are too hard or too dangerous for people" received predominantly high agreement, indicating a broadly positive outlook on the societal benefits of robotics.

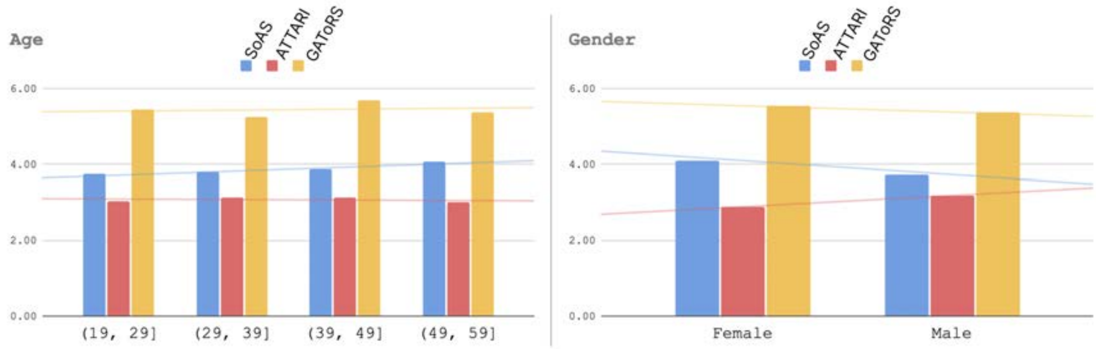

Figure 4.5 Left bar graph showing age group differences of average responses to the three scale questionnaires. Right bar graph looks at gender differences.

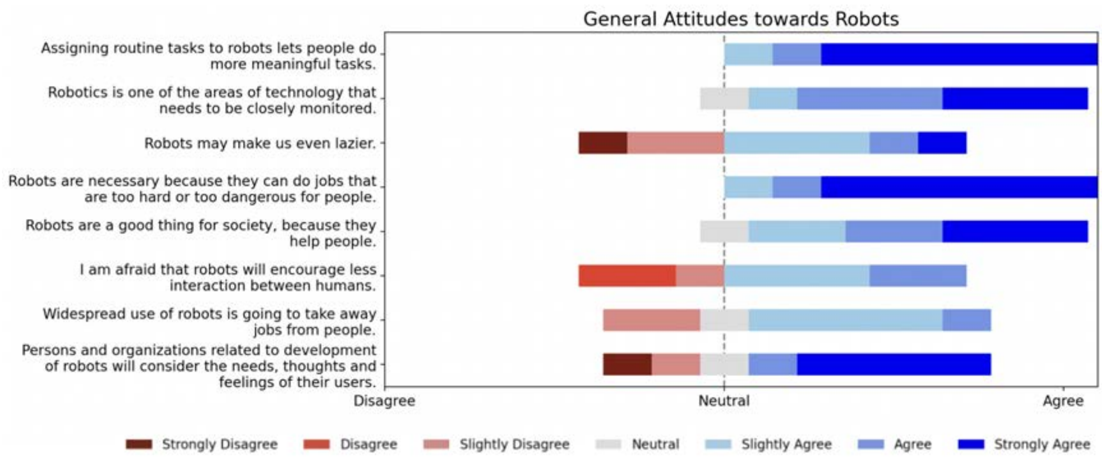

Figure 4.6 Bar graph of responses to the GAToRS (General Attitudes Toward Robots) scale. Each item reflects a different attitude toward robotics, with responses ranging from 1-7 likert scale from strong disagreement (dark red) to strong agreement (dark blue)

There was a clear recognition of the need for caution. The statement "Robotics is one of the areas of technology that needs to be closely monitored" also received high agreement, reflecting a cautious optimistism where participants welcomed robotic technologies but believed their development should be carefully overseen.

Responses were more mixed on issues of behavioral and societal displacement. For example, participants were divided on whether "robots may make us even lazier" and whether "robots will encourage less interaction between humans." This suggests a degree of ambivalence, with participants expressing excitement about the creative possibilities of embodied AI while also remaining sensitive to the risks of human passivity or social detachment.

Concerns over automation's impact on labor were present but not dominant. While some agreed with the statement that "robots will take away jobs," many participants responded neutrally or disagreed, pointing to a moderate level of anxiety around job displacement.

Finally, there was strong endorsement for human-centered values in the development of robotics. Most participants agreed that "persons and organizations related to development of robots will consider the needs, thoughts and feelings of their users," indicating a general trust in the ethical direction of the field. Responses to the ATTARI-12 scale indicate generally positive attitudes toward AI. Participants agreed with statements such as "AI will make this world a better place," "AI offers solutions to many world problems," and "I look forward to future AI developments," suggesting optimism about AI's societal role. Most disagreed with negative items like "I am afraid of AI" and "AI has more disadvantages than advantages," suggests that participants did not view AI as threatening or harmful. While some ambivalence was observed in responses to statements concerning AI-related risks, such as "AI creates problems rather than solving them," the overall response pattern reflects a forward-looking, open stance. This suggests that participants engaged with INDRA in a context of trust and curiosity, rather than fear or skepticism, which may have contributed to the system's positive reception.

Participant responses to the post-session Sense of Agency Scale (SoAS) indicate a balanced and multifaceted experience of control during interaction with Indra. Most users agreed with positively framed items such as "I was completely responsible for everything that results from my actions" and "I was in full control of

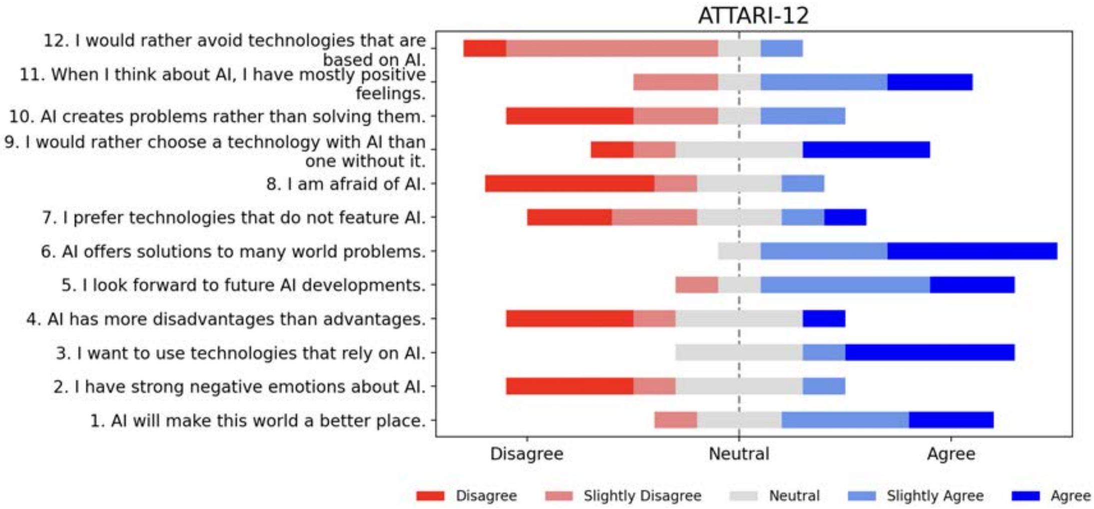

Figure 4.7 Participant responses to the 12 items of the ATTARI-12 scale (1-5 disagree to agree scale), which assesses attitudes toward artificial intelligence (AI).

what I do," suggesting that participants generally maintained a strong sense of personal authorship and control while co-creating with the system. At the same time, several responses indicate that this agency was not experienced as total or singular. The item "I am the creator of this painting" received a broad spread of responses with no extreme consensus and supports the notion of shared or distributed agency.

Importantly, strongly negative statements such as "Nothing I did is actually voluntary," "I was just an instrument in the hands of somebody or something else," and "While I was in action, I felt like I was a remote controlled robot" were largely disagreed with. This suggests that even though users were collaborating with a semi-autonomous AI system, they did not feel overridden or manipulated. However, neutral to slightly positive agreement with statements like "My movements were automatic—my body simply makes them" and "The outcomes of my actions generally surprised me" indicate some degree of emergent behavior or unpredictability, likely arising from the improvisational nature of the collaboration. These findings reflect a dynamic interplay between control and surprise that characterizes creative human-AI interaction. Together, the SoAS responses confirm

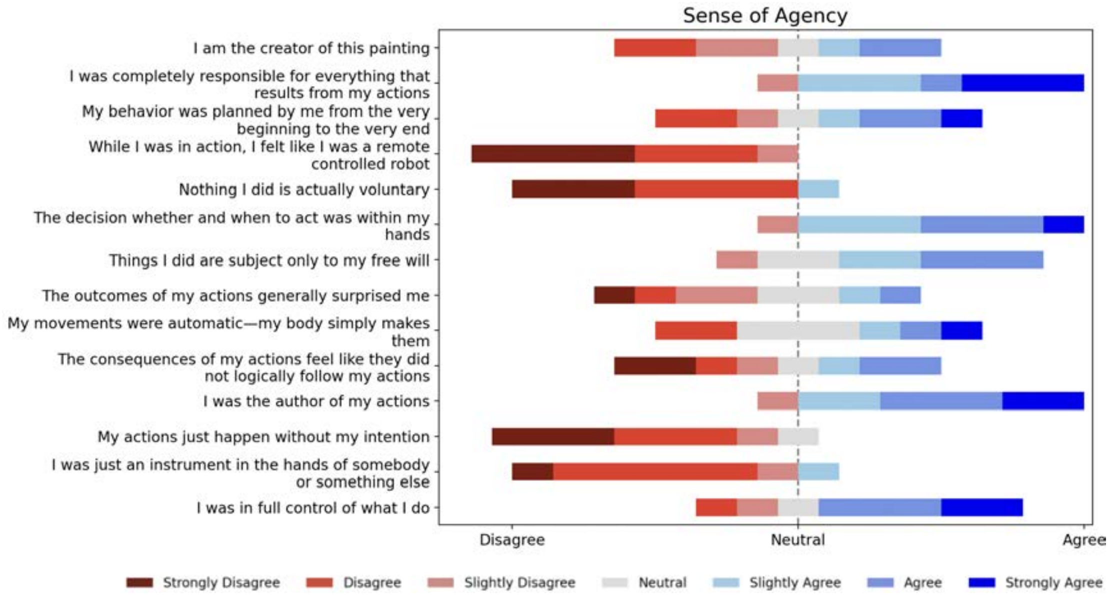

Figure 4.8 Bar chart presents participant responses to the Sense of Agency Scale (SoAS) following interaction with INDRA. Items are rated on a 7-point scale from strongly disagree (dark red) to strongly agree (dark blue).

that users maintained a meaningful sense of agency throughout their engagement with Indra, while also experiencing moments of shared authorship and creative emergence.

Participant Reflections

In addition to the quantitative measures, open-ended feedback from participants offered valuable insights into their perceptions of authorship, system behavior, and the nature of collaboration. Several participants expressed enjoyment in observing the AI's interpretations and responses. One noted:

> "It was exciting to watch and see how the robot interpreted my drawing and reacted to it. It saw something different in my drawing than I intended, but that was part of the fun."

Others raised questions about the system's transparency, revealing ambiguity in how users perceived the underlying mechanisms:

Figure 4.9 Child and guardian painting with Indra at CHI.

> "It was not clear to me what the algorithm behind it was—is it AI or just hard coded?"

More philosophically oriented responses pointed to broader themes around authorship, ownership, and the evolving meaning of art when co-created with machines. As one participant wrote:

> "My interaction was a bit limited due to demo issues, but I started wondering: who owns the painting in this scenario? What does art mean when expression becomes more than human? How is this AI–human collaboration different from human–human collaboration? Also, is the power dynamic still leaning toward the human, since the system is ultimately built by humans and trained on preexisting datasets?"

These reflections suggest that even brief interactions with the system can provoke complex considerations of agency, authorship, and trust—supporting the premise that embodied, turn-based AI can serve not only as a technical tool but also as a catalyst for critical reflection on the nature of creative labor.

Chapter 5
Discussion

5.1. Overall Findings

This research aims to address two central questions: how human agency changes when artificial intelligence is embodied in physical form and engaged in real-time co-creation, and how such systems can be designed to promote a sense of collaboration, control, and authorship in skill-based tasks. Early interactions with Indra and Asura in settings such as the Miraikan Open Lab, KMD Forum, and CHI Interactivity demonstrate promising affordances of embodied AI systems in shaping user experience.

Data collected from both the Encounters exhibition and CHI Interactivity provide early insight into how participants interpret their role in shared creative output with a robotic collaborator. Figure 5.1 shows participants' responses to the statement, "I am the creator of this painting," on a 7-point Likert scale. Notably, responses from the Encounters exhibition (top, blue) span the full range of agreement, but tend toward the upper-middle of the scale, with a concentration around neutral to agree (ratings 4–7). A substantial portion of users (68.5%) selected a 4 or above, suggesting that most participants felt at least partially responsible for the outcome. However, the distribution also reflects a diversity of interpretations, which aligns with qualitative feedback describing Indra as both a partner and an independent agent.

In contrast, responses from the CHI Interactivity study (bottom, maroon) show a wider divergence in perceived authorship. Here, participants were more likely to choose low or high extremes—half of the users selected 2 or 3 (disagree to slightly disagree), while the remainder selected values 4–6, with no responses at the highest level of agreement. This polarization may stem from the more formal, research-framed setting of CHI compared to the open-ended and spontaneous tone

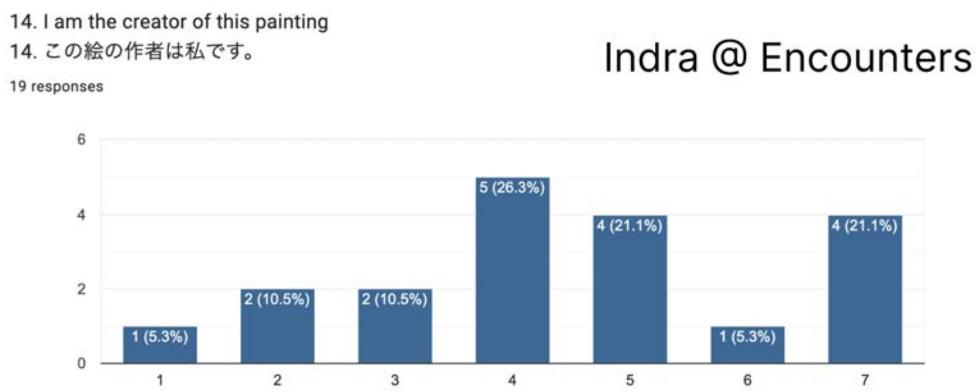

Figure 5.1 Participant responses (strong disagree 1, to 7 strong agree) from Encounters exhibition, showing the large range of sense of ownership of the produced painting.

of the Encounters exhibition.

These contrasting patterns raise compelling questions about context and framing: how do public exhibition settings versus structured academic environments influence perceived authorship? And how might this perception shift with longer engagement, greater control, or deeper understanding of the system's internals?

Across both datasets, the findings indicate that participants do not experience embodied AI as a mere tool, nor do they relinquish authorship entirely. Instead, they often occupy a middle ground, viewing the robot as a semi-autonomous collaborator with a creative voice of its own. This reinforces the project's hypothesis that embodiment and real-time responsiveness foster unique forms of shared authorship, especially when systems avoid prescriptive behavior and allow for interruption, redirection, and ambiguity.

The study's findings suggest that embodied, turn-based AI collaboration can meaningfully support users' sense of authorship and agency in creative tasks. Indra's design encouraged a shared authorship model, where users retained a sense of control while remaining open to the system's contributions. Participants did not report feeling overridden or passive, but instead described the collaboration as emergent, responsive, and occasionally surprising.

This dynamic was reflected not only in user comments but also in the final paintings, which often combined symbolic forms and abstract expression. The

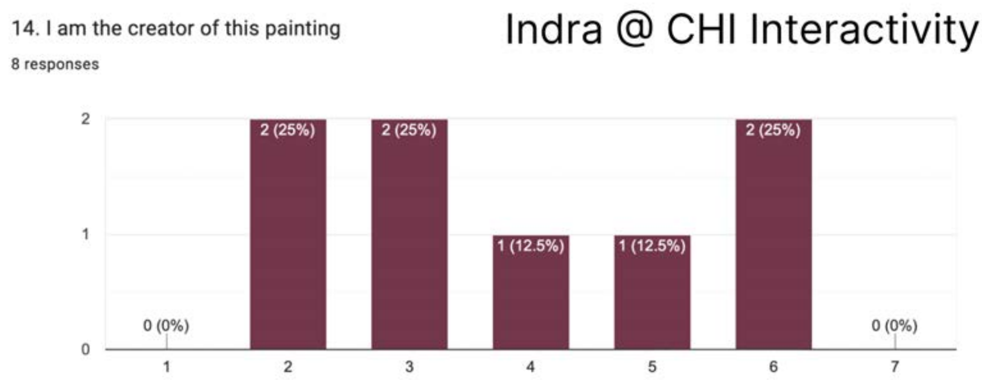

Figure 5.2 Participant responses (strong disagree 1, to 7 strong agree) from the CHI exhibition, showing a sense of fused ownership of the painting, with no users seeing either Indra or themselves as the a sole creator.

works revealed an improvisational rhythm between human and machine, shaped by mutual responsiveness rather than output optimization.

Across age and gender groups, participants expressed overall comfort and curiosity when engaging with the system. Older participants and female-identifying users showed slightly higher agency scores, while general attitudes toward robots remained broadly positive. Importantly, participants distinguished between creative augmentation and displacement, expressing optimism about AI's potential while also signaling the need for ethical, human-centered development.

Taken together, these findings point toward a productive design space for embodied AI systems that prioritize interpretability, user control, and collaborative engagement.

5.2. Limitations

Several limitations should be acknowledged when interpreting the findings. First, the participant sample was relatively small and demographically limited. Although efforts were made to include individuals with varying artistic experience, the study primarily drew from a self-selecting population of visitors at a HCI conference exhibition hall. This may introduce a selection bias toward participants with pre-existing interest in human–AI collaboration, limiting the generalizability of the results.

Second, the time available for interaction was constrained by the exhibition setting. Each participant engaged with the system for approximately 15 minutes, which allowed for meaningful first impressions and spontaneous co-creation, but limited the depth of engagement and the potential for longitudinal reflection. As a result, the data reflects initial interactions rather than sustained use or learning curves over time.

Third, all users contributed to a shared physical canvas. While this collective format encouraged a sense of collaborative authorship and continuity across sessions, it introduced difficulties in attributing specific visual outcomes to individual participants. The evolving nature of the canvas also influenced user choices, as participants were responding not only to Indra, but to the marks of previous users.

Fourth, the study focused primarily on interaction quality and sense of agency rather than the aesthetic evaluation of the visual outputs. While this aligns with the system's goal of supporting improvisational co-creation, it means that conclusions cannot be drawn about artistic skill development, style adaptation, or the refinement of final artifacts.

Finally, several technological constraints impacted system performance. Although safeguards were implemented to ensure stability, these limitations may have influenced user perceptions of fluency and control.

These limitations suggest opportunities for future work that includes longer-term deployments, isolated individual outputs, more diverse participant demographics, and extended interaction durations.

5.3. Future Works

This thesis presents systems that demonstrate early potential as their current forms only scratch the surface of what embodied co-creative AI can become. Looking forward, several directions suggest how these platforms could grow both technically and culturally.

Expanding Artistic Modalities. While Indra is currently focused on painting, future versions could engage with other art forms such as sculpture, performance, or interactive installation. Projects like Hiroshi Ishiguro's *Alter3* [65]

Figure 5.3 Possible use scenarios of Indra for disability support and in artist studios for painting large scale.

highlight how embodied agents can inhabit expressive, time-based, and spatial contexts beyond static media. Likewise, Asura's control architecture, built around timing and dexterity, could extend to domains like choreography, physical puppetry, or improvisational music—fields where timing, rhythm, and coordination define the art form.

Personalization and Co-adaptive Learning. As generative AI matures, future iterations could incorporate adaptive learning mechanisms that personalize their responses over time. Tools like Google's Magenta Studio [66] and RunwayML's style transfer engines suggest the value of systems that can recall prior user behavior to generate more resonant outputs. In our systems, this might mean learning an artist's pacing, gesture habits, or visual style, fostering a sense of mutual familiarity and ongoing collaboration.

Remote and Multi-user Collaboration. The next step for these systems could involve expanding beyond solo use to support shared creative sessions. This could involve networked co-creation between remote users or co-located collaboration across multiple control surfaces. Participatory platforms like *Teachable Machines* [67] or artistic systems like *Mirror Ritual* [68] demonstrate how distributed AI interfaces can enable new forms of shared authorship. Such designs

Figure 5.4 Abstract oil painter painting with Indra. The evolving artwork displays layered brushstrokes, abstract forms, and vibrant color contrasts, reflecting both human input and AI-generated contributions in a co-creative process.

might involve simultaneous co-control of the robot or a collective negotiation of intent between users and AI.

Therapeutic and Educational Applications. Indra show potential for use in therapy and learning contexts. Studies on social robots such as PARO [69] suggest that embodied AI can offer emotional support, reduce anxiety, or facilitate engagement for neurodiverse populations. Indra's turn-based structure could be adapted into creative tools for children, neurodivergent individuals, or people in recovery from injury or trauma. These systems might serve not only as creative partners but also as mediators of emotional expression and embodied interaction in therapeutic environments.

Design Frameworks for Embodied Co-creation. As embodied co-creative systems gain traction, there is a growing need for frameworks that evaluate not only their technical performance but their affective and relational impact. Existing work in Human-AI Interaction (HAII), explainability, and meaningful human control [70, 71] offers a foundation for such analysis. Future studies should prioritize criteria like emotional resonance, interpretability, shared authorship, and trust. These frameworks will be vital for guiding artists, technologists, and educators in deploying AI systems that support human agency and imagination.

Chapter 6
Conclusion

This thesis introduced two embodied AI systems, Indra, as case studies in designing artificial intelligence that works alongside humans in creative and physical tasks.

Indra's design demonstrates that embodied AI can support creative agency when users are given flexible control over timing and modality. Its physical presence and real-time feedback mechanisms made the AI's decisions more legible, which helped users interpret the AI as a partner rather than a tool. Asura explored how hybrid control systems can scaffold user performance without removing agency. The technical architecture developed for Asura helped validate a design approach in which AI provides minimal but meaningful assistance during physically demanding tasks. Both systems also contribute to technical methods for controlling robotic arms through real-time pipelines, including path planning, vision integration, and interaction timing. They show how generative AI and robotics can be combined to create novel shared interaction between humans and machines.

Research Question 1: Can Indra support human agency in creative tasks?

Findings from the user study suggest that Indra's design effectively supports a sense of agency in participants. The combination of embodied presence, visible planning processes, and flexible turn-taking enabled users to feel responsible for the outcomes, while also recognizing the system as an active collaborator. Most participants reported a strong sense of control and authorship, supported by the system's responsiveness to their input and the ability to override or redirect the AI's actions at any time. Participants did not perceive the AI as dominating the creative process, but rather as augmenting and responding to their intentions. These results indicate that agency can be preserved and even enhanced in co-

Figure 6.1 Painting done during development by myself and Indra

creative systems when the AI is designed to operate within a legible, interruptible, and user-paced framework.

Research Question 2: Can Indra extend human skill through real-time physical interaction?

While Indra does not automate painting tasks, it supports users by translating generative outputs into precise physical actions through a robotic arm. The system's real-time path planning and multimodal integration allow users to interact with high-complexity outputs, such as imagined visual compositions from diffusion models, without needing to manually execute every detail. In doing so, INDRA lowers the barrier for expressive experimentation, especially for participants without traditional artistic training. The presence of the robot performing their inputs, or interpreting their verbal and visual suggestions into physical strokes, gave users access to new forms of expression.

Looking ahead, this research points toward a future in which AI systems are embedded not just in functional contexts, but in cultural, performative, and artistic environments. The long-term impact of this work lies in its potential to inform design principles for embodied AI that center human values: especially creativity, agency, and play. As AI systems enter more aspects of daily life, their design must be informed by the contexts in which people already create, move, and express themselves.

References

[1] Batya Friedman and David G. Hendry. *Value Sensitive Design: Shaping Technology with Moral Imagination.* MIT Press, 2019.

[2] N. Caporusso. Generative AI and creative displacement anxiety. *Research Directs in Psychology and Behavior*, 1(1):1–9, 2023.

[3] Ziv Epstein, Aaron Hertzmann, Memo Akten, Hany Farid, Jessica Fjeld, Morgan R. Frank, Matthew Groh, Laura Herman, Neil Leach, Robert Mahari, Alex "Sandy" Pentland, Olga Russakovsky, Hope Schroeder, and Amy Smith. Art and the science of generative ai. *Science*, 380(6650):1110–1111, June 2023.

[4] Christian Then, Erickar Soewandi, Muhammad Danial, Said Achmad, and Rhio Sutoyo. The impact of artificial intelligence on art - a systematic literature review. pages 1–7, 10 2023.

[5] Alyse Marie Allred and Cecilia Aragon. Art in the machine: Value misalignment and ai "art". In *Cooperative Design, Visualization, and Engineering*, pages 31–42, 2023.

[6] The Mainichi. Fake suggestions from ChatGPT hamper AI's potential, Microsoft Japan official warns. `https://mainichi.jp/english/articles/20230516/p2a/00m/0et/014000c`, 2023. [Accessed: 21-Jan-2025].

[7] Stuart Russell. *Human Compatible: Artificial Intelligence and the Problem of Control.* Viking, 2019.

[8] Filippo Santoni de Sio and Jeroen van den Hoven. Meaningful human control over autonomous systems: A philosophical account. *Frontiers in Robotics and AI*, Volume 5 - 2018, 2018. URL: `https://www.frontiersin.org/`

journals/robotics-and-ai/articles/10.3389/frobt.2018.00015, doi:10.3389/frobt.2018.00015.

[9] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. A survey of socially interactive robots: Concepts, design, and applications. *Robotics and Autonomous Systems*, 42(3–4):143–166, 2003.

[10] Tathagata Chakraborti, Sarath Sreedharan, Yu Zhang, and Subbarao Kambhampati. Plan explanations as model reconciliation: Moving beyond explanation as soliloquy. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2017.

[11] Michael A. Goodrich and Alan C. Schultz. Human–robot interaction: A survey. *Foundations and Trends® in Human–Computer Interaction*, 1(3):203–275, 2007.

[12] Anca D. Dragan, Kenton C. T. Lee, and Siddhartha S. Srinivasa. Legibility and predictability of robot motion. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 301–308, 2013.

[13] Hannah Limerick, David Coyle, and James W. Moore. The experience of agency in human-computer interactions: a review. *Frontiers in Human Neuroscience*, 8(643):1–10, 2014.

[14] Donghee Shin. The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable ai. *International Journal of Human - Computer Studies*, 146, 2021. doi:https://doi.org/10.1016/j.ijhcs.2020.102551.

[15] Janna Anderson and Lee Rainie. The future of human agency. *Pew Research Center*, 2023. URL: https://www.pewresearch.org/internet/2023/02/24/the-future-of-human-agency/.

[16] Jenna Burrell. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1):1–12, 2016.

[17] Brent Daniel Mittelstadt, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2):1–21, 2016.

[18] Ben Shneiderman. Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, 38(1):1–22, 2022.

[19] Saleema Amershi, Daniel Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Justin Suh, Vicki Sinha, Q. Vera Lee, Paul Bennett, Kori Inkpen, Jaime Teevan, Rob Kikin-Gil, and Eric Horvitz. Guidelines for human-ai interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2019.

[20] Google DeepMind. Veo - google deepmind, 2025. Accessed: 2025-05-26. URL: `https://deepmind.google/models/veo/`.

[21] Google DeepMind. Meet flow: Ai-powered filmmaking with veo 3, 2025. Accessed: 2025-05-26. URL: `https://blog.google/technology/ai/google-flow-veo-ai-filmmaking-tool/`.

[22] Harold Cohen. The further exploits of aaron, painter. In *Stanford Humanities Review*, volume 4, pages 141–158, 1995.

[23] Thomas Lindemeier, Jan Metzner, Elke Pollak, and Bodo Urban. Image stylization with a painting machine using semantic hints. In *International Conference on Computational Intelligence in Music, Sound, Art and Design*, pages 83–95. Springer, 2015.

[24] Alison Flood. Robot artist to perform ai generated poetry in response to dante. *The Guardian*, 2021. Published online, `https://www.theguardian.com/books/2021/nov/26/robot-artist-to-perform-ai-generated-poetry-in-response-to-dante`.

[25] Pindar Van Arman. Van arman, reflective creativity, 2018. Available at `https://www.vanarman.com/reflection`.

[26] Leonel Moura. Rap—robotic action painter. *Generative Art Conference*, 01 2007.

[27] Peter Schaldenbrand, James McCann, and Jean Oh. Frida: A collaborative robot painter with a differentiable, real2sim2real planning environment, 2022. URL: `https://arxiv.org/abs/2210.00664`, `arXiv:2210.00664`.

[28] Takahiro Kizaki and Akio Namiki. Two ball juggling with high-speed hand-arm and high-speed vision system. *Proceedings - IEEE International Conference on Robotics and Automation*, pages 1372–1377, 05 2012. `doi:10.1109/ICRA.2012.6225090`.

[29] Tadayoshi Aoyama, Takeshi Takaki, Takumi Miura, Qingyi Gu, and Idaku Ishii. Realization of flower stick rotation using robotic arm. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5648–5653, 2015. `doi:10.1109/IROS.2015.7354179`.

[30] Kazutoshi Tanaka, Masashi Hamaya, Devwrat Joshi, Felix von Drigalski, Ryo Yonetani, Takamitsu Matsubara, and Yoshihisa Ijiri. Learning robotic contact juggling. 09 2021. `doi:10.1109/IROS51168.2021.9636790`.

[31] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever. DALL·E: Zero-shot text-to-image generation. Technical report, 2021. URL: `https://arxiv.org/abs/2102.12092`, `arXiv:2102.12092`.

[32] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of CVPR*, pages 10674–10685, 2022.

[33] Ai-Da Robot Official. Ai-da robot official. `https://www.ai-darobot.com/`. [Accessed: 2025-01-01].

[34] Marian Mazzone and Ahmed Elgammal. Art, creativity, and the potential of artificial intelligence. *Arts*, 8(1), 2019. URL: `https://www.mdpi.com/2076-0752/8/1/26`, `doi:10.3390/arts8010026`.

[35] Peter Schaldenbrand, Gaurav Parmar, Jun-Yan Zhu, James McCann, and Jean Oh. Cofrida: Self-supervised fine-tuning for human-robot co-painting. 2024. `arXiv:2402.13442`.

[36] H. Knight. Eight lessons learned about non-verbal interactions through robot theater. *International Journal of Social Robotics*, 7(5):697–708, 2015.

[37] H. Ishii and B. Ullmer. Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proceedings of the ACM SIGCHI Conference*, pages 234–241, 1997.

[38] Steffen Stange, Yvonne Funk, Nils Hörning, and Jessica Heesen. How transparency modulates trust in artificial intelligence: The relationship between user expectations, transparency design, and trust recovery. *Journal of Behavioral and Experimental Economics*, 101:101972, 2023.

[39] Ibad Ali. Ai transparency and explainability. 12 2024.

[40] Jan-Philipp Stein, Tanja Messingschlager, Timo Gnambs, Fabian Hutmacher, and Markus Appel. Attitudes towards ai: measurement and associations with personality. *Scientific Reports*, 14, 02 2024. `doi:10.1038/s41598-024-53335-2`.

[41] Mika Koverola, Anton Kunnari, Jukka Sundvall, and Michael Laakasuo. General attitudes towards robots scale (gators): A new instrumen t for social surveys. *International Journal of Social Robotics*, 14:1559–1581, year=2022, publisher=Springer.

[42] Saleh Afroogh, Ali Akbari, Emmie Malone, Mohammadali Kargar, and Hananeh Alambeigi. Trust in ai: progress, challenges, and future directions. *Humanities and Social Sciences Communications*, 11(1), Nov 2024. URL: `https://www.nature.com/articles/s41599-024-04044-8`, `doi:https://doi.org/10.1057/s41599-024-04044-8`.

[43] Kwan Min Lee, Namkee Park, and Hayeon Song. Can robots manifest personality? an empirical test of personality recognition, social responses, and social presence in human-robot interaction. *Journal of Communication*, 61(4):758–779, 2011.

[44] Mei Yii Lim, David A. Robb, Bruce W. Wilson, and Helen Hastie. Feeding the coffee habit: A longitudinal study of a robo-barista. *arXiv preprint arXiv:2309.02942*, 2023.

[45] Mei Yii Lim, José David Aguas Lopes, David A. Robb, Bruce W. Wilson, Meriam Moujahid, Emanuele De Pellegrin, and Helen Hastie. We are all individuals: The role of robot personality and human traits in trustworthy interaction. *arXiv preprint arXiv:2307.15568*, 2023.

[46] Aikaterina Manoli, Janet V. T. Pauketat, and Jacy Reese Anthis. The ai double standard: Humans judge all ais for the actions of one. *arXiv preprint arXiv:2412.06040*, 2024.

[47] Gabriel Lima, Nina Grgić-Hlača, and Meeyoung Cha. Blaming humans and machines: What shapes people's reactions to algorithmic harm. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery. URL: `https://doi.org/10.1145/3544548.3580953`, `doi:10.1145/3544548.3580953`.

[48] Shivani Kapania, Oliver Siy, Gabe Clapper, Azhagu Meena SP, and Nithya Sambasivan. "because ai is 100% right and safe": User attitudes and sources of ai authority in india. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery. URL: `https://doi.org/10.1145/3491102.3517533`, `doi:10.1145/3491102.3517533`.

[49] Zihan Liu, Han Li, Anfan Chen, Renwen Zhang, and Yi-Chieh Lee. Understanding public perceptions of ai conversational agents: A cross-cultural analysis. CHI '24, New York, NY, USA, 2024. Association for Computing Machinery. URL: `https://doi.org/10.1145/3613904.3642840`, `doi:10.1145/3613904.3642840`.

[50] Saumya Pareek, Eduardo Velloso, and Jorge Goncalves. Trust development and repair in ai-assisted decision-making during complementary expertise. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and*

Transparency, FAccT '24, page 546–561, New York, NY, USA, 2024. Association for Computing Machinery. URL: https://doi.org/10.1145/3630106.3658924, doi:10.1145/3630106.3658924.

[51] Rebecca Stower, Arvid Kappas, and Kristyn Sommer. When is it right for a robot to be wrong? children trust a robot over a human in a selective trust task. *Computers in Human Behavior*, 157:108229, 04 2024. doi:10.1016/j.chb.2024.108229.

[52] Sonia Sousa, David Lamas, José Cravino, and Paulo Martins. Human-centered trustworthy framework: A human–computer interaction perspective. *Computer*, 57(3):46–58, 2024. doi:10.1109/MC.2023.3287563.

[53] Oleksandra Vereschak, Fatemeh Alizadeh, Gilles Bailly, and Baptiste Caramiaux. Trust in ai-assisted decision making: Perspectives from those behind the system and those for whom the decision is made. CHI '24, New York, NY, USA, 2024. Association for Computing Machinery. URL: https://doi.org/10.1145/3613904.3642018, doi:10.1145/3613904.3642018.

[54] Rodney A. Brooks. A robust layered control system for a mobile robot. *IEEE Journal on Robotics and Automation*, 2(1):14–23, 1986.

[55] Erann Gat. On three-layer architectures. In *Artificial Intelligence and Mobile Robots*, pages 195–210. MIT/AAAI Press, 1998.

[56] R. Peter Bonasso, David Kortenkamp, David P. Miller, and Marc Slack. Experiences with an architecture for intelligent, reactive agents. *Journal of Experimental & Theoretical Artificial Intelligence*, 9(2-3):237–256, 1997.

[57] George Konidaris, Leslie Pack Kaelbling, and Tomas Lozano-Perez. From skills to symbols: Learning symbolic representations for abstract high-level planning. *Journal of Artificial Intelligence Research*, 61(https://doi.org/10.1613/jair.5575):215–289, Jan 2018. doi:https://doi.org/10.1613/jair.5575.

[58] David Ha and Douglas Eck. A neural representation of sketch drawings. arXiv preprint arXiv:1704.03477, 2017.

[59] Rosi Braidotti. *The Posthuman*. Polity Press, 2013.

[60] How the u.s. public and ai experts view artificial intelligence, 2025. Accessed June 17, 2025. URL: https://www.pewresearch.org/internet/2025/04/03/how-the-us-public-and-ai-experts-view-artificial-intelligence/.

[61] Hiroo Ide, Sayuri Suwa, Yumi Akuta, Naonori Kodate, Mayuko Tsujimura, Mina Ishimaru, Atsuko Shimamura, Helli Kitinoja, Sarah Donnelly, Jaakko Hallila, Marika Toivonen, Camilla Bergman-Kärpijoki, Erika Takahashi, and Wenwei Yu. A comparative study to elucidate factors explaining willingness to use home-care robots in japan, ireland, and finland. *Scientific Reports*, 14(1), Nov 2024. doi:https://doi.org/10.1038/s41598-024-79414-y.

[62] Chaolan Lin, Selma Šabanović, Lynn Dombrowski, Andrew D. Miller, Erin Brady, and Karl F. MacDorman. Parental acceptance of children's storytelling robots: A projection of the uncanny valley of ai. *Frontiers in Robotics and AI*, Volume 8 - 2021, 2021. URL: https://www.frontiersin.org/journals/robotics-and-ai/articles/10.3389/frobt.2021.579993, doi:10.3389/frobt.2021.579993.

[63] Bruno Latour. *We Have Never Been Modern*. Harvard University Press, 1993.

[64] Rosi Braidotti. *Posthuman Knowledge*. Polity Press, 2019.

[65] Takahide Yoshida, Atsushi Masumori, and Takashi Ikegami. From text to motion: grounding gpt-4 in a humanoid robot "alter3". volume 12. Frontiers Media SA, May 2025. doi:https://doi.org/10.3389/frobt.2025.1581110.

[66] Google Brain Team. Magenta studio tools, 2018. Available at: https://magenta.tensorflow.org/studio.

[67] teachable machine by google creative lab - experiments with google. URL: https://experiments.withgoogle.com/teachable-machine.

[68] Nina Rajcic and Jon McCormack. Mirror ritual: An affective interface for emotional self-reflection. pages 1–13, 04 2020. doi:10.1145/3313831.3376625.

[69] Takanori SHIBATA and Kazuyoshi WADA. Therapeutic seal robot “paro” and its effective operation. *Journal of The Society of Instrument and Control Engineers*, 51(7):640–643, 2012. doi:10.11499/sicejl.51.640.

[70] Tim Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267:1–38, 2019.

[71] Ben Shneiderman. Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, 36(6):495–504, 2020.

[72] Milena Nikolova, Femke Cnossen, and Boris Nikolaev. Robots, meaning, and self-determination. *Research Policy*, 53(5):104987, 2024. URL: https://www.sciencedirect.com/science/article/pii/S0048733324000362, doi:https://doi.org/10.1016/j.respol.2024.104987.

Appendices

A. Asura: AI-Assisted Juggling System

A.1 System Overview

Asura (Adaptive Skill-based Unification of Robotic Assistance) is a system that assists users in learning or performing bimanual juggling through collaborative robot arms. The system is designed to offer physical and visual feedback during coordination-intensive tasks, adjusting its assistance based on the user's movement and performance.

Architecture: Asura operates using a real-time motion capture pipeline integrated with predictive movement modeling. The system uses vision-based object tracking to monitor the trajectory of the balls and the user's hands, feeding data into a control module that plans robotic intervention or mirroring. Machine learning components predict throw consistency and timing misalignments, and adapt robot behavior accordingly.

Interface Mechanisms: The user interacts with Asura through two 3D mice to control the robot arms engagement of throwing and catching pucks. Feedback is offered visually (through projected pathlines) and physically (by robotic arms assisting or mirroring throws). A visual display provides real-time analysis of the user's rhythm and timing to support reflective learning.

Design Rationale: Asura is based on the concept of "adaptive scaffolding," where the robot supports rather than takes over the user's skill expression. The AI's role is to enable extended control and smoother coordination without diminishing the user's felt sense of contribution. Unlike Indra's narrative domain, Asura operates in a kinetic and temporal medium, where agency is felt in the rhythm and responsiveness of shared motion.

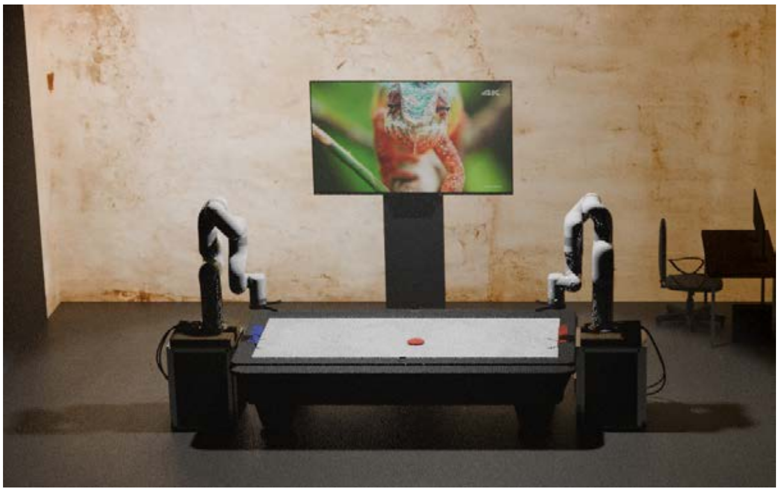

Figure A.1 Render of the Asura layout used during the KMD Forum showing two robot arms mounted at each end of the hockey table.

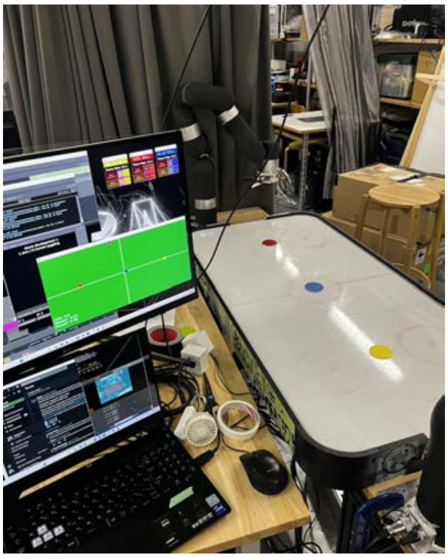

Figure A.2 Prototype of Asura using color blob tracking for 3 pucks.

Figure A.3 The SpaceMouse by 3Dconnexion, a 3D joystick designed for CAD software.

A.2 User Interaction

Asura provides a responsive interaction experience designed to support physical coordination and quick dexterous object manipulation through real-time robotic control. Users operate two robotic arms via dual 3D mice, which are mapped to a virtual environment displayed on a monitor. This setup enables intuitive teleoperation, letting users perform complex juggling motions without needing to see the actual robot arms or physical table. Asura adapts to different levels of experience and dexterity. For users familiar with robotics or 3D interaction, it offers high responsiveness and nuanced control, allowing them to explore advanced manipulations and improvisation. For novices, the system reduces the learning curve through visual trajectory guides and automated corrections that support timing and spatial awareness. Asura transforms juggling, a skill often seen as difficult or inaccessible, into an interactive and learnable experience by augmenting rather than replacing user input. For researchers and performers, it is a platform to investigate shared agency and bodily coordination between human and machine. For younger users or those with limited motor skills, it becomes a playful tool for learning motion, rhythm, and precision through embodied engagement. By blending tactile input with AI-enhanced feedback, Asura encourages users to experiment, make mistakes, and improve through embodied trial and error.

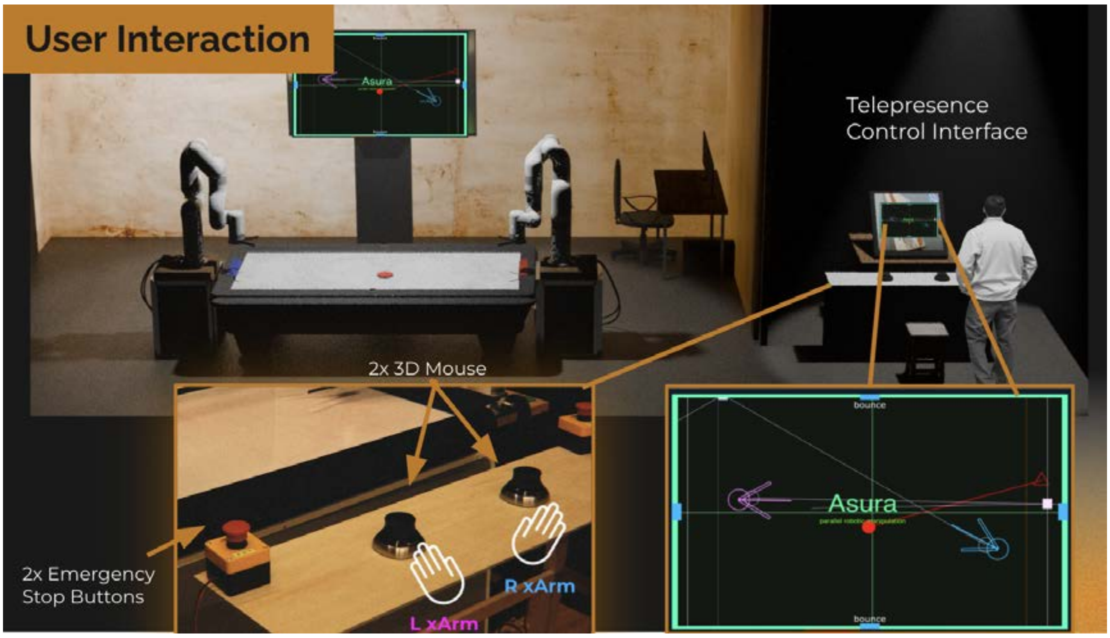

Figure B.1 Diagram of Asura's user interaction setup, showing a telepresence station with dual 3D mouse controllers and a live interface. The user operates two robotic arms remotely, aided by real-time trajectory visualization and emergency stop buttons for safety.

B. Brief of Asura Implementation

Asura is a teleoperative robotic system designed to investigate how AI can augment human dexterity in fast-paced, physically demanding tasks such as juggling, catching, and throwing. It builds on several foundational components developed for Indra, particularly in real-time control, system architecture, and spatial reasoning, adapting them to a different interaction modality focused on sensorimotor performance rather than symbolic co-creation.

B.1 Hardware Configuration

Asura features two 6-degree-of-freedom (6-DOF) `xArm` robotic arms by UFactory, mounted on a shared physical table. These arms are tracked using an OptiTrack motion capture system with six high-speed infrared cameras, which track retroreflective markers affixed to the robot arms and lightweight pucks. This

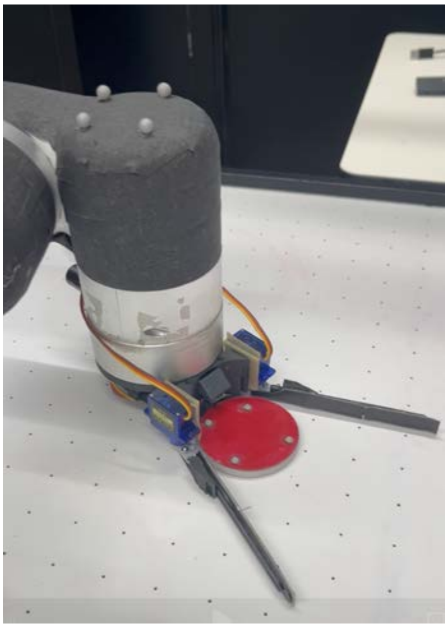

Figure B.2 Close-up of the gripper design using two servo motors for grip, proximity sensing, and shoot actuator, with optitrack marker balls attached to the top.

external tracking provides millimeter-accurate, low-latency spatial data for both targets and effectors. The user operates each `xArm` in real-time using `3Dconnexion SpaceMouse` controllers. These 6-DOF devices allow independent manipulation of position and rotation, granting users precise control over arm trajectories.

Real-Time Control (Inherited from INDRA)

The real-time motion control system is adapted directly from Indra, where it was originally developed to enable smooth robotic brushstrokes. Asura uses the same architectural approach: a `TouchDesigner`-based controller communicates directly with the robot arms via the `xArm Python SDK`, bypassing ROS in favor of a low-latency, high-frequency control loop.

- The `TouchDesigner` xArm controller streams real-time Cartesian or joint-space commands via TCP, updating position and velocity at frame rates between 30–110 Hz.

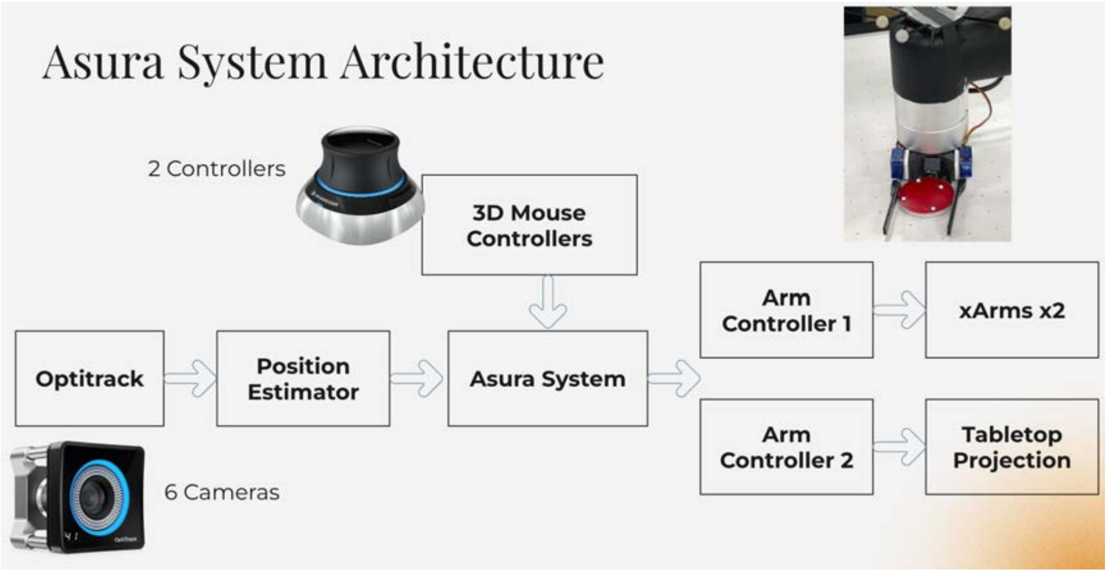

Figure B.3 System architecture of ASURA, showing how dual 3D mouse controllers and OptiTrack camera input are processed by the central Asura system to control two xArm robots and a tabletop projection interface. The flow illustrates the coordination between user input, motion tracking, and robotic response.

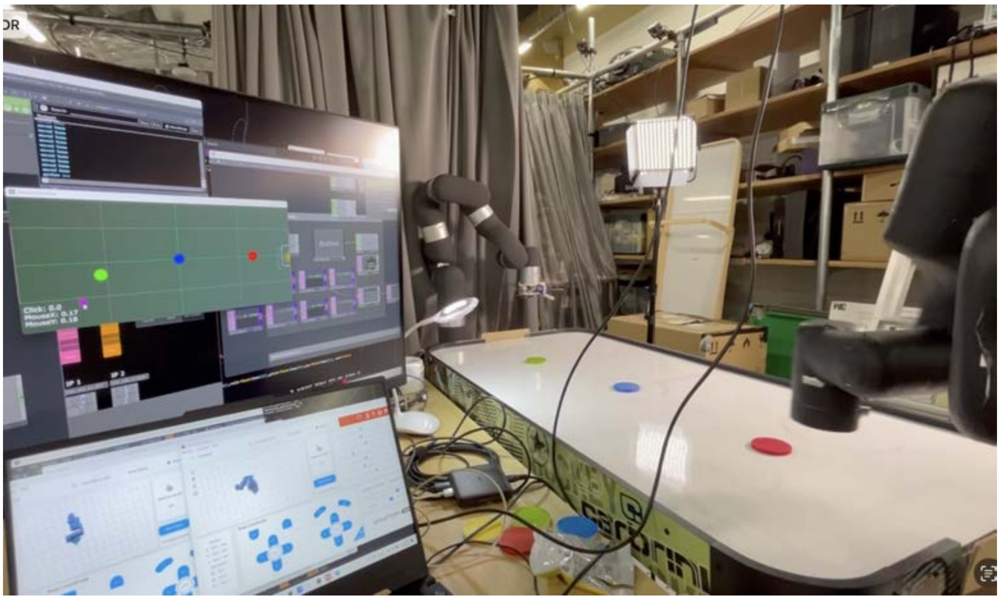

Figure B.4 The Asura prototype setup shows two xArm robotic arms interacting with an air hockey table, with color-coded pucks used for target tracking. The system is controlled via a TouchDesigner interface displayed on the external monitor, showing real-time puck positions. The bottom screen shows UFactory xArm Studio which provides direct control when needed.

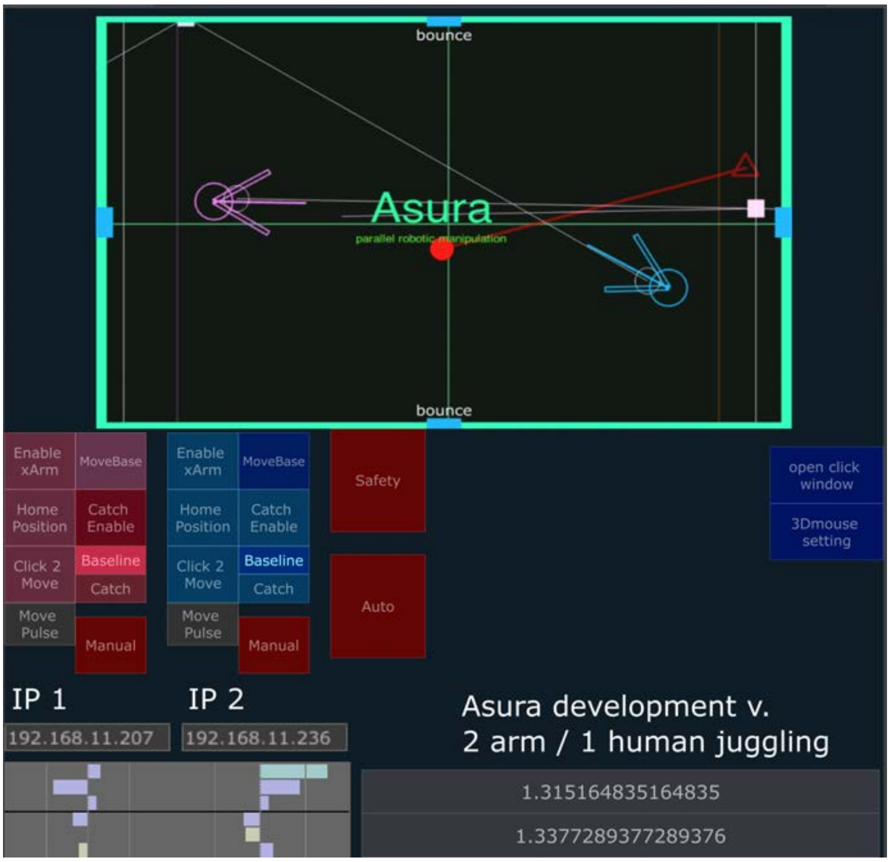

Figure B.5 User interface for the Asura system administrator, showing the trajectory lines, toggle-able features, and safety buttons.

- Servo mode control allows continuous motion interpolation with safety enforcement and boundary checks.
- This controller maintains sub-50 ms latency from user input to robot response, critical for tasks requiring timing accuracy, such as catching a moving object.

The path planning framework, originally developed for painting trajectories in Indra, was re-implemented to process and respond to real-time object positions in Asura. While Indra's planner interpreted pixel-level stroke paths, Asura's version uses 3D trajectory estimation and predictive targeting for interception.

Trajectory Prediction and Visual Feedback

Asura includes a trajectory estimator module, which processes real-time OptiTrack data of puck motion using Newtonian physics models to estimate future positions. These projected trajectories are visualized in real-time:

- A short-throw projector above the table overlays visual guides (puck trails, predicted arcs, and catch windows) directly onto the physical surface.
- These overlays act as an augmented reality (AR) feedback system, guiding users without removing their focus from the task.

This form of spatialized visual feedback draws conceptual inspiration from Indra's avatar and 3D planning visualizations, though adapted here for real-time motor coordination.

Assistance Logic

Asura is not fully autonomous but offers task-specific AI assistance, balancing human control with subtle support: **Auto-Catch Mechanism:** When a puck enters a configurable proximity zone around the end effector, the gripper automatically engages, ensuring successful grasps during high-speed interactions. **Reactive Interception:** If a puck is moving too quickly for user response, the system predicts its path and repositions the arm in anticipation. This lightweight AI intervention is designed to support, not override, the user's agency, paralleling Indra's turn-taking logic, but in a continuous control paradigm.

Interaction Model

Unlike Indra's turn-based symbolic collaboration, Asura operates in a continuous control loop. Users receive:

- Visual feedback through projected trajectories.
- Rhythmic flow enabled by minimal UI and uninterrupted input.

Table B.1 Shared Components between Indra and Asura

Component	**Adaptation in Asura**
Real-time xArm control loop	Used with dual-arm simultaneous input
TouchDesigner-based controller	Extended to handle two arms and 3D puck data
Path planning infrastructure	Modified from stroke generation to trajectory tracking
Visualization pipeline	Re-imagined as physical AR overlay for immediate feedback

B.2 Shared Components between Indra and Asura

C. Asura @ KMD Forum

The Asura system was designed to investigate how human agency is reconfigured when artificial intelligence and robotic systems are embodied as extensions of the user, particularly in high-skill, real-time tasks. Unlike Indra, which presents the robot as a co-creative partner with its own autonomy and turn-based interaction model, Asura positions the user in direct control of multiple robotic limbs. This setup allows one individual to operate a maximum of four robotic arms in parallel using dual 3D input devices, effectively extending their body into a coordinated multi-limb robotic system.

The primary aim is to explore how such direct and embodied control of AI-assisted robotic systems could inform future applications in domains like teleoperation, manufacturing, and collaborative physical labor. By engaging participants in a skill-based activity like juggling (which inherently requires dexterity, rhythm, and precision) the system tests whether AI augmentation can support, rather than supplant, the human experience of mastery.

Asura also serves to challenge common narratives of human-AI disconnection by framing robotics not as external or alien actors, but as intimately integrated tools of human expression and utility. In this view, technology does not replace human skill but rather elevates it. Through Asura, the research seeks to understand

Figure C.1 Two children using Asura against each other at KMD Forum 2024.

whether embodying AI-driven robotics can foster a reinforced sense of control, authorship, and self-extension, rather than diminished agency.

C.1 Demo Setup

Specific attention is given to physical safety and psychological comfort due to the embodied nature of the system.

Physical Safety Measures

Asura involves the teleoperation of four robotic arms, creating potential physical hazards if proper precautions are not taken. To mitigate this, the study implemented multiple layers of safety. Four physical emergency stop buttons were placed around the hockey table, alongside a software-based stop system that automatically halts all movement when predefined speed thresholds or spatial boundaries are violated. The robotic arms were confined to operate only within a programmed safety zone, specifically limited to the vertical volume above the hockey table surface. In addition, collision detection was enabled and tuned to be responsive to rapid or unexpected movement. These layers collectively ensure that participants remain physically safe throughout the session.

Minimizing Psychological Risk

While participants do not directly interface with the robot arms in close physical proximity, the embodied nature of Asura introduces the potential for psychological discomfort. Prior studies in human-robot interaction suggest that anxiety may arise when users feel a lack of control, experience performance-related stress, and feel unsafe when near a robot coworker, especially when engaging in novel, skill-based tasks with intelligent or semi-autonomous agents [72]. To address this, Asura's design includes a spatial separation between the participant and the physical system: the user operates from a teleoperation station positioned approximately one meter from the actual table, with a partitioning wall in between. This setup reduces feelings of intimidation while still allowing users to embody and influence the system.

Efforts were also made to communicate the experimental nature of the task clearly and emphasize that there were no “right” or “wrong” outcomes. Participants were encouraged to explore the system rather than perform under pressure. Facilitators were available throughout the study to answer questions, provide assistance, and de-escalate any distress.

C.2 Demo Content

The Asura system was first publicly demonstrated at the KMD Forum 2024 “Voyage” event, held over two days on October 12–13. The forum, hosted by Keio Media Design Graduate School, featured a wide array of student research prototypes. Asura was presented as an early-stage system within a shared exhibition space that included both academic visitors and members of the public, including families and children attending a concurrent event in the building. Approximately 15 visitors interacted with the system over the course of the exhibition.

Each session was short and informal, around 5 to 10 minutes and did not contain structured tasks. This open-ended format enabled observation of how people naturally approached and interpreted the dual-arm teleoperation setup. Visitors were invited to engage with Asura by grasping two 3D mice positioned side-by-side on a control table. Above the table, a digital twin was visualized on a television monitor, showing the hockey table surface and projected puck trajectories as they were computed in real time. No instruction was provided unless asked for support, allowing participants to intuitively explore the system's capabilities and limitations.

Notably, the interaction styles varied by demographic. Children tended to treat the system playfully, using each robot arm in turn to pass the puck back and forth, often engaging in a game-like exchange. However, adults experimented with more complex maneuvers such as bouncing the puck off the wall to simulate a juggling arc between arms. This behavior suggested an emergent understanding of the system's control dynamics and an intuitive effort to test its coordination potential.

Participant reactions indicated high engagement and enjoyment. Many commented on the smoothness of the control system, citing the responsiveness of the 3D mice as a highlight. The system's short latency allowed for fluid interaction, which appeared to promote a sense of agency and bodily extension. However, a recurring technical issue emerged during use: when users rotated the end effector excessively, joint errors occurred, requiring the arm to be reset. This led to a subsequent design modification that imposed improved rotation limits and enhancing overall system robustness.

This informal deployment offered valuable design insight for both Asura and future iterations of Indra. Observing how users gradually discovered fluent control in Asura highlighted the importance of embodied, pressure-based interfaces. The direct mapping between user hand motion and robotic arm behavior appeared intuitive, especially as users adjusted their grip pressure to modulate movement direction and speed. This inspired the development of Indra's "direct control" mode, which mirrors the user's iPad brush input as Indra physically paints on the canvas. In both systems the capacity to see one's own motion reflected or extended in real time proved central to user engagement and perceived control.

Another critical takeaway was the difference in pacing between the two systems. Asura's continuous interaction loop (without turn-taking or breaks) led to quicker fatigue, with many users disengaging earlier than in sessions with Indra. This contrast reinforced the value of structured rhythm and alternating control as a means of sustaining engagement, particularly in co-creative or collaborative contexts.

Overall, the KMD Forum prototype session served as a low-pressure exploratory deployment that validated key aspects of Asura's interaction design, while also revealing opportunities for refinement and cross-system inspiration. These observations underscore the importance of embodied interaction, visual feedback, and adaptable pacing in the design of human-AI systems for physical collaboration.