AIoT Anthropomorphic Little Monster

Cheng-Hao Chang, Taipei YouHua High School, Taiwan
Su-Chu Hsu, National Tsing Hua University, Taiwan
Shih-Ta Liu, National Tsing Hua University, Taiwan

The Kyoto Conference on Arts, Media & Culture 2025
Official Conference Proceedings

Abstract

To attract children's attention, this study utilizes anthropomorphic little monsters with traits such as timidity, gluttony, playfulness, and adventurousness. These cute, approachable characters help children project their emotions and imagination onto them. Our research integrates AI, the Internet of Things, and interactive design to develop the AIoT Anthropomorphic Monster Interaction System, planned for implementation at Hsinchu Science Theme Park during Halloween. We designed a "palm monster generator box" that allows children to insert their hands, and the AI program then generates a unique, cute monster based on the shape and posture of the child's palm. The system projects a nighttime forest scene onto the wall. Children can use a mobile phone with an app to walk in front of the nighttime forest wall and shake the phone to awaken the monster they created, and play and interact with it. They can also take their little monster home through a "scoop-up" action using the mobile phone and share it on social media. Through the development and implementation of this system, we aim to explore the emotional impact of Kawaii Theory on children, specifically verifying whether anthropomorphic monster design enhances joy and engagement. Additionally, we strive to examine whether our interactive design can provide children with a "safe sense of adventure." Monsters usually symbolize the "unknown," but when anthropomorphized, this "adventure" becomes "controllable" and "cute," allowing children to experience adventure and challenges in a safe environment.

Keywords: AIoT, interactive design, Kawaii Theory, anthropomorphism, safe sense of adventure

Introduction

In contemporary design for children, the "Kawaii" (cute) style has evolved beyond visual aesthetics into a key element for fostering emotional connection and engagement. Originating from Japan, "Kawaii" represents a psychological and social phenomenon that elicits care and empathy. While previous studies have predominantly focused on visual or character design, few have explored how Kawaii elements influence children's emotions and behaviors within AIoT-based interactive experiences. This study introduces "Anthropomorphic Little Monsters" as emotional mediators, combining AI generation and IoT interaction to create a digital experience characterized by a "safe sense of adventure." The primary objectives are:

- To explore the application of Kawaii Theory in interactive design.
- To verify whether anthropomorphic monsters influence children's emotions and motivation.
- To construct a "Cute and Adventurous Safe Experience Model," allowing children to face challenges within a safe framework.
- To implement an AIoT-integrated system in a real-world setting (a science theme park) to observe engagement and collaborative learning.

In summary, while the "Kawaii" culture has been widely recognized for its ability to foster emotional connection, its application in AIoT-based interactive environments for children remains underexplored. Existing designs often focus on visual aesthetics but lack a deeper integration of psychological mechanisms that can support children's exploration of the unknown. To bridge this gap, this study proposes an innovative approach by integrating generative AI and IoT technology with the "Anthropomorphic Little Monster" concept. By transforming the abstract and potentially frightening concept of "adventure" into a tangible, "controllable," and "cute" interactive experience, we aim to establish a "Safe Sense of Adventure" model. This research not only seeks to verify the emotional impact of Kawaii Theory in digital interaction but also provides empirical evidence on how such systems can enhance children's engagement and motivation to learn in a real-world setting.

Related Work

Originating from Japanese culture, "Kawaii" has transcended its regional roots to become a significant cross-cultural phenomenon, influencing aesthetics and consumer behavior globally, from Asia to Europe and the Americas. Historically viewed primarily as a visual style synonymous with "cuteness," contemporary research has redefined it as a complex psychological trigger.

Nittono (2016) proposed a behavioral science framework describing Kawaii as a two-layer model: an Affective Dimension, which evokes intrinsic positive emotions such as affection and the desire to protect; and a Social Value Dimension, which promotes interpersonal relationships and social attachment. Building on this theoretical foundation, this study integrates Kawaii not merely as a visual element, but as a strategic "emotional mediator" within an AIoT interactive system. By applying Kawaii traits to "monster" characters—which traditionally symbolize the unknown and fear—we aim to leverage the psychological mechanism of "care and empathy." This transformation allows children to perceive the unknown not as a threat, but as a companion, thereby facilitating a "Safe Sense of Adventure" that encourages active exploration and social sharing.

Anthropomorphism as a Transformation Mechanism

Anthropomorphism is defined as the attribution of human-like emotions, intentions, or mental states to non-human entities. Epley et al.'s (2007) Three-Factor Theory suggests that this psychological tendency is driven by a basic human need to comprehend uncertainty and establish social connections. In the context of this study, "monsters" traditionally symbolize the unknown, fear, and anxiety for children. We introduce anthropomorphism not merely as a design style, but as a critical transformation mechanism to address this anxiety. By imbuing these "unknown" entities with distinct human personality traits—specifically focusing on the "Experience" dimension (the ability to feel emotions like timidity, gluttony, and playfulness) from Gray et al.'s (2007) Mind Perception Model. We fundamentally alter the child's perception. According to the Media Equation (Reeves & Nass, 1996), children naturally treat these human-like virtual characters as social partners rather than threats. Consequently, the terrifying "unknown" is recontextualized into an approachable "social companion." This theoretical application allows us to convert a potential source of fear into a "Safe Sense of Adventure," where children are motivated to explore and interact with the monsters because they perceive them as relatable beings requiring care, rather than dangerous objects to be avoided.

Theoretical Synthesis: Constructing a Safe Sense of Adventure

This study postulates that a Safe Sense of Adventure is achievable through a strategic integration of the theories. It directly addresses the core research problem: how to encourage children to explore the unknown without the paralysis of anxiety. By synthesizing the Affective Dimension of Kawaii Theory (which triggers approach motivation and care) and the Cognitive Transformation of Anthropomorphism (which reframes the unknown into a social partner), we construct a psychological safety net. In this framework, the "unknown"—represented by the monster—is no longer a threat to be avoided but a companion to be cared for.

Therefore, the Safe Sense of Adventure is defined in this study not merely as a gameplay feature, but as a theoretically grounded experience model. It leverages the Kawaii and Anthropomorphic traits to create a zone of proximal development (Vygotsky, 1978), where the child's natural fear of the unknown is counterbalanced by the desire for social interaction. This confirms that adventure becomes "safe" not by removing the challenge, but by fundamentally transforming the child's emotional relationship with the challenger.

Methodology and Implementation

This study proposes a distributed AIoT architecture designed to bridge the gap between physical interaction and generative AI in edge computing environments. The system integrates computer vision, utilizing Latent Diffusion Models (Rombach et al., 2022), and IoT communication protocols to create an immersive "in-the-wild" interactive experience. The following sections detail the system design, generative pipelines, and the implementation of the latency-masking interaction mechanism.

System Design and AI Generation

The visual synthesis core of the system is built upon ComfyUI (comfyanonymous, 2023), a node-based interface for Stable Diffusion. We adopted this modular architecture to construct a precise, reproducible generative workflow that integrates disparate technical components—

ranging from computer vision inputs to style-transfer outputs—into a unified processing pipeline. The generative process is architected into two distinct sub-systems: Gesture-Driven Character Synthesis and Structure-Preserved Environmental Rendering.

Gesture-Driven Character Synthesis (The Palm Monster)

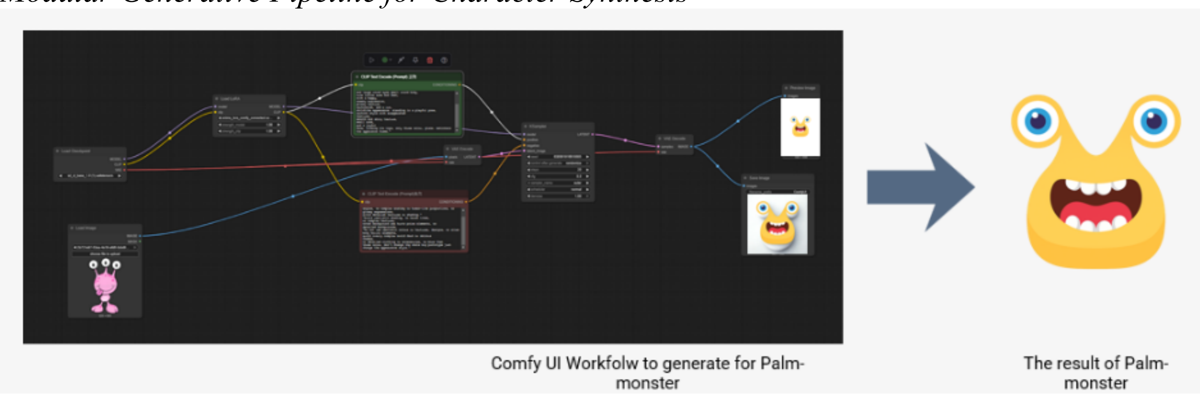
To ensure the generated agents (monsters) resonate with the user's physical actions, we implemented a Semantic Mapping Strategy rather than relying on randomized seed generation. The system utilizes a custom-built "Palm Monster Generator Box" equipped with a camera and MediaPipe (Google AI, 2020) for real-time hand tracking. The technical pipeline operates as follows:

Vision-to-Prompt Translation. The system employs a deterministic mapping algorithm where specific hand gestures are translated into distinct morphological constraints in the text prompt. For instance, a closed fist (detected as gesture "Zero") signifies compactness, which the system maps to prompts for a "limb-less, round-shaped" creature. Conversely, a "Two-finger" gesture introduces complexity, triggering prompts for "two distinct features" (e.g., long ears or dual legs).

Style Consistency via LoRA. To maintain a cohesive visual identity across diverse user inputs, we inject a custom-trained LoRA (Low-Rank Adaptation) model into the diffusion process. This ensures that regardless of the structural variability in hand gestures, the output strictly adheres to the pre-defined Kawaii artistic style, stabilizing the aesthetic quality of the generated content.

This two-part generation process, combining deterministic semantic mapping with aesthetic stabilization, forms the core of the monster synthesis pipeline. The full workflow, detailing the node structure from input prompt to final output, is illustrated in Figure 1.

Figure 1
Modular Generative Pipeline for Character Synthesis



Structure-Preserved Environmental Rendering

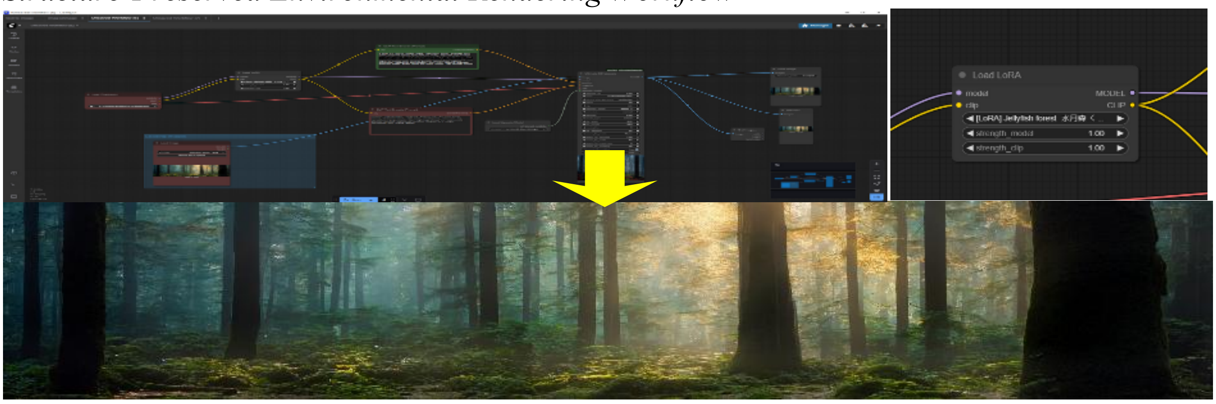
The immersive projection environment is constructed using an Image-to-Image (Img2Img) workflow. Unlike standard Text-to-Image generation, which creates scenes *ab initio*, our approach utilizes a photograph of the actual physical site as a Structural Baseline.

Context-Aware Generation. By applying the SDXL 1.5 model combined with a specific atmospheric LoRA (e.g., "Jellyfish Forest"), the system hallucinates a fantasy layer over the real-world geometry.

Physical-Digital Alignment. Crucially, this Img2Img method allows for Geometric Structure Preservation. Physical elements, such as tree trunks used for NFC interaction, retain their positions and shapes in the digital projection. This ensures that the digital content aligns perfectly with the physical props, preventing immersion-breaking discrepancies during the user's interaction.

The resulting environment, demonstrating the fusion of real-world structure with the fantasy aesthetic generated via the Img2Img workflow, is presented in Figure 2.

Figure 2
Structure-Preserved Environmental Rendering Workflow



Communication Framework

The system utilizes a decentralized communication network based on the MQTT protocol, specifically designed to handle the real-time asynchronous communication challenges among heterogeneous components, including physical sensing nodes (IoT), the image processing core, and the mobile interface. This architecture establishes a central message broker, coordinating data flow among all entities.

Network Core and Data Distribution. The communication backbone is anchored by an MQTT Broker hosted on a dedicated Raspberry Pi 5 edge computing unit. The MQTT Broker serves as the central message exchange, allowing all system components (clients) to communicate indirectly via a low-overhead Publish/Subscribe mechanism. This design effectively decouples the control flow from the data processing units, enhancing system stability and scalability in a multi-node environment. The Broker manages various data types, including Sensing Data (Gesture IDs, Location IDs), Command Data (generation requests, mobile shake commands), and State Data (monster generation completion notifications).

Interaction Flow and Game Synchronization. The system synchronizes user actions and system state through a two-stage process mediated by the central MQTT Broker. First, the Palm Monster Generator Box (RPi/MediaPipe) translates a user's gesture into a unique Monster ID, which the coordinator (Mini PC) uses to initiate the AI generation pipeline and bind the ID to the player's session. Second, all subsequent interactions are routed through a local network structure, where a local server (Node.js/Flask) manages user data via the on-site

4G/WiFi access point. This local server receives high-level commands from the mobile App (e.g., “shake” or “scoop-up” commands). Both the mobile commands and the sensor nodes’ (NFC) physical location data are fed back to the broker. The coordinator constantly subscribes to these streams, executing the core game logic by verifying if the physical Location ID matches the unique Monster ID or if a valid mobile command has been received. This process ensures that the visual feedback on the projection screen is triggered consistently and with low latency, completing the physical-digital interaction loop.

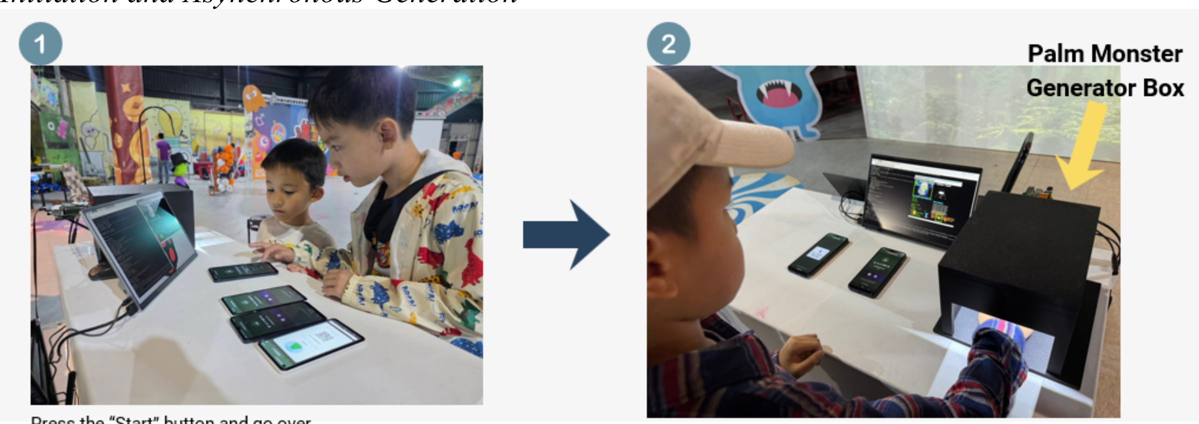
Interaction Flow

The user experience is structured around a five-step sequential process, designed to integrate the necessary AI computation time into the gameplay loop (Latency Masking Strategy).

Step 1: Participants **“Start”** the application on their smartphones, establishing a unique session with the local server.

Step 2: The user places their hand in the Palm Monster Generator Box. The system scans the palm gesture and generates a unique monster.

Figure 3
Initiation and Asynchronous Generation

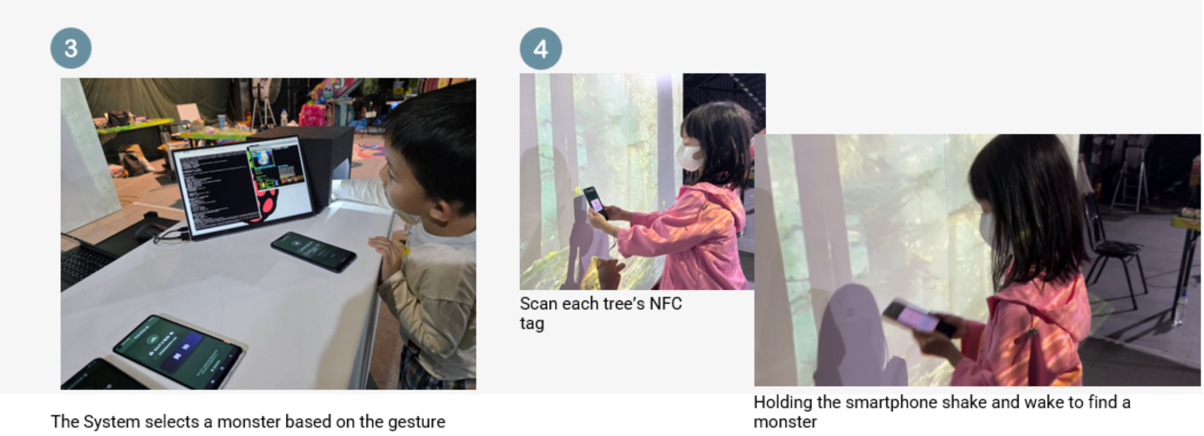


Press the “Start” button and go over.
Use your hand to reach into the box to create little monsters

Step 3: The generated monster is virtually “hidden” behind a specific physical tree trunk based on the gesture ID.

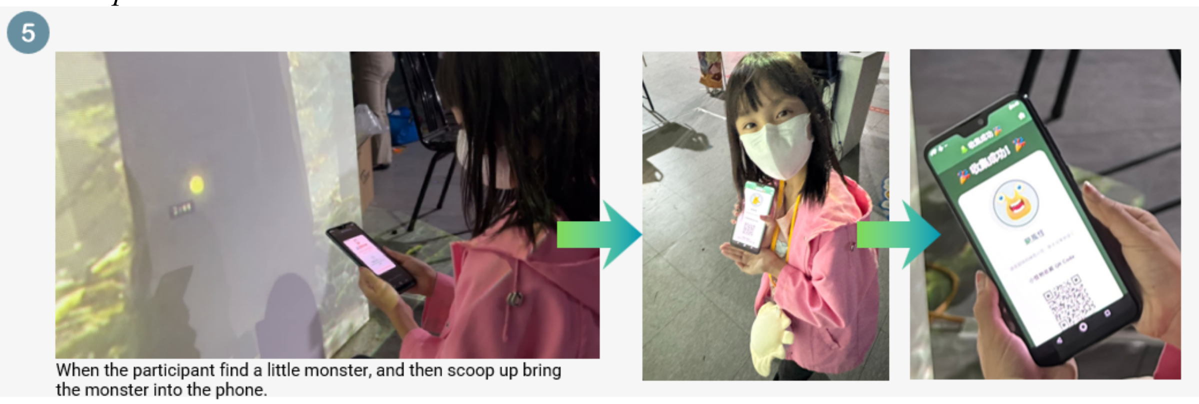
Step 4: Participants near the projection area (forest scene) and use their smartphones to scan NFC tags on trees. They utilize a “Shake-Wake” motion to find their monster.

Figure 4
Virtually Hidden and Locational Synchronization



Step 5: Once found, the monster appears on the projection. Users perform a "scoop up" motion with their phone to transfer the monster from the wall to their device.

Figure 5
Final Capture and Data Persistence



This five-step sequence represents the complete user experience loop, extending from physical input to final digital asset capture. The core engineering and design achievement of this flow is the successful implementation of the Latency Masking Strategy. By strategically embedding the high-latency generation time (Steps 1 & 2) within the physical movement and search actions (Steps 3 & 4), the system successfully mitigates the perceived effect of computational delays. The entire process validates the orchestration of heterogeneous components (computer vision, MQTT communication, NFC sensing) into a continuous and cohesive physical-digital experience, effectively resolving the interaction discontinuity typically found in edge-based generative AI applications.

Conclusions

Results and Core Design Contributions

The AIoT Anthropomorphic Little Monster Interactive System was successfully deployed as an "in-the-wild" exhibit at the HsinChu Science Park during the 2025 Halloween event. The deployment served as a critical validation of the system's stability and the effectiveness of the interaction model. We also successful deployment of this system provided empirical validation for its design and technical architecture. The system demonstrated robust stability under public usage, successfully processing over 320 interactions from more than 100 children. This high engagement confirmed the viability of the AIoT Framework and its capacity to maintain asynchronous communication between heterogeneous components (AI, IoT, Mobile). The synthesis of Kawaii Theory and Anthropomorphism successfully transforms potential fear of the unknown into controllability and cuteness. This design fosters curiosity and motivation for growth through the creation of a “Safe Sense of Adventure.” Furthermore, the system addresses a critical engineering challenge by integrating a complex ComfyUI-based generative pipeline with a Latency Masking Strategy. The physical “Shake-to-Wake” ritual and exploration sequence are strategically utilized to conceal the necessary AI generation time, effectively resolving the issue of high computational latency in real-time edge AI applications.

Future Works

We will focus on deepening the system’s interactive and educational capabilities by developing three primary extensions. First, the platform’s emotional feedback loop will be closed through the design of an AI emotion recognition model. This model will enable the anthropomorphic little monster to react to children’s real-time emotional states, such as smiling, surprise, and fear, directly using these states as inputs into the monster’s behavior. Second, we plan to enhance multi-player interaction by exploring the integration of diverse sensory stimuli, including sound, images, haptic feedback via a Tangible mobile phone grip, and NFC sensing. This expansion will be supported by an online extension platform (AIoT Cross-Media System), a specialized APP/Web bridge utilizing the combined potential of AI and IoT technology to form a cohesive cross-media interactive environment. Finally, the system’s utility will be broadened through educational applications, integrating the technology with STEM curricula, emotional values, and creativity. This will involve collaboration with actual educational institutions to conduct rigorous on-site teaching and data analysis to evaluate learning outcomes.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

The author declares that Grammarly, an AI-assisted writing software, was used in proofreading and refining the language used in the manuscript. The usage was limited to correcting grammatical and spelling errors and rephrasing statements for accuracy and clarity. The author further declares that, apart from Grammarly. The ideas, design, procedures, findings, analyses, and discussion are originally written and derived from careful and systematic conduct of the research.

References

comfyanonymous. (2023). *ComfyUI: The most powerful and modular stable diffusion GUI* [Computer software]. https://github.com/comfyanonymous/ComfyUI

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review, 114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Google AI. (2020). *MediaPipe* (Version 0.8+). https://github.com/google-ai-edge/mediapipe.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science, 315*(5812), 619. https://doi.org/10.1126/science.1134475

Nittono, H. (2016). The two-layer model of 'kawaii': A behavioural science framework for understanding kawaii and cuteness. *East Asian Journal of Popular Culture, 2*(1), 79–98. https://doi.org/10.1386/eapc.2.1.79_1

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. CSLI Publications; Cambridge University Press.

Rombach, S., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). *High-Resolution Image Synthesis with Latent Diffusion Models*. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 10674–10685. https://doi.org/10.48550/arXiv.2112.10752

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Contact email: shihta@mx.nthu.edu.tw