

Designing Personalized and Situated Conversational AI in Urban Public Space

Taery Kim and Kyuha Shim

Korea National University of Arts, Seoul, South Korea

ABSTRACT

Conversational AI has largely evolved within private, device centered environments that assume stable attention and extended engagement. When such systems enter shared urban space, these assumptions break down. This paper analyzes a real world intervention at COEX, now GangnamEyes, in Seoul, where a digital character named Lee Sol appears on a large media facade and invites passersby to initiate a mobile conversation through a QR code. The facade establishes a visible public AI presence, while each visitor privately engages in a short, five turn session on their smartphone. Presence is shared; interaction is individualized and time constrained. The project introduces a hybrid interaction model in which AI visibility and AI dialogue are spatially decoupled. Mobility, environmental noise, economic limits, and institutional goals shaped a constrained conversational architecture that favored tap based progression and minimal situational personalization over open ended dialogue. While this structure supported legibility and flow in a high density setting, it also exposed limitations in durational continuity and agent identity. The paper argues that contextual framing must guide conversational AI in shared environments. Meaningful spatial presence depends on clearly defining an agent's scope and temporal role rather than expanding conversational capability alone.

Keywords: Spatial AI presence, Hybrid interaction model, Conversational agents, Human centered design, Contextual framing

INTRODUCTION

Conversational AI has matured within the domain of personal devices. Most assistants are designed for environments in which users are stationary, attentive, and alone or semi private. Interaction unfolds on a single screen that houses both the agent and the exchange. Designers optimize for natural language flow, increased personalization, and richer multimodal input, often assuming that conversational depth corresponds with experiential quality. This trajectory has shaped expectations about what conversational systems should strive toward, including longer dialogues, persistent memory, and seamless speech interaction. Prior research has shown that user expectations frequently exceed what conversational systems can meaningfully sustain, especially when dialogue is framed as open ended interaction (Luger & Sellen, 2016).

When conversational AI extends into shared urban environments, these assumptions begin to fracture. Public space is structured by movement, distraction, noise, and social exposure. Attention is fragmented and time is compressed.

Encounters are often unplanned. Interaction in such settings unfolds within situated and embodied conditions rather than within stable device contexts, as Dourish argues in his account of embodied interaction (Dourish, 2001). An AI agent introduced at architectural scale must address heterogeneous audiences simultaneously while accommodating highly individual trajectories. The interface is no longer a contained screen but a distributed condition spanning built surface and personal device. The question therefore shifts from how to deepen dialogue to how to choreograph presence across spatial layers.



Figure 1: GangnamEyes Media Facade in urban context. (Image courtesy of CJ CGV, 2025).

The intervention examined in this paper took place at COEX, now GangnamEyes, a high density commercial and event complex in Seoul. A digital character named Lee Sol appears periodically on multiple media facades and invites passersby to scan a QR code. The facade establishes visual recognition at architectural scale. Once scanned, users enter a mobile web session where Lee Sol offers guidance about events, activities, and places within the complex. The interaction is limited to five turns and concludes with a satisfaction prompt. The system collects contextual information including who the visitor came with and what types of suggestions were requested. The client's intention was to position Lee Sol as an AI companion integrated into the identity of GangnamEyes while also generating useful backend insight into visitor behavior.

This paper approaches the intervention as a design inquiry into how conversational agents behave when presence and dialogue no longer occupy the same space. The central contribution is a hybrid interaction model in which AI presence is publicly staged while conversational exchange is privately instantiated. Through analysis of design decisions, constraints, and

shortcomings, the paper develops a critical perspective on spatial AI presence. It argues that contextual framing, defined as the deliberate articulation of an agent's scope, duration, and relational role within a specific environment, must guide design before capability expansion is pursued, aligning with contemporary calls for human centered AI (Shneiderman, 2020).



Figure 2: Transitional screen between public display and private session. (Image courtesy of CCID, 2026).

FRAMING CONVERSATIONAL AI IN SHARED ENVIRONMENTS

Conversational systems have historically been shaped by the logic of device centered interaction. The agent appears where the user inputs and receives output. Identity, responsiveness, and interface share a single frame. This arrangement simplifies mental models and concentrates accountability within one visible surface. Media facades, by contrast, have evolved primarily as broadcast systems. They operate at scale, address collective audiences, and privilege visibility over interactivity. Even when interactive elements are introduced, they typically involve lightweight participation or aggregated feedback rather than individualized dialogue. Research on urban media facades describes them as architectural interfaces that transform buildings into communicative surfaces while maintaining their collective orientation (Dalsgaard & Halskov, 2010).

The GangnamEyes intervention occupies a space between these established categories. Lee Sol is neither a conventional chatbot nor a traditional facade animation. The agent is publicly visible yet privately responsive. Each user initiates a separate session that unfolds asynchronously from others. The facade does not display user inputs or adapt to individual exchanges. The conversational system does not maintain persistent memory across visits. What emerges is a distributed condition in which visibility and dialogue are separated, and yet must remain conceptually coherent.

This separation produces a design problem that is not reducible to interface layout. The public layer confers legitimacy and scale, while the private layer performs responsiveness and limited personalization. The agent's identity must remain stable across both, even though its embodiment shifts and its

conversational depth is constrained. The hybrid model thus requires designers to treat spatial distribution as an architectural property of interaction rather than as a mere technical integration between screen and phone.

SPATIAL DECOUPLING AND THE ARCHITECTURE OF PRESENCE

Spatial decoupling describes the condition in which the agent's presence is anchored in public space while its dialogue is executed privately on personal devices. In the GangnamEyes intervention, Lee Sol's human form appears on large facades as a recognizable ambassador for the complex. The QR code remains visible throughout her appearance, inviting voluntary transition into a mobile session. Once on the phone, her face fades after a brief introductory animation, and the interface shifts to animated gradients that indicate system activity. Voice output continues in Lee Sol's pre-generated tone, but visual embodiment recedes.

This redistribution of embodiment reflects a deliberate calibration. Sustaining a fully anthropomorphic figure across both layers could have amplified expectations of relational depth that the five turn structure could not fulfill. By limiting visual embodiment in the private session while maintaining vocal continuity, the design signals connection without overcommitting to simulation. The gradient animations function as computational feedback rather than as character performance. The facade establishes recognition and the mobile interface manages exchange.

Decoupling also shapes temporality. The facade operates on a schedule visible to all. Private sessions begin and end independently. There is no shared conversational event, only multiple parallel exchanges unfolding under one visible identity. This asynchronous structure complicates the notion of collective interaction: the public layer invites and the private layer responds. Neither acknowledges the other in real time. The agent therefore exists simultaneously as a spatial figure and as many ephemeral conversational instances. This architecture succeeded in making AI presence legible at scale while enabling individualized engagement. At the same time, it also revealed limitations. Without synchronization or durable artifacts, sessions left no trace beyond memory. The agent's identity was recognizable but not cumulative. The spatial decoupling that allowed hybrid presence also prevented durational continuity and presence was achieved as encounter rather than as sustained relationship.

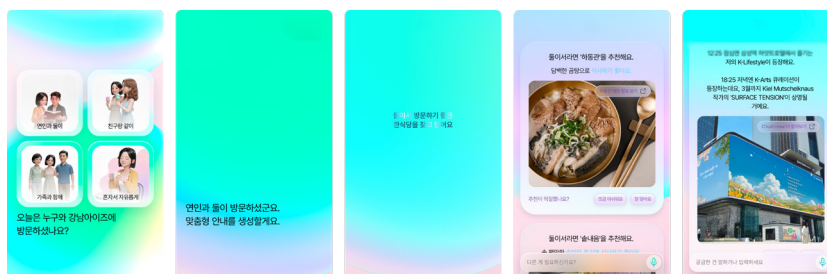


Figure 3: Conversational interface of Lee Sol. (Image courtesy of CCID, 2026).

MOBILITY, MODALITY, AND CONSTRAINT

Public environments impose kinetic and auditory conditions that challenge conversational depth. Visitors at GangnamEyes are typically in motion, navigating crowds and competing stimuli. Speech input is technically available through dictate functionality, and free text entry remains possible at each turn. At the same time, every step of the interaction presents three curated tap options that guide the exchange. The system speaks responses aloud while displaying text. These layered modalities coexist within a structure optimized for rapid progression.

The presence of speech output alongside tap based input does not represent inconsistency but hierarchy. Guided interaction is structurally reinforced. The five turn architecture was shaped partly by economic constraint and partly by flow considerations. Longer exchanges would have increased computational cost and cognitive burden. The result is a conversation that functions more as navigational orientation than as exploratory dialogue. This constraint preserved coherence under mobility. Users could abandon the session at any moment without confusion. Suggestions were delivered concisely and feedback was requested before closure. Yet, the compression also limited depth and the guidance remained general because there was insufficient time to refine preferences. Satisfaction prompts occurred before suggested experiences unfolded, capturing reaction to interaction rather than evaluation of outcome. The structure optimized throughput but constrained meaning.

The project demonstrates that capability does not automatically translate into appropriateness. Speech to text and extended generative dialogue were feasible, but environmental conditions rendered them unstable as primary channels. Tap based progression supported walking users: constraint became an organizing principle. However, constraint alone does not define value, and it must be paired with clarity about what the system is meant to accomplish within the moment it occupies.

IDENTITY, INSTRUMENTALITY, AND THE QUESTION OF WHAT THE AGENT IS

The initial brief positioned Lee Sol as a guidance companion integrated into the identity of GangnamEyes while also serving institutional interests through data collection. This dual objective shaped tone, phrasing, and backend architecture. Lee Sol asks whether visitors are alone, with friends, with a partner, or with family, and the system responses incorporate fragments of user language to create subtle personalization. Suggestions sometimes adopt relational phrasing, as when responding to a user expressing loneliness with an invitation framed as shared activity. At the final turn, the system requests satisfaction feedback and logs contextual data about engagement.

These layered functions expose an unresolved question. Is Lee Sol primarily a spatial host, a conversational guide, a brand ambassador, or a data instrument? The interaction structure supports aspects of each role but does not fully inhabit any. The companion tone gestures toward relational presence, yet the absence of memory or follow up limits emotional continuity. The guidance function provides orientation, yet suggestions

remain broad due to time constraints. The data collection function benefits the institution, yet remains peripheral in user awareness. This ambiguity can be a valuable design insight: spatial AI presence demands ontological clarity. An agent that occupies architectural scale carries symbolic weight. If its conversational depth is limited, that limitation must align with its declared role. Without explicit definition, hybrid systems risk producing spectacle without durable meaning. In this case, Lee Sol became visually integral to GangnamEyes while conversationally episodic. The encounter was engaging yet rarely transformative. The system greeted visitors effectively, but it did not accompany them beyond the immediate exchange.

CONTEXTUAL FRAMING AS FOUNDATIONAL DESIGN PRACTICE

The GangnamEyes intervention illustrates that contextual framing must precede technological expansion when conversational AI enters shared space. Contextual framing involves defining the agent's scope, duration, relational boundaries, and institutional function before optimizing capability. In this project, mobility and environmental noise were addressed through interaction constraint. Economic limits shaped turn length, visual identity reinforced spatial recognition; yet, the deeper framing question of what Lee Sol should be within the ecology of GangnamEyes remained partially open. Future spatial AI systems must confront this framing question directly. If the agent is a momentary guide, its value should lie in precision and clarity rather than in affective performance. If it is a recurring host, mechanisms for memory and follow up must be designed. If it serves as a data interface, transparency around collection and purpose becomes central. Each direction implies different structural commitments. Hybrid presence alone does not resolve these commitments.

The success of the project lies in demonstrating that publicly visible and privately instantiated conversational systems are viable and legible within dense urban environments. Its shortfall lies in revealing that legibility does not guarantee durability. The hybrid model is a foundation, and meaningful spatial AI presence requires further articulation of identity, duration, and consequence. Contextual framing, understood as the disciplined definition of what an agent is allowed to become within its environment, offers a path forward.

CONCLUSION

As conversational AI expands beyond personal devices, designers must reconsider the relationship between presence, dialogue, and space. The GangnamEyes case demonstrates a hybrid architecture in which public visibility and private interaction coexist through spatial decoupling. It shows how mobility aware constraint and layered modality can support engagement under high density conditions. It also reveals the limitations of compressed and episodic exchange when identity and duration remain under-defined.

Designing for meaningful spatial AI presence requires more than scaling existing assistants onto architectural surfaces. It requires treating spatial

distribution as interaction architecture and contextual framing as foundational design practice. By critically examining both the strengths and shortcomings of this intervention, the paper contributes a grounded perspective on how conversational systems may inhabit shared environments without defaulting to technological maximalism (Greenfield, 2017). The challenge ahead lies not in expanding capability for its own sake, but in defining what kind of presence is appropriate, sustainable, and responsible within the spaces we share.

ACKNOWLEDGMENT

The authors would like to acknowledge the design researchers at the CCID Center for Creative Intelligence Design who contributed to the design and development of the GangnamEyes intervention. We also thank Naver Cloud for their technical collaboration and infrastructure support, and CJ CGV for their partnership and trust as the primary client in realizing this project.

REFERENCES

- Dalsgaard, P. and Halskov, K., 2010. Designing urban media facades: Cases and challenges. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. New York: ACM, pp. 2277–2286.
- Dourish, P., 2001. *Where the action is: The foundations of embodied interaction*. Cambridge, MA: MIT Press.
- Greenfield, A., 2017. *Radical technologies: The design of everyday life*. London: Verso.
- Luger, E. and Sellen, A., 2016. Like having a really bad PA: The gulf between user expectation and experience of conversational agents. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. New York: ACM, pp. 5286–5297.
- Shneiderman, B., 2020. Human-centered artificial intelligence: Reliable, safe and trustworthy. *International Journal of Human–Computer Interaction*, 36(6), pp. 495–504.