Collective Object Intelligence: Coordinating AI-Driven Household Objects with Large Language Models

Ethan Chang
MIT
Cambridge, USA
echang25@mit.edu

Matthieu J. Huss
Mosaic AI
San Francisco, USA
mhuss@mosaicai.ca

Alexander Htet Kyaw
MIT
Cambridge, USA
alexkyaw@mit.edu

Yuqing Lucy Li
MIT
Cambridge, USA
lucy_li@media.mit.edu

Keunwook Kim
Harvard University
Cambridge, USA
keunwook@kimkeunwook.com

Brady Lin
Northwestern University
Evanston, USA
bradyyqlin@gmail.com

Cristiano Da Silva
Harvard University
Cambridge, USA
cristiano_da@seas.harvard.edu

Marcelo Coelho
MIT
Cambridge, USA
marceloc@mit.edu

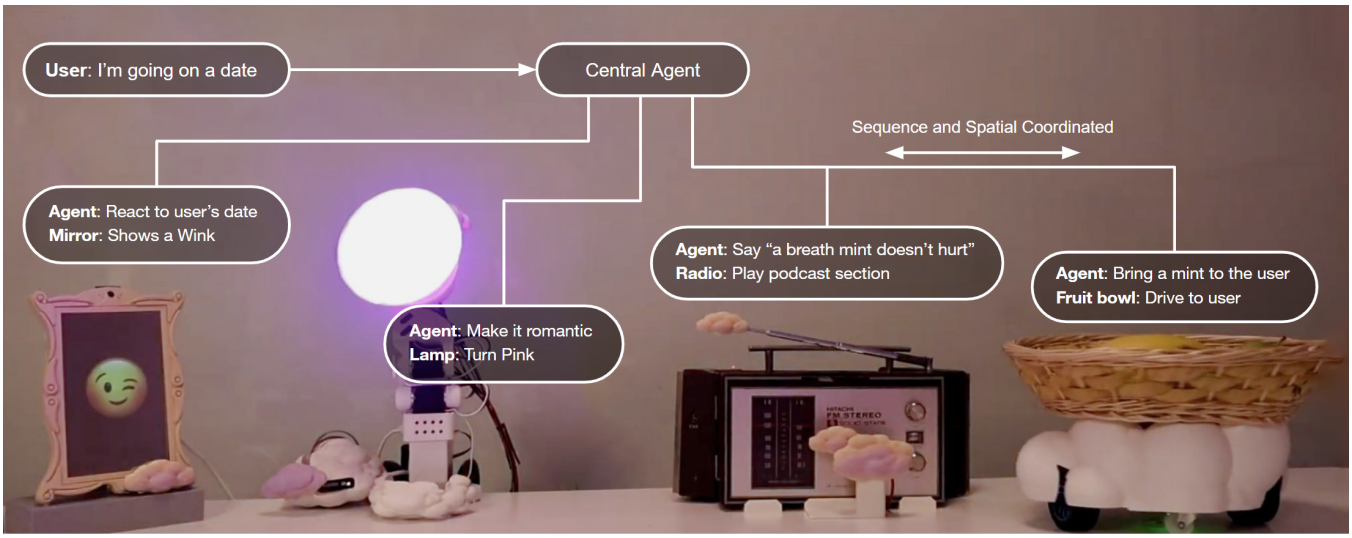


Figure 1: Example scenario of Collective Object Intelligence. The user says they are going on a date, and AI-augmented household objects coordinate through their own modalities to respond.

Abstract

Current smart home systems largely rely on fixed automation, predefined triggers, and centralized control interfaces. We present Collective Object Intelligence, a two-tier orchestration approach for coordinating AI-augmented household objects. A central agent generates a shared plan from live user and environmental context, while local object agents interpret and execute that plan through object-specific sensing and action. Our prototype connects a lamp, radio, mirror, and fruit bowl whose roles remain grounded in their physical affordances. This paper presents three core elements of the approach: an orchestration architecture for multi object coordination, an affordance role mapping that links object form and function to the intelligence it can express, and an analysis of failure modes. We implement this system and demonstrate four application scenarios. Our research suggests that in future domestic environments, intelligence may be expressed through the different modalities of collective objects themselves rather than through a centralized interface.

CCS Concepts

• **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**; **Interaction devices**; • **Computing methodologies** → *Artificial intelligence.*

Keywords

AI-augmented objects, smart homes, embodied interaction, orchestration, ubiquitous computing, demonstration

ACM Reference Format:
Ethan Chang, Matthieu J. Huss, Alexander Htet Kyaw, Yuqing Lucy Li, Keunwook Kim, Brady Lin, Cristiano Da Silva, and Marcelo Coelho. 2026. Collective Object Intelligence: Coordinating AI-Driven Household Objects with Large Language Models. In *The 39th Annual ACM Symposium on User Interface Software and Technology (UIST Adjunct '26), November 02–05, 2026,*

This work is licensed under a Creative Commons Attribution 4.0 International License.

UIST Adjunct '26, Detroit, MI, USA

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2855-6/2026/11
https://doi.org/10.1145/3830397.3842458

Detroit, MI, USA. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3830397.3842458

1 Introduction

Objects in the home carry meaning through both function and form: a radio speaks, a lamp creates ambience, a mirror supports reflection, and a bowl holds or delivers small items. AI augmentation complicates this legibility. Open-ended computational capabilities can detach behavior from physical form, but they also create an opportunity for familiar objects to coordinate rich behavior while expressing intelligence through their own modalities.

Current smart-home systems commonly centralize interaction in assistants such as Alexa or Google Home, or coordinate appliances through fixed automations such as IFTTT applets. These systems often treat devices as endpoints for commands rather than expressive participants in an ecology. Inspired by Large Language Objects [8] and artifact ecologies [1], we investigate how heterogeneous AI-augmented objects can act collectively without losing their recognizable identities.

This raises a central interaction-design question: how can generative systems coordinate heterogeneous objects while keeping each action attributable to a familiar physical identity?

Coordination here is more than executing several device commands. The system must decide which objects participate, how actions unfold across time and space, and whether intervention is appropriate. This creates a design tension between generative flexibility and material legibility. A response may be feasible yet confusing when object roles, timing, or audience are unclear.

We present **Collective Object Intelligence**, a two-tier orchestration approach that separates centralized coordination from distributed embodiment. In our prototype, a central agent interprets live context and assigns goals to local agents embedded in a lamp, radio, mirror, and mobile fruit bowl. Each local agent translates its goal into hardware-specific behavior such as lighting, movement, speech, display, or navigation. This paper contributes: (1) a two-tier architecture for live multi-object coordination; (2) an affordance-role mapping that connects object form and function to agent behavior; and (3) an exploratory evaluation of four scenarios with 17 participants.

2 Related Work

Smart homes and intelligible coordination. Smart-home research has examined context-aware behavior [9, 11], how people appropriate connected-device ecosystems [27, 29], and the need for intelligible control and accountability [5, 13, 16]. These systems typically coordinate devices through predefined capabilities and rules. Collective Object Intelligence instead uses generative agents to select and express object-specific actions from live context, making the legibility of both individual roles and collective behavior a central design concern. Rule-based smart-home ecosystems offer predictability because actions and relationships are specified in advance [10]. Generative coordination relaxes this requirement but creates uncertainty about why devices act, who is responsible, and how users can interrupt an emerging plan. We therefore treat intelligibility as a property of orchestration, not only of individual devices.

Artifact ecologies and augmented objects. Artifact-ecology research shows how meaning and activity emerge across heterogeneous artifacts rather than a single interface [1, 2, 17, 18]. Recent systems give individual objects conversational, expressive, or generative capabilities [3, 4, 6, 12, 15, 25, 26, 28]. These works establish objects as computational and social actors, but primarily examine one augmented object at a time. We extend this framing from an intelligent object to an ecology whose members retain distinct material identities. Moving from one augmented object to an ecology also introduces questions of composition: whether actions reinforce or compete with one another, how attention shifts among objects, and how users attribute intent across the collective.

Generative orchestration. Proactive assistants increasingly use multimodal context to infer intent and select actions [7, 22, 24]. Related generative systems translate language, gesture, movement, and scene context into digital or physical outcomes [14, 19–21]. Our work addresses the underexplored middle ground between rule-based multi-device control and open-ended single-object agents: live generative coordination across heterogeneous embodied objects. The contribution is not merely routing one command to several devices, but mapping a shared goal onto roles that remain appropriate to each object's sensing, actuation, form, and social meaning. This makes orchestration an interaction-design problem of deciding which objects should act, in what order, and when no action is appropriate.

3 System Design

Collective Object Intelligence uses a two-tier architecture in which a central agent coordinates heterogeneous objects, while local agents ground shared goals in device-specific capabilities (Figure 2). The central agent integrates user speech, visual observations, and device state; interprets the situation; and generates a coordinated plan. Reasoning is therefore centralized, while interpretation and physical execution are distributed across the objects.

Each object hosts a local agent that receives a high-level goal and selects hardware-specific actions. Local agents retain device parameters and constraints, such as display dimensions, speaker volume, actuator range, and available tools. This division lets the central agent determine *what* contribution is needed while each local agent determines *how* its object can express that contribution.

The prototype uses concise natural-language messages between the central and local agents rather than requiring one shared action schema. Local agents interpret these goals and call device tools for setting light, actuating motors, playing or generating audio, displaying imagery, and navigating. Natural language does not guarantee interoperability; instead, it reduces the need to encode

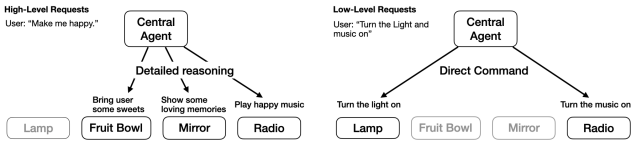


Figure 2: Two-tier reasoning process. A central agent interprets live context and assigns complementary goals; local agents translate those goals into object-specific tool calls.

	Lamp	Radio	Mirror	Fruit Bowl
Capabilities	• Change light color • 6-DoF arm movement	• Play songs or dialogue clips • Generate speech	• Capture images • Display images or text	• Drive on flat surfaces • Carry objects
Applications	• Control the room's mood • Blink to gain attention • Express through motion	• Provide audio information • Insert sound effects • Interject with dialogue	• Provide emoji feedback • Show generated images • Display memories	• Bring items to the user • Nudge a user physically • Manipulate desk objects

Table 1: Physical capabilities and application affordances of each object in the system.

every cross-device action centrally while preserving device-specific control locally. An event-driven pipeline routes speech, camera observations, and device feedback through the central reasoning loop. During a demonstration, both the coordinated plan and local actions are generated and executed live.

4 Augmented Household Objects

The prototype connects four Raspberry Pi-based objects to the orchestration layer. Table 1 makes the affordance-role mapping explicit: each local agent selects actions that are technically available and legible for its object.

Lamp. The lamp is built around a servo-driven articulated arm and can point, gesture, and perform full-body expressive movement. It modulates brightness and color to convey emotional or ambient context and can use movement to direct attention.

Radio. The radio is a vintage Hitachi unit that communicates through brief remarks, music, dialogue clips, and generated speech. A servo mechanism physically rotates the tuning dial, linking AI-generated audio behavior to the familiar material language of a radio.

Mirror. The mirror uses an LCD display behind one-way mirror film to produce visual overlays on a reflective surface. An onboard camera allows it to capture the user and environment, which can be passed to generative AI models to produce visual suggestions or contextual imagery.

Fruit Bowl. The fruit bowl is a wheeled platform with differential steering. It can navigate toward users or objects, deliver small items, and perform rocking or spinning movements in response to music or cues from other objects.

5 Demonstration and Applications

The demonstration presents four connected scenarios: (1) a user is nervous before a date; (2) the user prepares for the date and asks for help; (3) the user and a second person deal with conflict after a late arrival; and (4) the users leave the house for the date (Figure 4). Across these scenarios, we explore both reactive and proactive behavior. Reactive behaviors follow direct engagement, such as the user asking the objects for help. Proactive behaviors begin when the system observes a situation or contextual cue and elects to respond without an explicit request. The scenarios also vary sequential and simultaneous coordination. Although the situations and user prompts define each vignette, the responses are not replayed or manually triggered: the central and local agents generate and execute the coordination live during each run.

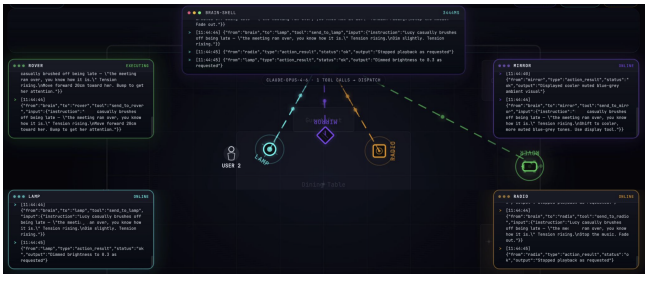


Figure 3: Coordination interface showing shared context, the central plan, and live object-level actions.

For example, in figure 4, when the user announces an upcoming date, the lamp creates a warm ambience, the radio responds and plays music, and the bowl moves expressively. When asked for help, the radio suggests a mint, the bowl delivers it, and the mirror generates an accessory visualization. When a second person arrives late, the radio comments and the bowl moves toward that person. Finally, as the users leave, the objects generate a celebratory response. Together, the scenarios exercise ambient, requested, sequential, and simultaneous coordination while preserving distinct object roles.

These scenarios demonstrate how different household objects can contribute to a shared interaction using their own modalities. The radio provides speech and sound, the fruit bowl provides physical delivery and movement, the mirror provides visual reflection and imagery, and the lamp provides ambience and gesture. Rather than treating the objects as interchangeable devices, the system assigns roles based on physical affordance and communicative fit.

Beyond the date scenario, the same orchestration approach could support broader applications in domestic life. A coordinated object ecology could help users prepare for daily routines, remember tasks, create ambient transitions between activities, assist with lightweight household logistics, or make social gatherings more expressive.

6 Preliminary Evaluation

We conducted an exploratory video-based evaluation with 17 participants recruited without screening for demographic background or prior smart-home and AI experience. Participants watched recorded demonstrations of the four live-agent scenarios. After each video, they answered five seven-point Likert-type items addressing whether the system understood the situation, object fit, whether the responses made sense, whether coordination worked, and whether behavior aligned with expectations. Each questionnaire also included an optional open-ended prompt asking participants to explain their ratings. We use these responses as descriptive context for interpreting the numerical results rather than as a formal thematic analysis.

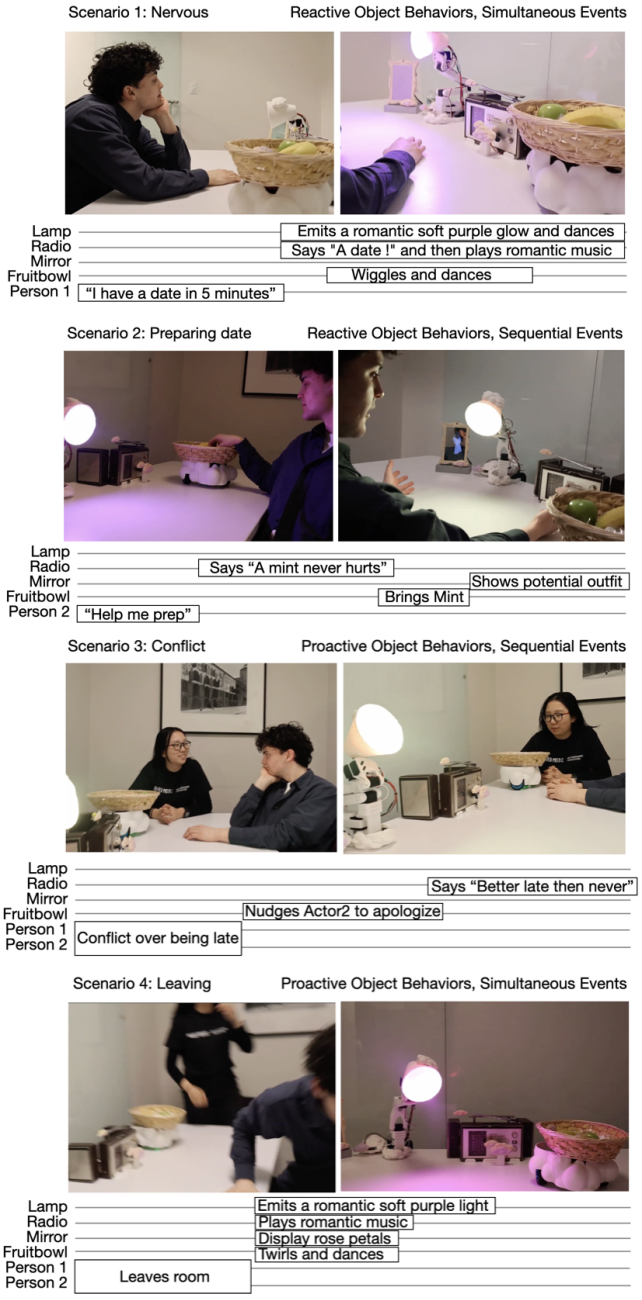


Figure 4: Object orchestration across four scenarios

7 Results

Table 2 reports each item separately. Scenario 2 received the highest rating on all five items, while Scenario 3 received the lowest ratings overall. Open-ended responses described the system as playful and human-like and identified the radio's speech and the bowl's delivery as recognizable contributions. These descriptive results suggest a difference between reactions to practical assistance and autonomous social intervention.

	Q1	Q2	Q3	Q4	Q5
S1: Nervous	5.29	5.35	5.35	5.29	4.94
S2: Preparing	**6.29**	**6.24**	**6.25**	**6.29**	**5.76**
S3: Conflict	5.00	5.06	5.00	4.71	4.29
S4: Leaving	5.71	5.82	5.65	5.65	5.24

Table 2: Mean ratings by scenario on a 1–7 scale. Q1 = system understood the situation; Q2 = object fit; Q3 = response made sense; Q4 = coordination worked; Q5 = behavior aligned with expectations. The shaded row is the highest-rated scenario.

8 Discussion

The demonstration suggests that collective behavior remains legible when each contribution is grounded in a familiar capability: the radio speaks, the bowl delivers, the mirror displays, and the lamp shapes ambience. The affordance-role mapping provides a practical constraint on open-ended generation. The evaluation further suggests that practical assistance may be a safer initial domain than autonomous social mediation. Participants rated lightweight preparation support highest and intervention in a late-arrival interaction lowest. This pattern should be treated as a design hypothesis rather than a causal finding. Future studies should compare multiple generated responses to the same situation and examine how users redirect, approve, or stop an emerging collective plan.

The contrast between Scenarios 2 and 4 suggests that successful coordination depends not simply on how many objects participate, but on whether each action has a legible purpose, recipient, and moment. In Scenario 2, participants recognized a sequential handoff: the radio suggested a mint and the bowl delivered it. P3 described the objects as working together, while P1 compared them to parts of a shared body. In Scenario 4, however, the objects continued performing after the users had left, prompting P16 to question who the objects were performing for. This contrast expands the role of the central agent beyond distributing goals. It must also manage attention, determine whether actions should be sequential or simultaneous, and suppress contributions that no longer serve the interaction. For collective object systems, choosing not to act may therefore be as important as selecting an appropriate action.

The study is limited by its small unscreened sample, video-based presentation, four-object prototype, and single scenario theme. Although the videos captured live agent runs, watching a recording cannot reproduce the timing, proximity, sound, and unpredictability of sharing a room with moving objects. Longer deployments should also examine privacy, safety, failure recovery, and how much authority users want to delegate to central and local agents.

9 Conclusion

Collective Object Intelligence coordinates familiar household objects through centralized planning and distributed, object-specific execution. The two-tier architecture and affordance-role mapping allow generative agents to act collectively while preserving legible material roles. Across four live-agent demonstrations, exploratory participant responses were most positive for practical assistance and least positive for social intervention. These results motivate further study of how people supervise and shape generative coordination across embodied object ecologies.

References

[1] Susanne Bødker and Clemens Nylandsted Klokmose. 2012. Dynamics in Artifact Ecologies. In *Proceedings of NordiCHI '12*. ACM, 448–457. doi:10.1145/2399016.2399085.

[2] Susanne Bødker, Peter Lyle, and Joanna Saad-Sulonen. 2017. Untangling the Mess of Technological Artifacts: Investigating Community Artifact Ecologies. In *Proceedings of C&T '17*. ACM, 246–255. doi:10.1145/3083671.3083675.

[3] Ethan Chang, Zhixing Chen, Jb Labrune, and Marcelo Coelho. 2025. Be the Beat: AI-Powered Boombox for Music Suggestion from Freestyle Dance. In *Proceedings of TEI '25*. ACM, Article 67, 1–6. doi:10.1145/3689050.3705995.

[4] Ethan Chang, Quincy Kuang, and Marcelo Coelho. 2025. The Stochastic Parrot: A Physical AI Cohabitant. In *The Thirty-Ninth Annual Conference on Neural Information Processing Systems, Creative AI Track: Humanity*. openreview.net/forum?id=8QJWS9e6V5.

[5] Xiaohong Chen, Shi Chen, Zhi Jin, Han Bian, Zihan Chen, and Haotian Li. 2025. Expressing the Needs in Smart Home: What Is the End Users' Favorite Way. *ACM Transactions on Computer-Human Interaction* 32, 2, Article 16. doi:10.1145/3715114.

[6] Hyungjun Cho and Tek-Jin Nam. 2025. Living Alongside Areca: Exploring Human Experiences with Things Expressing Thoughts and Emotions. In *Proceedings of CHI '25*. ACM, 1–16. doi:10.1145/3706598.3713228.

[7] Hyunsung Cho, Jacqui Fashimpaur, Naveen Sendhilnathan, Jonathan Browder, David Lindlbauer, Tanya R. Jonker, and Kashyap Todi. 2025. Persistent Assistant: Seamless Everyday AI Interactions via Intent Grounding and Multimodal Feedback. In *Proceedings of CHI '25*. ACM, 1–19. doi:10.1145/3706598.3714317.

[8] Marcelo Coelho and Jean-Baptiste Labrune. 2024. Large Language Objects: The Design of Physical AI and Generative Experiences. *Interactions*.

[9] Devleena Das, Yasutaka Nishimura, Rajan P. Vivek, Naoto Takeda, Sean T. Fish, Thomas Pløtz, and Sonia Chernova. 2023. Explainable Activity Recognition for Smart Home Systems. *ACM Transactions on Interactive Intelligent Systems* 13, 2, Article 7. doi:10.1145/3561533.

[10] Scott Davidoff, Min Kyung Lee, Charles Yiu, John Zimmerman, and Anind K. Dey. 2006. Principles of Smart Home Control. In *Proceedings of UbiComp '06*. Springer, 19–34. doi:10.1007/11853565_2.

[11] Anind K. Dey. 2001. Understanding and Using Context. *Personal and Ubiquitous Computing* 5, 1, 4–7. doi:10.1007/s007790170019.

[12] Mustafa Doga Dogan, Eric J. Gonzalez, Karan Ahuja, Ruofei Du, Andrea Colaco, Johnny Lee, Mar Gonzalez-Franco, and David Kim. 2024. Augmented Object Intelligence with XR-Objects. In *Proceedings of UIST '24*. ACM, 1–15. doi:10.1145/3654777.3676379.

[13] W. Keith Edwards and Rebecca E. Grinter. 2001. At Home with Ubiquitous Computing: Seven Challenges. In *Proceedings of UbiComp '01*. Springer, 256–272.

[14] Rodrigo Gallardo, Oz Fishman, and Alexander Htet Kyaw. 2025. Scene-Aware Urban Design: A Human-AI Recommendation Framework Using Co-Occurrence Embeddings and Vision-Language Models. In *The Thirty-Ninth Annual Conference on Neural Information Processing Systems, Creative AI Track: Humanity*. doi:10.48550/arXiv.2511.06201.

[15] Nozomu Iwai and Fabrice Matulic. 2025. Bringing Everyday Objects to Life in Augmented Reality with AI-Powered Talking Characters. In *Proceedings of CHI EA '25*. ACM, 1–7. doi:10.1145/3706599.3719978.

[16] Timo Jakobi, Gunnar Stevens, Nico Castelli, Corinna Ogonowski, Florian Schaub, Nils Vindice, Dave Randall, Peter Tolmie, and Volker Wulf. 2018. Evolving Needs in IoT Control and Accountability: A Longitudinal Study on Smart Home Intelligibility. *Proc. ACM IMWUT* 2, 4, Article 171. doi:10.1145/3287049.

[17] Heekyoung Jung, Erik Stolterman, William Ryan, Tonya Thompson, and Marty Siegel. 2008. Toward a Framework for Ecologies of Artifacts: How Are Digital Artifacts Interconnected within a Personal Life? In *Proceedings of NordiCHI '08*. ACM, 201–210. doi:10.1145/1463160.1463182.

[18] Henrik Korsgaard, Peter Lyle, Joanna Saad-Sulonen, Clemens Nylandsted Klokmose, Midas Nouwens, and Susanne Bødker. 2022. Collectives and Their Artifact Ecologies. *Proc. ACM Human-Computer Interaction* 6, CSCW2, Article 432. doi:10.1145/3555533.

[19] Alexander Htet Kyaw, Miana Smith, Se Hwan Jeon, and Neil Gershenfeld. 2025. Speech to Reality: On-Demand Production using Natural Language, 3D Generative AI, and Discrete Robotic Assembly. In *Proceedings of SCF '25*. ACM, Article 16, 1–12. doi:10.1145/3745778.3766670.

[20] Alexander Htet Kyaw, Yifei Li, Jin Gao, Maggie Chao, Keyi Zhang, Yudian Xu, and Hiroshi Ishii. 2026. Tangible AI Mind-map: A Graph-Based Interface for Human-AI Brainstorming using Projection-Based AR, Gesture Recognition and Large Language Models. In *Proceedings of SIGGRAPH Posters '26*. ACM, Article 81, 1–3. doi:10.1145/3799825.3818764.

[21] Alexander Htet Kyaw, Richa Gupta, Dhruv Shah, Anoop Sinha, Kory Mathewson, Stefanie Pender, Sachin Chitta, Yotto Koga, Faez Ahmed, Lawrence Sass, and Randall Davis. 2025. Text to Robotic Assembly of Multi Component Objects using 3D Generative AI and Vision Language Models. In *The Thirty-Ninth Annual Conference on Neural Information Processing Systems, Creative AI Track: Humanity*. doi:10.48550/arXiv.2511.02162.

[22] Geonsun Lee, Min Xia, Nels Numan, Xun Qian, David Li, Yanhe Chen, Achin Kulshrestha, Ishan Chatterjee, Yinda Zhang, Dinesh Manocha, David Kim, and Ruofei Du. 2025. Sensible Agent: A Framework for Unobtrusive Interaction with Proactive AR Agents. In *Proceedings of UIST '25*. ACM, 1–22. doi:10.1145/3746059.3747748.

[23] Avital Meshi and Adam Wright. 2025. in(A)n(I)mate: AI-Mediated Conversations with Inanimate Objects. *Proceedings of the ACM on Computer Graphics and Interactive Techniques* 8, 3, Article 36. doi:10.1145/3736787.

[24] Jeesun Oh, Wooseok Kim, Sungbae Kim, Hyeonjeong Im, and Sangsu Lee. 2024. Better to Ask Than Assume: Proactive Voice Assistants' Communication Strategies That Respect User Agency in a Smart Home Environment. In *Proceedings of CHI '24*. ACM, 1–17. doi:10.1145/3613904.3642193.

[25] Marco C. Rozendaal, Boudewijn Boon, and Victor Kaptelinin. 2019. Objects with Intent: Designing Everyday Things as Collaborative Partners. *ACM Transactions on Computer-Human Interaction* 26, 4, Article 26. doi:10.1145/3325277.

[26] Xuetong Wang, Ching Christie Pang, and Pan Hui. 2025. Talking Spell: A Wearable System Enabling Real-Time Anthropomorphic Voice Interaction with Everyday Objects. In *Proceedings of UIST '25*. ACM, 1–17. doi:10.1145/3746059.3747617.

[27] Mikołaj P. Woźniak, Sarah Vöge, Ronja Krüger, Heiko Müller, Marion Koelle, and Susanne Boll. 2023. Inhabiting Interconnected Spaces: How Users Shape and Appropriate Their Smart Home Ecosystems. In *Proceedings of CHI '23*. ACM, 1–18. doi:10.1145/3544548.3581497.

[28] Ziwen Yan, Xuanxuan Liu, Jb Labrune, and Marcelo Coelho. 2025. Colorimia: Color Picking and Image Generation with a Physical AI. In *Proceedings of TEI '25*. ACM, 1–6. doi:10.1145/3689050.3705989.

[29] Yuan Yao, Li Huang, Yi He, Zhijun Ma, Xuhai Xu, and Haipeng Mi. 2023. Reviewing and Reflecting on Smart Home Research from the Human-Centered Perspective. In *Proceedings of CHI '23*. ACM, 1–21. doi:10.1145/3544548.3580842.