

Augmenting Human Cognition through Everyday AR

Abstract

As spatial computing and multimodal LLMs mature, AR is tending to become an intuitive "thinking tool," embedding semantic and context-aware intelligence directly into everyday environments. This paper explores how always-on AR can seamlessly bridge digital cognition and physical affordances, enabling proactive, context-sensitive interactions that enhance human task performance and understanding.

Keywords

Augmented Reality; Mixed Reality; Virtual Reality; Generative AI; Large Language Models; Computer Vision; Machine Learning; Human-AI Interaction

ACM Reference Format:

. 2025. Augmenting Human Cognition through Everyday AR. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 3 pages. <https://doi.org/XXXXXXX.XXXXXXX>

1 INTRODUCTION

A persistent theme in technology evolution is that once a foundational layer matures and becomes commoditized, the real competition moves to the "interface layer." Historically, we have seen this in the shift from mainframe computing to personal computers, then from web browsers to mobile devices. At each transition, an enabling technology becomes sufficiently widespread and affordable, opening space for a new, more intuitive interface.

We are now witnessing a similar pivot in the age of spatial computing and multimodal large language models (MLLMs). As head-worn devices shrink and edge connectivity matures, Augmented Reality stands to become the predominant "interface layer" that seamlessly merges the digital realm with everyday physical contexts. Paired with AI that can interpret speech, gaze, gestures, and environmental signals, AR can go beyond overlaying disjointed information to become a cognitive amplifier—a "thinking tool for reality." [3]

2 REALITY AS THE CANVAS

In this shifting paradigm, augmented reality moves from being a mere overlay of graphics to an immersive layer of intelligence woven seamlessly into our daily environment.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Conference acronym 'XX, Woodstock, NY

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-XXXX-X/2018/06
<https://doi.org/XXXXXXX.XXXXXXX>

2.1 AR as an Always-On Knowledge Layer

In a smartphone-centric world, we measure "engagement" by screen time: the number of minutes an app occupies in the foreground. But with always-on AR glasses, attention transforms. In AR, the user might glance at an object, utter a partial phrase, or make a subtle gesture—and instantly the system offers assistance. Our "context sharing"—the exchange of situational information between user and system—shifts from traditional apps into an *ambient, environment-embedded dialogue*. [4, 6]

For example, instead of having to switch to a separate translation app, you can simply look at text through your AR glasses, and the system will automatically translate or summarize it for you. If you say something like "I need more info on this brand" while viewing a product label, the AI in your glasses quickly retrieves the relevant information.

2.2 Moving from "Screens" to "Scenes"

The fundamental difference is that we are no longer locked into 2D windows. AR can unify your physical space into a single interactive "scene" where every wall, tabletop, or building façade becomes a potential surface for dynamic content. Our human attention is guided not by explicit screen boundaries, but by the inherent salience of physical and social cues.

This shift hints at deeper cognitive augmentation: the device can unobtrusively sense your context, reduce superfluous mental overhead, and highlight the most relevant information or actions at the exact moment you need them.

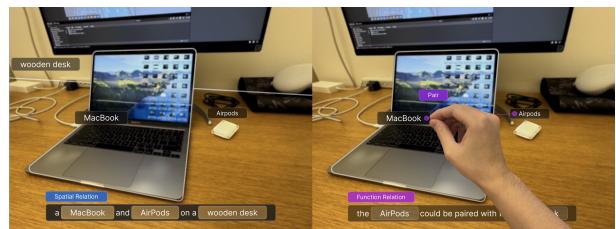


Figure 1: A scenario in AR where a MacBook and AirPods on a desk are labeled, with a 'Pair' prompt suggesting their connection, allowing users to initiate pairing with a gesture by directly manipulating the virtual connector.

Figure 1 demonstrates a basic example of how *always-on knowledge* can facilitate immediate insights about physical objects. Instead of manually searching for Bluetooth settings, a quick gesture or voice command in context triggers a pairing workflow.

3 THE SEMANTIC LAYER FOR REALITY

Reality computing involves embedding a semantic layer over the real world. This is not just about labeling objects or overlaying text; it is about endowing the AR system with an interpretive, context-sensitive understanding that lets it act as an extension of the user's cognition.

3.1 Natural and Contextual Interaction

By combining the user's gaze vector with voice or gesture input, AR systems can infer user intent even if commands are ambiguous. For instance, users might say "Change that," while glancing at a lamp, and the system interprets that they want to dim the light. Large language models—upgraded to handle images, 3D scans, or sensor data—can identify objects, interpret vague pronouns, track user references, and handle natural conversation. They become the semantic "bridge" tying real-world context to digital intelligence. [1, 5]

3.2 Proactive Thinking Tools

Adaptive guidance involves using an AR system to proactively walk the user through step-by-step tasks. Instead of requiring manual step-switching or referencing static instructions, the system "sees" what the user is doing in real time and autonomously advances through a tutorial. Figure 2 illustrates this idea with our "RealiTIPS" prototype.



Figure 2: RealiTIPS providing step-by-step guidance in AR.

What distinguishes RealiTIPS from traditional tutorials is that it advances instructions based on actual user progress, creating a more natural learning flow. Predictive suggestions further reduce mental overhead by anticipating user needs. For instance, if a user sets a half-assembled IKEA shelf on the floor, the system might proactively highlight the needed screwdriver and the correct set of screws.



Figure 3: A library scenario showing AI-assisted semantic search. The AR system recognizes multiple books, such as *Software Testing Foundations* and *Seven Databases in Seven Weeks*, allowing the user to ask for summaries, related topics, or best practices in context.

Figure 3 offers an example of combining *contextual interaction* and *semantic augmentation* in a library. The user can hover a finger

over a book's bounding box (e.g., *Software Testing Foundations*) and verbally request "summarize," "about the author," or "related books." The system harnesses large language models to provide concise text overlays or spoken feedback. This transforms a physical shelf-browsing experience into a context-rich knowledge discovery process.

4 LEVERAGING THE AFFORDANCES OF THE PHYSICAL WORLD

Heads-up AR displays uniquely exploit the "affordances" of the physical world [2]. By presenting spatially anchored information directly in one's field of view, these systems overlay rich digital context onto real-world cues (e.g., a handle that invites grasp, a knob that invites rotation), effectively scaffolding user cognition. Instead of memorizing sequences or referencing an external screen, the user can rely on direct, in-place guidance.

Figure 4 shows a prototypical scenario where a user assembles a small robotic platform with an Arduino. Each assembly step appears over the relevant part, with arrows, labels, or textual prompts. The system detects partial completion (e.g., whether screws are attached), updating the user's progress in real time. By integrating these instructions into the user's field of view and mapping them to physical objects, AR reduces the cognitive effort needed to understand or recall tasks, allowing users to better focus on the activity at hand. [7, 8]



Figure 4: Step-by-step instructions for assembling an Arduino-based robotic platform. The AR system tracks the user's progress and displays targeted prompts (e.g., "Attach the V5 Expanding Board"), reducing errors and accelerating learning.

This real-time interplay between the physical affordances of reality and contextual AI feedback not only accelerates task performance but also deepens understanding, as users learn how objects and actions connect in both physical and digital realms. The environment itself becomes a "living manual," where guidance and knowledge are actively anchored to items and surfaces, freeing users from distracting interface layers while still providing moment-to-moment assistance.

5 CONCLUSION

AR with MLLMs transcends being a mere interface, evolving into a potent "thinking tool" that bridges physical tasks and digital knowledge. From pairing headphones to exploring books and guiding cooking or robot assembly, reality computing seamlessly integrates with daily cognition. Ultimately, by embedding intelligence into our physical world, AR will enhance our understanding and extend cognitive capabilities, once the realm of science fiction.

References

- [1] Alex Endert, Patrick Fiaux, and Chris North. 2012. Semantic Interaction for Visual Text Analytics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, 473–482. doi:10.1145/2207676.2207741
- [2] James J Gibson. 2014. The theory of affordances:(1979). In *The people, place, and space reader*. Routledge, 56–60.
- [3] Eric Horvitz. 1999. Principles of Mixed-Initiative User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. ACM, 159–166. doi:10.1145/302979.303030
- [4] Hiroshi Ishii and Brygg Ullmer. 1997. Tangible Bits: Towards Seamless Interfaces Between People, Bits and Atoms. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '97)*. ACM, 234–241. doi:10.1145/258549.258715
- [5] Annie Kelly, R. Benjamin Shapiro, Jonathan de Halleux, and Thomas Ball. 2018. ARcadia: A Rapid Prototyping Platform for Real-time Tangible Interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, 582:1–582:8. doi:10.1145/3173574.3173983
- [6] Michael J. McCuffin. 2019. Augmented Reality Knowledge Work: Towards a Research Agenda. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, Paper LBW0264, 6 pages. doi:10.1145/3290607.3313074 Workshop on Immersive Analytics.
- [7] Kyzyl Monteiro, Ritik Vatsal, Thobthai N. Chulpongsatorn, Aman Parnami, and Ryu Suzuki. 2023. Teachable Reality: Prototyping Tangible Augmented Reality with Everyday Objects by Leveraging Interactive Machine Teaching. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. ACM, 330:1–330:16. doi:10.1145/3544548.3581449
- [8] Zhengze Zhu, Ziyi Liu, Tianyi Wang, Youyou Zhang, Xun Qian, Pashin F. Raja, Am Villanueva, and Karthik Ramani. 2022. MechARspace: An Authoring System Enabling Bidirectional Binding of Augmented Reality with Toys in Real-time. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology (UIST '22)*. ACM, 50:1–50:16. doi:10.1145/3526113.3545668