HARI-AR: A Heritage Agent Reasoning Interface for Autonomous Outdoor Navigation and Semantic Spatial Reasoning

Prathik M Hadagali*
HCI Lab, IIIT Sri City, India

Himangshu Sarma †
HCI Lab, IIIT Sri City, India

Mrinmoy Ghorai‡
HCI Lab, IIIT Sri City, India

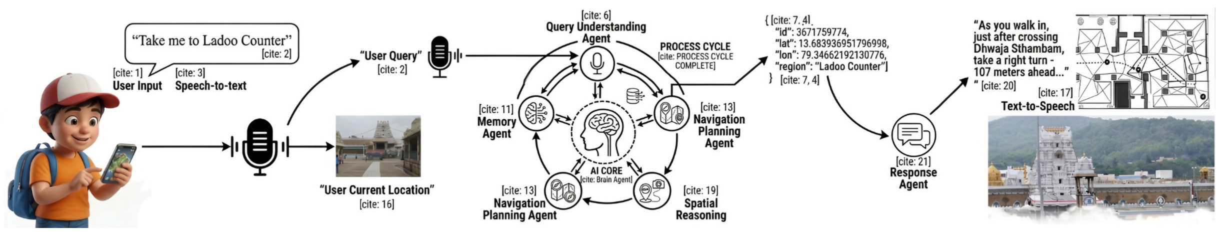

Figure 1: HARI-AR system overview. A user's spoken query and GPS location are processed through a five-agent pipeline Query Understanding, Memory, Navigation Planning, Spatial Reasoning, and Response to deliver landmark-anchored AR navigation guidance over the Tirumala temple complex.

Abstract

Outdoor heritage and pilgrimage sites present navigation challenges that existing AR systems have not addressed: GPS-grade localization without structured geometric models, culturally specific destinations expressed in informal or vernacular language. We present HARI-AR, an AR navigation system deployed at the Tirumala Venkateshwara temple complex, integrating a five-agent LLM orchestration pipeline over a custom 3,650-node OSM navigation graph enriched with 155 hand-curated semantic POIs to deliver landmark-anchored AR guidance via natural language voice queries. Rather than proposing a new AR rendering paradigm or novel agent architecture, the primary contribution is the deployment, empirical evaluation, and validation of semantically grounded landmark-anchored guidance in a real outdoor heritage environment at scale. A within-subjects study (N=24) showed landmark-referenced instructions significantly reduced directional errors compared to GPS-arrow-only guidance (21/24 vs. 14/24 correct turns; Fisher's exact p¡0.05), with a SUS score of 74.2, establishing that natural language landmark navigation is both feasible and measurably effective for first-time users in complex cultural environments.

Index Terms: AR navigation, multi-agent AI, language-guided goal retrieval, Openstreetmap

1 Introduction

AR navigation systems have advanced from static directional overlays [2] toward voice-interactive and LLM-driven interfaces [20, 21]. However, a critical deployment gap persists: virtually all prior work targets structured indoor environments where geometric models (e.g., BIM) support precise localization [22], or outdoor environments with standard street-level semantics. Large-scale outdoor heritage and pilgrimage sites visited by millions of users annually introduce a qualitatively distinct set of challenges that no prior deployed system addresses:

*e-mail: prathik.m@iiits.in, hadagali2002@gmail.com
†e-mail: himangshu.sarma@iiits.in, himangshu.tezu@gmail.com
‡e-mail: mgre04@gmail.com

- GPS localization causes quantization drift of 3-9 m, worsening in narrow temple corridors with partial sky occlusion [19]
- No structured geometric model (e.g., BIM) exists; semantic spatial data must be custom-curated for cultural navigation contexts [22]
- Pilgrims express destinations in informal, abbreviated, or vernacular language that keyword matching cannot resolve [8]

To address these challenges, HARI-AR combines semantic retrieval, OpenStreetMap-based routing, multi-agent reasoning, and landmark-aware instruction generation to enable natural-language outdoor AR navigation in a deployed heritage environment.

2 Related Work

2.1 AR Navigation Systems and Interface Design

AR navigation research has progressed from static marker-based overlays toward AI-driven, context-aware platforms. Foundational indoor systems [25, 7] established spatial overlay and landmark anchoring as core design principles, while subsequent outdoor approaches using ARCore and ARKit demonstrated the feasibility of mobile AR navigation despite localization limitations [2, 17].

Recent work bridges the expressiveness gap through natural language and LLM integration. Xu et al. [20, 19] improved wayfinding via egocentric AR cues and LLM-enhanced cognitive support, while Yuan et al. [24] enabled adaptive instruction playback through gaze and location tracking. Most relevant to HARI-AR, Yang et al. [22] integrated BIM with RAG for language-guided indoor navigation, demonstrating that semantically grounded natural language outperforms fixed-command interfaces. Hiraki et al. [26] showed that tool-augmented LLM agents can generate complete AR application logic from natural language, and Yanagawa et al. [21] explored on-demand navigation agents in metaverse environments. Despite these advances, no prior system applies multi-agent LLM orchestration to real-world, GPS-grade outdoor navigation at heritage-environment scale the gap HARI-AR addresses.

2.2 OpenStreetMap for Outdoor Navigation

OSM has become the dominant open geospatial substrate for pedestrian navigation, with over 10.5 million contributors [11]. Prior work has demonstrated its utility for accessibility-aware routing [6], indoor-outdoor integration [18], preference-based route generation [13], and temporally restricted areas [14]. However, applying OSM at heritage scale introduces persistent challenges: Graser [10] identified failures in modeling open-plaza traversal, Seetharam et al. [1] showed the need for federated visual positioning beyond GPS, and graph-theoretic analyses [12, 11] consistently reveal disconnected components and incomplete pedestrian annotations. HARI-AR addresses these through LCC filtering, yielding a unified 3,650-node walkable subgraph for reliable A* routing.

2.3 LLMs and Multi-Agent AI Architectures

The ReAct framework [23] established the reasoning-action paradigm enabling LLMs to iteratively query external tools the mechanism underpinning HARI-AR's LangGraph orchestration. Park et al. [16] demonstrated that memory-equipped LLM agents sustain coherent behavior across extended interactions, motivating HARI-AR's dedicated Memory Agent. Multi-agent surveys [5, 3] consistently show role-specialized architectures outperform monolithic systems for tasks requiring diverse expertise. Critically, Chen et al. [4] and a broader spatial survey [8] identified systematic topological errors and fictitious path generation in LLMs reasoning without grounded spatial references limitations HARI-AR mitigates by anchoring all spatial reasoning to a verified OSM graph and POI database via RAG [9]. Oh et al. [15] further validated that spatially aware LLM agents effectively interpret physical environments when supplied with appropriate context, supporting HARI-AR's design.

3 System Design

HARI-AR is designed to address three primary research questions:

- **RQ1:** How accurately does the pipeline retrieve navigation destinations from informal, vernacular, and multi-target queries?
- **RQ2:** Do landmark-anchored instructions reduce wayfinding errors compared to GPS-arrow-only guidance?
- **RQ3:** How usable is the full system for pilgrims with no prior AR experience?

Given a spoken query Q_t and GPS coordinates $\phi_t = (\text{lat}_t, \text{lon}_t)$, HARI-AR produces a response tuple (P^*, I^*, R^*), consisting of a computed AR waypoint path, a sequence of landmark-anchored navigation instructions, and a natural language text-to-speech response. Five specialized agents operate in a sequential pipeline, where each agent consumes structured output from its predecessor.

3.1 Navigation Graph

Raw OpenStreetMap (OSM) data for the Tirumala complex is preprocessed using OSMnx. The graph is exported as a directed MultiDiGraph and pruned using Largest Connected Component (LCC) filtering, yielding a walkable subgraph containing 3,650 nodes and 3,912 edges. Each graph edge stores: Haversine distance, Compass heading, Road type metadata. A semantic Point-of-Interest (POI) index containing 155 named locations is generated using dense vector embeddings produced by the `all-MiniLM-L6-v2` sentence transformer model.

3.2 A* Path Planning

Route planning is performed using the A* shortest-path algorithm. The traversal cost of an edge e_{ij} is defined as:

$$c(e_{ij}) = dist_{ij} \tag{1}$$

where $dist_{ij}$ denotes the Haversine distance between connected nodes. Source and destination nodes are resolved using OSMnx's KD-tree nearest-node lookup. In cases where A* fails due to graph disconnection, Dijkstra's algorithm is used as a fallback mechanism to ensure a valid route is returned within the connected subgraph. The resulting node sequence is post-processed by removing waypoints located within 3,m of their predecessor. This simplification reduces the average path length from approximately 312 raw graph nodes to 47 AR navigation waypoints.

3.3 Query Understanding Agent

The Query Understanding Agent is implemented using the `llama-3.1-8b-instant` model hosted through Groq. Its primary responsibilities include: Extracting destination phrases from user utterances, Resolving vernacular and informal references, Handling ordered multi-target navigation requests.

For example:

> "Take me to Ladoo Counter and then Anna Prasadam."

POI retrieval is performed using cosine similarity between the query embedding and stored POI embeddings:

$$k^* = \arg\max_k \frac{\mathbf{q} \cdot \mathbf{v}_k}{|\mathbf{q}|, |\mathbf{v}_k|} \tag{2}$$

where $\mathbf{q}$ denotes the query embedding and $\mathbf{v}_k$ denotes the embedding of POI k. A confidence threshold of 0.20 is employed to trigger fallback handling when semantic matches are weak.

3.4 Memory Agent

The Memory Agent maintains both short-term and long-term contextual memory. **Short-term memory:** In-process event history maintained during the active session. **Long-term memory:** Persistent JSON-based storage indexed by user identifier. This memory layer enables the resolution of referential navigation commands such as:

> "Take me back."

The agent executes asynchronously to prevent blocking the navigation response pipeline.

3.5 Spatial Reasoning Agent

The Spatial Reasoning Agent transforms the A* route into a heading-aware AR navigation path:

$$P^* = \{(\text{lat}_i, \text{lon}_i, \theta_i)\} \tag{3}$$

where θ_i represents the heading at waypoint i. For every significant turn point (bearing change greater than 20°), the agent queries nearby POIs within an 80,m radius and generates landmark-anchored instructions.

For example:

> "As you walk in, just after crossing Dhwaja Sthambam, take a right turn-107 metres ahead."

The instruction text is refined using an LLM to produce pilgrim-friendly language. If the LLM service is unavailable, a geometric rule-based fallback generates the navigation instructions.

3.6 Response Agent

The Response Agent assembles the final navigation output and converts it into a JSON structure compatible with the Unity AR client. The response contains: AR waypoint path, Navigation instruction sequence, Audio response URL. Communication between client and server is performed through a RESTful API interface.

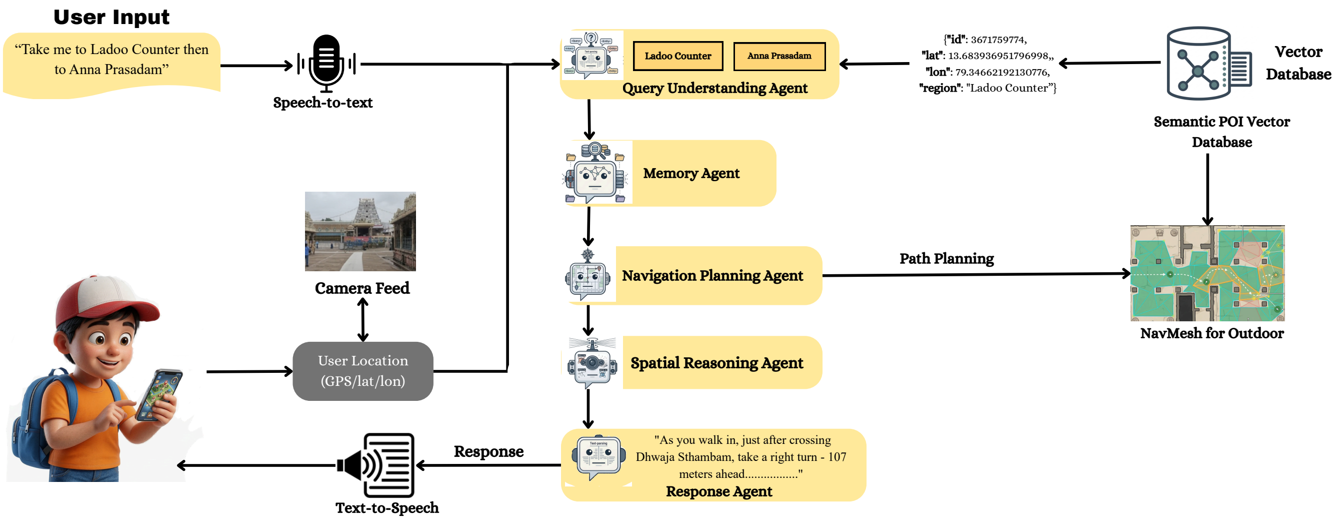

Figure 2: An overview of the HARI-AR system integrating a five-agent pipeline with an OSM-derived navigation graph and Semantic POI Vector Database for natural language AR navigation in heritage environments. Five specialized agents process the user's spoken query and GPS coordinates to produce AR-anchored waypoint guidance: **(a)** the Query Understanding Agent extracts destinations via semantic vector search; **(b)** the Memory Agent maintains session-aware navigation history; **(c)** the Navigation Planning Agent computes optimal routes via A* search over the OSM NavMesh; **(d)** the Spatial Reasoning Agent generates heading-annotated AR waypoints and landmark-anchored instructions; and **(e)** the Response Agent delivers structured guidance to the Unity AR client via RESTful API.

3.7 Client-Server Architecture

HARI-AR adopts a client-server architecture consisting of a Unity-based Android application and a Python backend. The **Unity 2022.3 Android client** is responsible for: GPS acquisition, Voice input through Android Speech-to-Text (STT), AR waypoint visualization using the ARCore Geospatial API. The **Python FastAPI backend** executes the complete multi-agent navigation pipeline. The sentence-transformer embedding model and OSMnx navigation graph are loaded once during system startup and shared across requests, reducing latency and improving runtime efficiency.

4 Evaluation

4.1 Study Design

Within-subjects study at Tirumala complex ($N = 24$; age 18–65; 13F, 11M; 9 university, 15 secondary educated). No prior AR experience. Four navigation tasks under two counterbalanced conditions: Baseline: GPS-arrow-only navigation HARI-AR: Full system with landmark-anchored instructions. Tasks covered: service facility, religious structure, transport stop, and multi-target sequence all starting from GNC Tollgate.

4.2 Goal Retrieval (RQ1)

Semantic embedding retrieval achieved cosine similarity > 0.80 across direct destination queries, successfully resolving vernacular Telugu inputs, abbreviated references, and multi-target requests. Failure cases were limited to STT phonetic errors on heritage-specific nouns and spatial qualifier queries (e.g., *“shrine near the main gate”*).

4.3 Landmark Instructions (RQ2)

HARI-AR: 21/24 correct orientations at Garuda Statue junction Baseline: 14/24 correct orientations Fisher's exact test: $p < 0.05$ 88% of route segments matched with nearby landmarks; 12% defaulted to cardinal-direction guidance Largest gains at junctions with visually similar pathways, where abstract compass directions proved insufficient.

4.4 Usability (RQ3)

Metric	Result
SUS Score	74.2 (SD=8.6) - *Good*
Task Completion	24/24 (100%)
Wayfinding Errors	1.8/session (baseline) - 0.6 (HARI-AR)
Perceived Workload	3.1 (baseline) - 2.3 (HARI-AR)
Landmark Utility	83.3% rated *useful* or *very useful*
Trust Score	3.8/5 (SD=0.7)

Primary usability issue: press-and-hold voice input was uncomfortable for older participants during outdoor walking.

5 Discussion

The results support our central claim: the decisive factor in outdoor heritage wayfinding is not AR rendering fidelity but the semantic density of the underlying spatial index. The orientation result at the Garuda Statue junction, together with the drop in wayfinding errors, isolates a single manipulation naming a visible landmark instead of a compass bearing as the source of the gain. That the largest improvements occurred at junctions with visually similar pathways indicates the mechanism: landmarks disambiguate routes where geometric cues alone become ambiguous, precisely the condition GPS drift makes worst. Decomposition made failures diagnosable rather than diffuse, localizing each observed failure mode to a specific stage STT phonetics, embedding-only retrieval, POI coverage and pointing to bounded fixes rather than prompt-level retuning. We do not claim that specialization outperforms a monolithic agent, as our study did not test that comparison; what the deployment supports is the central design choice of HARI-AR, that combining semantic retrieval, structured navigation, and landmark-aware spatial reasoning produces guidance grounded in the physical environment while remaining understandable to first-time users through a single voice interaction. The practical cost is curation. The 155 semantic POIs required manual annotation using local domain knowledge, bounding how quickly the system transfers to a new site. Partial automation via geotagged imagery and crowd-sourced place names could reduce this, making semi-automatic POI extraction the most valuable next step for replication at other heritage sites.

6 LIMITATIONS & FUTURE WORK

GPS drift (3–9 m) causes nearest-node misresolution in open plazas integrate VPS or inertial dead reckoning Spatial qualifier queries (e.g., *"nearest shrine"*) unsupported by embedding-only retrieval two-stage semantic + Haversine re-ranking OSM gaps in plaza traversal coverage synthetic crossing edges and plaza-aware graph augmentation. STT errors on heritage terminology fine-tune bilingual models on pilgrimage vocabulary.

7 CONCLUSION

This work presented HARI-AR, a multi-agent AR navigation framework designed for large-scale outdoor heritage environments. By combining large language model orchestration, semantic retrieval, OpenStreetMap-based routing, and landmark-aware AR guidance, the system enables natural-language navigation to culturally significant destinations. Unlike prior AR navigation systems that primarily focus on indoor environments, HARI-AR addresses challenges unique to outdoor heritage settings, including informal voice queries, culturally specific landmarks, and GPS-based localization constraints. The within-subjects evaluation demonstrated statistically significant improvements in wayfinding performance when landmark-anchored instructions were provided, validating the effectiveness of contextual navigation guidance. Furthermore, usability results indicate that the system is accessible to first-time users with no prior AR experience. The findings suggest that integrating multi-agent AI reasoning with semantically enriched spatial knowledge can substantially improve navigation experiences in complex cultural environments. Future work will focus on proximity-aware semantic retrieval, robust localization, multilingual interaction, and deployment on next-generation AR wearable devices.

ACKNOWLEDGMENTS

This work was supported by the project titled "Building and Development an AR Platform for Tirumala Darshan" under the grant IITTNiF/TPD/2024-25/P15. The authors sincerely thank the funding agencies and associated institutions for their valuable support and resources that contributed to the successful execution of this research.

REFERENCES

[1] S. Bharadwaj, H. Williams, L. Wang, M. Liang, T. Jin, S. Seshan, and A. Rowe. Openflame: Federated visual positioning system to enable large-scale augmented reality applications. In *2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 706–716. IEEE, 2025.

[2] L. Bibbò, F. La Foresta, S. Calcagno, E. Genovese, and V. Barrile. Indoor navigation: Augmented reality as case studio for cognitive inclusion. *WSEAS Transactions on Information Science and Applications*, 22:28–44, 2025.

[3] R. Buyya et al. Agentic artificial intelligence (ai): Architectures, taxonomies, and evaluation of large language model agents. *arXiv preprint arXiv:2601.12560*, 2026.

[4] B. Chen, Z. Xu, S. Kirmani, B. Ichter, D. Sadigh, L. Guibas, and F. Xia. Spatialvlm: Endowing vision-language models with spatial reasoning capabilities. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14455–14465, 2024.

[5] S. Chen, Y. Liu, W. Han, W. Zhang, and T. Liu. A survey on llm-based multi-agent system: Recent advances and new frontiers in application. *arXiv preprint arXiv:2412.17481*, 2024.

[6] A. Cohen, A. Natapov, and S. Dalyot. Leveraging openstreetmap to investigate urban accessibility and safety of visually impaired pedestrians. *Editors*, p. 66, 2022.

[7] E. D. Fajrianti, N. Funabiki, S. Sukaridhoto, Y. Y. F. Panduman, K. Dezheng, F. Shihao, and A. A. Surya Pradhana. Insus: Indoor navigation system using unity and smartphone for user ambulation assistance. *Information*, 14(7):359, 2023.

[8] J. Feng, J. Zeng, Q. Long, H. Chen, J. Zhao, Y. Xi, Z. Zhou, Y. Yuan, S. Wang, Q. Zeng, et al. A survey of large language model-powered spatial intelligence across scales: Advances in embodied agents, smart cities, and earth science. *arXiv preprint arXiv:2504.09848*, 2025.

[9] Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, Y. Bi, Y. Dai, J. Sun, M. Wang, and H. Wang. Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*, 2023.

[10] A. Graser. Integrating open spaces into openstreetmap routing graphs for realistic crossing behaviour in pedestrian navigation. *GI_Forum*, 4(1):217–30, 2016.

[11] R. Hosseini, D. Tong, S. Lim, G. Sohn, and G. Gidófalvi. A framework for performance analysis of openstreetmap data in navigation applications: the case of a well-developed road network in australia. *Annals of GIS*, 31(2):233–250, 2025.

[12] T. Hu, T. Du, Z. Qu, and M. Gorlatova. Xr reality check: What commercial devices deliver for spatial tracking. In *2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 932–942. IEEE, 2025.

[13] T. Novack, Z. Wang, and A. Zipf. A system for generating customized pleasant pedestrian routes based on openstreetmap data. *Sensors*, 18(11):3794, 2018.

[14] H. Ochoa-Ortiz, G. Gartner, and A. Graser. Pedestrian routing of periodically changing areas using volunteered geographical information (openstreetmap). *Abstracts of the ICA*, 5:92, 2022.

[15] S. Oh, N. An, Y. Cho, M. Jung, and K. K. Kim. When llms recognize your space: Research on experiences with spatially aware llm agents. *IEEE Transactions on Visualization and Computer Graphics*, 2025.

[16] J. S. Park, J. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*, pp. 1–22, 2023.

[17] S. Singh, J. Singh, B. Shah, S. S. Sehra, and F. Ali. Augmented reality and gps-based resource efficient navigation system for outdoor environments: Integrating device camera, sensors, and storage. *Sustainability*, 14(19):12720, 2022.

[18] Z. Wang and L. Niu. A data model for using openstreetmap to integrate indoor and outdoor route planning. *Sensors*, 18(7):2100, 2018.

[19] F. Xu, T. Zhou, T. Nguyen, H. Bao, C. Lin, and J. Du. Integrating augmented reality and llm for enhanced cognitive support in critical audio communications. *International Journal of Human-Computer Studies*, 194:103402, 2025.

[20] F. Xu, T. Zhou, H. You, and J. Du. Improving indoor wayfinding with ar-enabled egocentric cues: A comparative study. *Advanced Engineering Informatics*, 59:102265, 2024.

[21] H. Yanagawa, Y. Hiroi, S. Tokida, Y. Hatada, and T. Hiraki. Navigation pixie: Implementation and empirical study toward on-demand navigation agents in commercial metaverse. In *2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 1137–1147. IEEE, 2025.

[22] H.-K. Yang, T.-C. Hsiao, R. Oka, R. Nishino, S. Tofukuji, and N. Kobori. An embodied ar navigation agent: Integrating bim with retrieval-augmented generation for language guidance. In *2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 1461–1471. IEEE, 2025.

[23] S. Yao, J. Zhao, D. Yu, N. Du, I. Shafran, K. Narasimhan, and Y. Cao. React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*, 2022.

[24] B. Yuan, H. Cho, T. Teo, G. A. Lee, and M. Billinghurst. Focus-aware task guidance: Adaptive ar instruction playback via gaze and location tracking. In *2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 954–964. IEEE, 2025.

[25] J. Zhang, X. Xia, R. Liu, and N. Li. Enhancing human indoor cognitive map development and wayfinding performance with immersive augmented reality-based navigation systems. *Advanced Engineering Informatics*, 50:101432, 2021.

[26] C. Zhu, S.-K. Hsia, X. Hu, Z. Liu, J. Shi, and K. Ramani. agentar: Creating augmented reality applications with tool-augmented llm-based autonomous agents. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–23, 2025.