Demonstrating *TOM*: A Development Platform For Wearable Intelligent Assistants

Nuwan Janaka
nuwanj@u.nus.edu
Smart Systems Institute, National University of Singapore
Synteraction Lab
Singapore

Shengdong Zhao*
shengdong.zhao@cityu.edu.hk
Synteraction Lab
School of Creative Media & Department of Computer Science, City University of Hong Kong
Hong Kong, China

David Hsu*
dyhsu@comp.nus.edu.sg
School of Computing, National University of Singapore
Smart Systems Institute, National University of Singapore
Singapore

Sherisse Tan Jing Wen
sherisse_tjw@u.nus.edu
School of Computing, National University of Singapore
Singapore

Chun Keat Koh
idmkck@nus.edu.sg
Smart Systems Institute, National University of Singapore
Singapore

ABSTRACT

Advanced wearable digital assistants can significantly enhance task performance, reduce user burden, and provide personalized guidance to improve users' abilities. However, developing these assistants presents several challenges. To address this, we introduce *TOM* (*The Other Me*), a conceptual architecture and open-source software platform (https://github.com/TOM-Platform) that supports the development of wearable intelligent assistants that are contextually aware of both the user and the environment. Collaboratively developed with researchers and developers, *TOM* meets their diverse requirements. *TOM* facilitates the creation of intelligent assistive AR applications for daily activities and supports the recording and analysis of user interactions, integration of new devices, and the provision of assistance for various activities.

CCS CONCEPTS

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

KEYWORDS

context-aware system, wearable, AI assistance, smart glasses, HMD

ACM Reference Format:
Nuwan Janaka, Shengdong Zhao, David Hsu, Sherisse Tan Jing Wen, and Chun Keat Koh. 2024. Demonstrating *TOM*: A Development Platform For Wearable Intelligent Assistants. In *Companion of the 2024 ACM International Joint Conference on Pervasive and Ubiquitous Computing Pervasive and Ubiquitous Computing (UbiComp Companion '24), October 5–9, 2024, Melbourne, VIC, Australia.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3675094.3677551

*Corresponding Authors.

UbiComp Companion '24, October 5–9, 2024, Melbourne, VIC, Australia

ACM ISBN 979-8-4007-1058-2/24/10
https://doi.org/10.1145/3675094.3677551

1 INTRODUCTION AND RELATED WORK

With recent advancements in Machine Learning (ML) and Artificial Intelligence (AI) technologies, intelligent digital assistants are becoming an integral part of daily life. These include traditional voice assistants like Siri or Google, and emerging wearable assistants like Humane Ai Pin [1] and Rabbit R1 [27]. Intelligent digital assistants can practically aid users in performing both familiar and new tasks, reduce task load and errors, and enhance task performance [12]. Moreover, these assistants can offer personalization, optimize support for individual needs, and broaden accessibility.

However, developing wearable intelligent assistants presents challenges for stakeholders such as users, developers, and researchers. Despite existing interaction paradigms such as Heads-Up Computing [37] aiming to realize such assistance in daily activities with a focus on users, there is a lack of understanding of the required system capabilities and development guidance. While Augmented and Mixed Reality (AR/MR) assistive systems that enhance user performance have been developed [6, 7, 32], most are tailored to specific tasks (e.g., ARGUS [12] for immersive analytics, Project Aria [16] for data collection) and lack adaptability for various daily activities. Although the Platform for Situated Intelligence (\psi) [5, 8] enables accelerated research and development in traditional interactive systems, it lacks support for wearable, user-centered applications [37] that facilitate task assistance while minimizing interference by understanding user and context. Emerging wearable intelligent assistants such as Rabbit R1 [27], which support specific activities (e.g., booking taxis, querying objects), do not provide easy development or research support (e.g., analyzing/visualizing data) and have limited user interactions and understanding.

To tackle these challenges, we introduce *TOM* (*The Other Me*), a software platform developed by identifying the needs of users, researchers, and developers. *TOM* facilitates the creation and analysis of wearable assistive applications, integrates new devices, enables understanding of context and users, and supports multimodal interactions with AR/MR devices and ML/AI technologies. Through developing several proof-of-concept services (e.g., running coach assistance, translation, and querying assistance), we showcase the utility of *TOM* in supporting different daily activities and highlight the necessary future improvements. For additional evaluation, please refer to our original paper, *TOM* [22].

2 *TOM*: THE OTHER ME

2.1 Envisioned Usage Scenarios

Consider Jane, who regularly uses *Jerry*, **a digital assistant developed using *TOM***, in her daily life. *Jerry* sees what Jane sees, hears what Jane hears, knows her preferences, and understands her emotional and physical conditions.

Scenario. Unable to decide on a dish to prepare for herself and her toddler and wishing to try something new, she opens the refrigerator and asks, "Hey, *Jerry*. Can you suggest a new dish for us?" *Jerry* scans and identifies the ingredients in the refrigerator, finds possible dishes, and renders three new dish suggestions with their images. Jane finds the second suggestion appealing and inquires about the preparation process. *Jerry* then guides her through preparing the new dish, providing real-time, step-by-step feedback superimposed in real-world objects.

Later, Jane receives a delivery of a play table set for her toddler, ordered through *Jerry* during an online browsing session. She notices her toddler's eagerness to assist in assembling the set. Examining the package, she asks, "Hey *Jerry*, can you help me build this?" By identifying the play table set and retrieving instructions, *Jerry* displays step-by-step virtual instructions superimposed on the physical parts, which Jane follows while involving her toddler. Suddenly, her toddler accidentally drops a piece of the set, striking his leg and causing him to cry. Jane becomes panicked. Sensing the situation, *Jerry* instructs her to remain calm and inspect her toddler's leg. As Jane consoles her child, *Jerry* assesses the situation and provides first-aid instructions. During the first aid, *Jerry* asks whether to contact her husband, family doctor, or hospital for further treatment. Upon request, *Jerry* connects with the family doctor via video call to further observe the toddler's leg.

2.2 System Capabilities

In our quest for an envisioned intelligent wearable assistant, *Jerry*, we observed that while certain capabilities are supported by existing context-aware and assistive AR/MR systems, a complete integration of these capabilities into a single system is lacking. The Heads-Up Computing Paradigm [37], while theoretically supporting our envisioned use cases, does not provide guidance on implementing such a system or the capabilities required to further research optimal Human-AI interactions during daily activities. Drawing from literature, our experience working with AR/MR and AI researchers, and testing assistive Human-AI interfaces (including early prototypes of *TOM*) with participants and their feedback, we have formulated the following system capabilities. These are categorized based on three major stakeholders' requirements, which, though distinct, have overlapping capabilities.

Just-in-time Assistance for Users. Users should be able to **interact** with the system (i.e., provide input and receive feedback) naturally and optimally to obtain the desired assistance [18, 37]. Such assistance should be delivered just in time to match the user's current needs or proactively when users have limited knowledge of system capabilities [3, 29, 34], with minimal interference in the user's ongoing activities while accommodating the user's cognitive capabilities [4, 23]. To achieve this, the system should **understand the user and context** to provide the most appropriate feedback to support the user's ongoing activities [14, 37]. Such understanding aids in modeling the human and the world to minimize awareness mismatch between user expectations and system feedback and maintaining profiles [30].

Data Recording and Analysis for Researchers. To understand user interactions with such a system and to design optimal interactions, researchers need to **record, visualize, and analyze** the data and develop models [12, 16, 24]. This involves collecting data to support real-time and retrospective observations, training models to predict optimal feedback and analyzing their performance, and understanding the underlying reasons for user and system behaviors [12, 16, 26].

Ease of Development for Developers. Considering the variety of activities users may engage in and their unique assistance requirements, the system should enable developers to create different assistive features easily. This requires that developers can **integrate new devices** easily (e.g., sensors to understand new contexts or actuators to provide optimal feedback), **deploy new assistance and models** (e.g., to predict optimal feedback), and **access and control current data** (e.g., from existing devices or models).

2.3 Conceptual Architecture

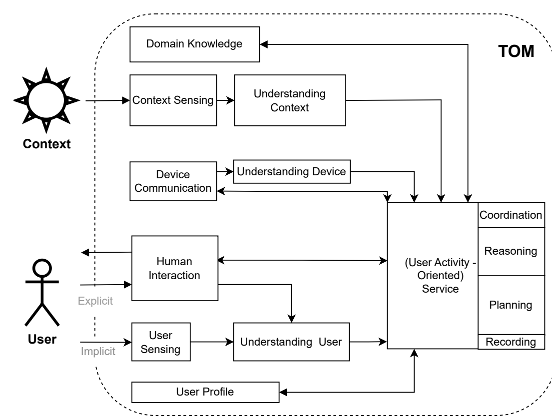

Figure 1: High-level conceptual architecture of *TOM*. Arrow directions represent the communication (e.g., data/interaction) flow.

To support the above requirements, we consider three main entities: *user* (i.e., the individual receiving assistance), *context* (i.e., the user's perceptual space and associated tasks), and the *system*, *TOM*, as illustrated in Figure 1, following the high-level context sources [17].

Separating the *user* from the *context* enables us to develop user interaction models [37]. These models sense and understand the user (e.g., cognitive states, affective states, physical states [17]) to provide personalized feedback. Thus, *TOM* maintains user profiles to cater to individual preferences and capabilities.

Given that daily activities, such as cooking, typically involve both digital (e.g., viewing a recipe) and physical tasks (e.g., selecting the proper portion), *TOM* offers system-level support to connect the digital world with the physical world by understanding the context (e.g., physical environment) and utilizing pervasive augmented reality [17]. This involves a multi-modal and multifaceted understanding of the environment (e.g., understanding the ongoing activities, associated objects, and relationships) as well as understanding the devices that facilitate interactions (e.g., device resource availability).

In terms of input, *TOM* supports the user's explicit multi-modal inputs (such as voice and gesture) as well as implicit inputs (like gaze and physiological data), in addition to processing multi-modal context information.

After understanding the context (e.g., ongoing activity) and user (e.g., intention), *TOM* activates a context-aware service, employing domain knowledge to generate real-time proactive suggestions through reasoning and planning. These suggestions are conveyed to users as multi-sensory feedback, tailored to their cognitive capacity, including visual, auditory, and/or haptic modalities. The feedback is dynamically updated based on the user's actions; for instance, if the user does not follow a given suggestion, *TOM* formulates the next appropriate suggestion, considering the user's current status and context, facilitating a closed-loop control system.

System Architecture: Implementation. Refer to TOM [22] and https://github.com/TOM-Platform for client-server implementation of the system architecture, which satisfies the above requirements and (partial) envisioned use cases. To overcome potential latency issues between the client and server, time-critical processing can also be implemented on-device.

3 DEMONSTRATION DURING DAILY ACTIVITIES

We have implemented several proof-of-concept services to support daily activities, realizing our vision of an intelligent, wearable, proactive assistant.

3.1 During Exercise: Running Assistance

Scenario. Jack uses *Jerry* (implemented using *TOM*) to assist with his running exercises. He wears an OHMD and a smartwatch (connected to a Server operating on a laptop[1]). He initiates his running (speed or distance) training using voice interactions. *Jerry* provides route options, and he selects one using either voice commands or mid-air gestures. During his run, *Jerry* provides personalized running coach instructions (e.g., speeding up or slowing down based on his current speed, training plan, and user profile) and proactive suggestions (e.g., encouraging feedback based on duration, alerting about potential dangers like traffic lights based on environment sensing, giving direction cues based on location, indicating waterpoints based on location) using either visual or auditory modality when required (i.e., by default, *Jerry* will provide only essential details, such as the time, to reduce the display clutter and information overload). At the end of his run, he receives a summary of the exercise.

System. As shown in Figure 2, the current running assistance is implemented as a Running Service. This service processes sensor data from the smartwatch, user interactions, and the egocentric camera view from the OHMD, and route information from external API to determine the next running coach instruction and provide feedback. Then, the service sends the feedback to the OHMD using a pre-configured display layout (Figure 3) designed based on user testing. If the user does not specify certain details required for running (e.g., expected speed), the system uses the user's profile to determine them.

[1]In the future, we plan to run the Server in the cloud

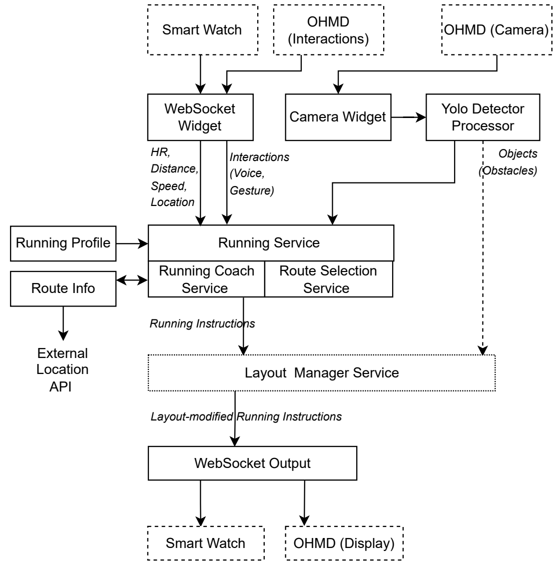

Figure 2: The system components of the running assistance implemented using *TOM.* Dashed-line boxes indicate implemented Client components, solid lines represent implemented Server components, and dotted lines denote Server components under development.

Limitations. During preliminary user testing, we identified several device and technical limitations. These include impaired visibility of the OHMD's visual feedback in outdoor environments, misrecognition of voice commands due to background noise and user fatigue during running, the OHMD's weight affecting the exercise experience, and instability of visual feedback from frequent head movements (content jumps) [19, 21]. Additionally, participants requested adaptive user interfaces tailored to their preferences and environment [25] for enhanced visibility.

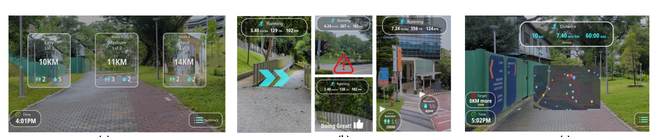

Figure 3: The running assistance UI supports voice and mid-air gesture input interactions. (a) The user starts the running assistance and is prompted to select a route. (b) *Jerry* provides personalized training guidance, proactive feedback on potential dangers or encouragement, and details about water points while running. (c) In the end, *Jerry* presents the user with a run summary.

3.2 During Dining and Shopping: Translation and Querying Assistance

Scenario. Jack visits a new restaurant and discovers that the menu is only in Mandarin, which he cannot understand (See Figure 4 for details). He verbally requests *Jerry*'s assistance, and it displays the translated menu in English, superimposed on the original menu. When he shows prolonged gaze duration on a particular menu item, *Jerry* automatically displays an image with a brief description. Upon inquiring whether it is vegetarian, *Jerry* informs him that it contains some non-vegetarian ingredients. Consequently, Jack orders the recommended dish.

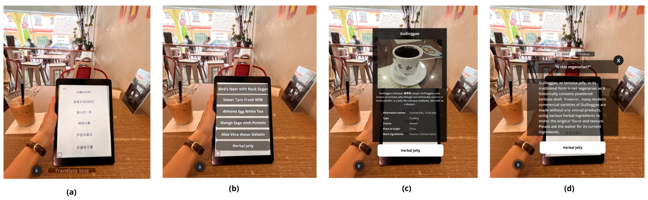

Figure 4: The translation assistance UI supports voice, mid-air gesture, and gaze input interactions. (a) The user is prompted for the action they wish to take and chooses the 'Translate Text' option. (b) *Jerry* translates the Mandarin text on the screen into English and overlays the translated information onto the location of the Mandarin text. (c) The user shows interest in 'Herbal Jelly' and seeks more general information. (d) The user verbally inquires, "Ok *Jerry*, is this vegetarian?"

After dining, Jack goes to a supermarket. He notices some medical supplements that are different from his usual purchases. Jack asks *Jerry* about them while looking at (or pointing to) them, and *Jerry* provides detailed information about them and their usage, assisting him in making an informed decision before purchasing.

System. The current translation assistance system uses an OHMD camera to scan and recognize texts, translates them, and employs a Large Language Model (LLM) to adapt them to the local context. Subsequently, *TOM* displays the translations aligned with the original menu. Based on the user's explicit (e.g., voice) or implicit (e.g., gaze) inputs, *TOM* presents the menu images and detailed descriptions and verbally answers questions. Similarly, the querying assistance system scans and recognizes objects and texts using the OHMD camera, based on the user's explicit interactions (e.g., gaze plus voice, gesture plus voice). It then interacts with an LLM to provide responses to the user's verbal queries.

Limitations. Similar to the running assistance, additional limitations were noted in the translation and querying assistance. These include the inability to recognize small text in dim lighting, occasional retrieval of incorrect images for specific dishes, and delays (2-8 seconds) in displaying feedback, which is attributable to the response times of external APIs (e.g., ChatGPT, Bing Image Search).

4 LIMITATIONS AND FUTURE IMPROVEMENTS

In addition to the identified technical limitations from specific demonstrations, interaction design challenges surfaced in daily activities. Situational impairments, especially in dynamic environments, restrict certain user interactions (e.g., diminished voice command accuracy in outdoor wind), underscoring the need for methods to support seamless input modality transitions [21]. Inaccuracies in AI-generated suggestions also contribute to user mistrust, necessitating more transparent AI explanations [11, 15, 33].

Moreover, *TOM*'s current implementation exhibits limitations. A notable area is the automatic switching of services based on user inputs to optimize service execution for ongoing activities, when multiple services match expected assistance, requiring further research in this area. Despite *TOM*'s support for the Large Language Model (LLM) in facilitating human-like conversations and tasks [9, 10, 28, 35], integrating Large Action Models (LAM) [27] could enhance interaction efficiency with external applications and improve user action understanding. Effective live monitoring is available, but *TOM* needs better visualizations for comprehensive retrospective analysis of long-term user behaviors. Aggregated visualization techniques, similar to ARGUS [12], and retrospective analysis such as PilotAR [20] could aid in this. Additionally, advancing the system's grasp of users' cognitive states and their activity correlations requires sophisticated modeling and simulation [24, 26]. Similarly, facilitating effective multi-agent collaboration while preserving user autonomy in multi-user *TOM* scenarios remains a significant research challenge.

Finally, developing such systems implicates privacy, security, safety, and ethical challenges. Despite Institutional Review Board (IRB) approval for user studies, real-world deployment raises critical privacy, safety, and social acceptability concerns, considering both users and bystanders [2, 13, 31]. Issues include monitoring and recording users' physical and cognitive states, capturing bystanders' behaviors without consent, securely handling sensitive data, and anonymizing data for aggregate analysis. Although on-device/edge computing provides partial solutions [36], the limitations of current devices necessitate further advancements.

5 CONCLUSION

We have presented the anticipated capabilities of developing an intelligent wearable assistive system and introduced *TOM*, an architecture and open-source implementation (https://github.com/TOM-Platform) that enables researchers and developers to create and analyze assistive applications for supporting daily activities. We welcome contributions from the community to expand its supported devices and usage scenarios. We envision that *TOM* will serve as a software platform for researchers and developers to develop innovative, intelligent assistance in various tasks, facilitating human-computer, human-AI, and human-robot interactions. Our future plans include extending *TOM*'s capabilities to enable remote robot interactions, where humans can share information (e.g., intentions) with a remote robot to execute tasks.

ACKNOWLEDGMENTS

We would like to express our gratitude to the volunteers who participated in our studies and to the interns of *TOM* project, including Teo Yun Yew Jarrett and Taufiq Bin Abdul Rahman, who developed specific system components. We also wish to thank the anonymous reviewers for their valuable time and insightful comments, which helped to improve this paper.

This research is supported by the National Research Foundation, Singapore, under its AI Singapore Programme (AISG Award No: AISG2-RP-2020-016). The CityU Start-up Grant 9610677 also provides partial support. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not reflect the views of the National Research Foundation, Singapore.

REFERENCES

[1] 2024. Ai Pin Overview. https://hu.ma.ne/aipin

[2] Fouad Alallah, Ali Neshati, Yumiko Sakamoto, Khalad Hasan, Edward Lank, Andrea Bunt, and Pourang Irani. 2018. Performer vs. observer: whose comfort level should we consider when examining the social acceptability of input modalities for head-worn display?. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology (VRST '18)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3281505.3281541

[3] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300233

[4] Christoph Anderson, Isabel Hübener, Ann-Kathrin Seipp, Sandra Ohly, Klaus David, and Veljko Pejovic. 2018. A Survey of Attention Management Systems in Ubiquitous Computing Environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (July 2018), 1–27. https://doi.org/10.1145/3214261

[5] Sean Andrist, Dan Bohus, Ashley Feniello, and Nick Saw. 2022. Developing Mixed Reality Applications with Platform for Situated Intelligence. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 48–50. https://doi.org/10.1109/VRW55335.2022.00018

[6] E. Z. Barsom, M. Graafland, and M. P. Schijven. 2016. Systematic review on the effectiveness of augmented reality applications in medical training. *Surgical Endoscopy* 30, 10 (Oct. 2016), 4174–4183. https://doi.org/10.1007/s00464-016-4800-6

[7] Mark Billinghurst, Adrian Clark, and Gun Lee. 2015. A Survey of Augmented Reality. *Foundations and Trends® in Human–Computer Interaction* 8, 2-3 (March 2015), 73–272. https://doi.org/10.1561/1100000049

[8] Dan Bohus, Sean Andrist, Ashley Feniello, Nick Saw, Mihai Jalobeanu, Patrick Sweeney, Anne Loomis Thompson, and Eric Horvitz. 2021. Platform for Situated Intelligence. https://doi.org/10.48550/arXiv.2103.15975 arXiv:2103.15975 [cs].

[9] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. https://doi.org/10.48550/arXiv.2005.14165

[10] Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio Ribeiro, and Yi Zhang. 2023. Sparks of Artificial General Intelligence: Early experiments with GPT-4. https://doi.org/10.48550/arXiv.2303.12712

[11] John M. Carroll. 2022. Why should humans trust AI? *Interactions* 29, 4 (June 2022), 73–77. https://doi.org/10.1145/3538392

[12] Sonia Castelo, Joao Rulff, Erin McGowan, Bea Steers, Guande Wu, Shaoyu Chen, Iran Roman, Roque Lopez, Ethan Brewer, Chen Zhao, Jing Qian, Kyunghyun Cho, He He, Qi Sun, Huy Vo, Juan Bello, Michael Krone, and Claudio Silva. 2023. ARGUS: Visualization of AI-Assisted Task Guidance in AR. https://doi.org/10.48550/arXiv.2308.06246

[13] Tamara Denning, Zakariya Dehlawi, and Tadayoshi Kohno. 2014. In situ with bystanders of augmented reality glasses: perspectives on recording and privacy-mediating technologies. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*. ACM Press, Toronto, Ontario, Canada, 2377–2386. https://doi.org/10.1145/2556288.2557352

[14] Anind K. Dey, Gregory D. Abowd, and Daniel Salber. 2001. A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications. *Human–Computer Interaction* 16, 2-4 (Dec. 2001), 97–166. https://doi.org/10.1207/S15327051HCI16234_02

[15] Malin Eiband, Hanna Schneider, Mark Bilandzic, Julian Fazekas-Con, Mareike Haug, and Heinrich Hussmann. 2018. Bringing Transparency Design into Practice. In *23rd International Conference on Intelligent User Interfaces (IUI '18)*. Association for Computing Machinery, New York, NY, USA, 211–223. https://doi.org/10.1145/3172944.3172961

[16] Jakob Engel, Kiran Somasundaram, Michael Goesele, Albert Sun, Alexander Gamino, Andrew Turner, Arjang Talattof, Arnie Yuan, Bilal Souti, Brighid Meredith, Cheng Peng, Chris Sweeney, Cole Wilson, Dan Barnes, Daniel DeTone, David Caruso, Derek Valleroy, Dinesh Ginjupalli, Duncan Frost, Edward Miller, Elias Mueggler, Evgeniy Oleinik, Fan Zhang, Guruprasad Somasundaram, Gustavo Solaira, Harry Lanaras, Henry Howard-Jenkins, Huixuan Tang, Hyo Jin Kim, Jaime Rivera, Ji Luo, Jing Dong, Julian Straub, Kevin Bailey, Kevin Eckenhoff, Lingni Ma, Luis Pesqueira, Mark Schwesinger, Maurizio Monge, Nan Yang, Nick Charron, Nikhil Raina, Omkar Parkhi, Peter Borschowa, Pierre Moulon, Prince Gupta, Raul Mur-Artal, Robbie Pennington, Sachin Kulkarni, Sagar Miglani, Santosh Gondi, Saransh Solanki, Sean Diener, Shangyi Cheng, Simon Green, Steve Saarinen, Suvam Patra, Tassos Mourikis, Thomas Whelan, Tripti Singh, Vasileios Balntas, Vijay Baiyya, Wilson Dreewes, Xiaqing Pan, Yang Lou, Yipu Zhao, Yusuf Mansour, Yuyang Zou, Zhaoyang Lv, Zijian Wang, Mingfei Yan, Carl Ren, Renzo De Nardi, and Richard Newcombe. 2023. Project Aria: A New Tool for Egocentric Multi-Modal AI Research. https://doi.org/10.48550/arXiv.2308.13561

[17] Jens Grubert, Tobias Langlotz, Stefanie Zollmann, and Holger Regenbrecht. 2017. Towards Pervasive Augmented Reality: Context-Awareness in Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics* 23, 6 (June 2017), 1706–1724. https://doi.org/10.1109/TVCG.2016.2543720

[18] Kasper Hornbæk and Antti Oulasvirta. 2017. What Is Interaction?. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, Denver Colorado USA, 5040–5052. https://doi.org/10.1145/3025453.3025765

[19] Yuta Itoh, Tobias Langlotz, Jonathan Sutton, and Alexander Plopski. 2021. Towards Indistinguishable Augmented Reality: A Survey on Optical See-through Head-mounted Displays. *Comput. Surveys* 54, 6 (July 2021), 120:1–120:36. https://doi.org/10.1145/3453157

[20] Nuwan Janaka, Runze Cai, Ashwin Ram, Lin Zhu, Shengdong Zhao, and Yong Kai Qi. 2024. PilotAR: Streamlining Pilot Studies with OHMDs from Concept to Insight. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (Sept. 2024). https://doi.org/10.1145/3678576

[21] Nuwan Janaka, Jie Gao, Lin Zhu, Shengdong Zhao, Lan Lyu, Peisen Xu, Maximilian Nabokow, Silang Wang, and Yanch Ong. 2023. GlassMessaging: Towards Ubiquitous Messaging Using OHMDs. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 3 (Sept. 2023), 100:1–100:32. https://doi.org/10.1145/3610931

[22] Nuwan Janaka, Shengdong Zhao, David Hsu, Sherisse Tan Jing Wen, and Chun Keat Koh. 2024. TOM: A Development Platform For Wearable Intelligent Assistants. In *Companion of the 2024 ACM International Joint Conference on Pervasive and Ubiquitous Computing Pervasive and Ubiquitous Computing*. ACM. https://doi.org/10.1145/3675094.3678382

[23] D. Scott McCrickard and C. M. Chewar. 2003. Attuning notification design to user goals and attention costs. *Commun. ACM* 46, 3 (March 2003), 67. https://doi.org/10.1145/636772.636800

[24] Roderick Murray-Smith, Antti Oulasvirta, Andrew Howes, Jörg Müller, Aleksi Ikkala, Miroslav Bachinski, Arthur Fleig, Florian Fischer, and Markus Klar. 2022. What simulation can do for HCI research. *Interactions* 29, 6 (Nov. 2022), 48–53. https://doi.org/10.1145/3564038

[25] Jason Orlosky, Kiyoshi Kiyokawa, and Haruo Takemura. 2013. Dynamic text management for see-through wearable and heads-up display systems. In *Proceedings of the 2013 international conference on Intelligent user interfaces - IUI '13*. ACM Press, Santa Monica, California, USA, 363. https://doi.org/10.1145/2449396.2449443

[26] Antti Oulasvirta, Jussi P. P. Jokinen, and Andrew Howes. 2022. Computational Rationality as a Theory of Interaction. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3491102.3517739

[27] rabbit research team. 2023. Learning human actions on computer applications. https://rabbit.tech/research

[28] Alec Radford, Jeff Wu, Rewon Child, D. Luan, Dario Amodei, and Ilya Sutskever. 2019. Language Models are Unsupervised Multitask Learners. https://api.semanticscholar.org/CorpusID:160025533

[29] B. J. Rhodes and P. Maes. 2000. Just-in-time information retrieval agents. *IBM Systems Journal* 39, 3.4 (2000), 685–704. https://doi.org/10.1147/sj.393.0685

[30] Albrecht Schmidt. 2014. *Context-Aware Computing*. https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/context-aware-computing-context-awareness-context-aware-user-interfaces-and-implicit-interaction

[31] Mel Slater, Cristina Gonzalez-Liencres, Patrick Haggard, Charlotte Vinkers, Rebecca Gregory-Clarke, Steve Jelley, Zillah Watson, Graham Breen, Raz Schwarz, William Steptoe, Dalila Szostak, Shivashankar Halan, Deborah Fox, and Jeremy Silver. 2020. The Ethics of Realism in Virtual and Augmented Reality. *Frontiers in Virtual Reality* 1 (2020). https://doi.org/10.3389/frvir.2020.00001

[32] X. Wang, S. K. Ong, and A. Y. C. Nee. 2016. A comprehensive survey of augmented reality assembly research. *Advances in Manufacturing* 4, 1 (March 2016), 1–22. https://doi.org/10.1007/s40436-015-0131-4

[33] Xuhai Xu, Anna Yu, Tanya R. Jonker, Kashyap Todi, Feiyu Lu, Xun Qian, João Marcelo Evangelista Belo, Tianyi Wang, Michelle Li, Aran Mun, Te-Yen Wu, Junxiao Shen, Ting Zhang, Narine Kokhlikyan, Fulton Wang, Paul Sorenson, Sophie Kim, and Hrvoje Benko. 2023. XAIR: A Framework of Explainable AI in Augmented Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–30. https://doi.org/10.1145/3544548.3581500

[34] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376301

[35] Zhengyuan Yang, Linjie Li, Kevin Lin, Jianfeng Wang, Chung-Ching Lin, Zicheng Liu, and Lijuan Wang. 2023. The Dawn of LMMs: Preliminary Explorations with

GPT-4V(ision). https://doi.org/10.48550/arXiv.2309.17421

[36] Jiale Zhang, Bing Chen, Yanchao Zhao, Xiang Cheng, and Feng Hu. 2018. Data Security and Privacy-Preserving in Edge Computing Paradigm: Survey and Open Issues. *IEEE Access* 6 (2018), 18209–18237. https://doi.org/10.1109/ACCESS.2018.2820162

[37] Shengdong Zhao, Felicia Tan, and Katherine Fennedy. 2023. Heads-Up Computing Moving Beyond the Device-Centered Paradigm. *Commun. ACM* 66, 9 (Aug. 2023), 56–63. https://doi.org/10.1145/3571722