On the Promises and Challenges of AI-Powered XR Glasses as Embodied Software

Ruizhen Gu
rgu10@sheffield.ac.uk
University of Sheffield
Sheffield, UK

Jingqiong Zhang
jingqiong.zhang@sheffield.ac.uk
University of Sheffield
Sheffield, UK

José Miguel Rojas
j.rojas@sheffield.ac.uk
University of Sheffield
Sheffield, UK

Donghwan Shin
d.shin@sheffield.ac.uk
University of Sheffield
Sheffield, UK

ABSTRACT

AI-powered Extended Reality (XR) glasses represent the next frontier in software interface, integrating spatial computing with foundation models (FMs) to interact with physical environments in real-time. This technology promises a rich, immersive, and interactive user experience with seamless integration in real-world scenarios while at the same time introducing unprecedented challenges at the AI-Software Engineering (SE) intersection. This vision paper aims to catalyze the development of robust spatial software by characterizing XR glasses as a distinct software paradigm through a conceptual framework and defining its advanced capabilities. We identify critical research problems, including security and privacy, validation of spatial capabilities, and explainability, while highlighting broader societal implications spanning ethics, accessibility, inclusivity, and open development ecosystems. Finally, we outline pathways for reliable, trustworthy XR systems in the FM era.

CCS CONCEPTS

• **Computing methodologies** → **Artificial intelligence**; **Mixed / augmented reality**; • **Software and its engineering**;

KEYWORDS

Extended Reality, AI-Powered XR Glasses, Embodied AI, Human-AI Collaboration, Spatial Intelligence, Security and Privacy

ACM Reference Format:
Ruizhen Gu, Jingqiong Zhang, José Miguel Rojas, and Donghwan Shin. 2018. On the Promises and Challenges of AI-Powered XR Glasses as Embodied Software. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 6 pages. https://doi.org/XXXXXXX.XXXXXXX

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Rapid advancements in AI and spatial computing technologies are accelerating the evolution of wearable Extended Reality (XR) devices—encompassing Augmented, Mixed, and Virtual Reality (AR, MR, and VR, respectively) [18]. These devices range from *XR head-mounted displays* (XR HMDs), such as the Meta Quest 3[1], which provide immersive experiences for entertainment and productivity, to display-free, AI-enabled *smart glasses*, such as the Ray-Ban Meta[2]. While smart glasses integrate features like image recognition and voice assistance, these capabilities are increasingly powered by foundation models (FMs) that can handle diverse tasks [5].

Academic and industry analysis highlight the convergence of spatial computing from XR HMDs and AI features from smart glasses into a new device category: *AI-powered everyday XR glasses* (hereafter, XR glasses) [19, 31]. These devices seamlessly superimpose digital content onto physical environments to generate realistic MR experiences. Positioned as the next evolution beyond smartphones, they are projected to bring a paradigm shift in human-digital interaction and social communication [1]. Industry prototypes like Meta Orion[3] and XREAL's Project Aura[4] exemplify the trajectory of fusing multimodal AI with XR capabilities for imminent commercial release. Crucially, their integration with FMs enables sophisticated contextual understanding and adaptive content and behavior generation that extends beyond traditional AI-enabled tasks [33].

Amid rapid commercial development, advancing XR glasses will require academic research to tackle underexplored challenges. Existing literature predominately falls within AI and human–computer interaction (HCI) domains, including topics such as multimodal context-aware fusion for user gaze prediction [28] and interaction design frameworks [34]. However, the unique SE demands of embodied spatial computing, such as software design and evolution, are largely unexplored. While some efforts have examined SE aspects for XR systems (e.g., requirements engineering, testing), they often lack insights from modern AI capabilities and fail to meet the challenges in the FM era [6, 12]. In this vision paper, we bridge the divides by synthesizing the latest academic and industrial insights to: (1) establish a conceptual overview of XR glasses as software systems; (2) map their unique capabilities; (3) identify research problems spanning AI and SE; and (4) highlight societal implications requiring broader attention. By proactively addressing these

[1] https://www.meta.com/quest/quest-3/ [2] https://www.meta.com/gb/ai-glasses/ray-ban-meta/
[3] https://www.meta.com/emerging-tech/orion/ [4] https://www.xreal.com/aura

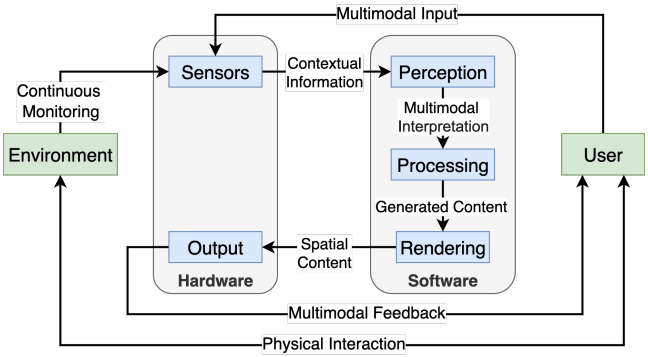

Figure 1: A conceptual overview of XR glasses systems

dimensions, we equip AI and XR researchers and practitioners with critical insights into the upcoming software paradigm evolution.

2 CONCEPTUAL FRAMEWORK

We propose a conceptual framework for XR glasses as integrated software systems that merge advanced spatial and AI capabilities. Inspired by embodied AI agents and wearable computing, we position XR glasses through key distinctions: Unlike autonomous robots or vehicles, users directly wear and interact with XR glasses, creating continuous perception-action feedback loops that emphasize human-AI collaboration. Similarly, they extend beyond conventional wearables (e.g., smartwatches) through rich multimodal rendering (e.g., visual, audio) and persistent spatial awareness.

Figure 1 presents our human-centered framework, adapting the paradigm of perception, planning, and control from autonomous systems [23]. The system begins with environmental data captured by *Sensors* (e.g., camera, mic) that flows into the context-aware *Perception* layer for spatial interpretation. The information then enters the AI-powered *Processing* layer, where contextual digital content is synthesized. Next, the *Rendering* layer organizes this content into interactive media outputs (e.g., visual, audio), which drive *Output* devices (e.g., lens display, speakers) to guide user interactions with both physical environments and digital content. Crucially, these user interactions create a closed-loop system: human response (e.g., gestures, voice commands) combined with environmental context generate new sensor inputs. This framework repositions human cognition as the primary actuator, transforming traditional embodied AI into collaborative human-AI partnerships.

3 ADVANCED CAPABILITIES

This section examines selected advanced capabilities inherent to XR glasses that distinguish them from other computing devices. Though not exhaustive, these features highlight unique opportunities while introducing novel AI and SE research challenges. They establish the foundation for subsequent discussions in the paper.

3.1 Spatial Intelligence

Spatial intelligence represents the core cognitive component in XR glasses, transforming raw contextual awareness from multimodal sensors into actionable spatial understanding of physical environments [36]. This capability enables devices to both *interpret* spatial relationships and dynamically *generate* digital content with physical attributes (e.g., collision) [41]. FM integration has significantly expanded these capabilities, facilitating the simulation of interactive 3D environments from minimal inputs. For instance, Google Deepmind's Genie 2 [22] demonstrates how a single prompt image can generate playable virtual worlds with realistic object physics, serving as training environments for embodied agents.

For XR glasses, spatial intelligence enables significant functional advances through the following aspects: (1) **Streamlined environmental understanding:** Spatial relationships are interpreted directly from camera frames, eliminating the need for traditional tracking infrastructure in spatial computing [8, 16]. This can accelerate the development process while reducing computational demands. (2) **Physics-compliant content generation:** Generated digital content gain realistic physical properties [30]. For example, a virtual ball can bounce on real surfaces and respond to user actions like hitting and throwing, creating realistic and intuitive user interactions. (3) **Affordance recognition**: Affordance, the functional potentials of objects, can be interpreted for actionable possibilities in physical environments [32]. For instance, a table surface is identified as supporting "placement" actions, providing contextual recommendations for user activities [15].

Building upon these capabilities, spatial intelligence elevates environmental *understanding* to *generation*. As depicted in the *Rendering* layer in our conceptual framework (Figure 1), this transcends traditional 3D model rendering using computer graphics techniques. Instead, FMs dynamically synthesize digital content that interacts physically with the real world. This enables seamless reality-virtuality integration while minimizing computational overhead (e.g., using minimal inputs like 2D images).

3.2 Multimodal Interfaces

XR glasses represent the next evolution in HCI through integrated multimodal interfaces. These systems incorporate natural input channels including gesture, gaze, and voice, while delivering output through visual displays and audio feedback [14]. Critically, these multimodal interactions serve dual purposes: enabling intuitive engagement with both digital content and physical environments, while simultaneously functioning as primary input channels for FM agents. This dual-role design positions multimodal interfaces as the essential mediator between human intent and spatial intelligence.

Gesture and gaze interactions extend beyond conventional voice assistants to provide XR glasses with novel, natural paradigms for engaging both physical environments and FMs. For example, a user might gesture to delineate real-world birds in their field of view with a voice query to identify the type of birds. Subsequently, the FM agent will process the spatial-visual context and display digital bird information on the glass lens while the voice assistant explains the species. This exemplifies the interface loop central to our conceptual framework (§ 2): multimodal inputs (gesture and voice) initiate FM-mediated environmental interpretation, generating corresponding digital output (visual and audio) that completes the interaction cycle. Unlike 2D interfaces, this spatial interplay blurs boundaries between environmental interactions and AI responses.

3.3 Distributed Software Architecture

XR glasses typically employ distributed architectures similar to traditional wearable devices [35], spanning three hardware layers: (1) **On-device layer:** Lightweight operating systems (OSs) handle real-time sensor fusion, rendering, and display management on the glasses; (2) **Edge layer:** Companion devices like smartphones providing low-latency compute for intensive tasks like FM inference for environmental understanding; and (3) **Cloud layer:** Offers virtually unlimited resources for data storage and processing, FM training, and hosting. On the software front, specialized OS platforms like Android XR[5] serve as the nexus, providing multimodal interfaces for user interaction, host environments for third-party apps, and deep integration with FMs (e.g., Gemini as an AI assistant).

A key architectural innovation is that Android XR provides functions that allow developers to convert traditional 2D mobile apps to spatial Android XR apps with minimal effort. This accelerates ecosystem flourishing while introducing a transformative software paradigm: *agent-mediated interaction*. In this model, we envision two types of FM agents, i.e., OS-level and domain-specific agents, coexisting and cooperating. The OS-level agent acts as an enhanced AI assistant, translating user commands into system functions (e.g., launching Google Maps via voice command). The domain-specific agents handle app-level tasks (e.g., restaurant reservations).

This cooperation enables complex workflows, transcends traditional app boundaries toward fluid agent ecosystems. This paradigm may eliminate the need for dedicated app stores; users no longer need to install apps before using them. Agent features could be packaged and sold as separate capabilities–for instance, a "Bird Watching Pro" feature for £5. Systems like Manus[6] demonstrate how autonomous agents can execute multi-step real-world tasks without continuous human guidance. For XR glasses prioritizing seamless interactions, these agents minimize the barriers between app boundaries, delivering intuitive experiences to users.

4 RESEARCH PROBLEMS

This section identifies key emerging research problems for XR glasses from both AI and SE perspectives. While the challenges fall into established categories such as security, validation, and explainability, their application to XR glasses introduces novel complexities requiring distinct solutions. Specifically, our analysis focuses on problems arising from the unique characteristics of XR glasses. These mainly include sensitive data leakage from XR glasses' always-on nature, reliability of virtual-physical integration, and explainability of FM outcomes for user decision making.

4.1 Security and Privacy

Security and privacy concerns are the most critical challenges for XR glasses, given the inherent risks in both AI and XR domains [25, 40]. While hardware vulnerabilities exist, we emphasize software risks. Security threats include severe safety risks from external attacks and internal faults. While less immersive than VR HMDs, XR glasses face analogous attacks, including overlaying malicious content, cybersickness, and manipulating user physical movements [7]. These risks may become more severe due to XR's integration with the physical world, and further exacerbated by vulnerabilities in embodied AI systems that can generate harmful behaviors. Defects include perceptual failures (e.g., misclassifying critical objects like humans) or vulnerabilities to deceptive prompting techniques [24, 42]. To ensure user safety, mitigation strategies are required across stakeholders. XR OSs should embed core safety mechanisms, such as collision detection and emergency features, allowing users to disable the display and use the devices as standard glasses. These safeguards should be enforced at the OS level to prevent override by third-party apps.

While security threats differ between AI and XR, both share critical risks akin to personalized systems (e.g., sensitive data leakage) [21, 43]. XR glasses amplify these concerns through their persistent environmental perception. The continuous data capture (e.g., visual, audio, environment) introduces inherent input privacy vulnerabilities, especially when raw data are transmitted to untrusted apps or remote servers without proper safeguards (an architectural risk illustrated in § 3.3) [44]. Moreover, this persistent environmental monitoring introduces unique *bystander privacy* risks absent in conventional systems [25]. Features like "object recognition" or "social interaction" often rely on facial and biometric data, potentially collecting sensitive information from non-consenting individuals. For example, activating an "identify friend" feature in a crowd may collect bystanders' biometric signatures, which can cause mass identification without consent or notification.

Traditional permission systems would be ill-suited for XR glasses due to their multimodal, always-on nature. Even app store policies (e.g., Google Play's data safety disclosures) may lack timely enforcement, allowing apps to launch without thorough privacy assessments [39]. While prior work has proposed mitigations such as restricting the visual processing within specified areas to avoid unintentional data capture [44], such approaches remain insufficient. In features like "identify friend", exhaustive biometric processing is unavoidable. It makes architectural safeguards, e.g., on-device processing, essential to prevent transmission of sensitive data. For instance, Apple Vision Pro adopts a private cloud computing design to address this concern [2].

4.2 Validation of Spatial Capabilities

XR glasses face significant validation challenges due to their real-world interactivity. Core to this challenge is the dynamic blending of real and virtual elements within AR/MR environments, causing environment-dependent software failures that are difficult to reproduce and debug. This contrasts fundamentally from traditional context-aware systems like mobile apps, where environmental triggers typically yield discrete and reproducible behavioral changes, such as geolocation-based recommendations [17].

The integration of reality and virtuality creates digital content coupled with persistent spatial capabilities [31], significantly increasing validation complexity. Testers should account for infinite real-world permutations that affect virtual object behavior, while ensuring contextual appropriateness in dynamic physical environments. Existing XR software testing work has largely focused on VR apps, overlooking the critical real-world interplay in AR/MR [12]. While pioneering studies addressed test oracle prediction for virtual object misplacement in AR [38], current approaches still rely on

[5] https://www.android.com/xr/ [6] https://manus.im/

unscalable manual intervention. Thus, it is essential to adapt and extend automated testing methodologies to support AR/MR.

To address these challenges, we advocate for developing specialized simulation systems that systematically replicate diverse real-world conditions. Inspired by autonomous driving platforms like CARLA, which facilitate the development and validation of complex systems through simulated scenarios [9], it would be beneficial to adapt similar infrastructure for XR glasses to enable scalable testing of spatial capabilities across dynamic real-world conditions.

4.3 Explainability

Spatial intelligence (§ 3.1) and multimodal interfaces (§ 3.2) that define XR glasses can suffer from the "black-box" nature of underlying AI models. This poses safety risks when untransparent decisions trigger malicious behaviors (§ 4.1) [3]. Given their context-aware, always-on operation and deep integration into daily life, XR glasses necessitate Explainable AI (XAI) techniques that clearly reveal decision rationales to users. As these devices are worn directly and provide critical real-time guidance based on user intent and environmental context, XAI techniques should ensure AI outputs are reliable and trustworthy, especially in safety-critical scenarios [37].

As discussed in § 4.2, validating XR glasses' functionality in dynamic environments is exceptionally challenging, and XAI requirements amplify this. Explainability should be reliable and well aligned with users' needs, despite FMs' non-determinism [20] and their limited interpretability. Personalized XAI further complicates this, as individual preferences (e.g., prefer concise explanations) introduce additional variability [37]. Thus, ensuring the accuracy and consistency of the explanation across diverse environments and users is critical for XR glasses' safety and usability. However, current XAI evaluation methods are limited to specific models [26, 27], inapplicable to scale to the vast, open-ended scenarios in XR contexts. We identify this as a core SE challenge of adapting XAI to diverse applications. To address this, it is promising to extend the simulation environment described in § 4.2 to enable end-to-end explainability testing throughout the development pipeline. For example, simulation-based causal analysis can be employed to trace and explain system misbehaviors [29]. Integrating such techniques into AI development frameworks, like Machine Learning Operations (MLOps), positions XAI validation as a critical component of trustworthy XR development.

5 BROADER CHALLENGES

Beyond technical challenges, the responsible evolution of XR glasses should address broader societal concerns. We discuss three key pillars: ethical governance, user accessibility and inclusivity, and open development ecosystems. We argue that advancing these domains is critical to empowering end users, supporting developers, and aligning industry innovation with collective human values.

5.1 Ethics Governance

The pervasive computing nature of XR glasses poses risks to human cognition. Avatar-centric apps with social features (e.g, communications with other users' avatars) may induce dissociative identity disorders, where users struggle to distinguish between physical and virtual selves, potentially impairing cognitive functioning [11].

In addition, XR glasses threaten significant exploitation of human attention. Unlike existing addictive interfaces (e.g., personalized YouTube feeds), their always-on displays can inject unskippable ads, and behaviorally manipulative content directly into users' field of view. This "cognitive hijacking" leads to significant information overloads which can degrade human mental autonomy at neurological levels [4]. To prevent these harms, we urge AI and XR communities to work with policymakers on regulatory frameworks, like the General Data Protection Regulation (GDPR)[7], to establish attention sovereignty principles. These would restrict manipulative patterns and empower users to regain cognitive control.

5.2 Accessibility and Inclusivity

XR glasses should adapt to diverse user capabilities, such as replacing gesture inputs with gaze dwell-time for motor-impaired users. Their multimodal interfaces offer unique accessibility benefits. For instance, voice control supports users with motor impairments, and visual-to-audio translation assists the visually impaired [33]. A multimodal agent, SceneScout from Apple, provides accessible interactions with street view imagery for blind users [13].

While XR glasses show promise for disability support, addressing accessibility and inclusivity issues requires coordinated efforts from various stakeholders. The emerging market leaves key questions unresolved: the scale of disabled user adoption remains unknown. Furthermore, independent developers may not prioritize inclusive design despite likely compliance from major tech companies. We expect that accessibility testing research [10] will transfer effectively to XR glasses once they mature as personal computing platforms.

5.3 Open Development Ecosystems

The nascent XR glasses ecosystems suffer from limited open resources, with few available apps. Most of the apps are closed-source or paid and lack documented industrial best practices. Unlike mature platforms like Android, which are supported by vibrant open repositories (e.g., F-Droid[8]), XR lacks community-driven infrastructures for knowledge sharing. These platforms not only accelerate learning for practitioners but also enable SE research communities to develop and evaluate new techniques using open-source apps.

This scarcity of open-source XR resources stems partly from competitive dynamics, where companies withhold proprietary advancements. While Android XR may introduce more transparent industrial standards, the research community should proactively track industrial progress, including tools, frameworks, and technical stacks to maintain relevance. We advocate for curated knowledge platforms, such as research-focused newsletters that analyze emerging industry insights, fostering timely innovation in XR AI/SE research communities. Openness is essential given XR's security and privacy risks. Trust in XR glasses depends on transparency, which is ideally through open-source code that is verifiable and monitored by independent third parties. Stronger academia-industry collaboration is essential to ensure safe and user-centered XR software.

[7] https://gdpr-info.eu/ [8] https://f-droid.org/

6 CONCLUSION

XR glasses represent a fundamental shift in human-computer interaction, promising to transform everyday software through AI-powered embodied spatial intelligence and multimodal interfaces. This vision paper has charted their unique capabilities while exposing critical research frontiers and broader societal concerns. These challenges necessarily demand interdisciplinary collaboration across AI, SE, and HCI.

Though consumer-grade XR glasses are still emergent, precursor technologies (smart glasses, XR HMDs) already demonstrate the field's research viability. As the XR SE research community is gradually gaining momentum (e.g., a recent special issue of *Virtual and Augmented Reality Software Engineering* in the *Automated Software Engineering* journal), we anticipate a rapid expansion of the domain in the FM era. By establishing solid foundations before mass adoption, the research communities can steer XR glasses development toward becoming equitable, human-centered extensions of our cognitive and physical realities—not merely new devices, but responsible partners in augmenting human potential.

REFERENCES

[1] Michael Abrash. 2021. Creating the Future: Augmented Reality, the next Human-Machine Interface. In *2021 IEEE International Electron Devices Meeting (IEDM)*. https://doi.org/10.1109/IEDM19574.2021.9720526

[2] Apple Inc. 2024. Apple Vision Pro Privacy Overview. https://www.apple.com/privacy/docs/Apple_Vision_Pro_Privacy_Overview.pdf. Accessed: June 2025.

[3] Shahin Atakishiyev, Mohammad Salameh, Hengshuai Yao, and Randy Goebel. 2024. Explainable Artificial Intelligence for Autonomous Driving: A Comprehensive Overview and Field Guide for Future Research Directions. *IEEE Access* (2024). https://doi.org/10.1109/ACCESS.2024.3431437

[4] Stuart J. Barnes, Andrew D. Pressey, and Eusebio Scornavacca. 2019. Mobile ubiquity: Understanding the relationship between cognitive absorption, smartphone addiction and social network services. *Computers in Human Behavior* 90 (2019), 246–258. https://doi.org/10.1016/j.chb.2018.09.013

[5] Rishi Bommasani et al. 2022. On the Opportunities and Risks of Foundation Models. arXiv:2108.07258 [cs.LG] https://arxiv.org/abs/2108.07258

[6] Ingo Börsting, Markus Heikamp, Marc Hesenius, Wilhelm Koop, and Volker Gruhn. 2022. Software Engineering for Augmented Reality - A Research Agenda. *Proc. ACM Hum.-Comput. Interact.* EICS (2022). https://doi.org/10.1145/3532205

[7] Peter Casey, Ibrahim Baggili, and Ananya Yarramreddy. 2021. Immersive Virtual Reality Attacks and the Human Joystick. *IEEE Transactions on Dependable and Secure Computing* (2021). https://doi.org/10.1109/TDSC.2019.2907942

[8] Ralf Doerner, Wolfgang Broll, Paul Grimm, and Bernhard Jung (Eds.). 2022. *Virtual and Augmented Reality (VR/AR): Foundations and Methods of Extended Realities (XR)*. Springer. https://doi.org/10.1007/978-3-030-79062-2

[9] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. 2017. CARLA: An Open Urban Driving Simulator. arXiv:1711.03938 [cs.LG] https://arxiv.org/abs/1711.03938

[10] Marcelo Medeiros Eler, Jose Miguel Rojas, Yan Ge, and Gordon Fraser. 2018. Automated Accessibility Testing of Mobile Apps. In *2018 IEEE 11th International Conference on Software Testing, Verification and Validation (ICST)*. https://doi.org/10.1109/ICST.2018.00021

[11] Gregory P. Garvey. 2010. Dissociation in virtual reality: depersonalization and derealization. In *The Engineering Reality of Virtual Reality 2010*, Ian E. McDowall and Margaret Dolinsky (Eds.), Vol. 7525. International Society for Optics and Photonics, SPIE, 75250A. https://doi.org/10.1117/12.843641

[12] Ruizhen Gu, José Miguel Rojas, and Donghwan Shin. 2025. Software Testing for Extended Reality Applications: A Systematic Mapping Study. *Automated Software Engineering* (2025).

[13] Gaurav Jain, Leah Findlater, and Cole Gleason. 2025. SceneScout: Towards AI Agent-driven Access to Street View Imagery for Blind Users. arXiv:2504.09227 [cs.HC] https://arxiv.org/abs/2504.09227

[14] Martin Lachmair, Martin Fischer, and Peter Gerjets. 2022. Action-control mappings of interfaces in virtual reality: A study of embodied interaction. *Frontiers in Virtual Reality* 3 (11 2022), 976849. https://doi.org/10.3389/frvir.2022.976849

[15] Kit Yung Lam, Lik Hang Lee, and Pan Hui. 2021. A2W: Context-Aware Recommendation System for Mobile Augmented Reality Web Browser. In *Proceedings of the 29th ACM International Conference on Multimedia*. ACM. https://doi.org/10.1145/3474085.3475413

[16] Shuqing Li, Binchang Li, Yepang Liu, Cuiyun Gao, Jianping Zhang, Shing-Chi Cheung, and Michael R. Lyu. 2024. Grounded GUI Understanding for Vision Based Spatial Intelligent Agent: Exemplified by Virtual Reality Apps. arXiv:2409.10811 [cs.SE] https://arxiv.org/abs/2409.10811

[17] Chu Luo, Jorge Goncalves, Eduardo Velloso, and Vassilis Kostakos. 2020. A Survey of Context Simulation for Testing Mobile Context-Aware Applications. *ACM Comput. Surv.* (2020). https://doi.org/10.1145/3372788

[18] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. 1994. Augmented reality: A class of displays on the reality-virtuality continuum. *Telemanipulator and Telepresence Technologies* (1994). https://doi.org/10.1117/12.197321

[19] Omdia (Informa Tech). 2023. XR Market in 2035 and Beyond: Forecast, Challenges, and the Road to Mass Adoption. https://omdia.tech.informa.com/om135790/xr-market-in-2035-and-beyond-forecast-challenges-and-the-road-to-mass-adoption. Accessed: June 2025.

[20] Shuyin Ouyang, Jie M. Zhang, Mark Harman, and Meng Wang. 2025. An Empirical Study of the Non-Determinism of ChatGPT in Code Generation. *ACM Trans. Softw. Eng. Methodol.* (2025). https://doi.org/10.1145/3697010

[21] Xudong Pan, Mi Zhang, Shouling Ji, and Min Yang. 2020. Privacy Risks of General-Purpose Language Models. In *2020 IEEE Symposium on Security and Privacy (SP)*. https://doi.org/10.1109/SP40000.2020.00095

[22] Jack Parker-Holder et al. 2024. Genie 2: A Large-Scale Foundation World Model. (2024). https://deepmind.google/discover/blog/genie-2-a-large-scale-foundation-world-model/

[23] Scott Drew Pendleton et al. 2017. Perception, Planning, Control, and Coordination for Autonomous Vehicles. *Machines* (2017). https://doi.org/10.3390/machines5010006

[24] Kui Ren, Qian Wang, Cong Wang, Zhan Qin, and Xiaodong Lin. 2020. The Security of Autonomous Driving: Threats, Defenses, and Future Directions. *Proc.*

IEEE (2020). https://doi.org/10.1109/JPROC.2019.2948775

[25] Franziska Roesner and Tadayoshi Kohno. 2023. Security and Privacy for Augmented Reality: Our 10-Year Retrospective. In *VR4Sec: 1st International Workshop on Security for XR and XR for Security*.

[26] Ahmed M. Salih, Zahra Raisi-Estabragh, Ilaria Boscolo Galazzo, Petia Radeva, Steffen E. Petersen, Karim Lekadir, and Gloria Menegaz. 2025. A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME. *Advanced Intelligent Systems* (2025). https://doi.org/10.1002/aisy.202400304

[27] Gokula Krishnan Santhanam, Ali Alami-Idrissi, Nuno Mota, Anika Schumann, and Ioana Giurgiu. 2020. On Evaluating Explainability Algorithms. https://openreview.net/forum?id=B1xBAA4FwH

[28] Lukas Stappen, Georgios Rizos, and Björn Schuller. 2020. X-AWARE: ConteXt-AWARE Human-Environment Attention Fusion for Driver Gaze Prediction in the Wild. In *Proceedings of the 2020 International Conference on Multimodal Interaction*. ACM. https://doi.org/10.1145/3382507.3417967

[29] Huijia Sun, Christopher M. Poskitt, Yang Sun, Jun Sun, and Yuqi Chen. 2024. ACAV: A Framework for Automatic Causality Analysis in Autonomous Vehicle Accident Recordings. In *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering*. ACM. https://doi.org/10.1145/3597503.3639175

[30] Ryo Suzuki, Mar Gonzalez-Franco, Misha Sra, and David Lindlbauer. 2023. XR and AI: AI-Enabled Virtual, Augmented, and Mixed Reality. In *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM. https://doi.org/10.1145/3586182.3617432

[31] Ryo Suzuki, Mar Gonzalez-Franco, Misha Sra, and David Lindlbauer. 2025. Everyday AR through AI-in-the-Loop. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. ACM. https://doi.org/10.1145/3706599.3706741

[32] Yihe Tang, Wenlong Huang, Yingke Wang, Chengshu Li, Roy Yuan, Ruohan Zhang, Jiajun Wu, and Li Fei-Fei. 2025. UAD: Unsupervised Affordance Distillation for Generalization in Robotic Manipulation. arXiv:2506.09284 [cs.RO] https://arxiv.org/abs/2506.09284

[33] Ethan Waisberg, Joshua Ong, Mouayad Masalkhi, Nasif Zaman, Prithul Sarker, and Alireza Tavakkoli. 2023. Meta smart glasses—large language models and the future for assistive glasses for individuals with vision impairments. *Eye* 38 (12 2023). https://doi.org/10.1038/s41433-023-02842-z

[34] Zhi-Min Wang, Mao-Hang Rao, Shang-Hua Ye, Wei-Tao Song, and Feng Lu. 2025. Towards spatial computing: recent advances in multimodal natural interaction for Extended Reality headsets. *Frontiers of Computer Science* 19 (06 2025). https://doi.org/10.1007/s11704-025-41123-8

[35] Ayman Wazwaz, Khalid Amin, Noura Semary, and Tamer Ghanem. 2024. Dynamic and Distributed Intelligence over Smart Devices, Internet of Things Edges, and Cloud Computing for Human Activity Recognition Using Wearable Sensors. *Journal of Sensor and Actuator Networks* (2024). https://doi.org/10.3390/jsan13010005

[36] Diankun Wu, Fangfu Liu, Yi-Hsin Hung, and Yueqi Duan. 2025. Spatial-MLLM: Boosting MLLM Capabilities in Visual-based Spatial Intelligence. arXiv:2505.23747 [cs.CV] https://arxiv.org/abs/2505.23747

[37] Xuhai Xu et al. 2023. XAIR: A Framework of Explainable AI in Augmented Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM. https://doi.org/10.1145/3544548.3581500

[38] Xiaoyi Yang, Yuxing Wang, Tahmid Rafi, Dongfang Liu, Xiaoyin Wang, and Xueling Zhang. 2024. Towards Automatic Oracle Prediction for AR Testing: Assessing Virtual Object Placement Quality under Real-World Scenes. In *Proceedings of the 33rd ACM SIGSOFT International Symposium on Software Testing and Analysis*. ACM. https://doi.org/10.1145/3650212.3680315

[39] Xiaoyi Yang and Xueling Zhang. 2023. A Study of User Privacy in Android Mobile AR Apps. In *Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering*. ACM. https://doi.org/10.1145/3551349.3560512

[40] Yifan Yao, Jinhao Duan, Kaidi Xu, Yuanfang Cai, Zhibo Sun, and Yue Zhang. 2024. A survey on large language model (LLM) security and privacy: The Good, The Bad, and The Ugly. *High-Confidence Computing* (2024). https://doi.org/10.1016/j.hcc.2024.100211

[41] Baichuan Zeng. 2025. Recent Advances and Future Directions in Extended Reality (XR): Exploring AI-Powered Spatial Intelligence. arXiv:2504.15970 [cs.HC] https://arxiv.org/abs/2504.15970

[42] Hangtao Zhang et al. 2025. BadRobot: Jailbreaking Embodied LLM Agents in the Physical World. In *International Conference on Learning Representations (ICLR)*.

[43] Xiaolu Zhang, Tahmid Rafi, Yuejun Guan, Shuqing Li, and Michael R. Lyu. 2025. Understanding the privacy-realisticness dilemma of the metaverse. *Automated Software Engineering* (2025). https://doi.org/10.1007/s10515-025-00516-6

[44] Xueling Zhang, Rocky Slavin, Xiaoyin Wang, and Jianwei Niu. 2019. Privacy Assurance for Android Augmented Reality Apps. In *2019 IEEE 24th Pacific Rim Intl. Symposium on Dependable Computing (PRDC)*. 114–1141. https://doi.org/10.1109/PRDC47002.2019.00037

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009