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

Ruizhen Gu*, Jingqiong Zhang†, José Miguel Rojas* and Donghwan Shin*
*School of Computer Science, University of Sheffield, Sheffield, UK
{rgu10, j.rojas, d.shin}@sheffield.ac.uk
†School of Electrical and Electronic Engineering, University of Sheffield, Sheffield, UK
jingqiong.zhang@sheffield.ac.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 simultaneously introducing unprecedented challenges at the intersection of AI and Software Engineering (SE). This vision paper aims to catalyse the development of robust spatial software by characterising 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 developing reliable and trustworthy XR systems in the FM era.**

***Index Terms*—AI-powered XR glasses, embodied AI, extended reality, human-AI collaboration, security and privacy, spatial intelligence**

I. Introduction

Rapid advancements in AI and spatial computing technologies are accelerating the evolution of wearable extended reality (XR) devices, which encompass augmented, mixed, and virtual reality (AR, MR, and VR, respectively) [1]. These devices range from *XR head-mounted displays* (XR HMDs), such as Meta Quest 3[1], which provide immersive experiences for entertainment and productivity, to display-free, AI-enabled *smart glasses*, such as 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 [2].

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) [3]. These devices seamlessly superimpose digital content onto physical environments to generate realistic MR experiences. Some consider them as a plausible successor to smartphones, possibly leading to significant changes in human-digital interaction and social communication [4]. Industry prototypes like Meta Orion[3] 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 behaviour generation that extends beyond traditional AI-enabled tasks [5].

Amid the rapid commercialisation of XR glasses by major technology companies, sustained academic attention is essential to tackle underexplored challenges that underpin their reliable and ethical deployment. Current literature remains largely confined to AI and human–computer interaction (HCI) domains, covering topics such as multimodal context-aware fusion for gaze prediction [6] and spatial interaction design frameworks [7]. 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 [8, 9].

This vision paper positions AI-SE for XR glasses as a distinct and timely research frontier, contributing the *first* foundational roadmap to inform future research and development in this emerging domain. By synthesising advances from spatial computing, intelligent systems, and software engineering, we:

1) Establish a conceptual overview of XR glasses as a new class of intelligent software systems.
2) Map their unique computational and interaction capabilities within a unified AI-SE context.
3) Identify possible interdisciplinary research problems at the intersection of AI and SE.
4) Highlight potential societal and ethical implications.

By proactively addressing these dimensions, we equip AI-SE researchers and practitioners with critical insights into the evolving software paradigm.

II. Conceptual Framework

Intending to underpin the unique architectural and interaction paradigms of XR glasses, we propose a conceptual framework to characterise them as integrated software systems that merge advanced spatial and AI capabilities. Inspired by embodied AI agents and wearable computing, our framework highlights two key distinctions: First, unlike autonomous robots, XR glasses are worn directly by users, creating continuous perception-action feedback loops that emphasise human-AI collaboration. Second, they also extend beyond conven-

[1] https://www.meta.com/quest/quest-3/ [2] https://www.meta.com/ai-glasses/
[3] https://www.meta.com/emerging-tech/orion/

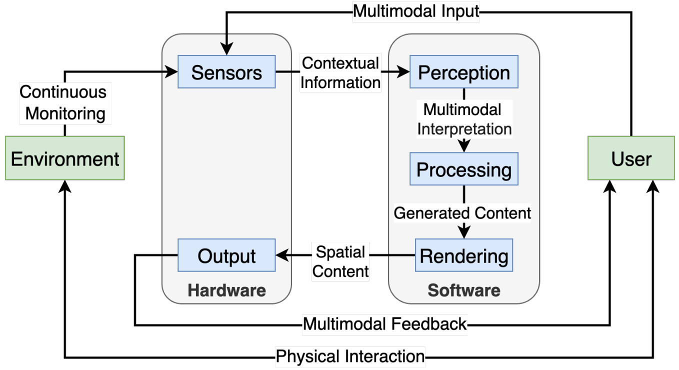

Fig. 1: A conceptual overview of XR glasses systems

tional wearables (e.g., smartwatches) through rich multimodal rendering (e.g., visual, audio) and persistent spatial awareness.

Figure 1 presents our human-centred framework, adapting the paradigm of perception, planning, and control from autonomous systems [10]. The framework advances the field through two key contributions: *AI-powered spatial intelligence* and *human-in-the-loop feedback*. Environmental data from *Sensors* (e.g., camera, mic) flows to the context-aware *Perception* layer for spatial interpretation. The AI-powered *Processing* layer uses FMs, generating dynamic, context-aware behaviours rather than pre-designed content. Next, the *Rendering* layer organises this content into interactive media outputs (e.g., visual, audio), which drive *Output* devices (e.g., lens display, speakers) to guide user interactions. Crucially, user responses (e.g., gestures, voice commands) and updated environmental context then feed back as new sensor inputs, forming a continuous perception-action cycle. This human-in-the-loop feedback positions human cognition as the primary actuator. By integrating these components, our framework could transform traditional embodied AI into collaborative human-AI partnerships for XR glasses.

III. 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.

A. Spatial Intelligence

Spatial intelligence forms the core cognitive capability of XR glasses, transforming raw sensor data into actionable spatial understanding of physical environments [11]. This allows devices not only to *interpret* spatial relationships but also to dynamically *generate* digital content with physical attributes (e.g., collision) [12]. These functions correspond to the *Perception* and *Processing* components of our conceptual framework (Figure 1). FM integration has significantly advanced these capabilities, enabling the generation of interactive 3D environments from minimal inputs. For example, Google DeepMind's Genie 2 [13] can generate playable virtual worlds with realistic object physics, from a single image prompt.

For XR glasses, spatial intelligence represents the foundation of their significant functional advances, enabling more natural, context-aware interaction between users and their surroundings. Its capabilities can be summarised through the following core aspects:

1) **Streamlined environmental understanding:** Spatial relationships and scene semantics can be inferred directly from camera frames, eliminating the need for traditional tracking infrastructure [14, 15]. This can reduce computational demands while improving portability, thereby accelerating software development and deployment pipelines.
2) **Physics-compliant content generation:** Generated digital content gains realistic physical properties [16] (e.g., material, motion), enhancing the believability of XR experience. For example, a virtual ball can respond to gravitational and collision dynamics, bouncing naturally off real surfaces and reacting to user gestures like hitting or throwing.
3) **Affordance recognition**: Spatially intelligent systems can detect the actionable possibilities of real-world objects, a concept known as affordance [17]. It can enable more adaptive and intuitive interactions aligned with human expectations. By recognising roles such as "placement" or "support" on surfaces (e.g., identifying a table as suitable for placing virtual items), XR glasses can generate context-aware recommendations [18].

Building upon these capabilities, spatial intelligence elevates environmental *understanding* to *generation*. As depicted in the *Rendering* layer in our framework (Figure 1), this transcends traditional 3D model rendering using computer graphics techniques. Instead, FMs dynamically synthesise digital content that interacts physically with the real world. This enables seamless reality-virtuality integration while minimising computational overhead. However, it may introduce challenges for SE, such as testing system behaviour across virtually infinite real-world contexts (discussed in § IV-B).

B. Multimodal Interfaces

XR glasses represent an innovation in HCI through integrated multimodal interfaces. These systems incorporate natural input channels, including gesture, gaze, and voice, while delivering corresponding visual and auditory outputs [19]. The multimodal capabilities allow users to interact seamlessly with both digital content and the physical world, while also serving as input channels for FM agents. This dual role 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 more intuitive engagement with both physical environments and FMs. For example, a user might gesture to delineate real-world birds in their field of view (FOV), followed by 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 framework (§ II): multimodal inputs (gesture and voice) initiate FM-mediated environmental interpretation, generating corresponding digital output (visual and audio) that completes the interaction cycle. Beyond established interaction modes, emerging hardware innovations are redefining how input can be captured. A leading example is Meta's *Neural Band* technology, introduced with the *Meta Ray-Ban Display*[4] glasses. The Neural Band employs electromyography (EMG) sensors [20] that detect minute electrical signals from wrist and finger muscles, allowing users to achieve specific functionalities by performing subtle gestures (e.g, pinching, tapping).

Unlike 2D interfaces, this spatial interplay blurs boundaries between environmental interactions and AI responses, which could raise new questions about transparency and user trust. As multimodal systems increasingly rely on FM inference, understanding and explaining AI decision processes to users becomes a challenge (discussed in § IV-C).

C. Distributed Software Architecture

XR glasses typically employ a distributed architecture similar to the multi-layered structure of modern wearable devices [21]. This architecture spans 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.
3) **Cloud layer:** Offers virtually unlimited resources for data storage and processing, FM training, and hosting.

On the software front, specialised 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).

This architectural approach introduces a potential shift toward an *agent-mediated interaction* paradigm. In this model, we envision two types of FM agents that coexist and cooperate:

- **OS-level agent**, acts as an enhanced AI assistant embedded within the OS. It interprets user intent and translates commands into system functions (e.g., launching Google Maps via voice command).
- **Domain-specific agents**, deployed at the application layer, specialise in targeted user scenarios (e.g., restaurant reservations, travel planning).

This cooperation enables complex workflows, transcends traditional app boundaries toward fluid agent ecosystems. It may eliminate the need for dedicated app stores, allowing users to access agent capabilities on demand without installation. 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 multistep real-world tasks without continuous human guidance. For XR glasses prioritising seamless interactions, these agents can minimise the barriers between app boundaries, delivering intuitive experiences to users. Yet, this agent-centric approach may introduce security and privacy challenges, as the continuous data exchange could create new attack surfaces that must be addressed (discussed in § IV-A).

IV. Research Problems

This section identifies possible emerging research problems for XR glasses from both AI and SE perspectives. While challenges like security, validation, and explainability are established in isolation, their manifestation in XR glasses can introduce novel complexities that demand distinct solutions. We focus specifically on problems arising from the unique characteristics of XR glasses, including sensitive data leakage due to always-on sensing, reliability of virtual-physical integration, and explainability needs for FM-guided decision-making. For each challenge, we discuss both the research problem and its envisioned mitigation approaches.

A. Security and Privacy

Security and privacy present critical concerns for XR glasses, given inherent risks in both AI and XR domains [22, 23]. While hardware vulnerabilities exist, we emphasise software risks. Security threats can include severe safety risks from external attacks and internal faults. While less immersive than VR, XR glasses may face similar attacks, such as malicious overlays, cybersickness, or manipulating user behaviours [24]. These risks can become more severe due to XR's integration with the physical world and vulnerabilities in embodied AI systems that might generate harmful actions. Defects can include perceptual failures (e.g., misclassifying critical objects) or vulnerabilities to deceptive prompting techniques [25, 26]. Mitigation strategies should involve multiple stakeholders. XR OSs could embed core safety mechanisms, such as collision detection and emergency features, allowing users to disable displays and use devices as standard glasses. These safeguards should be enforced at the OS level to prevent override by third-party apps.

Regarding privacy, both AI and XR share risks similar to personalised systems, such as sensitive data leakage [27, 28]. XR glasses may amplify these concerns through their persistent environmental perception. Continuous data capture (e.g., visual, audio) 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 § III-C) [29]. This persistent monitoring could introduce unique *bystander privacy* risks absent in conventional systems [23]. Features like "object recognition" could rely on facial and biometric data, collecting sensitive information from non-consenting individuals. For example, an "identify friend" feature in a crowd may collect bystanders' biometric signatures without consent or notification. While prior work has suggested mitigations like restricting visual processing within specified FOV areas to avoid unintentional

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

data capture [29], such measures may not fully address the problem. Architectural safeguards, such as on-device processing, could be essential to prevent sensitive data transmission. For instance, Apple Vision Pro adopts a private cloud computing design to address this concern [30].

B. Validation of Spatial Capabilities

XR glasses face significant SE challenges due to their real-world interactivity, especially in software testing. Core to this challenge is the dynamic blending of real and virtual elements within AR/MR environments, which can lead to environment-dependent failures that are difficult to reproduce and debug. This contrasts fundamentally with traditional context-aware systems like mobile apps, where environmental triggers typically yield discrete and reproducible behavioural changes (e.g., geolocation-based recommendations [31]).

The integration of physical and digital content with persistent spatial capabilities could substantially increase testing complexity, as virtual content behaviour depends on virtually infinite real-world variations [3]. Effective testing should ensure contextually appropriate responses in dynamic environments. Although recent tools have demonstrated the ability to generate realistic 3D user interactions for XR testing purposes [32, 33], these efforts have largely focused on VR apps, overlooking the critical real-world interplay central to AR and MR [9]. Recent work has begun addressing this gap by predicting test oracles for virtual object misplacement in AR [34], yet these approaches still rely on manual intervention. To address these challenges, we suggest developing specialised simulation frameworks that can systematically emulate diverse real-world conditions. Inspired by autonomous driving simulators like CARLA, which support testing of complex systems through controlled scenarios [35], similar infrastructure could be adapted for XR glasses to enable scalable testing of spatial capabilities across various real-world conditions.

C. Explainability

Spatial intelligence (§ III-A) and multimodal interfaces (§ III-B) that define XR glasses often rely on complex AI models, which can suffer from their "black-box" nature, making their decision processes opaque to users. This poses safety risks when untransparent decisions lead to unintended or malicious behaviours (§ IV-A) [36]. Given their always-on operation and deep integration into daily life, XR glasses require Explainable AI (XAI) techniques that clearly reveal decision rationales to users, especially in safety-critical scenarios [37].

Testing and validating XR glasses pose significant challenges due to the dynamic and unpredictable environments in which these devices operate (§ IV-B), and XAI requirements may amplify this. XAI outputs should be reliable and aligned with users' needs, despite FMs' non-determinism [38] and limited interpretability. Personalised XAI could further complicate this, as individual preferences (e.g., concise explanations) introduce additional variability [37]. Thus, ensuring accurate and consistent explanations across diverse environments and users is critical for the safety and usability of XR glasses. However, current XAI evaluation methods are often model-specific [39, 40] and may not scale to the vast, open-ended scenarios in XR contexts. We identify this scalability gap as a core SE challenge. Future research could include extending the simulation environment described in § IV-B to support end-to-end explainability testing throughout the development pipeline. For example, simulation-based causal analysis can be used to trace and explain system misbehaviours [41]. Integrating such techniques into Machine Learning Operations (MLOps) pipelines could position XAI validation as a fundamental component of trustworthy XR development.

V. Broader Challenges

This section examines three areas beyond technical challenges: ethical governance, user accessibility and inclusivity, and open development ecosystems. Progress in these domains could be valuable for XR glasses to benefit diverse users, support developer communities, and align with societal values.

A. Ethics Governance

The pervasive nature of XR glasses could pose risks to human cognition. Avatar-centric apps with social features (e.g, communications with other users' avatars) might contribute to dissociative identity disorders, where users may struggle to distinguish between physical and virtual selves, potentially affecting cognitive functioning [42].

Moreover, XR glasses could enable new forms of human attention exploitation. Unlike current addictive interfaces (e.g., personalised social media feeds), their always-on displays can inject unskippable ads and behaviourally manipulative content directly into users' FOV. This latent "cognitive hijacking" could lead to information overload, which may degrade human mental autonomy at neurological levels [43]. To mitigate these harms, we suggest collaboration between AI-SE communities to work with policymakers on regulatory frameworks, like the General Data Protection Regulation (GDPR)[7], to establish attention sovereignty principles. These could restrict manipulative patterns and support user cognitive control.

B. Accessibility and Inclusivity

XR glasses could adapt to diverse user capabilities through multimodal interfaces. For example, gaze dwell-time might benefit users with motor impairments, while voice control and visual-to-audio translation could assist visual-impaired users [5]. Systems like Apple's SceneScout demonstrate how multimodal agents can provide accessible interactions with spatial content for blind users [44].

While XR glasses show promise for supporting individuals with disabilities, addressing these accessibility issues may require coordinated efforts from various stakeholders. Key challenges in the emerging XR market may include uncertain adoption rates among users with disabilities and varying commitment to inclusive design among software developers and vendors. SE could play a crucial role in addressing these challenges, where established accessibility testing techniques [45] could be adapted for XR platforms.

[7] https://gdpr-info.eu/

C. Open Development Ecosystems

The nascent XR glasses ecosystems face challenges with limited open resources and a scarcity of available apps. Many existing apps are closed-source or commercial, with limited documentation of best practices. Unlike mature platforms like Android, which benefit from vibrant open repositories (e.g., F-Droid[8]), XR currently lacks strong community-driven infrastructures for knowledge sharing. Such ecosystems could be beneficial for practitioner learning and SE research communities to develop and evaluate new techniques.

This scarcity of open-source XR resources may partly result from competitive market dynamics, where companies tend to protect proprietary advancements. Research communities can benefit from actively monitoring industrial development and utilising curated knowledge platforms (e.g., research newsletters that analyse emerging tools and technology stacks) to foster timely innovation. Openness also supports the security and privacy concerns associated with XR. Transparency, verifiable open-source code, likely monitored by independent third parties, could help build trust in XR systems. Enhanced academia-industry collaboration could be valuable for developing safer, more user-centred XR software.

VI. Conclusion

XR glasses represent a conceivable shift in human-computer interaction, combining AI-powered spatial intelligence and multimodal interfaces to create new software paradigms. This vision paper has synthesised the emerging field of AI-SE for XR glasses, establishing a conceptual framework, mapping unique capabilities, identifying research challenges, and highlighting broader societal implications. Addressing these challenges may benefit from interdisciplinary collaboration across AI, software engineering, and HCI communities.

Though consumer-grade XR glasses are still emergent, precursor technologies (e.g., smart glasses, XR HMDs) demonstrate the field's research potential. Given the XR SE research community is gradually gaining momentum (e.g., a recent special issue on *Virtual and Augmented Reality Software Engineering* in *Automated Software Engineering*), we expect a rapid growth of this domain in the FM era. By addressing these technical and societal considerations early, researchers can help guide XR glasses toward equitable, human-centred applications that responsibly augment human capabilities.

Acknowledgment

This work was partially supported by the Engineering and Physical Sciences Research Council (EPSRC) [EP/Y014219/1].

References

[1] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino, "Augmented reality: A class of displays on the reality-virtuality continuum," *Telemanipulator and Telepresence Technologies*, 1994.

[2] R. Bommasani *et al.*, "On the opportunities and risks of foundation models," 2022. [Online]. Available: https://arxiv.org/abs/2108.07258

[3] R. Suzuki, M. Gonzalez-Franco, M. Sra, and D. Lindlbauer, "Everyday AR through AI-in-the-Loop," in *Proc. of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. ACM, 2025.

[4] M. Abrash, "Creating the future: Augmented reality, the next human-machine interface," in *IEEE Intl. Electron Devices Meeting (IEDM)*, 2021.

[5] E. Waisberg, J. Ong, M. Masalkhi, N. Zaman, P. Sarker, and A. Tavakkoli, "Meta smart glasses—large language models and the future for assistive glasses for individuals with vision impairments," *Eye*, 2023.

[6] L. Stappen, G. Rizos, and B. Schuller, "X-AWARE: ConteXt-AWARE human-environment attention fusion for driver gaze prediction in the wild," in *Proc. of the 2020 Intl. Conference on Multimodal Interaction (ICMI)*. ACM, 2020.

[7] Z.-M. Wang, M.-H. Rao, S.-H. Ye, W.-T. Song, and F. Lu, "Towards spatial computing: recent advances in multimodal natural interaction for extended reality headsets," *Frontiers of Computer Science*, 06 2025.

[8] I. Börsting, M. Heikamp, M. Hesenius, W. Koop, and V. Gruhn, "Software engineering for augmented reality - a research agenda," *Proc. ACM Hum.-Comput. Interact.*, no. EICS, 2022.

[9] R. Gu, J. M. Rojas, and D. Shin, "Software testing for extended reality applications: A systematic mapping study," *Automated Software Engineering*, 2025.

[10] S. D. Pendleton *et al.*, "Perception, planning, control, and coordination for autonomous vehicles," *Machines*, 2017.

[11] D. Wu, F. Liu, Y.-H. Hung, and Y. Duan, "Spatial-MLLM: Boosting MLLM capabilities in visual-based spatial intelligence," 2025. [Online]. Available: https://arxiv.org/abs/2505.23747

[12] B. Zeng, "Recent advances and future directions in extended reality (XR): Exploring AI-Powered spatial intelligence," 2025. [Online]. Available: https://arxiv.org/abs/2504.15970

[13] J. Parker-Holder *et al.*, "Genie 2: A large-scale foundation world model," 2024. [Online]. Available: https://deepmind.google/discover/blog/genie-2-a-large-scale-foundation-world-model/

[14] S. Li, B. Li, Y. Liu, C. Gao, J. Zhang, S.-C. Cheung, and M. R. Lyu, "Grounded gui understanding for vision based spatial intelligent agent: Exemplified by virtual reality apps," 2024. [Online]. Available: https://arxiv.org/abs/2409.10811

[15] R. Doerner, W. Broll, P. Grimm, and B. Jung, Eds., *Virtual and Augmented Reality (VR/AR): Foundations and Methods of Extended Realities (XR)*. Springer, 2022.

[16] R. Suzuki, M. Gonzalez-Franco, M. Sra, and D. Lindlbauer, "XR and AI: Ai-enabled virtual, augmented, and mixed reality," in *Adjunct Proc. of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, 2023.

[8] https://f-droid.org/

[17] Y. Tang, W. Huang, Y. Wang, C. Li, R. Yuan, R. Zhang, J. Wu, and L. Fei-Fei, "UAD: Unsupervised affordance distillation for generalization in robotic manipulation," 2025. [Online]. Available: https://arxiv.org/abs/2506.09284

[18] K. Y. Lam, L. H. Lee, and P. Hui, "A2W: Context-aware recommendation system for mobile augmented reality web browser," in *Proc. of the 29th ACM Intl. Conference on Multimedia (MM)*. ACM, 2021.

[19] M. Lachmair, M. Fischer, and P. Gerjets, "Action-control mappings of interfaces in virtual reality: A study of embodied interaction," *Frontiers in Virtual Reality*, 2022.

[20] M. B. I. Reaz, M. S. Hussain, and F. Mohd-Yasin, "Techniques of emg signal analysis: detection, processing, classification and applications," *Biological procedures online*, 2006.

[21] A. Wazwaz, K. Amin, N. Semary, and T. Ghanem, "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.

[22] Y. Yao, J. Duan, K. Xu, Y. Cai, Z. Sun, and Y. Zhang, "A survey on large language model (llm) security and privacy: The good, the bad, and the ugly," *High-Confidence Computing*, 2024.

[23] F. Roesner and T. Kohno, "Security and privacy for augmented reality: Our 10-year retrospective," in *VR4Sec: 1st Intl. Workshop on Security for XR and XR for Security*, 2023.

[24] P. Casey, I. Baggili, and A. Yarramreddy, "Immersive virtual reality attacks and the human joystick," *IEEE Trans. on Dependable and Secure Computing*, 2021.

[25] K. Ren, Q. Wang, C. Wang, Z. Qin, and X. Lin, "The security of autonomous driving: Threats, defenses, and future directions," *Proc. of the IEEE*, 2020.

[26] H. Zhang *et al.*, "Badrobot: Jailbreaking embodied LLM agents in the physical world," in *Intl. Conference on Learning Representations (ICLR)*, 2025.

[27] X. Zhang, T. Rafi, Y. Guan, S. Li, and M. R. Lyu, "Understanding the privacy-realisticness dilemma of the metaverse," *Automated Software Engineering*, 2025.

[28] X. Pan, M. Zhang, S. Ji, and M. Yang, "Privacy risks of general-purpose language models," in *2020 IEEE Symposium on Security and Privacy (SP)*, 2020.

[29] X. Zhang, R. Slavin, X. Wang, and J. Niu, "Privacy assurance for android augmented reality apps," in *2019 IEEE 24th Pacific Rim Intl. Symposium on Dependable Computing (PRDC)*, 2019.

[30] Apple Inc., "Apple vision pro privacy overview," https://www.apple.com/privacy/docs/Apple_Vision_Pro_Privacy_Overview.pdf, 2024, accessed: June 2025.

[31] C. Luo, J. Goncalves, E. Velloso, and V. Kostakos, "A survey of context simulation for testing mobile context-aware applications," *ACM Comput. Surv.*, 2020.

[32] R. Gu and J. M. Rojas, "A test automation framework for user interaction in extended reality applications," in *40th IEEE/ACM Intl. Conference on Automated Software Engineering Workshops (ASEW)*, 2025.

[33] R. Gu, J. M. Rojas, and D. Shin, "XRintTest: An automated framework for user interaction testing in extended reality applications," in *40th IEEE/ACM Intl. Conference on Automated Software Engineering (ASE)*, 2025.

[34] X. Yang, Y. Wang, T. Rafi, D. Liu, X. Wang, and X. Zhang, "Towards automatic oracle prediction for ar testing: Assessing virtual object placement quality under real-world scenes," in *Proc. of the 33rd ACM SIGSOFT Intl. Symposium on Software Testing and Analysis (ISSTA)*. ACM, 2024.

[35] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," 2017. [Online]. Available: https://arxiv.org/abs/1711.03938

[36] S. Atakishiyev, M. Salameh, H. Yao, and R. Goebel, "Explainable artificial intelligence for autonomous driving: A comprehensive overview and field guide for future research directions," *IEEE Access*, 2024.

[37] X. Xu *et al.*, "XAIR: A framework of explainable ai in augmented reality," in *Proc. of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, 2023.

[38] S. Ouyang, J. M. Zhang, M. Harman, and M. Wang, "An empirical study of the non-determinism of ChatGPT in code generation," *ACM Trans. Softw. Eng. Methodol. (TOSEM)*, 2025.

[39] G. K. Santhanam, A. Alami-Idrissi, N. Mota, A. Schumann, and I. Giurgiu, "On evaluating explainability algorithms," 2020. [Online]. Available: https://openreview.net/forum?id=B1xBAA4FwH

[40] A. M. Salih, Z. Raisi-Estabragh, I. B. Galazzo, P. Radeva, S. E. Petersen, K. Lekadir, and G. Menegaz, "A perspective on explainable artificial intelligence methods: SHAP and LIME," *Advanced Intelligent Systems*, 2025.

[41] H. Sun, C. M. Poskitt, Y. Sun, J. Sun, and Y. Chen, "ACAV: A framework for automatic causality analysis in autonomous vehicle accident recordings," in *Proc. of the IEEE/ACM 46th Intl. Conference on Software Engineering (ICSE)*. ACM, 2024.

[42] G. P. Garvey, "Dissociation in virtual reality: depersonalization and derealization," in *The Engineering Reality of Virtual Reality 2010*, I. E. McDowall and M. Dolinsky, Eds., Intl. Society for Optics and Photonics. SPIE, 2010.

[43] S. J. Barnes, A. D. Pressey, and E. Scornavacca, "Mobile ubiquity: Understanding the relationship between cognitive absorption, smartphone addiction and social network services," *Computers in Human Behavior*, 2019.

[44] G. Jain, L. Findlater, and C. Gleason, "Scenescout: Towards ai agent-driven access to street view imagery for blind users," 2025. [Online]. Available: https://arxiv.org/abs/2504.09227

[45] M. M. Eler, J. M. Rojas, Y. Ge, and G. Fraser, "Automated accessibility testing of mobile apps," in *IEEE 11th Intl. Conference on Software Testing, Verification and Validation (ICST)*, 2018.