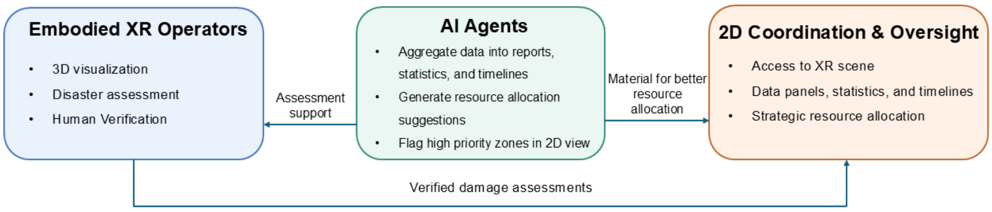

Figure 1: Proposed asymmetric XR–AI collaboration enabling calibrated trust and resilience in disaster management scenarios. Human-in-the-loop verification by XR operators before reaching 2D managers supports calibrated trust, while role-based conflict resolution and offline state caching ensure resilience under network failures.

Calibrated Trust and Resilience in Many-to-Many AI-Integrated XR Collaboration for Disaster Response

Mohammad Javadian Farzaneh[1,*], Selim Balcisoy[1]

[1]*Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul, Turkey*

Abstract

Recent demand for remote collaboration has grown significantly; however, most existing systems target one-to-one or one-to-many interactions, leaving many-to-many (MTM) collaboration underexplored. This limitation becomes critical in high-risk disaster management scenarios, where AI-supported remote collaboration enables timely and informed decision-making. Extended reality (XR) offers strong support for spatial understanding and remote coordination, yet technical and functional constraints make it insufficient in isolation. We argue that MTM AI+XR systems must be designed around calibrated trust, resilience, and situated explainability. This position paper addresses this gap by presenting a WebXR-based testbed and a conceptual architecture that integrates these three design principles into AI-supported XR disaster management for high-stakes environments. Through asymmetric collaboration between embodied field coordinators in XR and remote emergency managers in 2D interfaces, our architecture demonstrates how these design shifts address the unique challenges of MTM coordination under time pressure.

Keywords

Extended Reality (XR), AI Agents, Many-to-Many Collaboration, Disaster Management, Decision-Making

1. Introduction

Collaboration involves joint, interdependent activities toward a common goal [1], co-located or remote [2]. Remote collaboration has grown for emergencies [3, 4], yet existing tools often fail to convey body language, spatial context, and social cues essential for face-to-face closeness [5]. Extended Reality (XR) technologies leverage motion, face, and eye tracking to enable richer human communication similar to face-to-face interactions [5, 6]. However, XR adoption faces challenges including hardware availability, spatial constraints, and cybersickness. Asymmetric scenarios where some users join via mobile devices, tablets, or laptops can help address these limitations [5, 7]. In XR-supported collaboration, latency and incomplete non-verbal cues can cause miscommunication, which large language models (LLMs) can mitigate by grounding communication, clarifying intent, and supporting collaborative sense-making.

Currently, most scenarios deploy AI agents in one-to-one communication settings [8]. However, real-world collaborations increasingly occur in many-to-many (MTM) contexts. Our literature review

CHI'26: XR4CE workshop, April 14, 2026, Barcelona, Spain

*Corresponding author.
mohammad.javadian@sabanciuniv.edu (M. J. Farzaneh); balcisoy@sabanciuniv.edu (S. Balcisoy)
https://analyticslab.sabanciuniv.edu/ (S. Balcisoy)
0009-0003-1999-7939 (M. J. Farzaneh); 0000-0002-6495-7341 (S. Balcisoy)

reveals that most collaboration scenarios remain designed for one-to-one/many interactions, with MTM collaboration remaining at the conceptual or prototype stage. Unlike one-to-one/many settings, MTM XR collaboration requires concurrent multi-user state manipulation, role-dependent authority handling, and conflict resolution over shared resources, challenges that remain only partially addressed in current XR collaboration systems [9, 3].

This highlights a critical gap: the need to better support MTM collaboration, where multiple remote participants must coordinate concurrently, negotiate control of shared artifacts, and maintain situational awareness of both people and system state [9]. This gap becomes particularly acute in high-stakes applications such as disaster management, where MTM remote collaboration combined with AI technologies can enable rapid decision-making, swift response, and efficient damage mitigation. Decision-making in such high-risk situations is inherently challenging due to time pressure and uncertain, dynamic environments [10, 11]. While some studies suggest that virtuality does not affect decision quality [12, 13], these findings focus on tasks manageable through communication alone. In contrast, disaster management requires 3D visualizations, multi-party collaboration, and information modeling, capabilities that XR is uniquely positioned to support [11].

This position paper presents an in-progress WebXR-based MTM collaboration testbed that synchronizes object creation, manipulation, and deletion across VR headsets and 2D web clients. Building on this technical foundation, we propose a conceptual architecture demonstrating how AI-integrated XR-supported MTM environments can enhance decision-making in disaster management. The architecture explores how AI agents can perceive and function within spatial XR interfaces to coordinate multiple participants while maintaining transparency and user oversight. We discuss key design decisions around trust, explainability, and resilience, and use them to motivate future work examining how AI-supported MTM XR collaboration compares to one-to-one settings in decision quality and response time.

2. Related Work

This section reviews XR-supported remote collaboration technologies, structured around constraint classes of the FURPS+ framework [14]. FURPS+ provides a systematic method for organizing both functional capabilities and system-level constraints in software systems [15], making it well-suited for categorizing the technical and experiential dimensions of XR collaboration systems relevant to MTM collaboration. We focus on XR Type, Modalities, and AI Use, that can have application in MTM collaborations for disaster management.

XR supports stronger social and spatial presence than audio or video by conveying embodied cues such as gaze, gesture, and depth, enabling both human–human and human–object interaction in shared spaces [16]. XR modality choice is domain-dependent: AR maintains physical environment awareness in LLM-mediated conversations [17] and office-based analytics scenarios [18], while VR is often preferred for focused remote collaboration in immersive analytics [18], education [19], and expert–worker industrial settings [16]. Avatar representation is critical for establishing social presence and trust. Realistic avatars that convey non-verbal cues such as eye gaze, body movements, and facial expressions improve co-presence and reduce misunderstandings [19, 5]. These non-verbal signals establish trust, regulate conversational flow, and signal attention or disengagement.

Integrating AI agents, particularly LLMs, helps mitigate technical limitations in immersive environments by providing contextually relevant responses. Prior work shows that LLMs enhance creativity, innovation, team dynamics, and communication quality [20, 21]. We observed diverse uses of AI, which can be classified into four roles: Perception, Dialogue, Control/Agency, and Governance. Perception refers to AI systems that sense and interpret multimodal inputs to support awareness and decision-making [22, 23]. Dialogue captures AI agents that mediate conversational interactions through natural language or speech pipelines [8, 5]. Control/Agency concerns AI agents that guide task execution through instructions or embodied cues [24]. Governance includes AI systems managing ethical, privacy, or orchestration aspects to ensure transparency [25].

Table 1
Mapping between design principles and corresponding system mechanisms. Each mechanism is designed to directly operationalize its corresponding principle within the MTM collaboration layer.

Principle	Mechanism
Explainability	Explanation service + evidence store
Trust	Provenance + confidence metadata
Resilience	Offline log + reconciliation
MTM Collaboration	Shared state + role permissions

3. Conceptual Architecture

We propose an AI-integrated XR architecture for MTM disaster management collaboration comprising three interconnected components: an immersive layer (VR/AR), a 2D web interface, and AI agents. The asymmetric architecture defines two primary roles to facilitate MTM disaster response: Embodied XR Operators and 2D Coordinators, as conceptually illustrated in Figure 1. XR operators navigate a 3D digital twin of the disaster site, performing high-fidelity spatial tasks such as tagging hazards, verifying victim locations, and organizing the resources. Simultaneously, 2D coordinators, serving as decision-makers and remote emergency managers, use a synchronized, dashboard-style view of the XR scene to oversee the broader operation, accessing aggregated data panels, timelines, and resource allocation tools to make strategic resource allocation decisions across multiple zones based on XR field assessments and AI recommendations.

AI agents bridge these roles by synthesizing raw spatial data from XR operators into actionable heatmaps, statistics, reports, and timelines for the 2D oversight team, significantly saving time. They will operate within the spatial XR environment to support damage assessment and information management; in the 2D view, they will monitor resource allocations, alert emergency managers to resource conflicts, suggest optimal distributions based on constraints, and flag high-priority zones using data from the XR environment. Consider an earthquake response scenario with three field coordinators in VR assessing different damaged zones simultaneously, two emergency managers in 2D monitoring overall resource constraints, and one AI agent. VR users independently evaluate structural damage and identify survivors in parallel across geographic areas.

All human and AI actions are treated as events in a shared workspace and immediately reflected for all participants, following an event-sourced, optimistic concurrency model with role- and timestamp-based reconciliation. Each update includes source, interface, timestamp, and confidence metadata, with conflicts resolved using role-based priority rules. AI recommendations follow a review-first, human-in-the-loop process with uncertainty indicators, ensuring human control and positioning the AI as a collaborative partner rather than a black-box decision-maker. During network instability, the system uses localized state synchronization and a priority-based merge protocol that resolves MTM conflicts by prioritizing ground-truth tags from nearby XR operators and timestamps, maintaining a consistent global view across modalities. Table 1 summarizes how the proposed design principles map to system mechanisms. To assess whether the proposed architecture achieves calibrated trust and resilience, we will evaluate with domain experts completing disaster scenarios, comparing MTM and one-to-many baselines in a within-subjects design using measures of decision quality (expert-rated allocation optimality), time to stable plan, conflict rate, NASA-TLX workload [26], and trust calibration via post-task surveys using established trust scales [27], measuring dimensions of reliability, competence, and faith in AI recommendations across both XR and 2D roles. We also implemented a WebXR-based MTM testbed using Meta's Interaction SDK (IWSDK) for hand and controller input, Three.js for 3D scene rendering, and Socket.IO for real-time state synchronization, demonstrating synchronized object creation, manipulation, and deletion across VR headsets and 2D web clients, establishing technical feasibility for asymmetric MTM collaboration. The current position paper focuses on the collaboration architecture and design principles; digital twin formation methods fall within the domain of sensing and reconstruction research, which is independent of the collaboration layer addressed here.

4. Discussion: Designing for Trust, Resilience, and Explainability

4.1. Explainability: How to present AI recommendations differently in current modalities?

AI recommendations must be presented differently across modalities to match user roles and cognitive contexts. In XR, AI agents assess initial damage using historical and real-time incident data, presenting findings with confidence scores (0-100%) and supporting evidence that field coordinators can rapidly verify against direct spatial observation. In the 2D interface, AI agents synthesize information from multiple sources, including XR field assessments and verified external data to generate comprehensive incident reports for emergency managers. AI also provides resource allocation suggestions with detailed rationale explaining the basis for each recommendation, supporting informed decision-making under time pressure. This asymmetric presentation strategy ensures each role receives appropriately scoped information: XR users get verification-focused feedback during assessment, while 2D managers receive decision-support analytics for strategic resource coordination, enabling calibrated trust between humans and AI. To avoid overtrust, AI agent outputs are presented through spatially anchored 2D overlays in XR and structured panel summaries in the 2D interface, rather than anthropomorphic avatars. In XR, confidence will be shown with visual cues (e.g., confidence percentages and evidence overlays) to support rapid verification in the field, whereas in the 2D interface it is communicated through summaries and explanatory rationales to support strategic decision-making.

4.2. Trust: How individuals should trust AI in their recommendations?

Establishing appropriate trust is crucial when emergency managers must act on AI recommendations under time pressure without direct field visibility. Our architecture addresses this through a two-tier trust model: damage assessments undergo human-in-the-loop verification by the XR field coordinators before reaching managers while AI provides resource allocation suggestions with supporting rationale directly to emergency managers in 2D view. This separation allows managers to distinguish verified facts from AI-generated suggestions, calibrating trust appropriately for each information type. The key remaining challenge is enabling rapid evaluation of AI suggestions when managers lack time for detailed analysis, requiring careful design of confidence indicators and rationale presentation. Confidence scores should be calibrated using expert-labeled disaster scenarios and iterative simulation-based evaluation, rather than presented as fixed model outputs. When uncertainty remains high, the system can surface multiple independent AI agents offering alternative assessments, shifting the recommendation from prescriptive guidance to decision support and prompting additional human verification.

4.3. Resilience: How the system will handle network failure?

Network failures must not stop disaster response operations. Our architecture maintains resilience by caching the last synchronized state on all devices, so disconnected users have access to recent data and can continue working. XR headsets embed small language models locally, allowing field coordinators to assess damage and make resource allocation decisions for their immediate zone when offline. Upon reconnection, decisions made by offline XR users take priority over allocations made by emergency managers during the same period, since field coordinators have direct observation of ground conditions. Moreover, they will get recent updates from other parties, too. However, this creates coordination challenges requiring explicit reconciliation: upon reconnection, the system applies a priority-based merge protocol where (1) field coordinator decisions override manager decisions for the same resources, (2) conflicts between same-tier users resolve by timestamp (earliest wins), and (3) all affected parties receive notifications showing which decisions were superseded, why, and by whom. The system highlights outdated data with visual indicators and provides rollback options for overridden allocations.

Declaration on Generative AI

During the preparation of this work, the authors (Mohammad Javadian Farzaneh and Selim Balcisoy) used ChatGPT in order to: Grammar and spelling check, Paraphrase and reword. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] B. Marques, S. Silva, J. Alves, T. Araújo, P. Dias, B. S. Santos, A conceptual model and taxonomy for collaborative augmented reality, IEEE Transactions on Visualization and Computer Graphics 28 (2022) 5113–5133. doi:10.1109/TVCG.2021.3101545.

[2] B. Marques, S. Silva, P. Dias, B. S. Santos, One-to-many remote scenarios: The next step in collaborative extended reality (xr) research, in: Workshop on Analytics, Learning & Collaboration in eXtended Reality (XR-WALC). ACM International Conference on Interactive Media Experiences (IMX 2022), 2022, pp. 1–6.

[3] P. Wang, H. Yang, M. Billinghurst, S. Zhao, Y. Wang, Z. Liu, Y. Zhang, A survey on xr remote collaboration in industry, Journal of Manufacturing Systems 81 (2025) 49–74. URL: https://www.sciencedirect.com/science/article/pii/S0278612525001256. doi:https://doi.org/10.1016/j.jmsy.2025.05.016.

[4] R. Druta, C. Druta, P. Negirla, I. Silea, A review on methods and systems for remote collaboration, Applied Sciences 11 (2021). URL: https://www.mdpi.com/2076-3417/11/21/10035. doi:10.3390/app112110035.

[5] C. Merz, C. Wienrich, M. E. Latoschik, Does voice matter? the effect of verbal communication and asymmetry on the experience of collaborative social xr, in: 2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2024, pp. 1127–1136. doi:10.1109/ISMAR62088.2024.00129.

[6] H. J. Smith, M. Neff, Communication behavior in embodied virtual reality, in: Proceedings of the 2018 CHI conference on human factors in computing systems, 2018, pp. 1–12.

[7] E. Chang, Y. Lee, M. Billinghurst, B. Yoo, Efficient vr-ar communication method using virtual replicas in xr remote collaboration, International Journal of Human-Computer Studies 190 (2024) 103304. URL: https://www.sciencedirect.com/science/article/pii/S1071581924000880. doi:https://doi.org/10.1016/j.ijhcs.2024.103304.

[8] J. Bendarkawi, A. Ponce, S. C. Mata, A. Aliu, Y. Liu, L. Zhang, A. Liaqat, V. N. Rao, A. Monroy-Hernández, Conversar: Exploring embodied llm-powered group conversations in augmented reality for second language learners, in: Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, CHI EA '25, Association for Computing Machinery, New York, NY, USA, 2025. URL: https://doi.org/10.1145/3706599.3720162. doi:10.1145/3706599.3720162.

[9] Y. Wang, D. Gong, R. Xiao, X. Wu, H. Zhang, A systematic review on extended reality-mediated multi-user social engagement, Systems 12 (2024). URL: https://www.mdpi.com/2079-8954/12/10/396. doi:10.3390/systems12100396.

[10] W. Sun, P. Bocchini, B. D. Davison, Applications of artificial intelligence for disaster management, Natural Hazards 103 (2020) 2631–2689. URL: https://doi.org/10.1007/s11069-020-04124-3. doi:10.1007/s11069-020-04124-3.

[11] S. Khanal, U. S. Medasetti, M. Mashal, B. Savage, R. Khadka, Virtual and augmented reality in the disaster management technology: A literature review of the past 11 years, Frontiers in Virtual Reality Volume 3 - 2022 (2022). URL: https://www.frontiersin.org/journals/virtual-reality/articles/10.3389/frvir.2022.843195. doi:10.3389/frvir.2022.843195.

[12] I. Moser, S. Chiquet, S. K. Strahm, F. W. Mast, P. Bergamin, Group decision-making in multi-user immersive virtual reality, Cyberpsychology, Behavior, and Social Networking 23 (2020) 846–853.

[13] M. Sinnemann, M. Weiss, Team virtuality and innovation: A meta-analysis of the moderating role of team design, Journal of Organizational Behavior 46 (2025) 867–888. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/job.2889. doi:https://doi.org/10.1002/job.2889. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/job.2889.

[14] P. Eeles, Capturing architectural requirements, IBM Corp., Armonk, NY, USA, White Paper, 2001. [Online]. Available: https://www.researchgate.net/publication/329760910_Capturing_Architectural_Requirements.

[15] J. Motejlek, E. Alpay, Taxonomy of virtual and augmented reality applications in education, IEEE Transactions on Learning Technologies 14 (2021) 415–429. doi:10.1109/TLT.2021.3092964.

[16] L. Ziegfeld, M. Michel, I. Stuldreher, A framework for assessing and enhancing social and spatial presence in mediated communication to support remote collaboration, in: T. Ahram, R. Taiar (Eds.), Human Interaction and Emerging Technologies (IHIET-AI 2024): Artificial Intelligence and Future Applications, volume 120 of *AHFE Open Access*, AHFE International, USA, 2024. URL: http://doi.org/10.54941/ahfe1004576. doi:10.54941/ahfe1004576.

[17] Y. Fujimoto, Chatar: Conversation support using large language model and augmented reality, arXiv preprint arXiv:2506.16008 (2025).

[18] M. Borowski, P. W. S. Butcher, J. B. Kristensen, J. O. Petersen, P. D. Ritsos, C. N. Klokmose, N. Elmqvist, Dashspace: A live collaborative platform for immersive and ubiquitous analytics, IEEE Transactions on Visualization and Computer Graphics 31 (2025) 7034–7047. doi:10.1109/TVCG.2025.3537679.

[19] A. Mayer, K. Kastner, J. Reichwald, J.-R. Chardonnet, J. Ovtcharova, Comparing avatar and face-to-face collaboration in vr education: Concept and preliminary insights, in: 2023 IEEE 2nd German Education Conference (GECon), 2023, pp. 1–5. doi:10.1109/GECon58119.2023.10295113.

[20] G. Vallasciani, L. Stacchio, P. Cascarano, G. Marfia, Creaixr: Fostering creativity with generative ai in xr environments, in: 2024 IEEE International Conference on Metaverse Computing, Networking, and Applications (MetaCom), 2024, pp. 1–8. doi:10.1109/MetaCom62920.2024.00034.

[21] F. Rahimi, A. Sadeghi-Niaraki, S.-M. Choi, Generative ai meets virtual reality: A comprehensive survey on applications, challenges, and future direction, IEEE Access 13 (2025) 94893–94909. doi:10.1109/ACCESS.2025.3574779.

[22] B. Marques, S. Silva, C. Ferreira, S. Oliveira, A. Santos, P. Dias, B. S. Santos, Mixed reality collaboration: How expert representations shape user experiences, IEEE Computer Graphics and Applications 45 (2025) 143–151. doi:10.1109/MCG.2025.3537316.

[23] H. Moon, S. Yu, J. Y. Lee, Smart task assistance through deep learning-based visual guidance in asymmetric xr remote collaboration, IEEE Access 12 (2024) 126899–126914. doi:10.1109/ACCESS.2024.3455014.

[24] S. Srinidhi, E. Lu, A. Rowe, Xair: An xr platform that integrates large language models with the physical world, in: 2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2024, pp. 759–767. doi:10.1109/ISMAR62088.2024.00091.

[25] Y. Kim, Z. Aamir, M. Singh, S. Boorboor, K. Mueller, A. E. Kaufman, Explainable xr: Understanding user behaviors of xr environments using llm-assisted analytics framework, IEEE Transactions on Visualization and Computer Graphics 31 (2025) 2756–2766. URL: https://doi.org/10.1109/TVCG.2025.3549537. doi:10.1109/TVCG.2025.3549537.

[26] S. G. Hart, Nasa-task load index (nasa-tlx); 20 years later, Proceedings of the Human Factors and Ergonomics Society Annual Meeting 50 (2006) 904–908. URL: https://doi.org/10.1177/154193120605000909. doi:10.1177/154193120605000909. arXiv:https://doi.org/10.1177/154193120605000909.

[27] J.-Y. Jian, A. M. Bisantz, C. G. Drury, Foundations for an empirically determined scale of trust in automated systems, International Journal of Cognitive Ergonomics 4 (2000) 53–71. URL: https://doi.org/10.1207/S15327566IJCE0401_04. doi:10.1207/S15327566IJCE0401_04.