

Toward Governing Perception in Safety-Critical Mediated Reality on the Move

Pascal Jansen

pascal.jansen@uni-ulm.de

Institute of Media Informatics, Ulm University
Ulm, Germany

Abstract

Wearable Augmented Reality (AR) is increasingly deployed in on-the-move contexts such as automated driving, cycling, and pedestrian navigation. To date, most systems rely on additive overlays that highlight hazards, intentions, or predictions without altering the scene itself. However, advances in head-mounted displays and computer vision now enable Diminished and Modified Reality techniques that suppress, transform, or substitute scene elements. These capabilities conceptually extend AR into Mediated Reality (MR), shifting the design space from “what to add” to “what is perceptually available.” Because such mediation reshapes the evidential basis for situation awareness and trust calibration, it raises novel interaction challenges. This position paper argues that MR on the move must become governable, as users need mechanisms to configure, inspect, and understand mediation without compromising safety. Additionally, this position paper outlines design challenges related to governance granularity, epistemic signaling, and accountability, and frames MR on the move as a research agenda for governable perceptual mediation in dynamic, safety-critical environments.

CCS Concepts

• **Human-centered computing** → **Human computer interaction (HCI)**; *Interaction paradigms*; Ubiquitous and mobile computing; Mixed / augmented reality.

Keywords

augmented reality, mediated reality, situation awareness, trust calibration, mobility

ACM Reference Format:

Pascal Jansen. 2026. Toward Governing Perception in Safety-Critical Mediated Reality on the Move. In *Proceedings of CHI 2026 Workshop on Next Steps for Augmented Reality On-the-Move: Challenges & Opportunities*. ACM, New York, NY, USA, 3 pages.

1 Introduction

Wearable Augmented Reality (AR) systems are increasingly used in on-the-move contexts such as pedestrian navigation, cycling, and automated vehicles [17]. In these contexts, users move through environments that change continuously [6]. For example, vehicles

approach at different speeds, pedestrians hesitate before crossing, cyclists appear from behind obstacles, and traffic density varies. Under such conditions, users must quickly judge what is present, what others intend to do, and what will likely happen next. These judgments directly affect safety [4].

To support such decisions, much automotive AR research has focused on additive overlays. These overlays highlight detected objects, visualize predicted trajectories, or display maneuver intentions to increase transparency (e.g., [2, 8, 9]). Similarly, AR systems for vulnerable road users mainly rely on additive visual signals layered onto the environment without altering it [17, 18].

However, recent advances in computer vision models (e.g., object detection and inpainting [10]) have extended the capabilities of Head-Mounted Displays (HMDs) beyond AR. Following the definition of Mediated Reality (MR) by Mann [15], HMD systems can now additionally suppress or remove parts of the scene (Diminished Reality, DR) or transform and substitute them (Modified Reality, ModR) [1, 14]. The potential applications are manifold. For example, roadside clutter (e.g., advertisements) could be removed, intersections could be visually simplified (e.g., by removing non-relevant vehicles for drivers or by simplifying visually cluttered crossings for pedestrians), or predicted trajectories could be emphasized for specific road users (e.g., pedestrians at road crossings). Jansen et al. [5] recently demonstrated that AR, DR, and ModR can be integrated as Automotive Mediated Reality (AMR) and deployed in real traffic within a moving vehicle. Expert evaluations indicated that AMR effects were interpreted not merely as visual refinements, but as interventions that influenced judgments of environmental complexity, risk, and system capability [5].

These observations suggest that DR and ModR are not only technical extensions, but interaction-level variables that reshape how users interpret the scene. While AR primarily influences saliency, DR and ModR influence perceptual availability, which is the evidential basis from which users construct SA [4]. SA unfolds in three stages: perceiving relevant elements (Level 1), comprehending their meaning (Level 2), and projecting future states (Level 3). For example, if a DR application removes an element, it never enters SA Level 1. If predicted trajectories are emphasized without distinction from sensor observations, users may interpret forecasts as certain rather than probabilistic.

As mobility contexts involve automated decision-making, these shifts in SA extend to users’ appraisals of the traffic scene and road users’ capabilities (e.g., automated vehicles). For instance, to form trust in automation, users evaluate whether the vehicle reliably detects relevant actors, accurately predicts their motion, and plans appropriate responses [12]. Trust calibration depends on aligning reliance with actual capability under uncertainty [12]. Consequently,

Position paper presented at Next Steps for Augmented Reality On-the-Move: Challenges & Opportunities - Workshop at CHI '26.



This work is licensed under a Creative Commons Attribution 4.0 International License. *CHI 2026 Workshop on Next Steps for Augmented Reality On-the-Move: Challenges & Opportunities, Barcelona, Spain*

© 2026 Copyright held by the owner/author(s).

when MR interfaces mediate environmental complexity (e.g., removing road users), users may misjudge what the automation can detect, forecast, or handle, leading to non-calibrated trust [9].

This position paper, therefore, argues that AR in on-the-move contexts will evolve into MR interfaces, and must incorporate mechanisms that make perceptual mediation *governable*. Rather than treating scene element suppression and transformation as hidden rendering choices, interfaces can expose the epistemic status of scene elements, for example, by differentiating directly sensed road users, algorithmic interpretations, predicted trajectories, and intentionally filtered content through visual encodings or interactive inspection techniques. Such mechanisms enable users not only to see what is emphasized, altered, or removed, but also to understand how the scene has been constructed and, where appropriate, adjust mediation levels towards forming appropriate SA and appraisal. Insights from this setting may inform the broader design of MR systems in other dynamic, safety-relevant mobility domains, such as pedestrian navigation and micromobility.

2 Design Challenges for Governable Mediated Reality on the Move

Governance of MR in on-the-move contexts concerns interaction design choices that define (i) when users can influence mediation, (ii) what they can control, (iii) how epistemic status is communicated, and (iv) how responsibility is attributed when mediation changes what was perceivable in the first place.

2.1 When and How Can Users Govern Mediation?

Continuous in-situ governance of MR during driving, walking, or cycling is typically infeasible because interaction is mechanically constrained, and additional controls risk increasing workload or diverting attention from primary safety tasks. This contrasts with AR systems in stationary rooms [14], where exploratory inspection is often feasible.

An interaction model with three loci of control might address when and how governance can occur. First, **pre-ride configuration**: users select high-level mediation profiles (e.g., “reduced clutter,” “highlight vulnerable road users,” “conservative uncertainty display”) via low-demand interfaces such as a companion app, a dashboard menu, or a voice-based setup dialogue before entering motion. Interaction modalities may include touch- or gesture-based selection and conversational voice interaction (e.g., “prioritize pedestrians today”). Second, **minimal in-situ override**: during motion, only lightweight interventions are permitted, such as a single voice command (“show full scene”), a gaze-triggered reveal of a suppressed object, or a context-attached affordance directly anchored to the mediated element (e.g., a subtle icon next to an in-painted vehicle that expands on fixation). These mechanisms might constrain interaction cost while preserving agency. Third, **post-ride inspection**: users review a structured mediation summary that explains when and why transformations occurred.

This temporal separation aligns with prior work on inspectability and reviewability in adaptive systems [20], yet mobility particularly introduces a binding problem: users must connect a post-hoc explanation to a moment experienced while in motion and under

limited attention. To support this, mediation events could be logged and exposed through event-based replay mechanisms that bind each intervention to spatiotemporal anchors (timestamp, location, maneuver context; see [6]). The interface could provide synchronized dual views (mediated vs. unmediated, or delta visualization), annotated with triggers and system confidence levels. Hence, (i) at what moments (e.g., low traffic density, low speed, straight-road segments) is it safe and appropriate to surface even lightweight mediation controls without degrading performance, and (ii) which replay and annotation formats best support recognition and understanding of past mediation events that were originally experienced under attentional load?

2.2 Granularity of Mediation Governance

To govern AR and visualization systems, users often select between presets (e.g., overview vs. detail mode) or adjust individual visualization parameters [21]. In mobility, however, granularity is coupled with safety because users cannot meaningfully “debug” the world while moving. Moreover, non-expert users may not understand the MR system’s internal models or training data [8]. Consequently, overly fine control risks misconfiguration, while overly coarse control risks epistemic opacity.

Coarse control (e.g., “simplified environment” vs. “full detail”) could be used. For instance, a “simplified environment” mode might suppress parked vehicles, reduce texture and facade detail, and collapse distant road users into abstract glyphs, whereas “full detail” would render the scene without suppression or abstraction. These modes might be easier to communicate, but they can obscure which aspects are being altered, such as object density (number of rendered agents) or semantic emphasis (which object classes are highlighted). Users may therefore not know whether, for example, background clutter reduction can also affect peripheral hazard visibility.

Fine-grained control (e.g., “suppress background clutter only,” “de-emphasize predicted trajectories,” “highlight vulnerable road users”) may increase agency and allow alignment with individual risk preferences, but it risks complexity and misconfiguration [7]. A mixed-initiative approach could allow the system to propose adjustments based on context (e.g., high traffic density or reduced visibility), with the user confirming or overriding the proposal. However, this raises governance demands, as proposals should be intelligible (e.g., “why is clutter suppression suggested now?”), reversible, and attributable to concrete triggers (e.g., “activated due to high pedestrian density”) to achieve user acceptance [13, 16].

This framing yields two research questions: (i) how does control granularity affect SA and appraisal (e.g., trust, perceived safety) when mediation changes perceptual availability under time pressure? (ii) Which UI mechanisms, such as delta visualizations, explanations, or context-attached annotations, make “what exactly changed” sufficiently legible that coarse controls do not become epistemically opaque?

2.3 Epistemic Signaling of Mediated Reality in On-the-Move Contexts

Signaling the epistemic status of MR must be compatible with motion constraints. Persistent legends or dense encodings may clutter the scene and compete with hazard perception. Prior work on

uncertainty visualization and explainable interfaces has proposed various techniques, such as persistent vs. on-demand views and progressive disclosure (e.g., see [8, 11]).

These point to interaction techniques that minimize sustained visual load, such as gaze-triggered inspection and transient overlays on demand. However, these techniques may fail in some scenarios. For example, on-demand inspection assumes users detect uncertainty early enough to request clarification, which may not hold in fast-moving scenes, while persistent encodings can produce visual noise and inattention blindness [19]. This motivates the question: how can epistemic signaling remain legible and behaviorally effective (i.e., actually change interpretation and action) without degrading scene clarity in dynamic, safety-critical environments?

2.4 Responsibility and Attribution of Mediated Reality in Automated Mobility

In automated mobility, MR that suppresses or transforms environmental cues changes the SA available to human intervention. This can also affect how responsibility is attributed between the user and the automation. Prior work focuses on communicating internal model reasoning and decision processes, typically assuming a stable perceptual representation of the environment [3, 8, 9]. Governable MR adds a distinct layer because the interface can visually alter the world itself. This is critical in mobility because small perceptual changes might alter users' hazard detection and emergency response times.

MR systems could log suppression/transformation events, associated triggers, confidence levels, and user-facing signals displayed at the time. Users (and potentially third parties) should be able to retrospectively inspect what was mediated during an incident, ideally via a delta that clarifies what was removed, modified, or emphasized. This yields a research question: what logging and replay abstractions provide actionable accountability without overwhelming users, while remaining faithful enough to support post-incident interpretation of whether sufficient perceptual access existed?

3 Conclusion

This position paper argues that increasing advances in HMD technology and computer vision models (e.g., DR and ModR techniques) will conceptually extend AR in on-the-move contexts into MR (e.g., see AMR [5]). While AR overlays remain effective for highlighting hazards and intent, they can increase clutter and compete for limited user attention. Hence, MR complements AR by enabling selective suppression, transformation, or abstraction of elements in safety-critical contexts. This could enable semantically richer applications that can improve users' SA and, thereby, positively influence traffic safety and user appraisal (e.g., trust calibration) in on-the-move contexts such as pedestrian navigation and micromobility.

However, MR introduces challenges for interaction, as users interact with a world representation constructed by the MR system. Governability, intelligibility, and auditability, therefore, can become central design requirements toward MR technology acceptance, as users must be able to understand what was altered, why it was altered, and how to influence these processes. This reframing positions MR in on-the-move contexts as a broader research agenda on governable perceptual mediation in dynamic environments.

References

- [1] Yi Fei Cheng, Hang Yin, Yukang Yan, Jan Gugenheimer, and David Lindlbauer. 2022. Towards understanding diminished reality. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–16.
- [2] Mark Colley, Benjamin Eder, Jan Ole Rixen, and Enrico Rukzio. 2021. Effects of semantic segmentation visualization on trust, situation awareness, and cognitive load in highly automated vehicles. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–11.
- [3] Upol Ehsan, Q Vera Liao, Michael Muller, Mark O Riedl, and Justin D Weisz. 2021. Expanding explainability: Towards social transparency in ai systems. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–19.
- [4] Mica R. Endsley. 1995. Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors* 37, 1 (1995), 32–64. doi:10.1518/001872095779049543
- [5] Pascal Jansen, Julian Britten, Mark Colley, Markus Sasalovici, and Enrico Rukzio. 2026. MIRAGE: Enabling Real-Time Automotive Mediated Reality. *arXiv preprint arXiv:2601.19385* (2026).
- [6] Pascal Jansen, Julian Britten, Alexander Häusele, Thilo Segsneider, Mark Colley, and Enrico Rukzio. 2023. Autovis: Enabling mixed-immersive analysis of automotive user interface interaction studies. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–23.
- [7] Pascal Jansen, Mark Colley, Svenja Krauß, Daniel Hirschle, and Enrico Rukzio. 2025. Opticarvis: Improving automated vehicle functionality visualizations using bayesian optimization to enhance user experience. In *Proceedings of the 2025 CHI conference on human factors in computing systems*. 1–23.
- [8] Pascal Jansen, Mark Colley, Tim Pfeifer, and Enrico Rukzio. 2024. Visualizing imperfect situation detection and prediction in automated vehicles: Understanding users' perceptions via user-chosen scenarios. *Transportation Research Part F: Traffic Psychology and Behaviour* 104 (2024), 88–108.
- [9] Pascal Jansen, Mark Colley, Max Rädler, Jonas Schwedler, and Enrico Rukzio. 2025. Longitudinal effects of visualizing uncertainty of situation detection and prediction of automated vehicles on user perceptions. *Transportation Research Part F: Traffic Psychology and Behaviour* 113 (2025), 536–553.
- [10] Mohamed Kari, Tobias Grosse-Puppenthal, Luis Falconeri Coelho, Andreas Rene Fender, David Bethge, Reinhard Schütte, and Christian Holz. 2021. TransforMR: Pose-Aware Object Substitution for Composing Alternate Mixed Realities.. In *ISMAR*. 69–79.
- [11] Matthew Kay, Tara Kola, Jessica R Hullman, and Sean A Munson. 2016. When (ish) is my bus? user-centered visualizations of uncertainty in everyday, mobile predictive systems. In *Proceedings of the 2016 chi conference on human factors in computing systems*. 5092–5103.
- [12] John D. Lee and Katrina A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors* 46, 1 (2004), 50–80. doi:10.1518/hfes.46.1.50.30392
- [13] Q Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: informing design practices for explainable AI user experiences. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–15.
- [14] David Lindlbauer and Andy D Wilson. 2018. Remixed reality: Manipulating space and time in augmented reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [15] Steve Mann. 2002. Mediated reality with implementations for everyday life. *Presence Connect* 1 (2002), 2002.
- [16] Forough Poursabzi-Sangdeh, Daniel G Goldstein, Jake M Hofman, Jennifer Wortman Vaughan, and Hanna Wallach. 2021. Manipulating and measuring model interpretability. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–52.
- [17] Helen Stefanidi, Markus Tatzgern, and Alexander Meschtscherjakov. 2024. Augmented reality on the move: A systematic literature review for vulnerable road users. *Proceedings of the ACM on Human-Computer Interaction* 8, MHCI (2024), 1–30.
- [18] Wilbert Tabone, Riender Happee, Jorge García, Yee Mun Lee, Maria Luce Lupetti, Natasha Merat, and Joost de Winter. 2023. Augmented reality interfaces for pedestrian-vehicle interactions: An online study. *Transportation research part F: traffic psychology and behaviour* 94 (2023), 170–189.
- [19] Christopher D Wickens. 2008. Multiple resources and mental workload. *Human factors* 50, 3 (2008), 449–455.
- [20] Xuhai Xu, Anna Yu, Tanya R Jonker, Kashyap Todi, Feiyu Lu, Xun Qian, João Marcelo Evangelista Belo, Tianyi Wang, Michelle Li, Aran Mun, 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*. 1–30.
- [21] Stefanie Zollmann, Tobias Langlotz, Raphael Grasset, Wei Hong Lo, Shohei Mori, and Holger Regenbrecht. 2020. Visualization techniques in augmented reality: A taxonomy, methods and patterns. *IEEE transactions on visualization and computer graphics* 27, 9 (2020), 3808–3825.