Perceptual Scaffolding for Medical Triage in Mission-Critical XR: Design Materials for Multi-Agent AI*

Jihyun Kim[1], Leah Tellier[1,†], Kara Sealock[1,†] and Junho Park[1,*]

[1]University of Calgary, Calgary, Alberta, Canada

Abstract

Mission-critical XR systems increasingly integrate AI to support high-stakes decision-making, yet existing approaches prioritize recommendation accuracy while paying limited attention to how recommendations are perceptually structured under stress. Using medical triage as an analytically revealing case, this position paper reframes multi-agent AI recommendations as a design materials problem. We define perceptual scaffolding as the deliberate structuring of visual, temporal, spatial, and embodied design materials to guide attention, comparison, and judgment under constraint and uncertainty. When multiple AI judgments are presented in parallel, design choices such as color, motion, persistence, and spatial arrangement shape reasoning strategies, trust calibration, and decision resilience. We identify key material-level tensions, including differentiating reaction and deliberation, representing disagreement, organizing judgment flow over time, regulating attention through persistence, and balancing spatial meaning with embodied cost. By foregrounding design materials as the unit of analysis, this paper offers a shared framing for resilient and explainable XR assistance.

Keywords

extended reality (XR), perceptual scaffolding, XR Interface Design, emergency triage, multi-agent AI, human–AI interaction, XR decision support

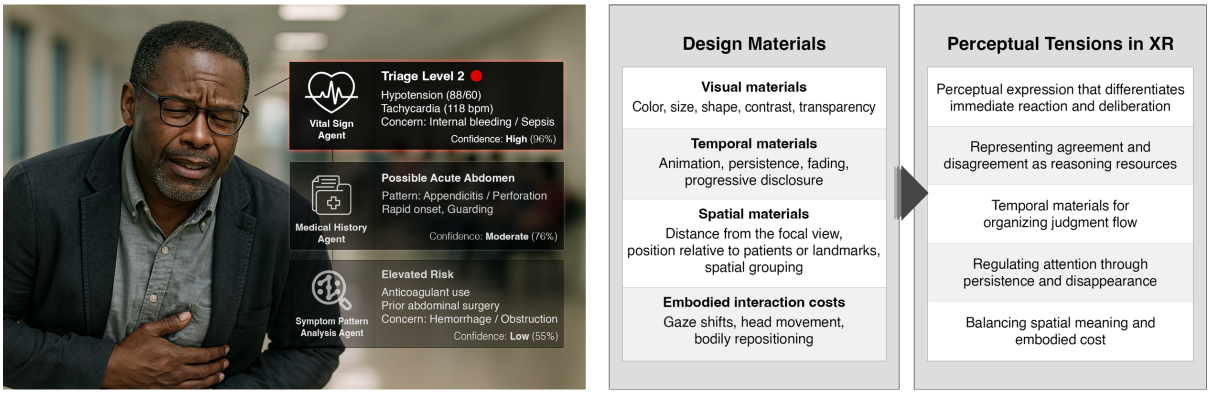

Figure 1: Conceptual Overview: Design Materials and Perceptual Tensions in Multi-Agent XR Triage

1. Motivation

Medical triage is a critical clinical decision-making process in which patients' acuity is assessed to determine treatment priority and allocate limited resources [1, 2]. Although standardized protocols exist, real-world triage rarely relies on rules alone [3]. Clinicians must integrate multiple cues under uncertainty and time pressure, often drawing on experience and intuition. This integration can be

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

*Corresponding author.

†These authors contributed equally.

jihyun.kim3@ucalgary.ca (J. Kim); leah.tellier1@ucalgary.ca (L. Tellier); klsealoc@ucalgary.ca (K. Sealock); junho.park@ucalgary.ca (J. Park)

https://wpsites.ucalgary.ca/hero-lab/author/junho-park/ (J. Park)

0009-0008-3756-8157 (J. Kim); 0000-0001-6436-2009 (L. Tellier); 0000-0002-2145-8534 (K. Sealock); 0000-0001-6348-3765 (J. Park)

cognitively demanding even when relevant information is available, and incorrect judgments can directly compromise patient safety [4, 5, 6]. This suggests that triage is not only a question of whether sufficient information is provided, but also a design problem in which the same information can lead to different decision processes and outcomes depending on how it is perceived and organized into action. To alleviate the inherent cognitive demands of triage, recent work has explored digital technologies, including Extended Reality (XR) and Artificial Intelligence (AI), as supports for triage in time-critical clinical settings [7, 8]. However, recent scoping evidence suggests that these technologies have been evaluated predominantly through outcome-oriented metrics. As a result, while system performance has been well documented, comparatively less attention has been given to how clinicians integrate information into a coherent and temporally unfolding reasoning process in practice [9]. Furthermore, many AI-based systems are designed to produce a single "optimal" recommendation, even when multiple models or agents are involved [10, 11]. Although this approach supports efficient decision-making, it can be problematic in mission-critical settings. A single recommendation may strongly channel attention toward one decision path, obscure uncertainty, and encourage rapid acceptance or post hoc rationalization [12, 13]. This creates additional burden when clinicians must calibrate their reliance on AI under high-stakes conditions. These limitations arise less from the correctness of recommendations than from how they are presented.

These risks are amplified when time and peer support are limited, reducing opportunities for cross-checking and allowing a single AI output to effectively shape the decision itself. Rural emergency departments illustrate this dynamic, as triage nurses often work with limited staffing and heterogeneous expertise, making reflective comparison more difficult [14]. Motivated by this setting as an analytically revealing case rather than a narrow application domain, we are designing an AR-based decision support prototype for triage nurses working in rural emergency departments. The system employs multiple AI agents that generate distinct recommendations from different informational perspectives (e.g., vital signs, medical history, and prior disease progression patterns from similar emergency visits). By exposing these perspectives side by side, the system supports comparison and interpretation while keeping clinicians as the final decision authority, promoting resilient decision-making even when uncertainty or system limitations are visible.

However, implementing multi-agent recommendations in XR is not merely a matter of deciding what information to present. Design materials such as color, size, motion, persistence, and spatial arrangement can substantially influence attention allocation and the flow of reasoning, even when the underlying recommendation content remains unchanged. The issue, therefore, is not the quantity of information, but how that information is perceptually organized [15]. In this paper, we define perceptual scaffolding as the deliberate structuring of perceptual cues to support attention, comparison, and judgment under conditions of constraint and uncertainty. Unlike existing approaches such as ecological interface design or distributed cognition, which primarily analyze how information representations reflect underlying system constraints, perceptual scaffolding focuses on how perceptual design materials organize human judgment when multiple AI-generated perspectives are presented simultaneously [16, 17]. This perspective shifts the design problem from representing system information to structuring how clinicians reason across multiple algorithmic judgments.

This position paper therefore asks: **How should multi-agent recommendations be structured in mission-critical XR environments to scaffold human judgment under stress?** In this framing, trust calibration, decision resilience, and situated explainability are understood as effects emerging from perceptual organization, not as separate theoretical programs to be independently resolved. We aim to offer a shared design framing that allows researchers and practitioners to compare and reflect on design choices shaping human judgment under stress.

2. Design Materials as the Unit of Discussion

This position paper does not focus on algorithms, model architectures, or the internal logic by which recommendations are generated. Instead, it focuses on how multiple AI-generated recommendations are experienced through perceptual and spatial form. This perspective assumes that even identical recommendations can lead to different attentional patterns, interpretations, and judgment strategies depending on how they are rendered [17].

In XR environments, this issue becomes particularly important. XR is not simply a display medium; recommendations persist spatially, incur embodied costs such as gaze shifts and movement, and support heads-up, at-a-glance perception [18, 19, 20, 21]. Consequently, designing multi-agent recommendations in XR differs fundamentally from designing conventional dashboards. Beyond informational accuracy or quantity, designers must consider the perceptual and spatial conditions under which recommendations are encountered.

In this paper, multi-agent recommendations refer to independent judgment units generated in parallel by AI agents with distinct reasoning perspectives. Each recommendation may convey different levels of importance, uncertainty, and trust cues. Because these judgments are perceived simultaneously in XR, clinicians must compare and interpret them directly. Presenting such recommendations therefore inevitably involves decisions at the level of design materials.

Design materials are treated as analytical dimensions that shape how perception, timing, spatial interpretation, and embodied interaction unfold during judgment. Drawing inspiration from human factors research on attentional resource structures, particularly Multiple Resource Theory [22], we organize design materials into four analytical lenses: visual, temporal, spatial, and embodied.

- Visual materials: color, size, shape, contrast, transparency
- Temporal materials: animation, persistence, fading, progressive disclosure
- Spatial materials: distance from the focal view, position relative to patients or landmarks, spatial grouping
- Embodied interaction costs: gaze shifts, head movement, bodily repositioning

These lenses are not mutually exclusive. Many design materials operate across multiple perceptual channels—for example, fading involves both visual and temporal modulation, while spatial grouping relies on visual cues to express spatial relations. The purpose of this framing is therefore not strict classification, but to reveal how materials jointly structure attention, reasoning, and interaction in real-time decision making.

No design material should be assumed inherently superior. Instead, this perspective highlights the tensions and trade-offs different materials introduce for attention, reasoning, and trust in mission-critical contexts. Rather than prescribing solutions, the framework offers a shared way to analyze how XR interfaces structure human judgment in practice.

3. Position: From Information Selection to Perceptual Scaffolding

This position argues that design challenges in mission-critical XR environments extend beyond information selection to how information is perceived, organized, and used within human judgment processes. In multi-agent systems, the central issue is not the quantity of information but how multiple judgments and uncertainties are structured through design materials to support human reasoning. Interfaces should therefore be understood not as neutral delivery channels, but as perceptual scaffolds that structure how clinicians compare, weigh, and integrate multiple AI-generated judgments in real time. This perspective differs from traditional interface approaches that focus primarily on selecting or prioritizing information for display. In multi-agent AI environments, the design challenge shifts from deciding what information to present toward structuring how multiple algorithmic judgments become perceptually available for comparison and reasoning. From this perspective, XR interfaces actively shape trust calibration, resilience under breakdown, and situated explainability in high-stress, embodied work.

4. Design Material Tensions in Multi-Recommendation XR Interfaces

Building on this perspective, this section outlines key material-level tensions that arise when implementing multi-agent recommendations in mission-critical XR environments. The focus is not on algorithms, but on how perceptual and spatial conditions influence reasoning, trust calibration, and decision resilience under time pressure.

Multi-recommendation XR interfaces require decisions about color, size, contrast, motion, persistence, and spatial layout. These choices directly shape attention and judgment flow. Drawing on cognitive psychology and human factors, this paper abstracts recurring material-level tensions as a conceptual framework.

Perceptual Expression That Differentiates Immediate Reaction and Deliberation One key challenge in multi-recommendation XR interfaces is differentiating between representations that prompt immediate reaction and those that invite deliberate consideration. When multiple AI judgments are presented simultaneously, it is easy for all recommendations to appear equally urgent—an effect driven not by information quantity, but by perceptual emphasis. Materials such as color contrast, size variation, and animation support rapid detection, but excessive emphasis can lead to attentional tunneling and cognitive overload. The design question, therefore, is not what to emphasize, but which mode of judgment a representation is intended to evoke.

Representing Agreement and Disagreement as Reasoning Resources In multi-agent environments, agreement and disagreement among recommendations are inevitable. The key design question is not whether to hide these differences, but how to make their relationships interpretable through material structure. Visual similarity, shape alignment, and spatial grouping can structure relationships among recommendations, allowing disagreement to function as a resource for reflective reasoning. Conversely, unstructured exposure of disagreement can result in confusion and decision paralysis.

Temporal Materials for Organizing Judgment Flow Triage judgment unfolds over time, yet many interfaces assume static visual hierarchies. Multi-recommendation XR design requires decisions about whether recommendations should be presented simultaneously, sequentially, or reorganized dynamically in response to evolving conditions. While animation and progressive disclosure can help organize reasoning flow, in high-stress environments dynamic change itself can impose additional cognitive burden. This is not merely a problem of information update, but of how to maintain stability in judgment flow.

Regulating Attention Through Persistence and Disappearance Decisions about whether recommendations should remain persistently visible or appear only when needed are closely tied to attentional regulation. Persistent representations can reduce memory burden but may induce alarm fatigue, while transient displays preserve visual simplicity at the risk of missing critical information. This challenge should be understood not as a matter of informational importance, but as a design problem of when and how attention is summoned.

Balancing Spatial Meaning and Embodied Cost XR enables spatial placement of information, allowing location itself to carry meaning. Recommendations may be positioned relative to patients, clinicians' bodies, or environmental reference points, supporting embodied reasoning. However, spatial meaning incurs embodied costs in the form of gaze shifts and bodily movement. In mission-critical environments, these costs can directly undermine decision efficiency. Designers must therefore balance meaningful spatialization against minimal physical burden.

5. Summary and Workshop Contribution

This position paper reframes multi-agent recommendation in mission-critical XR as a design materials problem. We argue that how multiple AI judgments are perceptually structured is central to trust calibration, decision resilience, and at-a-glance explainability under stress. By foregrounding design materials as the unit of discussion, this work offers a shared framing to support cross-domain workshop discussion on mission-critical XR assistance.

Declaration on Generative AI

During the preparation of this work, the author used ChatGPT (OpenAI) for grammar and spelling checks and minor language improvements. ChatGPT was also used to generate the background patient illustration in Figure 1. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the final manuscript.

References

[1] K. V. Iserson, J. C. Moskop, Triage in Medicine, Part I: Concept, History, and Types, Annals of Emergency Medicine 49 (2007) 275–281. doi:10.1016/j.annemergmed.2006.05.019.

[2] C. C. Yancey, M. C. O'Rourke, Emergency department triage, 2023. URL: https://www.ncbi.nlm.nih.gov/books/NBK557583/, in *StatPearls*. Updated August 28, 2023.

[3] M. Fry, G. Burr, Current triage practice and influences affecting clinical decision-making in emergency departments in NSW, Australia, Accident and Emergency Nursing 9 (2001) 227–234. doi:https://doi.org/10.1054/aaen.2001.0268.

[4] A. Surendran, L. Beccaria, P. McIlveen, Cognitive Burden in Emergency Care: A Qualitative Exploratory, Journal of Nursing Management 2025 (2025) 6332434. doi:10.1155/jonm/6332434.

[5] A. H. Soola, S. Mehri, I. Azizpour, Evaluation of the factors affecting triage decision-making among emergency department nurses and emergency medical technicians in Iran: a study based on Benner's theory, BMC Emergency Medicine 22 (2022) 174. doi:10.1186/s12873-022-00729-y.

[6] J. Park, M. Zahabi, Cognitive Workload Assessment of Prosthetic Devices: A Review of Literature and Meta-Analysis, IEEE Transactions on Human-Machine Systems 52 (2022) 181–195. doi:10.1109/THMS.2022.3143998.

[7] S. Levin, M. Toerper, E. Hamrock, J. S. Hinson, S. Barnes, H. Gardner, A. Dugas, B. Linton, T. Kirsch, G. Kelen, Machine-Learning-Based Electronic Triage More Accurately Differentiates Patients With Respect to Clinical Outcomes Compared With the Emergency Severity Index, Annals of Emergency Medicine 71 (2018) 565–574.e2. doi:10.1016/j.annemergmed.2017.08.005.

[8] C. R. Nelson, J. L. Gabbard, J. B. Moats, R. K. Mehta, User-Centered Design and Evaluation of ARTTS: an Augmented Reality Triage Tool Suite for Mass Casualty Incidents, in: 2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2022, pp. 336–345. doi:10.1109/ISMAR55827.2022.00049.

[9] J. Kim, J. Park, Guidelines for Developing Digital Triage Tools to Mitigate Workload Challenges: A Scoping Review, Applied Ergonomics 134 (2026) 104742. doi:10.1016/j.apergo.2026.104742.

[10] Z. Araouchi, M. Adda, TriageIntelli: AI-Assisted Multimodal Triage System for Health Centers, Procedia Computer Science 251 (2024) 430–437. doi:10.1016/j.procs.2024.11.130.

[11] A. Choi, K. Lee, H. Hyun, K. J. Kim, B. Ahn, K. H. Lee, S. Hahn, S. Y. Choi, J. H. Kim, A novel deep learning algorithm for real-time prediction of clinical deterioration in the emergency department for a multimodal clinical decision support system, Scientific Reports 14 (2024) 30116. doi:10.1038/s41598-024-80268-7.

[12] K. Goddard, A. Roudsari, J. C. Wyatt, Automation bias: a systematic review of frequency, effect mediators, and mitigators, Journal of the American Medical Informatics Association : JAMIA 19 (2012) 121–127. doi:10.1136/amiajnl-2011-000089.

[13] E. Efendic, P. P. F. M. Van de Calseyde, S. Bahnik, M. A. Vranka, Taking Algorithmic (Vs. Human) Advice Reveals Different Goals to Others, International Journal of Human–Computer Interaction 40 (2024) 45–54. doi:10.1080/10447318.2023.2210886, _eprint: https://doi.org/10.1080/10447318.2023.2210886.

[14] B. Kennedy, K. Curtis, S. Kourouche, L. Casey, D. Hughes, A. McCloughen, Development of an evidence-based strategy to implement the HIRAID® emergency nursing framework in rural emergency departments using behaviour change theory, International Emergency Nursing 77 (2024) 101538. doi:https://doi.org/10.1016/j.ienj.2024.101538.

[15] N. Zhang, J. Zhang, S. Jiang, W. Ge, The Effects of Layout Order on Interface Complexity: An Eye-Tracking Study for Dashboard Design, Sensors (Basel, Switzerland) 24 (2024) 5966. doi:10.3390/s24185966.
[16] D. Wood, J. S. Bruner, G. Ross, The role of tutoring in problem solving, Journal of Child Psychology and Psychiatry, and Allied Disciplines 17 (1976) 89–100. doi:10.1111/j.1469-7610.1976.tb00381.x.
[17] D. Kirsh, The intelligent use of space, Artificial Intelligence 73 (1995) 31–68. doi:10.1016/0004-3702(94)00017-U.
[18] J. Wither, S. DiVerdi, T. Höllerer, Annotation in outdoor augmented reality, Computers & Graphics 33 (2009) 679–689. doi:10.1016/j.cag.2009.06.001.
[19] L. Sidenmark, H. Gellersen, Eye, Head and Torso Coordination During Gaze Shifts in Virtual Reality, ACM Trans. Comput.-Hum. Interact. 27 (2019) 4:1–4:40. doi:10.1145/3361218.
[20] S. Feiner, B. Macintyre, D. Seligmann, Knowledge-based augmented reality, Commun. ACM 36 (1993) 53–62. doi:10.1145/159544.159587.
[21] J. Jerald, The VR Book: Human-Centered Design for Virtual Reality, Association for Computing Machinery and Morgan & Claypool, 2015.
[22] C. Wickens, Multiple Resources and Mental Workload, Human factors 50 (2008) 449–55. doi:10.1518/001872008X288394.