Advances in Medical Education and Practice

Dovepress
Taylor & Francis Group

Open Access Full Text Article

REVIEW

# Integration of Artificial Intelligence Into Extended Reality Debriefing in Healthcare Simulation: A Narrative Review

Abdullah Saeed Khan[1], Selina Hasan[1], Faisal Wasim Ismail[1,2]

[1]Center for Innovation in Medical Education, Aga Khan University, Karachi, Pakistan; [2]Department of Medicine, Aga Khan University, Karachi, Pakistan

Correspondence: Faisal Wasim Ismail, Center for Innovation in Medical Education, Aga Khan University, Stadium Road, Karachi, 74800, Pakistan, Email faisal.ismail@aku.edu



**Abstract:** Extended reality (XR) is increasingly used in healthcare simulation; however, debriefing, the phase in which performance is translated into learning through guided reflection, faces distinct challenges in immersive, data-rich, and sometimes asynchronous environments. This focused narrative review examined how artificial intelligence (AI) can augment debriefing within XR/virtual reality (VR) simulation. Peer-reviewed literature published between 2015 and 2025 was identified through structured searches of PubMed/MEDLINE and Google Scholar, supplemented by reference screening. Search strategies combined terms related to XR (eg, "extended reality", "virtual reality", "augmented reality", "mixed reality"), simulation and education, and AI-enabled debriefing (eg, "debriefing", "feedback", "learning analytics", "natural language processing", "large language models", "conversational agents"). Studies were included when AI was directly applied to debriefing processes or generated debrief-relevant feedback within XR/VR healthcare simulations; studies focused solely on scenario generation, automated grading without reflective components, or non-XR contexts were excluded. Four recurring AI functions were identified: (1) automated performance analytics that convert XR telemetry (eg, timestamps, trajectories, error logs, and in some systems gaze or physiological data) into structured metrics to support procedural feedback; (2) natural language processing to enable transcription and discourse analysis that can surface communication patterns and candidate moments for team reflection; (3) conversational agents and large language model–enabled systems that scaffold reflective dialogue and summarize performance; and (4) multimodal fusion approaches that integrate action, speech, gaze, and physiological signals to deliver adaptive feedback. The evidence base remains dominated by feasibility studies, pilots, and prototypes, with limited controlled comparisons, psychometric validation, or evidence of sustained behavior change or clinical transfer. A pragmatic near-term approach is a hybrid model in which AI prepares transparent debriefing artefacts, while human facilitators retain responsibility for psychological safety, meaning-making, and high-stakes interpretation. Future research should prioritize validation, bias and privacy safeguards, faculty development, and longitudinal educational outcomes.

**Keywords:** education, medical, virtual reality, augmented reality, debriefing, artificial intelligence, natural language processing

## Introduction

Debriefing, which translates performance into learning through guided reflection and facilitator-led inquiry, is commonly acknowledged as the pedagogical core of simulation-based education (SBE).[1] Experienced facilitators employ organized debriefing methods (such as advocacy-inquiry and plus-delta) in more conventional simulation contexts (manikins, task trainers, and team-based scenarios) to encourage learners to consider their actions, mental processes, and results and to make connections to clinical practice.[2] New opportunities such as reproducible immersive scenarios, comprehensive telemetry (actions, gaze, physiology), and asynchronous access are brought about by the introduction of extended reality (XR), which includes virtual reality (VR), augmented reality (AR), and mixed reality (MR), in healthcare simulation. However, there are also new challenges for debriefing, as the facilitator might not be able to watch every action in real time, learners might interact asynchronously or remotely, and complex data streams could be overwhelming.[3,4]

https://doi.org/10.2147/AMEP.S578857
Received: 3 November 2025
Accepted: 19 February 2026
Published: 20 March 2026

Beyond their technical capabilities, XR environments present unique pedagogical contexts for debriefing. Immersive simulations modify perceptual focus, embodiment, and emotional involvement, thereby generating learning experiences that are spatially dispersed and individually experienced via head-mounted displays, as opposed to being collectively observed within a shared physical setting.[4,5] In traditional mannequin-based simulation, facilitators can directly observe the majority of learner behaviors and team interactions in real time, facilitating debriefing that reconstructs events from a largely shared experiential perspective.[2] Conversely, XR interactions may transpire within partially opaque virtual environments, necessitating that facilitators depend on telemetry, replay functionalities, and learner narrative accounts to establish a shared mental model of events.[1,4] Furthermore, immersive XR experiences have been demonstrated to augment presence and cognitive engagement, potentially intensifying emotional significance and influencing how learners encode and recall events.[5] These characteristics necessitate deliberate strategies during debriefing to scaffold reflection, reconstruct distributed experiences, and uphold psychological safety while navigating high-fidelity performance traces.[2,6]

These differences extend beyond increased data volume. XR simulations frequently enable asynchronous participation, distributed teams, and remote facilitation, thereby altering the social and temporal structure of debriefing compared with traditional co-located simulation.[1,4] The availability of granular replay data, including gaze tracking, trajectory mapping, and timestamped event logs, introduces the possibility of detailed performance analysis.[7] While such precision may enhance objective feedback, it also risks overemphasizing micro-level technical deviations at the expense of broader clinical reasoning, teamwork processes, and affective dimensions of learning.[2] Additionally, embodied interaction metrics, such as movement smoothness or force application, represent novel performance signals that require interpretation within pedagogically meaningful frameworks.[5,8,9] Without careful alignment to debriefing models grounded in reflective inquiry, the presence of objective telemetry may shift debriefing toward technical audit rather than reflective meaning-making.[2,6] Consequently, XR contexts demand thoughtful integration of data-rich analytics within established facilitation frameworks to preserve the core educational intent of debriefing.

XR debriefing can be understood through the lens of existing learning theories that inform simulation-based education. From the perspective of reflective practice, debriefing offers a structured approach to reflection-on-action, enabling learners to examine their reasoning, underlying assumptions, and actions within a psychologically safe environment that promotes professional growth.[10] Reflective dialogue has been identified as the key process through which simulation experiences are transformed into enduring learning.[11] Moreover, experiential learning theory views simulation as the concrete experience phase, with debriefing functioning as the reflective observation and abstract conceptualization stages that support transfer to future clinical practice.[12] Within immersive XR environments, the increased sense of presence and environmental complexity may elevate cognitive demands.[5] Cognitive load theory suggests that excessive sensory input or unstructured performance data can increase extraneous load and impede schema construction.[8,13] Therefore, AI systems that filter telemetry, highlight critical decision points, and scaffold structured inquiry may serve as cognitive supports that enhance, not replace, facilitator-guided reflection.[1,13] Evaluating AI-augmented XR debriefing through these theoretical frameworks ensures that technological advancements align with established principles of reflective depth, cognitive efficiency, and transfer-oriented learning.

Konzelmann et al, in their recent review, highlighted the "integration of artificial intelligence (AI) into XR debriefing" as a potentially valuable, yet insufficiently investigated, domain.[1] Considering the escalating interest in AI-driven simulation, this narrative review specifically examines AI's function in debriefing within XR/VR simulation, deliberately excluding the wider simulation process. This review is based on peer-reviewed publications from 2015 to 2025, sourced through structured searches of PubMed/MEDLINE and Google Scholar, and supplemented by reference screening of relevant articles. Search strategies employed combinations of terms such as "extended reality", "virtual reality", "augmented reality", "mixed reality", "XR", "healthcare simulation", "medical education", "debriefing", "feedback", "artificial intelligence", "natural language processing", "large language models", "conversational agents", and "learning analytics". Studies were deemed eligible if they directly applied artificial intelligence to debriefing processes or provided debriefing-relevant feedback within XR/VR healthcare simulations; conversely, studies concentrating solely on scenario generation, automated grading devoid of reflective elements, or non-XR environments were excluded. Specifically, we examine: the functional roles AI can play in debriefing (feedback generation, spoken/dialogue analysis, conversational agents, multimodal fusion); synthesize empirical evidence from the past decade; discuss pedagogical alignment and technical design considerations; and offer a roadmap for education-developers and researchers.

The primary contribution of this review is a functional taxonomy of AI roles in XR debriefing to guide educators, developers, and researchers. Secondarily, we map the maturity of the empirical evidence and provide pragmatic implementation considerations aligned with debriefing standards. Where possible, findings are interpreted using commonly applied medical education outcomes hierarchies, noting that most existing studies report learner perceptions or performance metrics (lower-level outcomes), with limited evidence of sustained behavior change or patient-level impact.[1,9,14]

## Utility of AI in XR Debriefing

XR simulation platforms produce rich, high-fidelity datasets encompassing timestamps of learner actions, gesture and trajectory logs, audio and speech recordings, gaze tracking, and in some cases, physiological signals.[7] Artificial intelligence (AI) enables the transformation of these raw data into meaningful debriefing artefacts through four primary mechanisms. First, *automated performance metrics and analytics* use AI models to analyze event logs and generate objective performance indicators, such as time-to-task completion, deviation from optimal trajectories, and error counts, thereby facilitating effective, data-driven feedback for procedural skills.[15,16] Second, *Natural Language Processing (NLP) for dialogue analysis* allows AI to transcribe speech, code communication for key themes like decision-making or coordination, detect tone or emotion, and identify critical “turning-point” utterances that can enrich reflective discussions during debriefing.[17,18] Third, *conversational agents, embodied conversational agents (ECAs), and large language models (LLMs)* can emulate virtual facilitators or participants, engaging learners in guided reflective dialogue—such as by asking “What were you thinking when…?”—and providing immediate, context-specific feedback derived from simulation logs.[19–21] Finally, *multimodal fusion* integrates multiple data streams, including learner actions, speech transcripts, gaze, and physiological signals, to deliver adaptive, individualized feedback, such as by highlighting issues in team-role allocation or adjusting the pacing of reflection according to learner needs.[22,23] Collectively, these AI-driven functions directly address persistent debriefing challenges, such as limited faculty availability, variability in facilitation quality, and the demand for objective feedback, while capitalizing on XR’s intrinsic advantages of rich telemetry, asynchronous accessibility, and scalability.[1] Table 1 attempts to summarize the functional roles of AI in XR debriefing keeping in mind objectives, evidence maturity, and limitations.

**Table 1** Functional Roles of AI in XR Debriefing: Objectives, Evidence Maturity, and Key Considerations

| AI Function | Primary Debriefing Objectives Supported | Evidence Maturity | Representative Study Types | Key Considerations and Limitations |
|---|---|---|---|---|
| Automated performance analytics (telemetry-based scoring) | Identification of performance gaps; procedural feedback; deliberate practice planning | Feasibility studies; limited RCTs in procedural VR contexts | Correlation with expert ratings; novice–expert differentiation[5,8,9] | Translation of technical metrics into pedagogically meaningful feedback; limited transfer evidence |
| NLP-driven discourse analysis | Team learning; communication patterns; reflection depth (surface-level) | Pilot and proof-of-concept studies | Automated transcript coding; agreement with human raters[24–27] | Focus on surface metrics; limited emotion detection; accent and bias concerns |
| Conversational agents / LLM-guided reflection | Structured reflection; immediate feedback; rehearsal of communication | Prototype systems; early implementation studies | Virtual AI patient simulators; dialogue scaffolding[19–21] | Limited controlled comparisons; affective nuance limitations; risk of hallucinated feedback |
| Multimodal fusion (action + speech + gaze + physiology) | Adaptive feedback; personalization; cognitive load adjustment | Conceptual and emerging experimental work | Integrative XR-AI frameworks[22,23] | Limited validation; interpretive complexity; data governance challenges |

 https://doi.org/10.2147/AMEP.S578857 

# Current State of Evidence

## NLP and Automated Discourse Analysis

Early work in medical education applied NLP to simulate clinical reasoning and discourse coding, showing feasibility for automated coding of transcripts and reflective statements (Chary et al reviewed NLP in medical education).[17] Trauma simulation research showed relationships between human rater ratings and NLP-derived metrics (eg, amount of advocacy-style utterances) by comparing pre- and post-debrief conversation and quantifying teamwork behaviors.[24,25] NLP has been employed in proof-of-concept research more recently (2024–2025) to extract sentiment patterns, decision points, and the frequency of reflecting utterances from team simulations. These studies have shown agreement with human-rater judgments.[26,27] However, limitations remain: the majority of NLP research concentrates on surface metrics (turn count, keywords) rather than more profound interpretive concepts (emotional understanding, depth of thought).[17,24] Emotion detection and multimodal alignment (eg, speech + gaze + action) are underdeveloped.

## Automated Scoring and Analytics for Procedural Debriefing

XR and VR procedural simulators (eg, for surgery) capture kinematic and timing data; AI anomaly detectors have been used to compute performance dashboards that correlate with expert ratings and distinguish between competence levels.[5,8,9] For example, a haptics-enabled VR drill classified users as novice or expert based on automatic criteria of smoothness and mistakes.[9] However, the majority of these indicators (force profiles and trajectories) are still technical, and they do not always correspond to instructional messages that are pertinent to the debrief (eg, team coordination and decision-making). The major challenge lies in translating lower-level metrics into pedagogically meaningful feedback (eg, "You applied excess force at step X, consider slower hand-off").[5,9]

## Conversational Agents, LLMs, and ECAs for Debrief Dialogue

Emerging prototypes embed LLMs or ECAs in VR simulations to conduct interactive virtual patient encounters and to provide automated debriefing prompts.[19,20] Research shows that these systems have the ability to ask guided reflection questions, role-play difficult conversations before providing feedback, and summarize performance logs.[14] For example, a virtual AI patient simulator (VAPS) powered by LLMs allowed learners to complete a simulated scenario, then receive immediate feedback and reflection scaffolding through a conversational agent.[19] There are not many controlled studies that directly compare AI-only debriefing with hybrid or human facilitator models. Key concerns include AI's lack of affective nuance, contextual sensitivity, and risk of generating misleading feedback if the model is improperly trained or domains are outside it's scope.[3,14,20]

## AI Model Comparison Synthesis

Across studies, three broad debriefing configurations can be identified: AI-only, hybrid (AI-assisted human facilitation), and traditional human-led models. AI-only approaches offer scalability, consistency, and immediate feedback delivery, particularly in asynchronous XR environments; however, empirical support remains largely prototype-based, and concerns persist regarding affective nuance, contextual sensitivity, and the risk of misleading feedback.[14,19–21] Traditional human-led debriefing remains the gold standard for psychological safety, emotional processing, and nuanced meaning-making, yet is resource-intensive and subject to facilitator variability.[2] Hybrid models, where AI prepares structured artefacts such as transcripts, performance dashboards, and suggested advocacy–inquiry prompts for facilitator use, appear most aligned with current evidence and educational theory.[1,14] In this configuration, AI augments analytic capacity while preserving human oversight for high-stakes interpretation, emotional attunement, and adaptive questioning.

## Pedagogical Alignment

AI-augmented debriefing needs to follow accepted debriefing pedagogy (such as the International Nursing Association for Clinical Simulation and Learning or SSH best practice guidelines) in order to be considered educationally sound.[6] Three alignment strategies are particularly relevant:

 https://doi.org/10.2147/AMEP.S578857 

a) Scaffolded AI prompts mimicking facilitators: AI should start reflection with neutral, curiosity-driven questions (eg, "Tell me what you were thinking when…") rather than judgmental statements, then follow up according to learner responses.[5,20]
b) Hybrid models: AI-assisted facilitator workflows: Instead of replacing human debriefers, AI can prepare documentation (transcripts, event highlights, suggested questions) freeing up facilitators to concentrate on meaningful, high-value conversations.[1,14]
c) Learner-centered personalization: AI can adjust feedback and prompts based on a learner's past performance, cognitive load (eg, shorter prompts when physiological stress is high), and competence, to maximize reflection and support germane load.[13,28] However, there is still little proof of long-term learning, behavior modification, and transfer to clinical settings (Kirkpatrick levels 3 and 4).[1,9,14]

## Ethical Considerations

AI-augmented debriefing raises ethical issues. Audio, video, and telemetry data are sensitive. Recording, secondary analysis, retention, and anonymization must all be covered by explicit learner consent.[3,7] NLP models may mis-transcribe non-native accents or under-represent non-western communicative styles, disadvantaging diverse learners.[17] AI should not be used for decision-making, but rather as an adjunct. Human facilitators must continue to be in charge of making crucial decisions (like competency).[1,14] Feedback should be transparent and comprehensible, demonstrating the methods used to arrive at conclusions (eg, event times and transcript excerpts).[29]

## Key Gaps

Despite promising developments, major gaps remain. Few controlled trials comparing debrief modes have been conducted, as a result, very limited data exist on long-term learning or patient-care transfer.[9,14,15] Psychometric validation of AI metrics should be prioritized (eg, does "hesitation index" predict clinical decision error?).[9,29] Implementation research needs to be conducted in order to determine the kind of training required for faculty to integrate AI into debrief operations, interpret AI outcomes, and minimize over-reliance[1,6] Empirical work on bias (voice/accent), privacy risk, and learner perceptions remains scarce.[3,17]

## Practical Recommendations

Educators and simulation developers should adopt a hybrid model in which artificial intelligence (AI) is used to prepare debriefing artefacts (such as transcripts, event highlights, and suggested questions) while preserving the human facilitator as the central reflective coach who guides meaning-making and emotional processing.[1,14] AI-generated outputs should be deliberately designed to align with familiar debriefing frameworks by structuring automated feedback in advocacy-inquiry-compatible language and including transparent references such as quotations or timestamps to support credibility.[6,29] Before full implementation, institutions should pilot and localize AI systems by validating AI-derived scores against local expert raters and checking for transcription errors or biases, including those related to accent or language variability.[3,9] Safeguarding learner privacy is essential; programs must build opt-in consent processes, ensure telemetry data are anonymized, and retain data only for the minimum necessary duration.[3] Finally, faculty should receive dedicated training to develop the skills needed to interpret AI outputs, integrate them appropriately into debriefing practice, and maintain effective facilitation when automated systems generate unexpected or ambiguous feedback.[1,6]

# Conclusion

XR platforms generate high-resolution performance data, and AI systems can transform these data into structured analytics and feedback artefacts, with feasibility studies demonstrating technical viability and preliminary alignment with expert ratings. Promising applications include scalable, personalized debrief preparation, and adaptive reflective prompting in immersive contexts. However, AI-only debriefing models, the predictive validity of AI-derived metrics, and demonstrable effects on behavior change or patient outcomes remain unproven.

Unlike broader XR or AI-in-simulation reviews that emphasize scenario design or automated assessment, this review centers on debriefing as the pedagogical core of simulation-based education. By proposing a functional taxonomy of AI roles in XR debriefing and synthesizing current evidence through an outcomes lens, it clarifies both the potential and the present limitations of AI-augmented reflection.

Future work should extend beyond technological development to examine how AI reshapes debriefing scholarship, particularly its effects on reflective depth, psychological safety, facilitator judgment, and human–AI co-facilitation. Robust validation and longitudinal outcome studies will be essential for translating innovation into evidence-based practice.

## Abbreviations

AI, artificial intelligence; AR, augmented reality; ECA, embodied conversational agent; LLM, large language model; MR, mixed reality; NLP, natural language processing; SBE, simulation-based education; VAPS, virtual AI patient simulator; VR, virtual reality; XR, extended reality.

## Acknowledgments

Chat-GPT 5 was used for language improvement for this submission.

## Disclosure

The authors report no conflicts of interest in this work.

## References

1. Konzelmann J, Belk WB, Mahdy ZA, Mallala SB, Smeltzer S, Carter ML. Extended reality—a narrative review of current state and best use in health care simulation. *J Med Ext Real*. 2025;2:174–182. doi:10.1089/jmer.2024.0022
2. Duff JP, Morse KJ, Seelandt J, et al. Debriefing methods for simulation in healthcare: a systematic review. *Simul Healthc*. 2024;19(suppl 1):S112–S121. doi:10.1097/SIH.0000000000000778
3. Hamilton A. Artificial intelligence and healthcare simulation: the shifting landscape of medical education. *Cureus*. 2024;16(5):e59747. doi:10.7759/cureus.59747
4. Herur-Raman A, Almeida ND, Greenleaf W, Williams D, Karshenas A, Sherman JH. Next-generation simulation—integrating extended reality technology into medical education. *Front Virtual Real*. 2021;2:693399. doi:10.3389/frvir.2021.693399
5. Makransky G, Mayer RE. Benefits of taking a virtual field trip in immersive virtual reality: evidence for the immersion principle in multimedia learning. *Educ Psychol Rev*. 2022;34(3):1771–1798. doi:10.1007/s10648-022-09675-4
6. Decker S, Sapp A, Bibin L, et al. Healthcare simulation standards of best practice®: the debriefing process. *Clin Simul Nurs*. 2025;105:101775. doi:10.1016/j.ecns.2025.06.008
7. Anthamatten A, Holt JE. Integrating artificial intelligence into virtual simulations to develop entrustable professional activities. *J Nurse Pract*. 2024;20(9):105192. doi:10.1016/j.nurpra.2024.105192
8. Andersen SA, Mikkelsen PT, Konge L, Cayé-Thomasen P, Sørensen MS. The effect of implementing cognitive load theory-based design principles in virtual reality simulation training of surgical skills: a randomized controlled trial. *Adv Simul*. 2016;1(1):20. doi:10.1186/s41077-016-0022-1
9. Takac M, Collett J, Conduit R, De Foe A. Addressing virtual reality misclassification: a hardware-based qualification matrix for virtual reality technology. *Clin Psychol Psychother*. 2021;28(3):538–556. doi:10.1002/cpp.2624
10. Sandars J. The use of reflection in medical education: AMEE guide no. 44. *Med Teach*. 2009;31(8):685–695. doi:10.1080/01421590903050374
11. Fanning RM, Gaba DM. The role of debriefing in simulation-based learning. *Simul Healthc*. 2007;2(2):115–125. doi:10.1097/SIH.0b013e3180315539
12. Yardley S, Teunissen PW, Dornan T. Experiential learning: transforming theory into practice. *Med Teach*. 2012;34(2):161–164. doi:10.3109/0142159X.2012.643264
13. Carberry DE, Bagherpour K, Beenfeldt C, Woodley JM, Mansouri SS, Andersson MP. A roadmap for designing eXtended reality tools to teach unit operations in chemical engineering: learning theories & shifting pedagogies. *Digit Chem Eng*. 2023;6:100074. doi:10.1016/j.dche.2022.100074
14. PwC. Understanding the effectiveness of virtual reality soft skills training in the enterprise: a study. PwC; 2020. https://www.pwc.com/us/en/services/consulting/technology/emerging-technology/assets/pwc-understanding-the-effectiveness-of-soft-skills-training-in-the-enterprise-a-study.pdf. Accessed October 31, 2025.
15. Gani A, Pickering O, Ellis C, Sabri O, Pucher P. Impact of haptic feedback on surgical training outcomes: a randomised controlled trial of haptic versus non-haptic immersive virtual reality training. *Ann Med Surg*. 2022;83:104734. doi:10.1016/j.amsu.2022.104734
16. Sainsbury B, Łącki M, Shahait M, et al. Evaluation of a virtual reality percutaneous nephrolithotomy (PCNL) surgical simulator. *Front Robot AI*. 2019;6:145. doi:10.3389/frobt.2019.00145
17. Chary M, Parikh S, Manini AF, Boyer EW, Radeos M. A review of natural language processing in medical education. *West J Emerg Med*. 2019;20(1):78–86. doi:10.5811/westjem.2018.11.39725
18. Jerfy A, Selden O, Balkrishnan R. The growing impact of natural language processing in healthcare and public health. *Inquiry*. 2024;61:469580241290095. doi:10.1177/00469580241290095
19. Zhu XT, Cheerman H, Cheng M, Kiami SR, Chukoskie L, McGivney E. Designing VR simulation system for clinical communication training with LLMs-based embodied conversational agents. In: Extended Abstracts of the 2025 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery; 2025:1–9. doi:10.1145/3657604.3664642.
20. Cook DA, Overgaard J, Pankratz VS, Del Fiol G, Aakre CA. Virtual patients using large language models: scalable, contextualized simulation of clinician-patient dialogue with feedback. *J Med Internet Res*. 2025;27:e68486. doi:10.2196/68486

21. Alon Y, Naimi E, Levin C, Videl H, Saban M. Leveraging natural language processing to elucidate real-world clinical decision-making paradigms: a proof of concept study. *J Biomed Inform*. 2025;166:104829. doi:10.1016/j.jbi.2025.104829
22. Krilavičius T, De Paolis LT, De Luca V, Spjut J. eXtended reality and artificial intelligence in medicine and rehabilitation. *Inf Syst Front*. 2025;27 (1):1–6. doi:10.1007/s10796-024-10523-9
23. Halman J, Tencer S, Siemiński M. Artificial intelligence and extended reality in the training of vascular surgeons: a narrative review. *Med Sci*. 2025;13(3):126. doi:10.3390/medsci13030126
24. Rosser AA, Qadadha YM, Thompson RJ, Jung HS, Jung S. Measuring the impact of simulation debriefing on the practices of interprofessional trauma teams using natural language processing. *Am J Surg*. 2023;225(2):394–399. doi:10.1016/j.amjsurg.2022.09.018
25. Borg A, Georg C, Jobs B, et al. Virtual patient simulations using social robotics combined with large language models for clinical reasoning training in medical education: mixed methods study. *J Med Internet Res*. 2025;27:e63312. doi:10.2196/63312
26. Neo NW, Gunawan J, Levett-Jones T, Khoo ET, Chua WL, Liaw SY. Generative artificial intelligence in healthcare simulation-based education: a scoping review. *Clin Simul Nurs*. 2025;108:101819. doi:10.1016/j.ecns.2025.101819
27. Dasa D, Board M, Rolfe U, Dolby T, Tang W. Evaluating AI-driven characters in extended reality (XR) healthcare simulations: a systematic review. *Artif Intell Med*. 2025;170:103270. doi:10.1016/j.artmed.2025.103270
28. Willett J, Adelman-Mullally T, Ng H, Chung SY. Virtual reality simulation integration in a prelicensure nursing program: lessons learned. *Nurse Educ*. 2024;49(4):217–221. doi:10.1097/NNE.0000000000001609
29. Azher S, Mills A, He J, et al. Findings favor haptics feedback in virtual simulation surgical education: an updated systematic and scoping review. *Surg Innov*. 2024;31(3):331–341. doi:10.1177/15533506241242608