

Original Paper

Exploring Ambient Artificial Intelligence to Enhance Learning and Feedback During Operating Room-to-Intensive Care Unit Handoffs: Co-Design and Simulation Study

Laleh Jalilian¹, MD, MS; Federico Lorenzo Barra², MD; Tristan Grogan³, MS; Jason Lee¹, MD; Achuta Kadambi⁴, PhD

¹Department of Anesthesiology and Perioperative Medicine, University of California, Los Angeles, Los Angeles, CA, United States

²Department of Education - General Directorate, Azienda Ospedaliero-Universitaria "Maggiore della Carità", Novara, Italy

³Department of Medicine Statistics Core, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, United States

⁴Electrical and Computer Engineering, University of California, Los Angeles, Los Angeles, CA, United States

Corresponding Author:

Laleh Jalilian, MD, MS
Department of Anesthesiology and Perioperative Medicine
University of California, Los Angeles
757 Westwood Plaza, Suite 3325
Los Angeles, CA 90095
United States
Phone: 1 (310) 267-6629
Email: ljalilian@mednet.ucla.edu

Abstract

Background: Operating room (OR)-to-intensive care unit (ICU) handoffs are among the most complex and high-risk communication events in perioperative care. Despite the implementation of structured checklists, trainees often receive limited feedback on their communication skills, and simulation-based education rarely provides objective data on communication performance and checklist adherence. This study explores how an ambient artificial intelligence (AI) handoff assistant used during simulation-based training of OR-to-ICU handoff discussions can enhance clinical communication training and AI literacy by mapping spoken handoff discussions to handoff checklist items, providing immediate feedback on checklist item omissions, and generating a structured handoff note that functions as a feedback-rich learning artifact.

Objective: This study aims to co-design and evaluate an ambient AI handoff assistant that transcribes spoken OR-to-ICU handoff communication, maps the discussion to handoff checklist items, generates a structured handoff note for educational review, and provides immediate feedback on handoff completeness during simulated OR-to-ICU handoff discussions in a low-fidelity educational setting.

Methods: A 2-phase mixed-methods study was conducted within the University of California, Los Angeles, Department of Anesthesiology and Perioperative Care (July-October 2025). Phase 1 comprised co-design interviews with 4 clinician educators to identify limitations of current handoff training and inform AI feature development. Phase 2 involved an error analysis, as well as evaluations of usability, workload, and educational impact, conducted through ten 60-minute simulation sessions with pairs of medical students and first-year residents. Quantitative measures included the Physician Task Load Index, System Usability Scale, and a postsimulation survey; qualitative data from co-design sessions and simulation debrief interviews were thematically analyzed.

Results: Educators highlighted inconsistent checklist use and the absence of objective feedback on learners' communication skills as key areas that could benefit from structured documentation of handoff discussions using AI. Error analysis of the ambient AI handoff assistant revealed a mean of 3.6 (SD 1.2) errors per note, with incorrect output being the most frequent error type. There was no statistically significant difference between the ambient AI handoff assistant and the paper checklist with respect to the Physician Task Load Index and System Usability Scale measures. Trainees valued real-time transcripts and structured handoff notes for reflection of communication practices, and exposure to AI documentation errors enhanced critical thinking and awareness of AI technology limitations.

Conclusions: The ambient AI handoff assistant mapped simulated handoff discussions to checklist items and generated a structured handoff note, facilitating reflection on team-based communication skills in handoff education. Imperfections in

the AI's output encouraged critical appraisal of its capabilities and prompted discussion about automation complacency, suggesting that AI-assisted simulations can foster both communication and digital literacy skills essential for future AI-enabled clinical practice.

JMIR Med Educ 2026;12:e85666; doi: [10.2196/85666](https://doi.org/10.2196/85666)

Keywords: simulation training; handoff communication; human-computer interaction; ambient artificial intelligence; prompt engineering

Introduction

Effective handoff and transitions of care discussions are essential for ensuring safe and high-quality patient care [1-4]. Operating room (OR)-to-intensive care unit (ICU) handoffs are particularly complex and high risk, requiring coordinated communication among surgical, anesthesia, and ICU teams under conditions of cognitive load, time pressure, and frequent interruptions [5-7]. Incomplete communication transfer during OR-to-ICU handoffs remains a preventable contributor to medical errors and adverse patient outcomes [6,8,9], and structured handoff checklists have been developed to standardize communication and improve information transfer [10-15]. Despite the introduction of checklists, adherence to checklists remains variable [16], and trainees often receive limited instruction or feedback on communication skills [17]. Handoff skills are frequently learned informally through observation, with variable role modeling, inconsistent feedback, and few opportunities for deliberate practice [18]. As a result, trainees may be unaware of omissions or deficiencies in their communication and may not learn important relational communication skills. Without objective assessment, gaps between knowledge and performance may persist, contributing to preventable harm. Simulation-based education offers a safe environment to practice handoffs, but most simulations rely on subjective observation. Learners may overestimate the completeness of their handoffs, and without objective records or data-driven feedback, meaningful self-assessment and improvement are limited [19].

Recent advances in ambient artificial intelligence (AI) present new opportunities to address this educational gap. Ambient AI systems use natural language processing and large language models (LLMs) to passively capture spoken communication and generate summarized documentation of the transcript. Applied to checklist-based handoffs, such systems are being explored for their potential to identify which checklist items are verbalized, map spoken communication to structured checklist elements [20], and provide feedback on the completeness of information exchange. This capability may enhance handoff education from subjective assessment to measurable performance analytics, enabling iterative practice and self-reflection. Despite growing interest in AI applications for documentation and workflow support, no prior studies have examined the use of ambient AI for enhancing handoff communication skills to medical trainees [21,22]. This gap is particularly striking given that standardized handoff checklists already provide structured frameworks that align well with LLM-based mapping through prompt engineering, and simulation-based handoff training

is widely implemented yet lacks robust, objective assessment tools [23] of verbal discussions [24].

To address this need, we developed and evaluated an ambient AI-assisted handoff simulation platform designed to support learning and feedback during OR-to-ICU handoff educational sessions. The system listens to trainee handoff discussions, maps verbalized content to checklist items, and generates structured documentation identifying which items were omitted. The objectives of this study were to (1) identify and co-design features of an ambient AI handoff assistant for OR-to-ICU handoff education through needs assessment and interviews with clinician educators, focusing on limitations of current handoff training and design requirements for AI-mediated feedback, (2) assess the performance and accuracy of the LLM-based checklist validation in identifying discussed and omitted handoff items, and (3) evaluate the usability, workload, and learning impact of the AI-assisted handoff simulation platform during low-fidelity sessions, while exploring learner perceptions of the AI-generated feedback, specifically its influence on communication, reflection, and perceived educational and clinical value.

Methods

Overview

A 2-phase, mixed-methods study was conducted within the University of California, Los Angeles (UCLA), Department of Anesthesiology and Perioperative Care (July-October 2025). The COREQ (Consolidated Criteria for Reporting Qualitative Research) [25] (Checklist 1) framework was used to guide research reporting.

Study Design

We used a human-centered, co-design methodology to develop an ambient AI handoff assistant. Co-design was selected for its suitability in simulation-based health care research, as it enables clinicians to surface workflow challenges and limitations, shape system features, and envision technology integration in high-stakes team communication [26,27]. The study was conducted in two phases: (1) needs assessment and co-design interviews with OR-to-ICU quality improvement leaders, aiming to (i) identify limitations of the current handoff educational modules, (ii) elicit expectations and perceptions of automated handoff discussion capture using an ambient AI handoff assistant, and (iii) prioritize design features for the ambient AI handoff assistant for handoff education; and (2) low-fidelity simulation sessions to evaluate the educational use of the ambi-

ent AI handoff assistant, with mixed-methods assessment of experience, acceptance, usability, and workload.

Participants and Setting

For the Needs Assessment and Co-Design workshop, 4 clinician educators participated in semistructured 45-minute interviews (remote or in person). Their years of professional experience ranged from 9 to 24, with a mean of 16 (SD 4.9) years. The decision to interview educators individually was dictated by their availability to meet with the study

team. In the simulation phase, 10 simulation sessions were conducted in a low-fidelity setting. Each session involved 2 medical students or first-year interns who had no prior exposure to the OR-to-ICU handoff process at UCLA, who played the roles of surgery and anesthesia, and played the role of the ICU attending. New participants were recruited for each simulation session to minimize bias from repeated exposure. Demographic breakdown of simulation participants is listed in [Table 1](#).

Table 1. Overview of simulation study participants (n=20).

Participant type	Number of participants
What is your current training level?	
Medical student	7
Resident	13
Sex	
Male	7
Female	10
Nonbinary	2
Prefer not to answer	1
Do you have experience using an ambient AI ^a scribe? (An AI scribe refers to a tool that automatically generates documentation from ambient audio or speech input during clinical encounters, such as DAX Copilot or Nabla.)	
Yes	2
No	18
Do you have experience using generative AI technologies? (Examples include ChatGPT, Claude, Med-PaLM, Microsoft Copilot, or other tools that use artificial intelligence to generate content, summarize information, or support documentation.)	
No, I have never used any AI-generated technologies.	1
Yes, I've tried one briefly out of curiosity.	3
Yes, I've used them occasionally for general tasks (eg, answering questions, drafting emails).	9
Yes, I've used them occasionally for clinical or work-related tasks (eg, summarizing notes, generating documentation).	1
Yes, I use them frequently and consider them an integral part of my workflow.	6

^aAI: artificial intelligence.

Needs Assessment and Co-Design Interviews

The primary study author interviewed 4 clinician educators in a 45-minute semistructured interview fashion using the Clinician Educator Interview guide in [Multimedia Appendix 1](#) after obtaining verbal consent. The interviewer was a clinician trained in the use of qualitative research techniques. Given the interviewer's role as a faculty peer, reflexivity was emphasized to reduce potential bias. Interviews began with reflections on recent OR-to-ICU handoffs and handoff education to ground discussions in real clinical practice. Clinician educators identified limitations of current handoff educational methods and potential roles for ambient AI in encouraging educational discussion on checklist completeness and communication quality. The interviewer then introduced a design probe, an ambient AI handoff assistant to be used during handoff education simulation sessions with junior trainees. A functional prototype was used to ground discussions in concrete interactions. Using whiteboards/Miro and the ambient AI handoff assistant prototype, participants sketched desired AI functions (eg, transcription verification,

structured note review, and interaction styles) and collaboratively prototyped ideal AI-supported simulation workflows. Transcripts were member-checked and coded in Atlas.ti (ATLAS.ti Scientific Software Development GmbH) using a grounded theory approach until thematic saturation was reached.

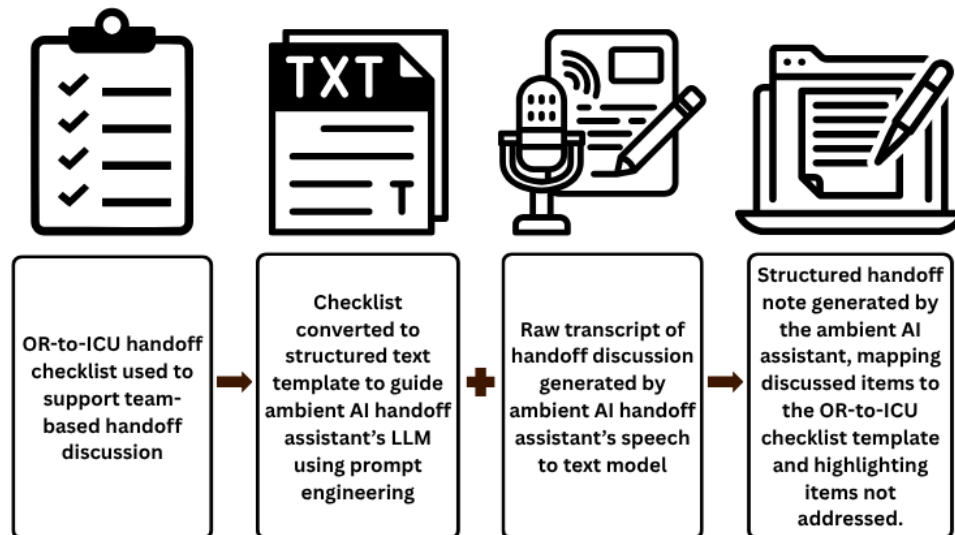
Ambient AI Handoff Assistant Technology Development

Based on insights from the needs assessment and co-design interviews, we developed an ambient AI handoff assistant to support both clinician educators and learners during OR-to-ICU handoff simulation sessions. The assistant digitally displays the UCLA OR-to-ICU Handoff Checklist, provides real-time visual transcription of the clinical discussion, and automatically generates a structured handoff note where the discussion is mapped to individual checklist items for learner review. It also highlights checklist items that were not discussed, helping educators and learners identify omissions and communication gaps. A commercial ambient AI scribe was not used because existing systems lack the

capability to map spoken handoff dialogue directly to a structured handoff checklist for the purpose of assessing checklist adherence. Meeting this requirement necessitated custom prompt engineering of an LLM aligned specifically to

UCLA's OR-to-ICU handoff checklist. [Figure 1](#) illustrates the technical framework used to map handoff discussion content to corresponding checklist items.

Figure 1. Standardized handoff checklists, such as the University of California Los Angeles operating room-to-intensive care unit Handoff Checklist, can be converted into structured text templates to guide the ambient artificial intelligence handoff assistant in generating documentation aligned with the desired format. Figure adapted from Jalilian et al [28]. AI: artificial intelligence; ICU: intensive care unit; LLM: large language model; OR: operating room.



Google AI Studio was used to develop the ambient AI handoff assistant, a single-page application built with HTML5, CSS3, and TypeScript, running entirely within the client's web browser to ensure low latency and data privacy. Real-time audio capture and transcription are handled by the Gemini 2.5 Flash Live application programming interface (API), configured for continuous listening to capture the full clinical dialogue. The captured transcript is processed by the Gemini 2.5 Flash LLM via the @google/genai SDK. No fine-tuning of the model was performed; instead, a zero-shot prompt engineering approach was used. The prompt requirements and handoff template were specified by clinical domain experts to align with UCLA's OR-to-ICU handoff checklist and are listed in [Multimedia Appendix 2](#). The prompt uses explicit "Chain of Thought" instructions, requiring the model to extract factual data, resolve conflicting information by prioritizing the most recent statement, and strictly mark missing fields as "N/A." To ensure structured and reliable data extraction, the system leverages the model's

JSON mode with a strict responseSchema, forcing the output into a predictable JSON object containing a formatted OR-to-ICU handoff note. The organization of transcription content into the final checklist structure and the identification of items labeled "not discussed" are performed via deterministic post-processing logic in TypeScript, rather than a separate predictive model. This logic parses the AI-generated JSON, compares it against a master checklist of 38 clinical items, and flags any item as 'missed' if the value is empty or remains "N/A." The application's core logic is managed by a finite state machine that guides the user through distinct phases: listening, processing, verifying, and summarizing. Auditory feedback and status prompts are delivered via the Gemini 2.5 Flash Live API's SpeechSynthesis interface, with a preselected "en-GB" female voice to ensure a consistent user experience. Screenshots of the different interface views of the ambient AI Handoff Assistant are visible in [Figure 2](#) and [Figure 3](#).

Figure 2. Screenshot of the ambient artificial intelligence handoff assistant pairing digital checklist display with visualization of real-time transcription. The University of California Los Angeles operating room-to-intensive care unit handoff checklist is digitally displayed as a visual cognitive aid for learners, who can simultaneously view a real-time transcription of the discussion in the right-hand panel. AI: artificial intelligence; ICU: intensive care unit; OR: operating room.

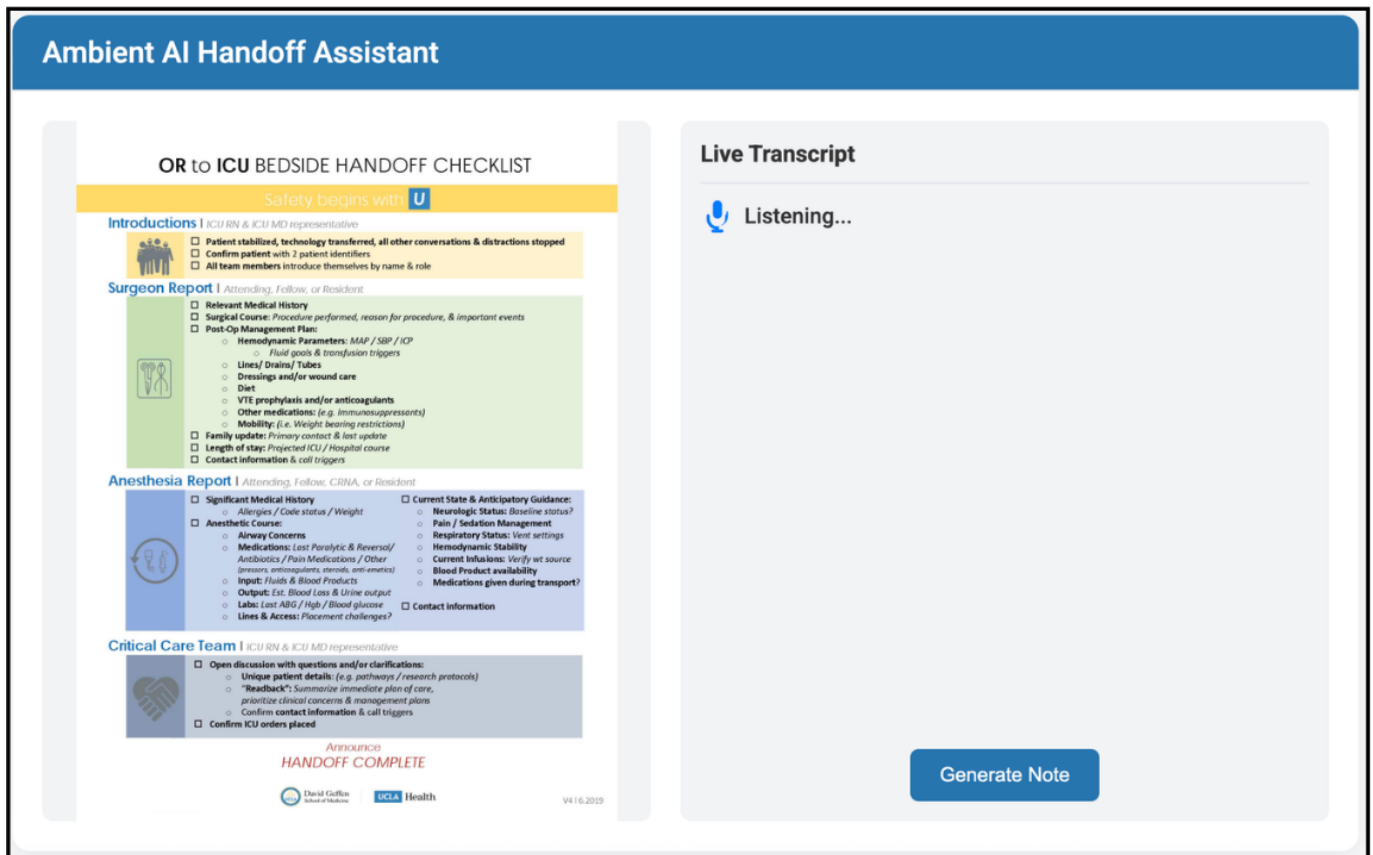
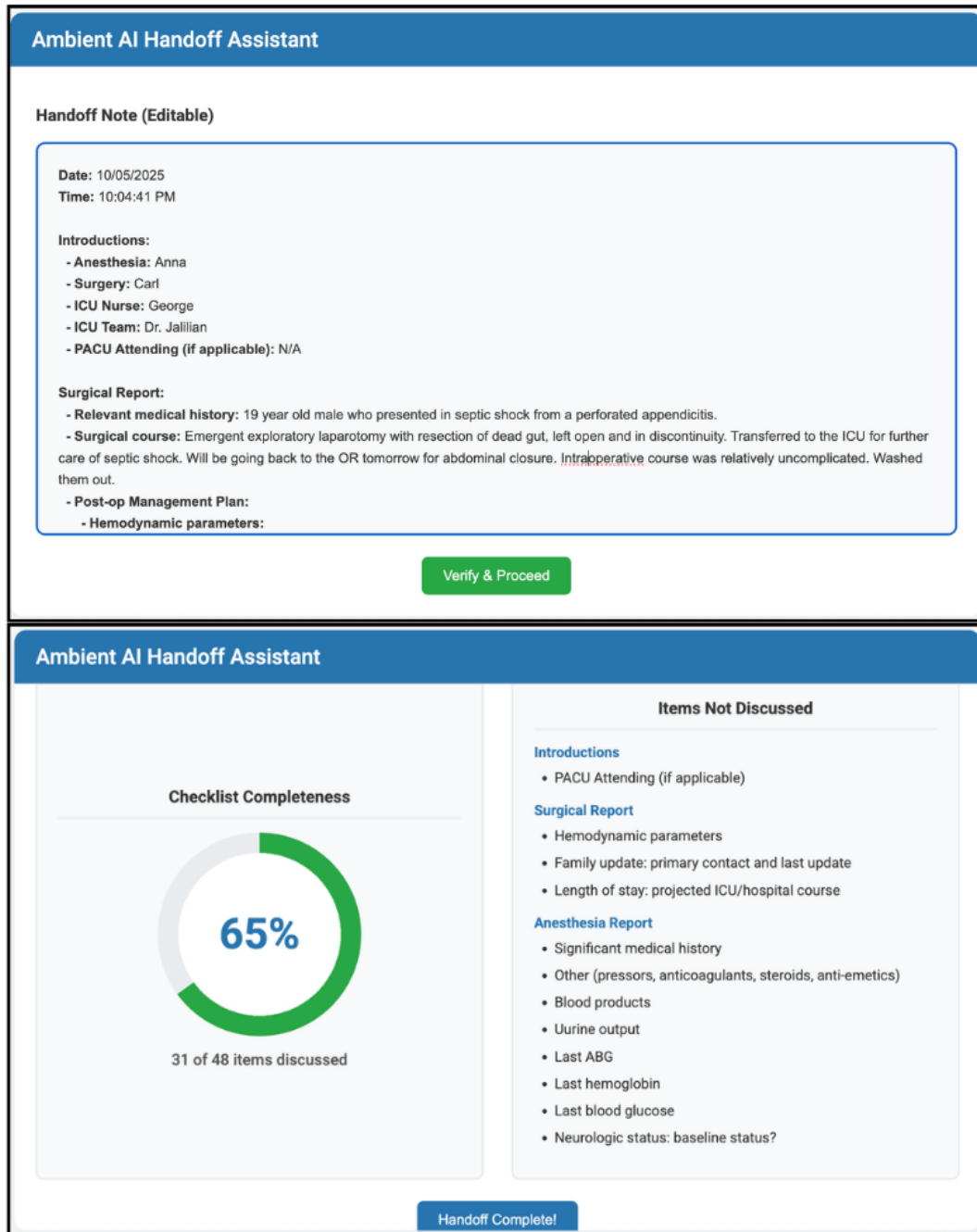


Figure 3. Screenshot of the ambient artificial intelligence handoff assistant as a supportive tool for checklist omission detection. In the top panel, handoff participants view an editable note organized in a checklist-based format, in which discussed content is mapped to its corresponding checklist items, facilitating post-hoc identification of omissions compared with a traditional narrative format. In the bottom panel, items flagged by the artificial intelligence Handoff Assistant as “not discussed” are displayed at the end of the session, providing explicit automated feedback to further support gap identification. Both the checklist-based note in the top panel and the AI-flagged omissions in the bottom panel operate after the discussion has concluded, serving complementary roles by enabling structured human review while also providing explicit, automated feedback to reinforce omission identification. AI: artificial intelligence; ICU: intensive care unit; OR: operating room.



Error Analysis of the AI Handoff Assistant

Five simulated OR-to-ICU handoff scenarios were developed using a previously established agentic framework [29] to evaluate the ambient AI handoff assistant in generating handoff notes and displaying undiscussed items. The scenarios represented a range of surgical cases commonly encountered in the ICU. Two clinician educators (LJ and JL) simulated the surgeon and anesthesiologist roles by reading the scenarios aloud while the ambient AI handoff assistant

operated in the background. After each session, the clinician educators reviewed the AI-generated handoff note for errors. Each identified error was categorized into one of the following error types: omission, irrelevant or misplaced text, wrong output, or addition. This categorization was coded independently by 2 clinician educators (LJ and JL), and disagreements were discussed until consensus was reached.

Simulation Testing of the Ambient AI Handoff Assistant

Ten 60-minute simulation sessions in a nonclinical, low-fidelity setting were conducted with pairs of medical students or first-year resident physicians who played the roles of anesthesiology and surgery team members during an OR-to-ICU handoff. Participants were recruited through voluntary email invitations and had no academic or evaluative relationship with the study faculty. Demographic characteristics are presented in [Table 1](#). Each simulation was facilitated by faculty experts in handoff communication who held no supervisory role over participants. Reflexivity was emphasized to mitigate potential bias. Two distinct simulation scenarios (A and B) were generated from a previously developed agentic workflow [29]. The study used a within-team crossover design in which each dyad completed two simulated handoffs: (1) Control: a standard verbal handoff using the institutional paper checklist, and (2) Intervention: a handoff supported by the ambient AI handoff assistant. Each team served as its own control to reduce variability in communication skills, teamwork, and experience. Condition order (AI-first vs. control-first) and case scenario (A vs B) were randomly assigned and counterbalanced to control for order and case effects. Following both simulations, participants completed a postsimulation survey and a semistructured debriefing interview ([Multimedia Appendix 3](#)). Debriefing interviews were transcribed and returned to participants for member checking. All participants provided verbal informed consent.

Statistical Analysis

For the needs assessment and co-design phase, inductive content analysis was used to synthesize design insights. Two researchers (LJ and JL) independently coded transcripts, sketches, and field notes to identify (1) limitations in current handoff education and (2) desired ambient AI handoff assistant features. Coding discrepancies were resolved by consensus, and themes were consolidated into preliminary design requirements for the handoff AI assistant. The Educator Co-design codebook is listed as Codebook A in [Multimedia Appendix 4](#).

Quantitative data from the 4-item Physician Task Load Index (PTL) [30] and System Usability Scale (SUS) were analyzed descriptively. Each participant completed both the control (paper checklist) and intervention (ambient AI assistant) conditions in a within-subjects crossover design. For each measure, paired-samples analyses were conducted to compare participant scores between conditions. The normality of paired differences was evaluated using the Shapiro–Wilk test. When assumptions of normality were satisfied, paired

t-tests were used to compare mean scores between the control and AI conditions. Results are reported as means (standard deviations) for normally distributed data. All estimates include 95% CIs for the mean differences. Effect sizes for paired 2-tailed *t* tests were expressed as Cohen *d* (standardized mean difference for paired data). Although 2 participants worked together in each of 10 simulation sessions, the paired 2-tailed *t* test was selected as the primary analytic approach because each participant served as their own control, and between-team clustering was minimal. All analyses were conducted using Python (version 3.11; Python Software Foundation), with statistical significance set at $P < .05$ (2-tailed).

For the simulation debriefings, qualitative data from debriefings were coded independently by two researchers (LJ and JL) using [Atlas.ti](#), with themes refined through iterative consensus. Structured observations were analyzed to capture communication dynamics and were triangulated with interview findings to assess human-technology interaction. The Learner Simulation Codebook is listed as Codebook B in [Multimedia Appendix 4](#).

Ethical Considerations

This study was approved by the UCLA Institutional Review Board #25-1998. Verbal informed consent was obtained from all participants before data collection. Participants were informed of the study purpose, procedures, potential risks and benefits, and their right to decline participation or withdraw at any time. Measures were taken to protect participant privacy and maintain confidentiality, including deidentification of data and secure storage of study materials.

Results

Needs Assessment and Co-Design Thematic Analysis

Themes from the co-design interviews with educators are listed in [Table 2](#). Educators described ongoing challenges in teaching closed-loop communication and ensuring complete information transfer during handoffs. The OR-to-ICU checklist was viewed as a valuable cognitive aid for structuring handoff discussions across surgical, anesthetic, and ICU domains, yet there was recognition that its real-world use was inconsistent and that trainees could inadvertently skip over checklist items. Educators emphasized the need for greater situational awareness and systematic progression through each checklist item during real-time communication.

Table 2. Thematic analysis of needs assessment and co-design sessions on ambient AI handoff assistants for operating room-to-intensive care unit handoff education. Representative quotes are drawn from clinician educator interviews conducted during the co-design phase (n=4).

Theme	Key insight	Representative quote
Promoting professional communication	Awareness of AI ^a documentation may encourage more mindful, deliberate, and professional communication among trainees.	<i>"I'd be more careful of what I'm saying."</i>
Enhancing reflective learning	AI-generated transcripts and note summaries may support self-assessment and reflection, helping learners review and improve their communication.	<i>"Having a transcript to go back and refer to adds another level of feedback."</i>
Exposure and readiness for AI	Exposure in simulation may allow trainees to experience AI-assisted documentation and assess its accuracy and use in realistic settings.	<i>"Ambient AI scribes might be one of the first AI technologies that perioperative doctors might be exposed to. Would be interesting to see how accurate the handoff assistant is in documenting the discussion."</i>
Calibration of trust in AI output for learners	EAmbient AI Handoff Assistant may be more inaccurate than commercial ambient AI scribes, but exposure to AI transcription inaccuracies may enable trust calibration in AI.	<i>"[Learners] need to see the AI make mistakes. It doesn't matter how accurate [the ambient AI handoff assistant] is, because exposure to the errors is instructive in and of itself."</i>
Reducing documentation burden in simulation environments	Automating note-taking and checklist reconciliation may allow educators to focus on observation and higher-level feedback rather than manual documentation during teaching.	<i>"It'd be nice to focus on the students instead of taking notes."</i>
Ethical and privacy considerations in training	Educators emphasize the ethical obligation to protect learners' privacy, psychological safety, and autonomy when using AI in training.	<i>"If they know they're being recorded, they might not ask questions they think will sound dumb."</i>
Trust and validation of AI output for educators	Educators are uncertain of the accuracy of AI-generated content and need experience to understand its educational use.	<i>"If it can't tell which handoff checklist items the trainee did or did not mention, I wouldn't rely on it. It might not be the point of a tool like this though."</i>
Design requirements for ambient AI handoff assistant	Digital checklist display, real-time display of transcription, mapping transcript content to checklist categories in the format of a structured checklist-aligned handoff note, and explicitly labeling checklist items as "discussed" or "not discussed" based on transcript evidence would support clinician educators.	<i>"I don't want to bring a paper checklist to one of these sessions, so we should digitally display it as part of the module."</i>

^aAI: artificial intelligence.

We've taught the residents to use the checklist for years, but in the heat of the moment, they skip parts of the checklist. [Clinician Educator]

Educators felt that many trainees struggled with the interpersonal dimensions of team communication, such as speaking hesitantly, not introducing themselves, or not making eye contact with other members of the handoff team. Educators highlighted that in high-stakes transitions of care, clarity, confidence, and engagement are essential safety and communication behaviors that promote shared understanding. Developing these expressive and relational communication skills was seen as critical yet under-taught.

We don't teach the residents how to speak in a handoff. We focus on what checklist items need to be discussed, not so much on delivery, tone, or engagement with the team. [Clinician Educator]

Traditional feedback was described as subjective and infrequent, offering limited improvement, especially if an attending physician does not accompany a trainee to the ICU for a handoff.

Without a record of what was actually said, feedback is just opinion. Everyone remembers the handoff a little bit differently. [Clinician Educator]

Educators perceived AI-generated documentation as a promising way of creating an objective record of what was actually communicated during handoffs. In developing the ambient AI handoff assistant, they noted that AI-assisted documentation technologies could more concretely identify communication omissions and translate them into specific, actionable learning points, while maintaining a safe environment for learners to critically evaluate the limitations of AI.

My understanding is that AI output isn't perfect, but I guess that would be part of the learning. Trainees would see how the AI interprets them. Maybe it might make them question it or think critically about accuracy. [Clinician Educator]

Educators felt that the design requirements of the ambient AI handoff assistant should include digital display of the checklist, real-time visualization of transcription, and an AI-generated handoff note where each checklist item is populated with what the learners discussed to enable educator validation of checklist completeness. Beyond the educational

module, they suggested that commercially available ambient AI scribe technologies could potentially enhance handoff processes in the future through the generation of a handoff note for the receiving ICU team to use as a reference of the discussion. Introducing ambient AI first in a simulated setting would allow educators and Handoff QI leaders to evaluate their use, feasibility, and user acceptance in simulated handoff workflows before considering any potential real-world use.

Error Analysis of the Ambient AI Handoff Assistant

There were 18 errors noted in 5 of 5 draft notes (average of 3.6 errors per draft note). Error type by product is detailed in Table 2, with “wrong output” being the most frequent error across products.

Physician Task Load and System Usability Score

The paired-samples 2-tailed *t* test showed no significant difference in perceived task load between the paper checklist (mean 12.9, SD 16) and the AI handoff assistant (mean 15.8, SD 8.3), a difference of 2.9 (95% CI -5.9 to 11.7; *P*=.50). This suggests that the AI tool did not meaningfully change participants’ subjective workload during simulated handoffs. Participants rated the AI assistant and paper checklist as similarly usable, with mean SUS scores of 81.4 (SD 11.3) for the paper checklist and 74.3 (SD 13) for the AI assistant, a difference of 2.9 (95% CI -0.6 to 14.8; *P*=.06). Although the AI assistant trended slightly lower in usability ratings, this difference was not quite statistically significant (Tables 3 and 4).

Table 3. Physician task load and System Usability Scale (n=20).

Metric	Paper checklist, mean (SD)	Ambient AI ^c handoff assistant, mean (SD)	<i>t</i> test (df)	<i>P</i> value	Cohen <i>d</i>
PTL ^a	12.9 (16)	15.8 (8.3)	0.68 (19)	.50	0.15
SUS ^b	81.4 (11.3)	74.3 (13.0)	-1.93 (19)	.06	-0.43

^aPTL: Physician Task Load Index.

^bSUS: System Usability Scale.

^cAI: artificial intelligence.

Table 4. Postsimulation survey results^a (n=20).

Postsimulation survey results	Mean (SD)	95% CI
The AI ^b -generated handoff note and missed-items list helped me recognize strengths and weaknesses in my handoff communication.	3.9 (0.9)	3.5-4.3
The AI assistant’s feedback complemented faculty debriefing and peer discussion.	4.3 (0.6)	4-4.6
I was able to distinguish between accurate and inaccurate information in the AI-generated documentation.	3.8 (1)	3.2-4.3
I exercised appropriate caution in interpreting and relying on AI-generated documentation.	3.9 (0.9)	3.5-4.3
Reviewing the AI-generated note and missed-items list increased my confidence in performing OR ^c -to-ICU ^d handoffs.	4 (1)	3.5-4.4
The AI feedback helped me feel more prepared to apply structured handoff practices in clinical settings.	4.2 (0.9)	3.8-4.6

^aResponse options of strongly disagree (1), disagree (2), neither agree nor disagree (3), agree (4), and strongly agree (5)

^bAI: artificial intelligence.

^cOR: operating room.

^dICU: intensive care unit.

A post-hoc power analysis was performed to estimate the minimum detectable effect size under our study’s sample size of 20 paired observations. Assuming a 2-sided $\alpha=.05$, this sample provides 80% power to detect a standardized mean difference of 0.66. In contrast, the observed effect sizes in our data were 0.43 for the SUS outcome and -0.15 for the task load outcome, which correspond to smaller magnitudes than this detectable threshold. To achieve 80% power to detect an effect size of 0.43, a sample size of approximately 45 would be required, whereas detecting an effect as small as 0.15 would require roughly 351 participants. These estimates may guide the design of future studies with adequate power to detect more modest effects.

Simulation Debrief Thematic Analysis

Themes from debrief interviews with trainees are captured in Table 5. Trainees emphasized the educational value of an ambient AI handoff assistant in simulated handoff training. AI-generated transcripts and structured notes were felt to provide objective feedback, allowing learners to identify omissions, reflect on their communication, and discuss the limits of AI. These outputs transformed team discussions into concrete learning records that supported debriefing and self-assessment, addressing a gap in handoff education often reliant on subjective recall (Table 6).

Table 5. Frequency counts, percentages, definitions, and examples of error types in the artificial intelligence-generated operating room-to-intensive care unit handoff note (n=5).

Error type	Ambient AI ^a handoff assistant, n (%)	Definition	Example
Omission	5 (27.7)	AI assistant leaves out key information that was discussed	<i>"No mention of estimated blood loss when it was discussed"</i>
Misplaced text	3 (16.7)	AI assistant output is technically correct but not in appropriate checklist item	<i>"Estimated blood loss was discussed but placed in the wrong checklist item"</i>
Wrong output	7 (38.9)	AI assistant provides an incorrect response	<i>"Phenylephrine was the spoken vasopressor medication but ephedrine was documented instead"</i>
Addition	3 (16.7)	AI assistant adds inappropriate or irrelevant information	<i>"Estimated blood loss was described only as 'minimal,' but the AI output documents it as 'minimal (~50 mL)."</i>
Total	18 (100)		

^aAI: artificial intelligence.

Table 6. Thematic analysis of simulation debrief. Representative quotes are drawn from learner participants conducted during the simulation phase (n=20).

Theme	Key insight	Representative quote
Exposure and readiness for AI ^a	Learners viewed the simulation as a valuable first exposure to AI tools in clinical communication, where they could evaluate AI capabilities and accuracy.	<i>"This was my first time using something like this. I was surprised at how well it actually captured what I said."</i>
Reflective and professional learning	Reviewing AI-generated transcripts encouraged self-assessment and helped learners recognize patterns in their communication and teamwork.	<i>"Seeing the transcript was good. I didn't realize I skipped over things when I talked through a case."</i>
Feedback integration and educational use	Learners discuss how they or faculty could use AI transcripts to guide debriefing, identify missed information, or support feedback loops.	<i>"I liked the feature that highlighted what we didn't discuss."</i>
Communication awareness	Awareness of being recorded made learners more deliberate in how they presented information during handoffs.	<i>"Knowing it was recording made me think about what I was saying."</i>
Peer dynamics and team communication	Learners sometimes adjust how they communicate with teammates when AI is present.	<i>"We were more formal because we knew the AI was picking up everything."</i>
Trust, accuracy, and error sensitivity	Learners noted AI's limitations and inaccuracies in speech-to-text transcription and note generation, emphasizing the need for human verification.	<i>"It made mistakes! It was interesting to see how some of the medical language was mistakenly transcribed, like 'mcgs' being transcribed as 'Mike's.' I think you will need to check it."</i>
Translation to real-world clinical practice	Learners connected the simulation to future practice, envisioning how AI scribes could improve communication and documentation in real OR ^b -to-ICU ^c handoff workflows.	<i>"If this worked in the real ICU, the generated note would keep everyone on the same page and be really helpful to refer back to."</i>
Emotional and psychological responses to AI	Learners sometimes felt self-conscious knowing AI is "listening."	<i>"At first it felt weird knowing it was recording everything. I got used to it though."</i>
Ethical and privacy awareness in training	Learners reflected about consent, data storage, and fairness in educational AI use.	<i>"I'd want to know who can access the data from these sessions."</i>

^aAI: artificial intelligence.

^bOR: operating room.

^cICU: intensive care unit.

Seeing the AI transcript right after the handoff was great. It was interesting to see what I actually said versus what I thought I said. [Learner]

Exposure to the ambient AI handoff assistant's errors also promoted critical thinking and awareness of the technology's limitations.

Seeing the mistakes the AI made in its transcription was helpful. We can't just go on auto-pilot with it. [Learner]

Trainees further envisioned how AI scribe technology might impact real-world OR-to-ICU workflows to document complex handoffs, enhance continuity, and strengthen accountability.

If it could document the handoff automatically, it would help with continuity, especially for the ICU team coming on shift. [Learner]

They did express reservations regarding data governance of ambient AI technologies.

If AI gets implemented in the real-world, it should be clear what data is actually being stored. [Learner]

Discussion

Principal Findings

This study designed and evaluated an ambient AI handoff assistant for documenting OR-to-ICU handoff discussions within a medical education context. Because OR-to-ICU transitions are among the most complex and high-risk discussions in perioperative care, they provided an ideal setting for exploring how ambient AI can support medical trainees in promoting awareness of their communication and checklist completion during high-stakes discussions. Introducing the ambient AI handoff assistant to transcribe participants' discussions appeared to enhance perceptions of learning and educational value, as reflected in both survey feedback and debrief interviews. Further, PTL scores did not differ significantly between the AI scribe and the traditional checklist. This may suggest that verbally delivering a complex OR-to-ICU handoff is inherently challenging for inexperienced trainees, regardless of whether they are supported by a paper checklist or the AI-based handoff assistant. SUS scores did not differ significantly between the paper checklist and the AI-based handoff assistant, suggesting that digitally displayed checklists may be acceptable to users as a cognitive aid. Participants reported that the AI-generated note functioned as a valuable record of the discussion, helping them verify checklist completion, promoting reflective conversations about communication, and enabling them to better understand the limitations of AI-assisted tools. These observations align with prior work suggesting that documentation of verbal discussions can enhance reflective learning and promote deliberate practice in team communication [31,32]. These findings also align with prior research showing that new technologies can be perceived as equally usable as established tools once users recognize their educational and workflow benefits [33].

Because learners were medical students and first-year interns with limited semantic understanding or familiarity with complex anesthesia and surgical terminology, they adhered closely to the scripted simulation scenarios. As a result, their handoffs were generally comprehensive, creating a ceiling effect on traditional performance measures such as completeness and limiting our ability to assess whether the ambient AI handoff assistant improved completeness. However, qualitative feedback indicated that learners valued the opportunity to practice structured team-based communication in a simulated setting and to engage with speech-to-text technology. This experience prompted broader discussions about communication strategies in high-stakes scenarios,

focusing on the importance of closed-loop communication, clear and confident speech, and shared situational awareness to ensure effective information transfer. Further, interacting with imperfect AI output proved instructive, as learners observed transcription and checklist-mapping errors firsthand. In one instance, a learner with a foreign accent observed that her articulation of medical terminology was frequently inaccurately transcribed by the ambient AI handoff assistant, prompting a discussion about biases in training data and the limitations of speech-to-text systems. Trainees discussed the importance of vigilance of stereotyped assumptions and biased outputs with AI-generated content. Further, the limitations of the ambient AI handoff assistant in recognizing complex perioperative medical terminology also led to transcription and checklist mapping errors. Learners felt these inaccuracies became a productive feature of the simulation, enabling them to calibrate their trust in ambient AI outputs and also to understand how these technologies are assistive but still require human oversight. This led to discussion on automation complacency and its implications for patient safety with the use of AI technologies, prompting learner participants to reflect on the importance of verifying AI outputs [34,35] in future AI-assisted clinical environments. This fostered critical appraisal skills and transformed the AI assistant from a passive documentation tool into an active learning aid that promoted vigilance, AI literacy, and reflective communication. Future iterations of AI-assisted handoff training could vary the fidelity of AI output to examine how learners calibrate trust and maintain situational awareness when interacting with interactive systems.

Although this project primarily examined how ambient AI could support handoff education through automated documentation and verbal checklist workflow capture, it also demonstrated that high-stakes, multi-speaker handoff discussions can be structured into clinically and pedagogically meaningful documentation. The ambient AI handoff assistant was designed to map spoken communication to predefined checklist elements, thereby supporting the Educator during handoff simulation sessions and allowing greater focus on the Learners rather than on checklist reconciliation. During the error analysis phase, we assessed the ambient AI handoff assistant's ability to map what was discussed to its corresponding checklist item. Most errors were attributed to difficulties in recognizing medical terminology and variability in speaker articulation. This finding was expected, given the inherent limitations of the Gemini 2.5 Flash Live API on which the system relies. Further, there were more transcription errors with the learners in simulation compared with the prior error analysis, likely reflecting the greater acoustic variability of real learning environments. Together, these factors contributed to downstream checklist mapping errors that resulted in imperfect checklist completeness assessment functionality of the ambient AI handoff assistant. Importantly, the goal of developing this system was to create an educational platform that would enable some transcription feedback of a high-stakes discussion, exposure to ambient AI, and assessment of whether AI could map spoken team communication to handoff checklist items as a feedback tool. Both

educators and learners reflected on whether ambient AI technologies could play a role in real-world handoffs and used the simulation experience as a basis for discussing whether commercial ambient AI scribes might provide value in real-world OR-to-ICU handoff clinical workflows.

Simulation as Essential Infrastructure for AI Evaluation in Health Care

Our work highlights the importance of simulation-based testing as essential infrastructure for preparing the medical workforce to engage safely with AI technologies [36]. Simulation provided a safe environment for experimentation, where medical trainees experienced an AI-assisted handoff workflow firsthand. It also allowed them to gain exposure to the limitations of AI technologies without exposing patients to risk. The findings also informed educators, prompting consideration of how AI assistance could be integrated into other educational modules. Simulation made it possible to observe behavioral and human-factor dynamics, such as how trainees modified their language when they knew a recording technology was in use. Further, it allowed Handoff QI leaders to consider whether to trial commercial ambient AI scribes in real-world handoff workflows and whether such technologies may play supportive roles in these workflows. These findings align with emerging frameworks for AI safety evaluation that emphasize controlled, risk-free environments where failure modes and sociotechnical evaluation can be systematically explored [37].

Clinician-Led Innovation as a Model for Health Care AI Development

This study demonstrates that clinician-led development of AI technologies can produce viable AI proof-of-concept platforms for medical education purposes. This approach contrasts with traditional technology-driven innovation models where external developers create solutions that may not align with educational needs. The participatory design methodology [38,39] used in our co-design interviews exemplifies how domain expertise can be integrated into AI system development for real-world educational needs. In this project, clinicians designed and developed the ambient AI handoff assistant, which enabled them to gain a more realistic understanding of the AI's capabilities and limitations. Furthermore, our experience suggests that clinicians can effectively translate clinical workflows into technical requirements. As an example, the development of the OR-to-ICU handoff note template using prompt engineering in Figure 3 [28], where items in the discussion are mapped to the checklist, demonstrates how clinical expertise can guide AI system design to produce outputs that are clinically meaningful.

Limitations and Future Directions

This study was limited by its exploratory design, small sample size, and simulated environment. Novelty effects and learning curves were evident, as participants included trainees with varying degrees of medical experience and

almost no exposure to the OR-to-ICU handoff process at UCLA. The single-site design and use of UCLA's handoff checklist may also limit generalizability to other institutions with differing handoff checklist structures. However, the methodological framework developed in this study was designed to be adaptable across institutional handoff formats. Further, the error analysis was limited in scope and designed to assess feasibility rather than comprehensively evaluate model performance. Transcription errors that are specific to this ambient AI handoff assistant may not generalize to commercial ambient AI scribe platforms, which differ in accuracy, editing workflows, and user interaction. These differences may influence learners' perceptions of ambient AI, particularly prior to direct exposure to commercial ambient AI scribe systems, and give them an inaccurate impression of how real-world commercial ambient AI scribe technologies work. Additionally, this analysis focused on checklist completeness and did not assess handoff sequence, which may affect comprehension and retention. Completeness alone may not reflect handoff quality; future studies should evaluate whether quantitatively more complete handoffs improve knowledge acquisition and retention among receiving clinicians. Additionally, further evaluation in larger samples and among more experienced trainee populations is needed to assess generalizability, usability, and effectiveness of the handoff assistant in supporting handoff education. Despite these limitations, simulation demonstrated the feasibility of using ambient AI technology to train learners in adhering to institutional checklists, reflecting on their communication practices, and gaining exposure to ambient AI technologies.

Conclusions

This study highlights the potential of an ambient AI handoff assistant to enhance handoff education by converting real-time team dialogue into structured documentation for reflection and feedback. Through participatory co-design and simulation-based evaluation, we identified design requirements for integrating AI into handoff training that preserve authentic team-based communication and enrich reflective learning. The ambient AI handoff assistant demonstrated that complex team-based discussions can be mapped to real-world checklists. Although scripted simulations limited the ability to assess the handoff assistant's impact on promoting improved checklist completeness, the AI-generated note emerged as a powerful learning tool that prompted critical discussion on communication clarity, calibration of trust in AI, and the boundaries of AI automation in human-AI collaboration. Imperfect AI output from the ambient AI handoff assistant encouraged learners to verify and question AI-generated information, fostering critical skills required for the safe use of AI in clinical settings. Taken together, these findings underscore how ambient AI technologies can support reflective, team-based learning while laying the groundwork for interactive AI systems that enhance safety, communication, and education in future clinical workflows.

Acknowledgments

ChatGPT (OpenAI) was used during manuscript preparation to assist with grammar refinement and to improve overall readability. The authors reviewed and approved all content and take full responsibility for the accuracy and integrity of the final text. The authors would like to acknowledge Selene Mak, PhD, and Emily Patterson, PhD, for their suggestions and feedback.

Funding

No external funding was provided.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

All authors meet the ICMJE criteria for authorship. Each made substantial contributions to the conception or design of the work. All authors participated in drafting the manuscript or revising it critically for important intellectual content, and approved the final version to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Needs assessment and co-design interview script with clinician educators.

[\[DOCX File \(Microsoft Word File\), 2491 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Representative prompt examples for the ambient AI assistant.

[\[DOCX File \(Microsoft Word File\), 2548 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Debrief interview guide for simulation learner participants who used the ambient AI handoff assistant.

[\[DOCX File \(Microsoft Word File\), 2489 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Final codebook for educator co-design and learner simulation sessions on the use of the AI in handoff education.

[\[DOCX File \(Microsoft Word File\), 2491 KB-Multimedia Appendix 4\]](#)

Checklist 1

COREQ checklist.

[\[PDF File \(Adobe File\), 113 KB-Checklist 1\]](#)

References

1. Singh H, Thomas EJ, Petersen LA, Studdert DM. Medical errors involving trainees: a study of closed malpractice claims from 5 insurers. *Arch Intern Med*. Oct 22, 2007;167(19):2030-2036. [doi: [10.1001/archinte.167.19.2030](https://doi.org/10.1001/archinte.167.19.2030)] [Medline: [17954795](https://pubmed.ncbi.nlm.nih.gov/17954795/)]
2. Borowitz SM, Waggoner-Fountain LA, Bass EJ, Sledd RM. Adequacy of information transferred at resident sign-out (in-hospital handover of care): a prospective survey. *Qual Saf Health Care*. Feb 2008;17(1):6-10. [doi: [10.1136/qshc.2006.019273](https://doi.org/10.1136/qshc.2006.019273)] [Medline: [18245212](https://pubmed.ncbi.nlm.nih.gov/18245212/)]
3. Kitch BT, Cooper JB, Zapol WM, et al. Handoffs causing patient harm: a survey of medical and surgical house staff. *Jt Comm J Qual Patient Saf*. Oct 2008;34(10):563-570. [doi: [10.1016/s1553-7250\(08\)34071-9](https://doi.org/10.1016/s1553-7250(08)34071-9)] [Medline: [18947116](https://pubmed.ncbi.nlm.nih.gov/18947116/)]
4. Petersen LA, Brennan TA, O'Neil AC, Cook EF, Lee TH. Does housestaff discontinuity of care increase the risk for preventable adverse events? *Ann Intern Med*. Dec 1, 1994;121(11):866-872. [doi: [10.7326/0003-4819-121-11-199412010-00008](https://doi.org/10.7326/0003-4819-121-11-199412010-00008)] [Medline: [7978700](https://pubmed.ncbi.nlm.nih.gov/7978700/)]
5. Odom-Forren J. Accurate patient handoffs: imperative for patient safety. *J Perianesth Nurs*. Aug 2007;22(4):233-234. [doi: [10.1016/j.jopan.2007.06.003](https://doi.org/10.1016/j.jopan.2007.06.003)] [Medline: [17666293](https://pubmed.ncbi.nlm.nih.gov/17666293/)]
6. Nagpal K, Arora S, Vats A, et al. Failures in communication and information transfer across the surgical care pathway: interview study. *BMJ Qual Saf*. Oct 2012;21(10):843-849. [doi: [10.1136/bmjqs-2012-000886](https://doi.org/10.1136/bmjqs-2012-000886)] [Medline: [22773891](https://pubmed.ncbi.nlm.nih.gov/22773891/)]
7. Krishnamoorthy V, Raghunathan K. Anesthesia care handovers and risk of adverse outcomes. *JAMA*. Jun 5, 2018;319(21):2235-2236. [doi: [10.1001/jama.2018.4442](https://doi.org/10.1001/jama.2018.4442)] [Medline: [29872855](https://pubmed.ncbi.nlm.nih.gov/29872855/)]

8. Dutra M, Monteiro MV, Ribeiro KB, Schettino GP, Kajdacsy-Balla Amaral AC. Handovers among staff intensivists: a study of information loss and clinical accuracy to anticipate events. *Crit Care Med*. Nov 2018;46(11):1717-1721. [doi: [10.1097/CCM.0000000000003320](https://doi.org/10.1097/CCM.0000000000003320)] [Medline: [30024429](https://pubmed.ncbi.nlm.nih.gov/30024429/)]
9. Abraham J, Meng A, Sona C, Wildes T, Avidan M, Kannampallil T. An observational study of postoperative handoff standardization failures. *Int J Med Inform*. Jul 2021;151:104458. [doi: [10.1016/j.ijmedinf.2021.104458](https://doi.org/10.1016/j.ijmedinf.2021.104458)] [Medline: [33932762](https://pubmed.ncbi.nlm.nih.gov/33932762/)]
10. Lane-Fall MB, Pascual JL, Peifer HG, et al. A partially structured postoperative handoff protocol improves communication in 2 mixed surgical intensive care units: findings from the handoffs and transitions in critical care (HATRICC) prospective cohort study. *Ann Surg*. Mar 2020;271(3):484-493. [doi: [10.1097/SLA.0000000000003137](https://doi.org/10.1097/SLA.0000000000003137)] [Medline: [30499797](https://pubmed.ncbi.nlm.nih.gov/30499797/)]
11. Catchpole KR, de Leval MR, McEwan A, et al. Patient handover from surgery to intensive care: using Formula 1 pit-stop and aviation models to improve safety and quality. *Paediatr Anaesth*. May 2007;17(5):470-478. [doi: [10.1111/j.1460-9592.2006.02239.x](https://doi.org/10.1111/j.1460-9592.2006.02239.x)] [Medline: [17474955](https://pubmed.ncbi.nlm.nih.gov/17474955/)]
12. Hebballi NB, Gupta VS, Sheppard K, et al. Standardization of pediatric noncardiac operating room to intensive care unit handoffs improves communication and patient care. *J Patient Saf*. Sep 1, 2022;18(6):e1021-e1026. [doi: [10.1097/PTS.0000000000000986](https://doi.org/10.1097/PTS.0000000000000986)] [Medline: [35985048](https://pubmed.ncbi.nlm.nih.gov/35985048/)]
13. Abraham J, King CR, Meng A. Ascertaining design requirements for postoperative care transition interventions. *Appl Clin Inform*. Jan 2021;12(1):107-115. [doi: [10.1055/s-0040-1721780](https://doi.org/10.1055/s-0040-1721780)] [Medline: [33626584](https://pubmed.ncbi.nlm.nih.gov/33626584/)]
14. Schmidt RF, Vibbert MD, Vernick CA, et al. Standardizing postoperative handoffs using the evidence-based IPASS framework through a multidisciplinary initiative improves handoff communication for neurosurgical patients in the neuro-intensive care unit. *J Clin Neurosci*. Oct 2021;92:67-74. [doi: [10.1016/j.jocn.2021.07.039](https://doi.org/10.1016/j.jocn.2021.07.039)] [Medline: [34509265](https://pubmed.ncbi.nlm.nih.gov/34509265/)]
15. Parent B, LaGrone LN, Albirair MT, et al. Effect of standardized handoff curriculum on improved clinician preparedness in the intensive care unit. *JAMA Surg*. May 1, 2018;153(5):464. [doi: [10.1001/jamasurg.2017.5440](https://doi.org/10.1001/jamasurg.2017.5440)] [Medline: [29299602](https://pubmed.ncbi.nlm.nih.gov/29299602/)]
16. Conn Busch J, Wu J, Anglade E, Peifer HG, Lane-Fall MB. So many ways to be wrong: completeness and accuracy in a prospective study of OR-to-ICU handoff standardization. *Jt Comm J Qual Patient Saf*. Aug 2023;49(8):365-372. [doi: [10.1016/j.jcjq.2023.05.001](https://doi.org/10.1016/j.jcjq.2023.05.001)] [Medline: [37316396](https://pubmed.ncbi.nlm.nih.gov/37316396/)]
17. Massa S, Wu J, Wang C, Peifer H, Lane-Fall MB. Interprofessional training and communication practices among clinicians in the postoperative ICU handoff. *Jt Comm J Qual Patient Saf*. Apr 2021;47(4):242-249. [doi: [10.1016/j.jcjq.2020.12.004](https://doi.org/10.1016/j.jcjq.2020.12.004)] [Medline: [33451897](https://pubmed.ncbi.nlm.nih.gov/33451897/)]
18. Lazzara EH, Keebler JR, Simonson RJ, Agarwala A, Lane-Fall MB. Navigating the challenges of performing anesthesia handoffs and conducting anesthesia handoff research. *Int Anesthesiol Clin*. 2020;58(1):32-37. [doi: [10.1097/AIA.0000000000000260](https://doi.org/10.1097/AIA.0000000000000260)] [Medline: [31800413](https://pubmed.ncbi.nlm.nih.gov/31800413/)]
19. Patterson ES, Wears RL. Patient handoffs: standardized and reliable measurement tools remain elusive. *Jt Comm J Qual Patient Saf*. Feb 2010;36(2):52-61. [doi: [10.1016/s1553-7250\(10\)36011-9](https://doi.org/10.1016/s1553-7250(10)36011-9)] [Medline: [20180437](https://pubmed.ncbi.nlm.nih.gov/20180437/)]
20. Gershov S, Braunold D, Spektor R, Ioscovich A, Raz A, Laufer S. Automating medical simulations. *J Biomed Inform*. Aug 2023;144:104446. [doi: [10.1016/j.jbi.2023.104446](https://doi.org/10.1016/j.jbi.2023.104446)] [Medline: [37467836](https://pubmed.ncbi.nlm.nih.gov/37467836/)]
21. Davis J, Roach C, Elliott C, Mardis M, Justice EM, Riesenber LA. Feedback and assessment tools for handoffs: a systematic review. *J Grad Med Educ*. Feb 2017;9(1):18-32. [doi: [10.4300/JGME-D-16-00168.1](https://doi.org/10.4300/JGME-D-16-00168.1)] [Medline: [28261391](https://pubmed.ncbi.nlm.nih.gov/28261391/)]
22. Blackmore A, Kasfiki EV, Purva M. Simulation-based education to improve communication skills: a systematic review and identification of current best practice. *BMJ Simul Technol Enhanc Learn*. 2018;4(4):159-164. [doi: [10.1136/bmjstel-2017-000220](https://doi.org/10.1136/bmjstel-2017-000220)] [Medline: [35519010](https://pubmed.ncbi.nlm.nih.gov/35519010/)]
23. Winkler-Schwartz A, Bissonnette V, Mirchi N, et al. Artificial intelligence in medical education: best practices using machine learning to assess surgical expertise in virtual reality simulation. *J Surg Educ*. 2019;76(6):1681-1690. [doi: [10.1016/j.jsurg.2019.05.015](https://doi.org/10.1016/j.jsurg.2019.05.015)] [Medline: [31202633](https://pubmed.ncbi.nlm.nih.gov/31202633/)]
24. Wartman SA, Combs CD. Medical education must move from the information age to the age of artificial intelligence. *Acad Med*. Aug 2018;93(8):1107-1109. [doi: [10.1097/ACM.0000000000002044](https://doi.org/10.1097/ACM.0000000000002044)] [Medline: [29095704](https://pubmed.ncbi.nlm.nih.gov/29095704/)]
25. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care*. Dec 2007;19(6):349-357. [doi: [10.1093/intqhc/mzm042](https://doi.org/10.1093/intqhc/mzm042)] [Medline: [17872937](https://pubmed.ncbi.nlm.nih.gov/17872937/)]
26. Bate P, Robert G. Experience-based design: from redesigning the system around the patient to co-designing services with the patient. *Qual Saf Health Care*. Oct 2006;15(5):307-310. [doi: [10.1136/qshc.2005.016527](https://doi.org/10.1136/qshc.2005.016527)] [Medline: [17074863](https://pubmed.ncbi.nlm.nih.gov/17074863/)]
27. Sanders EBN, Stappers PJ. Co-creation and the new landscapes of design. *CoDesign*. Mar 2008;4(1):5-18. [doi: [10.1080/15710880701875068](https://doi.org/10.1080/15710880701875068)]
28. Jalilian L, Lukac P, Lane-Fall M. Enhancing structured team communication in acute care settings with ambient AI scribes. *J Am Med Inform Assoc*. Nov 1, 2025;32(11):1785-1789. [doi: [10.1093/jamia/ocaf166](https://doi.org/10.1093/jamia/ocaf166)] [Medline: [41003645](https://pubmed.ncbi.nlm.nih.gov/41003645/)]

29. Barra FL, Rodella G, Costa A, et al. From prompt to platform: an agentic AI workflow for healthcare simulation scenario design. *Adv Simul*. May 16, 2025;10(1):29. [doi: [10.1186/s41077-025-00357-z](https://doi.org/10.1186/s41077-025-00357-z)]
30. Trockel M, Bohman B, Lesure E, et al. A brief instrument to assess both burnout and professional fulfillment in physicians: reliability and validity, including correlation with self-reported medical errors, in a sample of resident and practicing physicians. *Acad Psychiatry*. Feb 2018;42(1):11-24. [doi: [10.1007/s40596-017-0849-3](https://doi.org/10.1007/s40596-017-0849-3)] [Medline: [29196982](https://pubmed.ncbi.nlm.nih.gov/29196982/)]
31. Duff JP, Morse KJ, Seelandt J, et al. Debriefing methods for simulation in healthcare: a systematic review. *Simul Healthc*. Jan 1, 2024;19(1S):S112-S121. [doi: [10.1097/SIH.0000000000000765](https://doi.org/10.1097/SIH.0000000000000765)] [Medline: [38240623](https://pubmed.ncbi.nlm.nih.gov/38240623/)]
32. Eppich W, Cheng A. Promoting Excellence and Reflective Learning in Simulation (PEARLS): development and rationale for a blended approach to health care simulation debriefing. *Simul Healthc*. Apr 2015;10(2):106-115. [doi: [10.1097/SIH.0000000000000072](https://doi.org/10.1097/SIH.0000000000000072)] [Medline: [25710312](https://pubmed.ncbi.nlm.nih.gov/25710312/)]
33. Salwei ME, Carayon P, Hoonakker PLT, et al. Workflow integration analysis of a human factors-based clinical decision support in the emergency department. *Appl Ergon*. Nov 2021;97:103498. [doi: [10.1016/j.apergo.2021.103498](https://doi.org/10.1016/j.apergo.2021.103498)] [Medline: [34182430](https://pubmed.ncbi.nlm.nih.gov/34182430/)]
34. Dzindolet MT, Peterson SA, Pomranky RA, Pierce LG, Beck HP. The role of trust in automation reliance. *Int J Hum Comput Stud*. Jun 2003;58(6):697-718. [doi: [10.1016/S1071-5819\(03\)00038-7](https://doi.org/10.1016/S1071-5819(03)00038-7)]
35. Parasuraman R, Molloy R, Singh IL. Performance consequences of automation-induced "Complacency". *Int J Aviat Psychol*. Jan 1993;3(1):1-23. [doi: [10.1207/s15327108ijap0301_1](https://doi.org/10.1207/s15327108ijap0301_1)] [Medline: [34182430](https://pubmed.ncbi.nlm.nih.gov/34182430/)]
36. Bhutiani M, Hester DL, Lonsdale HJ. Artificial intelligence and the simulationists: more iterations needed. *Simul Healthc*. Aug 1, 2024;19(4):268-269. [doi: [10.1097/SIH.0000000000000803](https://doi.org/10.1097/SIH.0000000000000803)] [Medline: [38922444](https://pubmed.ncbi.nlm.nih.gov/38922444/)]
37. Biro JM, Handley JL, Mickler J, et al. The value of simulation testing for the evaluation of ambient digital scribes: a case report. *J Am Med Inform Assoc*. May 1, 2025;32(5):928-931. [doi: [10.1093/jamia/ocaf052](https://doi.org/10.1093/jamia/ocaf052)] [Medline: [40116912](https://pubmed.ncbi.nlm.nih.gov/40116912/)]
38. Spinuzzi C. The methodology of participatory design. *Tech Commun*. May 1, 2005;52(2):163-174. URL: <https://www.semanticscholar.org/paper/The-Methodology-of-Participatory-Design-Spinuzzi/3cbf8fd37d5502670f61ca06449f14df7a049a51> [Accessed 2026-05-14]
39. Clemensen J, Larsen SB, Kyng M, Kirkevold M. Participatory design in health sciences: using cooperative experimental methods in developing health services and computer technology. *Qual Health Res*. Jan 2007;17(1):122-130. [doi: [10.1177/1049732306293664](https://doi.org/10.1177/1049732306293664)] [Medline: [17170250](https://pubmed.ncbi.nlm.nih.gov/17170250/)]

Abbreviations

AI: artificial intelligence
API: application programming interface
COREQ: Consolidated Criteria for Reporting Qualitative Research
ICU: intensive care unit
LLM: large language model
OR: operating room
PTL: Physician Task Load Index
SUS: System Usability Scale
UCLA: University of California, Los Angeles

Edited by A Hasan Sapci; peer-reviewed by Andrea Vannucci, Md Shadab Mashuk; submitted 24.Oct.2025; final revised version received 23.Feb.2026; accepted 05.Apr.2026; published 02.Jul.2026

Please cite as:

Jalilian L, Barra FL, Grogan T, Lee J, Kadambi A

Exploring Ambient Artificial Intelligence to Enhance Learning and Feedback During Operating Room-to-Intensive Care Unit Handoffs: Co-Design and Simulation Study

JMIR Med Educ 2026;12:e85666

URL: <https://mededu.jmir.org/2026/1/e85666>

doi: [10.2196/85666](https://doi.org/10.2196/85666)

© Laleh Jalilian, Federico Lorenzo Barra, Tristan Grogan, Jason Lee, Achuta Kadambi. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 02.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on <https://mededu.jmir.org/>, as well as this copyright and license information must be included.