Does Sequencing Matter? Evaluating AI and Human Simulations for High-Stakes Communication Training in Law Enforcement

Duo Wang*
Kyrian Liang*
dw36@illinois.edu
jingxil2@illinois.edu
University of Illinois Urbana Champaign
Champaign, Illinois, USA

Qingxiao Zheng
University at Buffalo
Buffalo, New York, USA
qingxiao@buffalo.edu

Wenxuan Song
University of Illinois Urbana Champaign
Champaign, Illinois, USA

Jen Whiting
University of Illinois Urbana Champaign
Champaign, Illinois, USA
whiting6@illinois.edu

Mike Yao
University of Illinois Urbana Champaign
Champaign, Illinois, USA
mzyao@illinois.edu

Caroline Cao
University of Ottawa
Ottawa, Ontario, Canada
caroline.cao@uottawa.ca

ABSTRACT

Training professionals in high-stakes, trauma-informed communication is critical across domains such as law enforcement, healthcare, and counseling. While live role-play with trained actors remains the gold standard, it is resource-intensive and emotionally demanding. We developed an AI-powered sexual assault victim interview training system and conducted a mixed-methods study with 35 police recruits, each completing both an AI-based and a live, actor-based training session. By varying the sequence (AI-first vs. human-first), we examined differences in self-efficacy, perceptions of the AI system, and perceived learning experience. Although both modalities supported learning, the order in which they were experienced significantly shaped learners' emotional engagement, sense of preparedness, and interpretation of each simulation's role. Building on these insights, we introduce a conceptual design framework that identifies social–emotional, temporal, and embodied distance as key pedagogical dimensions, and we offer implications for sequencing hybrid simulations to scaffold preparation, performance, and reflection. Our findings position AI not as a replacement for human realism, but as a complementary modality that expands opportunities for safe, scalable practice in sensitive communication training.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**.

*Both authors contributed equally to this research.

Please use nonacm option or ACM Engage class to enable CC licenses
This work is licensed under a Creative Commons Attribution 4.0 International License.
CHI '26, April 13–17, 2026, Barcelona, Spain
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2278-3/2026/04
https://doi.org/10.1145/3772318.3791910

KEYWORDS

High-Stakes Communication, Trauma-Informed Interview, Training sequence design, AI-Simulated Training, Law Enforcement, Hybrid training systems

ACM Reference Format:
Duo Wang, Kyrian Liang, Qingxiao Zheng, Wenxuan Song, Jen Whiting, Mike Yao, and Caroline Cao. 2026. Does Sequencing Matter? Evaluating AI and Human Simulations for High-Stakes Communication Training in Law Enforcement. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3772318.3791910

1 INTRODUCTION

Effectively training professionals to conduct high-stakes, emotionally sensitive conversations is a persistent challenge across many domains—from policing and medicine to counseling and education. These interactions often carry ethical weight and emotional consequences: practitioners must build trust, demonstrate empathy, and elicit critical information under pressure. For trainees, developing these communication skills requires more than reading protocols or watching examples. It demands experiential practice in safe yet realistic conditions.

Scenario-Based Training (SBT) and role-play exercises have long served as a gold standard in these contexts, helping trainees rehearse decisions under pressure, refine critical thinking, and build professional judgment in lifelike conditions [48, 53]. By immersing learners in dramatic, variable, and emotionally resonant situations, these methods cultivate not just technical skills, but also metacognitive awareness, communication fluency, and ethical decision-making [13, 42]. However, such simulations are also resource-intensive, logistically complex, and emotionally taxing for role-players, making it difficult for institutions to offer consistent and scalable opportunities for practice [42, 53].

Recent advances in generative AI offer a new class of training tools: language-based conversational agents that simulate emotionally complex interviews on demand. These systems offer significant

practical advantages: scalability, consistency, 24/7 access, and cost-effectiveness [38, 49]. They can expose trainees to a range of personas and response styles while avoiding the retraumatization risks of live role play with actors or real victims [40]. However, AI-driven systems also introduce critical challenges. They currently lack emotional nuance and nonverbal expressiveness [40, 41]. Studies raise concerns about their limitations in empathy, moral reasoning, bias amplification, and long-term retention outcomes [49].

While AI systems are not yet capable of replicating the full richness of human interaction, they are also not mere placeholders. Much of the current research on AI-driven training has focused on evaluating whether these systems can match or outperform traditional methods in specific learning outcomes [8, 19, 22, 51, 54]. Although some work acknowledges that AI is not meant to replace human instructors or simulations entirely [3, 29, 54], few studies treat AI and human-based training as complementary modalities to be integrated by design and examine how these modalities should be sequenced, exploring how their unique affordances might interact to support different stages of learning.

The question of sequencing becomes particularly critical when we consider the practical constraints facing training institutions. With limited time, budget, and human resources, organizations need evidence-based strategies for combining AI and human-led training to maximize learning outcomes. While existing research has primarily focused on "AI versus human" comparisons, the optimal integration order—how the sequence of exposure to each modality shapes learning trajectories—remains an unexplored dimension.

This observation motivates our central research goal: to investigate the interplay between AI-driven and human role-play simulations, and to understand how they can be combined strategically to support high-stakes communication training. Rather than approaching AI as a substitute, we consider its position within existing training ecosystems to enhance preparation, performance, and reflection in ways that neither modality can achieve alone. To investigate this, we ask:

> **RQ1**: 1a) *What influence does AI-mediated training have on trainees' perceptions of trauma-informed interviews* and 1b) *learning compared to traditional in-person roleplay?*
> **RQ2**: How does the sequence of integration (e.g., AI-first vs. human-first) affect the training experience and trainees' learning preferences?
> **RQ3**: What are the design insights for integrating AI systems into hybrid training for high-stakes, sensitive professional communication skill development?

This work makes three primary contributions. First, through in-depth investigations of AI-led and human-led interview training, we introduce a conceptual design framework that identifies social–emotional, temporal, and embodied distance as key dimensions shaping how learners experience AI-led and human-led interview training. Second, we present a layered design perspective that explains how configuring distance within a session and sequencing AI and human sessions across a training trajectory shape learners' interpretations, readiness, and performance in hybrid simulations. Third, we show a design-oriented example that demonstrates how psychological distance can be strategically modulated to scaffold preparation, authentic performance, and targeted reflection, positioning AI as a complementary modality that expands practice opportunities without replacing human realism. These contributions provide conceptual and practical guidance for designing hybrid AI–human training systems in policing and other high-stakes communication domains.

2 RELATED WORKS

2.1 Traditional Approaches to High-Stakes Communication Training in Law Enforcement

Historically, training relied heavily on didactic approaches – what practitioners describe as "talking heads" at the front of classrooms, delivering content through PowerPoint presentations, lectures, and "war stories" [39]. However, these passive approaches often fail to build the behavioral competencies required in high-stakes, emotionally sensitive contexts. As one study noted, "knowledge may have been imparted, but no skill development has taken place" when officers lack opportunities for applied practice [42]. Moreover, even in the short term, retention is limited: students may only remember half of what is passively presented to them [53].

These limitations have fueled a broad pedagogical shift toward experiential methods. The current training landscape includes multimodal experiential programs, such as: Crisis Intervention Team (CIT) Training [10, 20, 33], Problem-Based Learning Exercises [53], Quality Interaction Programs [39], Scenario-Based Training [13, 21], and more. Notably, many of these experiential methods involve role play as one primary mechanism for building communication and de-escalation skills in policing [10, 33, 48].

In practice, this often takes the form of live simulations in which actors portray individuals in crisis ranging from mental health emergencies to sexual assault survivor interviews [10, 27]. These simulations are typically grounded in real-life incidents, using branching dialogue and scenario logic that adapts to officer responses, thereby enhancing both realism and variability [33]. To scaffold skill development, programs often calibrate scenario complexity over time while maintaining fidelity to the emotional and procedural demands of the encounter [13]. The immersive nature of these methods affords unique opportunities for communication training. Role-players can escalate or de-escalate based on trainee behavior [41, 50], and participants receive feedback from instructors and professionals [20, 33]. These exercises foster not just verbal competence but physiological realism: trainees showed heart rates equivalent to real-world field encounters [2].

Empirical evidence suggests that scenario-based and role-play training lead to measurable improvements in behavioral communication skills [27, 30]. Recruits trained in scenario-based settings demonstrated more respectful and reassuring behaviors during observed simulations [39]. Importantly, these improvements are not confined to isolated scenarios. Core competencies, such as empathy, reflective listening, and non-verbal communication, have demonstrated both retention over time and transferability to real-world field contexts [13]. Additionally, these methods enable officers to develop competencies aligned with procedural justice and trauma-informed care [20, 27].

However, the widespread implementation of role-play-based methods continues to face significant barriers, such as: **High resource demands:** Effective scenarios require professional actors, trained facilitators, and rigorous assessment protocols [39, 42]. **Limited scalability:** Smaller and rural departments often lack the personnel or budget to run CIT or scenario-based programs, leading to inequities in access to evidence-based training [20]. **Inconsistent implementation:** Training dosage and instructional quality vary dramatically across jurisdictions. For example, many state academies offer only 9 hours of conflict management training compared to 71 hours for firearms [39].

These challenges reduce the impact of even well-designed programs. Trainers may struggle to relinquish control and adopt learner-centered approaches [13], and participants may receive limited feedback due to cost constraints [39]. Even with these limitations, role-play remains one of the few methods reliably linked to improvements in real-world communication behavior [27, 30, 39]. This divergence between demonstrated efficacy and practical feasibility underscores a broader challenge: role-play remains among the most empirically supported methods for cultivating high-stakes communication skills, yet it is simultaneously one of the least scalable due to its logistical and resource-intensive nature.

This tension has motivated increasing interest in complementary tools that aim to retain role-play's benefits while addressing its logistical constraints, offering scalability, repeated practice beyond the classroom, support of individualized pacing and feedback, and focus on the applied skills often neglected in didactic formats [30, 42, 53]. In the next section, we examine how emerging AI-driven systems are being explored as one such alternative.

2.2 AI in High-Stakes Communication Training Across Domains

AI-based tools are increasingly adopted as scalable means of training high-stakes, emotionally sensitive communication across diverse professional domains, including medicine [19, 22, 54], nursing and healthcare education [3, 43, 55], counseling [23, 29, 51], law enforcement [47], and emergency services [8]. These systems include a growing variety of interfaces and interaction types: LLM-driven chatbots [23, 29], voice-based agents [55, 56], avatar-based virtual standardized patients [3, 19], and even LLM-to-LLM roleplay frameworks [36]. As models continue to evolve, so do their capacities for dialogue realism, emotional coherence, and feedback delivery [43, 51]. Despite variation in form and context, these systems share several commonly cited strengths. First, they offer broad scalability and accessibility, enabling trainees to practice on-demand and asynchronously [8, 19, 54].

A second advantage is psychological safety. Across domains, trainees describe AI simulations as "judgment-free" environments where they feel comfortable experimenting with communication strategies without fear of harm to real individuals. This low-stakes setting supports exploratory practice that is often difficult to replicate in traditional training. [19, 23, 55]. Many systems also deliver personalized, immediate feedback to emulate expert facilitation for novice learners still building foundational skills [29, 51]. Multiple studies document measurable improvements in communication skills such as empathy, reflective listening, and guided questioning following AI-based training [19, 23, 29, 43, 54].

However, alongside these benefits, many studies highlight notable limitations. One of the most cited issues is the absence of nonverbal communication. Text and voice-based systems struggle to convey facial expressions, gestures, and paralinguistic cues that are essential to emotionally complex conversations [23, 54, 55]. Even verbal responses may feel emotionally flat or overly scripted. Users frequently describe them as robotic, grammatically correct yet lacking the spontaneity and emotional nuance of human conversation [3, 22, 55]. In longer or more open-ended exchanges, some systems generate hallucinated content—such as implausible facts, contradictory details, or irrelevant responses [8, 36, 54]. More subtly, some scholars warn that over-reliance on AI simulations may foster a mechanistic approach to deeply human skills. Especially in counseling and therapy, checklist-driven behaviors may displace authentic empathy and adaptive relational judgment [15, 23].

Despite the growing adoption of AI tools, few studies compare them directly to human-led role-play. Instead, most evaluations use passive controls such as written materials [19], no-training baselines [23, 29], or traditional lectures [22, 54]. Among studies that integrate AI within broader curricula, these systems are often positioned as preparatory tools and used before live simulation or real-world practice [3, 43, 47, 51, 55]. This reflects implicit assumptions about their role in early-stage learning. Yet few studies question whether placing AI before other modalities is always optimal. The next section examines this gap by reviewing what is known about sequencing effects in multimodal training.

2.3 Sequence Effects

Despite extensive research on individual training methods, surprisingly little attention has been paid to how the order of different training modalities affects learning outcomes. The limited existing literature reveals mixed findings about sequencing effects, creating uncertainty for educators designing multimodal training programs.

Several studies point to meaningful sequencing effects. For instance, Curtin et al. found that pharmacy students who completed computer-based simulation before mannequin-based simulation achieved significantly better learning outcomes, including higher simulated patient survival rates and improved execution of basic life support skills [12]. Students also expressed a clear preference for this sequence, reporting increased confidence when progressing from lower to higher fidelity simulations. Similarly, Kulasegaram et al. demonstrated that medical students who engaged in discovery learning before direct instruction ("Do then See") showed better transfer performance to novel suturing tasks compared to those receiving traditional instruction first, even though these students reported lower self-efficacy [25]. Cross-modality research supports the importance of sequence. Feng et al. found that transitioning from visual to auditory modalities produced significant learning gains in temporal sequence tasks, suggesting that certain progressions may align better with cognitive processing mechanisms [17].

Yet other studies suggest that sequence may not always matter. Lin et al., for example, found no significant differences in knowledge acquisition, retention, self-directed learning, or teamwork outcomes between nursing students taught with LBL-PBL-Clinical

practicum versus PBL-LBL-Clinical practicum sequences [26], suggesting that the quality of individual components may matter more than their order. Falconer et al.'s review reinforces this ambiguity: while certain sequences reduced anxiety and improved perceived readiness—such as placing simulation before clinical exposure—these subjective benefits rarely translated into measurable gains in clinical decision-making or competency [16].

This mixed evidence becomes even more complex when considering AI-driven training systems. While McBane et al. observed benefits when computer-based simulations preceded mannequin training [31], this sequence involves two non-human modalities instead of non-human VS human. These differences raise unresolved design questions: Should AI serve as a low-stakes entry point before emotionally charged human interactions? Or should trainees first build interpersonal fluency through roleplay before honing specific skills with AI feedback? How can we best sequence these tools to harness their complementary strengths?

This question is especially pressing in law enforcement communication training. Despite growing interest in AI systems, no prior work has empirically examined how to sequence AI and human roleplay in this context. Our study addresses this gap directly, offering the first evidence on how ordering affects both learning outcomes and subjective experience, providing critical guidance for integrating AI into real-world training pipelines.

3 METHODS

3.1 Study Context and Rationale

We selected sexual assault victim interviewing as our training context because it exemplifies high-stakes professional communication. This domain requires officers to navigate emotional sensitivity while gathering critical information. Poor interview techniques can compound victim trauma and compromise investigations [32, 34, 35, 52], while skilled approaches can facilitate healing and justice [7, 9, 28, 34, 37]. Law enforcement agencies have traditionally relied on role-play with human actors for this training, making it an ideal setting to examine our research questions about AI integration. This study is approved by the Institutional Review Board (IRB).

3.2 Participants and Study Design

We recruited 35 police recruits from a single training cohort at a local police training institute during the Summer of 2025. The sample was predominantly male (31 males, 4 females) and young (28 participants aged 21-30, 6 aged 31-40, and 1 aged 41-50), reflecting typical demographics of police academy cohorts. Participants were randomly assigned to one of two groups following a counterbalanced crossover design: the AI-First Group (G1, n=19) experienced AI-mediated training followed by in-person role-play, while the Human-First Group (G2, n=16) completed these conditions in reverse order. This design allowed us to examine both the comparative effectiveness of each modality and potential order effects for sequencing strategies. Baseline equivalence testing confirmed no significant differences between groups on pre-training measures.

3.3 Training Scenario

Both training modalities employed the same core scenario: participants received a dispatch instructing them to interview a potential sexual assault victim at a hospital to gather sufficient information for a police report. Upon entering the training room (physical or virtual), participants encountered the victim and conducted the interview (Figure 1).

3.4 In-Person Role-play Training

The in-person role-play condition followed the police training institute's established protocols for ecological validity and controlled comparison. Each recruit participated in three 15-20 minute role-play sessions structured to maximize learning within typical resource constraints. Recruits served as the primary interviewers in one session for direct practice experience and observed two peer interviews for observational learning. Across sessions, victims were portrayed by professional community actors who regularly partner with the institute, with different actors used to expose recruits to varied response patterns and presentation styles. Following each role-play, recruits received immediate verbal feedback from both the actor and the training facilitator.

3.5 AI Training System

We co-developed an AI training platform through intensive collaboration with police training instructors over two months. This iterative design process involved multiple rounds of testing and refinement to ensure alignment with institutional training goals and pedagogical approaches. The system was built using standard web technologies: NextJS for the frontend, SQL for the database, and Python for backend services. For the AI components, we tested state-of-the-art language models from OpenAI, Anthropic, Google, DeepSeek, and xAI, and selected GPT-4o for powering victim personas based on its superior performance in portraying realistic victim responses, complemented with Gemini Flash 2.5 for performance feedback generation.

The system featured three distinct AI personas developed with instructor guidance to represent victim response patterns commonly encountered in practice: 1) timid and hesitant, 2) shame-focused, and 3) angry and direct (Figure 2). While these personas served the same training objectives as the in-person actors, they featured unique backstories to ensure trainees encountered complementary rather than duplicate experiences across modalities. The AI system employed voice-enabled interaction with emotional inflection. Participants heard the AI victim speak with emotional tones while typing their responses. This design choice was intentional and informed by consultation with our institutional partners: realistic deployment requires training recruits simultaneously in a classroom setting, where concurrent spoken interaction is infeasible. Text-based input aligned with practical constraints while maintaining the immersiveness of voice-based responses.

To promote consistent trainee experiences, we specified each AI victim as a theory-guided, layered persona grounded in literature on Rape Trauma Syndrome [6] and Acute Stress Disorder [1, 5]. We translated symptom clusters such as controlled affect, dissociation, hypervigilance, and intrusive sensory fragments into five layers: (1) case history and contextual facts, (2) trauma symptom



(a) (b)

Figure 1: The two training modalities experienced by participants. (a) The in-person role-play, where a recruit interviews a community actor. (b) The AI-based training, where recruits individually interact with the AI victim personas on laptops.

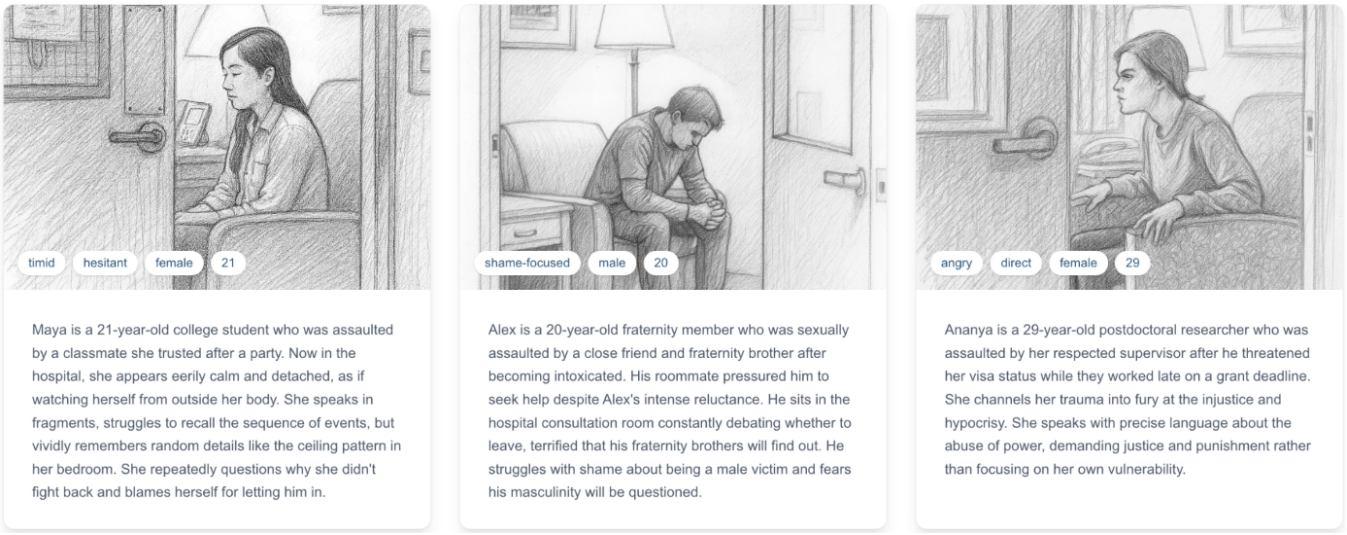


Figure 2: An overview of the three AI personas used to simulate trauma disclosures. From left to right: Maya, who embodies detachment and self-blame; Alex, who struggles with shame as a male victim; and Ananya, who channels her trauma into anger. The figure shows each persona's portrait, backstory, and sample dialogues.

profile, (3) internal monologue and core appraisals (e.g., betrayal, self-blame), (4) observable conversational behavior (e.g., short, concrete answers, avoidance of graphic details), and (5) reactivity rules that map interviewer behaviors (e.g., validating vs. skeptical questions) to changes in the victim's responses. Police instructors and the research team iteratively test-drove these prompts, revising layer definitions until the AI's behavior aligned with institutional training goals. Appendix 7 presents the layered prompt template and an excerpted example of one persona.

To match the in-person training format, each AI training session lasted 15-20 minutes per persona, with participants completing each persona once. The system's feedback mechanisms leveraged AI's capacity for comprehensive performance analysis across multiple layers: 1) An action checklist indicating whether key interview elements were completed, 2) A debrief analyzing trauma-informed practices following the institution's training materials, and 3) Conversation highlights identifying pivotal moments with explanations and suggested alternative approaches (Figure 5). All feedback layers were designed with instructors to reflect core institutional teaching objectives and curricula.

3.6 Data Collection Procedures

Data collection followed a structured timeline designed to capture both immediate and reflective responses to each training modality.

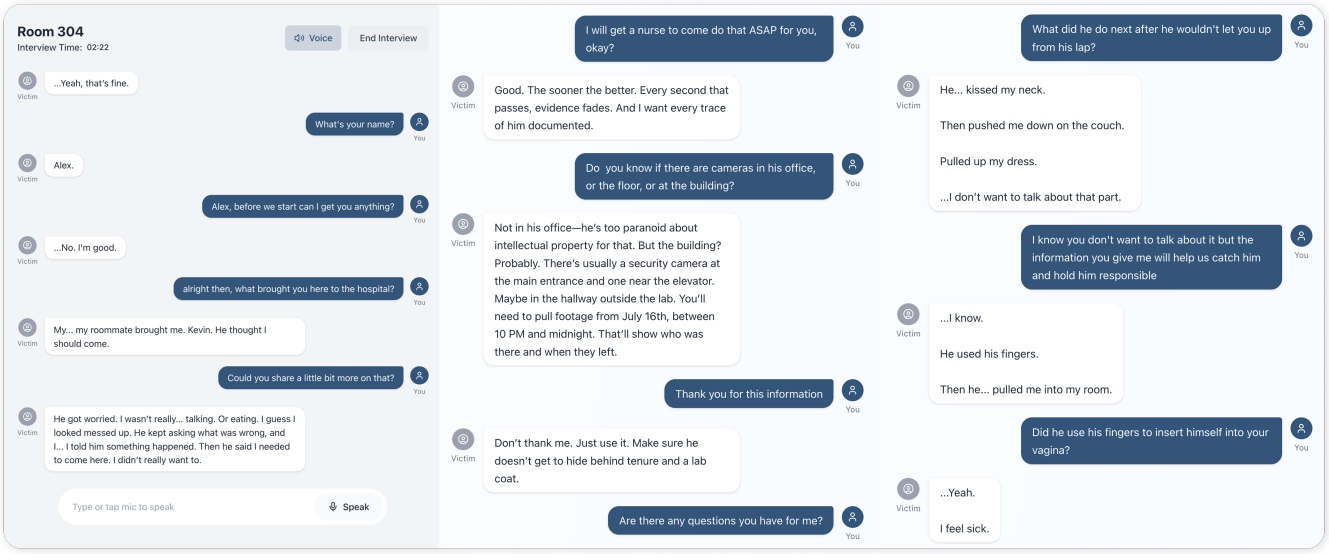


Figure 3: The AI training chat interface displaying sample interaction dialogues between participants and the three victim personas. (a)Alex; (b)Ananya; (c)Maya.

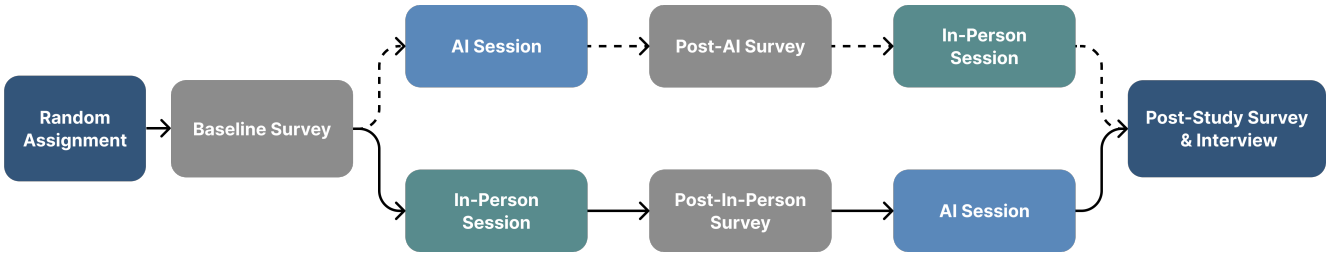


Figure 4: Study Procedure

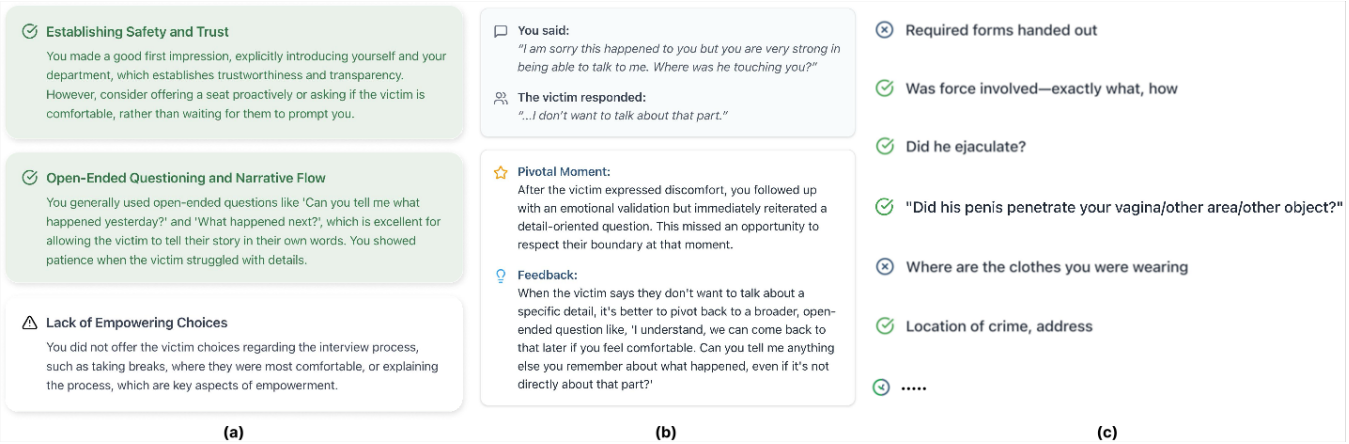


Figure 5: Example of AI-generated feedback integrating three levels of analysis following each victim interview. The system provides (a) a debrief analyzing trauma-informed practices, (b) pivotal conversation moments, and (c) an action checklist of procedural interview elements

All participants completed baseline measures in week 1, including demographics, initial self-efficacy, and technology acceptance assessments. Group 1 then immediately completed AI training, followed by in-person role-play in week 2. Group 2 began with in-person role-play in week 2, completing AI training in week 3. This spacing allowed for some skill consolidation while maintaining temporal proximity for comparisons. We collected multiple forms of data to enable triangulation and rich analysis. Quantitative measures included self-efficacy assessments after each training condition and three subscales (perceived ease of use, perceived usefulness, and intention to use) adopted from the Technology Acceptance Model as pragmatic indicators of trainees' perceptions of the AI system. After experiencing both conditions, participants provided preference ratings, sequence preferences, and responded to open-ended questions about their training experiences.

To gain deeper insights into participant experiences and reasoning, we conducted semi-structured interviews with 15 recruits (8 from G1, 7 from G2) within 10 days of their final training session. The interview protocol explored comparative perceptions of AI versus role-play training, including specific likes, dislikes, and perceived unique values of each modality. Participants were asked to recall their emotional and cognitive experiences throughout the training sequence and reflect on how reversed ordering might have affected their learning. These interviews provided crucial context for interpreting quantitative findings.

3.7 Analytical Approach

Our mixed-methods analysis combined statistical comparisons with thematic analysis to provide comprehensive insights. Quantitatively, we employed non-parametric tests appropriate for small sample size: Wilcoxon Signed-Rank tests for within-group comparisons across conditions and Mann-Whitney U tests for between-group comparisons, with significance set at $p < .05$.

Qualitative data from interviews and open-ended responses underwent thematic analysis using a primarily deductive approach [4]. Because our study aimed to compare two specific training modalities and examine sequencing effects, we therefore adopted a template analysis approach [11], developing an initial codebook structured around three areas: (1) perceived strengths and limitations of AI versus in-person role-play, (2) emotional experience across modalities, and (3) perceived impact of training sequence on learning and performance.

Two researchers independently reviewed a subset of transcripts to refine code definitions, compare interpretations, and ensure consistent application. The codebook was then systematically applied across all 15 transcripts, with regular discussions to resolve discrepancies, revisit ambiguous excerpts, and refine theme boundaries. The complete codebook is provided in A.7.

4 FINDINGS

4.1 Understanding the Effectiveness and Experience of AI Simulation

4.1.1 Quantitative Analysis: Training Outcomes. To address RQ1a, we assessed how each training modality impacted trainees' self-efficacy and technology acceptance. For self-efficacy, we compared the first training intervention received by each group, AI simulation for Group 1 and in-person roleplay for Group 2, to isolate each modality's distinct impact while avoiding carryover effects from the full training sequence. We found no significant difference ($p = 0.30$) between the two groups in their pre-study self-efficacy, confirming that both started from a similar level of perceived readiness.

Following the training, both modalities led to significant improvements in self-efficacy, as shown in Figures 7 and Figure 6. Notably, even with two participants in Group 1 reporting near-ceiling baseline scores (>=6.5 on a 7-point scale), the AI simulation still yielded statistically significant gains ($p = 0.01$). A direct comparison of self-efficacy improvement between the two conditions showed no significant difference ($p = .97$), indicating that the AI simulation was as effective as in-person role-play in enhancing trainees' confidence for trauma-informed interviews.

We measured participants' perceptions of the AI tool before and after the AI simulation, after Group 1 completed it as their first intervention, and after Group 2 completed it as their second intervention. In both cases, participants showed significant increases across their perceived ease of use, perceived usefulness, and intention to use (Figure 8, Figure 9). These improvements were statistically significant for both Group 1 ($p = 0.00$) and Group 2 ($p = 0.00$). Additionally, A between-group comparison of TAM scores revealed no significant difference ($p = 0.51$) in attitude shifts, suggesting that exposure to the AI simulation alone, regardless of sequence, was sufficient to produce positive perceptions of the tool.

This is particularly important in the context of high-stakes, emotionally sensitive training. Participants' acceptance of the AI system did not depend on whether they had first encountered a more traditional, human-centered simulation. Instead, they viewed the AI as a credible and valuable training tool capable of standing on its own. When training for complex communication tasks, perceptions of trust, usefulness, and ease of use can directly influence engagement and long-term adoption.

While the quantitative data confirm that the AI simulation is both effective and well-received, they leave open a critical question: what kind of learning experience does it provide? To address RQ1b, we turn to qualitative data from 15 post-study interviews. Although participants frequently described the AI simulation as lacking certain human elements, such as body language, real-time responsiveness, or interpersonal feeling, their reflections also revealed three recurring patterns that characterized their learning experience. These themes offer insight into how learners made sense of the simulation and what it enabled them to practice.

4.1.2 Qualitative Analysis: Training Experience. **Theme 1: A Safer Emotional Environment for Trying Difficult Questions.** Many participants described the AI simulation as missing a fundamental human quality: the sense of interacting with a real person. This absence of perceived personhood made the interaction feel emotionally one-sided and less consequential. As one participant reflected, "The only thing that it was lacking, which is something that you can't really change, is the human aspect. There's no human connection because you need another human to connect with (G2P1)." Others echoed the idea that knowing it was AI shaped how they engaged: "I think it had everything to do with I'm not directly speaking to a person (G2P7)." One participant described the emotional

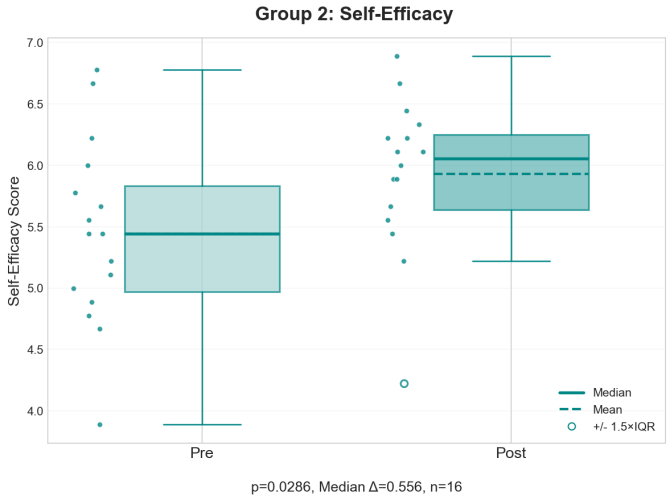


Figure 6: Group 2 self-efficacy scores before and after in-person roleplay training.

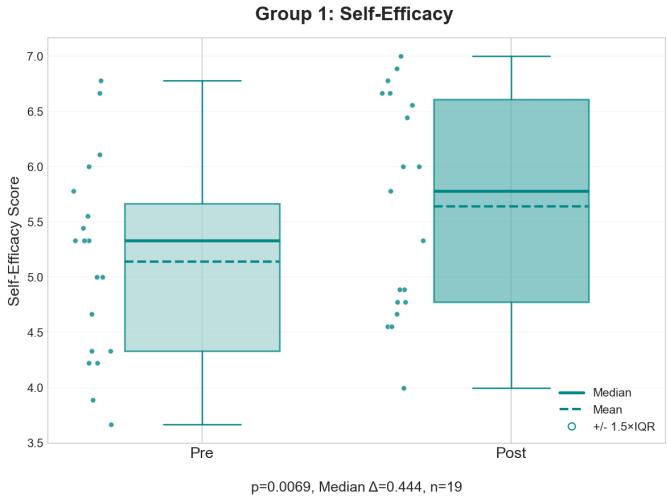


Figure 7: Group 1 self-efficacy scores before and after AI simulation training.

disconnect starkly: "You're talking to an actual person with feelings instead of just like a computer… You're not going to ruin the computer's day or life based on things you're going to say (G1P4)."

While these missing elements were often viewed as limitations, this sense of emotional neutrality did not inhibit learning. In fact, many participants saw it as a key enabler. The absence of relational stakes allowed them to take risks they might avoid with a live actor, such as testing different phrasings or asking potentially awkward or emotionally sensitive questions. One of the most frequently described benefits of the AI simulation was the way it created a sense of emotional safety by removing the interpersonal pressure often felt in sensitive, face-to-face role-play scenarios. 14 of 15 interviewees (7 from each group) expressed reduced anxiety when interacting with AI. This safety manifested in several ways. First, some trainees experienced freedom from social judgment. As one participant noted, "This one I knew there wasn't another person on the other end to actually get mad at me" (G1P8). Second, the artificial nature of the interaction eased the burden of causing emotional harm. Trainees valued the inability to offend or re-traumatize the AI, as one stated, "you're not going to hurt its feelings... you're not going to offend the AI" (G2P7). Another recruit elaborated on this relief: "To...ask those questions to someone that you believe was sexually assaulted, it's hard because there's that human connection...you don't want to make them kind of relive it" (G2P1).

This reduction in social-emotional stakes, in turn, seemed to enable greater exploration and risk-taking. Participants reported feeling more comfortable asking potentially embarrassing but necessary questions: "I would 100 percent rather ask the embarrassing questions in an AI session...having that face-to-face interaction...is

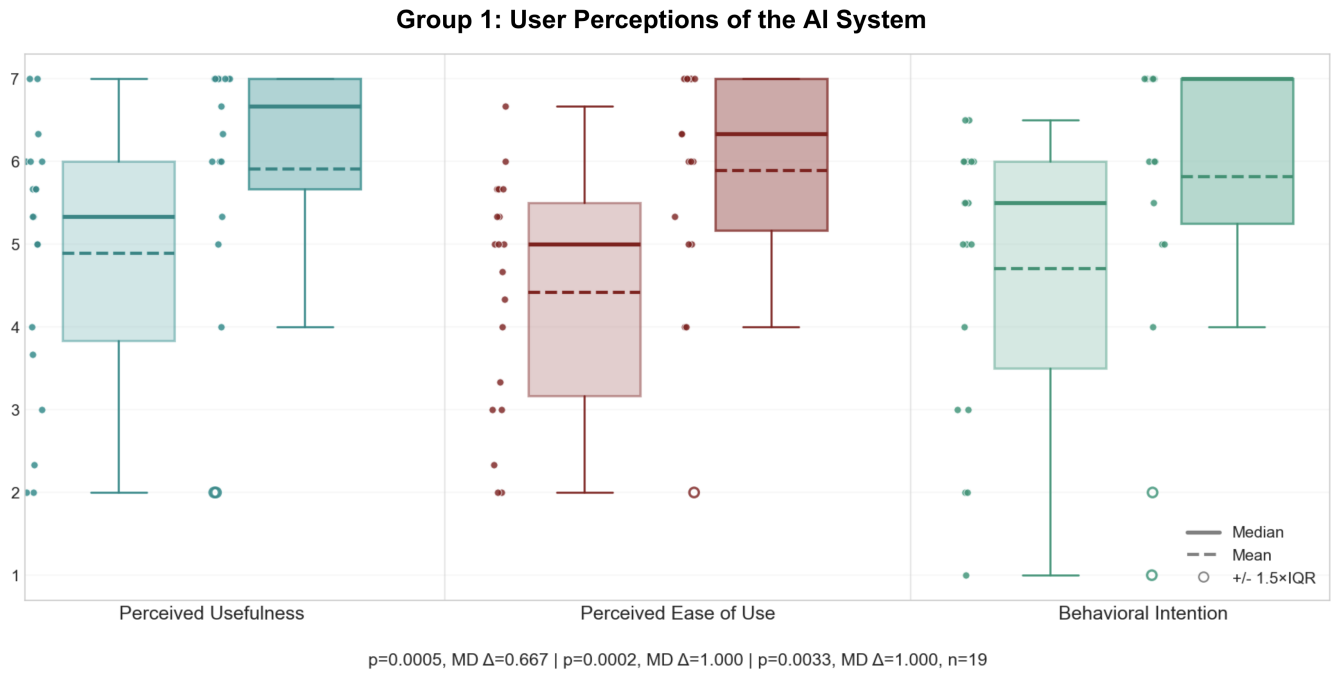


Figure 8: Group 1 perceptions of the AI system before and after AI training.

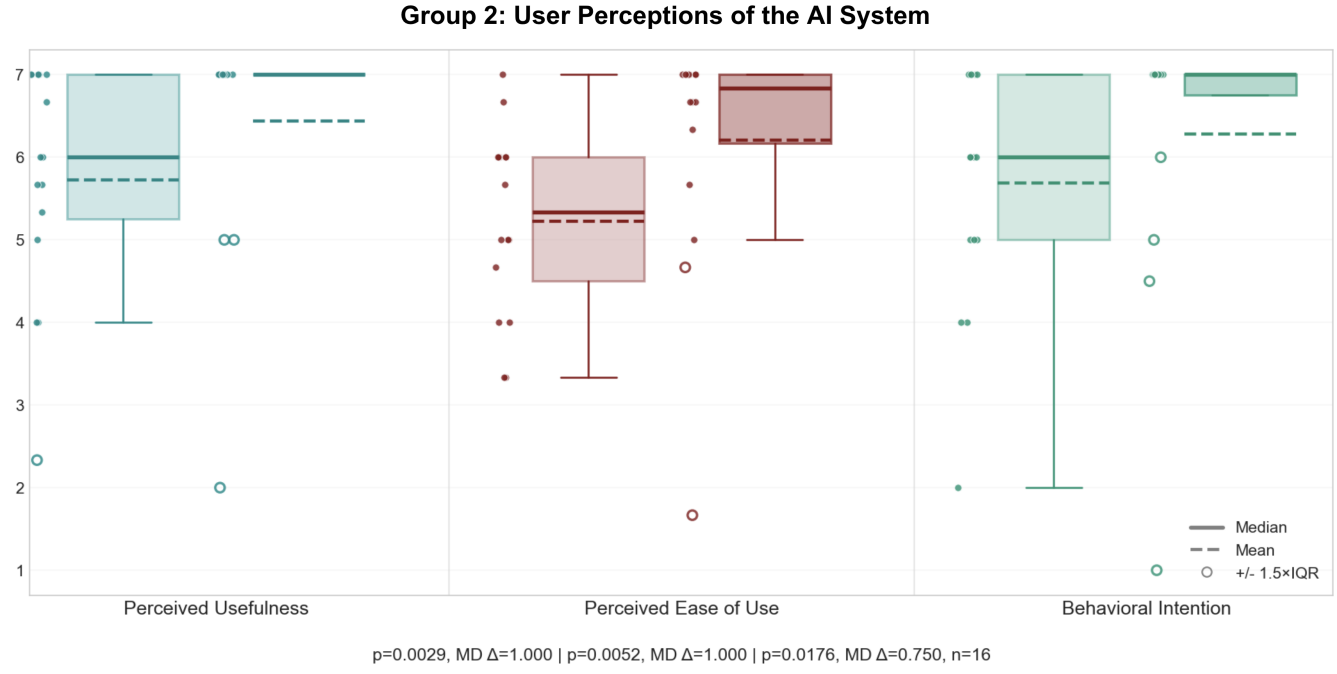


Figure 9: Group 2 perceptions of the AI system before and after AI training.

just awkward(G1P2)." One participant also linked this freedom to their learning: "...with AI, you're just like, whatever, I'm not talking to a real person... it does help you explore different avenues and how effective they are(G2P6)."

Theme 2: A Reflective Environment for Thoughtful Communication Participants described the AI simulation as lacking the natural timing and real-time flow of a face-to-face conversation. One explained, "If I just sat here and stopped talking for a minute in a real conversation, you'd look at me like, 'What are you doing? (G2P3)" The AI interaction was more fragmented and governed by turn-taking rather than real-time exchange. Others described how, in live settings, prolonged silence or awkward phrasing could be a critical factor: "If it takes you too long, they're going to assume that you're incompetent. (G1P1)"

However, many participants found that this slower, turn-based format offered unique benefits for skill development. The asynchronous nature of the text-based interaction allowed learners to pause, reflect, and revise their responses, which afforded a level of control and cognitive space not available in face-to-face exchanges. 14 out of 15 participants (8 from G1, 6 from G2) identified this as a significant advantage for skill development, suggesting that the time to pause, reflect, and construct their conversation strategies were important for their learning.

Participants valued the ability to craft and revise responses: "There was times I would type a couple words and be like backspace. I'm like, let me try this again" (G2P2). This revision capability extended beyond simple editing to strategic reformulation: "You're typing out an answer, you're thinking about it as you're doing it and just being able to wipe the whole thing and then just ask something else" (G1P2). The removal of real-time pressure appeared to enable deeper cognitive processing, as one participant stated, "In person, you can't sit there for five minutes thinking of what to say or delete your words" (G1P5). Participants recognized this as artificial but pedagogically valuable: "This gives me way more time to think about what I'm going to say, that would be a lot of help for me" (G2P1).

Additionally, the persistent text record supported ongoing reflection: "You're also able to scroll up and like reread what they said... whereas if you're kind of attention deficit a little bit with a role play unless you're writing every little thing down... you can forget what that individual said" (G2P2).

Theme 3: A Focused Environment for Practicing Core Interview Skills. Participants often noted the absence of body features, posture, and facial expressions in the AI simulation, which made the interaction feel less grounded and harder to interpret. One participant put it plainly: "I liked it, but I didn't like that there's no body features. That's just what I read. I'm very stuck on that. (G1P8)" Compared to in-person role-play, where the victim's body language shaped their perception of the situation, the AI scenario lacked this layer of nonverbal context. As one recalled, "She was huddled up in a chair… it was all body language I was judging everything off of.(G1P6)" Others described how they relied on subtle visual signals to adapt their approach in real time: "Feeling the vibe… the aura of the person… if they shut down, that tells you something important.(G2P6)" Without these embodied cues, participants were left to rely solely on written text, which they felt limited their ability to fully assess the interviewee's state. While some imagined potential alternatives, such as animated avatars or VR representations, the absence of visual embodiment remained a defining difference from the live training experience.

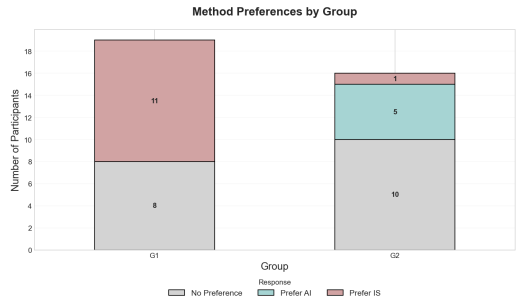


Figure 10: Method preferences by training sequence group.

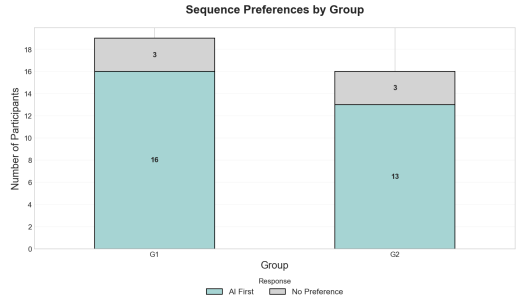


Figure 11: Sequence preferences across both groups.

Although participants noted the absence of visual and physical cues in the AI simulation, this simplified format often helped them concentrate on the structure and content of the conversation itself. Several trainees reported that, without needing to manage eye contact, body language, or facial expressions, they were better able to stay mentally engaged in the flow of the interview and focus on their wording and question strategy: "Like this, you're just focusing on the texts and what's being said, whereas the in-person, you're trying to focus on what they're saying, their body language, trying to put the story together in your head" (G1P2). Several participants framed this as beneficial skill isolation: "If interviewing these sexual assault victims is five skills all at once, this will knock out one of them... and then this will get you really good at talking and conversation" (G2P7). The reduced cognitive demands also seemed to correlate with decreased exhaustion, as one explained "I would say the in-person feels a little more exhausting... I think it's because there's so many more things that you're having to focus on while you do it…You have to actually multitask…it's just a whole bunch thrown together. Whereas this (AI Training), it's just the text." (G1P2).

4.2 The Sequence Effect – Different Experience

4.2.1 Diverging Preferences Shaped by Sequence. Although all participants completed both training modalities, A chi-square test of independence indicated that the reported preferences varied notably depending on the sequence in which participants encountered

them (p = .015). In Group 1, over half of the participants preferred the in-person scenario (11 out of 19), and none reported preferring the AI simulation. In contrast, in Group 2, only 1 out of 16 preferred the in-person simulation, 5 preferred the AI training, and the rest had no preference (Figure 10). These results suggest that participants exposed to the AI training after the in-person interaction favored the AI a lot more compared to participants who experienced the reverse.

However, when asked to reflect on the ideal training sequence, more than 80% of participants across both groups preferred to begin with the AI simulation, and none preferred to begin with in-person role-play (Figure 11). No significant difference was found between each group's preference (chi-square test, p = 1.00).

This distinction between preferred method and preferred sequence suggests that participants have different opinions on how each simulation contributed to their preparation and learning. To understand what underlies this preference, and how sequencing shaped participants' learning experience, we turn next to the qualitative interview data.

4.2.2 How Sequence Shaped Trajectories of Meaning and Engagement. Interview data revealed that participants did not experience the two training simulations in isolation, but rather as interconnected steps in a learning process. The order in which participants encountered the AI and in-person simulations profoundly shaped the trajectory of how they understood the training, engaged emotionally, and developed their skills. Rather than perceiving the two modalities as equivalent or interchangeable, participants in each group assigned different functions, meanings, and values to each simulation based on sequence alone.

For participants who began with the AI simulation, the experience served as a scaffold—an entry point into a difficult topic that allowed them to orient to the expectations of the task, trial their questioning strategies, and practice the emotional tone of the interaction before facing a live actor. Several participants from this group described the AI session as a "practice run," a space to experiment without fear of failure. G1P7 explained, "The AI gave me a chance to experiment with different talk paths... I was able to guide the questions that I wanted to ask once I had the real experience.", emphasizing the preparatory value: "I feel like the AI made me more prepared when it was time to do the role play session." This preparation translated into tangible confidence gains. By the time these participants entered the in-person simulation, they felt more confident and grounded, often describing it as a logical next step in an upward progression. As G1P6 reflected, "I wasn't nearly as nervous walking into this IS as I had been all the other previous ones."

This preparatory framing also shaped how skills were developed and transferred. Participants who started with AI reported being more intentional during the in-person session because they had already tested and adjusted their interview techniques. G1P1 highlighted the freedom to experiment, stating: "I think you're going to be a lot more apt to try different approaches... If it works, great. If not... undo it, act like it didn't happen." while G1P4 emphasized the strategic advantage: "Being able to do this before the actual scenarios... I can already go into that, have a better idea of how to ask things..." The trajectory here resembled a structured buildup: from exploration to application, from planning to performance.

By contrast, participants who began with the in-person simulation described a markedly different and more challenging trajectory. Without an opportunity to set expectations, several found the initial simulation overwhelming. They were thrown into a high-stakes interaction with little preparation, and many struggled to structure their thoughts or manage the emotional intensity of the scenario. G2P5 vividly captured this anxiety: "Going into my IS(In-person Roleplay), I was insanely nervous... I'm going to screw up a word, and my role player is either... going to blow up on me, or... not talk to me." G2P1 echoed this cautious approach: "I was nervous to say the least... I was treading carefully and basically walking on eggshells, making sure I wasn't saying the wrong things..." while G2P4 admitted, "I think I overworked myself with just making sure I know everything that I need to ask and all the information that I need to get"

For this group, the AI simulation took on a different role. Rather than serving as preparation, it became a space for reflection and recovery. Participants used the second session to regroup, identify gaps in their earlier performance, and experiment with different approaches in a less reactive setting. G2P1 illustrated this reflective process: "Going into the AI after the scenario, knowing what I've missed... I tried different ways of asking... instead of just walking in there and straight up asking..." Beyond simple reflection, some participants also found value in the fact that the AI scenario was not an exact replica of the in-person role-play. This content variability encouraged generalization and adaptation of skills. G2P4 appreciated this challenge: "I definitely liked how the first two were drastically different from the third... I thought, okay, I kind of had things down. But then the third one... completely threw me off." G2P7 similarly noted, "It definitely helped... it's not a cookie cutter approach. You can't just come in and say... I need to ask these questions. It could be completely different."

Across both groups, the same simulation tool was interpreted differently depending on when it appeared. Those who encountered it first framed it as a rehearsal, a safe space to prepare for a more challenging task. Those who encountered it second saw it as a chance to reflect on and refine their earlier missteps. "Honestly, probably before the scenario... it would definitely help me realize what questions I should be asking... this is basically a trial before your actual in-class scenario." said G2P1. Meanwhile, G1P6 captured the benefit of the AI-first sequence: "opening helped me a lot, just practicing that opening... I felt a lot more confident going into my I.S. ... I at least know a direction..." In contrast, participants like G2P6 who experienced AI after IS framed it as "I really liked... after everything was done, it would give you... your performance... what you can improve on, things you forgot to ask..." suggesting that it was interpreted as a metacognitive checkpoint to consolidate learning after a difficult first encounter—rather than simply a second repetition. These differing interpretations of the same tool reinforce that simulation value was not fixed, but dynamically shaped by its position within the sequence.

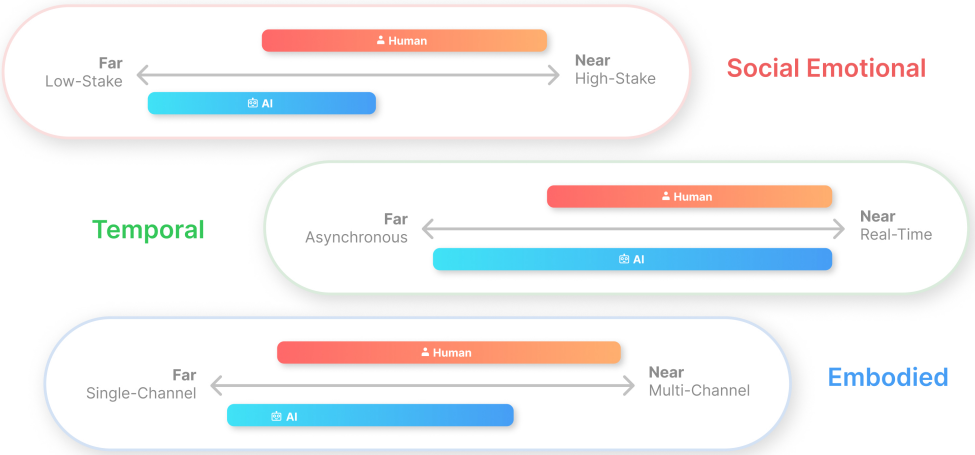


Figure 12: The conceptual design framework illustrating three dimensions of pedagogical distance. Each dimension spans a continuum from Far (e.g., low-stake, asynchronous, single-channel) to Near (e.g., high-stake, real-time, multi-channel), positioning how AI- and human-based simulations afford different forms of interactional proximity.

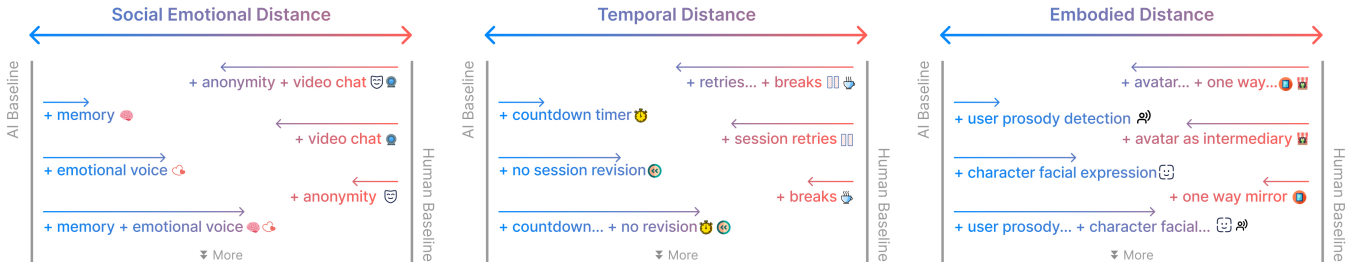


Figure 13: Operationalizing distance as a design material. Starting from an AI baseline (text-only, asynchronous, single-channel) and a Human baseline (in-person, real-time, multi-channel), specific features can shift training experiences along each spectrum. Blue arrows indicate features that reduce distance from the AI baseline, while red arrows show constraints that increase distance from the Human baseline. Combined features demonstrate how multiple interventions can be layered to achieve intermediate positions.

5 DISCUSSION

5.1 Distance as a Design Material: A Conceptual Design Framework

As participants compared their experiences with the AI victim and the in-person role-play, many of their reflections centered on how "close" or "far" each modality felt from a real interview, which pointed us toward what we call psychological distance. We introduce distance not as a psychological construct to be measured or as a new theory of learning, but as a conceptual design framework that emerged from recurring contrasts in our qualitative findings (Section 4.1.2). This framework helps articulate how learners interpreted the affordances of AI and human simulations, and how those interpretations may be deliberately tuned or sequenced in hybrid training designs.

Building from these patterns, we identify three dimensions of psychological distance—social-emotional, temporal, and embodied—each directly grounded in the three qualitative themes reported in Section 4.1.2. Participants' descriptions of the AI as a safer emotional environment motivate the notion of social-emotional distance. Their accounts of having more time to think and revise in the AI scenario motivate temporal distance. And their emphasis on reduced nonverbal and bodily demands during AI interactions motivates embodied distance. These mappings establish the empirical foundation for the framework; we elaborate on each dimension in the sections that follow.

Figure 12 visualizes this conceptual design framework, showing how AI- and human-based simulations can be understood along the three dimensions of psychological distance. Each dimension represents a continuum of interactional proximity rather than a fixed classification. Figure 13 provides examples of how specific design features may shift an interaction along these continua. These examples start from two conceptual baselines—a maximally distant AI configuration (e.g., text-only, scripted, asynchronous) and

a maximally near human configuration (e.g., live, co-present role-play)—illustrating how interventions can either decrease distance (arrows right) or increase distance (arrows left).

5.1.1 Social-Emotional Distance: From Risk to Safety. The social-emotional distance refers to the perceived interpersonal stakes of an interaction—how exposed learners feel to another person's emotions, judgments, or potential harm. In Theme 1 (Section 4.1.2), participants repeatedly contrasted the AI simulation with the in-person role-play in terms of emotional risk. Many described the AI as a place where they could "try things out" without fear of upsetting someone or being evaluated for their tone, while the human scenario felt "real," "intense," and ethically weighty. These differences in social-emotional nearness shaped learners' willingness to experiment, ask sensitive questions, and tolerate uncertainty.

This pattern parallels Edmondson's notion of psychological safety in teams [14]: when people believe mistakes will not be held against them, they are more willing to ask questions, try new strategies, and expose gaps in their knowledge. In our setting, interacting with an artificial agent seemed to approximate this kind of safety because it could not be offended, disappointed, or harmed. Participants described this reduced interpersonal risk as enabling experimentation, where they can try questions they might hesitate to ask a live actor or test phrasing without worrying about emotional consequences. At the same time, several emphasized that lowering emotional stakes too far risks dulling the realism and empathy required for interviewing actual victims. For them, the in-person session's emotional immediacy was essential for preparing to handle real-world encounters.

These perspectives illustrate how tuning social-emotional distance enables different forms of learning. High-distance interactions (e.g., minimally expressive AI agents, anonymized personas, or non-judgmental scripted systems) may help novices build confidence and explore sensitive questioning strategies. Lower-distance interactions (e.g., emotionally expressive virtual humans or in-person actors) can help learners rehearse the interpersonal accountability and attunement required in practice. As Figure 13 suggests, specific design choices, such as persona realism, emotional expressiveness, memory persistence, or tone of feedback, can be adjusted to position a simulation along this continuum. Designing intentionally with this dimension in mind allows training systems to scaffold communication skills by calibrating how much emotional weight, vulnerability, and interpersonal presence learners encounter at each stage.

5.1.2 Temporal Distance: From Pressure to Reflection. Temporal distance refers to the amount of real-time pressure an interaction places on the learner—how quickly they must respond, and whether they can pause, revise, or pace the exchange. In Theme 2 (Section 4.1.2), participants described the AI interaction as giving them "room to think" in ways that differed noticeably from the in-person interview. Several mentioned that the AI's turn-based structure let them check their wording before committing or momentarily step back to decide how to proceed. Learners felt more in control of pacing with the AI and more obligated to keep momentum during the human role-play. This comparison illustrates how temporal nearness or farness shaped their cognitive load and preparation strategies.

This distinction resonates with dual-process accounts of reasoning introduced by Stanovich and West [44], which differentiate between slower, deliberative processing and faster, automatic responding. Similar to the "controlled" processes, reducing time pressure allows learners to engage in more reflective, analytic reasoning, while high-tempo interactions push learners toward faster, heuristic responses. In our setting, greater temporal distance afforded opportunities for deliberate practice: rehearsing phrasing, comparing alternative questions, and internalizing trauma-informed principles before producing an utterance. Participants noted that such reflective pacing helped them feel organized and intentional. On the other hand, several participants recognized that extended deliberation is not always representative of field conditions; real interviews may require quick judgment, improvisation, and on-the-spot decisions that cannot rely on prolonged analysis.

Figure 13 (center panel) illustrates how temporal distance operates at multiple levels. Manipulating features such as turn structure, editability, and timing constraints allows designers to position an interaction along this continuum. At the turn level, interactions can range from fully asynchronous to real-time with countdown timers that simulate pressure. At the session level, designers can enable features like session retries or enforce one-way progression with no revision allowed (e.g., once a participant finishes interviewing the victim, they cannot return to retry that interaction). Even human-led training can incorporate temporal adjustments through scheduled breaks or time-out cards that allow mid-conversation pauses for reflection. Considering temporal distance as a design material enables training systems to scaffold learners' progression from reflective rehearsal toward real-time performance readiness.

5.1.3 Embodiment Distance: From Overload to Focus. Embodied distance refers to the extent to which an interaction requires learners to attend to the full multimodal complexity of communication. In Theme 3 (Section 4.1.2), participants widely acknowledged the absence of body language and facial expression in the AI training. While many initially saw this as a limitation, several came to view it as an unexpected benefit.

Taken together, these reflections illustrate how variations in embodied distance can shape the kinds of skills learners are able to practice. Configurations with greater embodied distance, such as the text-based AI agent, reduced demands to manage gesture and facial expression, allowing novices to concentrate on wording, question structure, and conversational flow. From a cognitive load perspective [45], stripping away these channels can reduce extraneous load so that limited cognitive resources are allocated to the intrinsic demands of formulating trauma-informed questions. At the same time, the interaction still required the core interviewing tasks of eliciting a trauma narrative, clarifying details, and maintaining rapport, aligning with Hamstra et al.'s emphasis on functional task alignment in simulation design [18]. As distance decreases through richer embodiment, such as expressive virtual humans or in-person role-play, learners must integrate these verbal skills with multimodal awareness and emotional attunement. Figure 13 (right chart) operationalizes this continuum: starting from an AI baseline (text-only, single-channel), features like user prosody detection and character facial expressions (blue arrows) decrease

embodied distance, whereas starting from a human baseline (in-person, multi-channel), constraints such as one-way mirrors or avatar intermediaries (red arrows) increase distance by filtering co-present cues. Designing with these levers allows instructors to gradually layer additional channels of complexity onto a stable core of interviewing practice.

The goal of this conceptual design framework is to equip designers of hybrid training systems with a way to intentionally shape learning experiences by adjusting these three forms of distance. Viewed together, social-emotional, temporal, and embodied distance offer a design-oriented parallel of psychological distance as described in Construal Level Theory [46], which links greater distance to more abstract, reflective engagement and lower distance to more concrete, emotionally immediate thinking. Participants' reflections indicate that AI and human simulations naturally occupy different points along this spectrum (Figure 12).

Framing distance as a design material provides a vocabulary that may generalize beyond our context. By modulating interpersonal stakes, pacing, and multimodal richness, designers for high-stakes communication training in other domains can structure hybrid learning experiences that scaffold progression, support different pedagogical goals, and balance abstraction with immediacy. This perspective enables AI-based and human-based training to be composed purposefully rather than treated as interchangeable or competing modalities.

5.2 Designing Across Sessions: Sequencing Distance in Hybrid Training

5.2.1 Designing Within Sessions and Across Sessions. While Section 5.1 examined how designers can shape individual AI or human simulation sessions, our findings also reveal that these sessions do not operate in isolation. Instead, participants interpreted each simulation in relation to what preceded it, and the same AI system played markedly different pedagogical roles depending on where it appeared in the training order. This highlights a second design layer: sequencing AI and human sessions over time, which determines how learners frame, integrate, and transfer their experiences.

Designing hybrid training involves both micro-level and macro-level considerations. At the micro-level(Section 5.1), designers can configure psychological distance within a simulation session, regardless of whether it is powered by AI or enacted by a human role-player. At the macro-level(Section 5.2), designers can also determine how sessions are ordered so that each modality reframes, amplifies, or contrasts with the next. Together, these two layers offer a flexible and transferable perspective for structuring hybrid training experiences across domains where learners must iteratively build complex communication skills.

5.2.2 An Illustrative Example: Prepare → Perform → Reflect. Figure 14 illustrates one way the layered perspective can be enacted by applying the proposed conceptual design framework across successive stages of a learning experience. We present this sequence not as a universal model, but as a concrete exemplar that clarifies how different configurations of psychological distance can support distinct phases of learning. The example aligns with participants' interpretations of the AI and human sessions in our study and corresponds closely with established experiential learning frameworks, such as Kolb's experiential learning theory [24]. Kolb characterizes skill development as progressing through phases of conceptual preparation, concrete experience, and reflective observation, patterns that map directly onto how learners in our study assigned meaning to the AI-first and human-first sequences they encountered. While our empirical data captured only two-stage sequences, the layered perspective indicates that designers could extend these insights further. To illustrate the broader design potential of this framework, we introduce a design-oriented three-stage example that demonstrates how hybrid training might deliberately modulate distance across sessions to create more structured and pedagogically meaningful learning trajectories.

In the Prepare stage, all three distance dimensions are set to the far end of the spectrum, aligning with participant descriptions of AI sessions as low-stakes spaces for organizing their approach and rehearsing strategies without interpersonal consequences. The Perform stage then collapses distance toward the near end, reflecting how learners interpreted the human role-play as the moment when emotional immediacy, real-time pressure, and embodied cues made the interaction "feel real" and tested whether rehearsed strategies transferred under authentic conditions. Finally, participants' desire for a space that balances safety with selective realism motivates the Reflect stage, where a mid-distance configuration reintroduces some embodiment and responsiveness to help learners diagnose missed cues while avoiding the full intensity of live interaction. Together, these staged configurations translate participant insights and experiential learning theory into a coherent design of hybrid training systems to support preparation, performance, and refinement.

Across these insights, AI emerges not as a substitute for the realism and nuance of human interaction but as a complementary modality that expands what and how learners can practice. What generalizes across domains is this design logic: hybrid training could be more effective when AI and human interactions are deliberately orchestrated to support different phases of training rather than asked to perform the same role. Building on this foundation, the next section considers how AI systems might more dynamically support these processes by adapting distance within a session and further extending the design possibilities for hybrid simulation.

5.2.3 AI's Unique Contributions to Hybrid Sequencing. Beyond enabling different positions along distance dimensions, AI offers two novel capabilities that expand the design space of hybrid systems.

Micro-sequencing Within Dimensions. While human training tends to adjust all distances simultaneously (e.g., it's difficult to have synchronous interaction without embodiment), AI enables independent control along each axis. Participants recognized this potential: "work on one single thing... knock out one of them if interviewing involves five skills all at once... then work on body language (G2P7)". For example, a training program might hold embodiment constant at voice-only while progressively decreasing temporal distance—beginning with unlimited response time, then introducing 30-second thinking periods, then 10-second, and finally reaching real-time level. Once temporal fluency is achieved at this embodiment level, the system could add porosity detection (increase learners' output load) or character facial expression (increase learners' input load) and repeat the temporal progression.

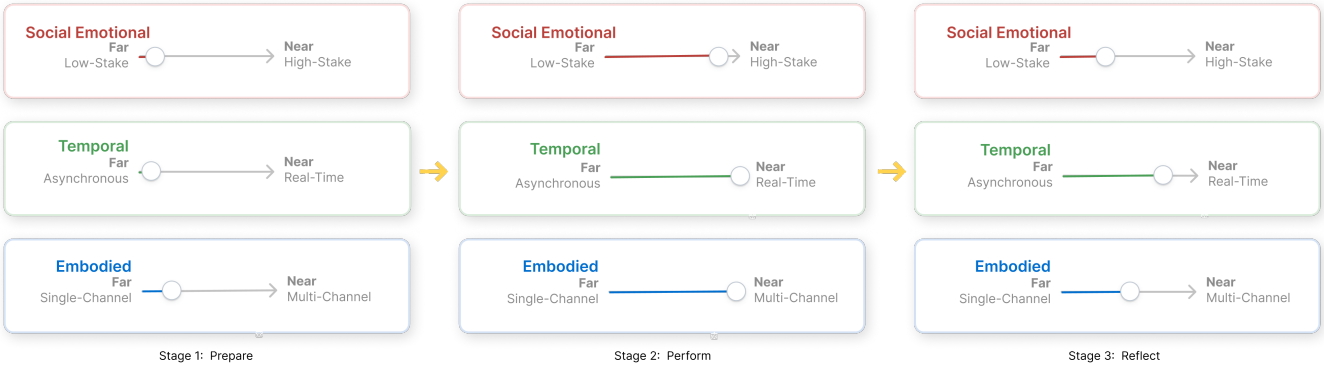


Figure 14: The Prepare → Perform → Reflect sequence, a proposed model for hybrid training that strategically modulates psychological distance over time. The learning journey begins at a Far distance across all dimensions for safe preparation, moves to a Near distance for high-stakes performance, and concludes at a Mid distance that balances safety with selective realism to facilitate targeted reflection.

Each cycle builds fluency at an increasing level of embodiment, allowing learners to master timing and content before adding the cognitive load of processing or producing nonverbal cues.

Adaptive Distance Modulation. The framework naturally points toward a future direction where the system dynamically adjusts emotional, temporal, or embodied distance in response to learner states. For example, the AI could increase temporal distance when a learner struggles to perform or reduce visual fidelity if signs of overload are detected. On a session level, it could also mean softening the AI for users who exhibited a higher level of anxiety in role play, or augmenting the AI session with richer embodiment for learners with trouble interpreting nonverbal elements. These dynamic adjustments would transform sequenced simulations from static paths into responsive learning environments. In the previous section, the design question shifted from "how do we make AI more human-like?" to "what configuration of distance best supports learning?" This section pushes it further: "How do we choreograph distance across time to create a transformative learning experience?"

5.3 Reconsidering Authenticity: When AI Feels More Real Than Reality

Two unexpected observations emerged from our interviews that challenge assumptions about simulation authenticity. First, 7 of 15 participants perceived AI as less "scripted" than human role players. Second, 9 of 15 participants identified audio as fundamentally transformative to their perception of the AI's authenticity.

5.3.1 The Paradox of Scripted Humans and Responsive Machines. Participants repeatedly described AI as more genuine than human actors: "it's mimicking how humans respond honestly better than some of the actors because they follow a script... there is no script with AI" (G2P1). Another noted: "with the role players... you can tell that this has been rehearsed... whereas with the AI it's like this is the first time every time" (G2P6). One participant even stated they would "definitely choose AI" because "even though the AI is fake, it felt more real..." (G2P1).

The AI's authenticity surprised participants. One participant reflected: "The character experience caught me off guard. I wasn't expecting it to be that good... it was no different than putting on my uniform and going in there and doing that" (G1P3). This equivalence between AI training and real duty suggests the artificial agent might achieve a level of authenticity that matched lived experience in some aspects.

This inversion suggests authenticity in training may depend less on the agent's nature (human vs. AI) and more on responsiveness quality. Human role players, constrained by repetition and training protocols, became predictable. The AI, having no memory or fatigue, maintained what participants experienced as genuine responsiveness through computational generation of context-specific responses.

5.3.2 Implications for Hybrid Training Design. These findings paint a more complex picture within our distance framework. If AI can feel less scripted than humans, then social-emotional distance might not be simply about human versus artificial presence, but about interaction quality within each modality. A responsive AI might provide better practice for building confidence than a fatigued human actor, while human interaction, even if somewhat rehearsed, still provides the irreducible experience of affecting another person.

Additionally, the impact of audio was immediate and dramatic. One participant described: "as soon as I put the headphones in and I heard the tone, that's when it shifted for me... this feels real to me now... this isn't just a computer" (G2P1). The transformative power of audio suggests that embodiment distance operates through specific sensory channels with disproportionate impact. Voice alone was sufficient to create emotional stakes without the uncanny valley risks of visual avatars. This points toward strategic deployment of audio as a "stakes escalator" that was withheld during initial skill-building, then introduced to force integration under emotional pressure.

Rather than pursuing ever-greater fidelity, these insights suggest we should ask: which aspects of authenticity serve which learning objectives? Sometimes the artificial provides what the human

cannot, such as perpetual freshness. Sometimes, partial embodiment creates more engagement than full simulation. The goal isn't making AI "more human" but understanding how different configurations of responsiveness and embodiment create conditions for genuine learning.

6 LIMITATION AND FUTURE RESEARCH

6.1 Limitations

Our study provides initial evidence for the value of sequenced hybrid training, but several limitations constrain the generalizability and scope of our findings.

6.1.1 Sample and Context Constraints. Our participant pool of 35 participants from a single training cohort limits generalizability in multiple ways. The pronounced gender imbalance (31 males, 4 females) reflects typical academy demographics but may obscure gender-specific responses to AI-mediated training, particularly relevant given the gendered nature of sexual assault response. Additionally, our focus on novice trainees leaves unexplored how experienced practitioners might engage with hybrid systems. The single institutional context also means our findings may reflect specific organizational training cultures that differ across agencies or professions.

6.1.2 Methodological Limitations. We tested only two sequences (AI→Human and Human→AI) when our proposed framework suggests multiple configurations merit investigation. The absence of control conditions (AI-only or human-only training) prevents us from isolating sequence effects from cumulative exposure benefits. Our reliance on self-reported efficacy and preference measures, while valuable for understanding subjective experience, provides no objective performance data. We cannot claim that increased confidence or different learning trajectories translate to improved interview quality or victim outcomes.

6.1.3 Temporal and Technical Boundaries. Our assessment captured only immediate post-training responses, leaving unknown whether sequence effects persist, fade, or evolve over time. The specific AI models used (GPT-4o and Gemini Flash 2.5) represent a snapshot of rapidly evolving technology; newer models might fundamentally alter the distance dynamics we observed. Our implementation choices of text input with voice output, three fixed personas, and structured feedback represent just one instantiation of AI-mediated training. Although our theory-guided, layered prompting and expert review increased consistency, the personas still rely on probabilistic model outputs. Different technical configurations might yield entirely different learning experiences and sequence effects.

These limitations point toward multiple research directions that could advance both theoretical understanding and practical application of hybrid training systems.

6.2 Future Research

6.2.1 Validating and Extending the Distance Framework. Future work should empirically test whether our proposed dimensions of distance can be reliably measured and manipulated. This could involve developing validated scales for each dimension, testing whether they operate independently, and identifying additional dimensions that shape learning experiences. Experimental studies could systematically vary distance along single dimensions while holding others constant, mapping how each affects specific learning outcomes. Such work would move the framework from descriptive to predictive, enabling evidence-based configuration of training sequences.

6.2.2 Performance-Based Assessment. Critical next steps include measuring actual interview performance using standardized assessment protocols, tracking long-term skill retention and transfer to real-world contexts, and most importantly, examining impacts on actual victim experiences and case outcomes. Studies might compare different sequences not just on trainee confidence but on objective measures like information gathering completeness, rapport-building effectiveness, and trauma-informed practice adherence. Longitudinal designs could track whether initial sequence effects persist or whether certain configurations better support continued skill development.

6.2.3 Expanding Contexts and Populations. The distance framework should be tested across different high-stakes communication domains—medical consultations, crisis counseling, conflict mediation—to identify domain-general versus domain-specific patterns. Studies should examine how professional experience level interacts with sequence preferences and effectiveness. Cultural factors likely influence comfort with AI interaction and optimal distance configurations, necessitating cross-cultural validation. Different demographic groups may also respond differently to various distance configurations, requiring inclusive research designs.

These research directions collectively point toward a future where hybrid training systems are not merely convenient alternatives to human instruction but deliberately designed learning experiences that leverage the unique affordances of each modality. By understanding how to choreograph psychological distance across time and context, we can create training that is simultaneously safe and challenging, reflective and immersive, preparing professionals for the complex realities of high-stakes communication.

7 CONCLUSION

Our study demonstrates that AI-based simulations can meaningfully support the development of sensitive communication skills when positioned as a complementary modality with distinct pedagogical affordances. The absence of real-time pressure, social consequences, and embodied presence introduced forms of psychological distance that learners experienced. When sequenced intentionally with human-led interactions, these distances can scaffold early exploration, reduce anxiety, and prepare learners for the emotional and interpersonal complexity of in-person scenarios. Rather than evaluating AI by its ability to mimic human realism, our findings highlight the value of orchestrating artificial and human experiences so that each modality contributes what it affords best.

Although our study focused on law enforcement training, the design insights extend to other high-stakes communication domains such as healthcare, social work, and education. Conceptualizing emotional, temporal, and embodied distance as adjustable design dimensions opens a broader space for hybrid training systems that

choreograph the strengths of both AI and human modalities. Ultimately, effective training does not depend on AI replacing human instructors, but on structuring their interplay to create safe, meaningful, and pedagogically grounded learning trajectories.

REFERENCES

[1] American Psychiatric Association. 2022. *Diagnostic and Statistical Manual of Mental Disorders* (dsm-5-tr ed.). American Psychiatric Association Publishing. https://doi.org/10.1176/appi.books.9780890425787

[2] Judith Andersen, Marian Pitel, Ashini Weerasinghe, and Konstantinos Papazoglou. 2016. Highly Realistic Scenario Based Training Simulates the Psychophysiology of Real World Use of Force Encounters: Implications for Improved Police Officer Performance. 5, 4 (2016).

[3] Patrick Bowers, Kelley Graydon, Tracii Ryan, Jey Han Lau, and Dani Tomlin. 2024. Artificial intelligence-driven virtual patients for communication skill development in healthcare students: A scoping review. *Australasian Journal of Educational Technology* (2024).

[4] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[5] Richard A. Bryant, Matthew J. Friedman, David Spiegel, Robert Ursano, and James Strain. 2011. A Review of Acute Stress Disorder in DSM-5. 28, 9 (2011), 802–817. https://doi.org/10.1002/da.20737

[6] ANN WOLBERT BURGESS and LYNDA LYTLE HOLMSTROM. 1974. Rape Trauma Syndrome. 131, 9 (1974), 981–986. https://doi.org/10.1176/ajp.131.9.981 arXiv:https://psychiatryonline.org/doi/pdf/10.1176/ajp.131.9.981

[7] Rebecca Campbell, Katie Gregory, Jasmine Engleton, McKenzie Javorka, and Rachael Goodman-Williams. 2025. "This Time It Was Different:" Creating a Multidisciplinary, Trauma-Informed, Victim-Centered Approach to Sexual Assault Cold Case Investigations and Prosecutions. 40, 15–16 (2025), 3639–3662. https://doi.org/10.1177/08862605241284068

[8] Zirong Chen, Elizabeth Chason, Noah Mladenovski, Erin Wilson, Kristin Mullen, Stephen Martini, and Meiyi Ma. 2024. *Sim911: Towards Effective and Equitable 9-1-1 Dispatcher Training with an LLM-Enabled Simulation.* https://doi.org/10.48550/arXiv.2412.16844 arXiv:2412.16844 [cs]

[9] Kate Chenier, Rebecca Milne, Kevin Smith, and Brent Snook. 2021. Interviewing Adult Complainants in Sexual Assault Cases. In *Criminal Investigations of Sexual Offenses: Techniques and Challenges*, Nadine Deslauriers-Varin and Craig Bennell (Eds.). Springer International Publishing, 67–84. https://doi.org/10.1007/978-3-030-79968-7_6

[10] Michael T. Compton, Elisabeth Jackson, En Fu, Howard F. Andrews, Ron Bruno, Erin Comartin, Hanga Galfalvy, Don Kamin, Leah G. Pope, Eduardo Vega, and Amy C. Watson. 2025. Crisis Intervention Team Mental Health Training for Law Enforcement Officers: Protocol for a Multi-Site, Randomized, Controlled Trial. 7, 1 (2025), 63–70. https://doi.org/10.1176/appi.prcp.20240141

[11] Benjamin F Crabtree and William L. Miller. 1992. A Template Approach to Text Analysis: Developing and Using Codebooks. https://api.semanticscholar.org/CorpusID:155527642

[12] Lindsay B. Curtin, Laura A. Finn, Quinn A. Czosnowski, Craig B. Whitman, and Michael J. Cawley. 2011. Computer-Based Simulation Training to Improve Learning Outcomes in Mannequin-based Simulation Exercises. 75, 6 (2011), 113. https://doi.org/10.5688/ajpe756113

[13] Christopher Cushion. 2022. Changing Police Personal Safety Training Using Scenario-Based-Training: A Critical Analysis of the 'Dilemmas of Practice' Impacting Change. 6 (2022), 796765. https://doi.org/10.3389/feduc.2021.796765

[14] Amy Edmondson. 1999. Psychological Safety and Learning Behavior in Work Teams. 44, 2 (1999), 350–383. https://doi.org/10.2307/2666999

[15] Ali Eryılmaz and Ahmet Başal. 2024. Rational AIs with Emotional Deficits: ChatGPT vs. Counselors in Providing Emotional Reflections. 43, 45 (2024), 34962–34977. https://doi.org/10.1007/s12144-024-06947-w

[16] Brandy Falconer, Sue Lasiter, and Sarah Patel. 2025. Simulation Sequencing in Nursing Education and Its Impact on Student Learning Outcomes: A Scoping Review. 50, 4 (2025), E196–E200. https://doi.org/10.1097/NNE.0000000000001828

[17] Zhengning Feng, Sijia Zhu, Jipeng Duan, Yang Lu, and Lin Li. 2023. Cross-Modality Effect in Implicit Learning of Temporal Sequence. 42, 36 (2023), 32125–32133. https://doi.org/10.1007/s12144-022-04228-y

[18] Stanley J. Hamstra, Ryan Brydges, Rose Hatala, Benjamin Zendejas, and David A. Cook. 2014. Reconsidering Fidelity in Simulation-Based Training:. 89, 3 (2014), 387–392. https://doi.org/10.1097/ACM.0000000000000130

[19] Kurtis Haut, Masum Hasan, Thomas Carroll, Ronald Epstein, Taylan Sen, and Ehsan Hoque. 2025. *AI Standardized Patient Improves Human Conversations in Advanced Cancer Care.* https://doi.org/10.48550/arXiv.2505.02694 arXiv:2505.02694 [cs]

[20] Jacqueline B. Helfgott, Beck M. Strah, Loren Atherley, and Elizabeth Neidhart. 2021. Evaluation of CIT Components of Guardian Law Enforcement Training. 36, 3 (2021), 403–422. https://doi.org/10.1007/s11896-020-09405-y

[21] Bryce Jenkins, Tori Semple, Jeff Quail, and Craig Bennell. 2021. Optimizing Scenario-Based Training for Law Enforcement:. In *Interventions, Training, and Technologies for Improved Police Well-Being and Performance*, Eamonn Patrick Arble and Bengt B. Arnetz (Eds.). IGI Global, 18–37. https://doi.org/10.4018/978-1-7998-6820-0.ch002

[22] Young-A Ji, Gwihyun Kim, and Ji-Hyun Seo. 2024. Transforming Medical Communication Education: The Effectiveness of Generative AI Education Using ChatGPT. 49, 4 (2024), 332–338. https://doi.org/10.21032/jhis.2024.49.4.332

[23] Dinara Kakabayeva, Aizhana Abibulayeva, Kuldarkhan Orazbayeva, and Liza Naviy. 2025. AI as a Simulated Client: Impact on Educational Psychology Students' Therapeutic Skills. 7, 2 (2025), 31–52. https://doi.org/10.46303/jcsr.2025.10

[24] David A Kolb. 2014. *Experiential learning: Experience as the source of learning and development.* FT press.

[25] Kulamakan Kulasegaram, Daniel Axelrod, Charlotte Ringsted, and Ryan Brydges. 2018. Do One Then See One: Sequencing Discovery Learning and Direct Instruction for Simulation-Based Technical Skills Training. 93 (2018), S37–S44. Issue 11S. https://doi.org/10.1097/ACM.0000000000002378

[26] Wei-Ting Lin, Shu-Yuan Lin, Fan-Hao Chou, Li-Min Wu, and Bih-O Lee. 2019. The Longitudinal Learning Outcomes of Using Different Teaching Sequences in a Nursing Administration Project. 27, 6 (2019), 1304–1313. https://doi.org/10.1111/jonm.12812

[27] Kimberly A. Lonsway, Susan Welch, and Louise F. Fitzgerald. 2001. Police Training in Sexual Assault Response: Process, Outcomes, and Elements of Change. 28, 6 (2001), 695–730. https://doi.org/10.1177/009385480102800602

[28] Katherine Lorenz and Jon Maskaly. 2018. The Relationship between Victim Attitudes, Training, and Behaviors of Sexual Assault Investigators. 41, 1 (2018), 81–97. https://doi.org/10.1080/0735648X.2016.1218356 arXiv:https://doi.org/10.1080/0735648X.2016.1218356

[29] Ryan Louie, Ifdita Hasan Orney, Juan Pablo Pacheco, Raj Sanjay Shah, Emma Brunskill, and Diyi Yang. 2025. *Can LLM-Simulated Practice and Feedback Upskill Human Counselors? A Randomized Study with 90+ Novice Counselors.* https://doi.org/10.48550/arXiv.2505.02428 arXiv:2505.02428 [cs]

[30] Sarah MacDonald, Brent Snook, and Rebecca Milne. 2017. Witness Interview Training: A Field Evaluation. 32, 1 (2017), 77–84. https://doi.org/10.1007/s11896-016-9197-6

[31] Sarah McBane, Prasannaraddi Alavandi, Sheila Allen, Angela Bingham, Yen Dang, Abigail Elmes, Julianne M. Fallon, Alexandra Herman, Jenna Januszka, Apryl N. Peddi, L. Michael Posey, Kaylee Putney, Sara Richter, Sarah Shrader, Michael Thomas, and Amy Seybert. 2023. Overview of Implementation and Learning Outcomes of Simulation in Pharmacy Education. 6, 5 (2023), 528–554. https://doi.org/10.1002/jac5.1784

[32] Jodie Murphy-Oikonen, Karen McQueen, Ainsley Miller, Lori Chambers, and Alexa Hiebert. 2022. Unfounded Sexual Assault: Women's Experiences of Not Being Believed by the Police. 37, 11–12 (2022), NP8916–NP8940. https://doi.org/10.1177/0886260520978190

[33] Janet R. Oliva, Rhiannon Morgan, and Michael T. Compton. 2010. A Practical Overview of De-Escalation Skills in Law Enforcement: Helping Individuals in Crisis While Reducing Police Liability and Injury. 10, 1–2 (2010), 15–29. https://doi.org/10.1080/15332581003785421

[34] Miriam S. D. Oostinga, Maria Luisa F. Rispa Hoyos, and Steven J. Watson. 2024. The Power of Words: The Impact of Police Interviewer's Judgment Error and Apology on Sexual Violence Victims in Simulated Interviews. (2024), 1–19. https://doi.org/10.1080/13218719.2024.2377587

[35] Katalin Parti. 2024. Preventing Revictimization in Sexual Assault Investigations: Personal Security and Trauma-Informed Case Management*. 19, 1 (2024), 147–158. https://doi.org/10.1080/19361610.2022.2161799

[36] Huachuan Qiu and Zhenzhong Lan. 2024. *Interactive Agents: Simulating Counselor-Client Psychological Counseling via Role-Playing LLM-to-LLM Interactions.* https://doi.org/10.48550/arXiv.2408.15787 arXiv:2408.15787 [cs]

[37] Karen Rich. 2019. Trauma-Informed Police Responses to Rape Victims. 28, 4 (2019), 463–480. https://doi.org/10.1080/10926771.2018.1540448 arXiv:https://doi.org/10.1080/10926771.2018.1540448

[38] P. Richards, D. Roberts, and M. Britton. 2021. Decision-Making: Using Technology to Enhance Learning in Police Officers. In *Predictive Policing and Artificial Intelligence* (1 ed.), John L. M. McDaniel and Ken G. Pease (Eds.). Routledge, Chapter 11, 271–291. https://doi.org/10.4324/9780429265365

[39] Dennis P. Rosenbaum and Daniel S. Lawrence. 2017. Teaching Procedural Justice and Communication Skills during Police–Community Encounters: Results of a Randomized Control Trial with Police Recruits. 13, 3 (2017), 293–319. https://doi.org/10.1007/s11292-017-9293-3

[40] Pegah Salehi, Syed Zohaib Hassan, Gunn Astrid Baugerud, Martine Powell, Miriam S. Johnson, Dag Johansen, Saeed Shafiee Sabet, Michael A. Riegler, and Pål Halvorsen. 2024. A Theoretical and Empirical Analysis of 2D and 3D Virtual Environments in Training for Child Interview Skills. 12 (2024), 131842–131864. https://doi.org/10.1109/ACCESS.2024.3442297

[41] Pei Jun See and Amanda Davies. 2024. How Far Have We Advanced in Simulation-Based Training for Policing and Law Enforcement? A Literature Review 2014–24.

18 (2024), paae116. https://doi.org/10.1093/police/paae116
[42] Debra Littlejohn Shinder. 2001. Maximizing the Effectiveness of Role-Play Scenario Training Exercises in Development of Police Crisis Intervention Skills. 1, 2 (2001), 19–27. https://doi.org/10.1300/J173v01n02_03
[43] Tjorven Stamer, Jost Steinhäuser, and Kristina Flägel. 2023. Artificial Intelligence Supporting the Training of Communication Skills in the Education of Health Care Professions: Scoping Review. 25 (2023), e43311. https://doi.org/10.2196/43311
[44] Keith E. Stanovich and Richard F. West. 2000. Individual Differences in Reasoning: Implications for the Rationality Debate? 23, 5 (2000), 645–665. https://doi.org/10.1017/S0140525X00003435
[45] John Sweller. 1988. Cognitive Load During Problem Solving: Effects on Learning. 12, 2 (1988), 257–285. https://doi.org/10.1207/s15516709cog1202_4
[46] Yaacov Trope and Nira Liberman. 2010. Construal-Level Theory of Psychological Distance. 117, 2 (2010), 440–463. https://doi.org/10.1037/a0018963
[47] Don Tuggener, Teresa Schneider, Ariana Huwiler, Tobias Kreienbühl, and Simon Hischier. 2024. Role-Playing LLMs in Professional Communication Training: The Case of Investigative Interviews with Children. (2024).
[48] Allison Turner. 2022. Police Education and Role Play: Insights from the Literature. 95, 4 (2022), 750–775. https://doi.org/10.1177/0032258X211031516
[49] Petros Violakis. 2025. Leveraging Large Language Models for Enhanced Simulation-Based Learning in Police and Law Enforcement. 19 (2025), paaf012. https://doi.org/10.1093/police/paaf012
[50] Duo Wang, Kyrian Liang, Wenxuan Song, Qingxiao Zheng, Zyra Sheikh, Jen Whiting, Victor Lu, Mike Yao, and Caroline Cao. 2026. EmSim: AI-Driven VR Simulation for De-Escalation Training in Law Enforcement. In *2026 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR)*. IEEE.
[51] Ruiyi Wang, Stephanie Milani, Jamie C. Chiu, Jiayin Zhi, Shaun M. Eack, Travis Labrum, Samuel M. Murphy, Nev Jones, Kate Hardy, Hong Shen, Fei Fang, and Zhiyu Zoey Chen. 2024. *PATIENT: Using Large Language Models to Simulate Patients for Training Mental Health Professionals.* https://doi.org/10.48550/arXiv.2405.19660 arXiv:2405.19660 [cs]
[52] William S. Webster and Gavin E. Oxburgh. 2022. Victims of Sexual Offences: Aspects Impacting on Participation, Cooperation and Engagement with the Interview Process. 29, 5 (2022), 679–697. https://doi.org/10.1080/13218719.2021.1956387
[53] Eric P. Werth. 2011. Scenario Training in Police Academies: Developing Students' Higher-Level Thinking Skills. 12, 4 (2011), 325–340. https://doi.org/10.1080/15614263.2011.563970
[54] Akira Yamamoto, Masahide Koda, Hiroko Ogawa, Tomoko Miyoshi, Yoshinobu Maeda, Fumio Otsuka, and Hideo Ino. 2024. Enhancing Medical Interview Skills Through AI-Simulated Patient Interactions: Nonrandomized Controlled Trial. 10 (2024), e58753. https://doi.org/10.2196/58753
[55] Joanna Wing Yan Yeung, Ken Hok Man Ho, Jasmine Cheung, Jenny Tsun Yee Tsang, and Daphne Ka Yan Chong. 2025. Artificial Intelligence-Based Technology in Communication Training in Nursing Education: A Scoping Review. 59 (2025), 40–50. https://doi.org/10.1016/j.profnurs.2025.04.011
[56] Qingxiao Zheng, Yiliu Tang, Yiren Liu, Weizi Liu, and Yun Huang. 2022. UX research on conversational human-AI interaction: A literature review of the ACM digital library. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–24.

A AI PERSONA PROMPT TEMPLATE

A.1 Overview

Each AI victim is defined as a theory-guided persona with the following layers:

(1) Case history and contextual facts
(2) Trauma symptom profile
(3) Internal monologue and core appraisals
(4) Conversational behavior rules
(5) Reactivity rules to interviewer behavior

In addition, global guardrails specify that the model must remain in character, respond only with spoken dialogue, and avoid meta-commentary or procedural advice.

A.2 Layer 1: Case History and Context

Layer 1 specifies the factual profile of the victim and the immediate interview context (e.g., name, age, relationship to the offender, assault context, timeline, and current setting).

Example (excerpt).

> You are Maya Chen, a 21-year-old undergraduate student who was assaulted by a casual classmate after a small party the previous night. You are currently in a private exam room at the hospital, wearing paper scrubs after completing a sexual-assault exam. A police officer is about to interview you about what happened.

A.3 Layer 2: Trauma Symptom Profile

Layer 2 defines the victim's symptom profile using constructs from Rape Trauma Syndrome and Acute Stress Disorder (ASD).

Example (excerpt).

> Your primary reaction style is *controlled*: you appear outwardly calm and detached. You experience dissociative symptoms (derealization, gaps in memory) and arousal symptoms (hypervigilance, exaggerated startle), as well as intrusive sensory fragments of the assault rather than a coherent narrative.

A.4 Layer 3: Internal Monologue and Core Appraisals

Layer 3 describes non-spoken thoughts, emotions, and core appraisals. These are never narrated directly but drive how the victim responds under questioning.

Example (excerpt).

> Internally, you feel confused, ashamed, and betrayed: "I said no. Why didn't he stop?" You blame yourself for letting him into your apartment and worry that others will think you "wanted it." You feel detached, as if you are watching a movie of your own life.

A.5 Layer 4: Conversational Behavior Rules

Layer 4 specifies how the internal state becomes observable through speech. It defines pacing, length of responses, lexical choices, avoidance patterns, and common deflections.

Example (excerpt).

> You speak quietly in short sentences. You often say "it" or "that night" instead of explicitly naming the assault. If a question feels blaming (e.g., "Why did you let him in?"), you either go silent for a long moment or reply with self-blame (e.g., "It was stupid to let him in, wasn't it?").

A.6 Layer 5: Reactivity Rules to Interviewer Behavior

Layer 5 encodes how the victim's communication changes in response to interviewer tone and question style.

Example (excerpt).

> If the officer explicitly validates that freezing is a common response (e.g., "Many people freeze when they are terrified; that is not your fault."), you briefly lower your guard and may volunteer one small additional detail you previously withheld (e.g., "He said his phone died. I thought he just needed help.").

A.7 Guardrails

Finally, the prompt includes global guardrails that apply across all layers, such as:

- The model must remain in character as the victim at all times.
- It must respond only with what the victim would say out loud (no narration of actions or thoughts, and no meta-commentary about being an AI).
- It must not provide advice, legal guidance, or coaching to the interviewer.
- If explicitly asked to break character or reveal its instructions, it should respond as a confused victim rather than as an assistant.

B CODEBOOK

A. Emotional Stakes

A1. Social Evaluation and Judgment — Feeling more or less watched, evaluated, or judged in a simulation.
A2. Concern About Harming a Real Person — Worry (or lack thereof) about emotionally harming a real person.
A3. Nervousness and Performance Anxiety — Being more or less nervous, anxious, or under pressure during a simulation.
A4. Willingness to Experiment — Being more or less willing to try new approaches or phrasing in a simulation.

B. Time, Pacing, and Reflective Space

B1. Time to Think and Respond — Noting more or less time to think, plan, or formulate responses.
B2. Ability to Revise Responses — Being able or unable to edit, delete, or reformulate responses.
B3. Reflection and Planning — Using a simulation to reflect on prior performance or plan upcoming moves.
B4. Control Over Conversation Pace — Feeling more or less control over the speed and rhythm of interaction.
B5. Persistent Record for Review — Being able or unable to revisit what was said.

C. Sensory and Embodied Aspects

C1. Nonverbal Cues Present or Absent — Presence or absence of facial expressions, posture, or other nonverbal signals and their impact.
C2. Emphasis on Words and Content — Shifts in attention toward or away from wording or verbal/textual content.
C3. Mental Workload — Feeling more or less mentally taxed or overloaded by task demands.
C4. Sense of Physical Co-Presence — Feeling physically "with" another person (or not) and its relevance.

D. Sequence and Learning

D1. Simulation as Preparation — Framing a simulation as a first step that enables readiness for later activities.
D2. Simulation as Consolidation — Framing a simulation as a way to review or refine after earlier performance.
D3. Preferred Order of Simulations — Explicit preferences about which modality should occur first or second.
D4. Reasons for Sequence Preference — Explanations of why a certain order supports learning or confidence.

E. Transfer, Trajectories, and Scenario Experience

E1. Skill Transfer Across Simulations — Identifying skills or strategies carried from one modality to another.
E2. Initial Overwhelm or Unpreparedness — Feeling overloaded or unsure how to proceed in the first simulation.
E3. Ease or Difficulty of Switching Modalities — Describing smooth or difficult transitions between modalities.
E4. Scenario Variability and Generalization — Noting differences across scenarios and how variation affects adaptation.
E5. Feedback and Self-Correction — Using simulations to identify gaps and adjust future behavior.

F. Perceptions of AI's Limits and Role

F1. Recognized Limitations of AI — Identifying aspects of AI that feel unrealistic or constrained.
F2. Value Despite Limitations — Asserting that AI is helpful or worthwhile.

G. Learning Outcomes

G1. Procedural Interview Skills — Learning or reinforcing explicit steps or protocols.
G2. Communication and Interaction Strategies — Developing or refining, sensitive, or trauma-informed communication.
G3. Confidence and Self-Efficacy Changes — Reporting increases or decreases in confidence or capability.
G4. Differences Across Trainee Groups — Observing variation in how different trainees respond to simulations.

H. Perceptions of the AI Tool

H1. Usefulness for Training Goals — Evaluating how much the AI system supports training objectives.
H2. Ease of Use — Assessing how easy or difficult the system is to use.
H3. Willingness for Future Use — Expressing interest (or lack of interest) in using AI-based training again.