

LLMs Can Talk, but Can They Behave? Modeling Emotional Dynamics in AI-Driven Behavioral Training Simulations

Rachel Benton, Dr. Gwen Campbell, Dr. Jennifer Riley
Design Interactive, LLC

Within military organizations, personnel must often navigate emotionally difficult interpersonal interactions where communication and behavioral judgment directly influence organizational trust, leadership effectiveness, and mission readiness. Within Defense Human Resources Activity (DHRA) training contexts, these interactions include emotionally charged Equal Opportunity (EO) conversations surrounding sensitive workplace violations. Traditionally, these behavioral skills have been trained through live roleplay, a resource-intensive approach constrained by facilitator cost, scheduling and scaling challenges, inconsistent scenario delivery, and limited opportunities for individualized, data-driven feedback. Large language models (LLMs) offer a promising foundation for adaptive conversational roleplay due to their ability to generate dynamic, open-ended dialogue. While LLMs demonstrate strong conversational fluency, modeling realistic emotionally responsive human behavior within interpersonal training scenarios remains an exploratory challenge.

In this work, trainees interacted with AI-driven virtual personas during emotionally difficult workplace communication scenarios. While LLMs effectively generate natural language, they do not inherently model the emotional persistence, ambiguity, escalation patterns, and interpersonal tension required for realistic roleplay. Virtual personas frequently accepted apologies unrealistically quickly, defaulted toward cooperation or mediation, avoided sustained confrontation, or reverted to emotionally flattened language despite adversarial conversational contexts. Personas were often capable of expressing transient nervousness or uncertainty, but struggled to sustain anger, defensiveness, resistance, or domineering interpersonal behavior across extended interaction.

Rather than treating conversations as unconstrained chat interactions, the system employed a layered behavioral architecture surrounding the LLM. Human behavior was operationalized into observable trainee competencies and role-dependent behavioral expectations grounded in historical training materials and subject matter expert (SME) guidance. Trainee performance against these competencies dynamically influenced persona behavior throughout interaction, allowing emotional intensity, resistance, and escalation patterns to evolve proportionally to conversational performance. Prompt engineering, behavioral state management, and persona motivation logic were used to reinforce emotional persistence and discourage sycophantic conversational tendencies commonly observed in general-purpose LLM behavior.

The work additionally addressed how emotional states manifest behaviorally beyond text generation alone. Human facilitators naturally communicate discomfort, hesitation, defensiveness, frustration, and ambiguity through conversational pacing, tone, incomplete responses, and nonverbal behavior that force trainees to actively navigate interpersonal tension.

Historical roleplay sessions and SME-guided demonstrations were analyzed to identify observable indicators associated with different emotional states and escalation patterns. These indicators were subsequently linked to both dialogue generation and avatar-based verbal and nonverbal expression systems to support more behaviorally coherent interactions.

Together, these approaches support a broader framework for emotionally adaptive AI roleplay through competency-driven persona behavior, emotional persistence modeling, escalation management, and multimodal behavioral expression. Beyond EO training contexts, these approaches may support broader applications including leadership development, negotiation, counseling, customer conflict resolution, and other forms of emotionally complex behavioral training. This presentation contributes an implementation-grounded framework for structuring emotionally adaptive and behaviorally coherent AI roleplay within complex interpersonal training environments.