Proceedings of the Human Factors and Ergonomics Society Annual Meeting
1–5


DOI: 10.1177/10711813251367362
journals.sagepub.com/home/pro


Adaptive De-escalation Trainer: Piloting a RAG-Enhanced, Emotionally Modulated AI Simulator for Police Training

Eshwara Prasad Sridhar[1], Jose Lopez[1], Mohammad Islam[1], and Shuchisnigdha Deb[1]

Abstract

Effective police de-escalation training requires realistic practice but is limited by the scalability and consistency of traditional methods. We present the Adaptive De-escalation Trainer, which combines a large language model with retrieval-augmented generation for scenario grounding and text-to-speech that modulates prosody based on officer keywords. We hypothesized that the simulator would yield high conversational coherence and responsiveness (H1), emotional realism (H2), keyword-driven emotional shifts (H3), and viability (in terms of latency and workload) (H4). In a study with three officers, high average ratings for conversational flow ($M = 6.00$), validity of persona ($M = 5.67$), and logical responses ($M = 6.67$) were found, supporting H1; high scores on emotional realism ($M = 5.75$) supporting H2; positive AI reactions after "success" phrases offering support for H3; and an average turn latency of 4.3 s with moderate NASA-TLX scores, confirming H4. These findings demonstrate feasibility and inform future work on latency reduction, scenario expansion, and advancing AI-based training.

Keywords

police de-escalation, conversational AI, simulation training, human-computer interaction

Introduction

Effective de-escalation is essential in modern law enforcement, significantly affecting the safety of officers and subjects, community relations, and the credibility of policing. The ability to calmly address volatile situations is crucial; however, providing consistent and realistic training for officers remains challenging. Traditional training methods, such as live role-playing and fixed scripted scenarios, have shown progress in officers' efforts to de-escalate situations (Oliva et al., 2010). Nevertheless, issues with confirmation bias and reduced adaptability from inconsistent and predictable scenarios, as well as the limited availability of role players, hinder officers' ability to practice nuanced communication skills effectively. This leads to poor decision-making and an increased risk of errors in real-world crises.

Advances in artificial intelligence (AI), especially in conversational systems, can help overcome training limitations. Creating realistic AI simulators for complex interactions, such as police de-escalation, relies on progress across several technical areas. At the core are Spoken Dialogue Systems (SDS), which form the foundation for interactive conversational agents by integrating speech recognition, dialogue management, and speech synthesis, often using unified approaches (Basit & Shafique, 2024; Ji et al., 2024).

Large Language Models (LLMs) have improved SDS by generating more fluent and human-like dialogue (Lee et al., 2025; Song & Xiong, 2025). LLMs create Role-Playing Language Agents (RPLAs) that can simulate various human personas and behaviors (J. Chen et al., 2024; Shanahan et al., 2023). Research shows that RPLAs can mimic cognitive functions like persuasion, decision-making, and emotional reasoning (Carrasco-Farre, 2024; Y. Liu & Long, 2025; Shao et al., 2023). Retrieval-Augmented Generation (RAG) is an important technique that ensures RPLAs give informed and consistent responses (Lewis et al., 2021). RAG improves LLMs output by retrieving relevant information from external sources; for example, expert-designed de-escalation scenarios for common crises, such as mental health issues or suicidal thoughts. This helps improve factual consistency and reduce inappropriate responses, known as "character hallucination" (Huang et al., 2024; Shao et al., 2023). Emotional Text-to-Speech (TTS) systems improve simulations by

[1]The University of Texas at Arlington, USA

Corresponding Author:
Eshwara Prasad Sridhar, Department of Industrial, Manufacturing, and Systems Engineering, The University of Texas at Arlington, 500 West First Street, Arlington, TX 76019-9800, USA.
Email: exs3645@mavs.uta.edu

adjusting pitch and speed to reflect the subject's emotions. This emotional tone and intensity greatly influence user interactions and the effectiveness of training (Q. Chen et al., 2024). Recent research has linked LLM outputs with TTS emotion by utilizing style control and emotion-guided dialogue to enhance emotional expression (Lee et al., 2025; C. Liu et al., 2024; Sigurgeirsson & King, 2023; Song & Xiong, 2025). Combining advanced SDS and RPLAs with steady emotional control allows AI simulators to create dynamic emotional interactions (Borsos et al., 2023; Q. Chen et al., 2024; Huang et al., 2024; Lee et al., 2025; Shanahan et al., 2023). This paper presents the "Adaptive De-escalation Trainer," a prototype AI conversation simulator designed to help law enforcement officers practice de-escalation techniques. The scenarios and the system were developed based on the instructional resources and live conversational transcripts provided by the Fort Worth Police Department in Texas. It uses an LLM combined with RAG to provide contextual scenarios. It also features dynamic TTS that adjusts the AI's emotional tone based on the officer's speech analysis of success or failure phrases, enabling responsive interactions.

This work describes the development of a simulator and presents preliminary findings from a pilot study involving three active-duty police officers. The study evaluated the simulator's feasibility and initial effectiveness through participant surveys that measured interaction quality, emotional realism, workload (using the NASA-TLX), and system performance logs. We hypothesized that: (H1) the simulator would show high perceived conversational coherence and responsiveness; (H2) it would achieve high perceived emotional realism; (H3) the keyword-based feedback loop would impact the simulator's emotional state; (H4) the system would demonstrate technical viability based on performance metrics like latency. Key contributions of this work include the integration of AI technologies for police de-escalation training, insights from the pilot study, and the identification of the system's potential as a scalable, on-demand training resource, as well as current limitations.

Methodology

System Architecture

The Adaptive De-escalation Trainer is a modular, real-time simulation system designed to engage police officers in emotionally responsive conversations. It operates as a closed-loop system (Figure 2) that processes officer speech, generates grounded AI responses, and adjusts emotional tone based on interaction flow. The system's core components are as follows:

Speech-to-Text (STT) interface: The system begins by capturing the officer's speech in real time and transcribing it using Google's Cloud Speech-to-Text API (*Speech-to-Text AI: Speech recognition and transcription*, 2025). This transcription provides the textual input needed for the conversation engine to interpret intent and generate a response.

Dialogue engine with RAG: The core dialogue engine uses LLM engine "gemini-2.0-flash-lite" (Comanici et al., 2025). To ensure responses are contextually grounded and relevant, the system uses RAG. Officer input is embedded using SentenceTransformer and matched against a FAISS index of expert-authored scenario information/backgrounds. The most relevant content is retrieved and added to the LLM prompt, helping to maintain coherent, scenario-consistent conversations and reducing off-topic or unrealistic replies that are common in LLMs.

Emotional state module and prosody control: The simulator models the emotional state of the virtual subject, tracking dimensions like agitation, fear, or calmness. Using keyword detection, it analyzes officer speech for specific "success" or "failure" phrases. Positive phrases (e.g., empathy, validation) raise the emotional score positively, while negative words (e.g., confrontation, dismissal) do the opposite. The current emotional state modulates the AI's vocal output using a dynamic TTS system, altering pitch and speed using weights to match the simulated emotion, such as distress, anger, or relief.

Scenario design and integration: Subject matter experts designed scenarios to reflect real-world challenges, such as mental health crises or suicidal ideation. These scenarios provide narrative context, subject history, emotional cues, and success and failure phrases. The RAG system utilizes this content to inform AI responses, ensuring that conversations remain relevant to the needs of law enforcement training.

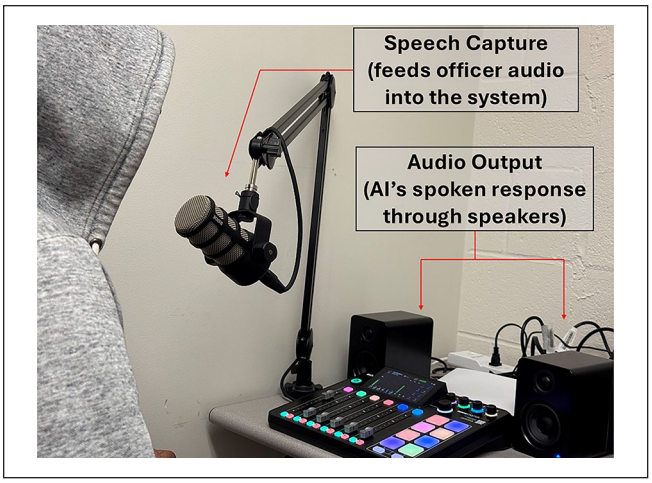

Figure 1. Pilot study—setup.

Pilot Study with Law Enforcement Officers

A pilot study was conducted with active-duty police officers ($N=3$) to evaluate the feasibility and perceived effectiveness of the simulator (Figure 1).

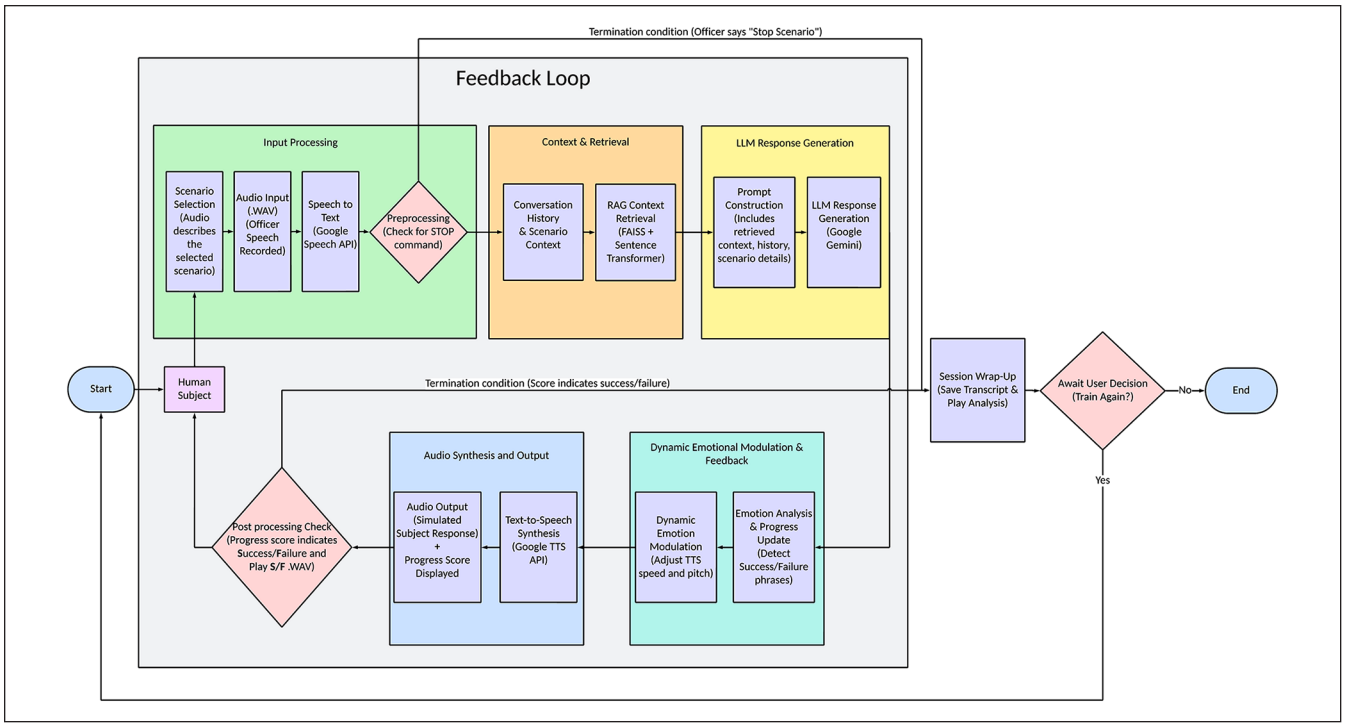

Figure 2. System architecture—AI simulator.

After obtaining informed consent, participants completed a brief demographic survey that captured their age, years of service, background in de-escalation training, and comfort level with technology. Each officer was given standardized instructions and a short briefing on the test scenario. They then engaged in one verbal interaction with the AI-simulated subject. The session was recorded, and officer speech, AI responses, and system performance data were logged.

Following the interaction, officers completed a custom questionnaire assessing the AI's conversational quality, including how natural, logical, and responsive it felt, as well as emotional realism, such as tone of voice, appropriateness of emotional reactions, and perceived individuality of the AI. They also completed the NASA-TLX, a six-dimensional rating scale (mental demand, physical demand, temporal demand, effort, performance, and frustration) scored 0 to 20 per subscale to yield an overall workload score, measuring perceived workload (Hart & Staveland, 1988).

Open-ended questions gathered feedback on misunderstandings or other issues during the conversation. Given the small sample size, results were analyzed descriptively using mean ratings, qualitative themes, and system logs (Figure 2).

Results

Participants

The three participating police officers were all male, with an average age of 46.3 years (range: 35–53) and an average of 19 years of service (range: 5–29). Two officers reported being "Very Confident" in generally de-escalating tense situations, while one reported being "Somewhat Confident." Comfort levels with learning and using new computer software or technology varied, with responses of "Neutral," "Slightly Uncomfortable," and "Comfortable."

Perceived Conversational Coherence, Responsiveness, and Emotional Realism

Officers rated the AI very positively on conversation quality and emotional realism (Figure 3). Key areas scored as follows (out of 7): conversational flow ($M = 6.00$), validity of persona ($M = 5.67$), and story consistency ($M = 6.33$). Logical responses ($M = 6.67$), understanding officer intent ($M = 5.67$), responsiveness to de-escalation tactics ($M = 5.33$), and minimal need to repeat statements ($M = 6.00$) all showed high means.

Notably, none of the officers experienced complete misunderstandings. For emotional realism, realistic expression ($M = 6.00$), appropriateness of reactions ($M = 5.67$), perceived individuality ($M = 5.67$), and clarity of emotional shifts ($M = 5.67$) all averaged above 5.5. This uniform set of high ratings shows the system produces smooth, believable dialogue and expressive emotions. These findings validate our approach of combining scenario grounding via RAG and dynamic, TTS-based emotion control to simulate realistic crisis dialogues. Overall, the results support the simulator's

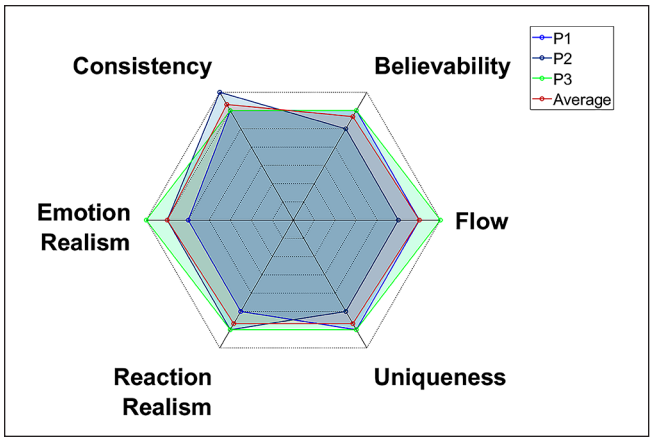

Figure 3. Average officer ratings of AI simulator interaction quality and emotional realism (7-point scale).

ability to provide reliable and realistic practice for de-escalation skills.

AI Response Sentiment Analysis

We looked at officer statements for "success" and "failure" keywords to see how they influence the AI's responses. Using the VADER score to gauge emotional responses (ranging from −1 for very negative to +1 for very positive), we found that in three pilot sessions, officers used "success phrases" before an AI response only twice, both of which resulted in positive AI reactions (average score of 0.56). There were no instances of "failure phrases" leading to a clear AI response. In contrast, out of 24 AI replies to comments without these keywords, 75% were positive (average score of 0.30) and 25% were negative.

System Performance and Workload

System performance logs (Figure 4) showed that the average wait time between conversation turns was 4.3 s. Figure 4 displays the latency distribution for key system components: STT, Emotion Analysis, RAG processing, LLM response generation, and TTS. The STT and TTS components had the longest median latencies and the most variability in all interactions.

Workload Assessment (NASA-TLX)

The perceived workload, measured using the NASA-TLX survey, indicated varied demand across dimensions (Figure 5). Average scores for the subscales were: Mental Demand (M=4.33, SD=4.93), Physical Demand (M=1.33, SD=1.53), Temporal Demand (M=5.33, SD=5.03), Performance (M=10.67, SD=9.29; lower scores indicate better participant-perceived performance), Effort (M=6.00, SD=3.61), and Frustration (M=2.33, SD=3.21).

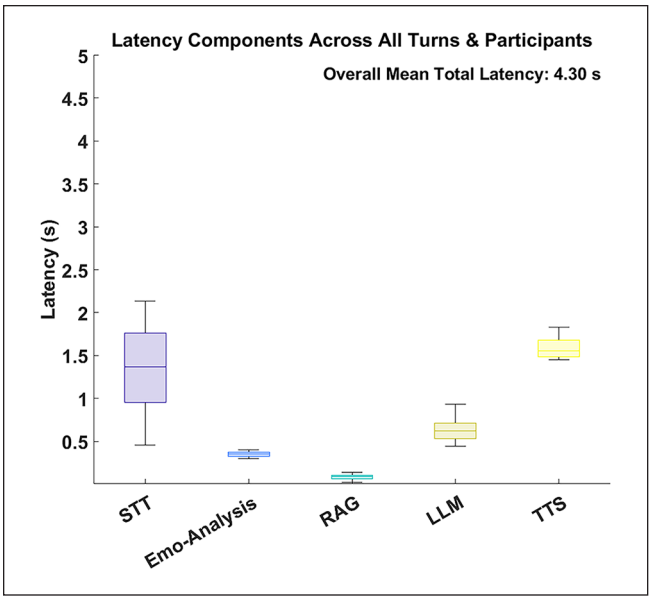

Figure 4. System response latencies (s) for core components.

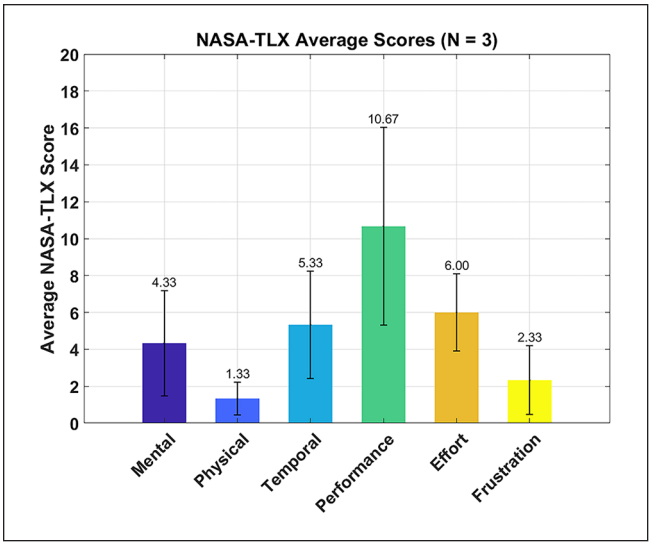

Figure 5. Average perceived workload (NASA-TLX scores) after AI simulator interaction.

Discussion and Conclusion

This pilot study indicates that the Adaptive De-escalation Trainer could effectively train police officers. The AI system performed reliably, with officers finding conversations logical and responsive. They felt the AI understood their intent during de-escalation efforts, with no major misunderstandings reported. Emotional realism was highly rated, with officers believing the AI's tone and reactions felt human-like. This supports the incorporation of emotional text-to-speech and an emotion model. The AI adjusts its emotional tone based on key phrases from the officer. While signs of effectiveness were

noted, more data are required for confirmation. A notable challenge was a 4.3-second average response delay due to cloud processing, which affected conversation flow. Local operation of these components could enhance responsiveness. Despite the small sample size (only three officers) and limited scenarios, the initial feedback highlights potential areas for improvement. Future work should expand scenario diversity, include a control group using standard role-play, and collect objective performance metrics (e.g., physiological stress, decision accuracy). Integrating virtual-reality avatars may further increase immersion.

In conclusion, the Adaptive De-escalation Trainer shows promise as an AI-based pedagogical model for scalable, on-demand de-escalation practice. Refining system responsiveness, broadening evaluation, and comparing against traditional methods will be key to its safe, effective deployment.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: We thank and acknowledge the Fort Worth Police Department and the Department of Justice for funding the study under grant 15PBJA-23-GG-06172-NTCP.

ORCID iD


Eshwara Prasad Sridhar https://orcid.org/0009-0004-4942-1815

References

Basit, A., & Shafique, M. (2024). *tinyDigiClones: A multi-modal LLM-based framework for edge-optimized personalized avatars* [Conference session]. 2024 International Joint Conference on Neural Networks (IJCNN), Yokohama, Japan (pp. 1–9). IEEE. https://doi.org/10.1109/IJCNN60899.2024.10649909

Borsos, Z., Sharifi, M., Vincent, D., Kharitonov, E., Zeghidour, N., & Tagliasacchi, M. (2023). *SoundStorm: Efficient parallel audio generation* (arXiv:2305.09636). arXiv. https://doi.org/10.48550/arXiv.2305.09636

Carrasco-Farre, C. (2024). *Large Language Models are as persuasive as humans, but how? About the cognitive effort and moral-emotional language of LLM arguments* (arXiv:2404.09329). arXiv. https://doi.org/10.48550/arXiv.2404.09329

Chen, J., Wang, X., Xu, R., Yuan, S., Zhang, Y., Shi, W., Xie, J., Li, S., Yang, R., Zhu, T., Chen, A., Li, N., Chen, L., Hu, C., Wu, S., Ren, S., Fu, Z., & Xiao, Y. (2024). *From persona to personalization: A survey on role-playing language agents* (arXiv:2404.18231). https://doi.org/10.48550/arXiv.2404.18231

Chen, Q., Gong, Y., Lu, Y., & Luo, X. (Robert). (2024). The golden zone of AI's emotional expression in frontline chatbot service failures. *Internet Research*. Advance online publication. https://doi.org/10.1108/INTR-07-2023-0551

Comanici, G., Bieber, E., Schaekermann, M., Pasupat, I., Sachdeva, N., Dhillon, I., Blistein, M., Ram, O., Zhang, D., Rosen, E., Marris, L., Petulla, S., Gaffney, C., Aharoni, A., Lintz, N., Pais, T. C., Jacobsson, H., Szpektor, I., Jiang, N.-J. . . . Ramabhadran, B. (2025). *Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities* (arXiv:2507.06261). arXiv. https://doi.org/10.48550/arXiv.2507.06261

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Advances in psychology* (Vol. 52, pp. 139–183). North-Holland. https://doi.org/10.1016/S0166-4115(08)62386-9

Huang, L., Lan, H., Sun, Z., Shi, C., & Bai, T. (2024). *Emotional RAG: Enhancing role-playing agents through emotional retrieval* (arXiv:2410.23041). arXiv. https://doi.org/10.48550/arXiv.2410.23041

Ji, S., Chen, Y., Fang, M., Zuo, J., Lu, J., Wang, H., Jiang, Z., Zhou, L., Liu, S., Cheng, X., Yang, X., Wang, Z., Yang, Q., Li, J., Jiang, Y., He, J., Chu, Y., Xu, J., & Zhao, Z. (2024). *WavChat: A survey of spoken dialogue models* (arXiv:2411.13577). arXiv. https://doi.org/10.48550/arXiv.2411.13577

Lee, J., Sim, Y., Kim, J., & Suh, Y.-J. (2025). EmoSDS: Unified emotionally adaptive spoken dialogue system using self-supervised speech representations. *Future Internet*, *17*(4), Article 4. https://doi.org/10.3390/fi17040143

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2021). *Retrieval-augmented generation for knowledge-intensive NLP tasks* (arXiv:2005.11401). arXiv. https://doi.org/10.48550/arXiv.2005.11401

Liu, C., Xie, Z., Zhao, S., Zhou, J., Xu, T., Li, M., & Chen, E. (2024). *Speak from heart: An emotion-guided LLM-based multimodal method for emotional dialogue generation* [Conference session]. Proceedings of the 2024 International Conference on Multimedia Retrieval, New York, NY, United States (pp. 533–542). ACM Digital Library. https://doi.org/10.1145/3652583.3658104

Liu, Y., & Long, Y. (2025). *EQ-negotiator: An emotion-reasoning LLM agent in credit dialogues* (arXiv:2503.21080). arXiv. https://doi.org/10.48550/arXiv.2503.21080

Oliva, J. R., Morgan, R., & Compton, M. T. (2010). A practical overview of de-escalation skills in law enforcement: Helping individuals in crisis while reducing police liability and injury. *Journal of Police Crisis Negotiations*, *10*(1–2), 15–29.

Shanahan, M., McDonell, K., & Reynolds, L. (2023). Role play with large language models. *Nature*, *623*(7987), 493–498. https://doi.org/10.1038/s41586-023-06647-8

Shao, Y., Li, L., Dai, J., & Qiu, X. (2023). *Character-LLM: A trainable agent for role-playing* (arXiv:2310.10158). arXiv. https://doi.org/10.48550/arXiv.2310.10158

Sigurgeirsson, A. T., & King, S. (2023). *Controllable speaking styles using a large language model* (arXiv:2305.10321). arXiv. https://doi.org/10.48550/arXiv.2305.10321

Song, Y., & Xiong, W. (2025). Large language model-driven 3D hyper-realistic interactive intelligent digital human system. *Sensors*, *25*(6), Article 6. https://doi.org/10.3390/s25061855

Speech-to-Text AI: Speech recognition and transcription. (2025). Google Cloud. Retrieved July 22, 2025, from https://cloud.google.com/speech-to-text