Advancing Dialogic Communication Skills
through LLM-Based Simulation
(Short paper)

Elad Yacobson
Technion – Israel Institute of Technology
eladyacobson@campus.technion.ac.il

Ofra Amir
Technion – Israel Institute of Technology
oamir@technion.ac.il

Ayelet Baram-Tsabari
Technion – Israel Institute of Technology
ayelet@technion.ac.il

קידום מיומנויות תקשורת דיאלוגית באמצעות סימולציה
מבוססת מודלי שפה גדולים (LLM)
(מאמר קצר)

אילת ברעם-צברי
הטכניון – מכון טכנולוגי לישראל
ayelet@technion.ac.il

עפרה עמיר
הטכניון – מכון טכנולוגי לישראל
oamir@technion.ac.il

אלעד יעקבסון
הטכניון – מכון טכנולוגי לישראל
eladyacobson@campus.technion.ac.il

Abstract

Communication skills are increasingly recognized as essential for science students. However, existing communication training programs face scalability challenges due to limited resources, time constraints, and the need for expert instructors. Consequently, these programs rarely provide scientists with sufficient opportunities to practice dialogic communication. This study explores the use of large language models (LLMs), specifically ChatGPT-4o, as a novel tool for scalable and personalized science communication training. We developed an LLM-based dialogue simulator grounded in *Prodigy*, a new theoretical framework that defines productive dialogue across four dimensions: Content, Interpersonal Rapport, Perspective Taking & Listening, and Integrity & Humility. In an exploratory study with 37 science students, participants conducted two voice conversations with ChatGPT 4o, receiving AI-generated feedback between sessions. Analysis of the conversations revealed significant improvements in dialogic performance across three of the four dimensions, along with high levels of user satisfaction. These results suggest that LLMs hold promise as effective tools for scalable communication training.

Keywords: large language models, science communication, soft skills training, computer-human interaction, AI in education.

Introduction

Communication skills are increasingly recognized as essential for science students (Akin et al., 2021). The benefits associated with improved communication skills span multiple levels. First, effective communication benefits society: Clear and engaging communication fosters public engagement with science, supports informed decision-making, and promotes a scientifically literate citizenry (Brownell et al., 2013). It also plays a key role in combating misinformation and building public trust, particularly in times of crisis, such as pandemics (Fischhoff, 2013). Second, stronger communication skills benefit scientists themselves, making researchers better positioned to advocate for science policy and funding (Baron, 2010), enhance the impact of their work, and advance their academic opportunities (Akin et al., 2021). Third, communication benefits science itself. It underpins interdisciplinary collaboration, which advances the collective growth of scientific knowledge. Moreover, engaging with the public and policymakers can help scientists align their research with societal needs, and learn from lay expertise, thereby enhancing the relevance and impact of their work (Besley et al., 2015).

Recent years brought an increase in science communication training programs for science professionals (Baram-Tsabari & Lewenstein, 2017; Besley et al., 2015; Coletti et al., 2023). However, while most existing training programs focus on one-way dissemination of knowledge (Reincke et al., 2020), research increasingly points to the value of practicing dialogue: engaging in two-way conversations that acknowledge the perspectives, concerns, and knowledge of diverse audiences (Lewenstein & Baram-Tsabari, 2022; Worthington et al., 2024). Yet meaningful practice in dialogic communication typically requires human actors and expert facilitators to provide feedback, resources that are costly, time-consuming, and often unavailable (Chu & Goodell, 2024; Kerr et al., 2020). Recent advances in generative AI, and especially the advent of large language models (LLMs), offer promising opportunities to overcome these barriers by providing accessible, adaptive, and scalable tools for practicing dialogic communication.

Research Goal and Questions

Our goal in this research was to explore whether an LLM-based simulator can offer a scalable, accessible alternative for dialogic science communication training. Specifically, our research questions were:

RQ1: Do students improve their dialogic communication performance following training with an LLM-based communication simulator?

RQ2: How do students experience the tool, and what benefits do they perceive in using it?

Theoretical Framework

For AI to be able to effectively support dialogic communication training, we need to first clearly define to the machine what productive dialogue entails, what its characteristics and features are. To this end, we developed *Prodigy* (Productive Dialogue for Generative AI), a theoretical framework that defines what constitutes productive dialogue by drawing on key ideas from science communication, academic productive dialogue, and nonviolent communication. Prodigy comprises 15 features organized into four dimensions: Content (e.g. clarity, credibility, reasoning), Interpersonal Rapport (e.g. empathy, respect, sharing personal details), Perspective Taking & Listening (e.g. inviting the partner to share their ideas, building on the partner ideas), and Integrity & Humility (e.g. expressing intellectual humility, transparency regarding one's own stance or agenda). It serves both as a guiding structure for LLM-generated feedback and as an analytic rubric for evaluating communicative performance.

Methodology

37 Israeli students from the Technion and the Weizmann Institute of Science participated in the study. Each student engaged in two voice conversations with an LLM (ChatGPT 4o). The scenario simulated a casual encounter with a layperson in a medical doctor's waiting room. The students were asked to explain their research to the LLM (See Figure 1 for an example excerpt of a conversation between a student and the LLM). Students first completed the initial conversation. They then received feedback from ChatGPT, and were asked to reflect on the feedback and consider how to improve. Immediately following this feedback, they participated in a second conversation based on the same scenario. Each conversation lasted between five to eight minutes and was transcribed by ChatGPT. After both conversations, students completed a reflective questionnaire about their experience with the training tool.

Two independent expert evaluators separately rated 36 transcripts (49% of the total number of transcripts, which was 74). Their scores reached 87% agreement (Cohen's k = 0.70). The remaining 38 transcripts were scored by one of the experts. To assess changes in dialogic performance, we conducted a Wilcoxon signed-rank test comparing students' scores across the four dimensions (Content, Interpersonal Rapport, Perspective Taking & Listening, Integrity & Humility) in their first and second conversations.

> **ChatGPT :**
> How are you doing? I'd love to hear about your research.
>
> **Student:**
> Thanks, that's great. I am researching NLP applications for science education assessment purposes.
>
> **ChatGPT:**
> Hmm, "NLP" sounds a bit technical – could you explain what that means in simple terms? And how does it connect to science education and assessment?
>
> **Student:**
> So, NLP stands for Natural Languages Processing. It's analysing free texts, and I am analysing students, high school students' free text responses to open-ended items about science education. The goal is to analyse their answers to open-ended questions and to be able to give them a formative feedback on their responses so they can learn from it.
>
> **ChatGPT:**
> Ah, got it—thanks! So you're using computer tools to read what students write in their own words and then help them improve by giving useful feedback. That actually sounds really helpful. But I wonder… if a computer is giving feedback, how do you make sure it really understands what the student meant? Isn't there a risk it might miss some important nuance or make a mistake?

Figure 1. Example excerpt from a student's dialogue with ChatGPT.

Results

Regarding RQ1, the results demonstrated a statistically significant improvement in students' overall dialogic performance from the first to the second conversation. Students improved across three of the four dimensions: Content (p = .015), Interpersonal Rapport (p < .001), and Perspective Taking & Listening (p = .007). The only dimension that did not show a statistically significant gain was Integrity & Humility (p = .38), which also received the lowest scores in both conversations relatively to the other dimensions. The overall average dialogic score increased from 0.20 to 0.30 (p < .001) on a (-1) to 1 scale, indicating a robust training effect. These results suggest that even a single feedback-guided interaction can lead to measurable improvements in key dialogic competencies.

Regarding RQ2, participants' responses to the post-task survey revealed high levels of satisfaction with the training experience. Most students (29 of 37) rated the simulation as helpful (ratings of 4 or 5 out of 5), and 75% reported that they felt they had improved their communication skills in the

second conversation. In response to open-ended questions about their learning experience, participants identified four key benefits of the LLM-based training: (1) Increased Confidence – Students reported that verbalizing their research aloud helped them clarify their message and gain confidence; (2) Safe, Non-Judgmental Environment – The AI partner provided a low-stakes setting for practice; (3) Actionable Feedback – Feedback was perceived as helpful and directly applicable to improvement; (4) Research Benefits – Some students reported that the dialogue even helped them think about new ideas and directions for their research.

This work provides an initial proof-of-concept that LLMs can effectively model dialogic interactions, support theoretically grounded feedback delivery, and facilitate communication skills improvement.

References

Akin, H., Rodgers, S., & Schultz, J. (2021). Science communication training as information seeking and processing: A theoretical approach to training early-career scientists. *Journal of Science Communication, 20*(5), A06. https://doi.org/10.22323/2.20050206

Baram-Tsabari, A., & Lewenstein, B. V. (2017). Science communication training: What are we trying to teach? *International Journal of Science Education, Part B, 7*(3), 285-300. https://doi.org/10.1080/21548455.2017.1303756

Baron, N. (2010). *Escape from the ivory tower: A guide to making your science matter*. Island Press.

Besley, J. C., Dudo, A., & Storksdieck, M. (2015). Scientists' views about communication training. *Journal of Research in Science Teaching*, *52*(2), 199-220. https://doi.org/10.1002/tea.21186

Brownell, S. E., Price, J. V., & Steinman, L. (2013). Science communication to the general public: Why we need to teach undergraduate and graduate students this skill as part of their formal scientific training. *Journal of Undergraduate Neuroscience Education, 12(*1), E6.

Chu, S. N., & Goodell, A. J. (2024). *Synthetic Patients: Simulating Difficult Conversations with Multimodal Generative AI for Medical Education* (No. arXiv:2405.19941). arXiv. https://doi.org/10.48550/arXiv.2405.19941

Coletti, A., McGloin, R., Embacher-Martin, K., Hamlin, E., & Turner, M. (2023). Examining Oral Communication Skills in Communication Training Programs for STEM Professionals: A Systematic Quantitative Literature Review. *Science Communication, 45*(6), 693-723. https://doi.org/10.1177/10755470231202773

Fischhoff, B. (2013). The sciences of science communication. *Proceedings of the National Academy of Sciences, 110*(supplement_3), 14033–14039. https://doi.org/10.1073/pnas.1213273110

Kerr, D., Ostaszkiewicz, J., Dunning, T., & Martin, P. (2020). The effectiveness of training interventions on nurses' communication skills: A systematic review. *Nurse Education Today, 89*, 104405.

Lewenstein, B. V., & Baram-Tsabari, A. (2022). How should we organize science communication trainings to achieve competencies? *International Journal of Science Education, Part B*, *12*(4), 289–308. https://doi.org/10.1080/21548455.2022.2136985

Reincke, C. M., Bredenoord, A. L., & Van Mil, M. H. (2020). From deficit to dialogue in science communication: The dialogue communication model requires additional roles from scientists. *EMBO Reports*, *21*(9), e51278. https://doi.org/10.15252/embr.202051278

Worthington, D., Graham, B., Gilliard, H., & Kelp, N. C. (2024). Development of a Scale to Measure Planned Behavior in Inclusive Science Communication: Validity Evidence in Undergraduate STEM Students. *CBE—Life Sciences Education*, *23*(4), ar48. https://doi.org/10.1187/cbe.24-01-0012