

Integrating AI into the Cadet Experience

Adam Fatebene, Andrew Howard, Colin Donelson, Xuan-Chen Du, and Jonathan Mellon

Department of Systems Engineering, West Point, New York, 10996

Corresponding author's Email: adam.j.fatebene.mil@army.mil

Author Note: The authors of this paper conducted a year-long research project through the Department of Systems Engineering at the United States Military Academy under the advisement of Dr. Jonathan Mellon (Associate Professor). The views expressed in this paper are those of the authors and do not reflect those of the United States Army, the United States Military Academy, or the Department of Systems Engineering.

Abstract: This paper aims to find effective uses of Artificial Intelligence (AI) in education and training at West Point. We developed AI bots for three use cases: language learning, stakeholder engagement, and practice with difficult military conversations such as counseling. We deployed them to five courses: Persian, Russian, Systems Design, Environmental Engineering, and a military dilemma's class (Leader's Challenge). After evaluating user surveys, we found that cadets perceived the bots to be highly usable and delivered superior educational experience to traditional methods. Our most promising product is the voice-mode AI bot for challenging conversations. Stakeholders and users perceived this use case as offering a unique training experience which is not currently possible within West Point or even the Army as a whole.

Keywords: Artificial Intelligence, Education, Language Learning Large Language Model, e-Learning

1. Introduction

Finding the most effective methods to instruct students has always been a core task for educators. The United States Military Academy must teach not only academic subjects, but also military and physical instruction, making efficient use of class time even more crucial. With the rise of AI in education, there have been questions as to whether these new technologies could better meet students' varying learning needs (Ayeni, 2024). Our research aims to find use cases for Artificial Intelligence (AI) at West Point that can deliver a better learning experience than traditional techniques within the academic, moral, military, and leadership domains. Our approach to addressing educational needs at West Point is to create AI products that support learning for various courses. We developed and tested three use cases: 1) text chatbots that act as language learning assistants, 2) text chatbots that roleplay as stakeholders to teach the skill of stakeholder analysis for engineering classes, and 3) audio bots that act as soldiers for cadets to practice in-person conversations, training, or counseling.

We implemented our designs across several courses: Russian II (LR204), Persian II (LZ204), Foundations of Engineering Design and Systems Management (SE301), Leader's Challenge, and Environmental Engineering for Community Development (EV450). We assessed the outcomes using surveys with cadets afterwards. As a part of the survey, we implemented the System Usability Score (SUS) standard to evaluate how usable our interface was for users (Lewis, 2017). The SUS is measured on scale of 0-100 where 68 is the benchmark for a system to be considered "usable". Cadets rated our use cases as usable with an overall average standardized System Usability Score (SUS) score of 73.5.

We also assessed cadets' perceptions on the extent to which the AI products added value when compared to traditional methods. Overall, we found that cadets preferred the introduction of AI education tools into their traditional curriculum. Those given the text-based ThayerAI platform (whether language or stakeholder analysis in this case) found it to be a positive addition to their learning process. Those given the voice-based AI bots for leader's challenge saw potential training values in their cadet career. Cadets largely found the systems for the AI products usable, and we tracked a positive trend of user acceptance across all use-cases.

2. Background

Artificial intelligence has been researched as a tutoring option for students as far back as the 1980s but faced obstacles due to the limitations of available data (Woolf, 1991). However, AI is now able to use a far larger set of information due to increased computational capacity and the availability of large quantities of data on the Internet (Gill, 2024). Trials of AI use in education can be seen in 2019 in China with the implementation of an AI-tailored studying platform known as Squirrel AI (Hao, 2019). Another study conducted in 2024 pre-assessed 300 students and 50 educators across various subjects and demographics (Sari, 2024). Subjects were given a pre-assessment test and given AI-powered tools such as "Smart Sparrow Platform" and real-time feedback tools such as "Open Essayist". The result was a 14-point

improvement in their post-assessments validating its effectiveness (Sari, 2024). However, there are concerns that the education sector could develop a reliance on these technologies and move away from believing in the necessity of human tutors (Holmes, 2022).

The previous capstone team built an AI platform (ThayerAI) specialized for building and hosting educational bots at West Point (Leggett et al., 2025). They tested a language tutoring AI use case with a West Point's Portuguese course and found a positive initial reception to the bots they built. While the previous team found strong initial results, they were not able to deploy the chatbots at scale, across multiple departments, or using voice models. We aim to fill these gaps in our project.

3. Stakeholder Analysis & Product Development

Dr. Richardson, Vice Dean of Academics at the United States Military Academy tasked our team with finding effective uses of AI across West Point. We were able to gauge interest in departments seeking to enhance their curriculum with AI. We initially started with five prototype use cases and gauged technical feasibility and stakeholder interest through initial interviews. Our stakeholder analysis and product development followed a cyclical process of prototyping, testing, and collaborating with stakeholders. Through this process we narrowed our final products down to three use cases: 1) bots to create daily assignments tailored to lesson objectives in foreign languages, 2) bots that act as stakeholders to engineering projects, and 3) programming ChatGPT's voice mode to simulate counseling conversations and provide practice in Army counseling.

Our first attempt at integrating AI into the cadet curriculum was through foreign languages, building on the previous team's success with Portuguese. We introduced our ideas to all foreign language courses, sparking interest from the Russian and Persian course directors. Our first prototype consisted of a text-based AI tool that would create conversation in the desired language and provide feedback to the user about their grammar and syntax. We had separate system prompts for the main Chatbot conversation and a "feedback" panel that critiqued the user's conversation. Following this prototype, both departments advocated for a tool to create daily assignments, conversations, and engaging questions with exercises that directly stem from lesson objectives and the syllabus. They wanted AI to pull questions that are tailored to the students' weakest performing topics. In developing these lesson tools, we recruited students from the Advanced Russian and Persian courses to more efficiently develop these bots and verify their proficiency in their respective language. Through continued stakeholder meetings, and empowering cadet language students to help develop new designs, our World Language use case was made public and used as a learning tool for each lesson of the intermediate level Russian course (LR204) and sees continued development in intermediate level Persian Course (LZ204).

Seeing the AI tools' capabilities for conversation, we introduced a stakeholder analysis bot to simulate interview and data collection of stakeholders in business, and environmental engineering projects. Prior to our prototype, West Point's Systems Engineering Department had introduced this idea to their curriculum in SE301 (which we also evaluate). We reached out to gauge interest in developing this AI tool and find where it could be beneficial in other parts of their department and across West Point. Our discussions with the Systems Engineering Department led us to seek out the Environmental Engineering Department and their curriculum tailored to stakeholder analysis. After providing them with an initial prototype, department members recognized its potential usability in the Mid-Course project of the Environmental Engineering for Community Development course (EV450). Environmental Engineering instructors saw potential in developing a more interactive approach to practicing stakeholder analysis. Following this discourse, we collaborated with the course instructors of EV450 and immediately began development of stakeholder bots.

Using the same interface as our language models, we developed system prompts to create characters, their backgrounds, and personalities for users to interview as a part of the Mid-Course Project. In our product development, we came across several challenges in the AI's ability to act personally and be congruent with general information across several bot characters. We recognized through our testing that AI, specifically ChatGPT, is programmed to act as an information provider and regurgitate information in a lengthy paragraph structure. We saw this as a potential limitation but through iterative changes to system prompting, we were able to eliminate this behavior and ensure the conversation between user and stakeholder was free flowing and showcased the personality and background that we programmed for each character. Continued testing and collaboration with instructors led us to create nine total stakeholder bots that have been implemented as a part of the stakeholder analysis section of the EV450 Mid-Course Project.

While most of our use cases revolved around academia, we also introduced AI platforms to enhance leadership development programs and training at West Point. From previous success with conversational AI tools in language learning, we introduced a text conversation bot to simulate an Army counseling scenario. Like past prompt engineering, we designed the bot to have specific backgrounds, personalities, preferences, and issues that the user would have to help handle in this counseling session. While the bot was able to hold conversations, the course director for the counselling class gave us feedback that the text mode detracted from the importance of face-to-face interaction and how tone of voice and the flow of conversation could play a role in the success of the counseling session. Fortunately, we were able to address these concerns with the introduction of ChatGPT's voice mode. With this new audio mode, we discovered we could develop system prompts for the voice mode the same way we developed them in all previous use cases.

Through product development, we created several counseling simulations spanning scenarios such as enlisted soldiers having financial issues, family and work-life balance concerns, and inter-unit tensions between soldiers. Once we developed three prototypes, we once again reached out to find stakeholders willing to give their input. We introduced these counseling simulations to members of the Simon Center for the Professional Military Ethic (SCPME) at West Point. Through stakeholder analysis we developed additional counseling simulations tailored to an Academy wide training known as Leader's Challenge. Leader's Challenge is a program conducted semi-annually that provides cadets with real world scenarios that current Army Officers have dealt with. These trainings are predominantly discussions about making difficult decisions as officers. With the help of our stakeholders at the Simon Center, we identified specific scenarios which we can simulate via ChatGPT's voice mode and immerse the cadets into the situation to provide more engaging training. An example recording is linked [here](#). Through meetings with stakeholders of the Simon Center, our efforts caught the attention of the Commandant of the USCC, BG Garcia, Command Sergeant Major of USCC, CSM Powers, and Command Sergeant Major of the United States Military Academy, CSM Barretto. Our discussions with these high-level leaders at West Point gave us deeper insights into their leadership experiences, which we can implement into our simulations. This helped us establish more scenarios applicable to training and facilitate better counseling practice to real life situations cadets may face as future officers.

While we found success in many of our use cases, our rejected use cases helped us recognize vulnerabilities in current AI technology and gave us a better perspective as to what is usable in the classroom. We rejected essay feedback/grading use case because stakeholders feared that the feedback component would be used as a crutch by cadets to take unwanted shortcuts to improve their essays. In terms of the grading component, our bot would often give different grades for identical essays. We found the failure rate in grading inconsistencies to be significant in comparison to grading by human instructors. In some cases, the same essay would score as low as a "C" and as high as a "B+". We rejected an International Relations (IR) simulation because stakeholders did not see it as adding value compared to traditional class discussions. In a class with a focus on how human behavior influences politics, stakeholders were not persuaded that students would gain greater insight into these nuances from an IR simulation.

4. User Acceptance Testing

User acceptance testing (UAT) is used to determine whether software meets user acceptance criteria (Santos et al., 2018). We designed and fielded a survey to 406 cadets across all the courses we involved in our research with the goal of capturing the usability and viability of these AI use cases. Those surveyed included cadets in LR204 (Russian II), LZ204 (Persian II), SE301 (Fundamentals of Engineering Design and Systems Management), EV450 (Environmental Engineering for Community Development), and a group of 16 junior (Cows) and senior (Firsties) cadets who tested our voice-based counseling bot for Leader's Challenge. We received 135 survey responses.

The survey included the system usability scale and custom indicators created by our group to capture the user's thoughts on usability and viability. Our survey consists of 23 questions total. The system usability portion of the survey encompasses the first ten questions, followed by ten custom-closed questions, and three optional open-ended questions. All closed-ended questions are on a 5-point scale from strongly disagree (1) to strongly agree (5). Scores above three therefore indicate agreement with a statement on average whereas scores below three indicate disagreement.

The custom indicator questions asked if AI bots 1) made the learning process engaging, 2) helped them feel more connected to the subject, 3) helped to understand relevant concepts, 4) identified areas for improvement, 5) supported their learning progress, 6) prepared them to handle real-world scenarios, 7) improved their overall performance, and 8) added significant value. Finally, we asked whether cadets would 9) like to use similar AI bots for similar tasks and whether they 10) would have preferred traditional methods.

5. Results

Figure 1 shows the system usability scores for our AI tools. Pooling all responses, system usability for our AI tools (73.5) is significantly higher than the standard benchmark score of 68 (Lewis, 2017). This is also the case for Leader's Challenge (77), Systems Engineering (76), and Russian (73). Persian (68) and Environmental Engineering (73) were not significantly better than the benchmark, mainly due to their smaller sample sizes of 6 and 17 survey responses, respectively. This also means we cannot reject the hypothesis that they were *more* usable than the rest of our AI tools. Overall, our results indicate that our tools are more usable than a typical system.

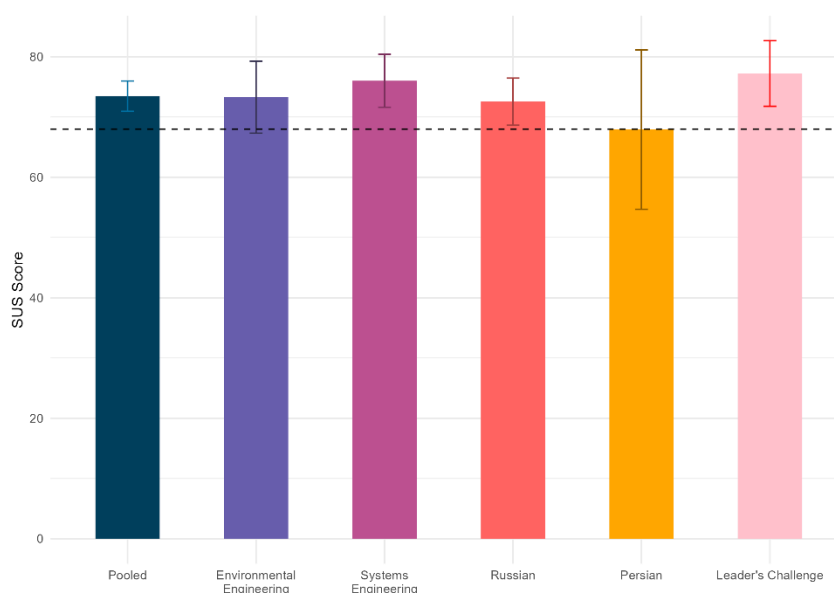


Figure 1: System Usability Comparison. Dashed line illustrates standard benchmark score of 68 against each class.

Figures 2 and 3 show average user acceptance scores for each class on our 10 custom indicators. In general, the scores for Persian are not significant (either in terms of being better than the neutral baseline or worse than the pooled score) because of the small sample size (6 cadets). We therefore omit further discussion of Persian's results in this section. For the learning engagement and connectivity to the subject, all classes scored above the neutral benchmark value of 3. Leader's Challenge scored significantly higher than the other use cases on engagement (4.5) and connection to the subject (4.2). Having been the only AI tool that was voice enabled; these results indicate that voice-based tools provide a more engaging experience that cadets feel more connected with than text-based models.

Combining all responses together cadets assess our AI tools as better than traditional methods (3.5), with Systems Engineering (3.8), Russian (3.5), and Leader's Challenge (3.7) separately achieving significance compared to the neutral value as well. Environmental Engineering scored above the neutral benchmark but was not significant. From this data, we can conclude that our AI tools were perceived as better learning tools than the traditional classroom methods. Cadets *disagreed* that they preferred traditional learning techniques. More specifically, Leader's Challenge scores were identified as significantly different (2.33) than the neutral benchmark, suggesting that current methods of practice are seen as less valuable than the AI-assisted approach.

Next, we address whether the bots helped students to recognize where they need to improve. Through our pooled average of all classes, we saw overall positive scores in this area (3.5), however this is predominantly driven by Russian which scored significantly higher than the benchmark. AI tools tailored to language learning appear to have helped language learners in Russian classes identify their struggles more so than other classrooms.

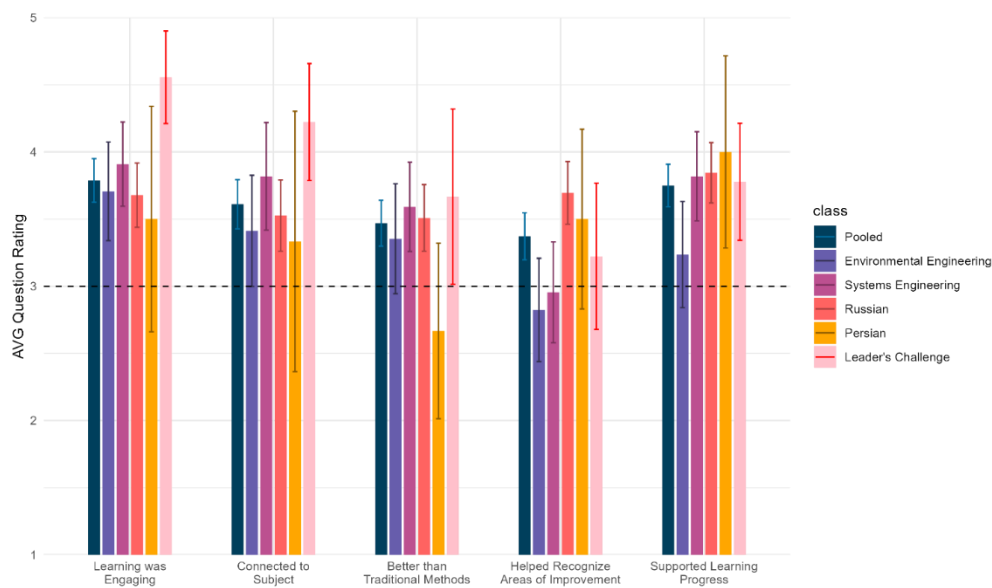


Figure 2: User Acceptance Rating (first set of questions). Dashed line represents neutral score for each question.

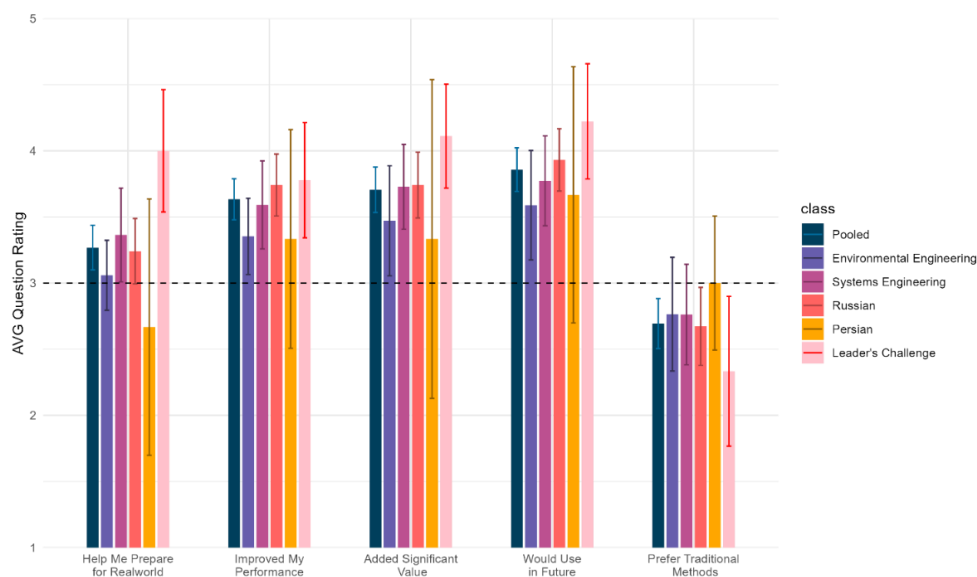


Figure 3: User Acceptance` Rating (second set of questions). Dashed line represents neutral score for each question.

All classes significantly exceed the neutral benchmark on adding significant value to the classroom. Likewise, all classes surveyed scored positively in improving their performance and supporting learning progress. This data once again shows that AI tools can enhance the learning environment.

The Leader's Challenge AI tool was the only case where AI was seen as being significantly above neutral in preparing students for real world scenarios. This may reflect the real-world nature of the scenarios cadets encounter during leader's challenge and may influence their scores.

Throughout our data, cadets have demonstrated a positive overall response to these AI tools. All classes surveyed scored in agreement to use these tools in the future and we hope continued development further benefits curriculum and the learning progress of cadets across West Point.

6. Conclusion

Our research on integrating AI into the cadet experience at USMA has had positive impacts. Our team built three types of AI bots: language learning bots, stakeholder engagement bots, and voice-mode AI chatbots to simulate difficult military conversations. We then deployed them to Persian, Russian, Environmental Engineering, Systems Engineering, and Leaders' Challenge. We conducted user acceptance testing through a survey of 135 cadets who used our AI tools. Cadets rated the AI tools as highly usable, rated the bots positively on effectiveness, and expressed an interest in using similar AI tools in future courses. These findings show that our products can efficiently support cadets' learning experiences. Our results were particularly strong for our voice bots which simulate realistic conversations covering different aspects of military life and provide useful training experiences.

Future teams should continue to identify viable use cases within West Point and other colleges that have high impact and user acceptance. We believe that the Russian use case presents a good model for integrating bots into language learning. The bots were created by a team of Russian majors and faculty who learned to craft the tools necessary to make quality bots. This collaboration should be scaled for more language classes at West Point in future. Additionally, text-mode stakeholder engagement bots could be rolled out to other engineering courses. Moreover, the positive feedback of our current voice-mode suggests there is potential for voice-mode bot for stakeholder analysis. This could entail integrating both text bots and voice mode bots to construct a more interactive user experience.

Currently, West Point already has programs such as Team Leader Academy and Leadership Training Program (LTP) where cadets are taught techniques to counsel subordinates but do get few opportunities to practice these skills. AI would allow cadets to practice these skills repeatedly during such programs. West Point should also consider voice mode applications where a conversation between cadet upperclassman and their subordinates are simulated to develop cadet leadership at West Point. Lastly, there is potential for extending AI use to broader topics such as negotiation, debate, and bidding processes, allowing cadets to gain a deeper understanding of these subjects through AI interactions. We also believe that our approach has applicability in education beyond West Point.

We believe that voice mode counseling could fill a key gap in Army training beyond West Point: difficult conversations is a core part of leadership throughout the Army. Practicing these skills would therefore enhance readiness and leadership. While the voice mode is currently expensive to deploy at scale, the potential gains in readiness may make this a price worth paying. Our hope and expectation are that proper implementation of this technology will fill in a key gap in current Army training.

References

- Ayeni, O. O., Al Hamad, N. M., Chisom, O. N., Osawaru, B., & Adewusi, O. E. (2024). AI in education: A review of personalized learning and educational technology. *GSC Advanced Research and Reviews*, 18(2), 261–271.
- Chiu, T. K., Meng, H., Chai, C. S., King, I., Wong, S., & Yam, Y. (2021). Creation and evaluation of a pretertiary artificial intelligence (AI) curriculum. *IEEE Transactions on Education*, 65(1), 30–39.
- Gill, S. S., Xu, M., Patros, P., Wu, H., Kaur, R., Kaur, K., ... & Buyya, R. (2024). Transformative effects of ChatGPT on modern education: Emerging era of AI chatbots. *Internet of Things and Cyber-Physical Systems*, 4, 19–23.
- Hao, K. (2019). China has started a grand experiment in AI education. It could reshape how the world learns. *MIT Technology Review*, 123(1), 1–9.
- Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education*, 57, 542–570.
- Sari, H. E., Tumanggor, B., & Efron, D. (2024). Improving educational outcomes through adaptive learning systems using AI. *International Transactions on Artificial Intelligence*, 3(1), 21–31.
- Leggett, C., Marchi, M., Muse, M., Wesley, S., & Mellon, J. (2024). Leveraging artificial intelligence for cadet education. Proceedings of the Annual General Donald R. Keith Memorial Conference, West Point, NY, USA. Society for Industrial and Systems Engineering.
- Lewis, R. J. (2017). Item benchmarks for the system usability scale. *Journal of User Experience*. Retrieved March 14, 2025, from <https://uxpajournal.org/item-benchmarks-for-the-system-usability-scale>
- Woolf, B. (1991). AI in education. University of Massachusetts at Amherst, Department of Computer and Information Science.