

Leveraging Artificial Intelligence for Cadet Education

Conner Leggett, Maximus Marchi, McKenzie Muse, Samuel Wesley, and Jonathan Mellon

Department of Systems of Engineering,
United States Military Academy,
West Point, NY 10996

Corresponding author's Email: conner.r.leggett@gmail.com

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 contributes to the ongoing discourse on integrating artificial intelligence (AI) technologies into educational settings while addressing the challenges within military institutions. Specifically, we investigate how the United States Military Academy (USMA) can integrate emerging technology into classrooms while upholding core military values. We conducted a comprehensive assessment of potential AI applications at USMA, culminating in the development of a use-case feasibility index for educational purposes. We developed an AI-powered platform that would enable cadets and faculty to create customizable chatbots aimed at enhancing learning experiences. We devised Portuguese AI chat and feedback bots aligned with Standard Portuguese (LP204) curriculum objectives, facilitating natural conversations, and delivering personalized feedback to users. Through a systematic test conducted across all LP204 sections, followed by a post-application survey, we examined the efficacy that AI can have on language learning programs. Findings reveal positive feedback, suggesting the potential utility of AI-driven educational tools.

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

1. Introduction

AI in education has begun to transform student-teacher interactions since the creation of Large Language Models (LLMs). ChatGPT, in its first months, rapidly gained popularity, reaching 100 million users globally and 30% of U.S. college students (Chen, 2023). The mission of USMA is “to build, educate, train, and inspire the Corps of Cadets to be commissioned leaders of character” (Gilland, 2024). AI might endanger the development of these core values in future leaders due to its potential for misuse and overreliance. Our team, however, explores the beneficial integration of AI technology into USMA education, focusing on its potential to be an educational tool rather than a crutch.

One way to align AI usage with institutional goals is to create closed-domain bots that facilitate critical thinking without eroding academic integrity. We built a platform (Thayer AI) that enables cadets and faculty with no technical skills to construct AI chatbots for specific subjects that give personalized feedback to students. To ensure these chatbots behave appropriately, users should iteratively employ effective prompt engineering techniques, such as specificity and clearly defined limits.

Assessing the utility of bots built on Thayer AI, we built Portuguese AI chat and feedback bots (this collective body of chat and feedback bots will be referred to as the “Bot”) to support LP204 curriculum objectives. We tested it with 134 cadets in LP204 and found that the performance of the Bot could be substantially improved by prompt engineering: iterating the prompt to eliminate unwanted behavior (identified by the team, LP204 cadets, and instructors) and keeping desirable responses. As students interact with the chatbot, a feedback bot offers critiques and insights throughout the conversation, allowing students to see areas to improve in real-time.

We evaluated the accuracy of the Portuguese chat and feedback bots’ outputs (certified by Portuguese professors) and conducted a user survey. From our final testing data, the accuracy checks revealed that 92% of the feedback responses were perfect and only 3% were clear errors. The survey results showed a positive reception of the Bot: 95% of cadets wanted it as a study tool and 95% felt it effectively corrected their mistakes. However, 9% of cadets reported it sometimes provided inaccurate information. Overall, our results demonstrate AI’s potential as a valuable educational tool.

2. Background

LLMs are deep-learning models that scan past text and predict what text is most likely to come next, based on processing large quantities of text data (Google, 2023). Text completion turns out to be widely useful including holding conversations and performing arbitrary tasks. AI offers much of this potential to transform the educational world with what is currently being used and future use cases such as personalized learning and customized feedback (Zhang, 2021). In a broader scope, there are parallels across education that point to general use cases where AI is currently being used from utilizing expert systems for pedagogical planning to creating systems for customized feedback (Zhang, 2021). Scholars have argued that AI can stimulate learning within the classroom (Zhang, 2021). The idea of students engaging with a new system that offers personalized learning could motivate them to be their own teacher (Mollick, 2023). Additionally, since AI can be prompted in a way that creates a unique learning experience, it makes the interaction stimulating for all student users (Mollick, 2023). By requiring students to articulate their ideas, AI can make them put cognitive effort into learning (Mollick, 2023).

The major challenge of AI in education lies in its potential misuse, often as a means to bypass human intuition rather than enhancing it (Elmessiry, 2023). An example of such misuse is AI's potential to limit in-person human interaction. Scholars have found that the use of AI inside of school hours has led to the fatigue of real-human interaction outside of school hours (Monash, 2019). Creating an isolated environment where students lean on AI too much can dissolve their ceiling for social skills that can be developed in a normal classroom environment through the guidance of instructors (Bryant, 2020). This lack of human interaction may disrupt traditional educational norms, affecting future generations negatively. Our goal is to develop AI use cases for USMA education that harness its potential benefits while mitigating associated risks.

3. Stakeholder Analysis

Our client, Dr. Susan Richardson, Vice Dean of Academic Affairs at USMA, expressed concerns about AI integration and its impact on critical thinking. Our consultations focused on identifying AI applications that enhance rather than replace critical thinking skills while meeting the diverse needs of USMA stakeholders. We conducted interviews with 14 stakeholders, including representatives from departments such as the Department of Foreign Languages (DFL), the Department of Electrical Engineering and Computer Science, and the Simon Center for Professional and Military Ethics. Through collaboration, we categorized potential use cases based on two main variables: Risk and Feasibility. Risk refers to what you should and should not develop (crutch vs. resource), while feasibility refers to how challenging a use case is for AI to execute effectively.

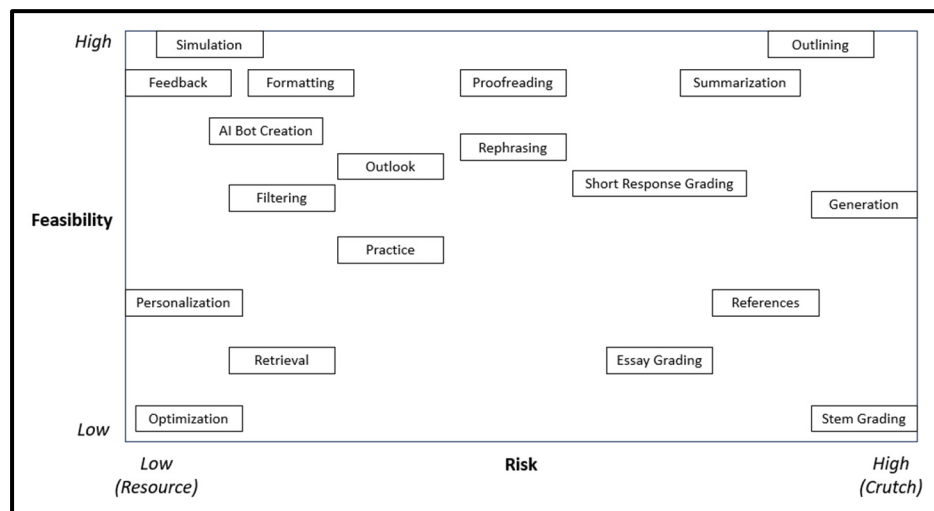


Figure 1. Use-Case Feasibility Index

The Use-Case Index (Figure 1) outlines our assessment of AI use cases based on risk and feasibility, informed by research, iterative prototyping, and stakeholder analysis. We identified 19 potential use cases and created beta versions to evaluate risk and feasibility. For example, we developed a STEM grading bot (High Risk & Low Feasibility) and a retrieval bot (Low Risk & Low Feasibility). The STEM Grading bot provided user input with a grade based on the prompted grading criteria. This use case was infeasible because AI was inconsistent and not advanced enough to give accurate grades or partial credit on complicated STEM assignments. We determined it was high risk because instructors would develop an overreliance on the tool and eliminate their ability to understand the shortcomings/needs of their students. The retrieval bot extracts information from

documents when prompted by a user. We found that this bot was not feasible because AI struggled to consistently retrieve exact information or direct quotes from documents. It would hallucinate and provide inaccurate data. However, we found it to be low risk because if it worked it would help users find information from sources more efficiently.

We focused on highly feasible and low-risk use cases: Simulation and Feedback. Simulation guides through tailored scenarios, fostering critical thinking, while Feedback offers constructive input based on instructor-provided criteria and user inputs. To assess their value, we developed a platform for instructors to quickly deploy these bots. Initially considering both language learning and USMA's Leader's Challenge (ethical dilemma discussions), we narrowed our focus due to project constraints, prioritizing the language learning use case for testing with cadets.

4. Product Development

4.1 Platform

The platform empowers faculty members to craft personalized chatbots (simulation) and feedback bots (feedback) for student engagement. Students access a faculty-provided URL to interact with the chatbot, sending messages and receiving responses in the conversation panel based on custom prompts. Additionally, they receive feedback on their interactions in a separate panel, also guided by faculty prompts.

Faculty can create bots through a create page (Figure 2) which is designed to have a low technical barrier to allow the creation of custom bots for specific use cases. Faculty will input the bot's name, provide a brief description for cadets, and input chatbot and feedback prompts that define its behavior. By default, the application only provides context for each individual response from the chatbot. However, selecting the 'use full conversation for feedback' option enables feedback to consider the entire conversation history. Faculty can choose between LLM models (GPT-3.5 Turbo or GPT-4) for each panel. Upon creation, the use case is assigned a unique 'botid' stored in our capstone's Microsoft Azure Cloud database for prompts.

Figure 3 demonstrates how our Unified Modeling Language (UML) Use Case Diagram chatbot application operates. When a chatbot is created it is assigned a specific value known as the "botid". To access a particular application this botid needs to be used within our user interface (Figure 4) to execute the faculty-designed prompts. The user will input a message that is assigned into a field known as the "user message". Upon the send request from the user, both the user message and botid will be sent to the server. The server will then locate the botid from our database and extract the prompts associated with this value. Next, these fields along with the conversation history will be submitted to the LLM API (OpenAI) and its AI responses will be delivered back to the user interface. The user interface will display the new user input and conversation output on the conversation panel while the new feedback on the conversation will be shown on the feedback panel.

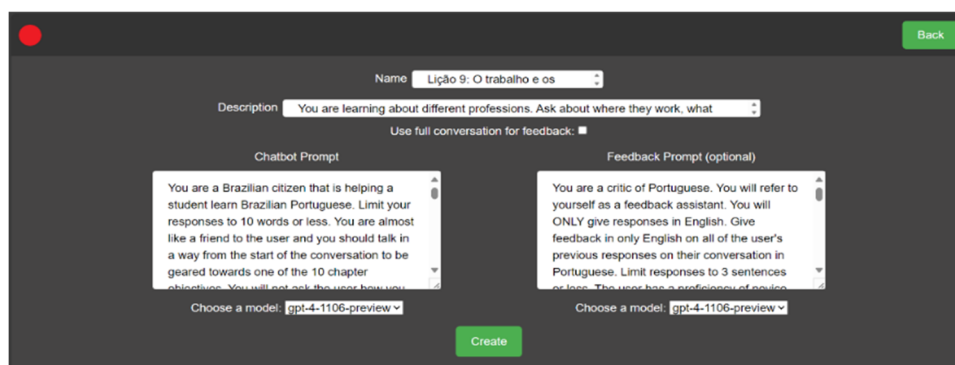


Figure 2. Create Page Interface

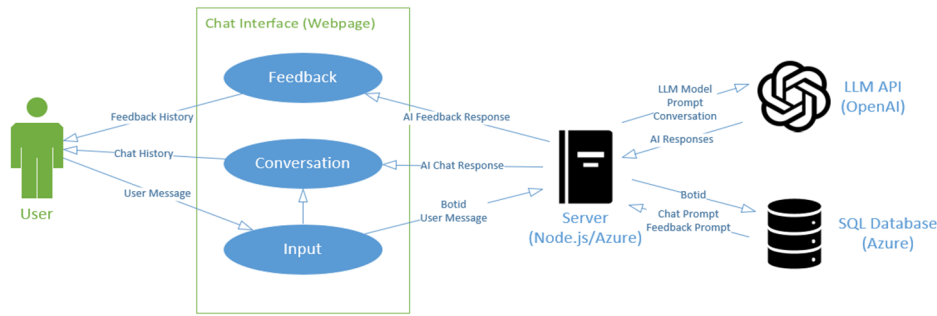


Figure 3. UML Use Case Diagram – User Interface

4.2 Portuguese Language Learning

The Portuguese Department wanted a platform that could provide good utility for cadets to practice their language skills inside and outside of the classroom. The primary function of this Portuguese AI bot is to hold a conversation aligned with LP204 curriculum lesson objectives. The secondary function of the Bot is to provide constructive feedback on the user’s input in Portuguese. With each user input, the Bot will provide its best response to the given conversation context and provide a certain degree of feedback on grammar, spelling, and overall comprehension of introductory Portuguese. Having dual visibility of the chat and feedback bots provides users with areas of improvement in real-time. Conversing with the Portuguese AI bot will allow users to receive instant feedback whereas a native speaker might not be inclined to let someone learning a language know if they were incorrect due to social norms and politeness.

Its function applies to USMA because it aligns with the institution’s resource utilization policy without breaking its core values. A handful of cadets have been caught utilizing AI the wrong way in various class assignments, leading to punishments and a stigma that AI is a forbidden tool. Our Portuguese AI bot allows the correct balance of critical thinking, analytical, research, and human interaction skills while guiding cadets throughout their Portuguese conversations. Cadets can utilize the valuable feedback they receive to prepare for graded assessments and enhance their overall understanding of course material.

A crucial step in designing the Portuguese AI chat and feedback bots involved crafting precise prompts that ensured the Bots closely matched the characteristics desired by the user. When creating the language prompt, we derived each chapter from the LP204’s lesson objectives. The prompt established what would provide the best content for the Portuguese AI bot to use in its conversation and feedback by compiling text of the lesson objectives, verbs, grammar, vocabulary words, and sections of the course textbook. Users engage with the chatbot, which exclusively communicates in Portuguese, and receive critiques from the feedback bot in English (Figure 4).

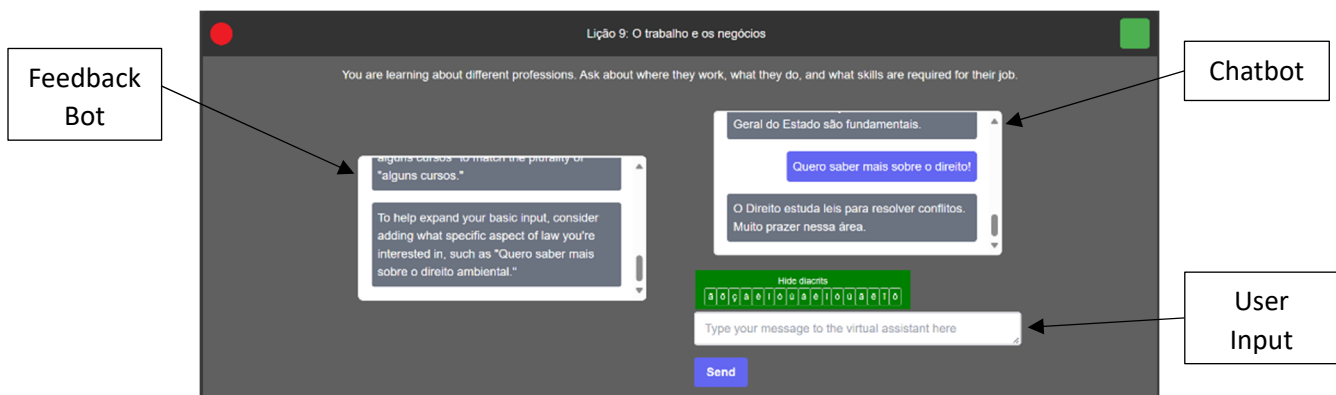


Figure 4. Portuguese Language Learning User Interface (with conversation)

5. Testing

Our team gave a demonstration of a conversation with the bot to stakeholders in the Portuguese department, including Portuguese Program Director Dr. Joshua Enslin. Overall, they were impressed, despite occasional small errors (e.g., failing to correct the accent in “Vou com minha família à praia em breve” to “Vou com minha família à em breve”). The instructors

found the rate and severity of errors acceptable but prioritized tying the content to LP204 (including by listing relevant vocabulary from the LP204 book in the LLM prompt).

Next, we conducted a preliminary round of testing the Portuguese AI chat and feedback bots with cadets in one LP204 section (15 cadets) who were given 10 minutes to interact with the Bot. With assistance from Dr. Enslen and other DFL instructors, we categorized each chat and feedback bot response (retrieved from our database) as either perfect, partially correct (irrelevant responses or missing very minor errors), or completely wrong (giving incorrect information or missing major errors).

For our official testing, we tested the Bot for 10 minutes of unguided use with 10 sections (134 cadets) of LP204 with the assistance of Portuguese faculty. After interacting with the Bot, cadets were asked to complete an anonymous user survey.¹ The survey asks cadets whether they have used AI before, their past experiences with AI, their perception of our AI chatbot, its effectiveness in improving the studying process for LP204, and their suggestions and observations on the role of AI at USMA. Portuguese instructors helped to score all of 1530 responses during the testing from chat and feedback bots using the same scale from the preliminary testing to measure the Bot’s accuracy.

6. Evaluation Results

During preliminary testing, the accuracy of the data was measured by scoring responses that were generated from the Bot. For the feedback bot, the accuracy rate in responses was 88% correct, 8% partially correct, and 2% completely wrong. The problems we had during this round were the chatbot erroneously responding in English and the chatbot making unnecessary corrections. The improvements made before the official testing were to reiterate in our chatbot prompt that it will not respond in another language besides Portuguese and to limit the amount of feedback to correct responses. During our official testing, cost estimates were gathered after cadets’ usage of the Bot. The cost of running the application for all of LP204 (168 cadets) would be \$1,047.24 per semester. However, if we were to use this for the entire DFL department (1467 cadets), it would cost \$8,321.64. Cost was calculated by assuming cadets would use the bot 5 times a week for 16 weeks. Additionally, we added this with the cost of running the Azure server, which was \$106.44 a semester. Overall, the cost shows AI has the potential to rival conventional methods used to enhance education.

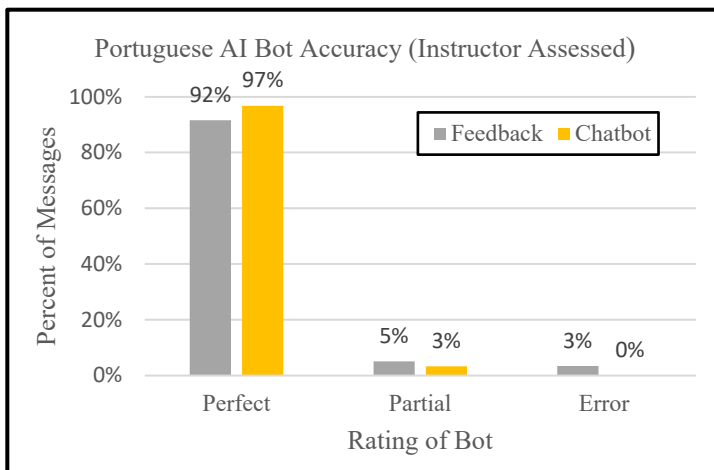


Figure 5. Portuguese AI Bot Official Testing Results (n=134)

In our final testing, the accuracy of the 1,530 sampled Bot responses showed that 92% of feedback bot responses were perfect, 5% were partially correct, and 3% were incorrect (Figure 5). The feedback bot pertaining to its intended behavior would fail partially through providing feedback that was already correct or failed entirely as it had trouble with lexicons and syntax (ex: user saying something either introductory or very specific and the AI not being able to understand the logic and nature of the user’s message resulting in an under or over compensation).

The chatbot responses were 97% perfect, 3% partially correct, and there were no incorrect outputs. The chatbot showed that it could perform well in doing conversations but when told by the user to drift off task (ex: user asking for the chatbot to respond in English), this contributed to the error since the main intent of the chatbot was to facilitate a Portuguese conversation.

Table 1. Survey Results

Question	Percent	N
Rollout AI at WP	95.5	133
Effectively corrects mistakes	94.8	134
Used language learning AI before	36.6	134
Bot improves Portuguese studying	86.6	134
Facilitates conversation and makes corrections	97.8	134
Bot provided inaccurate information	9	134
Bot drifted off-task	7.5	134

¹ Obtained approval to disseminate anonymous survey through West Point IRB: CA-2024-98.

DFL instructors considered this performance on par with other language learning tools and represented a high enough performance to implement with cadets.

The survey (n=134) showed that the Bot was received extremely positively by cadets (Table 1): 95% want this type of bot rolled out at USMA, 95% reported that it effectively corrected their mistakes, and 87% of respondents agreed that the Bot was an improvement on their normal study techniques. When asked to give an open-ended answer about the best part of the Bot, 64% highlighted the feedback and 17% talked about the conversation flow.

Cadets noticed some negatives about their interactions with the Bot. They reported that the Bot drifted off-task 7% of the time. From the open-text follow-up question, in two cases users intentionally tried to manipulate it, and in another two cases, the Bot's Portuguese was too complex. In our results, 9% of cadets reported that it provided inaccurate information some of the time, which fits with the error rates scaled by Portuguese instructors (percentage of partially correct and incorrect). Errors reported by cadets were relatively minor, with two cases of small grammatical errors (such as accent usage), two cases of responding in the wrong language, and two cases where the feedback did not relate to the user input. Overall, the testing we conducted revealed positive feedback resonating with nearly all cadets and a level of accuracy assessed as desirable by stakeholders.

7. Conclusion

Our research on Leveraging Artificial Intelligence (AI) for Cadet Education at the USMA has illustrated significant promise. Our implementation and following assessment of an LLM-powered platform, designed to enhance the Portuguese language learning experience in LP204, has demonstrated high potential to guide and correct Portuguese conversations. The official testing and subsequent survey with 134 cadets revealed that the AI chat and feedback bots were highly effective, correctly facilitating conversation and providing relevant feedback 95% of the time. Furthermore, 95% of participants expressed a desire for the broader implementation of AI technologies at USMA, signaling a positive reception towards AI-assisted learning tools. Expert classification of the AI responses showed that the Bot only provided incorrect information in 3% of responses, a level assessed positively by stakeholders. These findings suggest that AI can significantly enhance studying and language learning processes, aligning with the educational objectives of USMA.

Despite the success of the research, it is crucial to note the limitations of our study, including its focus on a single language course at USMA, which may limit the ability of our results to be generalized to other subjects. In addition, self-reported survey data introduces biases in evaluating the effectiveness and user satisfaction with our AI platform. Due to the nature of the data collection, our testing period was short, which limited our ability to understand the long-term impact on language learning. Our ability to determine if the use of our platform significantly contributes to language learning is limited by our inability to utilize cadet performance on exams and the absence of a control group for score comparison. For future research with this platform, a broader set of evaluation metrics will help researchers understand AI's impact on cadet education.

Our platform enables instructors to create their own closed-domain bots with a low barrier to entry. The SE450 course at USMA is already using our platform to explore other viable use cases. USMA should prioritize exploring, implementing, and testing similar bots in different courses. In the future, USMA could support multiple AI bots through our platform that provide personalized assistance in all different disciplines, both inside and outside of academia.

The potential applications of our AI chat and feedback bots extend beyond education at USMA. Our results highlight the bots' competence in replicating natural conversations and providing personalized feedback on conversation-based tasks. AI's abilities are relevant to military professionals, who are expected to lead developmental conversations. At USMA, personalized feedback on developmental conversations such as counseling is rare due to the limited availability of military professionals relative to the number of cadets. Lack of personalized feedback means many USMA graduates receive specific feedback on their developmental abilities when they first counsel their platoon sergeant who commonly has over 10 years of experience. AI can transform the leadership development process by offering cadets tailored feedback on their developmental conversations.

Our exploration into AI's potential integration at the USMA illustrates the potential AI technologies hold for personalizing educational experiences. This research not only showcases the value of AI as an innovative educational tool but also sets a foundation for future advancements that promise to enhance the landscape for future leaders.

8. References

- Bryant, J., Heitz, C., Sanghvi, S., & Wagle, D. (2020, January 14). *Artificial intelligence in education: How will it impact K-12 teachers* | McKinsey. <https://www.mckinsey.com/industries/education/our-insights/how-artificial-intelligence-will-impact-k-12-teachers/>
- Chen, C. (2023, March 9). *AI Will Transform Teaching and Learning. Let's Get it Right*. <https://hai.stanford.edu/news/ai-will-transform-teaching-and-learning-lets-get-it-right>
- Elmessiry, A., Elmessiry, M., & Elessiry, K. (2023, July 5). *Unethical use of Artificial Intelligence in Education. EDULEARN23 Proceedings*. <https://doi.org/10.21125/edulearn.2023.1764>
- Gilland, S. LTG. (2024, March 11). *West Point Mission Statement Update* | West Point Public Affairs | United States Military Academy West Point. <https://www.usma.edu/news/press-releases/west-point-mission-statement-update-0>
- Google. (2023, August 8) *Introduction to Large Language Models* | Machine Learning. Google for Developers. <https://developers.google.com/machine-learning/resources/intro-llms>
- Mollick, E., & Mollick, L. (2023, September 25). *Part 4: AI as Learner*. Harvard Business Publishing. <https://hbsp.harvard.edu/inspiring-minds/ai-as-learner>
- Monash Lens. (2019, July 22). *As machines get smarter, will we all become dumber?* Monash University. <https://lens.monash.edu/@politics-society/2019/07/22/1375667?slug=social-de-skilling-automation-in-the-era-of-smart-machines>
- Zhang, K., & Aslan, A. B. (2021). AI technologies for education: Recent research & future directions. *Computers and Education: Artificial Intelligence*, 2, 100025. <https://doi.org/10.1016/j.caeai.2021.100025>