Ethical Integration in Computer Science Education: Leveraging Open Educational Resources and Generative Artificial Intelligence for Enhanced Learning

Ranjidha Rajan
Computer Science Department
MSU Denver
rranjidh@msudenver.edu

Renato Cortinovis
Freelance Researcher
Italy
rmcortinovis@gmail.com

Abstract
In contemporary society, the extensive integration and dependence on computerized systems are evident across various aspects of everyday life. The training and education of the developers responsible for these systems should encompass more than just technical skills: a profound grasp of ethical considerations and the societal impact of their work is considered essential. This paper outlines an experimental approach utilizing adapted and newly developed Open Educational Resources (OER) to familiarize computer science students with the ACM Code of Ethics and Professional Conduct. These OERs, employing an underlying reusable pattern, propose assignments mandating the integration of ethical considerations into software development practices within an Inquiry-Based Learning (IBL) framework.

In the scope of these assignments, this study conducted a preliminary investigation into leveraging Generative Artificial Intelligence (gen-AI) to augment student learning and self-efficacy. This was achieved through the analysis of the data gathered from the assignments evaluation and a survey encompassing Likert scale ratings and open-ended inquiries. Factor analysis helped identifying the key themes 'Use', 'Tool Efficiency (TE)', 'Concerns (C)', 'Academic Integrity (AcI)', and 'Tool Convenience (TC)', which reflect various aspects of student engagement and perceptions of gen-AI tools. Structural Equation Modelling (SEM) further explored the relationships among these themes, suggesting that a combined 'TE' and 'TC' factor significantly enhanced user engagement with gen-AI tools. Conversely, the combined 'Concerns' and 'Academic Integrity' factors, i.e., concerns about reliability and academic dependency, did not significantly inhibit the willingness of the students to adopt gen-AI technologies.

Preliminary findings also indicate that gen-AI exhibits notable efficacy among students of moderate proficiency, albeit demonstrating underutilization among academically advanced students. Conversely, students categorized as lower-ranked tend to utilize gen-AI without exercising critical discernment. These results underscore the necessity to carefully tailor these OER to accommodate diverse student proficiency levels, thereby maximizing their educational efficacy.

1. Introduction
In contemporary society, the proliferation and reliance upon computerized systems pervades most facets of daily life. It is widely acknowledged that the training and education of the developers behind these systems extend beyond mere technical proficiency: a critical understanding of ethics and societal implications is deemed imperative. The ACM Code of Ethics and Professional Conduct is an invaluable resource that synthesises the key aspects in this regard. However, in our previous experience, many students who are strongly technically inclined tend to underestimate the importance of considering ethics and impact on social good of their work. Therefore, a first research question (RQ) for this study was:

> RQ1: What strategies could help familiarizing technically inclined computer science students with the ACM Code of Ethics and Professional Conduct?

As we will discuss, in response to this research question, assignments were designed to integrate ethical considerations into software development practices within an Inquiry-Based Learning (IBL) framework. Students were tasked with applying the ACM Code principles across progressively complex and broader scenarios. In order to support the students in these activities, we experimented the use of gen-AI tools, which prompted the following further research inquiries:

RQ2: Which are the main factors influencing positively or negatively the students' attitude in using gen-AI tool for inquiry-based learning assignments in CS classrooms?

RQ3: What is the correlation between the quality of students prompts to genAI tools and their performances in the specific IBL-base assignment and overall performance in learning?

The subsequent sections describe the learning assignments devised, the methodological approach employed to address our research inquiries, and the preliminary findings gleaned from the initial two pilot studies conducted.

2. Learning assignments

In order to familiarize students with the ACM Code, we mainly followed the approach of Fiesler et al. (2021), and Peck (2017) among others, integrating ethical considerations into traditional programming design and development assignments. This is also very similar to the CSG_ED approach of Goldweber et al. (2013), helping CS students learn concepts of computing for social good, that is, how computer and information technologies can be used to address social issues ranging from health, water resources, poverty, climate change, human rights, etc.

We considered that the IBL model was particularly appropriate to incrementally foster a deeper and broader understanding of ethical considerations and help students developing an autonomous research-oriented attitude. For this study we adopted the IBL5E variation (Duran, L. and Duran, E., 2004) articulated in the phases described in Table 1.

Phase	Purpose
Engage	Create interest and stimulate curiosity. Set learning within a meaningful context. Raise questions for inquiry.
Explore	Provide experience of the phenomenon or concept. Explore and inquire into students' questions and test their ideas. Investigate and solve problems.
Explain	Introduce conceptual tools that can be used to interpret the evidence and construct explanations of the phenomenon. Construct multi-modal explanations and justify claims in terms of the evidence gathered. Compare explanations generated by different students/groups.
Elaborate	Use and apply concepts and explanations in new contexts to test their general applicability. Reconstruct and extend explanations and understanding using and integrating different modes, such as written language, diagrammatic and graphic modes, and mathematics.
Evaluate	Provide an opportunity for students to review and reflect on their own learning and new understanding and skills. Provide evidence for changes to students' understanding, beliefs and skills.

Table 1 – Phases of the IBL5E model.

We have abstracted a generic schema for assignments based on the IBL5E model with guided and open enquiry, where gen-AI has been integrated into the Elaborate phase, to extend/improve artefacts previously developed in the Explore/Explain phases, as follows:

1. ENGAGE - Discuss the importance of ethics in computer science; critically read the ACM Code of Ethics and Professional Conduct; debate motivating and intriguing ethical dilemmas to realize that the application of the ACM Code is not necessarily straightforward: group decision making in autonomous vehicles (Awad at al., 2018), matching decisions to relevant aspects of the Code.
2. EXPLORE – Preliminarily develop an artifact [Program/UML, Diagram/ERD] about a system [on a specified topic].

3. EXPLAIN [guided enquiry] - Extend the previous artifact to cater for specified ethical implications – free support from the Internet, excluding gen-AI.
4. ELABORATE [open enquiry] – Freely identify further extensions/improvements to the artifact, integrating additional (unspecified) aspects concerning ethics – free support from gen-AI tools. Report your prompts to the tool.

The next two sections provide pertinent excerpts from the assignments utilized in our pilot studies, exemplifying the overarching schema just described.

2.1. Personalized Ads Programming

This first assignment, just slightly adapted from an existing OER published by Fiesler et al. (2021), asks students with very basic programming skills to incrementally develop a program to serve personalized ads on a social platform. In a sequence of scenarios of increasing complexity, the ads program prompts the user for information (in a real situation it would automatically extract information from the profile and posts of the users) and then return text that describes ads based on their inputs. In a first scenario, for example, the program provides text advertising dog food if the user has a dog. In another scenario, the program provides advertisements for any product that includes dogs to extrovert people, and advertisements for any product that includes cats to introvert people. In a third scenario the program provides advertisements about more or less expensive products, based on the age of the user and the estimated average income for the zip code where she lives. Here is a meaningful extract of the assignment:

> EXPLAIN – (Structured enquiry)
>
> Explain in a short report, helping yourself also with information you may search on the Internet (without using AI tools), how personalized ads work and the ethical implications, by answering the questions:
>
> - How do you feel about these kinds of inferences being used to influence your behaviour?
> - Are there ethical and unethical ways to use the technology of personalized advertisement?
> - […]
>
> ELABORATE – (partially guided enquiry)
>
> - Using the support of ChatGPT (or any genAI tool), create a short report (with your own words, avoiding cut & paste) where you:
>   - identify a new case where a personalized ad led to ethical dilemmas,
>   - pinpoint the ethical issues involved,
>   - match them to the ACM ethical code,
>   - and discuss potential solutions.
> - Include as an annex the specific prompts you submitted to ChatGPT and its responses.

2.2. Ethical Database Design

A second newly developed assignment asks students studying database design to develop an Entity Relationship Diagram (ERD) for a simplified database of a medical clinic, incrementally enhancing it by integrating aspects related to ethics. Here is a meaningful extract of the assignment:

> EXPLORE – Look into how ethical guidelines from the ACM Code of Ethics could affect the design of a database for a personalized healthcare clinic. Write a short report about it.
>
> EXPLAIN – Sketch a first ERD of the database, trying to incorporate relevant ethical aspects. Provide suitable comments to the schema, in particular making explicit any impact of the Code of Ethics on the ERD. You are allowed to make use of the Internet, excluding gen-AI tools.
>
> ELABORATE – Using ChatGPT or similar gen-AI tools, identify further enhancements to the ERD by integrating additional ethical considerations. Document the prompts (queries) submitted to the AI tool, present the extended ERD resulting from these interactions, and

provide clear comments showing how specific features of the ERD are linked to corresponding items in the ACM Code.

3. Methodology: data collection and analysis

To address RQ1, an evaluation of the students' assignment submissions was conducted to gauge their comprehension of ethical dimensions and their proficiency in effectively incorporating these principles into their software development tasks. For example, in the second pilot study, the assessment schema employed was as follows:

> EXPLORE, EXPLAIN: check ERD correctness, discussion of relevant ethical considerations, integration of ethical considerations into ERD.
>
> ELABORATE: check ethical considerations extensions, corresponding ERD extensions.

To address RQ2, we collected data from Likert questions concerning the students' attitude in using gen-AI tools, and their answers to open-ended questions about their experience with the assignments. Data from Likert questions were analysed and refined with factor analysis and subsequent thematic analysis to reveal the main factors affecting their perceptions.

In this study, we employed Principal Component Analysis (PCA) as the extraction method for factor analysis to identify underlying dimensions within the dataset (Wetzel, 2012). We utilized Promax rotation with Kaiser Normalization to allow the factors to be correlated, enhancing interpretability in a framework where constructs may be interrelated (Grieder & Steiner, 2022). The selection of variables and the number of factors retained were based on their ability to meaningfully explain the covariance among observed variables.

Using Structural Equation Modelling (SEM), the interrelationships among these themes were further elucidated (Goldberger, 1972). The Maximum Likelihood (ML) estimation method was used to identify the best-fitting model, a standard approach in SEM due to its efficiency and robustness.

Concerning RQ3, we collected the prompts submitted by the students to the gen-AI tool, which were classified as Descriptive, Comparative, Inquisitive/Exploratory, Ethical/Philosophical Inquiry, Case Study, Focused, and Instructional. Descriptive prompts aim at eliciting detailed narratives or explanations (Cave et al., 2020). Comparative are prompts that encourage comparison between concepts or examples (Sutton & Barto, 2018). Inquisitive/Exploratory are prompts designed to probe deeper understanding or exploration of a topic (Lake et al., 2018). Ethical/Philosophical Inquiry are prompts that delve into ethical considerations or philosophical questions (Bostrom, 2014). Case Study Focused are prompts asking for specific examples, case studies, or applications (Silver et al., 2016). Instructional are prompts that guide the AI in performing a specific task or generating content in a certain way, reflecting the few-shot learning capabilities mentioned in AI research (Schick & Schütze, 2022).

Accordingly, each prompt was analysed using specific criteria where descriptive prompts asked for explanations, comparative prompts involved comparisons, inquisitive prompts sought understanding or exploration, ethical prompts focused on moral or ethical considerations, and instructional prompts provided summaries or instructions. Walter (2024) claims that prompt classification is crucial for understanding student learning with gen-AI tools, as it enables the identification of specific areas where the tool enhances educational outcomes, allowing for targeted improvements and better support for diverse learning needs. The prompt classification process was automated based on text analysis to label the prompts accordingly. The classified data were then visualized using a pivot table and bar chart to illustrate the distribution of prompt classifications by grade, providing insights into the cognitive and analytical skills development of students at different educational levels. Finally, we correlated the classification scores with their grades in the assignment.

4. Preliminary results and discussion

The analysis was carried out on a limited sample of 27 students. Concerning RQ1, the assignments' evaluation showed that the proposed generic schema, providing students with the opportunity to concretely apply the ACM code of ethics in progressively complex scenarios and with tools of increasing power, supported the students in developing an increasing level of understanding moving through the assignment phases. Students also explicitly appreciated the integration of IBL5E and gen-AI. One student stated, for example: "*I appreciate the school's interest in embracing innovations and its dedication to enhancing teaching methods*", while another commented "*These new teaching methods, including the use of AI, should be incorporated more frequently into other lessons*".

Concerning RQ2, the examination of Likert scale data through factor analysis (Figure 1) facilitated the identification of several themes. The pattern matrix resulting from Principal Component Analysis (PCA) (Wetzel, 2012) with Promax rotation and Kaiser normalization reveals the underlying factor structure of the dataset (Grieder & Steiner, 2022).

Here is a meaningful extract of the assignment:

Pattern Matrix[a]

	Component				
	1	2	3	4	5
Q18	.912				
Q12	.856				
Q19	.842				
Q21	.718				
Q20	.600				
Q3		.899			
Q5		.784			
Q7		.775			
Q1		.728			
Q13			.901		
Q10			.834		
Q15				.925	
Q14				.694	
Q11					.816
Q8					.788

Extraction Method: Principal Component Analysis.
Rotation Method: Promax with Kaiser Normalization.

a. Rotation converged in 6 iterations.

Figure 1 – Pattern matrix after removing irrelevant questions

Five distinct components were extracted, each representing a unique factor. Component 1 shows strong loadings for Q12, Q18, Q19, Q20, and Q21, indicating a shared underlying factor. Q2, Q4, Q6, Q9, Q16, Q17 were removed for better factor loading. Component 2 is defined by high loadings on Q5, Q3, Q1, and Q7, suggesting another common factor. Component 3 includes significant loadings for Q10, and Q13, highlighting a third distinct factor. Component 4 is characterized by loadings on Q15, and Q14, with Q16 showing a particularly high negative loading. Component 5 has a notable loading for Q8 and Q11 pointing to additional unique factors. Each component represents a different underlying factor extracted from the data set, where the numbers indicate the strength of the association between each question and the corresponding component. For instance, Q18 has a strong loading on Component 1 (0.912), suggesting it is closely related to that factor, while Q3 has a high loading on Component 2 (0.899). The matrix indicates which questions are grouped together under each component, helping to identify patterns and underlying structures in the data. The rotation method used ensures that the components are more interpretable by allowing them to be correlated. This pattern matrix effectively elucidates the factor structure, with each component representing a distinct underlying dimension measured by the variables in the dataset.

The resulting themes identified were 'Use', 'Tool Efficiency (TE)', 'Concerns (C)', 'Academic Integrity (AcI)', and 'Tool Convenience (TC)'. These themes encompass diverse dimensions of student engagement and perceptions pertaining to gen-AI tools. The survey questions were grouped based on the previous themes identified. The sample for survey question grouping for TE is given in Figure 2.

Tool Efficiency (TE)

ChatGPT provided quick and relevant responses to my inquiries. (Q1)

Using ChatGPT saved me time during the research process. (Q3)

I felt more confident about my research findings after using ChatGPT. (Q5)

I would prefer using ChatGPT over traditional research methods for future assignments. (Q7

Figure 2 – Sample Thematic factors based on factor loadings

SEM analysis was conducted using the maximum likelihood estimation method. It yielded a log likelihood of -84.41, suggesting a relatively good model fit (Figure 3). The variable "te_tc" demonstrated a strong positive and statistically significant relationship with the dependent variable, with a coefficient of 0.88 and a p-value of 0.001. This indicates that changes in "te_tc" are likely to have a significant impact on “use”. In other words, the amalgamation of 'TE' and 'TC' significantly impact on user engagement with gen-AI tools. Conversely, the variable "c_ai" showed a positive relationship with a coefficient of 0.32, but this was not statistically significant with a p-value of 0.226, showing that the influence by “c_ai” on "Use" is not strongly supported by the current sample size of data. That is, the amalgamation of 'C' and ‘AcI' factors, indicative of apprehensions regarding reliability and academic dependence, did not markedly impede students' propensity to adopt gen-AI technologies.

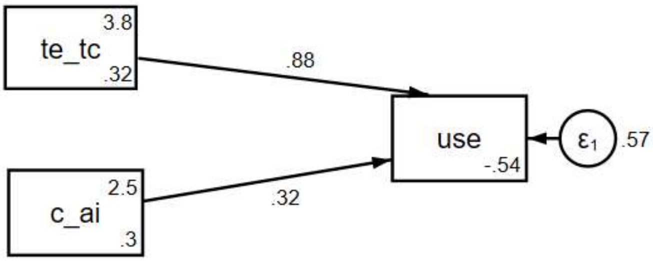

Figure 3 – Structural Equation Modelling

Concerning RQ3, the categorization of students’ prompts to ChatGPT resulted in the identification of three distinct groups: firstly, the highest-performing students effectively showcased a tangible integration of the ACM Code principles into their database design, solely relying on their own capabilities without resorting to ChatGPT assistance. Secondly, students with moderate proficiency levels attained comparable outcomes by utilizing ChatGPT, engaging in meaningful interactions that facilitated their progress. Lastly, a minority of students with very limited abilities demonstrated minimal interaction with GenAI: they just submitted the whole assignments directly to the tool without personalized interventions. This supports the claim that gen-AI exhibits notable efficacy among students of moderate proficiency, albeit demonstrating underutilization among academically advanced students. Conversely, students categorized as lower-ranked tend to utilize gen-AI without exercising critical discernment. These results underscore the necessity to carefully tailor these OER to accommodate diverse student proficiency levels, thereby maximizing their educational efficacy.

The analysis of prompt classifications across different grades revealed distinct trends in the focus and complexity of student inquiries (Figure 4). Grade 6 exhibits a higher frequency of descriptive prompts, reflecting a focus on foundational understanding and detailed descriptions. Ethical and philosophical inquiries are evenly distributed among Grades 4, 6, and 8, indicating a consistent engagement with moral and ethical considerations across these levels. Grade 8, however, shows a notable increase in inquisitive and exploratory prompts, suggesting a shift towards more critical and analytical thinking as students advance. Instructional prompts are unique to Grade 4, perhaps indicative of an emphasis on summarization and concluding thoughts at this stage. Interestingly, unclassified prompts appear predominantly in Grades 4 and 8, with Grade 8 having the highest number, which could reflect the more open-ended and complex nature of discussions at this level. These findings underscore the progression in cognitive and analytical skills development as students move through different educational stages.

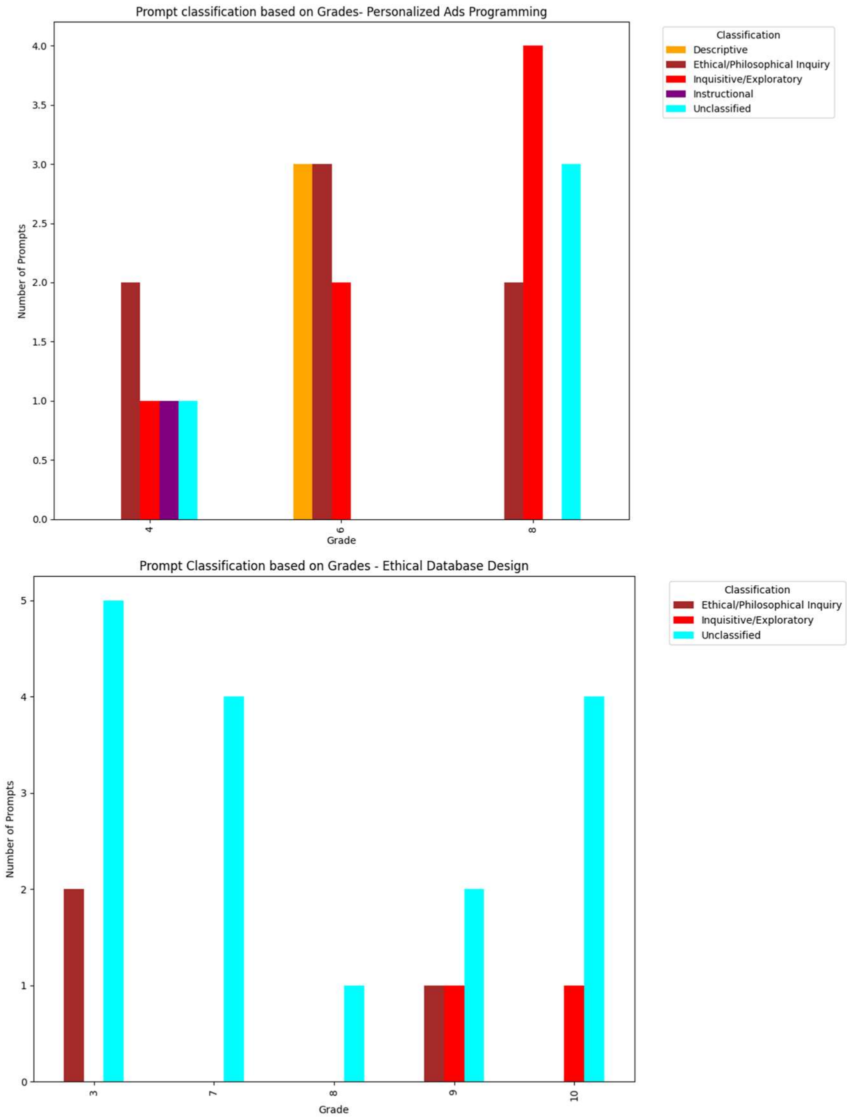

Figure 4 – Prompt classification samples

The data from the database design assignment and their analysis findings show that the majority of prompts at lower grades, especially Grade 3, remain unclassified, indicating a need for more specific guidance or focus in the questions posed. Higher grades tend to have more focused inquiries, with Grade 8 showing a balanced mix of inquisitive and ethical inquiries. This likely reflects the naturally expected deeper reasoning exhibited by the most proficient students. However, further research is needed to determine how best GenAI could contribute to improving critical thinking and understanding across all students.

This analysis helps in understanding the developmental trends in cognitive and analytical skills among students, and how they engage with ethical considerations in their academic tasks.

5. Conclusions

The utilization of an existing OER in the first pilot proved pivotal for the undertaken activity, underscoring the significance of OER in both research and educational endeavours. This prompted our decision to openly publish the resource developed in the second pilot as an OER, with further intentions to create additional resources, capitalizing on the identified seemingly effective overarching framework. The student reception of the assignments outlined has been positive, with active engagement noted and an explicit appreciation for the IBL5E model and the incorporation of gen-AI. The data analysis identified themes such as 'Use', 'Tool Efficiency', 'Concerns', 'Academic Integrity', and 'Tool Convenience', which represent various aspects of student engagement and perceptions of generative AI tools. SEM analysis revealed that the combination of 'Tool Efficiency' and 'Tool Convenience' significantly enhanced user engagement, while 'Concerns' and 'Academic Integrity' did not substantially deter students from adopting generative AI technologies. These insights highlight the complex interplay between efficiency, convenience, and apprehensions in shaping students' adoption of generative AI tools. Nonetheless, despite garnering some preliminary findings, these outcomes necessitate validation and expansion through subsequent studies involving students from diverse contexts, larger sample sizes, and varied assignments.

6. References

Awad, E., Dsouza, S., Kim, R. *et al*. (2018). The Moral Machine experiment. Nature 563, 59–64. https://doi.org/10.1038/s41586-018-0637-6

Bostrom, N. (2014). Superintelligence: Paths, Dangers, Strategies. Oxford University Press.

Cave, S., Dihal, K. & Dillon, S. (2020). AI Narratives: A History of Imaginative Thinking about Intelligent Machines. Oxford University Press.

Duran, L., and Duran, E. (2004). The 5E Instructional Model: A Learning Cycle Approach for Inquiry-Based Science Teaching. The Science Education Review, 3(2), p49–58

Fiesler, C., Friske, M., Garrett, N., Muzny, F., Smith, J., and Zietz, J. (2021). Integrating Ethics into Introductory Programming Classes. Proceedings of the ACM SIGCSE Conference on Computer Science Education.

Goldberger, A. S. (1972). Structural Equation Methods in the Social Sciences. Econometrica, 40(6), 979–1001. https://doi.org/10.2307/1913851

Goldweber, M., Barr, J., Clear, T., Davoli, R., Mann, S., Patitsas, E., & Portnoff, S. (2013). A Framework for Enhancing the Social Good in Computing Education: A Values Approach. ACM Inroads, 4(1), 58-79.

Grieder, S., & Steiner, M. D. (2022). Algorithmic jingle jungle: A comparison of implementations of principal axis factoring and promax rotation in R and SPSS. Behavior research methods, 54(1), 54–74.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. Behavioural and Brain Sciences, 40, E253.

Peck, E. (2017, July 5). The Ethical Engine: Integrating Ethical Design into Intro Computer Science. Blog, Bucknell HCI, 5.

Schick, T., & Schütze, H. (2022). True Few-Shot Learning with Prompts—A Real-World Perspective. Transactions of the Association for Computational Linguistics; 10 716–731.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. Nature, 529(7587), 484-489.

Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction. MIT Press.

Walter, Y. (2024). Embracing the future of Artificial Intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. International Journal of Educational Technology in Higher Education, 21(1), 1-29. https://doi.org/10.1186/s41239-024-00448-3

Wetzel A. P. (2012). Factor analysis methods and validity evidence: a review of instrument development across the medical education continuum. Academic medicine: journal of the Association of American Medical Colleges, 87(8), 1060–1069. https://doi.org/10.1097/ACM.0b013e31825d305d