Optimizing Routine Educational Tasks through Prompt Engineering: A Comparative Study of AI Chatbots

Svitlana Skvortsova[1], Tetiana Symonenko[2] and Kira Hnezdilova[3]
[1]*Department of Mathematics and Methods of its Teaching, The State Institution "South Ukrainian National Pedagogical University named after K. Ushynsky", Staroportofrankyvska Str. 26, 65020 Odesa, Ukraine*
[2]*Department of Ukrainian Philology and Social Communication, The Bohdan Khmelnytsky National University of Cherkasy, Shevchenko Avenue 81, 18031 Cherkasy, Ukraine*
[3]*Department of Primary Education, The Bohdan Khmelnytsky National University of Cherkasy, Shevchenko Avenue 81, 18031 Cherkasy, Ukraine*
skvo08@i.ua, irinka200888@i.ua, kiragnez@gmail.com

Keywords: AI in Education, Prompt Engineering, Chatbot Evaluation, Lesson Planning Automation, AI-Assisted Teaching.

Abstract: The rapid integration of artificial intelligence (AI) in education necessitates the development of effective strategies for optimizing routine teaching tasks. This study explores the relevance of prompt engineering as a tool for enhancing AI-generated responses, reducing educators' workload, and improving the efficiency of lesson planning, content creation, and assessment design. The primary objective of this research is to develop and evaluate a structured methodology for designing prompts that maximize the relevance, completeness, and applicability of AI-generated outputs. To achieve this goal, a three-phase methodology was employed: (1) a preparatory phase involving a literature review and the development of standardized educational prompts, (2) an experimental phase testing these prompts across multiple AI chatbot models (Claude, GPT, and Copilot), and (3) an analytical phase assessing chatbot responses based on predefined criteria, including relevance, accuracy, completeness, practicality, and structuredness. The results indicate significant differences in chatbot performance. Claude demonstrated superior contextual understanding, GPT provided well-balanced and structured responses, while Copilot exhibited high factual accuracy but required improvements in contextual adaptation. Statistical analysis using the Kruskal-Wallis H test confirmed these variations, highlighting the necessity of model-specific prompt optimization. The study's findings have both practical and theoretical significance. Practically, they provide educators with a structured approach to prompt engineering, enabling more effective use of AI tools in teaching. Theoretically, the research contributes to the growing field of AI-assisted education by offering insights into optimizing human-AI interaction. The conclusions emphasize the need for continued refinement of AI models and further exploration of prompt engineering techniques. Future research should focus on expanding testing across various disciplines and integrating AI-driven tools into digital learning environments to enhance personalized education.

1 INTRODUCTION

The rapid development of digital technologies has significantly impacted the cognitive processes of modern students, particularly Generation Alpha, who are accustomed to constant digital exposure. This has led to "clip thinking," where attention shifts quickly between content, posing challenges for deep focus and long-term retention [8]. Traditional educational methods may not align with their needs, necessitating innovative approaches. Integrating generative language models and prompt-based technologies into education can optimize routine tasks, enhance teaching efficiency, and create customized materials that resonate with digital learners [1], [5].

This study explores how prompt-writing technologies can optimize educational tasks and create engaging learning materials for Generation Alpha, leveraging AI models like GPT, Claude, and Copilot. The research demonstrates AI's potential in optimizing pedagogical activities, creating educational content, personalizing learning experiences, and increasing student engagement. Effective prompt engineering is crucial for successful

AI interaction, emphasizing the need for precise and structured prompts [2]. The study aims to develop and evaluate a structured methodology for creating prompts that optimize routine educational tasks.

Research Questions:

1) How do different AI chatbots (Claude, GPT, Copilot) respond to structured educational prompts?
2) What factors influence the accuracy, completeness, and applicability of AI-generated responses?
3) How can prompt engineering be refined to maximize its effectiveness in an educational setting?

2 PROMPT ENGINEERING IN EDUCATION

The process of designing, optimizing, and refining prompts to enhance the efficiency of generative AI models is known as prompt engineering. At the current stage of technological development, the prompt has evolved from a simple “query” to a crucial tool for interacting with LLMs (large language models). Prompt engineering now encompasses not only textual data but also other modalities, such as image, audio, and video processing. The prompt is input text or other forms of data provided to a language model (or another generative system) to initiate the generation of an appropriate response. The structure of a prompt may include instructions, questions, examples, context, or a combination of these elements. The functional purpose of a prompt is to set the parameters and context for the model’s operation, guiding it toward generating the desired response or behavior. There are two primary types of prompts: textual (instructions or queries) and multimodal (combinations of text with images, audio, etc.).

According to Schulhoff et al. [7], the prompt engineering process for maximizing the potential of LLMs and obtaining relevant, accurate, and useful results with minimal effort involves the following stages:

1) Design. Formulating a prompt for a specific task considering the goal, target audience, and context.
2) Optimization. Applying various techniques such as few-shot prompting (adding a few examples of desired responses to the main query), chain-of-thought prompting (encouraging the model to generate a sequential logical thought process before forming the final answer), or specifying response formats to improve accuracy.
3) Iteration. Experimental testing of different prompt variations and analyzing the obtained results.
4) Adaptation. Adjusting prompts based on the specific characteristics of the model and changes in task requirements.

The publication "The Prompt Report: A Systematic Survey of Prompting Techniques" [7] provides recommendations for creating prompts for LLMs like ChatGPT and discusses issues related to safety and reliability. Based on these recommendations, the following requirements for the prompt creation process can be formulated:

1) Clarity and specificity.
2) Contextualization.
3) Use of examples (few-shot prompting).
4) Task decomposition into steps (chain-of-thought prompting).
5) Specification of response format.
6) Avoiding ambiguous terms.
7) Testing and iteration.
8) Consideration of ethical aspects.

According to Wei et al. [9], prompt engineering is a tool for adapting language models to specific contexts. For example, in solving educational tasks, prompts should consider subject specificity and the students’ level of knowledge, among other factors.

Moreover, we agree with authors who view prompt engineering as an interdisciplinary skill that combines knowledge about query structure, AI tools, and task specifics. This position deserves attention since writing high-quality prompts in a particular field context requires subject knowledge, certain linguistic skills, logical thinking, and more. This is where the complexity and multidimensionality of prompt engineering lie. When developing prompts, it is essential to understand how linguistic nuances impact the capabilities of generative AI, ensuring the creation of authentic and well-adapted content for effective teaching and learning interactions [4].

Bozkurt and Sharma [3] emphasize the importance of developing prompt engineering skills among educators to effectively harness the full potential of generative AI in educational contexts.

They argue that co-creation involving generative AI represents a powerful approach in education, highlighting the significance of human-machine interaction facilitated by carefully crafted prompts.

Approaches to writing prompts often include using clear and specific instructions, keywords, tone, and response style settings. However, essential elements for creating a quality prompt must be highlighted:

1) Instruction. A specific task the user wants the AI to perform, such as “write a motivational speech for a lesson” or “create a story.”
2) Context. Information that can guide the model to provide more accurate answers, such as the lesson topic, objectives, or story theme and style.
3) Input data. Detailed contextual information, such as structural elements of the lesson or specific problem types, or detailed descriptions of characters in a story. Providing examples of desired outcomes is also appropriate.
4) Output type or format. For example, “a fantasy story of 400 words.” Currently, no definitive guidelines exist for creating quality prompts for educational purposes. However, it is essential to avoid unclear and ambiguous formulations since AI may misinterpret the request and produce unintended results. The absence of context or grammatical errors can also lead to misunderstandings by the AI. Experimenting with different formulations to find the best approach for your task is crucial.

Bozkurt and Sharma [3] suggest strategies for creating effective prompts:

1) Clearly define the goal. Specify the purpose, desired response type, or result.
2) Understand the AI model’s capabilities. Leverage the model’s strengths and limitations, creating prompts aligned with its expertise. Role prompting (assigning specific roles) can sometimes “break” the model’s default behavior.
3) Use concise and clear formulations. Avoid confusing or irrelevant prompts.
4) Provide sufficient context. Enable the AI model to better understand the task or subject.
5) Use examples of desired outcomes. Show the AI what kind of output is expected.
6) Fine-tune and debug prompts. Make adjustments for improved results.
7) Specify output format or structure. Provide clarity on how the answer should look.
8) Include key details. Ensure the AI receives all necessary information.
9) Test different prompt variations. Identify the most effective formulation.
10) Consider safety and ethical aspects. Maintain responsible AI usage practices.

According to the authors, adhering to these strategies allows educators, researchers, and users to optimize prompt engineering for meaningful and accurate responses from language models, aligning with their specific goals and requirements. They underscore that effective prompt engineering is not merely a technical skill but an art of communication that requires understanding AI’s technical capabilities and the nuances of human language and interaction [3].

2.1 Technology for Creating Prompts for Optimizing Routine Tasks in Education

Based on the analysis of scientific literature on the issues of prompt engineering, a technology for creating prompts has been developed, aimed at optimizing routine tasks in the educational sphere. Prompts, as instructions for artificial intelligence (AI) systems, allow automating such pedagogical processes as developing lesson plans, creating learning tasks, preparing didactic materials, checking work, etc. The prompt creation process is considered according to the main stages: design, optimization, iteration, and adaptation.

Design Stage. Forming a Basic Prompt The prompt creation stage involves designing an initial request, which includes formulating a clear requirement (determining the desired result), specifying the topic and goal, determining the response format and target audience. Additional details can increase the effectiveness of the prompt.

For example, “Create 5 mathematical problems for 2nd grade students on the topic ‘Addition and subtraction of two-digit numbers.’ The level of complexity is simple problems that are solved by arithmetic operations of addition and subtraction. The format is plot problems with practical content.”

Optimization Stage. Correction and Clarification of the Prompt The next stage is the evaluation of the result – the generated AI answer. The evaluation includes checking the correspondence of the response to the query, its correctness (truthfulness) and the logic of the information presented. In the case of partial or complete non-compliance of the result with the user's expectations, the prompt is optimized. Optimization may include: adding clarifications, breaking the task into subtasks, providing examples, indicating specific requirements. For example, the optimization of the previous prompt may look like this: "Additional requirements: each task must contain a context that describes situations from the school life of students; the formulation of the task must correspond to the canonical structure (first the condition, then the question); include visual elements (diagrams, tables)."

Usage and Iteration Stage After obtaining a satisfactory result, the prompt can be generalized (formed into a template) and used to perform similar routine tasks. For the purpose of effective use, it is advisable to save successful prompts for further use. In the process of reusing the prompt, there may be a need to improve it based on the accumulated experience. Experimenting with different prompt formulations is a process of iteration. This stage is not mandatory and is performed as needed.

2.1.1 Prompt Creation Algorithm

The prompt creation process can be presented as the following algorithm:

1) Formulation of the basic query with additional details (if necessary):
   - Clear requirement/question.
   - Topic and goal/objective.
   - Type and format of the response.
   - Target audience.
   - Additional details.
2) Evaluation of the quality of the response to the basic query:
   - Relevance of the query.
   - Completeness of the response.
   - Correctness of the information.
   - Practical applicability.
3) Optimization (if necessary):
   - Adding clarifications.
   - Breaking down into subtasks.
   - Providing examples.
   - Indicating chain-of-thought prompting.
   - Assigning the role of AI.
4) Checking the result (after optimization):
   - Relevance of the request.
   - Completeness of the answer.
   - Correctness of the information.
   - Practical applicability.
5) Iteration (if necessary):
   - Clarifying the requirements.
   - Adding details.
   - Changing the format.
   - Correcting the structure.

For educators new to creating prompts for AI systems, it is recommended to start with a basic prompt and gradually add contextual details. Using standardized templates can facilitate this process. The following template is suggested:

1) Task: [clearly formulate the required action or result].
2) Topic: [specify the subject area or topic of the task].
3) Target audience: [describe the characteristics of the audience (e.g., grade, level of knowledge, age)].
4) Format: [determine the desired format of the output (e.g., list, table, essay)].
5) Additional requirements: [specify specific requirements such as volume, style, level of detail, visual elements].

This template helps structure the request and provides necessary information for correct AI interpretation. Avoid the following mistakes:

1) Overly general query. Vague wording leads to ambiguous results. For example, instead of "Create a math problem," use "Create three problems to find the area of a rectangle for 5th graders."
2) Conflicting requirements. Contradictory instructions can confuse the AI and lead to incorrect answers.
3) Overloading with details. Too many details at the initial stage can complicate the prompt creation process. Add details gradually.
4) Examples of optimized prompts. Overly general wording:
   - Incorrect: "Create a math problem."
   - Optimized: "Create 5 math problems: Topic: adding fractions, Grade: 6, Difficulty level: medium, Problem types: 2 on calculations, 3 text problems."
5) Conflicting requirements:
   - Incorrect: "Solve the problem quickly, but explain each step in detail."
   - Optimized: "Solve the problem of finding the area of a triangle with sides 3, 4, 5. Use Heron's formula. Show intermediate calculations. Write down the answer with an explanation."

Note that various prompt optimization techniques can also be used at the stage of designing the basic prompt, depending on the task at hand:

1) Step-by-step process description: chain-of-thought prompting:
2) Sample response: few-shot prompting.
3) Specific role or expertise: role prompting.
4) Complex structure: structuring + chain-of-thought.
5) Creative task: role prompting + few-shot.
6) Solving complex problems: chain-of-thought + few-shot.

This approach allows teachers to master prompt engineering skills and effectively use AI in educational activities. The proposed algorithm optimizes the process of creating prompts for AI, ensuring relevant and useful results for routine educational tasks.

2.2 Methodology

2.2.1 Research Design

The study follows a three-phase approach:

1) Preparatory stage. Literature review, identification of routine educational tasks, and development of structured prompts.
2) Experimental stage. Testing prompts across Claude, GPT, and Copilot chatbots, documenting response quality.
3) Analytical stage. Comparative evaluation of chatbot responses using predefined assessment criteria.

2.2.2 Research Methods

At the preparatory stage, existing approaches to prompt engineering and typical errors when creating prompts were studied using the following methods:

1) Analysis of scientific publications, articles, reports, and API documentation of chatbots using search engines: Web of Science, Scopus, Google Scholar, Researchgate, and other scientific databases.
2) Analysis of the teacher's pedagogical activity to identify typical routine tasks in the lesson, develop prompt templates for different types of activities and stages of the lesson, and model the teacher's activity in developing tasks from different subject areas (e.g., mathematics, Ukrainian language).

At the experimental stage, the effectiveness of the developed technology was assessed by comparative analysis of the responses of different chatbots to the same prompts. Experimental testing involved applying identical prompts to different chatbot systems and further analysis of the generated responses with subsequent optimization and iteration.

Meta-prompts were developed to solve certain routine teacher tasks, and data on the responses of different chatbots to the same prompts were collected.

At the analytical stage, quantitative processing of the obtained results was carried out, comparative analysis of the responses of different chatbots by each criterion was conducted, and mathematical statistics methods were applied to identify statistically significant differences between the results.

Preparatory Stage of the Study. At this stage, chatbot models were selected for further testing. The results of a survey of teachers at Ukrainian universities [8], publications by scientists from other countries [1], [5] showed the high popularity of Chat GPT among Ukrainian educators, correlating with global trends. Given the integration of the Copilot chatbot into the Microsoft 365 corporate suite, widely used in Ukrainian universities, the choice of Chat GPT and Copilot for testing the developed prompt creation technology is justified.

Based on the survey results and subjective preferences, the Claude chatbot was chosen as the third object of the study [8]. At the time of the study, limited scientific publications were found on the spread of the Claude chatbot among teachers in other countries, while most research focuses on Chat GPT.

Thus, the result of the preparatory stage was the selection of three chatbots for experimental testing: Chat GPT, Copilot, and Claude.

A set of test tasks in mathematics and the Ukrainian language was created, for which the corresponding basic prompts were developed. The teacher's activities in organizing students' educational and cognitive activities in the lesson were analyzed based on the generally accepted structure of a combined lesson, which includes the following stages:

1) the motivational stage of the lesson;
2) the stage of updating knowledge and methods of activity;
3) the stage of formation of new knowledge and methods of action;
4) the stage of consolidation and formation of skills and abilities;
5) the stage of lesson results and reflection of educational and cognitive activity.

Within each stage, various routine tasks are implemented, related to different types of student activities (e.g., oral survey, mathematical or spelling dictation, individual survey, homework check) and different forms of organization of educational activity (collective, pair, group, individual). Considering the characteristics of modern students as representatives of the digital generation, one task a teacher can solve with AI is creating interactive learning environments, such as lesson shells or individual stages of lessons in the format of journeys, quests, competitions.

The following criteria were defined to assess the effectiveness of the prompt:

1) Relevance (K1) – alignment with the initial prompt.

2) Completeness (K2) – depth of the response.
3) Accuracy (K3) – consistency with educational standards.
4) Practicality (K4) – direct applicability in lesson planning.
5) Structuredness (K5) – logical organization of the response.

A five-point evaluation scale is used for these criteria. For example, for criterion K1 (Request compliance):

- 5 – full compliance with all prompt requirements;
- 4 – compliance with the main requirements with minor deviations;
- 3 – partial compliance with the requirements;
- 2 – significant deviations from the requirements;
- 1 – minimal compliance;
- 0 – complete non-compliance.

At the preparatory stage, a form was developed for recording the results of the responses of the three studied chatbots – Chat GPT, Copilot, and Claude. This form allows systematizing and comparing the obtained data, ensuring the objectivity and scientific validity of the study.

Using the developed technology, 12 generalized prompts were created to solve routine tasks that may arise in the professional activities of teachers of any subject. At this stage of the study, experimental testing was conducted, which consisted of applying these prompts to different chat models. For this, the generalized prompts were detailed on the material of two academic subjects: mathematics and the Ukrainian language.

2.2.3 Analytical Stage of the Study

At this stage, a quantitative analysis of the results was conducted, including a comparative evaluation of responses generated by different chatbot models based on the prompt evaluation criteria. Additionally, statistical processing of the testing results for the prompt creation technology was performed. This enabled the interpretation of the experimental data and the formulation of conclusions regarding the effectiveness of the developed technology.

To assess the differences in responses produced by various chatbot models (ChatGPT, Copilot, Claude) based on prompts designed using the proposed methodology, the Kruskal-Wallis H test was applied at a significance level of $\alpha = 0.05$. The selection of this statistical test is justified by its capability to determine variations in response evaluations when switching between chatbot models.

The general hypothesis of the study was formulated as follows: using the developed technology, prompts were generated that allow obtaining relevant responses from different chatbot models according to predefined criteria.

The following statistical hypotheses were tested at a significance level of $\alpha = 0.05$:

1) H0 (null hypothesis). There is no statistically significant difference in the scores of responses (by a specific criterion) generated by different chatbot models based on prompts developed using the proposed technology.
2) H1 (alternative hypothesis). There is a statistically significant difference in the scores of responses (by a specific criterion) generated by different chatbot models based on prompts developed using the proposed technology.

Kruskal-Wallis H test calculations were performed separately for responses to prompts used in Ukrainian language lessons and separately for mathematics lessons.

According to the results presented in Table 1, the null hypothesis was rejected in all cases, confirming that there are statistically significant differences in the evaluation of responses (by specific criteria) generated by different chatbot models based on prompts developed using the proposed technology at a significance level of $p \leq 0.01$.

Pairwise comparisons between chatbots for each criterion were also performed. The results indicated statistically significant differences ($p \leq 0.01$) in the assessment of prompts between Claude and Copilot for criteria K1 (Relevance to request) and K3 (Information accuracy). Additionally, significant differences were observed between Claude and ChatGPT for criterion K4 (Practical applicability), as well as between Claude and ChatGPT ($p \leq 0.01$) and Claude and Copilot ($p \leq 0.05$) for criterion K5 (Response structure). In the evaluation by criterion K2 (Completeness of the response) insignificant differences between Claude and Copilot were found ($p \leq 0.02$).

Table 1: Kruskal-Wallis test results (K) Ukrainian language.

Criterion (K)	Test statistic, asymptotic significance (2-sided)
K1	9.564, p= 0.08
K2	6.230, p=0.044
K3	7.582, p=0.023
K4	10.404, p=0.006
K5	9.406, p=0.009

In the case of testing prompts for mathematics lessons (Table 2) according to criteria K1, K2, K3, and K4, the null hypothesis (H_0) was accepted, indicating no statistically significant differences in responses generated by different chatbot models based on the developed prompts. Therefore, pairwise comparisons were not conducted for these criteria.

Table 2: Kruskal-Wallis test results for mathematics.

Criterion (K)	Test statistic, asymptotic significance (2-sided)
K1	2.722, p=0.256
K2	2.191, p=0.334
K3	1.444, p=0.486
K4	0.105, p=0.949
K5	6.106, p=0.047

However, for criterion K5 (Response structure), the null hypothesis was rejected ($p \leq 0.05$), confirming significant differences in the evaluations of chatbot responses. Pairwise comparisons revealed significant differences between Claude and Copilot ($p \leq 0.05$). The results of the calculations are shown in Table 3.

The statistical analysis confirmed the effectiveness of the developed prompt creation technology for chatbot responses. For Ukrainian language prompts, significant differences were found among chatbot responses for all evaluation criteria, particularly between Claude and Copilot, as well as Claude and ChatGPT. In contrast, for mathematics prompts, significant differences were observed only for the criterion of response structure (K5), specifically between Claude and Copilot.

Table 3: Pairwise comparisons for K5 (mathematics prompts).

Comparison	Test statistic, SE, z, p-value
Copilot - ChatGPT	4.083, SE = 2.958, z = 1.380, p = 0.167
Copilot - Claude	7.292, SE = 2.958, z = -2.465, p = 0.014
ChatGPT - Claude	3.208, SE = 2.958, z = -1.085, p = 0.278

These findings suggest that the effectiveness of the prompt creation methodology may vary depending on the subject area. Future research could explore refinement strategies for prompt engineering to enhance response consistency across different AI chatbot models.

3 CONCLUSIONS

The study confirmed the effectiveness of the developed technology for creating educational prompts across various subjects, including mathematics and the Ukrainian language. Statistical testing using the Kruskal-Wallis H test demonstrated that different chatbot models (ChatGPT, Copilot, Claude) exhibit varying abilities in generating responses.

For Ukrainian language prompts, significant differences were found in responses based on multiple criteria. For Criterion K1 (Response Relevance), differences were observed between Claude and Copilot (p = 0.002). Criterion K2 (Response Completeness) showed differences between Claude and Copilot (p = 0.02). For Criterion K3 (Correctness of information) differences are also observed between Claude and Copilot (p = 0.018). In the evaluation by criterion K4 (Practical applicability) there are statistical differences between Claude and ChatGPT (p = 0,002) and Claude and Copilot (p = 0,021).

Criterion K5 (Response Structure) revealed distinctions between Claude and ChatGPT (p = 0.005) and Claude and Copilot (p = 0,015).

For mathematics prompts, differences were mostly insignificant, except for Criterion K5 (Response Structure), where Claude and Copilot exhibited a statistically significant difference (p = 0,014). These findings suggest that while the prompt creation technology is effective, chatbot selection should align with educational objectives.

The results indicate that teachers can utilize this technology to streamline routine tasks and select chatbot systems based on response relevance, completeness, accuracy, practical applicability, and structure. However, further research is needed to:

1) Increase the number of experts evaluating chatbot responses.
2) Extend testing to additional educational subjects.

Our findings align with research by Yurchak et al. [10], which highlighted chatbot-specific strengths, weaknesses, and prompt optimization strategies. The study also confirms the necessity of integrating professional expertise with AI-generated outputs, as noted by Li et al. [6].

The proposed technology applies known prompt optimization techniques, supporting Bozkurt's [2] thesis on operational design as key to enhancing human-AI communication. This aligns with Cain's [4] conclusions on the role of prompts in fostering personalized and equitable learning experiences. The study confirms that structured, reusable, and adaptive prompts can reduce teacher workload and improve lesson planning efficiency, supporting the flexibility required for diverse educational contexts.

REFERENCES

[1] D. Baidoo-Anu and L. Owusu Ansah, "Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning," SSRN, 2023. [Online]. Available: https://dx.doi.org/10.2139/ssrn.4337484.

[2] A. Bozkurt, "Tell me your prompts and I will make them true: The alchemy of prompt engineering and generative AI," Open Praxis, vol. 16, no. 2, pp. 111–118, 2024. [Online]. Available: https://doi.org/10.55982/openpraxis.16.2.661

[3] A. Bozkurt and R. C. Sharma, "Generative AI and prompt engineering: The art of whispering to let the genie out of the algorithmic world," Asian Journal of Distance Education, vol. 18, no. 2, pp. i–vii, 2023. [Online]. Available: https://doi.org/10.5281/zenodo.8174941.

[4] W. Cain, "Prompting Change: Exploring Prompt Engineering in Large Language Model AI and Its Potential to Transform Education," TechTrends, vol. 68, no. 1, pp. 47–57, 2024. [Online]. Available: https://doi.org/10.1007/s11528-023-00896-0.

[5] G. Cooper, "Examining science education in ChatGPT: An exploratory study of generative artificial intelligence," Journal of Science Education and Technology, vol. 32, no. 3, pp. 444–452, 2023. [Online]. Available: https://doi.org/10.1007/s10956-023-10039-y.

[6] Y. Li, J. Shi, and Z. Zhang, "An approach for rapid source code development based on ChatGPT and prompt engineering," IEEE Access, vol. 12, 2024. [Online]. Available: https://doi.org/10.1109/ACCESS.2024.3385682.

[7] S. Schulhoff, M. Ilie, N. Balepur, K. Kahadze, A. Liu, C. Si, ... and P. Resnik, "The prompt report: A systematic survey of prompting techniques," arXiv preprint, arXiv:2406.06608, 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2406.06608.

[8] S. Skvortsova, T. Symonenko, T. Britskan, O. Onopriienko, and R. Romanyshyn, "Artificial intelligence in the professional activity of a university lecturer in Ukraine: Realities and prospects," Proceedings of the Second International Workshop on Artificial Intelligent Systems in Education, co-located with 23rd International Conference of the Italian Association for Artificial Intelligence (AIxIA 2024), vol. 3879, 2024. [Online]. Available: https://ceur-ws.org/Vol-3879/AIxEDU2024_paper_14.pdf.

[9] J. Wei, X. Wang, D. Schuurmans, M. Bosma, E. Chi, Q. Le, and D. Zhou, "Chain of thought prompting elicits reasoning in large language models," arXiv preprint, arXiv:2201.11903, 2022. [Online]. Available: https://arxiv.org/pdf/2201.11903.

[10] I. Y. Yurchak, O. O. Kychuk, V. M. Oksentyuk, and A. O. Khich, "Capabilities and limitations of large language models," Computer Systems and Networks, vol. 6, no. 2, pp. 286–300, 2024. [Online]. Available: https://doi.org/10.23939/csn2024.02.286.