УДК 81’32 DOI: 10.31249/chel/2025.02.12

Фирсанова В.И.

РАЗРАБОТКА ИНТЕРФЕЙСА ВИРТУАЛЬНОГО АССИСТЕНТА ПРЕПОДАВАТЕЛЯ НА ОСНОВЕ ТЕХНОЛОГИЙ ВЫЗОВА ФУНКЦИЙ И ИНЖЕНЕРИИ ИНСТРУКЦИЙ ДЛЯ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ©

Санкт-Петербургский государственныйы университет,
Россия, *Санкт-Петербург*, *st085687@student.spbu.ru*

Аннотация. Статья посвящена промпт-инжинирингу для снижения плагиата и повышения заинтересованности студентов. Было проведено интервью преподавателей, разработана первая коллекция пресетов для промпт-инжиниринга в академической среде, создан интерфейс виртуального ассистента. Качественный анализ результатов интервью показал, что сотрудники сферы образования скорее заинтересованы в применении больших языковых моделей при разработке форм контроля, адаптации учебного материала с целью предоставления доступности и создания иллюстративного материала. Однако в ходе исследования были выявлены такие ограничения, как порождение ложной информации языковыми моделями и высокий риск распространения плагиата среди студентов. В результате исследования был разработан пользовательский интерфейс виртуального ассистента преподавателя с такими функциями, как готовые наборы инструкций для больших языковых моделей, разработанные методами инженерии инструкций, а также функции для обеспечения кибербезопасности и обнаружения плагиата.

Ключевые слова: промпт-инжиниринг; генеративный искусственный интеллект; виртуальный ассистент; образовательные технологии; методология преподавания.

Получена: 17.07.2024 Принята к печати: 28.12.2024

Firsanova V.I.

The development of an interface for teaching virtual assistant based on function-calling and prompt engineering technologies©

Saint Petersburg State University,
Russia, *Saint Petersburg*, *st085687@student.spbu.ru*

Abstract. The paper highlights prompt engineering in academic setting to reduce plagiarism and increase students' interest. The research problem is the lack of a unified methodology for using artificial intelligence in education. The paper aims to create a generative artificial intelligence user interface, the Virtual Teaching Assistant. Teachers were interviewed, the first collection of presets for prompt engineering in an academic environment was developed, and a virtual assistant interface was created. The qualitative analysis revealed that educators are interested in applying large language models to developing assessment materials, adapting learning content for providing accessibility, and creating education illustrations. However, the study identified challenges, including hallucinations in generated content and the risk of plagiarism. The study resulted in the development of a virtual teaching assistant interface with features such as prompt presets, custom guardrails, and plagiarism detection. The prompt presets are simplified templates for generating assignments, quizzes, and educational materials. Guardrails are safety controls that ensure ethical and inclusive content generation. Plagiarism checker is a function that detects AI-generated content through analyzing students' submissions using neural network weights.

Keywords: prompt engineering; generative artificial intelligence; virtual assistant; educational technologies; teaching methodology.

Received: 17.07.2024 Accepted: 28.12.2024

Introduction

A large language model (LLM) is a machine learning model for natural language processing capable of general-purpose text language generation, or next token prediction, as well as zero-shot learning on a wide variety of tasks, such as text classification, question-answering, or natural language inference [Better language models …, 2019, p. 3]. For example, OpenAI GPT-4 [GPT-4 technical report, 2023], Microsoft Phi-3 [Phi-3 technical report …, 2024], or Google Gemma [Gemma …, 2024] are self-supervised broad-domain LLMs suitable for various downstream tasks also called foundation models [On the opportunities …, 2021, p. 4].

LLMs assist in academic tasks, such as proofreading, concept explanation, or personalized learning, increasing student educational engagement. However, there are concerns about content originality. For example, using LLMs in academia might increase plagiarism risks with legal impact [Elkhatat, 2023, p. 2]. In general and higher education, the misuse of LLMs might lead to broad cheat content generation among students [Cotton, Cotton, Shipway, 2024, p. 230].

It seems that the spread of LLM use in different settings, including academia, is associated with developing user-friendly interfaces. For example, ChatGPT, Google Gemini, and YandexGPT can be accessed through user-friendly chatbots and user interfaces. Making LLMs accessible allows for routine automation, brainstorming, and virtual assistance, however, it also increases the risks of irresponsible use of artificial intelligence technologies, caused by insufficient knowledge of how LLMs work on the user level and absence or lack of LLM output control, as well as prompt filtering.

The study aims to develop a generative artificial intelligence user interface for academic setting. The research problem is the lack of a unified methodology for artificial intelligence use in education. The study implies experimental interviews and questionnaires of educators, supported by qualitative results analysis to define the problems and needs of the members of the educational system. The paper describes a process of building a user interface layout in Figma and prototypes in Gradio and Google Colab, based on the qualitative analysis results. The study results include a demo version of a virtual assistant built with LangChain and LlaMA.CPP libraries and a novel educational methodology that describes virtual agent functions, LLM guardrails, and a first collection of presets for prompt engineering in academia.

Related work

Recently, the interest in using artificial intelligence in education has been rising. For example, novel proposals for using LLMs in language learning [Alenizi, Mohamed, Shaaban, 2023], inclusive education [Mujahid, Saha, 2023], and general or higher education enhancement [Zhang, Tur, 2024] are emerging. However, there's a lack of ready-made user interfaces enabling generative artificial intelligence use in education. Most of the studies explore the capacity of broad-domain systems, such as ChatGPT or Google Gemini, but do not propose an isolated educational tool based on LLMs.

For example, according to the results of educators' interviews conducted in [Alenizi, Mohamed, Shaaban, 2023], ChatGPT is a perspective tool for scaffolding, i.e. supporting students with special educational needs. This support can be provided by generating individualized instructions for students, creating collaborative activities, and enhancing peer-to-peer education. Other studies view ChatGPT as an accessibility tool. For instance, LLMs can provide language support by generating immediate translations and concept explanations. Generative models can be used for social-emotional learning developing interaction strategies for students with communicational difficulties. LLM text-to-text generation can provide accessibility for students with dyslexia or other learning difficulties [Mujahid, Saha, 2023, p. 92–93]. Overall, LLMs in general and higher education are often used for study preparation. For example, educators prefer using ChatGPT for workshop organization and assignment planning, while students use the model for problem-solving guidance and automated feedback [Zhang, Tur, 2024, p. 12].

It seems that most studies highlight the use of ChatGPT in inclusive education, however, they do not review the accessibility of the very ChatGPT user interface and the model outputs. There is a lack of research on ethical considerations around using LLMs in educational environments. For example, there is an insufficiency of strategies for plagiarism minimization. Another controversial aspect is prompt engineering accessibility. Considering that prompt engineering requires building instruction structures and applying domain knowledge, this technique might not be available for people with learning difficulties, meaning that novel strategies or prompt presets are needed in inclusive educational settings.

There is a lack of technical details on LLM applications in academia, such as guardrailing and special LLM user interfaces developed considering ethical issues, such as student cheating and plagiarism, as well as accessibility challenges, that will be discussed in the following section. In the Discussion section, a control in students' use of ChatGPT and W3C Accessibility Standards towards LLMs will be observed. Overall, the study declares that the modern educational system requires a special, i.e. closed-domain LLM-based tool, that is proposed in this paper.

Method

This study utilizes a qualitative approach, focusing on in-depth interviews with educators to gather insights into their experiences and

needs regarding LLMs and prompt engineering. The goal is to understand the practical applications of LLMs in academic settings and to develop a user-friendly interface for a Virtual Teaching Assistant based on the received feedback.

The study involves interviews with 18 educators from general and higher education institutions. The participants signed a formal agreement before participating in the experiment. The age of the participants varied from 25 to 50 years. 90% of the participants were females, and 10% were males. The participants represented different higher educational institutions, including Saint Petersburg State University and Higher School of Economics. The teaching experience of the participants varied from 1 to 25 years. The participants are selected from a range of general education subjects, specifically focusing on language learning, arts, and humanities. The key interview questions are about respondents' teaching experience, their experience with LLMs, and the types of tasks they would like to automate using generative artificial intelligence. The questions were as follows: (1) Characterize your level of awareness of AI technologies; (2) Describe your experience in applying LLMs for enhancing your teaching practices and skills; (3) Which tasks from your teaching routing would you like to automate with LLMs.

After the interview, the respondents were asked to take part in a prompt engineering experiment by forming three prompts relevant to their teaching disciplines and using them to generate answers from GPT-4o. The respondents were free to use English or Russian languages for prompting. The respondents were tasked to create prompts that would allow for generating various forms of control, such as quizzes, examination tasks and tests, as well as interactive tasks for seminars and workshops. At the final stage of the experiment, the respondents were asked to analyze the generated results for instances of hallucination (generic false information), evaluate the output accuracy, and decide whether the generated texts can be used in practice, and what specific changes are needed in the model result.

The insights gathered from the interviews are summarized to specify a foundation for a user interface of a novel academic LLM-based system. The key features of the interface include a prompt builder with presets, accessibility tools, and instruments for plagiarism detection. The system development comprises creating guardrails, and user interface layouts in Figma, as well as building a demo version of the tool with Gradio and Google Colab.

Interview Results

The qualitative analysis of interviews with 18 educators revealed several key themes related to their experiences and needs concerning LLMs and prompt engineering.

Firstly, many educators reported limited but growing use of LLMs in their teaching practice. For some respondents, this experiment was their first experience in prompt engineering. Several agreed to participate in this research to learn more about ChatGPT. Apart from ChatGPT, one of the participants claimed their active usage of GigaChat LLM, and several reported applying image generation models in their educational practice.

Secondly, educators identified several tasks they would like to automate using LLMs. Common applications included developing assessment tasks, adapting learning material through text-to-text generation, and creating illustrations for studied phenomena. Two participants reported they use ChatGPT to detect cheating by generating results for the assessment task and comparing the generated text with a student's answer. Table 1 shows examples of prompts created by the participants of the experiment.

Table 1

Examples of prompts created by the experiment participants

Prompt	Level of education	Displine
List 19 words of two or three syllables containing different stressed vowels according to IPA transcription	Higher education	Phonetics
Create a test in French on the topic Conditionnel of 20 tasks, including open and closed questions	Higher and general education	French as a foreign language
Create a test of 20 tasks on the topic of Indicative Tenses in French for levels B1-B2	Higher and general education	French as a foreign language
Rewrite the following text using more informal expressions and grammar	Higher and general education	Open domain (can be applied to any discipline)
Create questions for a quiz based on Tolstoy's story After the Ball	General education	Literature

Thirdly, most participants found hallucinations in the model answers reporting unacceptable fragments such as grammar mistakes in linguistic tasks. For example, in language learning tasks, some participants reported wrong verb tense usage in generating study illustrations for cer-

tain grammar phenomena. Precisely, for the prompt "Generate grammar mistakes in Russian" the model produced "Кот сидит на коврах" ("A cat sits on carpets") as an example of incorrect case usage. However, this sentence is grammatically acceptable, although it uses the plural form of the word "carpet" instead of a singular.

Hypothetically, hallucinations in linguistic phenomena generation are caused by LLM multilingualism. Such models as GPT-4o are trained on multilingual data, and the ratio of English data is many times higher than the amount of data in other languages, including Russian. That might result in mapping the linguistic structure of English to other languages, although this idea has yet to be proven through explainable artificial intelligence methods. In the following section, a novel LLM system based on the interview insights will be described. One of the solutions is to use LLMs fine-tuned specifically for the Russian language, such as Vikhr, GigaChat, or Saiga. However, the experiments show that the problem of hallucinations cannot be solved completely, since the generative mechanisms use stochastic processing in decision making, which cause probabilistic inaccuracies, such as grammar handling errors.

Virtual teaching assistant

The qualitative analysis resulted in building a user interface for the Virtual Teaching Assistant system, which uses LLMs to provide closed-domain assistance in academic setting. Figure 1 shows the demo version of the interface. The graphical interface supports English and Russian languages. One of the primary ethical challenges in developing an LLM-powered educational chatbot is minimizing the risk of plagiarism and cheating. To address this, an AI plagiarism checker function was developed. The function was developed through the function-calling mechanism, which is a novel LLM programming approach. The function concatenates user data with a pre-defined prompt "Is this AI generated?". The LLM weights allow for zero-shot AI plagiarism detection, since such models as OpenAI GPT-4 apply watermarks to the generated content. AI watermark is a set of weights tuned for immediate AI plagiarism detection. The watermark does not affect the quality of AI response, however, it makes them easily detectable through populating the model answer with words and collocations, which are not widespread in human written content. The model response is formatted as a JSON object containing on of the following model's decisions: (1) "Decision 1: This text is likely AI generated"; (2) "Decision 2: This text is likely

human written". The function is designed to compare LLM-generated outputs with students' works. This function leverages plagiarism detection algorithms to identify similarities and discrepancies, helping educators trace potential artificial intelligence cheating.

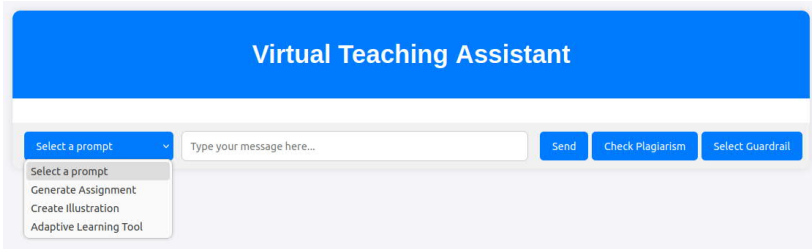

Fig. 1. Demo version of the Virtual Teaching Assistant user interface

The next stage of the study is guardrail development. Guardrails are the safety controls for LLM outputs dictating the model its behavior. The guardrails established in this study enforce discipline-specific guidelines, ethical standards, and inclusive and accessibility education principles. In general academic use, the guardrails lead the LLM to generate inclusive content adaptive for various educational needs, as well as culturally and socially diverse material.

For example, the guardrail for language learning practice in the presets developed in this study is the following: "Use clear and accessible language, particularly for audiences with varying levels of language proficiency or learning difficulties". The guardrails are concatenated with user prompts, for example: "User prompt: [Create a lesson plan for an inclusive classroom on the topic of climate change]. Guardrail: [Ensure the language is clear and accessible, the content is original, and the plan is sensitive to diverse cultural backgrounds]".

The study suggests both guardrails and prompt templates in simplified language to assist users with learning difficulties. The templates are designed for generating assignments, creating illustrations for learning materials, and adapting content for different languages or difficulty levels. For example, "Create a quiz with 10 questions, use [glossary], about [topic], in [language]" is a prompt preset developed after the interview analysis, which allows for generating diverse assignment tasks by setting the variables [glossary], [topic], and [language].

The prompt presets presented in this study are adaptable to diverse educational needs, ensuring that students with learning difficulties

can receive relevant content. This approach promotes inclusivity and enhances the accessibility of LLMs in education. The presented guardrails are customizable and provide a safe virtual educational environment. The developed user interface allows for tuning the system according to the needs of both educators and students.

Discussion

In this section, the perspectives of the study will be overviewed. In perspective, the proposed system should be reviewed according to W3C Web Content Accessibility Guidelines (WCAG) [Caldwell, 2008]. The guidelines is a set of recommendations for accessible web development. They cover the needs of individuals with various disabilities. For instance, WCAG includes providing text alternatives for images that, for example, can be reproduced by screen readers for people with visual impairments.

Universal Design [Goldsmith, 2000] is another approach that will be used for the system reviewing to ensure model accessibility. The principles of Universal Design allows creating digital and physical products, as well as environments that are both inclusive and aesthetic. For example, the Universal Design strategies in application development include customizable display settings, such as options to adjust text size, contrast, and color schemes. Another example is using simplified and diverse language that avoids jargon and complex terms, making the content accessible for students with varying levels of language proficiency and cognitive abilities. The system compatibility with assistive technologies, such as screen readers, speech recognition software, and alternative input devices is another aspect that should be resolved in the proposed system.

The next critical aspect of the proposed virtual assistant is enhancing student engagement through both control and interaction. The proposed user interface is designed to empower students by giving them control over their learning, at the same time providing a plagiarism and cheating checker that would motivate them to work independently, tailoring their studies with their interests.

In perspective, various LLM-based means of gamification and interactive elements will be explored. For example, the Virtual Teaching Assistant can be enhanced with the ability to integrate such tools as progress tracking features, quizzes generation, students discussions through virtual assistant user interface, and multimodal content generation.

By implementing WCAG and the principles of Universal Design the proposed system can become an instrument providing students control and interactive engagement in inclusive education setting.

Conclusion

The study addresses the need for a unified methodology for using large language models (LLMs) in academic setting through developing a user interface for a novel artificial intelligence system named Virtual Teaching Assistant. A qualitative research, involving interviews and prompt engineering experiments with 18 educators was conducted. The research resulted in building a set of tools comprising prompt presets, customizable guardrails for the educational environment, and a user interface layout aiming to reduce plagiarism and cheating risks, enhancing the student's engagement, and providing accessibility.

The findings revealed that educators are increasingly interested in using LLMs for tasks such as developing assessment materials, adapting learning content, and creating educational illustrations. However, challenges such as hallucinations in generated content and the risk of plagiarism were also highlighted. To mitigate the plagiarism and hallucination issues, customizable guardrails and plagiarism checker tools are proposed in the study. The checker compares the student text with generated content and evaluates the probability of artificial intelligence cheating.

The demo version of the Virtual Teaching Assistant uses Gradio and Google Colab, although advanced user interface development is planned. The study uses the principles of Universal Design and W3C Web Content Accessibility Guidelines (WCAG). The proposed approach ensures system accessibility, as well as studying material personalization to engage and empower students. Future work will focus on refining the system, incorporating more advanced features, and expanding its applicability to a broader range of educational contexts.

Bibliographic list

Alenizi M.A.K., *Mohamed A.M.*, *Shaaban T.S.* Revolutionizing EFL special education: how ChatGPT is transforming the way teachers approach language learning // Innoeduca: international journal of technology and educational innovation. – 2023. – Vol. 9, N 2. – P. 5–23.

Better language models and their implications / Radford A., Wu J., Amodei D., Clark J., Brun|dage M., Sutskever I. // OpenAI blog. – 2019. – Vol. 1, N 2. –

https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

Cotton D.R.E., *Cotton P.A.*, *Shipway J.R.* Chatting and cheating: ensuring academic integrity in the era of ChatGPT // Innovations in education and teaching international. – 2024. – Vol. 61, N 2. – P. 228–239.

Elkhatat A.M. Evaluating the authenticity of ChatGPT responses: a study on text-matching capabilities // International Journal for Educational Integrity. – 2023. – Vol. 19, № 1. DOI: https://doi.org/10.1007/s40979-023-00137-0. – URL: https://edintegrity.biomedcentral.com/articles/10.1007/s40979-023-00137-0#citeas

Gemma: open models based on gemini research and technology / Team G., Mesnard T., Hardin C., Dadashi R., Bhupatiraju S., Pathak S., Kenealy K. – 2024. – URL: https://storage.googleapis.com/deepmind-media/gemma/gemma-report.pdf?ysclid=m4mnvjuhxk346793320

Goldsmith S. Universal design. – London : Routledge, 2000. – 116 p.

GPT-4 technical report / Achiam J., Adler S., Agarwal S., Ahmad L., Akkaya I., Aleman F.L., McGrew B. – 2023. – URL: https://cdn.openai.com/papers/gpt-4.pdf

Mujahid K., *Saha R.* ChatGPT in inclusive education: a study of future potential // Emerging trends in Indian education and culture. – 2024. – Vol. 88. – P. 88–95.

On the opportunities and risks of foundation models / Bommasani R., Hudson D.A., Adeli E., Altman R., Arora S., von Arx S., Liang P. – 2021. – URL: https://arxiv.org/pdf/2108.07258

Phi-3 technical report: a highly capable language model locally on your phone / Abdin M., Aneja J., Awadalla H., Awadallah A., Awan A.A., Bach N., Zhou X. – 2024. – URL: https://arxiv.org/pdf/2404.14219

Web content accessibility guidelines (WCAG) 2.0 / Caldwell B., Cooper M., Reid L.G., Vanderheiden G., Chisholm W., Slatin J., White J. // WWW Consortium (W3C). – 2008. – Vol. 290, N 1–34. – P. 5–12.

Zhang P., *Tur G.* A systematic review of ChatGPT use in K-12 education // European Journal of Education. – 2024. – Vol. 59, N 2. – P. e12599.

References

Alenzini, M.A.K., Mohamed, A.M., Shaaban, T.S. (2023). Revolutionizing EFL special education: how ChatGPT is transforming the way teachers approach language learning. *Innoeduca. International Journal of Technology and Educational Innovation*, *9*(*2*), 5–23.

Radford, A., Wu, J., Amodei, D., Clark, J., Brundage, M., Sutskever, I. (2019). Better language models and their implications. *OpenAI blog*, *1*(*2*). Retrieved from: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

Cotton, D.R.E., Cotton, P.A., Shipway, J.R. (2024). Chatting and cheating: ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, *61*(*2*), 228–239.

Elkhatat, A.M. (2023). Evaluating the authenticity of ChatGPT responses: a study on text-matching capabilities. *International Journal for Educational Integrity*, *19*(*1*), 1–23. DOI: https://doi.org/10.1007/s40979-023-00137-0. Retrieved from: https://edintegrity.biomedcentral.com/articles/10.1007/s40979-023-00137-0#citeas

Team, G., Mesnard, T., Hardin, C., Dadashi, R., Bhupatiraju, S., Pathak, S., Kenealy, K. (2024). *Gemma*: *open models based on Gemini Research and Technology*. Retrieved from: https://storage.googleapis.com/deepmind-media/gemma/gemma-report.pdf?ysclid=m4mnvjuhxk346793320

Goldsmith, S. (2000). *Universal design*. London: Routledge.

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., McGrew, B. (2023). *GPT-4 Technical report*. https://cdn.openai.com/papers/gpt-4.pdf

Mujahid, K., Saha, R. (2024). ChatGPT in inclusive education: a study of future potential. *Emerging trends in Indian education and culture*, *88*, 88–95.

Bommasani, R., Hudson, D.A., Adeli, E., Altman, R., Arora, S., von Arx, S., Liang, P. (2021). *On the opportunities and risks of foundation models*. Retrieved from: https://arxiv.org/pdf/2108.07258

Abdin, M., Aneja, J., Awadalla, H., Awadallah, A., Awan, A.A., Bach, N., Zhou, X. (2024). *Phi-3 Technical report*: *a highly capable language model locally on your phone*. Retrieved from: https://arxiv.org/pdf/2404.14219

Caldwell, B., Cooper, M., Reid, L.G., Vanderheiden, G., Chisholm, W., Slatin, J., White, J. (2008). Web content accessibility guidelines (WCAG) 2.0. *WWW Consortium (W3C)*, *290*(*1–34*), 5–12.

Zhang, P., Tur, G. (2023). A systematic review of ChatGPT use in K-12 education. *European Journal of Education*, *59*(*2*), 1–22.

Об авторе

Фирсанова Виктория Игоревна – аспирант, Санкт-Петербургский государственный университет, Россия, Санкт-Петербург, st085687@student.spbu.ru

About the author

Firsanova Viktoriia Igorevna – PhD student, Saint Petersburg State University, Russia, Saint-Petersburg, st085687@student.spbu.ru