Advantages and limitations of large language models in chemistry education: A comparative analysis of ChatGPT, Gemini and Copilot

Yuliia V. Kharchenko, Olena M. Babenko

Sumy State Pedagogical University named after A. S. Makarenko, 87 Romenska Str., Sumy, 40002, Ukraine

Abstract

This study aims to explore the potential and limitations of large language models (LLMs) such as ChatGPT, Gemini, and Copilot, in the context of chemistry education. The primary objective of the study is to compare the effectiveness of LLMs in solving chemistry tasks and to identify the key challenges associated with their implementation in education. These LLMs were selected based on a survey of students which indicated their widespread use due to their free accessibility. To evaluate the potential of LLMs in chemistry education, we employed them to solve tasks corresponding to different levels of knowledge in different subfields of chemistry. A comparative evaluation of LLMs' performance against that of average Ukrainian students was conducted. The results indicate that while LLMs show promise mainly in tasks not demanding deep logical reasoning, they are generally inferior to students. Key challenges in using LLMs in chemistry education identified include understanding the nuances of chemistry as a complex and multifaceted science, abstract concepts used in chemistry, recognition of chemical compound formulas, chemical reaction equations, limitations in logical reasoning, language barriers, and the occurrence of AI hallucinations. Additionally, there is a need for students to develop skills in crafting effective queries and prompts to enhance the efficiency of working with LLM. While LLMs are promising, their implementation requires addressing the identified limitations.

Keywords

artificial intelligence, LLM, ChatGPT, Gemini, Copilot, chemistry education

1. Introduction

A characteristic feature of modern times is the continuous dynamics of change in all spheres of human life: economy, politics, science, and education. The dominant trend in the development of contemporary civilization is its transition into an information society, wherein information and communication technologies become the objects of human activity, providing all the necessary conditions for the formation and development of the personality of the new formation. The rapid development of the global Internet network has led to a computer revolution in the information world, where the computer serves as the primary means of telecommunication. Considering that the current stage of development of pedagogical science in the world is characterized by an intensive search for new ways to improve the quality of education, information and communication technologies have become powerful tools in this process [1]. Various forms of information and communication technologies (ICTs) have found active application in education, ranging from electronic textbooks [2], online learning technologies [3, 4, 5], mobile applications [6, 7] to augmented and virtual reality technologies [8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23]. The latest term in the digital revolution is artificial intelligence (AI) [24, 25], including generative AI. Generative artificial intelligence (GenAI) is a technology that automatically generates content in response to queries. GenAI actually creates new content using existing content. Its output may encompass formats including all symbolic representations of human though: texts, images

3L-Person 2024: IX International Workshop on Professional Retraining and Life-Long Learning using ICT: Person-oriented Approach, co-located with the 19th International Conference on ICT in Education, Research, and Industrial Applications (ICTERI 2024) September 23, 2024, Lviv, Ukraine
yuvlakhar@gmail.com (Y. V. Kharchenko); olena.babenko@sspu.edu.ua (O. M. Babenko)
https://scholar.google.com.ua/citations?user=zYiU4iMAAAAJ (Y. V. Kharchenko);
https://scholar.google.com.ua/citations?user=AeYIdfAAAAAJ (O. M. Babenko)
0000-0002-8960-2440 (Y. V. Kharchenko); 0000-0002-1416-2700 (O. M. Babenko)

(including drawings, photographs, videos and animations), music and software code. GenAI learns from data it collects from web pages, conversations in social networks, and other online media. It generates its content by statistically analyzing the order of words, pixels, or other elements in the data it has learned, and identifying and replicating common patterns (such as which words typically follow other words, and in what order).

Text-based generative artificial intelligence, utilizing a type of artificial neural network known as a general-purpose transformer, is particularly popular. This type of AI, often referred to as Large Language Models (LLMs), is commonly known as a generative pre-trained transformer, or GPT.

GPTs and their ability to automatically generate text became available to the global research community in 2018. The launch of ChatGPT in 2022, which offered free access and a user-friendly interface, became a sensation and led to active searches and technological solutions for other companies to create and launch new similar systems. By mid-2023, other alternatives to ChatGPT [26] became available, most of which were free (within certain limits), as well as services for generating images, videos, and music, some of which are listed in table 1.

Table 1
AI-based services.

Text GenAI models	Image GenAI models	Video GenAI models	Music GenAI models
Alpaca	Craiyon	Ela	Aiva
Bard (now Gemini)	DALL-E	GliaCloud	Boomy
Chatsonic	DreamStudio	Pictory	Soundraw
Ernie	Fotor	Runway	Voicemod
Hugging Chat	Midjourney		
Jasper	NightCafe		
Llama	Photosonic		
Open Assistant			
Tongyi Qianwen			
YouChat			

Many other tools based on the aforementioned LLMs are also emerging, such as ChatPDF, which can work with PDF documents and analyze them, or Perplexity, which serves as a knowledge hub and helps users find answers to queries based on their needs. Similarly, the process of integrating LLMs into other products, such as web browsers, is ongoing. In Ukraine, users gained access to artificial intelligence in 2023, and now the range of services based on generative artificial intelligence has significantly expanded.

2. Literature review

The increasingly deeper and active penetration and integration of AI into human activity could not fail to impact one of the most important components of human development, namely education. Worldwide, the initial concerns about the use of AI in education were linked to fears that students would use its capabilities primarily to cheat on academic tasks, thereby undermining the value of educational assessment, certification, and qualifications [27]. A clear consequence of such concerns was the prohibition by some educational institutions on students' use of AI [28]. However, in other institutions, a more optimistic approach was taken towards the use of AI [29], believing that it is more progressive not to prohibit its use, but to provide support for both teachers and students in utilizing tools based on generative AI [30]. Today, it should be noted that artificial intelligence is increasingly being integrated into education with the aim of enhancing student learning efficiency and improving teaching practices.

As indicated by the analysis of recent research, significant regarding the use of AI in the field of education, particularly in higher education, in educational and research activities, are occurring in the following key areas [31, 32]:

- assessment (including automatic assessment and evaluation of educational progress and students' attitude to learning, individual and group assessment, etc.);
- predicting learning status (predicting student withdrawals, at-risk groups, innovative abilities, career decisions), productivity or satisfaction, improving the learning experience;
- assistance (providing support to students in their educational pursuits, for example, anthropomorphic presence, which includes virtual agents and intervention through digital programs);
- tutoring (providing and supporting individual strategies and approaches to students, taking into account their characteristics and needs);
- learning management (learning analytics, sequence of educational plans and programs, development of instructions, and student allocation).

Educational tools based on artificial intelligence can offer personalized learning experiences, automate routine tasks, and provide real-time feedback and assessment. As numerous studies by scientists around the world have shown, AI has the potential to be a useful tool in teaching and learning Chemistry, particularly for creating interactive simulations, answering questions, and providing feedback on student work [33]. It can be used to create personalized learning experiences for students [34, 35].

The results of studying the attitudes of natural science teachers toward the use of artificial intelligence in teaching show that teachers are generally positive about integrating AI into the educational process. Key factors influencing their willingness to use AI include self-confidence, expected benefits, ease of use and general attitude towards AI technologies. Al Darayseh [36] found in his study that the easier teachers can integrate AI into teaching Natural Sciences, the more they believe in its benefits and are more willing to use it.

AI-based tools can be useful for teaching Chemistry by offering interactive simulations, answering questions, and providing feedback on student work. There are a number of studies regarding the use of AI in chemistry education, but it should be noted that the overwhelming majority of them focus on ChatGPT. For instance, dos Santos [37] explored the potential of generative chatbots based on artificial intelligence, including GPT-4 and BingChat, in chemical education. The study demonstrated that ChatGPT and BingChat act as "thinking agents" fostering critical thinking, problem solving, concept understanding, creativity, and personalized learning.

Another group of researchers explored the potential of using artificial intelligence to enhance Chemistry teaching in high schools in Vietnam, using ChatGPT as an example. They identified the potential of the ChatGPT text generation model from OpenAI. Their study showed that ChatGPT performed well on intermediate-level Chemistry exams, but struggled with questions that required a high level of knowledge application (e.g., analyzing and solving complex problems). Overall, ChatGPT's responses showed lower performance than most Vietnamese students in responses, indicating limitations in the application of this tool. At the same time, a number of advantages of using ChatGPT were highlighted, including increasing student engagement through interactive learning; providing immediate answers and explanations to student questions; personalizing learning by adapting responses to individual student needs; fostering critical thinking through open-ended questions and alternative points of view; providing access to additional learning materials such as links and examples; and facilitating self-directed learning [38].

In a similar study, Williams and Fadda [39] analyzed the responses of ChatGPT and ChatGPT Plus to questions related to carbohydrate chemistry, a topic frequently included in many chemistry curricula. The authors demonstrated that ChatGPT Plus performed significantly worse on test questions. Overall, both language models performed better on simple, common questions for which ample information is available.

Xuan-Quy et al. [40, 41, 42], Xuan-Quy and Le [43, 44] obtained interesting results in a series of their studies, identifying the potential of ChatGPT for different science subjects at the high school graduate level. It was found that the AI performed better on social science questions, such as Literature, History, Geography, and Civic Education, and showed slightly worse results in answering questions on natural science: Mathematics, Physics, Chemistry, and Biology.

Humphry and Fuller [45] discuss the potential use of ChatGPT in chemistry laboratories at the bachelor's education level. The authors propose using ChatGPT as a teaching tool for students to demonstrate their understanding of specific topics by detecting and correcting chatbot's errors. We agree with the authors' observation that ChatGPT is prone to conceptual errors in many of its chemistry-related responses and explanations and should not be used for general chemistry education. Overall, it is emphasized that ChatGPT can be useful for explaining basic concepts, but requires further development to effectively teach more complex topics.

Williams and Fadda [39] investigated the potential of integrating ChatGPT into a "flipped class" model for teaching chemistry. They demonstrated that this approach allowed students to take a more active role in their learning, while the teacher assumed a supervising or guiding role. Additionally, the authors tested ChatGPT's ability in writing annotations, abstracts, and essays on chemistry topics – tasks that are common and significant in the scientific community. However, the results indicated that the bot often produced vague and repetitive text, frequently including inaccurate fabricated information and invented references.

Given the importance of hands-on experience in chemistry education, researchers have turned their attention to integrating large language models (LLMs) into chemistry experimentation and laboratory training, which opens up new possibilities for enhancing the efficiency and safety of the learning process [46]. LLMs can be applied throughout the entire laboratory workflow, from preparation to result analysis, enabling a more holistic and personalized learning experience [47].

During the preparatory phase of laboratory work, LLMs can provide students with personalized assistance by clarifying procedures described in lab manuals and explaining the specifics of equipment setup. This allows students to better understand the objectives of the experiment and potential risks without the need to process lengthy manuals. Furthermore, LLMs can reinforce safety measures, ensuring students are fully aware of the importance of adhering to laboratory safety guidelines [48, 49].

During experiments, the integration of LLMs with augmented reality (AR) technologies opens up new possibilities for personalized guidance. These systems can collect real-time data on students' actions and provide instant feedback, suggestions, and warnings. Such interaction not only enhances the educational process but also significantly improves safety level in the laboratory by identifying and correcting unsafe practices in real time [50, 51].

During the analysis of experimental results, LLMs can serve as a powerful tool for bridging practical observations with theoretical knowledge. They can assist students in analyzing collected data, facilitate discussions about experimental outcomes, and link obtained data to relevant findings in the scientific literature. This enables students to better comprehend their research results and their significance within a broader scientific context [45].

Oh and Kang [52] proposed integrating AI into a laboratory experiment involving a carbon dioxide fountain by using AI-based technology to regulate the laboratory setup. This approach not only enhanced students' understanding of the underlying chemical process and laboratory experiment but also provided them with insights into the potential of AI in scientific research. Healy and Blade [53] developed a project based on a state-of-the-art AI model for studying organic molecules of interest due to their pharmaceutical significance and for further research, as well as for planning their synthesis. The authors also demonstrated the potential of this project for online education and enhancing student engagement in the learning process. In another study, Joss and Müller [54] assigned a task to students requiring them to create a correlation for predicting the boiling points of organic compounds. The processing of data for over 6,000 substances was conducted by the students using an artificial neural network. It should be noted that working with AI tools requires a distinct set of knowledge and skills compared to traditional search engines. One of the most crucial factors for effectively working with AI is crafting the correct prompt, i.e., formulating an appropriate input query [53, 55]. Researchers are currently paying significant attention to this issue. For instance, Tassoti [56] studied how chemistry students work with AI and demonstrated that most students lack skills in effective prompting and tend to simply copy and paste questions, thereby limiting the effectiveness of their interactions with AI.

Given the diversity of prompting strategies, ranging from basic approaches like "input-output" prompting [57] to more advanced methods such as "Chain-of-Thought" [58] and "Tree-of-Thought" prompting [59], and considering that students outside of digital technology fields are generally unaware of these approaches, there is a need to train students in creating effective queries and prompts to enhance their interactions with AI. Although the issue of prompt formulation was not the focus of our study, as it is more relevant for open-ended queries, it remains a pressing issue in the field.

In general, an analysis of existing scientific reports on the application of AI in education underscores the paramount importance and relevance of this topic. And the fact that the capabilities of AI have only begun to be utilized by both teachers and students attests to its pressing nature, indicating that it is currently at the forefront of attention. To properly establish the trajectory of the teacher-student-AI interaction, it is necessary to explore the functionality of AI that can be used in education for optimization purposes, as well as the limitations of AI, identifying its weaknesses and possible ways to correct them.

Analyzing the capabilities of AI will also help determine which specific needs of users (teachers and students) can be addressed using these technologies, as well as which features can best meet their needs.

Key challenges in using AI for chemical education can be identified. Firstly, this pertains to the complexity of chemical science, which encompasses numerous nuances [60] and makes it difficult to capture in AI algorithms. Secondly, accurate interpretation of chemical terms and concepts is crucial, as they play a significant role in comprehending scientific material. Many chemical concepts are abstract and challenging to comprehend, with multiple meanings, leading to potential misinterpretation by artificial intelligence. Thirdly, chemical formulas, structures, and equations have a clear structure that AI may not always recognize accurately, affecting its ability to generate correct answers or solve problems. Therefore, to obtain correct answers to chemistry-related questions and issues, it is essential to consider the context of the question, understand the meanings of chemical terms and concepts, and apply logical thinking and analysis.

When considering the use of generative AI in the context of Ukrainian education, another important aspect must be noted. It is crucial to remember that tools based on generative artificial intelligence perform much better with the English language than with any other [61]. This is because AI perceives queries as tokens, which are fundamental building blocks upon which the model learns, understands, and processes language. This is where the most significant and crucial language differences arise.

Firstly, the English language has a smaller vocabulary size compared to the Ukrainian language. This means that language processing algorithms can operate with fewer tokens (words), which simplifies their operation and improves speed and accuracy. Secondly, the English language has fewer word forms and word variations compared to the Ukrainian language. For example, many English words have only one form, while Ukrainian words can have multiple variations depending on context, gender, number, etc. This makes token analysis and recognition in Ukrainian more challenging. And thirdly, since English is used as a language of communication in many fields, including science, technology and business, there are far more sources of information available in English than in Ukrainian. Consequently, due to the smaller vocabulary and complexity of word-formation forms of the Ukrainian language, artificial intelligence algorithms can work more effectively with the English language in terms of token processing, and ChatGPT, like most other AI models, was trained on tokens tailored for the English language.

Our experience shows that Ukrainian students have recently started actively using AI in almost all types of their academic activities, believing that AI can provide the correct answer to any question. Often, they use AI as their primary source of information, placing complete trust in its outputs. As a result, students often fail to critically analyze the information and responses generated by AI, neglecting to verify them against reliable sources such as books, scientific journals, etc. Therefore, the aim of our research was to study the potential of AI to act as a tutor and assist students in the learning process, helping them find answers to questions of varying complexity, and to explore the potential of using some large language models in teaching Chemistry in the Ukrainian-language educational environment.

3. Research methods

To begin with, we conducted research to identify the most popular AI-based services among pre-service teachers at Sumy State Pedagogical University named after A. S. Makarenko, whose education programs include the study of chemical disciplines. We also analyzed the services used by teachers in preparing for chemistry classes. For this purpose, we developed a questionnaire using Google Forms and employed interview methodology.

The questionnaire included the following questions:

- Do you have experience using artificial intelligence?
- Rate your level of familiarity with the types and principles of artificial intelligence and neural networks on a scale from 1 to 5 (1 – low, 5 – high).
- What tools have you used?
- If you have used artificial intelligence, what was the purpose?
- Have you used AI-based services to solve tasks or find answers to questions in chemical disciplines?
- Have you encountered cases where AI-based services provided incorrect answers or solved tasks incorrectly?
- Have you experienced instances of AI hallucinations (when AI generated something that cannot be true)? Please specify the service where you noticed hallucinations.
- When did you notice AI hallucinations more frequently?

Additionally, the method of interviewing students about their expectations from AI was used.

The next step was to assess the potential of LLMs and their applicability in addressing chemistry-related questions and tasks. To achieve this, we identified specific chemical disciplines and types of questions that would help us analyze the limitations of LLM usage. Additionally, we conducted a comparative analysis of the capabilities of selected LLMs in solving different tasks related to various chemical disciplines studied by students at Sumy State Pedagogical University named after A. S. Makarenko, who are future chemistry teachers. We then compared the results of the LLMs to the average scores achieved by the students during testing. In total, the test results of 36 students were analyzed.

A review of studies on the topic has shown that large language models have limitations in understanding Chemistry as a subject despite their seemingly well-reasoned generated responses, as they lack the ability to think and reason logically or demonstrate understanding [62]. Fergus et al. [63] showed that AI is quite good at answering questions related simply to knowledge demonstration, but has limitations in processing questions that require interpretation of non-textual information, such as analysis of structural formulas. Taking these facts into account, we selected multiple-choice questions of varying complexity. The lower-order questions required only the correct description of specific concept or phenomenon. The tests also included a series of higher-order questions that required the ability to analyze the provided information, consider the context, and think logically.

Disciplines such as Structure of Matter, Organic Chemistry, Environmental Chemistry and Laboratory Chemical Practice (LCP) are among the compulsory courses for students studying Chemistry. It should be noted that our many years of experience show that the first three of these disciplines are quite challenging for students to study, since they encompass all modern concepts not only in the field of Chemistry, but also Physics and Mathematics (Structure of Matter), Biology and Ecology (Environmental Chemistry). Organic Chemistry, in particular, is actually based on a huge amount of data on the structures of substances, their properties and relationships, operating with substance formulas. Laboratory Chemical Practice involves the formation of students' knowledge about laboratory glassware, equipment, reagents, and basic operations with them.

4. Results

The results of the first stage of our study showed that, during the educational process, our students use LLMs, and not specific neural networks intended for the study of chemistry, which is understandable given their pedagogical specialization. As the survey revealed, the most popular tools among both students and teachers are the freely accessible ChatGPT 3.5, Gemini (Google Bard), and Copilot (figure 1).

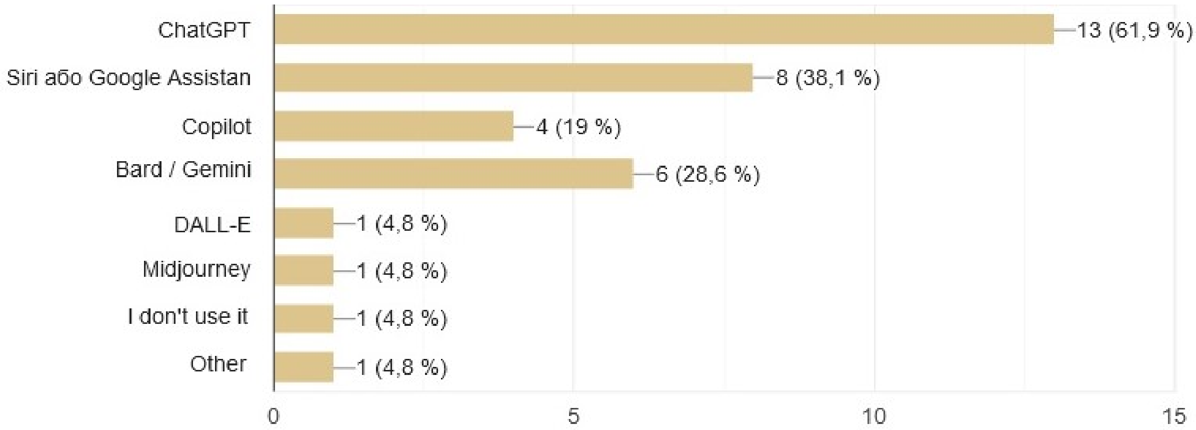

Figure 1: Percentage of answers to question “What tools have you used?”.

Therefore, we proceeded with further research using these specific large language models. These services were given identical questions related to the selected disciplines: Structure of Matter, Organic Chemistry, Environmental Chemistry and Laboratory Chemical Practice. Students provided answers to these same questions while studying these academic disciplines during the 2023-2024 academic year using the Moodle distance learning platform, which is used at our university. Examples of some test questions and the answers given by the LLMs are shown in table 2.

The test questions in Organic Chemistry were formulated to assess knowledge of nomenclature, the structure of various classes of organic compounds, methods of their synthesis and their chemical reactivity. We selected the topic “Alkenes and Alkadienes”. Let’s analyze the large language models responses to some of these questions. We chose several lower-order questions simply requiring demonstrating knowledge of certain facts or laws. Correct answers to such questions are well-known facts, presented in almost all textbooks on Organic Chemistry, even covered in high school Chemistry courses, and do not require additional logical reasoning. However, all LLMs provided incorrect answers to some of these questions. For example, in the question: “In the industry, buta-1,3-diene is obtained by the Lebedev method (dehydration). What substance is used for this as raw material?” all three AIs made a mistake. Moreover, both ChatGPT and Copilot provided their answers with explanatory comments. And for this question, the argumentation was based on the assertion that the Lebedev process involves the dehydration of acetylene. However, acetylene is an unsaturated hydrocarbon that does not contain a hydroxyl group and therefore cannot be dehydrated. Other questions required not just knowledge of a fact or law, but the ability to analyze and apply them in non-standard situations. Gemini failed to cope with the task because all its answers were incorrect and far from the truth. Particular attention is drawn to its response to the question about the reaction of propene with hydrochloric acid, with the formula HCI. Here, it was necessary to demonstrate knowledge of the reactivity of alkenes and take into account Markovnikov’s rule. The response provided by the LLM indicates that it considered this reaction as addition, which is correct. However, the product it indicated contains 2 chlorine atoms, which is nonsense, since the formula of hydrochloric acid contains only one chlorine atom. Therefore, it is evident that in this case, the AI hallucinated, as it replaced the reagent and performed the addition of a chlorine molecule instead of hydrochloric acid. ChatGPT provided an incorrect answer, while Copilot gave the correct answer with proper reasoning. Additionally, for some questions, ChatGPT offered two answer options, one of which was correct, but the justification provided was irrelevant to the correct answer.

Gemini performed slightly better than ChatGPT and Copilot on test questions in the discipline Structure of Matter. All three large language models mostly provided correct answers to lower-order questions. However, for questions requiring data analysis, comparison, and logical inference, they either

Table 2
Examples of questions and LLMs responses in chemical disciplines.

№	Questions	ChatGPT Responses	Gemini Responses	Copilot Responses
1	Indicate the final product of the reaction: propene + HCl a. 2-chloropropane (correct answer) b. 1,3-dichloropropane c. 2,2-dichloropropane d. 1,2-dichloropropane e. 1-chloropropane	In this reaction, potassium hydroxide (KOH) reacts with the halide (2,3-dibromobutane) under well-known conditions, such as nucleophilic substitution. Thus, the product of the reaction will be but-2-ene (a)	d. 1,2-dichloropropane	When propene reacts with HCl, a hydrogen atom attaches to a carbon atom that has more hydrogen atoms (Markovnikov's rule), resulting in the formation of product a (2-chloropropane)
2	Indicate in which aggregate state the potential energy of particles significantly exceeds the kinetic energy, Ekine.<Epoten. a. in the solid b. in the liquid (correct answer) c. in the gaseous d. in the plasma	Correct answer: b. in the liquid	a. in the solid	Correct answer: b. in the liquid
3	In river waters, the most abundant ions are a. Calcium and sulfates b. Sodium and sulfates c. Sodium and chlorides d. Sodium and carbonates e. Calcium and carbonates (correct answer)	a. Calcium and sulfates	c. Sodium and chlorides	d. Calcium and carbonates. This is due to the dissolution of carbonate rocks such as limestone, which is a common process in natural bodies of water
4	Crucible tongs should be placed on the table so that their curved ends are directed a. downwards b. upwards (correct answer) c. they are not placed on the table	b. upwards	a. downwards. Crucible tongs are placed on the table so that their curved ends are directed downwards to prevent accidental contact with the hot surface.	b. upwards The crucible tongs should be placed on the table with the zigzag ends facing upwards to prevent hot parts from coming into contact with the table surface and causing possible damage to the table or tongs

provided incomplete or incorrect responses. We aimed to investigate the factual basis utilized by the AI systems when answering more complex questions and why they formulated incorrect solutions. For example, in questions where it is necessary to select substances with a specific type of crystal lattice, and to provide the correct answer, it is necessary to analyze the structure of the substances offered in answers and correlate the structure with lattice types. We received correct answers to some questions, but not to others. For instance, in the question "Choose substances with an atomic crystal lattice: a) H_2O, CO_2, b) $NaNO_3$, Fe, c) KCl, NH_3, d) SiO_2, C," all three AIs chose answer b) $NaNO_3$, Fe, which is incorrect. The argument for the response to this question was as follows:

"Therefore, only $NaNO_3$ and Fe have an atomic crystal lattice. Here are some additional explanations: An atomic crystal lattice is characterized by a high melting point, hardness, brittleness, and lack of electrical conductivity in the solid state."

In response to our objection that "$NaNO_3$ is an ionic compound, can it have an atomic lattice?" we received similar responses from LLMs:

"You are correct, $NaNO_3$ is an ionic compound, not atomic. Ionic compounds have a crystal lattice consisting of cations and anions arranged in a specific order. This order is significantly different from the atomic crystal lattice, where atoms are bonded by covalent bonds. I made a mistake when I mentioned $NaNO_3$ as an example of a substance with an atomic crystal lattice. The correct answer to the question: Choose substances with an atomic crystal lattice. a) H_2O, CO_2 b) $NaNO_3$, Fe c) KCl, NH_3 d) SiO_2, C Answer: d) SiO_2, C SiO_2 has an atomic crystal lattice, where Si atoms are bonded to O atoms by covalent bonds. C also has an atomic crystal lattice. I apologize for the confusion."

In order to check whether the LLMs learning process is occurring, we asked the same question again after a few days using the same account. This time, ChatGPT provided the correct answer. However, when we asked the same question from a different account, ChatGPT again gave the same incorrect answer. When the question was asked to Gemini from the same account as before, it once again provided an incorrect answer. To ensure whether the AI "understands" questions in Ukrainian, we asked the same question in English but still received the same incorrect answer. This indicates that the essence of the question was understood, but the AI's analysis of the question was incorrect, resulting in an incorrect response.

The discipline Laboratory Chemical Practice plays a significant role in the training of chemistry specialists, providing them with the necessary theoretical knowledge, practical skills, and abilities for successful professional activity. We asked LLMs to provide answers to test questions for the Laboratory Chemical Practice exam.

ChatGPT showed a very low result, much lower than the average result of the students. Incorrect answers were related to understanding the properties of materials used in laboratory glassware, rules and principles of working with chemical reagents, principles and methods of substance purification operations, and the purpose and use of analytical balances, as well as the properties and structure of flames. Correct answers were provided for questions concerning the purpose and use of laboratory glassware and equipment for specific operations, which represent commonly known information and do not require cognitive processing.

Gemini and Copilot also showed lower results compared to the average result of students on the same test questions, although they had more correct answers than ChatGPT. Gemini performed better on questions about purification operations of substances, properties, and the structure of flames. At the same time, it made mistakes on questions where ChatGPT provided correct answers. Copilot, on the other hand, performed somewhat better overall.

It is noteworthy that Gemini and Copilot, while providing explanations to the questions, often give correct explanations but choose the wrong option in the test. For example, in the question: "Analyze the statement about the use of beakers: 1) for measuring the volume of precipitates formed during the settling of suspensions; 2) for determining the volumes of two immiscible liquid phases," both AIs indicate that beakers are used for measuring liquid volumes. However, they choose an answer that suggests beakers cannot measure the volume of two immiscible liquids or a liquid and a precipitate. A similar situation occurs with Gemini on the question: "Crucible tongs should be placed on the table so that their curved ends are directed a) downwards, b) upwards, c) they should not be placed on the table." In the explanation, Gemini correctly notes that the curved ends of the crucible tongs should not touch the table surface, yet selects an entirely different answer that contradicts this explanation. This indicates that while the AI can find information on separate aspects of a problem and even explain them, it fails to construct elementary (from a human point of view) logical chains and draw correct conclusions. Based on the given example: if beakers can measure the volumes of liquids, and immiscible liquids are still liquids, then beakers can measure the volumes of immiscible liquids. And so on.

Such inconsistency also exists in the choice of answers by ChatGPT. When answering question "What is the name of the method for purifying substances, which involves obtaining a saturated solution of the substance at high temperature followed by cooling?", the LLM selects the wrong answer, implying that extraction is a method of purifying substances involving obtaining a saturated solution at high temperature followed by cooling. After the chosen answer, we asked ChatGPT to justify its response. The AI justified its choice as follows:

"Extraction addresses the question of the method of purifying substances because extraction involves extracting a substance from a solution using another solvent."

Perhaps the presence of the word "solution" in the definition of both operations "confused" the AI, preventing it from making the correct answer choice.

The opposite problem is also observed. Copilot chooses the correct answer in the test but explains it incorrectly. For the non-reproductive question requiring analysis of the given situation, "Analyze the statements: 1) burettes without stopcocks are NOT used when working with iodine solutions; 2) burettes without stopcocks are NOT used when working with alkaline solutions," Copilot chose the correct answer: only the first statement is correct. However, it explained it as follows: iodine can evaporate and cause corrosion of the metal stopcock of the burette, while alkaline solutions do not have this problem, so burettes without stopcocks can be used with them. The following dialogue then took place: Me: "Can a burette have a metal stopcock? It would quickly corrode when interacting with various substances. A burette should be made of chemically inert materials in order to last a long time." Copilot: "You are absolutely right, and I apologize for the misunderstanding. A metal stopcock on a burette is really not a good idea. It is important that the burette is made of chemically inert materials, such as glass or special plastics, to avoid corrosion and interaction with reagents."

After this, the AI was asked to answer the same question again. This time, Copilot responded that neither solution could be used as a titrant in burettes without stopcocks, which is actually an incorrect statement. In its explanation, the AI stated that "alkaline solutions ... can cause corrosion of the metal parts of the burette," thus not abandoning its initial misconception even after clarification of the question.

The discipline Environmental Chemistry encompasses the study of physicochemical processes in the atmosphere, hydrosphere, lithosphere, and biosphere, familiarizing students with the chemical composition and patterns of chemical element migration in soils, water, and the atmosphere, as well as forming an understanding of the geochemical role of living matter and the influence of chemical substances.

As our study showed, when generating answers to test questions for the course "Environmental Chemistry," all three LLMs performed better compared to the average student results in this discipline. However, there were still errors in the generated answers, with the LLMs making mistakes on different questions. ChatGPT, which had the lowest percentage of correct answers, chose incorrect answers for questions regarding the prevalence of chemical elements in the Universe, the chemical composition of natural waters and the Earth, and the mechanism of the greenhouse effect. Gemini also gave incorrect answers about the prevalence of chemical elements in the Universe and the chemical composition of natural waters and the Earth. Unlike ChatGPT, Gemini correctly answered the question about the mechanism of the greenhouse effect but made a mistake on the question about the anomalous properties of water. Copilot provided the highest number of correct answers for both lower-level questions and more complex ones.

It is noteworthy that all three LLMs we analyzed gave incorrect answers to the test question "Among the chemical elements that make up the Earth, the most prevalent is...". For example, ChatGPT responded, "Silicon is one of the most abundant elements in the Earth's crust," demonstrating a misunderstanding of the question's essence. Gemini's answer, "Oxygen is the most abundant element in the Earth's crust," showed that the AI failed to distinguish between the terms "Earth" and "Earth's crust" due to the peculiarities of translation from English. This is probably why it was unable to select the correct answer "Iron". Copilot performed even worse, providing the incorrect answer "Oxygen", with the erroneous explanation that "Oxygen is the most abundant element forming our planet".

We were also interested in the answers from the three LLMs to the question that required completing the phrase, "In river waters, the most abundant ions are...". Only Copilot provided the correct answer and explanation: "...Calcium ions and carbonate ions. This is due to the dissolution of carbonate rocks, such as limestone." ChatGPT failed to choose the correct answer, while Gemini hallucinated and selected an answer and explanation for a completely different question: "...Sodium ions and chloride ions – these are the two most common ions in seawater".

Although generative artificial intelligence tools demonstrate significant success in processing and generating information, their accuracy in complex fields such as Chemistry remains limited. However, if, before posing questions, AI is first prompted to provide individual definitions and scientific facts related to the question content and then asked to provide the correct answer, the percentage of incorrect answers significantly decreases – to 10-15% of the total number of questions. These results indicate significant issues with the reliability of answers provided by modern AI services in response to chemical questions. This may be due to both the imperfections of the algorithms themselves and the complexity of the subject area.

5. Discussion

Received results of answers to test questions on all four disciplines, obtained from ChatGPT 3.5, Gemini, Copilot and students of Sumy State Pedagogical University named after A. S. Makarenko, studying chemical disciplines, were summarized by calculating the percentage of correct answers. The results are presented in table 3.

Table 3
Number of correct answers to test tasks on chemical disciplines.

	Structure of Matter (%)	Organic Chemistry (%)	Environmental Chemistry (%)	Laboratory Chemical Practice (%)	Hallucination
Students	63.3	48.3	76.8	69.7	
ChatGPT	58.3	30	86.7	33.3	—
Gemini	63.3	20	86.7	50.0	+
Copilot	56.7	40	96.7	60.0	—

As can be seen from the results obtained, Gemini showed the best average result in the disciplines of Structure of Matter and Laboratory Chemical Practice compared to ChatGPT and Copilot. However, it should be noted that Gemini immediately offers three answer options, so-called Drafts, which differ from each other. In one of the options, the LLM hallucinated and "invented" answer choices that were not among the original options in the task, and moreover, they were incorrect. We attempted to determine how the LLMs arrived at incorrect answers to some questions. LLMs provided facts that they used for answers, but they couldn't generalize them to reach a logical conclusion. After explanations, they agreed that they made mistakes. It is worth noting that Gemini attributed the reason for the error to incomplete mastery of the Ukrainian language and lack of access to some Ukrainian sources of information. However, Gemini showed the worst performance in Organic Chemistry. In Environmental Chemistry, Gemini and ChatGPT provided the same number of correct answers, although they made mistakes in different questions. Their performance in these tests was higher than that of the students. However, Copilot showed the best results. Providing answers to questions that comprehensively assess theoretical knowledge and practical skills in handling laboratory equipment, observing safety rules, and understanding the basic methodologies used in chemical practice (Laboratory Chemical Practice), Copilot achieved the best average result, while the other LLMs performed worse than the students. This fact overall correlates with our other results, confirming the inconsistency of both LLMs in solving more complex tasks.

As our interviews revealed, students often use AI expecting absolutely correct answers and solutions. More than half of the respondents indicated that they use LLM capabilities for studying, preparing homework, and finding materials for course works, presentations, and reports.

Our research showed that although generative AI-based tools demonstrate significant success in processing and generating information, their accuracy in complex fields such as chemistry remains limited and cannot consistently provide absolutely correct and accurate answers.

Overall, the results demonstrated by LLMs in tests across different chemical disciplines were predominantly lower than the results of the students. It is worth noting that the analysis of student responses showed that they often make mistakes in lower-level questions, which involve simply demonstrating the definition of a term or phenomenon. However, they perform better in solving questions that require cognitive activity and the ability to establish logical dependencies. A generalized analysis of LLMs responses to different types of questions showed that, on the contrary, it more often provides correct answers to simple questions that only require the reproduction of facts and information, but performs worse on tasks requiring productive thinking, data analysis, and the establishment of logical relationships between facts. Moreover, in the context of Chemical science, this is not always explained by linguistic factors. Nevertheless, there is a specificity of Chemical science related to the characteristics of chemical terminology and chemical laws and exceptions to these laws, regulating the relationships between terms and facts.

Considering the current insufficient "competence" (the ability to effectively utilize knowledge and demonstrate skills) of large language models such as ChatGPT, Gemini, or Copilot in handling chemical terminology, laws, and facts, they cannot yet be considered effective tools for tutoring. This raises the question of how AI capabilities can be utilized for chemical education.

The results of our survey of students at our faculty showed (figure 2) that 61.9% of respondents occasionally used AI services to solve problems and find answers to questions in chemical disciplines. The majority of respondents who used AI encountered incorrect answers or solutions provided by generative artificial intelligence. 42.9% of students indicated that this happened often, prompting them to check and correct the AI's answers. Additionally, 47.6% of respondents encountered the phenomenon of "hallucination" in AI, where it invented nonexistent things. Moreover, 52.4% of those who noticed hallucination observed it specifically in responses to chemical questions and problems. In our opinion, the phenomenon of "hallucination" is particularly dangerous because uncritical acceptance of such answers or information presented can lead to the formation of false beliefs among students.

Our experience as users of AI fully aligns with the results we obtained. Unfortunately, it is necessary to approach almost all answers provided by LLMs to questions or tasks in the field of Chemistry critically

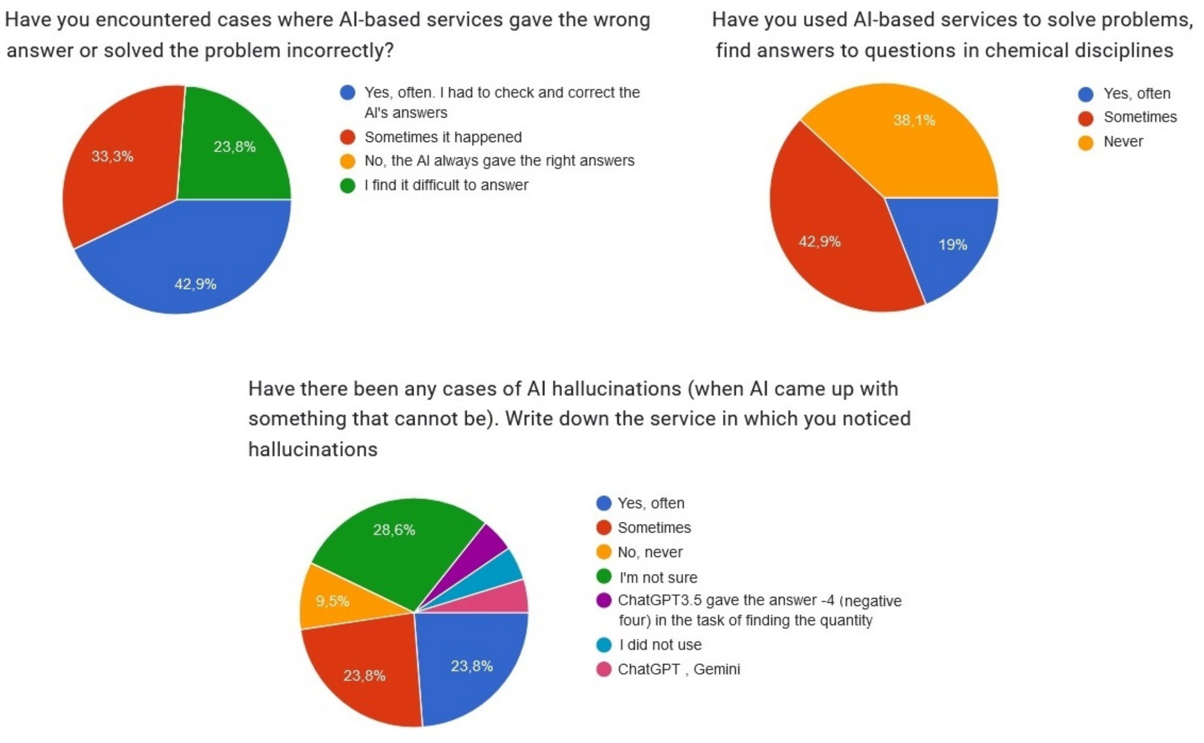

Figure 2: Students' responses.

and cautiously, verifying them. Additionally, the free version of ChatGPT 3.5 is currently unable to work with images, such as formulas, substance structures, or reaction equations. While Gemini can process graphical images, it often misinterprets them. However, there are some areas of pedagogical activity where these LLMs can be genuinely useful. This includes generating ideas of various kinds, from topics for reports or essays, creating plans for them, and even selecting theoretical material, to the topic of scientific research. For instance, we used LLMs to develop educational and methodological support for a practice-oriented course, "Equipment of the School Chemistry Classroom". With the help of AI, a comprehensive set of instructional materials was created, including lecture and practical session plans, and instructional materials for classes, which were designed based on a template provided by the teacher. AI can also assist in generating assignments and problem-based questions. For the mentioned course, test questions for individual student assignments were generated. It is important to note that all AI-generated materials were thoroughly reviewed and revised by the teacher to address any inaccuracies. Particular attention was given to verifying and adapting the test questions to ensure their accuracy and alignment with the curriculum and pedagogical requirements.

LLMs can also assist in generating tasks and problem questions. It can even be used to find ideas for possible solutions to various problem questions. However, this information still needs to be critically evaluated, meaning that the integration of AI into education does not negate the necessity of learning independently.

6. Conclusion

The conducted research has shown that while ChatGPT and Gemini are popular due to their accessibility, they are prone to errors, especially in more complex, productive-level chemistry questions. Overall, large language models based on artificial intelligence, such as ChatGPT, Gemini, and Copilot, undoubtedly have a certain potential for application in education, including chemistry education. They can provide satisfactory answers to questions at the reproductive level related to various branches of Chemistry, utilizing a vast array of information they have been provided with. In some cases, they even demonstrate better results than student answers.

Despite significant progress in the development of AI technologies, these systems still face a number of challenges and limitations in the context of chemistry education. The complexity and multifaceted nature of Chemistry, the presence of abstract concepts, specific formulas, and equations, create difficulties for the accurate processing and interpretation of information by artificial intelligence algorithms. Additionally, imperfect proficiency in the Ukrainian language and limited access to relevant sources of information can lead to errors or incorrect answers. As our research results show, the use of LLMs for organizing chemistry education is only possible in combination with careful control. Teachers should verify and evaluate the work of LLMs to ensure its accuracy.

A promising direction for the application of LLMs, in our opinion, is the organization and stimulation of discussions and debates, as LLMs are capable of generating interesting and unusual ideas that can serve as the basis for collaborative work and research. The errors and "hallucinations" made by AI tools can also be utilized by the teacher, as they can be offered to students for analysis. Through such generated errors, students can learn to critically evaluate information. When working with AI, students should understand and remember that LLMs are not infallible, and it is always necessary to verify information using other reliable sources. Another important aspect requiring the attention of researchers and educators is the improvement of LLM responses through prompt optimization.

Further research prospects in this field are as follows. Firstly, it is necessary to explore the opportunities offered by AI for chemistry education and learning in general. This will help understand which specific functions can be useful in the educational context, which tasks or processes can be automated or facilitated. These aspects may include personalized learning support, automated assessment and reporting, personalized learning, and recommendations for educational materials and methods. It is also crucial to develop students' understanding of prompt engineering strategies and to train them for more effective utilization of artificial intelligence.

Additionally, it is important to explore the impact of AI on students' motivation and engagement in learning, which is particularly relevant for chemistry education. Understanding which AI tools can stimulate students' interest and help them maintain motivation throughout the learning process is crucial for the successful integration of these technologies.

Furthermore, it is essential to explore the ethical aspects of using AI in education, including issues related to the privacy and security of student data, fairness of algorithms, and prevention of the emergence or exacerbation of inequalities in learning.

These aspects are crucial for the practical implementation of LLMs in the educational process and require in-depth study to ensure optimal performance and quality of educational outcomes. They necessitate detailed analysis, experimental research, and an interdisciplinary approach. Such a comprehensive approach to researching the effectiveness and potential applications of AI in education, including chemistry education, will enable the development of a balanced strategy for using these technologies to achieve the best outcomes for both educators and students.

References

[1] E. H. Fedorenko, V. Y. Velychko, A. V. Stopkin, A. V. Chorna, V. N. Soloviev, Informatization of education as a pledge of the existence and development of a modern higher education, CTE Workshop Proceedings 6 (2019) 20–32. doi:10.55056/cte.366.

[2] M. Lokar, The Future of E-Textbooks, International Journal for Technology in Mathematics Education 22 (2015) 101–106. doi:10.13140/RG.2.1.1142.8966.

[3] A. Sun, X. Chen, Online Education and Its Effective Practice: A Research Review, Journal of Information Technology Education: Research 15 (2016) 157–190. doi:10.28945/3502.

[4] H. E. Kentnor, Distance Education and the Evolution of Online Learning in the United States, Curriculum and Teaching Dialogue 17 (2015) 21–34. URL: https://digitalcommons.du.edu/law_facpub/24/.

[5] M. Sofi-Karim, A. Omar Bali, K. Rached, Online education via media platforms and applications as an innovative teaching method, Education and Information Technologies 28 (2023) 507–523. doi:10.1007/s10639-022-11188-0.

[6] V. Tkachuk, Y. Yechkalo, S. Semerikov, M. Kislova, Y. Hladyr, Using Mobile ICT for Online Learning During COVID-19 Lockdown, in: A. Bollin, V. Ermolayev, H. C. Mayr, M. Nikitchenko, A. Spivakovsky, M. Tkachuk, V. Yakovyna, G. Zholtkevych (Eds.), Information and Communication Technologies in Education, Research, and Industrial Applications, Springer International Publishing, Cham, 2021, pp. 46–67. doi:10.1007/978-3-030-77592-6_3.

[7] D. Oliveira, L. Pedro, C. Santos, The Use of Mobile Applications in Higher Education Classrooms: An Exploratory Measuring Approach in the University of Aveiro, Education Sciences 11 (2021) 484. doi:10.3390/educsci11090484.

[8] M. Abdinejad, B. Talaie, H. S. Qorbani, S. Dalili, Student Perceptions Using Augmented Reality and 3D Visualization Technologies in Chemistry Education, Journal of Science Education and Technology 30 (2020) 87–96. doi:10.1007/s10956-020-09880-2.

[9] Y. V. Kharchenko, O. M. Babenko, A. E. Kiv, Using Blippar to create augmented reality in chemistry education, in: S. H. Lytvynova, S. O. Semerikov (Eds.), Proceedings of the 4th International Workshop on Augmented Reality in Education (AREdu 2021), Kryvyi Rih, Ukraine, May 11, 2021, volume 2898 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. 213–229. URL: https://ceur-ws.org/Vol-2898/paper12.pdf.

[10] Y. V. Kharchenko, O. M. Babenko, O. G. Shvets, Y. V. Litsman, Possibilities of using augmented reality technology in chemistry education, Aktualni pytannia pryrodnycho-matematychnoi osvity [Topical Issues of Natural Science and Mathematics Education] (2021). URL: https://zenodo.org/record/5295793. doi:10.5281/ZENODO.5295793.

[11] I. S. Mintii, V. N. Soloviev, Augmented Reality: Ukrainian Present Business and Future Education, in: A. E. Kiv, V. N. Soloviev (Eds.), Proceedings of the 1st International Workshop on Augmented

Reality in Education, Kryvyi Rih, Ukraine, October 2, 2018, volume 2257 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2018, pp. 227–231. URL: https://ceur-ws.org/Vol-2257/paper22.pdf.

[12] T. H. Kolomoiets, D. A. Kassim, Using the Augmented Reality to Teach of Global Reading of Preschoolers with Autism Spectrum Disorders, in: A. E. Kiv, V. N. Soloviev (Eds.), Proceedings of the 1st International Workshop on Augmented Reality in Education, Kryvyi Rih, Ukraine, October 2, 2018, volume 2257 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2018, pp. 237–246. URL: https://ceur-ws.org/Vol-2257/paper24.pdf.

[13] N. O. Zinonos, E. V. Vihrova, A. V. Pikilnyak, Prospects of Using the Augmented Reality for Training Foreign Students at the Preparatory Departments of Universities in Ukraine, in: A. E. Kiv, V. N. Soloviev (Eds.), Proceedings of the 1st International Workshop on Augmented Reality in Education, Kryvyi Rih, Ukraine, October 2, 2018, volume 2257 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2018, pp. 87–92. URL: https://ceur-ws.org/Vol-2257/paper10.pdf.

[14] O. O. Lavrentieva, I. O. Arkhypov, O. P. Krupski, D. O. Velykodnyi, S. V. Filatov, Methodology of using mobile apps with augmented reality in students' vocational preparation process for transport industry, in: O. Y. Burov, A. E. Kiv (Eds.), Proceedings of the 3rd International Workshop on Augmented Reality in Education, Kryvyi Rih, Ukraine, May 13, 2020, volume 2731 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 143–162. URL: https://ceur-ws.org/Vol-2731/paper07.pdf.

[15] T. A. Vakaliuk, S. I. Pochtoviuk, Analysis of tools for the development of augmented reality technologies, in: S. H. Lytvynova, S. O. Semerikov (Eds.), Proceedings of the 4th International Workshop on Augmented Reality in Education (AREdu 2021), Kryvyi Rih, Ukraine, May 11, 2021, volume 2898 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. 119–130. URL: https://ceur-ws.org/Vol-2898/paper06.pdf.

[16] S. O. Semerikov, M. M. Mintii, I. S. Mintii, Review of the course "Development of Virtual and Augmented Reality Software" for STEM teachers: implementation results and improvement potentials, in: S. H. Lytvynova, S. O. Semerikov (Eds.), Proceedings of the 4th International Workshop on Augmented Reality in Education (AREdu 2021), Kryvyi Rih, Ukraine, May 11, 2021, volume 2898 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. 159–177. URL: https://ceur-ws.org/Vol-2898/paper09.pdf.

[17] O. B. Petrovych, A. P. Vinnichuk, V. P. Krupka, I. A. Zelenenka, A. V. Voznyak, The usage of augmented reality technologies in professional training of future teachers of Ukrainian language and literature, in: S. H. Lytvynova, S. O. Semerikov (Eds.), Proceedings of the 4th International Workshop on Augmented Reality in Education (AREdu 2021), Kryvyi Rih, Ukraine, May 11, 2021, volume 2898 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. 315–333. URL: https://ceur-ws.org/Vol-2898/paper17.pdf.

[18] R. O. Tarasenko, S. M. Amelina, S. O. Semerikov, V. D. Shynkaruk, Using interactive semantic networks as an augmented reality element in autonomous learning, Journal of Physics: Conference Series 1946 (2021) 012023. doi:10.1088/1742-6596/1946/1/012023.

[19] V. V. Babkin, V. V. Sharavara, V. V. Sharavara, V. V. Bilous, A. V. Voznyak, S. Y. Kharchenko, Using augmented reality in university education for future IT specialists: educational process and student research work, in: S. H. Lytvynova, S. O. Semerikov (Eds.), Proceedings of the 4th International Workshop on Augmented Reality in Education (AREdu 2021), Kryvyi Rih, Ukraine, May 11, 2021, volume 2898 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. 255–268. URL: https://ceur-ws.org/Vol-2898/paper14.pdf.

[20] S. P. Palamar, G. V. Bielienka, T. O. Ponomarenko, L. V. Kozak, L. L. Nezhyva, A. V. Voznyak, Formation of readiness of future teachers to use augmented reality in the educational process of preschool and primary education, in: S. H. Lytvynova, S. O. Semerikov (Eds.), Proceedings of the 4th International Workshop on Augmented Reality in Education (AREdu 2021), Kryvyi Rih, Ukraine, May 11, 2021, volume 2898 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021. URL: https://ceur-ws.org/Vol-2898/paper18.pdf.

[21] S. Papadakis, A. E. Kiv, H. M. Kravtsov, V. V. Osadchyi, M. V. Marienko, O. P. Pinchuk, M. P. Shyshkina, O. M. Sokolyuk, I. S. Mintii, T. A. Vakaliuk, L. E. Azarova, L. S. Kolgatina, S. M. Amelina, N. P. Volkova, V. Y. Velychko, A. M. Striuk, S. O. Semerikov, Unlocking the power of synergy: the

joint force of cloud technologies and augmented reality in education, in: S. O. Semerikov, A. M. Striuk (Eds.), Joint Proceedings of the 10th Workshop on Cloud Technologies in Education, and 5th International Workshop on Augmented Reality in Education (CTE+AREdu 2022), Kryvyi Rih, Ukraine, May 23, 2022, volume 3364 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2022, pp. 1–23. URL: https://ceur-ws.org/Vol-3364/paper00.pdf.

[22] D. A. Karnishyna, T. V. Selivanova, P. P. Nechypurenko, T. V. Starova, V. G. Stoliarenko, The use of augmented reality in chemistry lessons in the study of "Oxygen-containing organic compounds" using the mobile application Blippar, Journal of Physics: Conference Series 2288 (2022) 012018. doi:10.1088/1742-6596/2288/1/012018.

[23] M. M. Mintii, N. M. Sharmanova, A. O. Mankuta, O. S. Palchevska, S. O. Semerikov, Selection of pedagogical conditions for training STEM teachers to use augmented reality technologies in their work, Journal of Physics: Conference Series 2611 (2023) 012022. doi:10.1088/1742-6596/2611/1/012022.

[24] M. V. Marienko, S. O. Semerikov, O. M. Markova, Artificial intelligence literacy in secondary education: methodological approaches and challenges, in: S. Papadakis (Ed.), Proceedings of the 11th Workshop on Cloud Technologies in Education (CTE 2023), Kryvyi Rih, Ukraine, December 22, 2023, volume 3679 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2023, pp. 87–97. URL: https://ceur-ws.org/Vol-3679/paper21.pdf.

[25] O. M. Haranin, N. V. Moiseienko, Adaptive artificial intelligence in RPG-game on the Unity game engine, in: A. E. Kiv, S. O. Semerikov, V. N. Soloviev, A. M. Striuk (Eds.), Proceedings of the 1st Student Workshop on Computer Science & Software Engineering, Kryvyi Rih, Ukraine, November 30, 2018, volume 2292 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2018, pp. 143–150. URL: http://ceur-ws.org/Vol-2292/paper16.pdf.

[26] F. Miao, W. Holmes, Guidance for generative AI in education and research, UNESCO, 2023. doi:10.54675/EWZM9535.

[27] B. Anders, Is using ChatGPT cheating, plagiarism, both, neither, or forward thinking?, Patterns 4 (2023) 100694. doi:10.1016/j.patter.2023.100694.

[28] M. Yang, New York City schools ban AI chatbot ChatGPT, 2023. URL: https://www.theguardian.com/us-news/2023/jan/06/new-york-city-schools-ban-ai-chatbot-chatgpt.

[29] A. Tlili, B. Shehata, M. Adarkwah, A. Bozkurt, D. Hickey, R. Huang, B. Agyemang, What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education, Smart Learning Environments 15 (2023) 1–24. doi:10.1186/s40561-023-00237-x.

[30] D. Lee, M. Arnold, A. Srivastava, K. Plastow, P. Strelan, F. Ploeckl, D. Lekkas, E. Palmer, The impact of generative AI on higher education learning and teaching: A study of educators: perspectives, Computers and Education: Artificial Intelligence 6 (2024) 100221. doi:10.1016/j.caeai.2024.100221.

[31] H. Crompton, D. Burke, Artificial intelligence in higher education: the state of the field, International Journal of Educational Technology in Higher Education 20 (2023). doi:10.1186/s41239-023-00392-8.

[32] F. Ouyang, L. Zheng, P. Jiao, Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020, Education and Information Technologies 27 (2022) 7893–7925. doi:10.1007/s10639-022-10925-9.

[33] E. Sabzalieva, A. Valentini, ChatGPT and artificial intelligence in higher education: quick start guide, UNESCO, 2023. URL: https://unesdoc.unesco.org/ark:/48223/pf0000385146.

[34] T. M. Clark, Investigating the Use of an Artificial Intelligence Chatbot with General Chemistry Exam Questions, Journal of Chemical Education 100 (2023) 1905–1916. doi:10.1021/acs.jchemed.3c00027.

[35] M. Alam, M. Hasan, Applications and Future Prospects of Artificial Intelligence in Education, International Journal of Humanities & Social Science Studies (IJHSSS) 10 (2024) 197–206. doi:10.29032/ijhsss.v10.i1.2024.197-206.

[36] A. Al Darayseh, Acceptance of artificial intelligence in teaching science: Science teachers' perspective, Computers and Education: Artificial Intelligence 4 (2023) 100132. doi:10.1016/

j.caeai.2023.100132.
[37] R. P. dos Santos, Enhancing Chemistry Learning with ChatGPT and Bing Chat as Agents to Think With: A Comparative Case Study, 2023. doi:10.48550/arXiv.2311.00709.
[38] T. Nguyen, L. Cao, P. Nguyen, V. Tran, P. Nguyen, Capabilities, Benefits, and Role of ChatGPT in Chemistry Teaching and Learning in Vietnamese High Schools, 2023. doi:10.35542/osf.io/4wt6q.
[39] D. O. Williams, E. Fadda, Can ChatGPT pass Glycobiology?, Glycobiology 33 (2023) 606–614. doi:10.1093/glycob/cwad064.
[40] D. Xuan-Quy, N.-B. Le, T.-D. Vo, X.-D. Phan, B.-B. Ngo, V.-T. Nguyen, T.-M.-T. Nguyen, H.-P. Nguyen, VNHSGE: VietNamese High School Graduation Examination Dataset for Large Language Models, 2023. doi:10.48550/arXiv.2305.12199.
[41] D. Xuan-Quy, N.-B. Le, X.-D. Phan, B.-B. Ngo, An Evaluation of ChatGPT's Proficiency in English Language Testing of The Vietnamese National High School Graduation Examination, 2023. doi:10.2139/ssrn.4473369.
[42] D. Xuan-Quy, N.-B. Le, X.-D. Phan, B.-B. Ngo, T.-D. Vo, Evaluation of ChatGPT and Microsoft Bing AI Chat Performances on Physics Exams of Vietnamese National High School Graduation Examination, 2023. doi:10.48550/arXiv.2306.04538.
[43] D. Xuan-Quy, N.-B. Le, LLMs' Capabilities at the High School Level in Chemistry: Cases of ChatGPT and Microsoft Bing AI Chat, 2023. doi:10.26434/chemrxiv-2023-kxxpd.
[44] D. Xuan-Quy, N.-B. Le, Investigating the Effectiveness of ChatGPT in Mathematical Reasoning and Problem Solving: Evidence from the Vietnamese National High School Graduation Examination, 2023. doi:10.48550/arXiv.2306.06331.
[45] T. Humphry, A. Fuller, Potential ChatGPT Use in Undergraduate Chemistry Laboratories, Journal of Chemical Education 100 (2023) 1434–1436. doi:10.1021/acs.jchemed.3c00006.
[46] J. L. Araújo, I. Saúde, Can ChatGPT Enhance Chemistry Laboratory Teaching? Using Prompt Engineering to Enable AI in Generating Laboratory Activities, Journal of Chemical Education 101 (2024) 1858–1864. doi:10.1021/acs.jchemed.3c00745.
[47] Y. Du, C. Duan, A. Bran, A. Sotnikova, Y. Qu, H. Kulik, A. Bosselut, J. Xu, P. Schwaller, Large Language Models are Catalyzing Chemistry Education, 2024. doi:10.26434/chemrxiv-2024-h722v.
[48] J. Leinonen, P. Denny, S. MacNeil, S. Sarsa, S. Bernstein, J. Kim, A. Tran, A. Hellas, Comparing Code Explanations Created by Students and Large Language Models, in: Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1, ITiCSE 2023, Association for Computing Machinery, New York, NY, USA, 2023, p. 124–130. doi:10.1145/3587102.3588785.
[49] T. Kung, M. Cheatham, A. Medenilla, C. J. Sillos, L. Leon, C. Elepaño, M. Madriaga, R. Aggabao, G. Diaz-Candido, J. Maningo, V. Tseng, Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models, PLOS Digital Health 2 (2023) e0000198. doi:10.1371/journal.pdig.0000198.
[50] P. Maier, G. J. Klinker, Augmented Chemical Reactions: 3D Interaction Methods for Chemistry, Int. J. Online Biomed. Eng. 9 (2013) 80–82. doi:10.3991/ijoe.v9iS8.3411.
[51] M. Núñez, R. Quirós, I. Núñez, J. B. Carda, E. Camahort, Collaborative augmented reality for inorganic chemistry education, in: Proceedings of the 5th WSEAS/IASME International Conference on Engineering Education, EE'08, World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, Wisconsin, USA, 2008, p. 271–277.
[52] P.-K. Oh, S.-J. Kang, Integrating Artificial Intelligence to Chemistry Experiment: Carbon Dioxide Fountain, Journal of Chemical Education 98 (2021) 2376–2380. doi:10.1021/acs.jchemed.1c00004.
[53] E. F. Healy, G. Blade, Tips and Tools for Teaching Organic Synthesis Online, Journal of Chemical Education 97 (2020) 3163–3167. doi:10.1021/acs.jchemed.0c00473.
[54] L. Joss, E. A. Müller, Machine Learning for Fluid Property Correlations: Classroom Examples with MATLAB, Journal of Chemical Education 96 (2019) 697–703. doi:10.1021/

acs.jchemed.8b00692.

[55] W. J. D. Nascimento Júnior, C. Morais, G. Girotto Júnior, Enhancing AI Responses in Chemistry: Integrating Text Generation, Image Creation, and Image Interpretation through Different Levels of Prompts, Journal of Chemical Education (2024). doi:10.1021/acs.jchemed.4c00230.

[56] S. Tassoti, Assessment of Students Use of Generative Artificial Intelligence: Prompting Strategies and Prompt Engineering in Chemistry Education, Journal of Chemical Education 101 (2024) 2475–2482. doi:10.1021/acs.jchemed.4c00212.

[57] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, G. Neubig, Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing, ACM Computing Surveys 55 (2023). doi:10.1145/3560815.

[58] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. H. Chi, Q. V. Le, D. Zhou, Chain-of-thought prompting elicits reasoning in large language models, in: Proceedings of the 36th International Conference on Neural Information Processing Systems, NIPS '22, Curran Associates Inc., Red Hook, NY, USA, 2024. URL: https://dl.acm.org/doi/10.5555/3600270.3602070#.

[59] S. Yao, D. Yu, J. Zhao, I. Shafran, T. L. Griffiths, Y. Cao, K. Narasimhan, Tree of thoughts: deliberate problem solving with large language models, in: Proceedings of the 37th International Conference on Neural Information Processing Systems, NIPS '23, Curran Associates Inc., Red Hook, NY, USA, 2024. URL: https://dl.acm.org/doi/abs/10.5555/3666122.3666639.

[60] A. Kishimoto, B. Buesser, A. Botea, AI meets chemistry, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 32 of *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, AAAI Press, 2018, pp. 7978–7982. doi:10.1609/aaai.v32i1.12216.

[61] B. Pitaychuk, Spilkuytesya z chat botami angliyskoyu shchob ne otrimuvati u vidpovid nisenitnits. Chi bude ChatGPT krashche pratsyuvati z ukrainskoyu? Poyasnyue ShI ekspert Bogdan Pitaychuk [Communicate with chatbots in English to avoid receiving nonsense. Will ChatGPT work well in Ukrainian? AI expert Bohdan Pytaichuk explains], 2023. URL: https://forbes.ua/innovations/spilkuytesya-z-chat-botami-angliyskoyu-shchob-ne-otrimuvati-u-vidpovid-nisenitnits-chi-bude-chatgpt-krashche-pratsyuvati-z-ukrainskoyu-poyasnyue-shi-ekspert-bogdan-pitaychuk-01112023-17015.

[62] C. Nascimento, A. Pimentel, Do Large Language Models Understand Chemistry? A conversation with ChatGPT, Journal of Chemical Information and Modeling 63 (2023) 1649–1655. doi:10.1021/acs.jcim.3c00285.

[63] S. Fergus, M. Botha, M. Ostovar, Evaluating Academic Answers Generated Using ChatGPT, Journal of Chemical Education 100 (2023) 1672–1675. doi:10.1021/acs.jchemed.3c00087.