

Article

Artificial Intelligence Generative Tools and Conceptual Knowledge in Problem Solving in Chemistry

Wajeeh Daher ^{1,*} , Hussam Diab ² and Anwar Rayan ² 

¹ Mathematics Department, Al-Qasemi Academic College of Education, Baqa 3010000, Israel

² Science Department, Al-Qasemi Academic College of Education, Baqa 3010000, Israel; hussamono@gmail.com (H.D.); a_rayan@qsm.ac.il (A.R.)

* Correspondence: wdaher@qsm.ac.il

Abstract: In recent years, artificial intelligence (AI) has emerged as a valuable resource for teaching and learning, and it has also shown promise as a tool to help solve problems. A tool that has gained attention in education is ChatGPT, which supports teaching and learning through AI. This research investigates the difficulties faced by ChatGPT in comprehending and responding to chemistry problems pertaining to the topic of Introduction to Material Science. By employing the theoretical framework proposed by Holme et al., encompassing categories such as transfer, depth, predict/explain, problem solving, and translate, we evaluate ChatGPT's conceptual understanding difficulties. We presented ChatGPT with a set of thirty chemistry problems within the Introduction to Material Science domain and tasked it with generating solutions. Our findings indicated that ChatGPT encountered significant conceptual knowledge difficulties across various categories, with a notable emphasis on representations and depth, where difficulties in representations hindered effective knowledge transfer.

Keywords: artificial intelligence; ChatGPT; conceptual knowledge; chemistry; problem solving



Citation: Daher, W.; Diab, H.; Rayan, A. Artificial Intelligence Generative Tools and Conceptual Knowledge in Problem Solving in Chemistry. *Information* **2023**, *14*, 409. <https://doi.org/10.3390/info14070409>

Academic Editors: Surbhi Bhatia, M.K. Nallakuruppan and Jose Santamaria

Received: 25 June 2023
Revised: 6 July 2023
Accepted: 12 July 2023
Published: 16 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Artificial intelligence (AI), including the popular ChatGPT, has garnered significant attention from educational researchers due to its potential in the field of education [1,2]. ChatGPT has been recognized for its problem-solving capabilities across various disciplines [3]. This raises the question of whether ChatGPT can acquire a comprehensive conceptual understanding of concepts and relationships across diverse disciplines.

In a recent study, West [4] compared the performance of ChatGPT 3.5 and ChatGPT 4.0 in terms of their understanding of introductory physics. The author observed that ChatGPT 3.5 occasionally responded like an expert physicist but lacked consistency, often displaying both impressive mastery and complete incoherence of the same topics. In contrast, ChatGPT 4.0 demonstrated several facets of true “understanding” such as the ability to engage in metaphorical thinking and utilize multiple representations. However, the author cautioned against claiming that ChatGPT 4.0 exhibited a comprehensive understanding of introductory physics on a global scale.

In this research, we investigated the conceptual understanding of ChatGPT in the field of chemistry, specifically focusing on the topic of Introduction to Material Science. In this study, we explored how ChatGPT performed when presented with problems typically assigned to college students enrolled on an Introduction to Material Science course. Two of the main aspects of the present research, AI and conceptual knowledge, are related to sustainable education.

1.1. Theoretical Background and Literature Review

1.1.1. Artificial Intelligence and ChatGPT in Education

Gocen and Aydemir [1] highlighted several benefits of AI in education, including the ability to facilitate personalized learning at one's own pace and the potential to make informed decisions using vast amounts of data. However, they also acknowledged the concern that an overemphasis on utilitarian perspectives may overshadow humanistic values.

Dwivedi et al. [5] emphasized that AI, through its ability to predict and adapt to changes, offers systematic reasoning based on inputs and learning from expected outcomes. According to the Centre for Learning, Teaching and Development [3], ChatGPT serves various functions in the classroom, including its use in higher education to support instructional activities while maintaining academic integrity. These activities encompass creative writing, problem solving, research projects, group projects, and classroom debates. Moreover, Aljanabi et al. [6] argued that ChatGPT has opened possibilities regarding its use in education, including as an assistant for academic writing, a search engine, an assistant in coding, a detector of security vulnerabilities, and a social media expert and agent.

Firat [7] found that, according to their frequency of recurrence, educators perceived ChatGPT as enabling "evolution of learning and education systems", "changing role of educators", "impact on assessment and evaluation", "ethical and social considerations", "future of work and employability", "personalized learning", "digital literacy and AI integration", "AI as an extension of the human brain", and "importance of human characteristics".

Additionally, Dai et al. [2] highlighted the potential of ChatGPT to provide personalized support in real-time, delivering tailored explanations and scaffolding to sensitively and coherently address individual learners' difficulties.

Many educational applications of AI are available, including intelligent tutoring, technology-based learning platforms, and automated rating systems [8]. Research has shown that ChatGPT increases learning motivation and engagement among students [9]. In the study by Orrù et al. [10], ChatGPT demonstrated its problem-solving capabilities by achieving highly probable outcomes in both practice problems and pooled problems, ranking among the top 5% of answer combinations.

1.1.2. Conceptual Knowledge of Chemistry

Cracolice et al. [11] asserted that although learning algorithms can be valuable for problem solving, they should not be the sole method used to teach chemistry. Roth [12] argued that meaningful conceptual knowledge requires a deeper understanding that goes beyond factual knowledge and necessitates a fresh perspective on the subject. This understanding should encompass prediction and the ability to construct explanations. Puk and Stibbards [13] further emphasized that conceptual understanding involves recognizing the connections between different concepts. Additionally, Lansangan et al. [14] suggested that representations can be effective in assessing knowledge in a discipline, providing alternatives to traditional assessment methods. These representations could take the form of written or drawn symbols, iconic gestures or diagrams, and spoken, gestured, written, or drawn indices.

Holme et al. [15] conducted a study involving instructors who taught chemistry in higher-education settings, aiming to define conceptual understanding in chemistry. As a result, they identified five categories that define conceptual understanding:

- Transfer: the ability to apply core chemistry ideas to novel chemical situations.
- Depth: the capacity to reason about core chemistry ideas using skills that extend beyond rote memorization or algorithmic problem solving.
- Predict/explain: the capability of extending situational knowledge to predict and/or explain the behavior of chemical systems.
- Problem solving: demonstrating critical thinking and reasoning while solving problems, including those involving laboratory measurements.
- Translate: the ability to translate across different scales and representations.

These categories provide a framework to understand and evaluate conceptual understanding in the context of chemistry education. Considering the solution of chemistry problems by ChatGPT, we took into account that the problems could be from various levels of knowledge, according to Bloom's taxonomy.

1.1.3. Bloom's Knowledge Taxonomy

Researchers have utilized Bloom's taxonomy of knowledge since the suggestion of this taxonomy. The taxonomy consists of six levels: remembering, understanding, application, analysis, synthesis, and evaluation. Educational researchers have utilized this taxonomy to measure the level of students' knowledge in specific contexts. Prasad [16] discussed how Bloom's taxonomy could be used as a method to assess the critical thinking of students. A study of the textbook questions using Bloom's taxonomy found that about 40% emphasized higher-order thinking [17]. Daher and Sleem [18] used Bloom's taxonomy to examine the level of traditional, video-based, and 360-degree video-based learning in a social studies classroom. They found that students who learned from the video and the 360-degree video contexts had significantly higher scores in the synthesis knowledge level than students who learned in a traditional context.

1.2. Research Goals and Rationale

AI, including ChatGPT, has been recognized as a valuable tool to support students' learning in the field of science, particularly in chemistry [19,20]. However, for AI to effectively support learning, it needs to possess conceptual knowledge of scientific concepts and their relationships, specifically within the domain of chemistry. Previous research has examined various aspects related to conceptual knowledge in chemistry.

In this research, we aimed to investigate conceptual knowledge of ChatGPT in chemistry, specifically focusing on the topic of Introduction to Material Science. To guide our investigation, we adopted the theoretical framework proposed by Holme et al. [15], which identifies key components of conceptual knowledge in Chemistry, including depth, predict/explain, problem solving, and translate.

The present study contributes to the future implementation of AI and ChatGPT in education, particularly in the sciences, ensuring their effective use. Furthermore, the findings of this research could inform the development of AI tools that exhibit fewer difficulties in solving scientific problems and possess enhanced conceptual knowledge relevant to the sciences. All the previous arguments indicate the relationship between the present research with sustainable education, where a more profound performance of AI generative tools contributes to the prevalence of sustainable education.

1.3. Research Question

What are the conceptual knowledge difficulties encountered by ChatGPT when solving chemistry problems pertaining to the topic of Introduction to Material Science?

2. Methodology

2.1. Research Context

The research was conducted within the framework of an Introduction to Material Science course, which serves as an introductory course to fundamental topics and concepts in the field of chemistry. The course encompasses essential elements, including particles, atoms, molecules, ions, chemical bonds, the periodic table, and the chemical and physical properties of substances. Furthermore, it covers basic chemical reactions, including their definitions, properties, and the principles of chemical kinetics. By participating in this course, students acquire a solid foundation in the fundamental principles of chemistry and develop a comprehensive understanding of its key components.

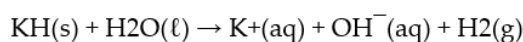
2.2. Data Collection Tools

We gathered responses from ChatGPT on a set of chemistry problems pertaining to the topic of Introduction to Material Science. The problem set consisted of two types: open-ended problems, as exemplified by Figure 1; and multiple-choice problems, as exemplified by Figure 2.

When you put a piece of chalk in a glass vessel that contains water and blow with a straw into the water for half an hour, you see the formation of gas bubbles on the surface of the chalk and a small reduction in the volume of the piece of chalk. Formulate the chemical reaction that caused the phenomena and explain in your own words what happened?

Figure 1. An open-ended chemistry problem.

If you know that the reaction that takes place when potassium hydride (KH(s)) is added to water is:



How does the reaction affect the pH of the solution?

1. It changes from an acidic solution to a basic one
2. It changes from a basic solution to an acidic solution
3. It changes from a neutral solution to a basic solution
4. It changes from a neutral solution to an acidic solution

Figure 2. A multiple-choice chemistry problem.

2.3. Data Analysis Tools

We conducted an analysis of ChatGPT's responses to the chemistry problems using the theoretical framework proposed by Holme et al. [15]. Specifically, we examined the conceptual understanding difficulties faced by ChatGPT in relation to the categories outlined in the framework: transfer, depth, predict/explain, problem solving, and translate. To provide a comprehensive overview, Table 1 presents the themes associated with each category, along with the corresponding difficulties observed during the analysis.

Table 1. Themes of the conceptual knowledge categories.

Category	Themes
Transfer	Transfer, transition, and implementation
Depth	Awareness of rules, awareness of reason, and awareness of nature
Predict/explain	Prediction, indication, explaining, because, and leading to
Problem solving	Critical thinking and procedural knowledge
Translate	Translate and representation

2.4. Validity and Reliability

We conducted an analysis of 30 chemistry problems pertaining to the topic of Introduction to Material Science, consisting of 15 open-ended problems and 15 multiple-choice

problems. From this dataset, we analyzed the responses generated by ChatGPT for 5 open-ended problems and 8 multiple-choice problems. Our analysis did not reveal any new types of difficulties related to conceptual understanding. This suggests that we reached saturation concerning the conceptual understanding categories and their properties [21,22].

To ensure a reliable data analysis, two experienced coders utilized the adopted conceptual framework to code the ChatGPT answers. Their coding process involved identifying sentences indicative of conceptual knowledge themes. The Cohen's Kappa coefficients obtained for the coder agreement ranged from 0.89 to 0.91 for the conceptual understanding categories, which was considered acceptable. These results further support the reliability of our data analysis.

3. Results

In the following section, we present the solutions provided by ChatGPT for various chemistry problems. Throughout the analysis, we highlighted the difficulties encountered by ChatGPT specifically related to the five aspects of conceptual understanding as outlined by Holme et al. [15].

Table 2 provides detailed information about the problem types, Bloom's level of the given chemical problems, and the specific conceptual difficulties encountered by the generative AI.

Table 2. Problem types, Bloom's level of the given problems, and the specific conceptual difficulties encountered by the generative AI.

Question Number	Bloom's Taxonomy Level	Type of Question (Open-Ended/True or False/Multiple Choice)	Type of Difficulty
1	Application	Multiple choice	-
2	Understanding	Multiple choice	D-reasoning
3	Analysis	Multiple choice	D-rules
4	Remembering	Multiple choice	-
5	Analysis	Multiple choice	D-understanding
6	Application	Multiple choice	D-understanding
7	Understanding	Multiple choice	-
8	Remembering	Multiple choice	-
9	Understanding	Multiple choice	-
10	Application	Multiple choice	Problem solving
11	Remembering	Multiple choice	-
12	Understanding	Multiple choice	-
13	Analysis	Multiple choice	D-understanding
14	Remembering	Multiple choice	-
15	Understanding	Multiple choice	D-reasoning
16	Application	Open-ended	D-understanding
17	Analysis	Open-ended	-
18	Analysis	Open-ended	Problem solving
19	Application	Open-ended	D-understanding
18	Application	Open-ended	D-understanding
21	Application	Open-ended	Problem solving

Table 2. Cont.

Question Number	Bloom's Taxonomy Level	Type of Question (Open-Ended/True or False/Multiple Choice)	Type of Difficulty
22	Application	Open-ended	-
23	Analysis	Open-ended	Problem solving
24	Application	Open-ended	D-Understanding
25	Remembering	Open-ended	-
26	Synthesis	Open-ended	Transfer
27	Synthesis	Open-ended	-
28	Synthesis	Open-ended	D-Understanding
29	Evaluation	Open-ended	-
30	Evaluation	Open-ended	Translation

D: Depth.

Table 2 shows that, generally speaking, ChatGPT encountered no conceptual difficulties in solving remembering problems; it encountered depth difficulties of the understanding type when solving understanding, application, and analysis problems; it encountered transfer difficulties when solving synthesis problems; and it encountered translation difficulties when solving evaluation problems.

The following section provides a comprehensive description of the various conceptual difficulties encountered by ChatGPT.

3.1. Difficulties Related to the Depth Issue

The depth category comprises two sub-categories that pertain to the conceptual challenges encountered by ChatGPT: awareness of chemical rules and understanding the nature of specific compounds. In the following sections, we provide a detailed description of each sub-category.

3.1.1. First Depth Issue: Awareness of the Chemical Rules

Table 3 features a problem in which ChatGPT's response exhibited a difficulty associated with the awareness of chemical rules.

Table 3. ChatGPT's difficulty with awareness of chemical rules.

Problem	Any unlikely reaction? 1. $F_2 + 2HCl \rightarrow 2HF + Cl_2$ 2. $Br_2 + 2HCl \rightarrow 2HBr + Cl_2$ 3. $F_2 + 2HI \rightarrow 2HF + I_2$ 4. $Br_2 + 2HI \rightarrow 2HBr + I_2$
Answer	All the given reactions involve halogens and hydrogen halides, which are expected to react to form halides of the respective elements and hydrogen gas. Therefore, all the reactions are likely to occur. However, in the third reaction, $F_2 + 2HI \rightarrow 2HF + I_2$, the product hydrogen fluoride (HF) is a strong acid and can react with iodine (I_2) to form hydrogen iodide (HI) and iodine pentafluoride (IF_5) as follows: $2HF + I_2 \rightarrow 2HI + IF_5$. This reaction is an unlikely side reaction that may occur along with the main reaction. Therefore, the correct answer is option 3, $F_2 + 2HI \rightarrow 2HF + I_2$, due to the potential side reaction of HF and I_2 forming IF_5 . "A halogen displacement reaction occurs when a more reactive halogen displaces a less reactive halogen from an aqueous solution of its halide".

Based on the explanation, it appeared that ChatGPT lacked awareness of the rules governing displacement reactions of halides. This highlighted a depth-related issue regarding ChatGPT's understanding of the chemistry rules associated with such reactions. Specifically, ChatGPT was not aware that the reactivity of halogens decreases in the following order: $F_2 > Cl_2 > Br_2 > I_2$.

3.1.2. Second Depth Issue: Awareness of the Nature of a Specific Compound

Table 4 includes a problem that, when ChatGPT answered, the answer included a difficulty related to the awareness of the nature of a specific compound.

Table 4. ChatGPT's difficulty with awareness of the nature of a specific compound.

Problem	If you know that the reaction that takes place when potassium hydride ($KH_{(s)}$) is added to water is:
	$KH_{(s)} + H_2O_{(l)} \rightarrow K^+_{(aq)} + OH^-_{(aq)} + H_{2(g)}$
Answer	How does the reaction affect the pH of the solution? 1. It changes from an acidic solution to a basic one 2. It changes from a basic solution to an acidic solution 3. It changes from a neutral solution to a basic solution 4. It changes from a neutral solution to an acidic solution
	The reaction of potassium hydride (KH) with water (H_2O) produces potassium hydroxide (KOH) and hydrogen gas (H_2) along with hydroxide ions (OH^-) in an aqueous solution. The hydroxide ions (OH^-) are responsible for the increase in pH of the solution, making it more basic. Therefore, the correct answer is option 1: it changes from an acidic solution to a basic one.

It appeared that there may have been a misunderstanding as KH should not be considered an acidic solution. It is right that most of the inorganic compounds with H are acidic, but KH (potassium hydride) reacts as a basic compound when added to water.

3.2. Difficulties Related to the Problem-Solving Issue

Table 5 includes a problem that, when ChatGPT answered, the answer included a difficulty related to the problem-solving category.

Table 5. ChatGPT's difficulty with problem solving.

Problem	In a closed vessel, 200 mL of 0.25 M acetic acid (CH_3COOH) solution was mixed with 100 mL of 0.5 M NaOH aqueous solution. The temperature of the liquid increased by 2.2 °C. What is the value of the heat of reaction (ΔH) for the reaction?
	Original: $q = (200\text{ g} + 100\text{ g}) \times 4.18\text{ J/g}\cdot\text{K} \times 2.2\text{ }^\circ\text{C}$, $q = 1983.6\text{ J}$, $\Delta H = q/\text{moles of NaOH} = 1983.6\text{ J}/0.05\text{ mol} = 39,672\text{ J/mol}$. The correct calculation would be $q = (200\text{ g} + 100\text{ g}) \times 4.18\text{ J/g}\cdot\text{K} \times 2.2\text{ }^\circ\text{C} = 2758.8\text{ J}$. $\Delta H = q/n$, where $n = \text{number of moles of NaOH or } CH_3COOH \text{ used}$. $\Delta H = 2758.8\text{ J}/0.05\text{ mol} = 55,176\text{ J/mol} = 55.176\text{ KJ/mol}$.

During the problem-solving process, ChatGPT demonstrated correct reasoning. However, there were errors in the arithmetic computations, leading to an incorrect answer.

Upon receiving the problem again after a 15 min interval, ChatGPT provided another incorrect answer. The mistakes were attributed to miscalculations in the determination of q and the number of moles of the product.

When informed of its mistake in calculating q , ChatGPT responded politely:

"I apologize for the mistake in my previous response. Thank you for bringing it to my attention. Here is the corrected calculation". The answer again included a

mistake in the calculation. We requested ChatGPT to reconsider the calculation: “are you sure in calculating q: $q = 0.3 \text{ kg} \times 4.18 \text{ J}/(\text{g} \cdot ^\circ\text{C}) \times 2.2 ^\circ\text{C} = 2.90 \text{ kJ}$?”. ChatGPT replied: “I apologize again for the mistake. Thank you for bringing it to my attention. The correct calculation for q is ... ”

In this instance, ChatGPT provided an incorrect summation. Instead of the accurate value of 2.7558 KJ, it mistakenly calculated the sum as 2.90 KJ.

ChatGPT acknowledged the repeated miscalculation and apologized, stating: “I apologize for any confusion my previous mistake may have caused”.

In this case, the difficulty encountered by ChatGPT related to problem solving and critical thinking applied to the given problem. The expression $(200 \text{ g} + 100 \text{ g}) \times 4.18 \text{ J}/\text{g} \cdot \text{K} \times 2.2 ^\circ\text{C}$ indicated that q should be greater than 2400 ($300 \times 4 \times 2$). Therefore, ChatGPT should have recognized that the previously provided answer was illogical.

3.3. Difficulties Related to the Explanation Issue

Table 6 includes a problem that, when ChatGPT answered, the answer included a difficulty related to the explanation category.

Table 6. ChatGPT’s difficulty with explanation.

Problem	Balance the following reaction: $\text{Cu}(\text{NO}_3)_2(\text{aq}) + \text{NaOH}(\text{aq}) \rightarrow \text{Cu}(\text{OH})_2(\text{s}) + \text{NaNO}_3(\text{aq})$
	Does the reaction give an acidic solution or a basic solution? Explain!
Answer	The balanced reaction is: $\text{Cu}(\text{NO}_3)_2(\text{aq}) + 2\text{NaOH}(\text{aq}) \rightarrow \text{Cu}(\text{OH})_2(\text{s}) + 2\text{NaNO}_3(\text{aq})$
	In this reaction, $\text{Cu}(\text{NO}_3)_2(\text{aq})$ and $\text{NaOH}(\text{aq})$ react to form $\text{Cu}(\text{OH})_2(\text{s})$ and $\text{NaNO}_3(\text{aq})$. The coefficient of $\text{NaOH}(\text{aq})$ is 2, which means that two moles of NaOH are required to react with one mole of $\text{Cu}(\text{NO}_3)_2$. When NaOH is added to a $\text{Cu}(\text{NO}_3)_2$ solution, it reacts with the HNO_3 produced by the dissociation of $\text{Cu}(\text{NO}_3)_2$ to form water and NaNO_3 . The resulting solution has a pH greater than 7, making it basic. Therefore, the reaction gives a basic solution.

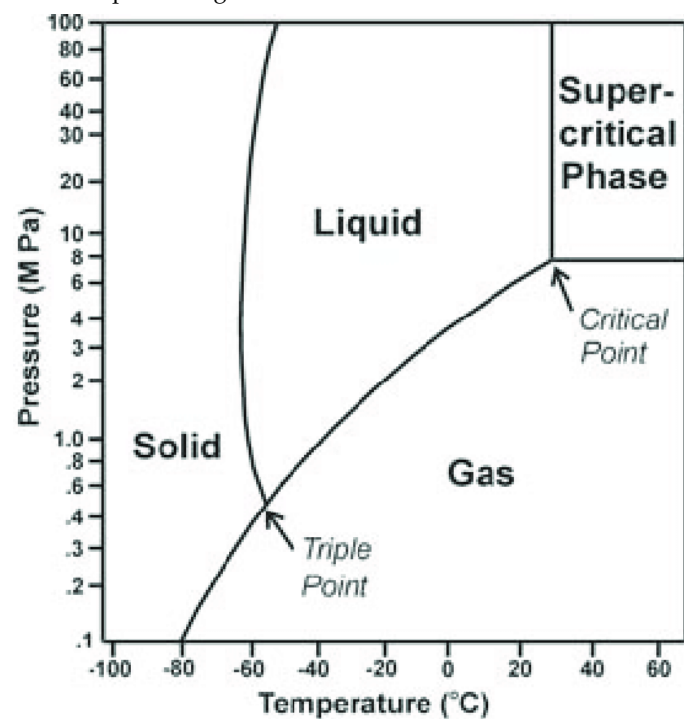
In the given solution, the chemical reaction balancing was correct. However, the conclusion drawn about the type of solution produced and the accompanying explanation were both incorrect. It should be noted that the dissociation of $\text{Cu}(\text{NO}_3)_2$ does not yield HNO_3 .

The accurate answer was that the solution should shift from basic to neutral due to the chemical reaction. $\text{Cu}(\text{OH})_2(\text{s})$ forms a precipitate and the K_{sp} ($\text{Cu}(\text{OH})_2$) at $25 ^\circ\text{C}$ is 1.6×10^{-19} , resulting in a negligible amount of OH^- ions. Consequently, the solution becomes neutral.

3.4. Difficulties Related to the Translation Issue

ChatGPT encountered challenges in the translation category, particularly when the problem involved an image. Table 7 presents a problem where ChatGPT’s response exhibited a difficulty associated with the translation category.

Table 7. ChatGPT's difficulty related to translation.

	<p>Here is a phase diagram of carbon dioxide.</p> 
<p>Problem</p>	<p>Please refer to the phase diagram of carbon dioxide and answer the following questions. (A) What do you think will happen if we heat solid carbon dioxide at a pressure of 0.2 Mpa from a temperature of -90°C to a temperature of -20°C while maintaining a constant pressure? (B) Another sample of solid carbon dioxide exists at initial conditions of 6 Mpa and -90°C; it is heated at constant pressure to -20°C. What will happen to this sample compared to the above sample?</p>
<p>Answer</p>	<p>To answer your questions, I need to refer to the phase diagram of carbon dioxide. However, as an AI text-based model, I do not have the ability to directly access or display images. I can provide you with a verbal description of the phase diagram and answer your questions based on the general understanding of carbon dioxide's phase behavior.</p> <p>.....</p> <p>.....</p> <p>(B) In the case of the second sample, starting with solid carbon dioxide at 6 MPa and -90°C and heating it at constant pressure to -20°C, the phase diagram suggests that carbon dioxide will undergo a phase transition from the solid phase to the gaseous phase, similar to the previous scenario. However, as the initial pressure is higher (6 MPa), the transition from solid to gas may occur at a slightly higher temperature compared to the previous sample. Nonetheless, both samples will eventually reach the gaseous phase at -20°C and their respective pressures.</p>

The last paragraph in the solution is not correct, as ChatGPT does not have the capability to read or process images. ChatGPT informed us that it is unable to perform image analyses due to its limitations. Therefore, ChatGPT faced difficulties in translating from a graphical representation to a symbolic representation.

3.5. Difficulties Related to the Transfer Issue

Table 8 presents a problem where ChatGPT's response exhibited a difficulty associated with the awareness of the transfer from the context of a chemical compound to that of an everyday compound such as chalk.

Table 8. ChatGPT's difficulty related to awareness of the transfer from a chemical to a chalk element.

Problem	When you put a piece of chalk in a glass vessel that contains water and blow with a straw into the water for half an hour, you see the formation of gas bubbles on the surface of the chalk and a small reduction in the volume of the piece of chalk. Formulate the chemical reaction that caused the phenomena and explain in your own words what happened.
Answer	<p>The phenomenon you described suggests that a chemical reaction occurred between the chalk and the water, resulting in the formation of gas bubbles and a reduction in the volume of the chalk. While I cannot provide an exact chemical equation without further information, I can explain what might have happened based on the given observations.</p> <p>Chalk, primarily composed of calcium carbonate (CaCO_3), is a common mineral found in sedimentary rocks and is widely used in classrooms as chalk sticks. When you blow air into water using a straw, you introduce carbon dioxide (CO_2) from your breath into the water. The water can dissolve some of the CO_2, forming carbonic acid (H_2CO_3), which is a weak acid.</p> <p>Calcium carbonate (chalk) can react with carbonic acid in the following reaction:</p> $\text{CaCO}_3 + \text{H}_2\text{CO}_3 \rightarrow \text{Ca}(\text{HCO}_3)_2$ <p>In this reaction, calcium carbonate reacts with carbonic acid to produce calcium bicarbonate ($\text{Ca}(\text{HCO}_3)_2$). This reaction is responsible for the observed reduction in the volume of the chalk.</p> <p>The formation of gas bubbles on the surface of the chalk indicates the release of carbon dioxide gas (CO_2), which is a byproduct of the reaction. The gas bubbles rise to the water's surface and escape into the atmosphere.</p> <p>Therefore, the overall process involves the dissolution of carbon dioxide in water, the formation of carbonic acid, and the subsequent reaction with calcium carbonate, leading to the formation of calcium bicarbonate, the release of carbon dioxide gas, and a decrease in the volume of the chalk.</p>

The above answer failed to mention the decomposition of the bicarbonate as the source of carbon dioxide released on the surface of the chalk. Instead of acknowledging this specific process, which would have shown ChatGPT's ability to transfer from the chemical compound context to the chalk context, the generative AI simply stated that carbon dioxide is a byproduct.

4. Discussion

Researchers have, since the advent of the computer, been interested in the potential of technology in various aspects of education [23–25]. Recently, educational researchers have shown increasing interest in the potential of AI, including ChatGPT, for teaching and learning [26–30]. The objective of this study was to examine the conceptual knowledge difficulties encountered by ChatGPT when addressing chemistry problems within a specific topic.

ChatGPT encountered no conceptual difficulties in solving remembering problems, which could have been due to the fact that the remembering problem entailed only returning to the content with which ChatGPT was supplied. It encountered depth difficulties of the understanding type when solving understanding, application, and analysis problems. ChatGPT encountered transfer difficulties when solving synthesis problems and it encountered translation difficulties when solving evaluation problems. The previous two types of difficulty may have been due the textual nature of ChatGPT as the two difficulties were conditioned by the ability to translate from one representation into another [31].

The findings revealed that although ChatGPT did not face conceptual difficulties in 13 out of 30 chemistry problems, it faced challenges across all five categories of conceptual knowledge, with a higher prevalence of difficulties observed in the depth category. According to Holme et al. [15], the depth category encompasses the ability to reason about chemistry concepts beyond simple memorization or algorithmic problem-solving. In our research, these challenges manifested as difficulties in understanding chemical rules, com-

prehending the nature of specific compounds, and grasping the reasons behind chemical phenomena. These characteristics can lead to an unclear understanding of the subject matter [32], highlighting the need to enhance the functioning of AI, including ChatGPT, to improve its capability to address scientific questions and minimize misconceptions.

Regarding representations, in addition to the argument above, ChatGPT, being text-based generative AI, can encounter difficulties in this aspect. The limitations stem from its inability to directly generate or display visual figures. To illustrate this, we engaged ChatGPT in a conversation about representations in chemistry. It mentioned various types of representations, including chemical formulas, structural formulas, ball-and-stick models, space-filling models, Lewis dot structures, electron configurations, reaction equations, and the periodic table. However, when we specifically requested a figure illustrating electron configurations, ChatGPT acknowledged its limitations and apologized for its inability to generate or display visual figures. Nevertheless, it offered to describe the concept and provide an example. For instance, the electron configuration of oxygen (O) is represented as $1s^2 2s^2 2p^4$.

Upon comparing this textual representation with the one in Figure 3, it became evident that ChatGPT would benefit from advancing beyond a purely text-based approach, incorporating symbol and figure-based capabilities to enhance its understanding and communication of chemistry concepts.

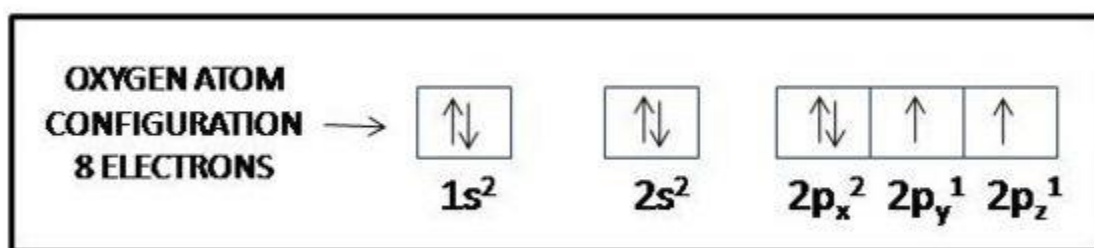


Figure 3. The electron configuration for oxygen.

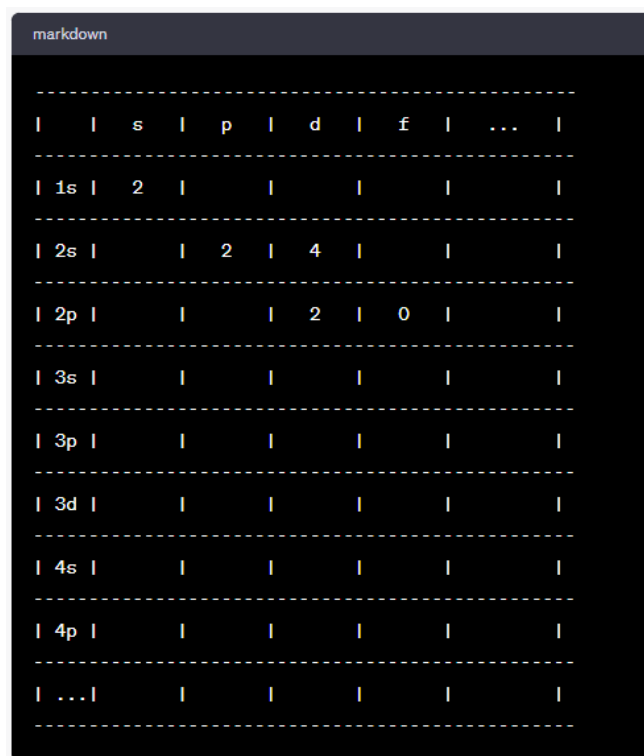
When we requested ChatGPT to provide a figure illustrating the electron configuration of oxygen once again, it presented the representation shown in Figure 4. This indicated that ChatGPT had the potential to generate visual representations and had improved its capabilities in this regard.

When prompted to provide contexts for using the multi-line representation of electron configurations, ChatGPT identified three scenarios: visualizing energy levels, comparing electron configurations, and identifying valence electrons.

To address the limitations of ChatGPT's text-based capabilities, researchers have proposed the use of converters that can transform text into graphical representations, as suggested by Jiang et al. [33].

In previous studies, there have been mixed findings regarding ChatGPT's performance as a learner. Huh [34] demonstrated that ChatGPT's performance in parasitology fell short compared with that of a Korean student, while Juhi et al. [35] found that ChatGPT's predictions and explanations for drug–drug interactions were only partially accurate. On the other hand, Kung et al. conducted a study in which ChatGPT successfully completed the United States Medical Licensing Examination without human assistance.

Although the present study did not directly compare ChatGPT's performance with that of college students in solving chemistry problems, it identified the conceptual knowledge difficulties that ChatGPT encountered when answering these problems. These difficulties could be attributed to ChatGPT's nature as generative AI as it had not previously been exposed to the specific chemistry problems presented. The future development of ChatGPT is expected to enhance its ability to handle new scientific problems, including those in the field of chemistry.



The screenshot shows a table representing the electron configuration of oxygen. The table has columns for orbitals (s, p, d, f, ...) and rows for principal quantum numbers (1s, 2s, 2p, 3s, 3p, 3d, 4s, 4p, ...). The electron counts are: 1s: 2, 2s: 2, 2p: 4, 3s: 0, 3p: 0, 3d: 0, 4s: 0, 4p: 0, ...: 0.

	s	p	d	f	...
1s	2				
2s	2	4			
2p		2	0		
3s					
3p					
3d					
4s					
4p					
...					

Figure 4. Electron configuration of oxygen provided by ChatGPT.

5. Conclusions

ChatGPT was requested to solve 30 chemistry problems, where it solved 13 out of 30 problems (43.33% of the problems) without showing conceptual difficulties. This showed that ChatGPT could be a promising companion of the learner in solving chemistry problems. Thus, it is recommended that ChatGPT be integrated into the chemistry classroom in order to support students' learning.

ChatGPT was also requested to answer questions regarding its ability in issues related to conceptual understanding in chemistry, which revealed that it could work with representations but not with sound ones. New generative AI has the potential to work with representations, which could lead to deeper knowledge of various disciplines, including chemistry.

The aim of this research was to examine the conceptual understanding challenges faced by ChatGPT when responding to chemistry questions on the topic of Introduction to Material Science. The findings revealed that ChatGPT encountered difficulties across all aspects of conceptual understanding, with particular challenges observed in the depth category and the translation category. As ChatGPT operates as text-based generative AI, it is recommended that the conversion of text models into pictorial or graphical models is explored using tools such as Graphologue, as suggested by Jiang et al. [32]. Implementing such recommendations would empower learners in the era of generative AI [36].

The difficulties encountered by ChatGPT in comprehending and responding to chemistry problems within the Introduction to Material Science domain highlight the limitations of AI tools in complex subject areas. Whilst ChatGPT shows promise as a teaching and learning resource, improvements are needed to enhance its conceptual understanding and analytical abilities, especially in representations and problem-solving skills. More advanced algorithms and training data encompassing a wider range of chemistry problems may be necessary to address these limitations.

This research sheds light on the conceptual understanding difficulties ChatGPT faced in the chemistry discipline, specifically in an Introduction to Material Science course. The tool encountered challenges in comprehending complex representations and generating in-depth, explainable solutions. These findings emphasize the need for ongoing research

and development to refine AI tools like ChatGPT and enhance their capabilities to maximize their potential as adequate educational resources in chemistry education.

In summary, our findings revealed that ChatGPT encountered significant conceptual knowledge difficulties across various categories, with a particular emphasis on representations and depth of understanding. Difficulties in representations hindered the effective transfer of knowledge. ChatGPT struggled to comprehend and interpret complex chemical structures or formulas, resulting in inaccurate or incomplete solutions. This limitation was most noticeable when problems required a high level of depth and critical thinking, suggesting a need for further improvements in ChatGPT's analytical capabilities. Furthermore, ChatGPT faced challenges in predicting and explaining the answers it provided. The tool often generated solutions that needed clear explanations or reasoning, making it difficult for users to understand the underlying principles behind the solutions. This limitation may hinder effective learning and its potential as an educational resource. Developers of generative AI could benefit from carefully reading epistemic studies on the role of transformers and knowledge, which could help them to develop more complex generative AI tools that address fewer conceptual understanding difficulties.

Author Contributions: Conceptualization, A.R.; methodology, W.D.; software, A.R.; formal analysis, W.D., H.D. and A.R.; investigation, H.D. and A.R.; data curation, A.R.; writing—original draft preparation, W.D., H.D. and A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gocen, A.; Aydemir, F. Artificial Intelligence in Education and Schools. *Res. Educ. Media* **2020**, *12*, 13–21. [\[CrossRef\]](#)
2. Dai, Y.; Liu, A.; Lim, C.P. Reconceptualizing ChatGPT and generative AI as a student-driven innovation in higher education. In Proceedings of the 33rd CIRP Design Conference, Sydney, Australia, 17–19 May 2023. [\[CrossRef\]](#)
3. The Centre for Learning, Teaching, and Development. *ChatGPT for Learning and Teaching*; University of the Witwatersrand: Johannesburg, South Africa, 2023.
4. West, C.G. AI and the FCI: Can ChatGPT project an understanding of introductory physics? *arXiv* **2023**, arXiv:2303.01067.
5. Dwivedi, Y.K.; Kshetri, N.; Hughes, L.; Slade, E.L.; Jeyaraj, A.; Kar, A.K.; Wright, R. “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *Int. J. Inf. Manag.* **2023**, *71*, 102642. [\[CrossRef\]](#)
6. Aljanabi, M.; Ghazi, M.; Ali, A.H.; Abed, S.A. ChatGpt: Open Possibilities. *Iraqi J. Comput. Sci. Math.* **2023**, *4*, 62–64. [\[CrossRef\]](#)
7. Firat, M. What ChatGPT means for universities: Perceptions of scholars and students. *J. Appl. Learn. Teach.* **2023**, *6*, 1. [\[CrossRef\]](#)
8. Adiguzel, T.; Kaya, M.H.; Cansu, F.K. Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *Contemp. Educ. Technol.* **2023**, *15*, ep429. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Muñoz, S.A.; Gayoso, G.G.; Huambo, A.C.; Tapia, R.D.; Incaluque, J.L.; Aguila, O.E.; Cajamarca, J.C.; Acevedo, J.E.; Rivera, H.V.; Arias-González, J.L. Examining the Impacts of ChatGPT on Student Motivation and Engagement. *Soc. Space* **2023**, *23*, 1–27.
10. Orrù, G.; Piarulli, A.; Conversano, C.; Gemignani, A. Human-like problem-solving abilities in large language models using ChatGPT. *Front. Artif. Intell.* **2023**, *6*, 1199350. [\[CrossRef\]](#)
11. Cracolice, M.S.; Deming, J.C.; Ehlert, B. Concept Learning versus Problem Solving: A Cognitive Difference. *J. Chem. Educ.* **2008**, *85*, 873. [\[CrossRef\]](#)
12. Roth, K.J. Developing Meaningful Conceptual Understanding in Science. In *Dimensions of Thinking and Cognitive Instruction*; Jones, B.F., Idol, L., Eds.; Lawrence Erlbaum Associates: Hillsdale, NJ, USA, 1990; pp. 139–175.
13. Puk, T.; Stibbards, A. Growth in Ecological Concept Development and Conceptual Understanding in Teacher Education: The Discerning Teacher. *Int. J. Environ. Sci. Educ.* **2011**, *6*, 191–211.
14. Lansangan, R.V.; Orleans, A.V.; Camacho, V.M.I. Assessing conceptual understanding in chemistry using representation. *Adv. Sci. Lett.* **2018**, *24*, 7930–7934. [\[CrossRef\]](#)
15. Holme, T.A.; Luxford, C.J.; Brandriet, A. Defining Conceptual Understanding in General Chemistry. *J. Chem. Educ.* **2015**, *92*, 1477–1483. [\[CrossRef\]](#)
16. Prasad, G.N.R. Evaluating student performance based on bloom's taxonomy levels. *J. Phys. Conf. Ser.* **2021**, *1797*, 012063. [\[CrossRef\]](#)
17. Assaly, I.R.; Smadi, O.M. Using Bloom's Taxonomy to Evaluate the Cognitive Levels of Master Class Textbook's Questions. *Engl. Lang. Teach.* **2015**, *8*, 100. [\[CrossRef\]](#)

18. Daher, W.; Sleem, H. Middle School Students' Learning of Social Studies in the Video and 360-Degree Videos Contexts. *IEEE Access* **2021**, *9*, 78774–78783. [CrossRef]
19. Cooper, G. Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence. *J. Sci. Educ. Technol.* **2023**, *32*, 444–452. [CrossRef]
20. Leon, A.; Vidhani, D. ChatGPT Needs a Chemistry Tutor Too. 2023. Available online: <https://chemrxiv.org/engage/api-gateway/chemrxiv/assets/orp/resource/item/642f2351a41dec1a5699bf9f/original/chat-gpt-needs-a-chemistry-tutor-too.pdf> (accessed on 1 July 2023).
21. Daher, W. Saturation in Qualitative Educational Technology Research. *Educ. Sci.* **2023**, *13*, 98. [CrossRef]
22. Daher, W.; Ashour, W.; Hamdan, R. The Role of ICT Centers in the Management of Distance Education in Palestinian Universities during Emergency Education. *Educ. Sci.* **2022**, *12*, 542. [CrossRef]
23. Daher, W.; Baya'a, N.; Jaber, O.; Shahbari, J.A. A Trajectory for Advancing the Meta-Cognitive Solving of Mathematics-Based Programming Problems with Scratch. *Symmetry* **2020**, *12*, 1627. [CrossRef]
24. Daher, W.; Mokh, A.A.; Shayeb, S.; Jaber, R.; Saqer, K.; Dawood, I.; Bsharat, M.; Rabbaa, M. The Design of Tasks to Suit Distance Learning in Emergency Education. *Sustainability* **2022**, *14*, 1070. [CrossRef]
25. Abuzant, M.; Ghanem, M.; Abd-Rabo, A.; Daher, W. Quality of Using Google Classroom to Support the Learning Processes in the Automation and Programming Course. *Int. J. Emerg. Technol. Learn. (ijET)* **2021**, *16*, 72–87. [CrossRef]
26. Toukiloglou, P.; Xinogalos, S. A Systematic Literature Review on Adaptive Supports in Serious Games for Programming. *Information* **2023**, *14*, 277. [CrossRef]
27. Munir, H.; Vogel, B.; Jacobsson, A. Artificial Intelligence and Machine Learning Approaches in Digital Education: A Systematic Revision. *Information* **2022**, *13*, 203. [CrossRef]
28. Lamerias, P.; Arnab, S. Power to the Teachers: An Exploratory Review on Artificial Intelligence in Education. *Information* **2022**, *13*, 14. [CrossRef]
29. Raschka, S.; Patterson, J.; Nolet, C. Machine Learning in Python: Main Developments and Technology Trends in Data Science, Machine Learning, and Artificial Intelligence. *Information* **2020**, *11*, 193. [CrossRef]
30. How, M.-L.; Cheah, S.-M.; Chan, Y.-J.; Khor, A.C.; Say, E.M.P. Artificial Intelligence-Enhanced Decision Support for Informing Global Sustainable Development: A Human-Centric AI-Thinking Approach. *Information* **2020**, *11*, 39. [CrossRef]
31. Rahmawati, D.; Purwanto, P.; Subanji, S.; Hidayanto, E.; Anwar, R.B. Process of Mathematical Representation Translation from Verbal into Graphic. *Int. Electron. J. Math. Educ.* **2017**, *12*, 367–381. [CrossRef] [PubMed]
32. Goris, T.V.; Dyrenfurth, M.J. How Electrical Engineering Technology Students Understand Concepts of Electricity. Comparison of misconceptions of freshmen, sophomores, and seniors. In Proceedings of the 2013 ASEE Annual Conference & Exposition, Atlanta, GA, USA, 23–26 June 2013; pp. 23–668. [CrossRef]
33. Jiang, P.; Rayan, J.; Dow, S.P.; Xia, H. Graphologue: Exploring Large Language Model Responses with Interactive Diagrams. *arXiv* **2023**, arXiv:2305.11473.
34. Huh, S. Are ChatGPT's knowledge and interpretation ability comparable to those of medical students in Korea for taking a parasitology examination?: A descriptive study. *J. Educ. Evaluation Health Prof.* **2023**, *20*, 1. [CrossRef]
35. Juhi, A.; Pipil, N.; Santra, S.; Mondal, S.; Behera, J.K.; Mondal, H.; Behera, J.K., IV. The capability of ChatGPT in predicting and explaining common drug-drug interactions. *Cureus* **2023**, *15*, e36272. [CrossRef]
36. Gašević, D.; Siemens, G.; Sadiq, S. Empowering learners for the age of artificial intelligence. *Comput. Educ. Artif. Intell.* **2023**, *4*, 100130. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.