geotechnics

MDPI

Article

An Investigation into the Utility of Large Language Models in Geotechnical Education and Problem Solving

Liuxin Chen [1], Amir Tophel [1,*], Umidu Hettiyadura [2] and Jayantha Kodikara [1]

[1] ARC Industrial Transformation Research Hub (ITRH)—SPARC Hub, Department of Civil Engineering, Clayton Campus, Monash University, Clayton, VIC 3800, Australia; liuxin.chen@monash.edu (L.C.); jayantha.kodikara@monash.edu (J.K.)

[2] Department of Civil Engineering, Clayton Campus, Monash University, Clayton, VIC 3800, Australia; ufer0001@student.monash.edu

* Correspondence: amir.tophel@monash.edu

Abstract: The study explores the capabilities of large language models (LLMs), particularly GPT-4, in understanding and solving geotechnical problems, a specialised area that has not been extensively examined in previous research. Employing a question bank obtained from a commonly used textbook in geotechnical engineering, the research assesses GPT-4's performance across various topics and cognitive complexity levels, utilising different prompting strategies like zero-shot learning, chain-of-thought (CoT) prompting, and custom instructional prompting. The study reveals that while GPT-4 demonstrates significant potential in addressing fundamental geotechnical concepts and problems, its effectiveness varies with specific topics, the complexity of the task, and the prompting strategies employed. The paper categorises errors encountered by GPT-4 into conceptual, grounding, calculation, and model inherent deficiencies related to the interpretation of visual information. Custom instructional prompts, specifically tailored to address GPT-4's shortcomings, significantly enhance its performance. The study reveals that GPT-4 achieved an overall problem-solving accuracy of 67% with custom instructional prompting, significantly higher than the 28.9% with zero-shot learning and 34% with CoT. However, the study underscores the importance of human oversight in interpreting and verifying GPT-4's outputs, especially in complex, higher-order cognitive tasks. The findings contribute to understanding the potential and limitations of current LLMs in specialised educational fields, providing insights for educators and researchers in integrating AI tools like GPT-4 into their teaching and problem-solving approaches. The study advocates for a balanced integration of AI in education to enrich educational delivery and experience while emphasising the indispensable role of human expertise alongside technological advancements.

Keywords: data science applications in education; human–computer interface; technology-enhanced learning; cooperative/collaborative learning

check for updates

Citation: Chen, L.; Tophel, A.; Hettiyadura, U.; Kodikara, J. An Investigation into the Utility of Large Language Models in Geotechnical Education and Problem Solving. *Geotechnics* **2024**, *4*, 470–498. https://doi.org/10.3390/geotechnics4020026

Academic Editors: Md Rajibul Karim, Md. Mizanur Rahman and Khoi Nguyen

Received: 28 March 2024
Revised: 28 April 2024
Accepted: 7 May 2024
Published: 9 May 2024

1. Introduction

The educational landscape is experiencing a profound transformation, propelled by the rapid advancement and integration of artificial intelligence (AI) technologies [1–4]. This shift is prominently marked by the emergence of Generative AI as a pivotal force in reshaping educational paradigms, offering innovative approaches to learning and problem-solving [5–7]. Central to this AI-driven revolution in education are generative AI models such as large language models (LLMs) like ChatGPT [8–13], which have demonstrated considerable potential in enhancing learning outcomes [14–16] and aiding in complex problem-solving scenarios. These models, powered by sophisticated algorithms and extensive data, can alter information delivery and fundamentally change how students interact with and comprehend complex topics [17].

The growing incorporation of AI into education, especially in STEM (science, technology, engineering, and math) fields, is underpinned by a burgeoning body of research [18].

Bai and Stede [19] conducted a survey on modern machine learning (ML) methods for automated assessment of students' natural language free-text responses. This study accentuates the growing adoption of AI in educational contexts and emphasises its transformative role in revolutionising traditional learning approaches. A notable area where AI has made significant inroads is in exam preparation and assessment. ChatGPT has been a focal point in various studies of AI applications in education, showcasing its effectiveness across multiple educational domains [20–24]. For instance, Schulze Balhorn et al. [25] systematically evaluated GPT-3.5's ability to answer questions in the natural sciences and engineering using 594 questions from 198 Delft University faculty members. Using a structured assessment, participants generally rated the answers as "mostly correct." However, ratings declined with higher educational levels of questions and when assessing skills beyond scientific knowledge, like critical thinking.

Zhang et al. [26] evaluated the programming capabilities of GPT-3.5 and GPT-4 using Swift-based exam questions from a third-year university course. Their study finds that both models generally outperform average student scores but do not consistently surpass top students. This comparison highlights areas where the GPT models excel and fall short, providing a nuanced view of their current programming proficiency. The study also reveals surprising instances where GPT-3.5 outperforms GPT-4, suggesting complex variations in AI model capabilities.

Katz et al. [27] assessed GPT-4's zero-shot performance on the entire Uniform Bar Examination (UBE), including the Multistate Bar Examination (MBE), the Multistate Essay Exam (MEE), and the Multistate Performance Test (MPT). GPT-4 significantly outperformed both humans and previous models on the MBE, achieving a 26% improvement over GPT-3.5 and higher scores in five out of seven areas. On the MEE and MPT, GPT-4 scored an average of 4.2/6.0, notably surpassing GPT-3.5. GPT-4 scored 297 points across the UBE, exceeding the passing threshold for all jurisdictions and demonstrating its potential to enhance legal service delivery.

Maitland et al. [28] evaluated GPT-4's performance on the MRCP (Membership of the Royal College of Physicians) Parts 1 and 2 practice questions. The study found 86.3% and 70.3% accuracy rates for Parts 1 and 2, respectively. The analysis identified eight error types, with factual, context, and omission errors being the most common. Overall, GPT-4 significantly outperformed the passing thresholds, offering insights into the effectiveness of GPT-4 in answering multiple-choice medical exams. Currie et al. [29] analysed GPT-3.5's effects on academic integrity and its use in medical imaging courses, testing it in exams and written assignments across six subjects in the medical radiation science undergraduate program. Evaluations using standardised rubrics and Turnitin showed that GPT-3.5, generally underperformed compared to students, particularly in advanced subjects, but excelled in basic exams. The study concluded that while GPT-3.5 can enhance learning for more straightforward tasks, it risks compromising academic integrity and is limited in handling complex, discipline-specific content. In addition, the study recommends cautious integration of AI tools to maintain academic standards. Ali et al. [30] evaluated GPT-3.5 and GPT-4 on a 500-question mock neurosurgical medical exam. The study finds both models surpass the passing threshold, with GPT-4 significantly outperforming GPT-3.5, scoring 83.4% compared to GPT-3.5's 73.4%. Both models showed lower performance on image-based questions, relying only on text clues. Overall, the study demonstrates the potential of LLMs in achieving passing scores on specialised medical exams, especially GPT-4's enhanced capabilities. Kung et al. [31] assessed GPT-3.5's capabilities in the exams of the United States Medical Licensing Examination (USMLE). This study also demonstrated the ability of GPT-3.5 to reach the passing threshold without needing specialised training, indicating its potential in medical education. Further corroborating these findings, Gilson et al. [32] and Antaki et al. [33] reported positive outcomes in applying ChatGPT within medical education, suggesting its significant role in this specialised field.

In business education, Terwiesch [34] investigated GPT-3.5's effectiveness in the final exam of a typical MBA (i.e., Master of Business Administration) core course, operations management. The study revealed that GPT-3.5 showed competence in essential operations management and process analysis questions, including those based on case studies. However, it made significant errors in simple calculations at a 6th-grade math level and struggled with advanced process analysis questions involving multiple products and stochastic effects. Despite these limitations, GPT-3.5 exhibited a notable ability to adapt its responses based on human hints, underscoring its potential as a valuable educational resource in business school settings. Eulerich et al. [35] expanded the scope of ChatGPT's evaluation to accounting certification exams. This study evaluated the capability of ChatGPT models to pass major accounting certification exams. Their findings suggest that while the initial GPT-3.5 scored an average of 53.1% and failed the exams, significant improvements were observed with the GPT-4. Enhancements, including additional training and using reasoning tools like calculators, raised the average score to 85.1%, allowing ChatGPT to pass all examined certifications.

However, despite these promising results, there is a noticeable gap in the literature regarding ChatGPT's performance in geotechnical engineering, an area explored by only a few researchers [36,37]. Fatahia et al. [38] discussed generative AI's role in geotechnical engineering, noting its potential and the need to explore tools like ChatGPT further in geotechnical engineering. This gap in the literature indicates the necessity for comprehensive research into ChatGPT's capabilities in geotechnical education and problem solving. Geotechnical engineering is inherently dependent on complex mathematical concepts and sophisticated problem-solving abilities, posing unique challenges yet to be thoroughly tested against the capabilities of ChatGPT and similar AI tools. A comprehensive investigation into ChatGPT's potential in geotechnical engineering is essential. Such an exploration would reveal both the current capabilities and limitations of ChatGPT in addressing complex geotechnical problems, setting the stage for future applications, such as:

1. Interactive tutoring where LLMs offer personalised instruction to enhance student understanding of complex topics, such as soil mechanics and foundation design.
2. Virtual lab developing where LLMs are combined with virtual reality. Students can perform virtual soil tests and experiments, with real-time guidance and results interpretation from the LLM.

This paper aims to fill the existing research gap by evaluating the depth and scope of GPT-4's capabilities in geotechnical engineering. Our study delves into the utility of GPT-4 for a diverse range of cohorts, including students, educators, researchers, AI developers, and practitioners, focusing on its proficiency in explaining concepts, answering questions, and crafting educational content. We rigorously scrutinise GPT-4's ability to handle fundamental undergraduate geotechnical topics and assess its performance in tackling problems of escalating complexity.

A critical part of our research is identifying major error types that hinder GPT-4's effectiveness in geotechnical contexts. We have categorised these errors into four groups: "Conceptual" errors arising from the inability to retrieve necessary concepts or facts; "Grounding" errors where retrieved concepts are incorrectly applied in equations or constraints; "Calculation" errors involving errors in algebraic and arithmetic manipulation; and "Deficiency", which include challenges in interpreting images, graphs, charts, and in drawing engineering images. Additionally, our study explores various prompting techniques such as zero-shot, chain-of-thought (CoT) processes, and tailored custom instructional prompts to optimise GPT-4's output.

We aim to offer a comprehensive overview of LLMs' role in geotechnical education and problem-solving by investigating these aspects. This research is structured around key questions that assess GPT-4's capabilities and limitations, providing a holistic view of its utility across different user groups in geotechnical engineering. We utilised the April 2023 version of GPT-4 for this purpose. Ultimately, we seek to contribute to the responsible and effective integration of AI models like GPT-4 into educational settings. Our findings are intended to provide valuable insights that could inform future teaching methodologies and guide generative AI development in geotechnical engineering. This approach involves leveraging these advanced AI tools while continuously refining and developing them to realise their full potential as invaluable educational resources.

2. Dataset

2.1. Data Selection

To ensure statistical validity in assessing the utility of GPT-4 in fundamental or undergraduate early-level geotechnical education and problem solving, we first calculated the requisite sample size according to the formula provided by Daniel and Cross [39]:

$$n = \frac{Z^2 P(1-P)}{d^2} \tag{1}$$

where n denotes the sample size; Z is the Z-score correlating to the chosen confidence level; P represents the anticipated prevalence or proportion; and d signifies the acceptable margin of error. For the present study, a 95% confidence level was chosen, corresponding to a Z-score of approximately 1.96, and a precision (margin of error) of 5%. Following the work of Lwanga and Lemeshow [40], setting P at 0.5 is recommended to obtain a sample size that is sufficiently large to account for variability. This approach dictates a sample question size of 385, which is conducive to a robust statistical analysis. It should be noted that by adopting the above formula, the following assumptions have been made:

1. A large population from which the sample is drawn, and the sample itself represents a minor fraction (less than 5%) of the total population;
2. The expected outcome has been simplified to a binary variable—the success or failure of the LLM in problem-solving tasks—which aligns with the typical dichotomous nature of the assessment outcomes;
3. The proportion's sampling distribution is presumed to be normal or approximately normal. This assumption is reasonable for the calculated sample size, invoking the central limit theorem to justify the normal approximation.

In alignment with the calculated sample size requirements, we compiled a diverse question bank comprising 391 questions from Das's authoritative textbook "Principles of Geotechnical Engineering, 8th Edition" in geotechnical engineering. The answers to these questions are obtained from the solution manual.

2.2. Data Categorisation

To effectively assess the capabilities of the LLM in geotechnical engineering education, we categorised our question bank along with topic relevance and cognitive complexity. These classifications allow for a comprehensive analysis of the model's performance across the breadth of geotechnical topics and the depth of problem-solving skills required.

2.2.1. Categorisation by Topic

Our question bank is structured around core topics essential to undergraduate geotechnical engineering, aligning with Chapters 2–17 from the above textbook. These chapters encompass a comprehensive range of topics, including soil origin, grain size analysis, weight–volume relationships, soil plasticity, classification, compaction, seepage, permeability, in situ stresses, compressibility, shear strength, and lateral earth pressures, as listed Table 1. Figure 1 delineates a comparative analysis of the prevalence of image-based and text-based questions across the sixteen chapters, labelled C2 through C17. This categorisation enables a targeted analysis of the LLM's proficiency in diverse geotechnical topics.

Table 1. Overview of Fundamental Geotechnical Engineering Topics Covered in the Question Bank.

Chapter Code	Topic Description
C2	Origin of Soil and Grain Size
C3	Weight–Volume Relationships
C4	Plasticity and Structure of Soil
C5	Classification of Soil
C6	Soil Compaction
C7	Permeability
C8	Seepage
C9	In Situ Stresses
C10	Stresses in a Soil Mass
C11	Compressibility of Soil
C12	Shear Strength of Soil
C13	Lateral Earth Pressure: At-Rest Rankine and Coulomb
C14	Lateral Earth Pressure: Curved Failure Surface
C15	Slope Stability
C16	Soil Bearing Capacity for Shallow Foundations
C17	Subsoil Exploration

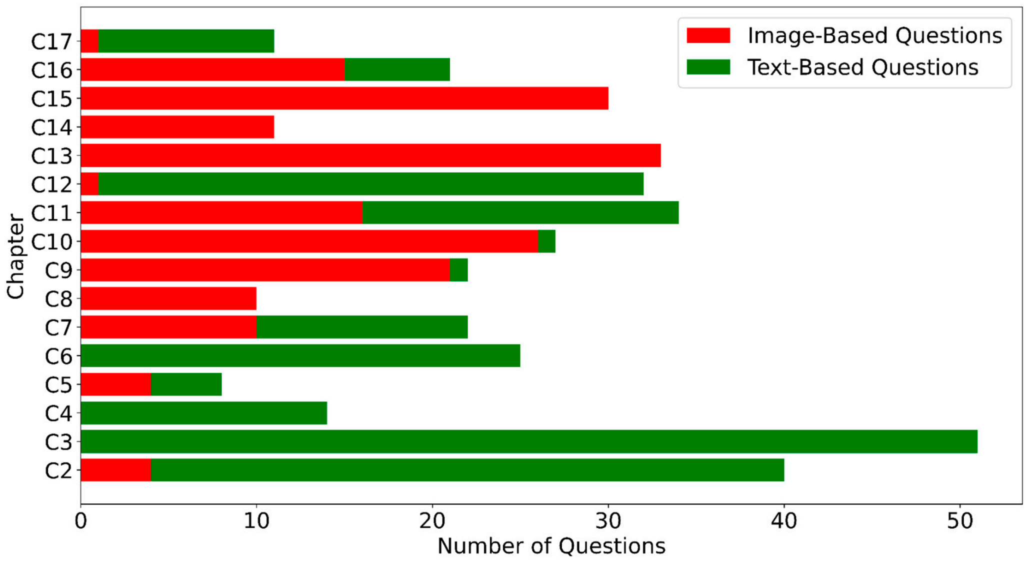


Figure 1. Topic-Wise Distribution of Our Question Bank.

2.2.2. Categorisation by Cognitive Complexity

Bloom's taxonomy, a framework consisting of categorised educational goals, helps educators define and distinguish between different levels of human cognition in learning environments [41]. Employing the updated Bloom's taxonomy proposed by Krathwohl [41], we classify questions into six levels of cognitive complexity: "Remember", "Understand", "Apply", "Analyse", "Evaluate", and "Create". Figure 2 illustrates this hierarchy and provides specific examples for each level. For instance, Level 1: "Remember" is exemplified by tasks such as 'Recalling and describing common laboratory tests in soil mechanics'. This systematisation aids in gauging the LLM's proficiency in addressing geotechnical challenges of varying difficulty. Figure 3 represents the typical distribution of cognitive complexity levels in an undergraduate setting. It shows that most tasks (52%) are at the "Apply" level, aligning with undergraduate studies' practical, application-focused nature. Lower proportions in the "Create" (1%) and "Evaluate" (4%) categories reflect the gradual development of higher-order thinking skills, often more emphasised in higher-level undergraduate and postgraduate education.

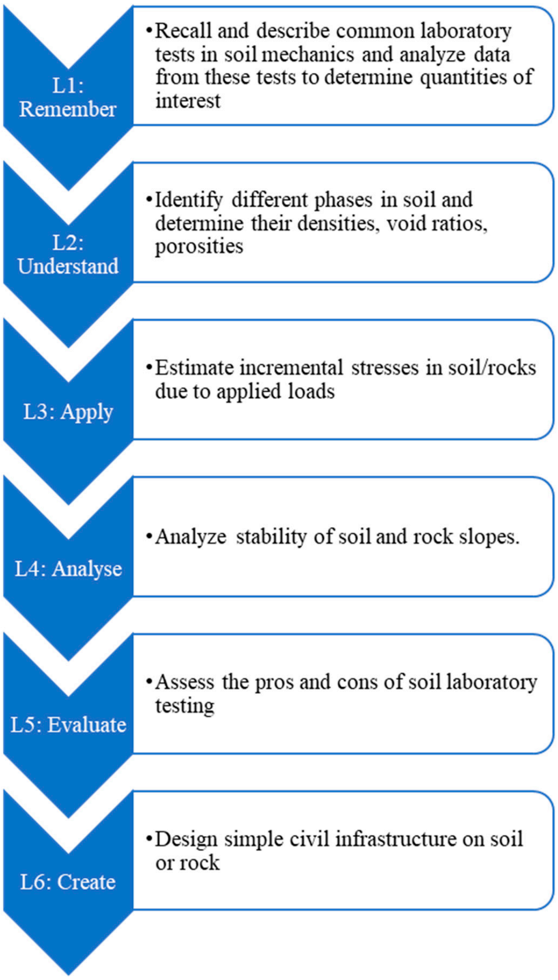


Figure 2. The Difficulty Level of the Questions in Our Data Bank According to the Revised Bloom's Taxonomy Proposed by Krathwohl [41].

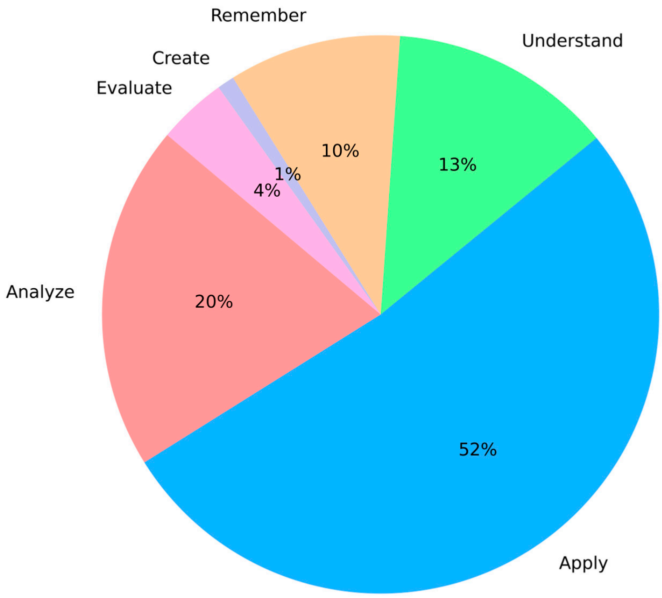


Figure 3. The Difficulty Level of The Questions in Our Data Bank According to The Revised Bloom's Taxonomy Proposed by Krathwohl [41].

3. Experimental Setup

This investigation employs GPT-4 [8], a state-of-the-art LLM, to interrogate our question bank. We have implemented three strategic prompting strategies to evaluate the model's problem-solving prowess: zero-shot learning, CoT prompting, and custom instructional prompting:

1. Zero-shot learning: This approach assesses GPT-4's baseline problem-solving skills without prior exposure to specific examples. It evaluates the model's ability to leverage pre-trained knowledge and innate reasoning capabilities.
2. CoT: In this approach, we prompted GPT-4 to detail its reasoning steps, akin to a human's approach to problem-solving. This approach aids in understanding the model's thought process and checks for logical coherence.
3. Custom instructional prompting: This technique proactively mitigates common errors and refines GPT-4's responses. Initially, we conducted a thorough analysis to identify the most frequent and impactful errors that GPT-4 encounters when tackling geotechnical problems. Armed with this insight, we crafted precise prompts containing targeted instructions addressing these shortcomings. These custom prompts were then used to guide GPT-4's problem-solving process, optimising its accuracy and relevance in response to geotechnical questions.

Each prompting strategy is designed to progressively improve GPT-4's problem-solving accuracy in geotechnical engineering. The Zero Shot strategy presents questions in their original form, testing the model's baseline capabilities without prior specific training. This strategy evaluates the model's ability to utilise pre-trained knowledge in new problems.

In the CoT approach, questions were again presented in their original form but with an additional directive to "solve the question and do it step by step". This approach prompted GPT-4 to explicitly outline its reasoning process, offering insight into its logical progression and facilitating the identification of any reasoning gaps or errors.

For custom instructional prompting, we first analysed GPT-4's performance in the Zero Shot approach to identify areas of improvement. Understanding these errors allowed us to create custom instructions, such as specific formulas or conceptual guidance, to address these identified issues directly. The examples of these custom instructions are detailed in the Appendix A. The custom instruction strategy and CoT were applied selectively, only for questions where the Zero Shot approach was insufficient. This approach represents a more targeted and refined strategy to overcome specific errors and enhance the model's response quality in complex geotechnical scenarios.

4. Results

4.1. Accuracy and Effectiveness of Prompting Strategies

Figure 4 compares accuracy rates among three prompting strategies: Zero Shot, CoT, and our proposed strategy, Custom Instruction. The accuracy rate of each strategy is marked on the figure, with Zero Shot achieving a 28.9% accuracy and CoT 34%, while Custom Instruction notably leads with an impressive 67% accuracy. This marked difference is represented by the linear progression of points, underscoring the superior effectiveness of Custom Instruction in answering our question bank.

The subsequent subsections will delve deeper into these strategies, offering a comprehensive comparison across various topics and difficulty levels as per Bloom's taxonomy and their performance in text-based versus image-based questions.

4.1.1. Different Topics

Figure 5 illustrates the accuracy rates of three prompting strategies—Zero Shot (blue), CoT (green), and Custom Instruction (red)—across chapters C2 to C17. Each chapter is represented by a trio of bars showing the performance of each strategy. The figure shows the variability of the strategies' success across chapters: while some chapters (e.g., C4) show high accuracy for all strategies, others (e.g., C8, C10) demonstrate low effectiveness across the board. This visualisation underlines the importance of strategy selection concerning topic-specific demand. In addition, it can be seen that Custom Instruction consistently outperforms other strategies, indicative of its superior effectiveness. Specifically, chapters C3, C4, C12, and C16 exhibit the pronounced advantage of Custom Instruction where it reaches a 100% correct rate, starkly contrasting with the other strategies, which show significantly lower rates in these chapters. In Chapter C2, Custom Instruction demonstrates a superior outcome with approximately 85% correctness, while Zero Shot and CoT hover around 35% and 37.5%, respectively. These results indicate that Custom Instruction substantially enhances correct response rates in specific chapters, potentially indicating that the content of these chapters may be more amenable to custom-tailored instructional strategies.

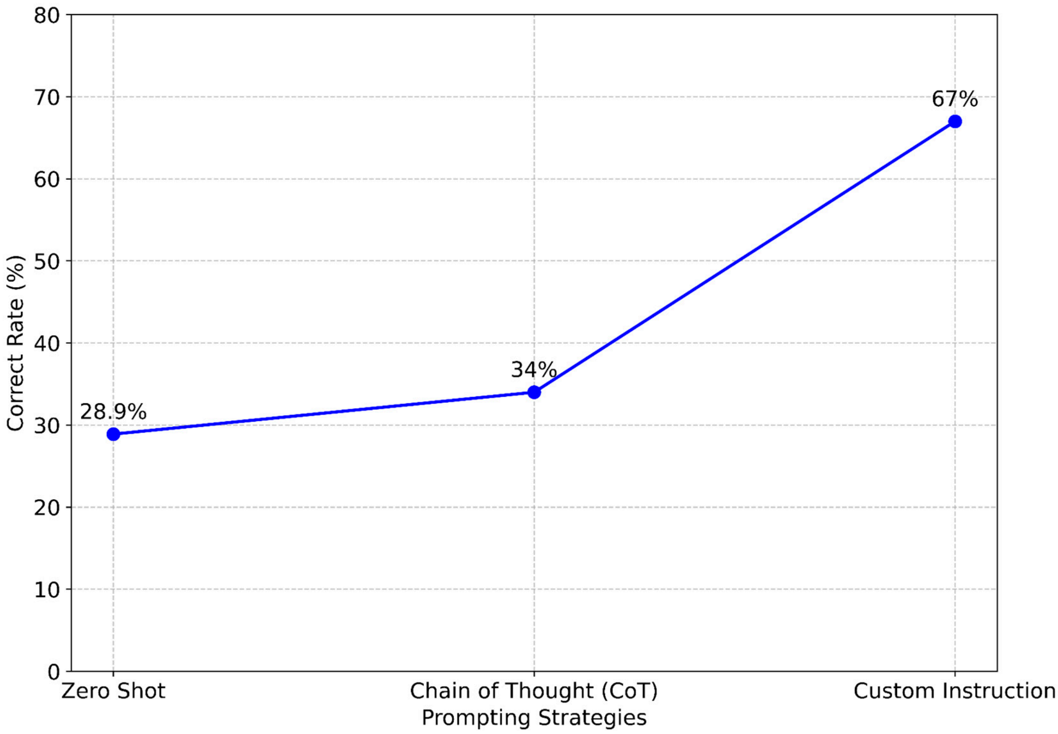


Figure 4. The Overall Accuracy of Each Prompting Strategy with Various Topics.

4.1.2. Difficulty Level According to Bloom's Taxonomy

Figure 6 illustrates the overall accuracy of responses categorised by Bloom's taxonomy levels using Zero Shot, CoT, and Custom Instruction. The accuracy for both the "Remember" and "Understand" levels is consistently high across all prompting strategies, standing at 100%. Notably, there is a marked increase in accuracy for the "Apply", "Analyse", and "Evaluate" levels when using Custom Instruction, with accuracy rates of 68.5%, 48.4%, and 50%, respectively, compared to the other strategies. While CoT performs marginally better than Zero Shot in the "Apply" and "Analyse" levels, Custom Instruction markedly outpaces both. At the higher cognitive levels of "Evaluate" and "Create", CoT does not enhance GPT-4's responses, yet Custom Instruction significantly elevates its responses. Remarkably, for the "Create" level, Custom Instruction achieves a 100% accuracy rate. In comparison, Zero Shot and CoT yield a 0% accuracy rate, underscoring the substantial impact of Custom Instruction at the highest level of cognitive demand in Bloom's taxonomy. This graphical analysis highlights the superior performance of Custom Instruction in facilitating higher-order thinking skills in GPT-4.

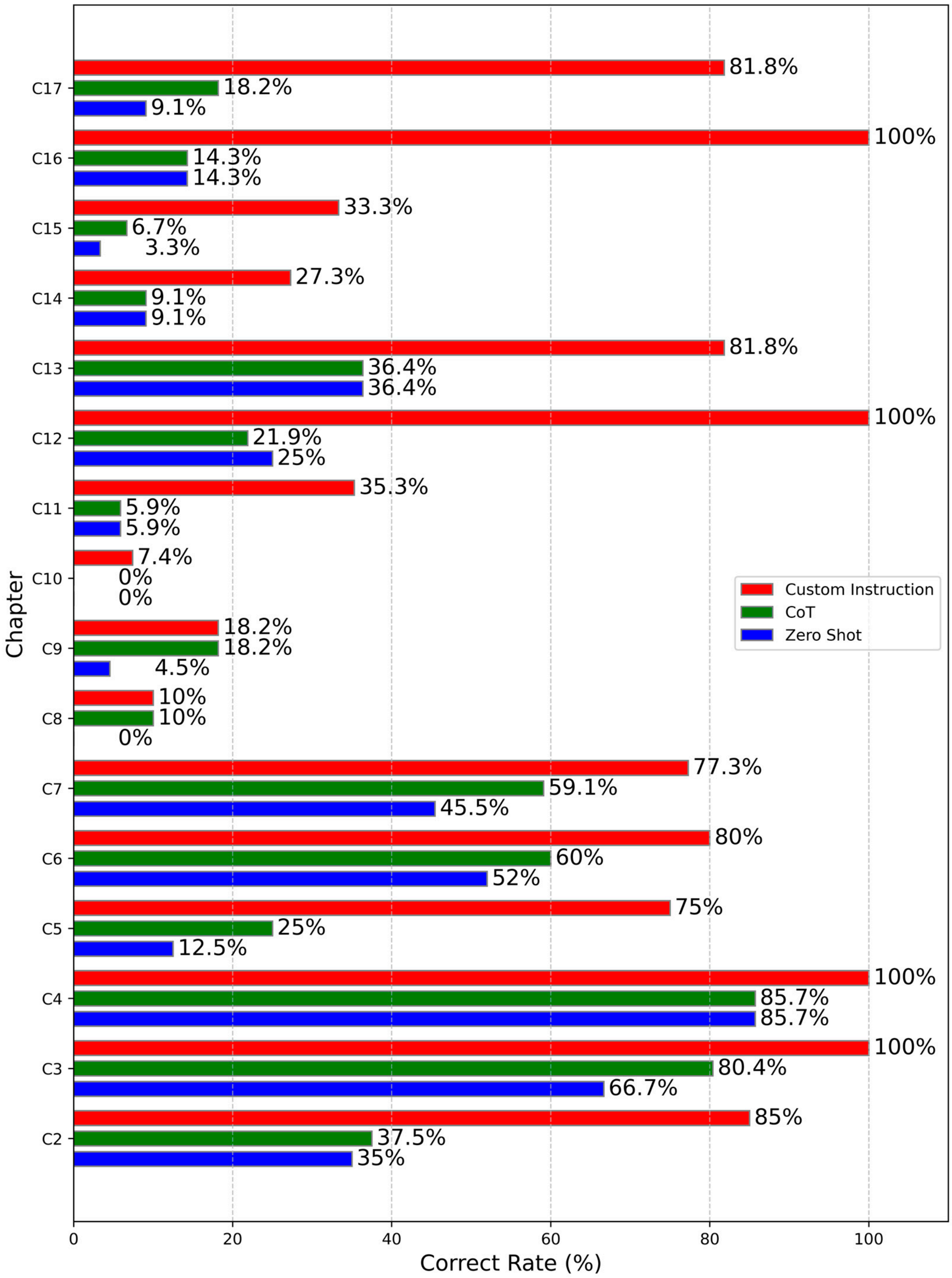


Figure 5. The Overall Accuracy of Each Prompting Strategy with Various Topics.

4.1.3. Question Type

Figure 7 portrays the accuracy of GPT-4 in answering questions with and without images using the three different prompting strategies, i.e., Zero Shot, CoT, and Custom Instruction. For text-based questions, the accuracy rates are 46.4% with Zero Shot, 52.6% with CoT, and markedly higher at 92.8% with Custom Instruction. In contrast, for image-based questions, the accuracies are significantly lower across all strategies, with Zero Shot at 8.8%, CoT at 12.6%, and Custom Instruction at 37.4%. Overall, the data succinctly underscore the better performance of GPT-4 in handling text-based questions and emphasise the pronounced escalation in precision afforded by the Custom Instruction strategy.

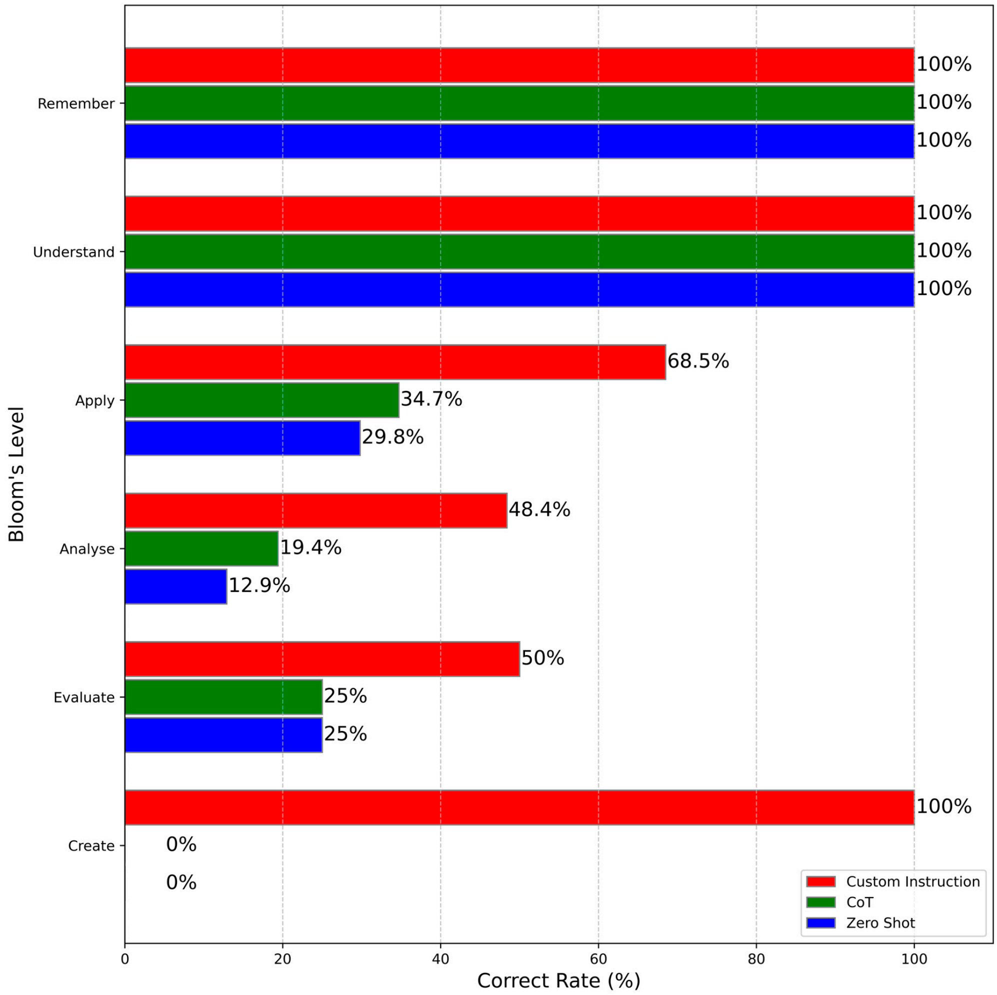


Figure 6. The Overall Accuracy of Each Prompting Strategy with Various Bloom's Levels.

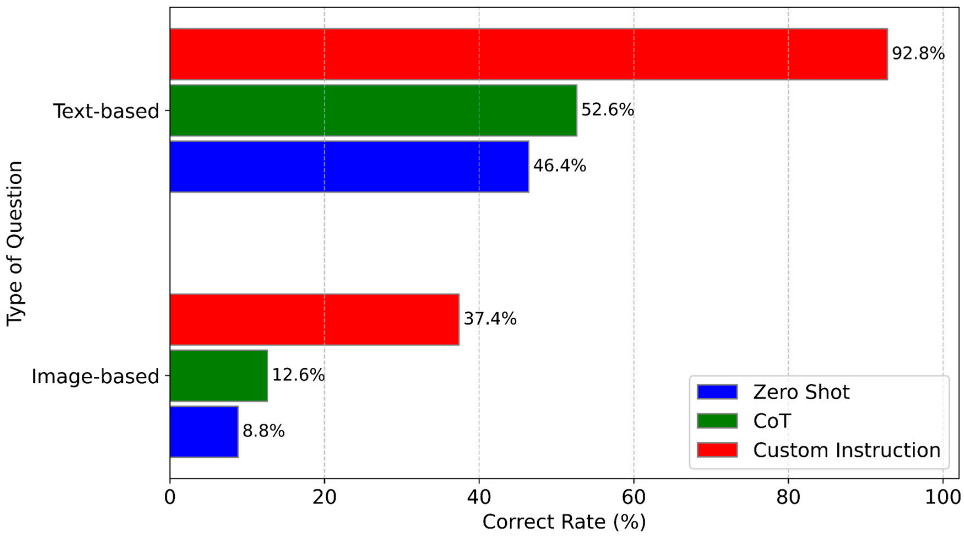


Figure 7. Accuracy of Each Prompting Strategy with Different Question Types.

4.2. Error Type Distribution

Subsequent to our analysis of accuracy rates, we investigated the error patterns associated with the utilisation of the three prompting strategies, as illustrated in Figure 8. This figure synthesises the occurrence rates of four principal error types—"Conceptual", "Grounding", "Calculation", and "Deficiency"—that were identified in the responses generated by GPT-4. In the Zero Shot strategy, "Deficiency" errors were the most prevalent at 44.2%, followed by "Grounding" errors at 39.2%, "Conceptual" errors at 12.6%, and "Calculation" errors at 4%. The CoT strategy shows a similar distribution, with "Deficiency" errors being the most common at 45.8%, "Grounding" errors at 40.1%, and equal rates for "Conceptual" errors at 12.6% and "Calculation" errors at 1.5%. However, the Custom Instruction strategy reveals a pronounced concentration of "Deficiency" errors, constituting 85.9% of its error profile, while "Calculation" errors emerge as the second most common at 9.4%. This observation potentially implies that the Custom Instruction strategy is more efficient in addressing "Conceptual" and "Grounding" errors than "Deficiency" and "Calculation" errors.

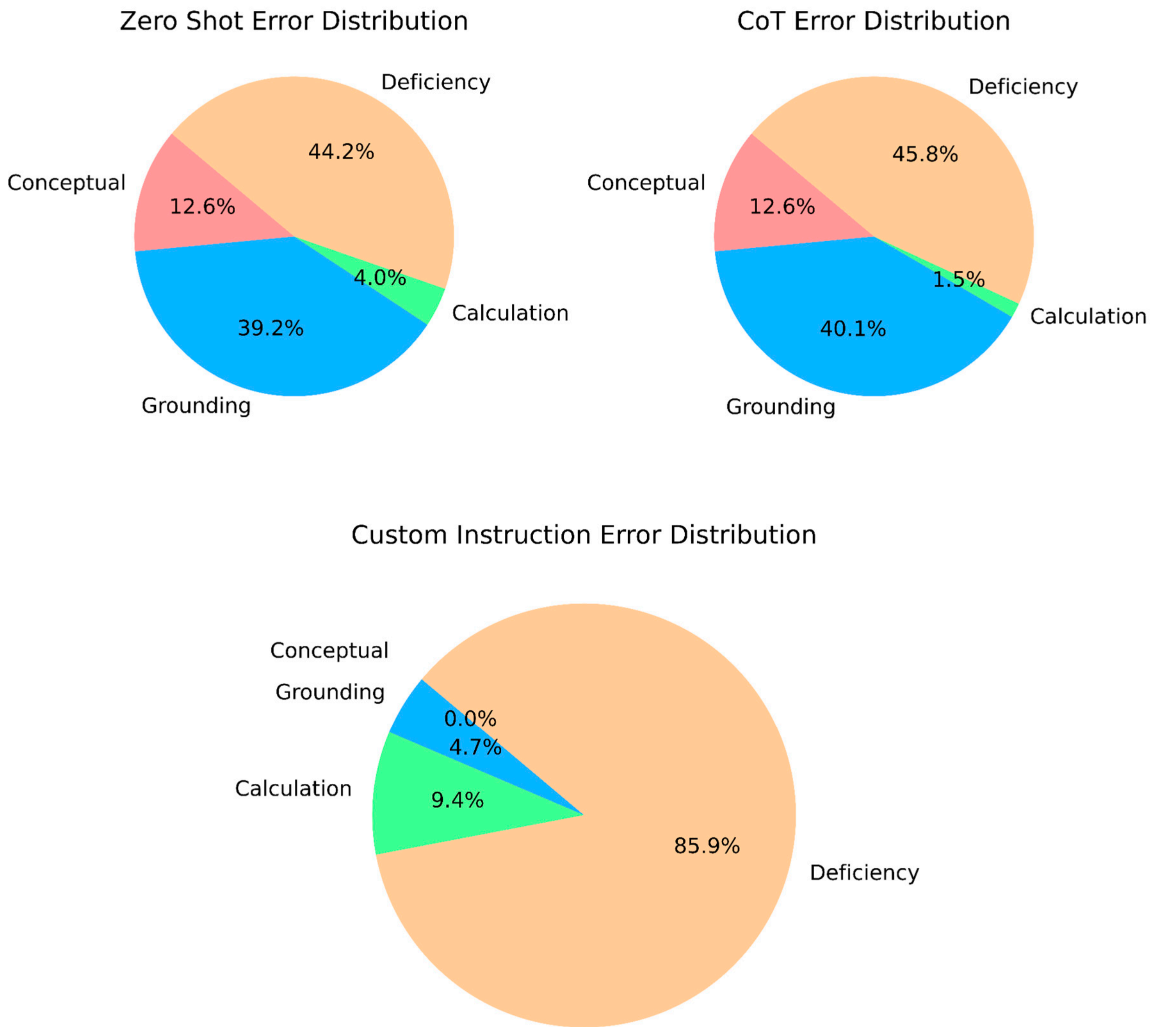


Figure 8. Error Types Across Various Prompting Strategies with GPT-4.

4.2.1. Different Topics

Figure 9 presents a comparative analysis of error frequencies across different chapters using the three prompting strategies—Zero Shot, CoT, and Custom Instruction. Organised into a tripartite heatmap structure, each column delineates one of the strategies, while the rows index the chapters (C2 through C17). Within each cell of the heatmaps, the number of errors is categorised into four types: "Conceptual", "Grounding", "Calculation", and "Deficiency". The colour intensity within the cells correlates with the count of errors, following a gradient from yellow (no errors) to dark blue (maximum error frequency).

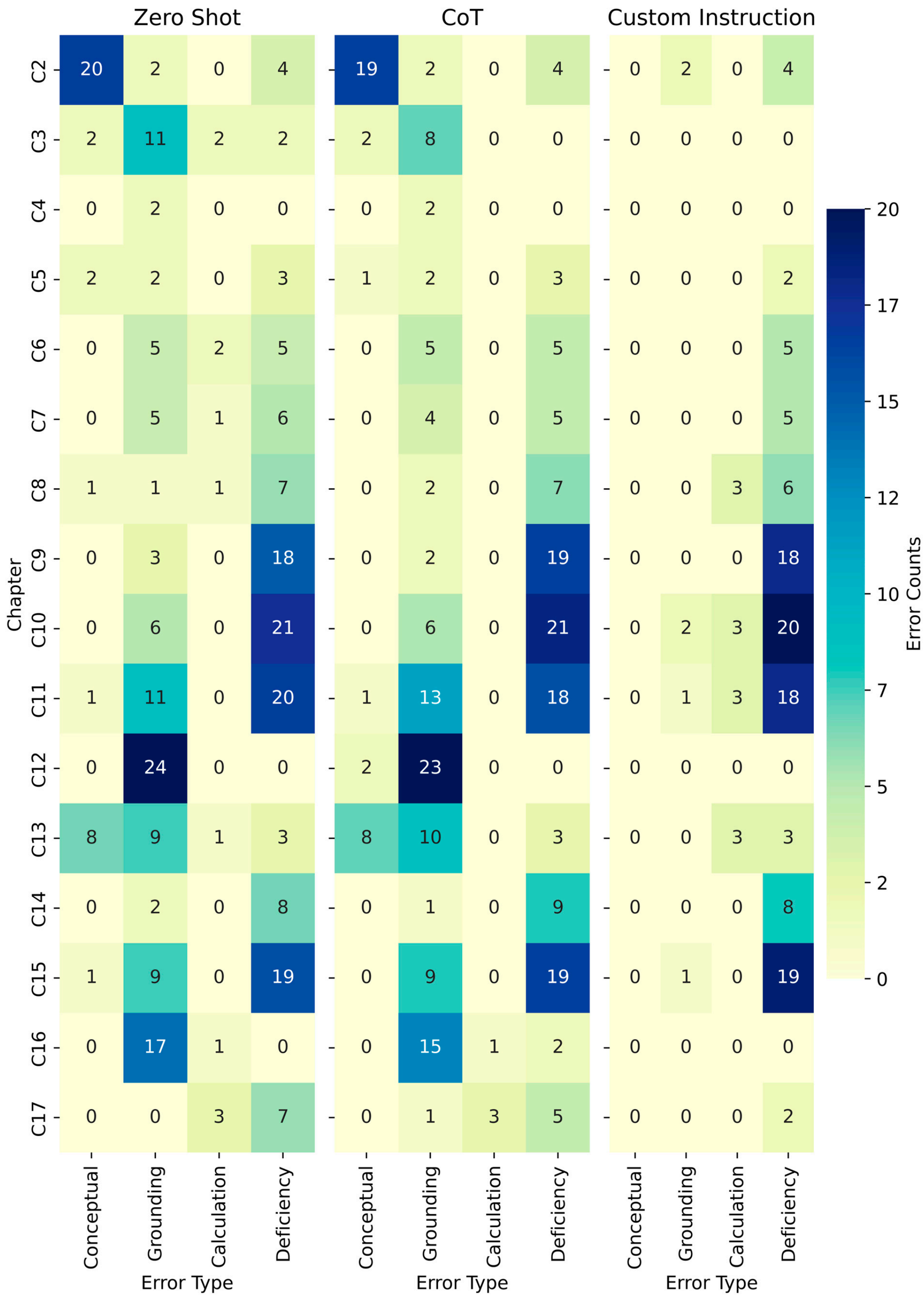


Figure 9. Comparative Heatmap Analysis of Error Types Across Chapters for Different Prompting Strategies.

For the Zero Shot strategy, the heatmap reveals a prominent concentration of "Conceptual" errors in the early chapters, with the highest count being 20 errors in C2—Origin of Soil and Grain Size, indicating challenges in grasping foundational concepts in these topics. "Grounding" errors are predominant in C12—Shear Strength of Soil, with 24 occurrences, suggesting difficulties in employing correct equations or constraints within this topic. Notably, "Calculation" errors are minimal, with no more than three instances observed, suggesting a relative strength of the model in handling algebraic and arithmetic operations. "Deficiency" errors related to visual data interpretation are more frequent in later chapters, such as C9 to C11. This trend could be attributed to the prevalence of image-based content in these chapters, posing interpretative challenges that the Zero Shot strategy does not adequately address.

Transitioning to the CoT strategy, a subtle decrement in "Conceptual" errors is discernible, yet "Grounding" and "Deficiency" errors persist with notable severity, especially in C12 and C10, respectively. The strategy shows its strength in curtailing "Calculation" errors, suggesting an aptitude for computational reasoning. However, the resilience of "Deficiency" errors across chapters indicates that the CoT approach may require further refinement to aid visual data comprehension.

The heatmap for Custom Instruction demonstrates a notable decrease in the "Conceptual" and "Grounding" errors, as evidenced by lower error counts in these categories across several chapters. This result indicates the effectiveness of the Custom Instruction strategy in enhancing understanding and context accuracy. However, "Deficiency" errors remain prevalent (e.g., 18–20 errors in later chapters like C9 and C10), highlighting ongoing challenges in dealing with visual information even with customised instructions.

Overall, the Zero Shot strategy shows a higher tendency for conceptual misunderstandings than the CoT and Custom Instruction strategies. The CoT strategy improves conceptual accuracy but does not fully resolve "Grounding" and "Deficiency" errors. The Custom Instruction strategy significantly reduces "Conceptual" and "Grounding" errors but struggles with "Deficiency" errors.

4.2.2. Difficulty Level According to Bloom's Taxonomy

Figure 10 represents the distribution of error types across different Bloom's Taxonomy levels for the Zero Shot, CoT, and Custom Instruction strategies. This figure reveals a concentration of errors, especially "Grounding" and "Deficiency", in more complex cognitive tasks like "Apply". This pattern suggests that the Zero Shot strategy might struggle with tasks that demand higher-order thinking and application of concepts.

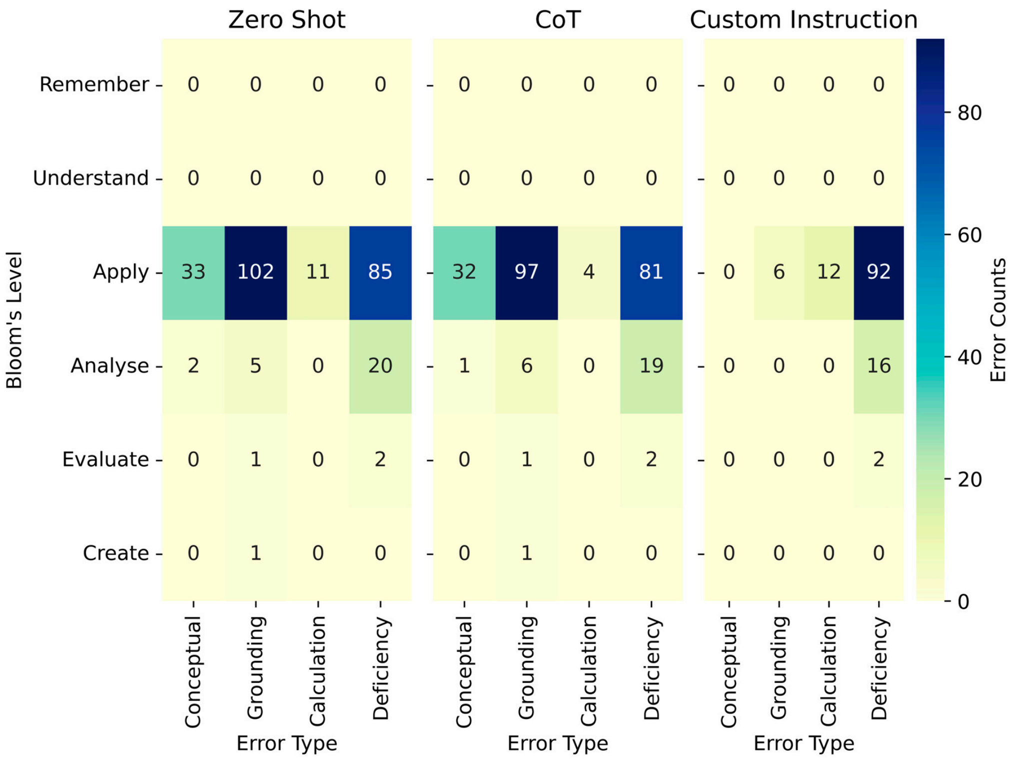


Figure 10. Comparative Heatmap Analysis of Error Types across Bloom's Taxonomy Levels for Different Prompting Strategies.

The CoT strategy, as depicted in the heatmap, similarly grapples with "Grounding" and "Deficiency" errors within the "Apply" level, although to a slightly lesser extent than Zero Shot, underscoring a common difficulty in tasks requiring the practical application of knowledge.

The Custom Instruction strategy demonstrates almost complete mitigation of "Conceptual" and "Grounding" errors across most levels of Bloom's taxonomy, signifying its effectiveness in reinforcing fundamental understanding and contextual application. Nonetheless, it confronts a pronounced incidence of "Deficiency" errors, particularly within the "Apply" category. This observation points to a persistent challenge this strategy faces with tasks demanding the synthesis of visual information and its application.

5. Discussion

5.1. Insights and Implications for Geotechnical Education

In our exploration of GPT-4's application in geotechnical education, several key insights and implications have emerged:

1. Errors and Problem Complexity: We observed a direct correlation between the complexity of geotechnical problems and the increase in errors from GPT-4. This trend underscores the model's challenges in processing and accurately responding to more complex tasks, indicating a potential area for future AI development and customised training.
2. Text-Based versus Image-Based Questions: GPT-4 exhibits proficiency in handling text-based queries, reflected in a lower occurrence of "Conceptual" errors. However, the performance of the version of GPT 4 used in image-based questions is limited due to an inherent inability to process or interpret images for geotechnical problems. This observation calls for integrated multimodal AI systems that can handle a variety of data formats.
3. Deficiencies and Limitations: Despite its strengths in processing textual information, GPT-4 exhibits notable deficiencies in tasks requiring interpreting images or generating complex visual representations. This limitation is particularly pronounced when the flow net diagrams are involved. The model struggles with creating or interpreting such visual content and occasionally fails to adhere strictly to the custom instructions provided. These shortcomings highlight two key areas in need of further development. First, there is a pressing need to enhance GPT-4's spatial and visual processing capability, enabling it to handle tasks beyond text-based information. Second, the model's occasional deviation from custom instructions underscores the ongoing requirement for refinement in its ability to understand and process user-specific directives accurately. Addressing these limitations is critical for advancing AI technologies like GPT-4 to improve their effectiveness as educational tools in geotechnical engineering and expand their applicability in solving complex, visually oriented problems.
4. Custom Instructions for Enhanced Accuracy: The application of tailored instructions has been shown to significantly improve the accuracy of GPT-4. The improvement is particularly notable in reducing "Conceptual" and "Grounding" errors and mitigating "Calculation" errors in tasks requiring numerical and logical reasoning. This aspect demonstrates the importance of context-specific guidance in enhancing AI performance in educational settings.

5.2. Demonstration of GPT-4's Question Generation Capabilities in Geotechnical Education

Utilising custom instructions, GPT-4 was tasked with generating educational questions and answers for specific chapters in geotechnical engineering, illustrating its application across various cognitive levels of Bloom's taxonomy. For Chapters 2, 3, and 12, GPT-4 was directed to create questions and provide answers for each level of Bloom's taxonomy: "Remember", "Understand", "Apply", "Analyse", "Evaluate", and "Create". The following examples in Table 2 showcase the model's ability after implementing custom instructions to formulate relevant questions and to provide correct answers. It should be noted that the

GPT-4's response is in italics. In addition, we explored GPT-4's proficiency as an AI tutor by evaluating its ability to correct errors in answering questions associated with geotechnical engineering. This evaluation is detailed in Table 3. This table shows that GPT-4 corrected the initial response, demonstrating its influential role as an AI tutor in identifying and rectifying errors in complex fundamental soil mechanics problems.

Table 2. Questions and Answers Generated by GPT-4 across Bloom's Taxonomy Levels.

Bloom's Level	Question and Answer Summary
Remember	**Question:** *What is the formula to calculate the degree of saturation (S) in soil mechanics?*
	Answer: *The degree of saturation (S) in soil mechanics is calculated using the formula* $S = \frac{(G_s\ w)}{e}$, *where* G_s *is the specific gravity,* w *is the water content in fraction, and* e *is the void ratio.*
Understand	**Question:** Explain why the maximum dry density or unit weight of soil is important in soil compaction and how it is related to the moisture content.
	Answer: *The maximum dry density or unit weight of soil is crucial in soil compaction as it represents the highest density a soil can be compacted to at a given moisture content. It's important for ensuring stability and support in engineering projects. The relationship with moisture content is significant as it determines the optimal water content at which the soil achieves its maximum dry density, leading to effective compaction and strength.*
Apply	**Question:** Given principal stresses σ_1 = 150 kPa and σ_3 = 50 kPa in a soil sample, and a friction angle (ϕ) of 30 degrees, calculate the normal stress (σn) and shear stress (τ) on a plane at an angle θ of 45 degrees using the relevant formulas.
	Answer: *To calculate the normal stress* (σ_n) *and shear stress* (τ) *on a plane at an angle* θ *of 45 degrees, we use the formulas:* $\sigma_n = \frac{(\sigma_1+\sigma_3)}{2} + \frac{(\sigma_1-\sigma_3)}{2} * cos(2 \times 45°)$ $\tau = \frac{(\sigma_1-\sigma_3)}{2} * sin(2 \times 45°)$ *Substituting the given values:* σ_1 *= 150 kPa,* σ_3 *= 50 kPa* *The calculated normal stress* (σ_n) *and shear stress* (τ) *on the plane at an angle of 45 degrees are:* *-Normal stress* (σ_n) *= 100 kPa* *-Shear stress* (τ) *= 50 kPa*
Analyse	**Question:** Given data for different soil samples including specific gravity, dry unit weight, and water content, determine the void ratio, degree of saturation, and wet unit weight for each sample.
	Answer: *To determine the void ratio, degree of saturation, and wet unit weight for each soil sample, the following steps are taken:* *-Calculate the void ratio (e) using the formula* $e = (G_s\ \gamma_w)/\gamma_d) - 1$. *-Calculate the degree of saturation (S) using* $S = (G_s w)/e$. *-Calculate the wet unit weight* (γ_{wet}) *using* $\gamma_{wet} = \gamma_d\ (1 + w)$. *Specific calculations would depend on the provided data for each soil sample.*
Evaluate	**Question:** Critically assess the applicability of Mohr-Coulomb Failure Theory in predicting shear failure in a clayey soil, considering its cohesion and angle of internal friction.
	Answer: *The Mohr-Coulomb Failure Theory is a widely used criterion for predicting shear failure in soils. For clayey soils, which typically have significant cohesion, this theory can be effective in predicting failure. However, the theory assumes linear increase of shear strength with normal stress, which may not always accurately represent the complex behaviour of clay, especially under varied moisture and loading conditions. Thus, while useful, the theory should be applied with consideration of its limitations and possibly supplemented with empirical data specific to the type of clay.*

Table 2. *Cont.*

Bloom's Level	Question and Answer Summary
Create	**Question:** Design a method to experimentally determine the maximum and minimum void ratios of a soil sample, and subsequently calculate its relative density. **Answer:** *To experimentally determine the maximum and minimum void ratios, the following method could be designed:* *-For the maximum void ratio (e_{max}) Place a known volume of loosely poured soil in a container and measure its mass. Calculate the void ratio using the soil's specific gravity and the volume of the container.* *-For the minimum void ratio (e_{min}) Place the same volume of soil in a container and compact it to its maximum density (e.g., using a mechanical compactor), then measure its mass. Calculate the void ratio as before.* *-With e_{max} and e_{min}, calculate the relative density (D_r) using $D_r = (e_{max} - e)/(e_{max} - e_{min})$.* *This experiment would provide a comprehensive understanding of the soil's packing characteristics.*

Table 3. Evaluation of GPT's Accuracy in Correcting Calculation Errors in Soil Mechanics Problem-Solving.

Prompted Question and Answer:
Question: Given principal stresses σ_1 = 150 kPa and σ_3 = 50 kPa in a soil sample and a friction angle (ϕ) of 30 degrees, calculate the normal stress (σ_n) and shear stress (τ) on a plane at an angle θ of 45 degrees using the relevant formulas. Answer: To calculate the normal stress (σ_n) and shear stress (τ) on a plane at an angle θ of 45 degrees, we use the formulas: $\sigma_n = \frac{(\sigma_1+\sigma_3)}{2} + \frac{(\sigma_1-\sigma_3)}{2}\cos(2\times 45^\circ)\tau = \frac{(\sigma_1-\sigma_3)}{2}\sin(2\times 45^\circ)$ Substituting the given values: σ_1 = 150 kPa, σ_3 = 50 kPa. The calculated normal stress (σ_n) and shear stress (τ) on the plane at an angle of 45 degrees are: -Normal stress (σ_n) = 120 kPa $-$ Shear stress (τ) = 60 kPa. Can you check if the solution is correct?
GPT's Evaluation and Response:
The correct normal stress (σ_n) and shear stress (τ) on the plane at an angle of 45 degrees, calculated using the given values of $\sigma_1 = 150$ kPa and $\sigma_3 = 50$ kPa, are: • *Normal stress (σ_n) = 100 kPa* • *Shear stress (τ) = 50 kPa* *The previously provided answer is incorrect.*

6. Conclusions

The paper investigates the application of large language models (LLMs), specifically GPT-4, in geotechnical engineering education. It assesses GPT-4's performance across various topics and cognitive complexity levels using different prompting strategies. The study reveals that while GPT-4 shows adeptness in handling basic geotechnical concepts and problems, its effectiveness varies with task complexity and prompting methods. Errors in GPT-4 responses are categorised into "Conceptual", "Grounding", "Calculation", and "Deficiency" related to visual interpretation. While the "Regenerate" function of ChatGPT has potential utility for improving the diversity of responses and correcting immediate errors, it should be used carefully in educational settings. Educators and students should employ this function as part of a broader strategy that includes critical evaluation and verification of AI-generated content, particularly in specialised fields such as geotechnical engineering. Our proposed prompting strategy—Custom Instruction—significantly enhanced GPT-4's performance, particularly in rectifying most errors, albeit with persistent challenges in calculations and visual interpretations, and hence such use is recommended.

A salient aspect of our investigation is the indispensable role of human intervention in refining GPT-4's outputs for complex tasks, underscoring the need for human–AI collaboration. The study demonstrates GPT-4's potential as a valuable tool in geotechnical education, such as generating questions, answers and feedback, and its emerging role as an interactive AI tutor for enriching students' learning experience.

However, our research also brings to light AI's limitations and reliability concerns in geotechnical engineering. The study suggests that despite ChatGPT's potential to assist in educational scenarios, its responses often need verification and refinement by knowledgeable individuals. These limitations stem from the model's general training approach and inherent constraints, often leading to inaccuracies or oversimplifications in addressing complex problems. The lack of domain-specific training and the model's inherent limitations sometimes lead to inaccuracies or oversimplifications in complex problem-solving scenarios.

Therefore, a significant future direction for AI in education is the development of more sophisticated, domain-tailored generative AI models. These models should be trained with discipline-specific datasets to enhance their accuracy and reliability. Until such advancements are achieved, human oversight remains essential in utilising AI tools like ChatGPT. Educators and students must critically evaluate AI-generated content, ensuring it aligns with factual and theoretical standards.

In summary, while generative AI in geotechnical education opens up exciting avenues for innovative learning and teaching, its current state demands a balanced integration, with human expertise playing a crucial role in guiding and overseeing its application. The future evolution of generative AI in education hinges on technological advancements and its responsible and informed use by educators and learners.

Author Contributions: Conceptualisation, A.T., L.C., U.H. and J.K.; methodology, L.C., A.T., U.H. and J.K.; software, A.T., L.C. and U.H.; validation, A.T. and L.C.; formal analysis, L.C., A.T. and U.H.; investigation, L.C., A.T., U.H. and J.K.; resources, A.T.; data curation, L.C.; writing—original draft preparation, L.C. and A.T.; writing—review and editing, J.K.; visualisation, A.T., L.C. and U.H.; supervision, J.K.; project administration, J.K.; funding acquisition, J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research work is sponsored by the SPARC Hub (https://sparchub.org.au) in the Department of Civil Engineering at Monash University funded by the Australian Research Council (ARC) Industrial Transformation Research Hub (ITRH) Scheme (Project ID: IH180100010). The financial and in-kind support received from Monash University and the SPARC Hub are gratefully acknowledged.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy reasons.

Acknowledgments: While preparing this work, the authors used ChatGPT to draft and edit the article. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication. Special thanks to Ace Mitrevski for his help in initial data collection.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of the data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

This appendix exemplifies different error categories observed when employing the Zero Shot approach, along with the improvements made through the CoT approach and the custom instructional prompting. Each example demonstrates a specific type of error made by GPT-4, followed by the CoT approach and the tailored instruction employed to rectify it. The errors made by GPT-4 have been underlined for clarity.

(1) "Conceptual" error

Question:

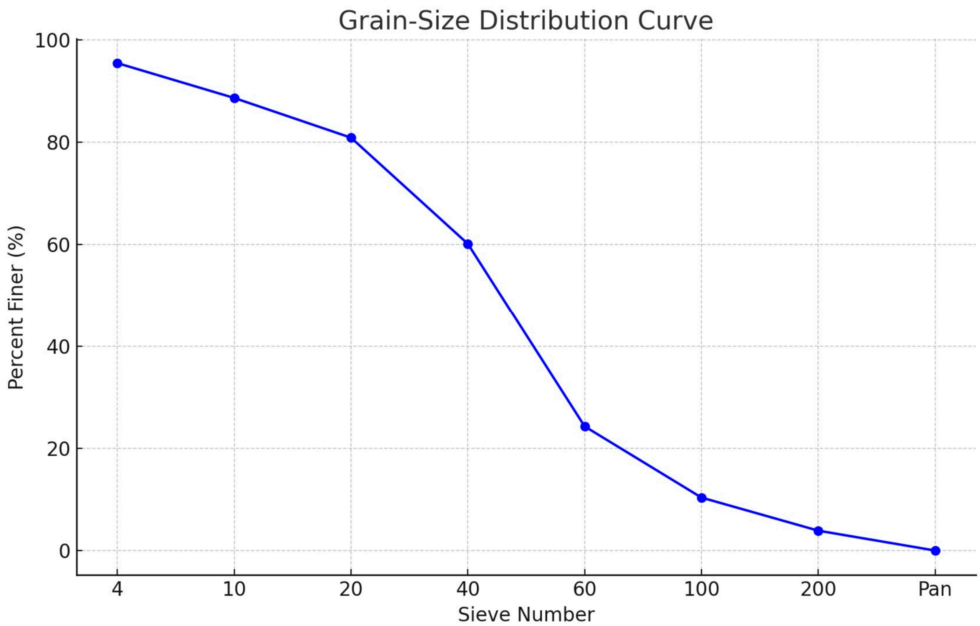
The following are the results of a sieve analysis:

US sieve no.	Mass of soil retained
4	28
10	42
20	48
40	128
60	221
100	86
200	40
Pan	24

Determine the percent finer than each sieve and plot a grain-size distribution curve.
Original GPT-4 answer:

US sieve no.	*Mass of soil retained*	*Cumulative mass retained*	*Percent finer (%)*
4	*28*	*28*	*95.46*
10	*42*	*70*	*88.65*
20	*48*	*118*	*80.88*
40	*128*	*246*	*60.13*
60	*221*	*467*	*24.31*
100	*86*	*563*	*10.37*
200	*40*	*593*	*3.89*
Pan	*24*	*617*	*0.00*

Next, let's plot the grain-size distribution curve using this data.



CoT prompting:
Solve the question and do it step by step.
GPT-4 answer after implementing the CoT prompting:

To determine the percent finer than each sieve and plot a grain-size distribution curve based on the given sieve analysis results, we'll follow these steps:

1. *Calculate the Total Mass: Sum all the mass retained on each sieve and in the pan to get the total mass of the soil sample.*
2. *Calculate the Mass Passing: For each sieve, determine the mass of soil that passes through by subtracting the cumulative mass retained up to and including that sieve from the total mass.*
3. *Calculate the Percent Finer: For each sieve, calculate the percent finer than that sieve by dividing the mass passing by the total mass and then multiplying by 100.*
4. *Plot the Grain-Size Distribution Curve: Plot the percent finer on the y-axis against the sieve size on the x-axis.*

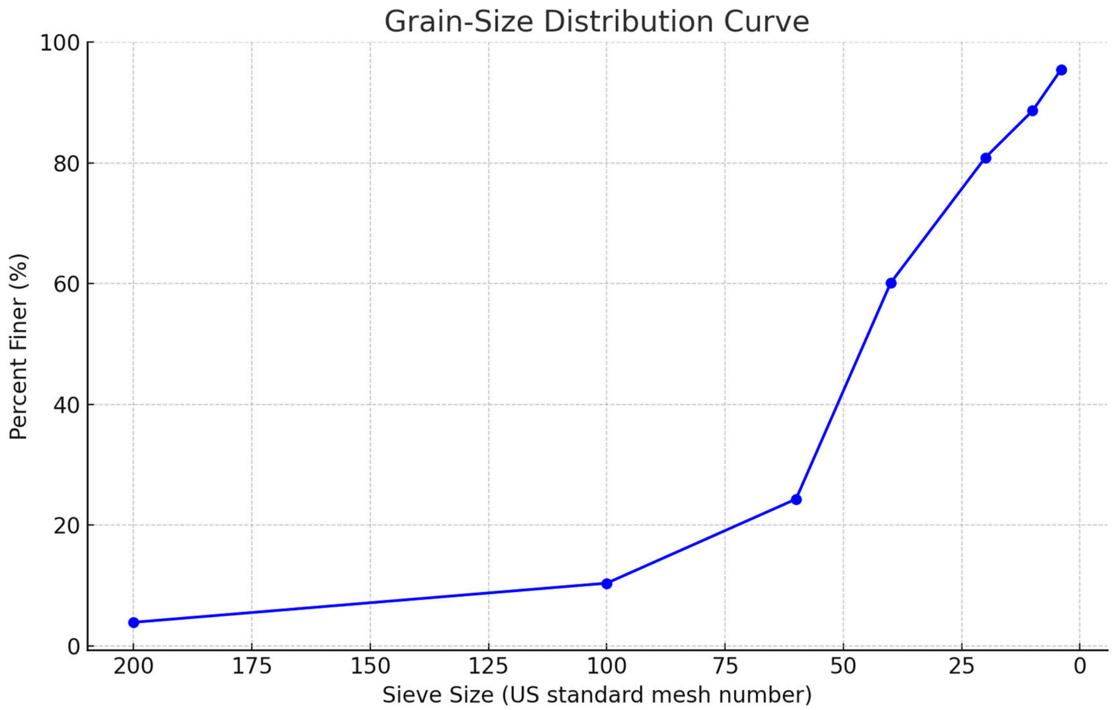
Let's start by calculating the total mass of the soil sample.
Step 1: Calculate the Total Mass
We have the following data:

- *US sieve no. 4: Mass retained = 28 g*
- *US sieve no. 10: Mass retained = 42 g*
- *US sieve no. 20: Mass retained = 48 g*
- *US sieve no. 40: Mass retained = 128 g*
- *US sieve no. 60: Mass retained = 221 g*
- *US sieve no. 100: Mass retained = 86 g*
- *US sieve no. 200: Mass retained = 40 g*
- *Pan: Mass retained = 24 g*

Let's calculate the total mass.
The total mass of the soil sample is 617 g.
Step 2: Calculate the Mass Passing Each Sieve
To find the mass passing through each sieve, we need to calculate the cumulative mass retained at each sieve and then subtract it from the total mass.
Step 3: Calculate the Percent Finer
For each sieve, the percent finer is calculated by dividing the mass passing by the total mass and multiplying by 100.
I'll now calculate the cumulative mass retained, the mass passing each sieve, and the percent finer for each sieve.
Here are the mass passing and percent finer than each sieve:

US Sieve No.	*Mass Passing (g)*	*Percent Finer (%)*
4	*589*	*95.46*
10	*547*	*88.65*
20	*499*	*80.88*
40	*371*	*60.13*
60	*150*	*24.31*
100	*64*	*10.37*
200	*24*	*3.89*
Pan	*0*	*0.00*

Step 4: Plot the Grain-Size Distribution Curve
I will now plot the grain-size distribution curve using the sieve sizes and the corresponding percent finer values. This plot will help visualise the distribution of particle sizes within the soil sample.
Here's the grain-size distribution curve for your soil sample. The curve shows the percent finer than each sieve size, providing a graphical representation of the soil's particle size distribution. Larger sieve numbers on the left correspond to finer particles, as the x-axis is inverted to reflect decreasing sieve sizes (increasing particle sizes).
If you have any more questions or need further analysis, feel free to ask!



Custom instruction:

US sieve number should be first converted to sieve sizes.

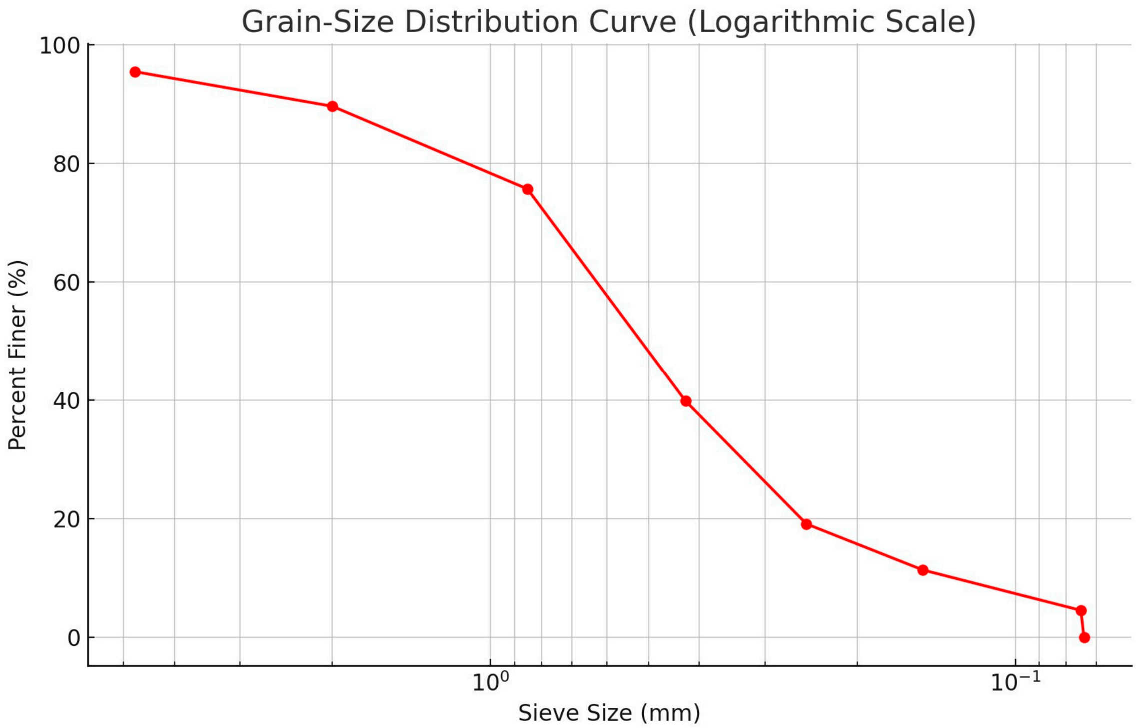
GPT-4 answer after implementing custom instructions:

Let's go through the steps needed to complete the sieve analysis and create the necessary table:

1. *Calculate the Percent Retained on Each Sieve: This is the mass of soil retained on each sieve as a percentage of the total mass.*
2. *Calculate the Percent Finer than Each Sieve: This is calculated by subtracting the cumulative percent retained up to that point from* 100%.
3. *Convert U.S. Sieve Numbers to Standard Sieve Sizes: This is done using a conversion table that correlates U.S. Sieve numbers to millimetre openings. We will perform these calculations and create the table with the following columns: U.S. Sieve Number, Sieve Size (mm), Mass of Soil Retained (g), Percent Retained on Each Sieve, and Percent Finer. Here is the completed table with US sieve numbers, corresponding standard sieve sizes in millimetres, the mass of soil retained on each sieve, percent retained on each sieve, and percent finer:*

U.S. Sieve No.	*Sieve Size (mm)*	*Mass of Soil Retained (g)*	*Percent Retained on Each Sieve (%)*	*Percent Finer (%)*
4	*4.75*	*28*	*4.54*	*95.46*
10	*2.00*	*42*	*6.81*	*89.63*
20	*0.85*	*48*	*7.78*	*75.69*
40	*0.425*	*128*	*20.75*	*39.87*
60	*0.25*	*221*	*35.82*	*19.12*
100	*0.15*	*86*	*13.94*	*11.35*
200	*0.075*	*40*	*6.48*	*4.54*
Pan	*0*	*24*	*3.89*	*0.00*

The total mass of the soil sample is 617 g.



(2) "Grounding" error

Question:

When the moisture content of a soil is 26%, the degree of saturation is 72%, and the moist unit weight is 16.98 kN/m^3. Determine:

a. Specific gravity of soil solids

b. Void ratio

Original GPT-4 answer:

The problem statement provided by the OCR results is:

"When the moisture content of a soil is 26%*, the degree of saturation is* 72%*, and the moist unit weight is* 16.98 kN/m^3*. Determine:*

a. Specific gravity of soil solids

b. Void ratio

Let's proceed to calculate each of these:

Specific gravity of soil solids (G)*: The specific gravity of the soil solids can be calculated using the relationship between the degree of saturation* (S)*, the moisture content* (w)*, the void ratio* (e)*, and the unit weight of water* γ_w*. We need to solve for the specific gravity* (G) *using the formula:*

$$S = \frac{w \cdot G}{e}$$

However, since we don't have the void ratio yet, we need to express the void ratio in terms of G *and the moist unit weight* (γ_m) *first. We can use the formula:*

$$e = \frac{G \cdot w}{\gamma_d / \gamma_w - G}$$

where:

- *w is the moisture content,*
- *γ_d is the dry unit weight, calculated by $\gamma_m / (1 + w)$,*
- *γ_w is the unit weight of water $(9.81\ kN/m^2)$.*
- *Void ratio e : After calculating G, we can use the above relationship to find the void ratio.*

where γ_d is the dry unit weight.

Let's start with the calculation of the specific gravity of soil solids, as it is fundamental to solving the other parts of the problem.

The calculations have resulted in complex numbers, which is not physically meaningful in this context. This indicates an error in the calculation process. I need to re-evaluate the equations and approach used to solve for the specific gravity of the soil solids.

Given:

- *The moisture content* $w = 26\%$ *or 0.26 in decimal,*
- *The degree of saturation* $S = 72\%$ *or 0.72 in decimal,*
- *The moist unit weight* $\gamma_m = 16.98\ kN/m^3$*,*
- *The unit weight of water* $\gamma_w = 9.81\ kN/m^3$*,*

We want to find the specific gravity G, void ratio e, and saturated unit weight γ_{sat}*.*

The formula for the degree of saturation in terms of specific gravity is:

$$S = \frac{w \cdot G}{e}$$

And we can relate the void ratio to the specific gravity using the dry unit weight γ_d*:*

$$e = \frac{G \cdot \gamma_w}{\gamma_d} - 1$$

The dry unit weight γ_d can be calculated from the moist unit weight γ_m and moisture content w:

$$\gamma_d = \frac{\gamma_m}{1 + w}$$

To calculate the specific gravity G, we can rearrange the equation for S in terms of and solve for G knowing S, w*, and* γ_d*.*

Let's rework these calculations correctly.

The recalculated values are now physically meaningful and are as follows:

a. Specific gravity of soil solids: 1.32

b. Void ratio: 7.06

CoT prompting:

Solve the question and do it step by step.

GPT-4 answer after implementing the CoT prompting:

To solve this question, we need to find:

a. The specific gravity of soil solids (G). b. The void ratio (e).

We start by using the given information:

- *Moisture content (w) = 26% or 0.26*
- *Degree of saturation* (S_r) *= 72% or 0.72*
- *Moist unit weight* (γ) *= 16.98 kN/m*^{3}

Step 1: Calculate the Dry Unit Weight (γ_d)

The dry unit weight (γ_d) *can be determined from the moist unit weight* (γ) *and the moisture content (w) using the following formula:*

$$\gamma_d = \frac{\gamma}{1 + w}$$

Step 2: Calculate the Specific Gravity of Soil Solids (G)

To calculate the specific gravity, we need the void ratio (e), which we derive from the relationship involving degree of saturation (S_r)*, moisture content (w), and specific gravity (G).The formula is:*

$$S_r = \frac{w \gamma_w G}{e}$$

where γ_w *is the unit weight of water (approximately 9.81 kN/m*^{3}*).*

We need to rearrange this formula to first find e and then G.

Step 3: Calculate the Void Ratio (e)

We rearrange the equation to:

$$e = \frac{w\gamma_w G}{S_r}$$

And using the fact that:

$$\gamma_d = \frac{G\gamma_w}{1+e}$$

We can combine and rearrange these equations to solve for e and G. However, as G appears in both, it typically requires iterative methods or initial guesswork to solve. Here, we'll make an initial estimate for G (commonly around 2.65 for mineral soils) and solve iteratively if necessary.

Let's begin by calculating the dry unit weight (γ_d) using the provided information. The dry unit weight (γ_d) is 13.48 kN/m^3.

Step 3: Estimate and Calculate Void Ratio (e) and Specific Gravity (G)

We will make an initial estimate for the specific gravity of soil solids (G). Let's assume G = 2.65 for mineral soils. We can then use this to calculate the void ratio (e) using the formula:

$$e = \frac{w\gamma_w G}{S_r}$$

Let's compute the void ratio with this initial guess for G.

There seems to have been an error in my calculation of the void ratio (e), resulting in an unusually high value, which suggests a mistake in the use of formulas. Let me correct that and calculate again.

We need to use:

$$e = \frac{w\gamma_w}{S_r\gamma_d}$$

instead, and then use e to solve for G using:

$$\gamma_d = \frac{G\gamma_w}{1+e}$$

Let's calculate e correctly now.

The corrected void ratio (e) is approximately 0.263.

Step 4: Calculate Specific Gravity (G)

Now, let's use the correct void ratio to calculate the specific gravity of soil solids (G) using the formula:

$$\gamma_d = \frac{G\gamma_w}{1+e}$$

We can rearrange this to solve for G:

$$G = \frac{\gamma_d(1+e)}{\gamma_w}$$

Let's compute G.

The specific gravity of the soil solids (G) is approximately 1.73. This value is somewhat low for typical mineral soils, which usually have G around 2.65. If this result is unexpected, it might be useful to verify the inputs or consider whether the soil has organic content or other characteristics affecting its specific gravity.

Here are the answers based on the calculations:

- *Specific Gravity of Soil Solids (G): 1.73*
- *Void Ratio (e): 0.263*

If you have any more questions or if something seems off, let me know!

Custom instruction:

The void ratio which is a dimensionless quantity is calculated using the following equation, $e = \frac{G_s \cdot \gamma_w}{\gamma_d} - 1$, where G_s is the specific gravity, γ_w is the unit weight of water and γ_d is the dry unit weight. In addition, the degree of saturation (S) can be calculated as $S = G_s w/e$ where, w is the water content in fraction.

GPT-4 answer after implementing custom instructions:

Based on the information provided and the equations for specific gravity, void ratio, and saturated unit weight, we can solve for the unknowns.

Given:

- *Moisture content* $(w) = 26\% = 0.26$ *(in fraction)*
- *Degree of saturation*$(S) = 72\% = 0.72$ *(in fraction)*
- *Moist unit weight*$(\gamma_m) = 16.98\ kN/m^3$

We will use the following equations to solve for the specific gravity of soil solids G_s*, void ratio (e), and saturated unit weight* γ_{sat}

1. *Void ratio is given by:*

$$e = \frac{G_s \cdot \gamma_w}{\gamma_d} - 1$$

2. *Degree of saturation (S) is given by:*

$$S = \frac{G_s \cdot w}{e}$$

We also know that the dry unit weight (γ_d) *can be found from the moist unit weight* (γ_w) *and the moisture content* (w) *using the equation:*

$$\gamma_d = \frac{\gamma_w}{1+w}$$

Given that the unit weight of water γ_w *is typically* $9.81\ kN/m^3$*, we can proceed to solve these equations. Let's first calculate the dry unit weight* γ_d*. Then, we can set up the equations to solve for the specific gravity* G_s *and the void ratio. Finally, we'll calculate the saturated unit weight* γ_{sat}

The calculated values for the given soil properties are:

a. Specific gravity of soil solids G_s*: approximately 2.73*

b. Void ratio: approximately 0.98

(3) "Calculation" error

Question:

The saturated unit weight of a soil is 19.8 kN/m^3. The moisture content of the soil is 17.1%. Determine the following: a. Dry unit weight b. Specific gravity of soil solids c. Void ratio.

Original GPT-4 answer:

Let's try to solve for G_s *using the relationship between* e, G_s*, and w for saturated soil, while keeping in mind that* $S = 1$.

From $S = \frac{G_s \times w}{e}$ *and* $e = \left(\frac{G_s \times \gamma_w}{\gamma_d}\right) - 1$*, we can set the two expressions for e equal to each other:*

$$\frac{G_s \times w}{S} = \left(\frac{G_s \times \gamma_w}{\gamma_d}\right) - 1$$

Simplifying, we get:

$$G_s \times w = G_s \times \gamma_w - \gamma_d \Longrightarrow G_s(\gamma_w - w) = \gamma_d \Longrightarrow G_s = \frac{\gamma_d}{\gamma_w - w}$$

CoT prompting:

Solve the question and do it step by step.

GPT-4 answer after implementing the CoT prompting:

To find the required soil properties based on the given data:

- *Saturated unit weight* (γ_{sat}) *= 19.8 kN/m*^{3}

- *Moisture content* (w) *= 17.1% or 0.171*

Step 1: Calculate the Dry Unit Weight (γ_d)

The dry unit weight (γ_d) *can be calculated using the relation between moist unit weight* (γ)*, moisture content* (w)*, and saturated unit weight* (γ_{sat})*. For saturated conditions:*

$$\gamma_{sat} = \gamma_d + w\gamma_w$$

Rearranging this formula to find the dry unit weight:

$$\gamma_d = \gamma_{sat} - w\gamma_w$$

Step 2: Calculate the Specific Gravity of Soil Solids (G)

To find G, we use the relationship involving the dry unit weight, the specific gravity, and the void ratio, which is derived from:

$$\gamma_d = \frac{G\gamma_w}{1+e}$$

Since we don't know the void ratio (e) directly, we first assume that the soil is saturated, and hence the degree of saturation (S_r) *is 100% or 1. We can use the following relationship:*

$$e = \frac{w\gamma_w}{\gamma_d}$$

And then use this e to determine G.

Step 3: Calculate the Void Ratio (e)

We'll start by calculating the void ratio using the above relation after computing γ_d*.*

Let's begin by calculating the dry unit weight (γ_d) *using the unit weight of water* (γ_w *= 9.81 kN/m*^{3}*).*

The dry unit weight (γ_d) *is approximately 18.12 kN/m*^{3}*.*

Step 2: Calculate the Void Ratio (e)

Using the derived formula:

$$e = \frac{w\gamma_w}{\gamma_d}$$

we'll compute the void ratio.

The void ratio (e) is approximately 0.093.

Step 3: Calculate the Specific Gravity of Soil Solids (G)

Using the relationship:

$$\gamma_d = \frac{G\gamma_w}{(1+e)}$$

we can solve for G by rearranging it:

$$G = \frac{\gamma_d(1+e)}{\gamma_w}$$

Let's compute the specific gravity (G).

The specific gravity of the soil solids (G) is approximately 2.02.

Here are the answers based on the calculations:

- *Dry Unit Weight* (γ_d)*: 18.12 kN/m*^{3}
- *Specific Gravity of Soil Solids (G): 2.02*
- *Void Ratio (e): 0.093*

These results are typical for mineral soils where the specific gravity often ranges from 2.65 for non-organic minerals to around 2.00 for organic or highly porous materials. If you need further details or another calculation, feel free to ask!

Custom instruction:

In some of the questions, you might have to solve simultaneous equations.

GPT-4 answer after implementing custom instructions:

The degree of saturation S is 100% *(or 1 in fraction) for a saturated soil, which gives us:*

$$S = \frac{G_s \cdot w}{e} \Longrightarrow G_s = \frac{S \cdot e}{w}$$

The void ratio e can be calculated using the formula:

$$e = \frac{G_s \cdot \gamma_w}{\gamma_d} - 1$$

where γ_w *is the unit weight of water, approximately* $9.81\ kN/m^3$.

We can solve these two equations to find G_s *and e.*

Next, let's calculate the specific gravity of soil solids (G_s) *using the degree of saturation and void ratio (e). Since the soil is saturated,* $S = 1$.

The equation for S is:

$$S = \frac{G_s \times w}{e}$$

We also have the equation for e:

$$e = \left(\frac{G_s \times \gamma_w}{\gamma_{dry}}\right) - 1$$

where $\gamma_w = 9.81\ kN/m^3$ (unit weight of water).

We can solve these equations simultaneously to find G_s *and e.*

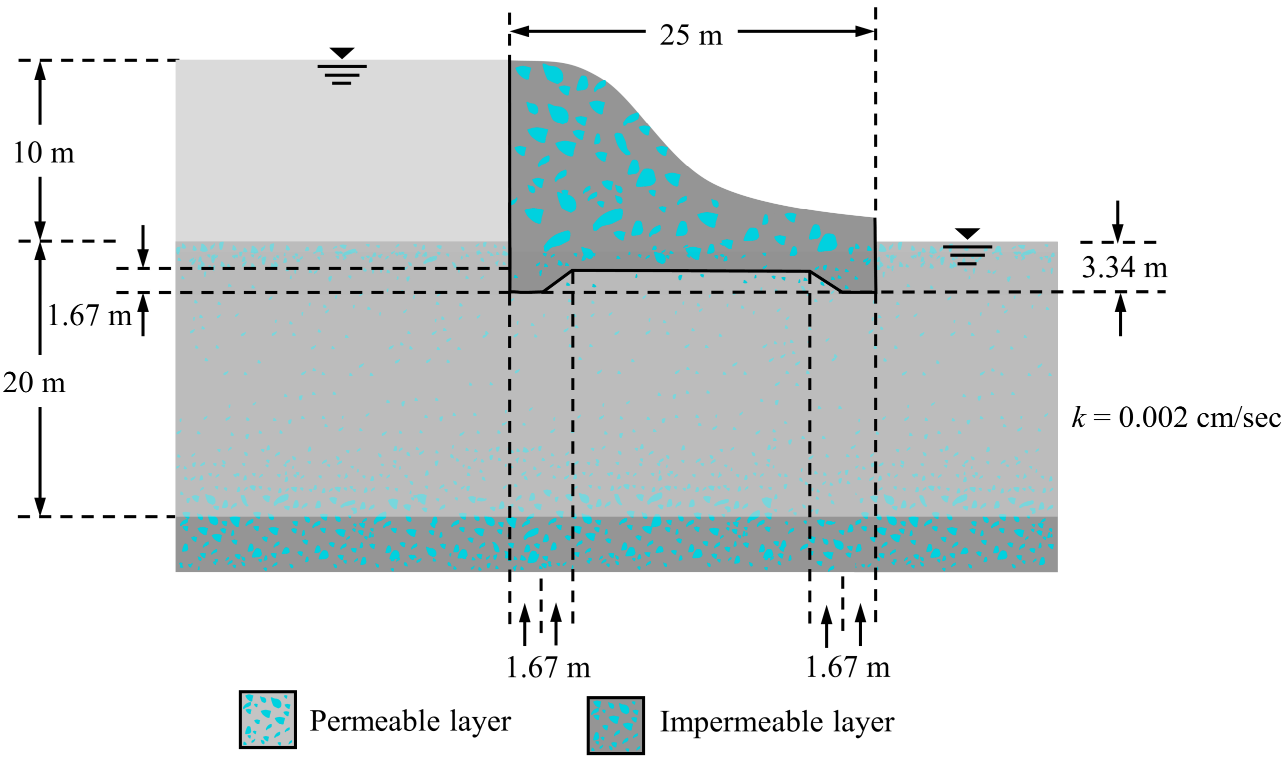
The calculated values are as follows:

The specific gravity of soil solids (G_s) *is approximately 2.44. The void ratio* (e) *of the soil is approximately 0.418.*

(4) Deficiencies:

Question:

For the hydraulic structure shown below, draw a flow net for flow through the permeable layer and calculate the seepage loss in day.



Original GPT-4 answer:

The image you've provided shows a hydraulic structure with a permeable layer over an impermeable layer, and various dimensions are provided for the geometry of the flow path.

To draw a flow net for the flow through the permeable layer, you would sketch flow lines that represent the path of the water particles, and equipotential lines that represent the locations where the hydraulic head is constant. The flow lines and equipotential lines should intersect at right angles, and ideally, the areas between them should resemble squares, indicating that the hydraulic gradient is the same in all parts of the flow field.

Given the complexity of this task, it is typically done manually on graph paper or using specialised software that allows you to draw curvilinear grids. The flow net is used to visualise the flow paths and to estimate the seepage loss by counting the number of flow channels and equipotential drops.

Here's how you would calculate the seepage loss:

1. *Count the number of flow channels (Ni) and equipotential drops (Nd) from the flow net.*
2. *Use Darcy's Law to calculate the seepage loss, with the formula* $Q = k \cdot N_f \cdot \frac{\Delta H}{N_d}$.
3. *Convert the seepage loss to*$m^3/day/m$ *by using the following conversion:*

$$Q_{m^3/day/m} = Q \cdot 86,400\ sec/day \cdot \frac{1\ m^2}{10,000\ cm^2}$$

Given that, you would need to provide the counts of Ni and Nd from your flow net to proceed with the seepage loss calculation. If you have these counts, please provide them, and we can continue with the calculation. If not, you will need to draw the flow net and count and Nd before we can proceed.

CoT prompting:

Solve the question and do it step by step.

GPT-4 answer after implementing the CoT prompting:

To solve this problem and calculate seepage loss through the permeable layer, we need to follow these steps:

1. *Draw a Flow Net: A flow net consists of flow lines and equipotential lines. Flow lines represent the path that water particles follow as they move through the soil, and equipotential lines are lines connecting points of equal hydraulic head.*
2. *Determine the Number of Flow Channels (Nf): This is the number of flow lines from the entry to the exit of the flow region.*
3. *Determine the Number of Equipotential Drops (Nd): This is the number of equipotential drops from the upstream to the downstream side.*
4. *Calculate the Hydraulic Gradient (i): It is the head loss per unit length of flow. The head loss is the difference in hydraulic head between two equipotential lines, and the length of flow is the perpendicular distance between two flow lines.*
5. *Use Darcy's Law to Calculate Seepage: Darcy's law in the form of Q = kiA can be used, where Q is the discharge, k is the hydraulic conductivity, i is the hydraulic gradient, and A is the cross-sectional area through which flow occurs.*

For the specific problem at hand, I'll first draw a flow net based on the given diagram:

1. *Drawing Flow Lines and Equipotential Lines: Flow lines should curve to intersect the impermeable boundaries at right angles, and the equipotential lines should intersect the flow lines at right angles as well.*
2. *Calculating the Seepage Loss: Once we have the flow net, we can determine Nf and Nd, and then use the values of k (given as 0.002 cm/s) and the dimensions of the permeable layer to calculate the seepage loss.*

However, since we're operating in a text-based environment, I'll describe the process of drawing the flow net conceptually, and we can then perform the necessary calculations for the seepage loss.

Conceptual Steps for Drawing a Flow Net*:*

- *Identify entry and exit points for water in the permeable layer.*

- *Draw flow lines starting from the water entry point, curving and ending at the water exit point. These lines represent the path of water through the permeable material.*
- *Draw equipotential lines starting from the upstream side (high head) to the downstream side (low head). The first equipotential line is at the upstream face where the head is maximum, and the last equipotential line is at the downstream face where the head is minimum.*

Once we have the flow net conceptually drawn, we can proceed to calculate the seepage loss. Since this is a visual task and we're working without visual aids here, we'll skip the drawing part and move on to the calculation. To do so, I will need the following information, which might be obtained from the flow net if it was drawn:

- *The number of flow channels, Nf*
- *The number of equipotential drops, Nd*
- *The length of the flow path, L (in meters)*
- *The difference in head, Δh, from the upstream to downstream (in meters)*

Since I cannot physically draw the flow net here, let's assume typical values for Nf and Nd to demonstrate the calculation method. Let's say Nf = 5 and Nd = 10. Now let's calculate the seepage loss assuming these values. We will also assume the total head loss is equal to the height of the water above the permeable layer, which we can obtain from the diagram provided. The cross-sectional area through which the water flows is also needed.

Custom instruction: Not applicable.

References

1. Bahroun, Z.; Anane, C.; Ahmed, V.; Zacca, A. Transforming Education: A Comprehensive Review of Generative Artificial Intelligence in Educational Settings through Bibliometric and Content Analysis. *Sustainability* **2023**, *15*, 12983. [CrossRef]
2. Rane, N.; Choudhary, S.; Rane, J. Education 4.0 and 5.0: Integrating Artificial Intelligence (AI) for Personalized and Adaptive Learning. *SSRN Electron. J.* **2023**. [CrossRef]
3. Tan, S. Harnessing Artificial Intelligence for Innovation in Education. In *Learning Intelligence: Innovative and Digital Transformative Learning Strategies: Cultural and Social Engineering Perspectives*; Springer Nature: Singapore, 2023; pp. 335–363.
4. Erduran, S. AI is transforming how science is done. Science education must reflect this change. *Science* **1979**, *382*, 2023. [CrossRef] [PubMed]
5. Dwivedi, Y.K.; Kshetri, N.; Hughes, L.; Slade, E.L.; Jeyaraj, A.; Kar, A.K.; Baabdullah, A.M.; Koohang, A.; Raghavan, V.; Ahuja, M.; et al. 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. Bandi, A.; Adapa, P.V.S.R.; Kuchi, Y.E.V.P.K. The Power of Generative AI: A Review of Requirements, Models, Input–Output Formats, Evaluation Metrics, and Challenges. *Future Internet* **2023**, *15*, 260. [CrossRef]
7. García-Peñalvo, F.J.; Vázquez-Ingelmo, A. What Do We Mean by GenAI? A Systematic Mapping of The Evolution, Trends, and Techniques Involved in Generative AI. *Int. J. Interact. Multimed. Artif. Intell.* **2023**, *8*, 7–16. [CrossRef]
8. OpenAI. GPT-4 Technical Report. *arXiv* **2023**, arXiv:2303.08774.
9. Noy, S.; Zhang, W. Experimental evidence on the productivity effects of generative artificial intelligence. *Science* **1979**, *381*, 2023. [CrossRef]
10. Epstein, Z.; Hertzmann, A.; Akten, M.; Farid, H.; Fjeld, J.; Frank, M.R.; Groh, M.; Herman, L.; Leach, N.; Mahari, R.; et al. Art and the science of generative AI. *Science* **1979**, *380*, 2023. [CrossRef]
11. Epstein, Z.; Hertzmann, A.; Herman, L.; Mahari, R.; Frank, M.R.; Groh, M.; Schroeder, H.; Smith, A.; Akten, M.; Fjeld, J.; et al. Art and the science of generative AI: A deeper dive. *arXiv* **2023**, arXiv:2306.04141.
12. Mao, J.; Chen, B.; Liu, J.C. Generative Artificial Intelligence in Education and Its Implications for Assessment. *TechTrends* **2024**, *68*, 58–66. [CrossRef]
13. Franceschelli, G.; Musolesi, M. Reinforcement Learning for Generative AI: State of the Art, Opportunities and Open Research Challenges. *J. Artif. Intell. Res.* **2024**, *79*, 417–446. [CrossRef]
14. Khosravi, H.; Viberg, O.; Kovanovic, V.; Ferguson, R. Generative AI and Learning Analytics. *J. Learn. Anal.* **2023**, *10*, 1–6. [CrossRef]
15. Steele, J.L. To GPT or not GPT? Empowering our students to learn with AI. Computers and Education: Artificial Intelligence. *Comput. Educ. Artif. Intell.* **2023**, *5*, 100160. [CrossRef]
16. Bewersdorff, A.; Seßler, K.; Baur, A.; Kasneci, E.; Nerdel, C. Assessing student errors in experimentation using artificial intelligence and large language models: A comparative study with human raters. *Comput. Educ. Artif. Intell.* **2023**, *5*, 100177. [CrossRef]
17. Farhi, F.; Jeljeli, R.; Aburezeq, I.; Dweikat, F.F.; Al-shami, S.A.; Slamene, R. Analyzing the students' views, concerns, and perceived ethics about chat GPT usage. *Comput. Educ. Artif. Intell.* **2023**, *5*, 100180. [CrossRef]

18. Relmasira, S.C.; Lai, Y.C.; Donaldson, J.P. Fostering AI Literacy in Elementary Science, Technology, Engineering, Art, and Mathematics (STEAM) Education in the Age of Generative AI. *Sustainability* **2023**, *15*, 13595. [CrossRef]
19. Bai, X.; Stede, M. A Survey of Current Machine Learning Approaches to Student Free-Text Evaluation for Intelligent Tutoring. *Int. J. Artif. Intell. Educ.* **2022**, *33*, 992–1030. [CrossRef]
20. Su, J.; Yang, W. Unlocking the Power of ChatGPT: A Framework for Applying Generative AI in Education. *ECNU Rev. Educ.* **2023**, *6*, 355–366. [CrossRef]
21. Jauhiainen, J.S.; Guerra, A.G. Generative AI and ChatGPT in School Children's Education: Evidence from a School Lesson. *Sustainability* **2023**, *15*, 14025. [CrossRef]
22. Eysenbach, G. The Role of ChatGPT, Generative Language Models, and Artificial Intelligence in Medical Education: A Conversation with ChatGPT and a Call for Papers. *JMIR Med. Educ.* **2023**, *9*, e46885. [CrossRef] [PubMed]
23. Rahman, M.M.; Watanobe, Y. ChatGPT for Education and Research: Opportunities, Threats, and Strategies. *Appl. Sci.* **2023**, *13*, 5783. [CrossRef]
24. Halaweh, M. ChatGPT in education: Strategies for responsible implementation. *Contemp. Educ. Technol.* **2023**, *15*, ep421. [CrossRef]
25. Schulze Balhorn, L.; Weber, J.M.; Buijsman, S.; Hildebrandt, J.R.; Ziefle, M.; Schweidtmann, A.M. Empirical assessment of ChatGPT's answering capabilities in natural science and engineering. *Sci. Rep.* **2024**, *14*, 4998. [CrossRef] [PubMed]
26. Zhang, Z.; Wen, L.; Jiang, Y.; Liu, Y. Evaluate Chat-GPT's programming capability in Swift through real university exam questions. In *Softw: Pract Exper*; Wiley: Hoboken, NJ, USA, 2024; pp. 1–15. [CrossRef]
27. Katz, D.M.; Bommarito, M.J.; Gao, S.; Arredondo, P. GPT-4 passes the bar exam. Philosophical Transactions of the Royal Society A: Mathematical. *Phys. Eng. Sci.* **2024**, *382*, 20230254. [CrossRef] [PubMed]
28. Maitland, A.; Fowkes, R.; Maitland, S. Can ChatGPT pass the MRCP (UK) written examinations? Analysis of performance and errors using a clinical decision-reasoning framework. *BMJ Open* **2024**, *14*, e080558. [CrossRef]
29. Currie, G.; Singh, C.; Nelson, T.; Nabasenja, C.; Al-Hayek, Y.; Spuur, K. ChatGPT in medical imaging higher education. *Radiography* **2023**, *29*, 792–799. [CrossRef] [PubMed]
30. Ali, R.; Tang, O.Y.; Connolly, I.D.; Zadnik Sullivan, P.L.; Shin, J.H.; Fridley, J.S.; Asaad, W.F.; Cielo, D.; Oyelese, A.A.; Doberstein, C.E.; et al. Performance of ChatGPT and GPT-4 on Neurosurgery Written Board Examinations. *Neurosurgery* **2023**, *93*, 1353–1365. [CrossRef]
31. Kung, T.H.; Cheatham, M.; Medenilla, A.; Sillos, C.; De Leon, L.; Elepaño, C.; Madriaga, M.; Aggabao, R.; Diaz-Candido, G.; Maningo, J.; et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLoS Digit. Health* **2023**, *2*, e0000198. [CrossRef]
32. Gilson, A.; Safranek, C.W.; Huang, T.; Socrates, V.; Chi, L.; Taylor, R.A.; Chartash, D. How Does ChatGPT Perform on the United States Medical Licensing Examination? The Implications of Large Language Models for Medical Education and Knowledge Assessment. *JMIR Med. Educ.* **2023**, *9*, e45312. [CrossRef]
33. Antaki, F.; Touma, S.; Milad, D.; El-Khoury, J.; Duval, R. Evaluating the Performance of ChatGPT in Ophthalmology. *Ophthalmol. Sci.* **2023**, *3*, 100324. [CrossRef]
34. Terwiesch, C. Would Chat GPT3 Get a Wharton MBA? A Prediction Based on Its Performance in the Operations Management Course. Mack Institute for Innovation Management at the Wharton School, University of Pennsylvania. 2023. Available online: https://mackinstitute.wharton.upenn.edu/wp-content/uploads/2023/01/Christian-Terwiesch-Chat-GTP.pdf (accessed on 2 December 2023).
35. Eulerich, M.; Sanatizadeh, A.; Vakilzadeh, H.; Wood, D.A. Is it All Hype? ChatGPT's Performance and Disruptive Potential in the Accounting and Auditing Industries. *SSRN Electron. J.* 2023. [CrossRef]
36. Kumar, K. Geotechnical Parrot Tales (GPT): Harnessing Large Language Models in Geotechnical Engineering. *J. Geotech. Geoenviron. Eng.* **2024**, *150*, 02523001. [CrossRef]
37. Rane, N.; Choudhary, S.; Rane, J. Contribution of ChatGPT and Similar Generative Artificial Intelligence in Geotechnical Engineering and Soil Mechanics. *SSRN Electron. J.* **2024**. [CrossRef]
38. Fatahi, B.; Khabbaz, H.; Xue, J.; Hadgraft, R. Generative AI as a Catalyst for Enhanced Learning Experience in Engineering Education. In Proceedings of the AAEE 2023 Griffith University, Gold Coast, QLD, Australia, 3–6 December 2023.
39. Daniel, W.W.; Cross, C.L. *Biostatistics: A Foundation for Analysis in the Health Sciences*; John Wiley & Sons: Hoboken, NJ, USA, 2018.
40. Lwanga, S.K.; Lemeshow, S. *Sample Size Determination in Health Studies: A Practical Manual*; World Health Organization: Geneva, Switzerland, 1991.
41. Krathwohl, D.R. A Revision of Bloom's Taxonomy: An Overview. *Theory Pract.* **2002**, *41*, 212–218. [CrossRef]