

Evaluating the Performance of Artificial Intelligence Chatbots and Large Language Models in the FE and PE Structural Exams

M. Z. Naser, M.ASCE

Assistant Professor, School of Civil and Environmental Engineering and Earth Sciences, Clemson Univ., Clemson, SC 29631; Artificial Intelligence Research Institute for Science and Engineering (AIRISE), Clemson Univ., Clemson, SC 29631 (corresponding author). ORCID: <https://orcid.org/0000-0003-1350-3654>. Email: mznaser@clemson.edu

Brandon Ross, M.ASCE

Associate Professor, School of Civil and Environmental Engineering, and Earth Sciences, Clemson Univ., Clemson, SC 29631. ORCID: <https://orcid.org/0000-0003-2418-7298>. Email: bross2@clemson.edu

Jennifer Ogle, M.ASCE

Professor, School of Civil and Environmental Engineering, and Earth Sciences, Clemson Univ., Clemson, SC 29631. Email: ogle@clemson.edu

Venkatesh Kodur, F.ASCE

Professor, Dept. of Civil and Environmental Engineering, Michigan State Univ., East Lansing, MI 48824. Email: kodur@egr.msu.edu

Rami Hawileh, F.ASCE

Professor, Dept. of Civil Engineering, American Univ. of Sharjah, Sharjah 26666, United Arab Emirates. Email: rhaweeleh@aus.edu

Jamal Abdalla, F.ASCE

Professor, Dept. of Civil Engineering, American Univ. of Sharjah, Sharjah 26666, United Arab Emirates. Email: jabdalla@aus.edu

Huu-Tai Thai, M.ASCE

Professor, Dept. of Infrastructure Engineering, Univ. of Melbourne, Melbourne, Parkville, VIC 3052, Australia. Email: tai.thai@unimelb.edu.au

Forum papers are thought-provoking opinion pieces or essays founded in fact, sometimes containing speculation, on a civil engineering topic of general interest and relevance to the readership of the journal. The views expressed in this Forum article do not necessarily reflect the views of ASCE or the Editorial Board of the journal.

<https://doi.org/10.1061/PPSCFX.SCENG-1369>

The engineering community has recently witnessed the emergence of chatbot technology with the release of OpenAI ChatGPT-4 and Google Bard. These chatbots have been reported to perform well and even pass various standardized tests, including medical and law exams, and this forum paper explores whether these chatbots can also pass the Fundamentals of Engineering (FE) and Principles and Practice of Engineering (PE) exams. A diverse range of civil and environmental engineering questions and scenarios that commonly are present in the FE and PE exams was used to evaluate the chatbots' performance. The chatbots' responses were analyzed

based on their relevance, accuracy, and clarity, and then compared with the recommendations of the National Council of Examiners for Engineering and Surveying (NCEES). Our report shows that the February 2023 editions of ChatGPT-4 and Bard scored 70.9% and 39.2% in the FE exam and 46.2% and 41% in the PE exam. The October 2023 edition of ChatGPT-4 is capable of scoring more than 70% in both exams. It is evident that the current version of ChatGPT-4 potentially could pass the FE and PE exams. Although future editions are much more likely to pass both exams, this study also highlights the potential of using chatbots as teaching assistants and guiding engineers.

What Is a Chatbot?

A chatbot is a computer program that is designed to be capable of simulating conversation with humans (Adamopoulou and Moussiades 2020). ChatGPT-4 and Bard were created and released recently by OpenAI (2023) and Google (2023), respectively. When asked to define a chatbot, ChatGPT-4 and Bard answered

A chatbot is an artificial intelligence (AI) software program designed to simulate human-like conversations with users, usually through text or voice interactions. Chatbots can understand and process natural language inputs, such as text messages or spoken words, and generate appropriate responses based on their underlying algorithms and knowledge base. They are typically used to provide customer support, answer frequently asked questions, assist with various tasks, or offer personalized recommendations. (ChatGPT-4)

Chatbots can be built using different natural language processing (NLP) techniques and machine learning algorithms, which allow them to understand and interpret the user's intent and context. Over time, chatbots can learn and improve their responses by analyzing user interactions and incorporating new knowledge into their systems. They are commonly integrated into messaging platforms, mobile applications, websites, and virtual assistants, providing users with a convenient and efficient way to access information and services. (Bard)

Furthermore, according to Bard

A chatbot is a computer program that simulates human conversation. It can be used to answer questions, provide information, or even complete tasks. Chatbots are often used in customer service, where they can answer questions about products or services without the need for a human representative. They can also be used in marketing, where they can generate leads or provide information about products or services.

Chatbots are typically programmed using a natural language processing (NLP) system. This allows them to understand and respond to human language. NLP systems are trained on large

amounts of data, such as text or transcripts of conversations. This allows them to learn how to understand and respond to human language in a natural way.

Chatbots are becoming increasingly popular as businesses look for ways to automate customer service and marketing tasks. They can be a valuable tool for businesses of all sizes, and they are likely to continue to grow in popularity in the years to come.

The same chatbots also were asked to define the Fundamentals of Engineering (FE) and Principles and Practice of Engineering (PE) exams. Their responses align well with the formal definitions of these exams, as noted by the National Council of Examiners for Engineering and Surveying (NCEES 2023b) and others (Johnston et al. 2007). According to ChatGPT-4

The Fundamentals of Engineering (FE) and Principles and Practice of Engineering (PE) exams are standardized examinations administered by the National Council of Examiners for Engineering and Surveying (NCEES) in the United States. These exams are designed to assess the knowledge and skills of engineering professionals and are part of the process to obtain a Professional Engineer (PE) license.

- Fundamentals of Engineering Exam: The FE exam is the first step in the process of becoming a licensed professional engineer. It is designed for recent graduates and students who are close to finishing an undergraduate engineering degree from an EAC/ABET-accredited program. The FE exam is a computer-based test that covers various engineering disciplines, such as civil, mechanical, electrical, and chemical engineering. Passing the FE exam signifies that the candidate has demonstrated a fundamental understanding of engineering principles and is ready to work as an Engineer-in-Training (EIT) or Engineering Intern (EI).
- Principles and Practice of Engineering Exam: The PE exam is the second step in the process of becoming a licensed professional engineer. It is designed for engineers who have gained a minimum of four years of post-college work experience in their chosen engineering discipline. The PE exam tests the candidate's competency.

According to Bard

The Fundamentals of Engineering (FE) and Principles and Practice of Engineering (PE) exams are the first and second steps, respectively, in the process of becoming a licensed professional engineer (PE). The FE exam is a computer-based exam that covers a broad range of engineering topics, while the PE exam is a more specialized exam that covers a specific area of engineering.

The FE exam is typically taken by recent graduates or students who are close to finishing an undergraduate engineering degree. The PE exam is typically taken after four years of professional engineering experience.

The FE exam is a required step for becoming a PE, but it is not a guarantee that you will pass the PE exam. The PE exam is a much more difficult exam, and it is important to be well-prepared before taking it.

There are many resources available to help you prepare for the FE and PE exams. These resources include books, online courses, and practice exams. It is important to find a resource that fits your learning style and that will help you cover the material that is on the exams.

It is also important to practice taking exams. The FE and PE exams are both multiple-choice exams, and it is important to get used to the format of the exams. There are many online practice exams available, and you can also find practice exams in books and online courses.

Taking the FE and PE exams is a challenging process, but it is a rewarding one. If you are committed to becoming a PE, then you should start preparing for the FE and PE exams as soon as possible.

Approach and Findings

The National Council of Examiners for Engineering and Surveying regulates the FE and PE exams (NCEES 2023b). According to the NCEES, the FE exam is a 6-h exam containing 110 questions, and exam takers often have 5 h 20 min to complete it. The FE exam covers a variety of topics spanning 18 areas. These are listed herein, along with the expected number of questions noted by the NCEES FE civil practice exam.

In contrast, the Principles and Practice of Engineering exam tests for a minimum level of competency in a particular engineering discipline. This exam is designed for engineers who have gained a minimum of 4 years of postcollege work experience in a particular civil engineering discipline. The PE exam includes 80 questions, and can be completed in 8 h. This study used the Structural Engineering discipline of the PE exam.

In our examination, we used the same questions recommended by the NCEES' practice exams available for FE and PE exam takers. Each chatbot was asked a single question at a time. Each answer then was compared with the recommended solutions provided by the NCEES FE and PE practice exams and graded separately. The grading followed a Correct–Not correct format (i.e., without partial points), which is similar to the grading scheme used in those exams. The total number of correct and incorrect answers was compiled at the end of the process. We did not apply any curving to adjust the scores. The complete examination process took 2 h.

The actual questions are not listed herein to maintain the NCEES' copyrights. However, a general discussion is provided to describe the outcome of this analysis. Both the FE and PE exams provide some visual-based questions. Some such questions simply were transcribed in order to be given to the chatbots using the February 2023 editions of ChatGPT-4 and Bard. With the release of the Vision edition of ChatGPT-4 (October 2023), some such questions were revisited to allow the chatbot to visualize and answer the questions. Overall, each chatbot was examined with 79 FE questions and 39 PE questions. Although this examination did not conform to actual FE and PE exams, the NCEES (2023a) states that

NCEES practice exams help you determine what you know—and what you don't know—before taking the exam. NCEES practice exams

- Contain questions from past exams
- Simulate the format, style, and level of difficulty
- Provide solutions or feedback on performance in each topic area

In our analysis, we provide some samples of correct and incorrect answers given by the chatbots. Some of the values were changed slightly to conceal the actual questions listed by the FE and PE practice exams.

FE Exam

The 18 areas tested in the FE exam are listed in Fig. 1. These areas cover general engineering themes. Altogether, each area has five different questions, and a handful of areas provide more than five questions. For example, Geotechnical Engineering, Hydraulics and Hydraulic Systems, Mathematics, Statics, and Mechanics of Materials account for the majority of the questions (Fig. 1).

Fig. 1 shows the number and distribution of questions in a typical FE exam and in our analysis. There was good convergence among most of the areas, with the exception of Surveying and Environmental Engineering, which aids in validating our analysis. This difference between these two areas was due to the fact that many of the questions presented in NCEES FE practice exam were in the form of images/charts and could not be easily transcribed to each chatbot.

Fig. 2 shows the performance of ChatGPT-4 and Bard in each area. ChatGPT-4 outperformed Bard in every entry. ChatGPT-4 did not perform below 50% except in the Structural Design and Geotechnical Engineering areas. Conversely, Bard consistently scored less than 50%, and did not answer any questions correctly in five areas (Probability and Statistics, Computational Tools,

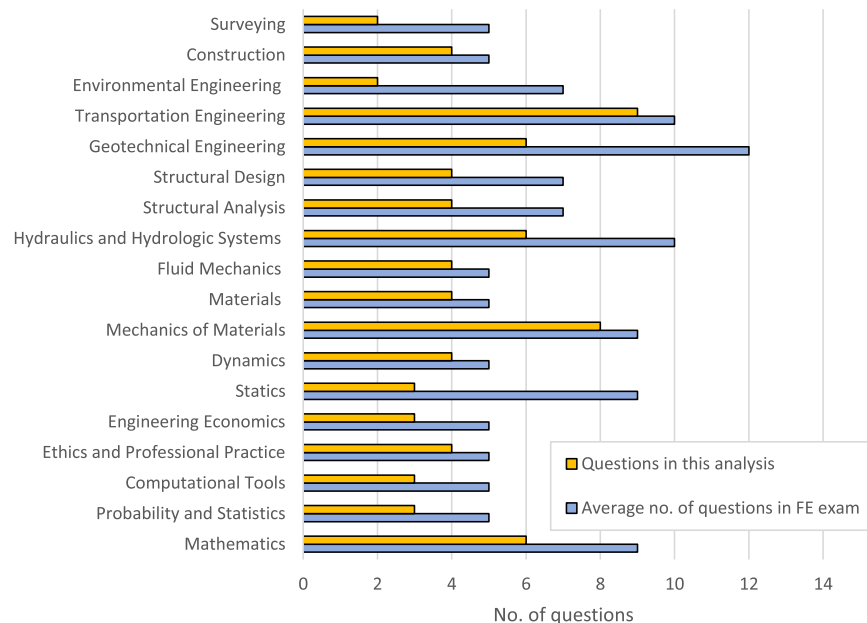


Fig. 1. Number of questions in a typical FE exam and in our analysis.

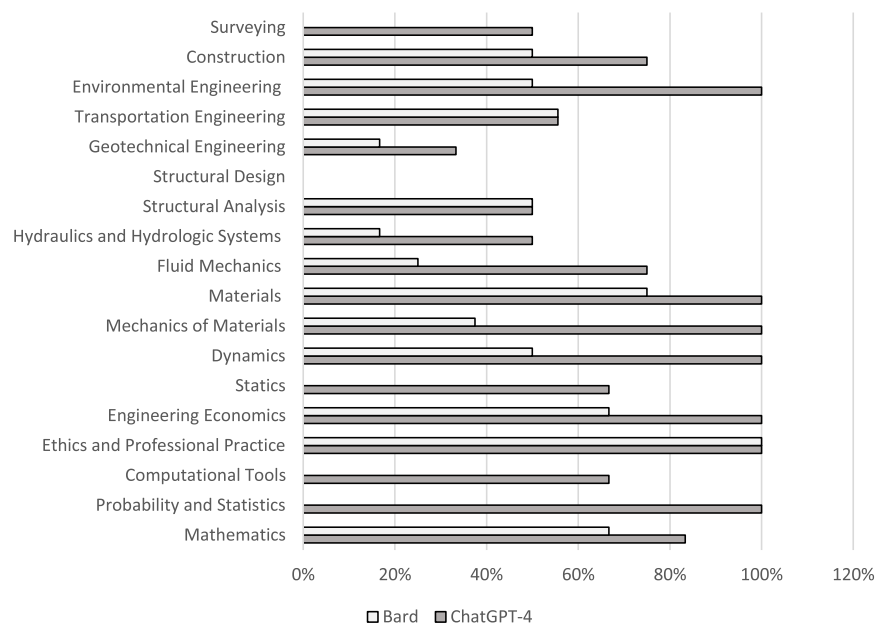


Fig. 2. Results of FE exam analysis.

Statics, Structural Design, and Surveying). Neither chatbot answered any question correctly in the area of Structural Design, but they answered all questions on Ethics and Professional Practice correctly.

The findings of our analysis indicate that ChatGPT-4 and Bard scored 70.9% and 39.2% in the FE exam, respectively. Although the passing score for the FE exam is not commonly announced, the score historically has been believed to be about 60%–70%. This implies that the current version of ChatGPT-4 potentially could pass this exam, and it is likely that future versions of ChatGPT-4 will be able to pass a complete and standard version of this exam as well. This rationale stems from the fact that six of the incorrect answers given by ChatGPT-4 were close to the correct answers listed by the practice exams—yet they were marked as wrong answers. For example, for one question, ChatGPT-4 laid out the complete process with correct formulae and logic, but had a calculation mistake that skewed the final answer. In another example, ChatGPT-4 did not convert the units properly. Each question addressed by ChatGPT-4 was answered with complete thoughts and step-by-step processes and equations. Two sample questions derived from those in the FE practice and the answers given by ChatGPT-4 are as follows:

Q1. The area of the shaded portion of the following equation ($y = 5x^3 + 1$) between $x = 4$ and $x = 9$ is most nearly:

To find the area of the shaded portion of the curve $y = 5x^3 + 1$ between $x = 4$ and $x = 9$, we need to integrate the function with respect to x over the given interval.

The definite integral of the function $y = 5x^3 + 1$ from $x = 4$ to $x = 9$ is:

$$\int (5x^3 + 1) dx \text{ from } 4 \text{ to } 9$$

To evaluate the integral, first find the antiderivative of the function:

Antiderivative of $5x^3 + 1$ is $(5/4)x^4 + x + C$, where C is the constant of integration. However, since we're computing a definite integral, we don't need to worry about the constant.

Now, evaluate the antiderivative at the given limits:

$$[(5/4)(9^4) + 9] - [(5/4)(4^4) + 4]$$

Calculating the values:

$$(5/4)(6561) + 9 - [(5/4)(256) + 4] \approx 8190.25 - 324 \approx 7866.25$$

The area of the shaded portion between $x = 4$ and $x = 9$ is approximately 7866.25 square units.

Q2. The magnitude (N) of the resultant of the three coplanar forces, A, B, and C, where has an angle of 115 degrees and force of 2N, B angle of 90 degrees and force of 4.13 N, and C with angle of 5 and force of 1.15 N is most nearly:

To find the magnitude of the resultant of the three coplanar forces A, B, and C, we can break down each force into its x and y components and then sum the components in each direction. Finally, we can calculate the magnitude of the resultant using the Pythagorean theorem.

Force A:

Magnitude: 2 N

Angle: 115 degrees

Force B:

Magnitude: 4.13 N

Angle: 90 degrees

Force C:

Magnitude: 1.15 N

Angle: 5 degrees

Force A components:

$$A_x = 2 * \cos(115^\circ) \approx -1.81 \text{ N}$$

$$A_y = 2 * \sin(115^\circ) \approx 1.91 \text{ N}$$

Force B components:

$$B_x = 4.13 * \cos(90^\circ) = 0 \text{ N}$$

$$B_y = 4.13 * \sin(90^\circ) = 4.13 \text{ N}$$

Force C components:

$$C_x = 1.15 * \cos(5^\circ) \approx 1.14 \text{ N}$$

$$C_y = 1.15 * \sin(5^\circ) \approx 0.10 \text{ N}$$

Sum of x -components:

$$R_x = A_x + B_x + C_x = -1.81 + 0 + 1.14 \approx -0.67 \text{ N}$$

Sum of y -components:

$$R_y = A_y + B_y + C_y = 1.91 + 4.13 + 0.10 \approx 6.14 \text{ N}$$

Now, we can calculate the magnitude of the resultant force (N) using the Pythagorean theorem:

$$N = \sqrt{(R_x)^2 + (R_y)^2} = \sqrt{((-0.67)^2 + (6.14)^2)} \approx \sqrt{(0.4489 + 37.6996)} \approx \sqrt{(38.1485)}$$

$$N \approx 6.17$$

The magnitude of the resultant of the three coplanar forces A, B, and C is most nearly 6.17 N.

These solutions are clear and concise.

PE Exam

Fig. 3 lists the areas covered in a typical PE exam. There are 11 different areas in the Structural-based PE exam. For transparency, the PE examination is much more complicated than that of the FE exam. This is due to the fact that the PE exam contains a larger number of questions with intricate visuals (i.e., design detailing, loading configurations, and so forth) that may not be transcribed easily. Therefore, the number of questions explored in this analysis was significantly less than the number expected in the actual PE exam (Fig. 3).

Fig. 4 breaks down the performance of ChatGPT-4 and Bard across each area. In the case of this exam, ChatGPT-4 and Bard performed equally (46.2% and 41%, respectively), and below the 50% mark. ChatGPT-4 performed exceptionally well in the area of Materials and Site Development, whereas Bard performed exceptionally well in Structural Mechanics and Site Development. In terms of Analysis of Structures, Design and Details of Structures, and Codes and Construction, ChatGPT-4 and Bard scores ranged from 25% to 67%. Unlike in the case of the FE exam, the current versions of ChatGPT-4 and Bard are unlikely to pass this exam. We anticipate that future versions of these chatbots will be able to pass this exam.

A few interesting observations were arrived at while examining the responses received from ChatGPT-4 and Bard. For example, both chatbots seemed to be able to understand the solution process behind a traditional design and analysis problem. However, the chatbots struggled to distinguish between uniform and point loads. On one occasion, Bard used the elastic deformation equation for a point load acting on a beam to calculate the elastic deflection of a uniform load. Conversely, ChatGPT-4 suffered from a simple calculation mistake (see Q4.) that led to a calculation error despite having the process procedure and equations. On a more positive note, both chatbots could cite ACI and ASCE properly on most occasions. Two sample questions derived from those in the FE practice exams are as follows:

Q3. For the following five 28-day average standard-cured concrete strength tests:

A, B, C, D, and E with a required strength of G ksi, select why the concrete is unsatisfactory:

- (A) Because the average of three consecutive tests is less than G ksi.
- (B) Because Test D did not meet the required strength.
- (C) Because the average of the first three tests is greater than the required strength.
- (D) Because Test A is within 500 psi of the required strength.

According to ACI 318, the concrete is considered satisfactory if both of the following conditions are met:

- The average of any three consecutive strength tests equals or exceeds the specified compressive strength (f'_c).
- No individual strength test result falls below the specified compressive strength by more than 500 psi.

Let's evaluate the test results:

Test A psi

Test B psi

Test C psi

Test D psi

Test E psi

We check the average of three consecutive strength tests.

The average of any three consecutive strength tests equals or exceeds the specified compressive strength (f'_c):

Average of Tests A, B, and C (satisfies the criterion)

Average of Tests B, C, and D (does not satisfy the criterion)

Average of Tests C, D, and e (does not satisfy the criterion)

- No individual strength test result falls below the specified compressive strength by more than 500 psi:
- Test C is 375 psi below the required strength (satisfies the criterion)

Based on these evaluations, the concrete is unsatisfactory because the average of the last two sets of three consecutive tests does not meet the required strength.

Thus, the correct statement is:

(A) Because the average of the last three consecutive tests is less than the required strength.

Q4. The maximum flexural stress (psi) in a 20 ft long simply supported rectangular member with a width and height of 4 in and 12 in, loaded with a uniform load of 200 lb/ft and a point of 1000 lb at 10 ft from the support is most nearly:

To find the maximum flexural stress, we'll first calculate the maximum moment for both the uniform load and point load and then use the flexural stress formula to find the maximum stress.

Maximum moment due to uniform load:

Uniform load, $w = 200$ lb/ft

Length, $L = 20$ ft

For a simply supported beam with a uniform load, the maximum moment (M_1) occurs at the center of the span:

$$M_1 = (w * L^2) / 8$$

$$M_1 = (200 \text{ lb/ft} * (20 \text{ ft})^2) / 8$$

$$M_1 = 10,000 \text{ lb}\cdot\text{ft}$$

Maximum moment due to point load:

Point load, $P = 1000$ lb

Distance from support, $a = 10$ ft

Length, $L = 20$ ft

For a simply supported beam with a point load, the maximum moment (M_2) is given by the formula:

$$M_2 = (P * a * (L - a)) / L$$

$$M_2 = (1000 \text{ lb} * 10 \text{ ft} * (20 \text{ ft} - 10 \text{ ft})) / 20 \text{ ft}$$

$$M_2 = 5,000 \text{ lb}\cdot\text{ft}$$

Now, we need to find the maximum combined moment. Since both loads are on the same beam, we'll simply add the two moments to get the maximum combined moment:

$$M_{\text{total}} = M_1 + M_2$$

$$M_{\text{total}} = 10,000 \text{ lb}\cdot\text{ft} + 5,000 \text{ lb}\cdot\text{ft}$$

$$M_{\text{total}} = 15,000 \text{ lb}\cdot\text{ft}$$

Now, we'll use the flexural stress formula to find the maximum stress:

$$\text{Flexural stress} = (M * S) / I \text{ [Note from authors: Flexural stress} = M / S]$$

Where:

M is the maximum moment

S is the section modulus

I is the moment of inertia

First, we'll calculate the moment of First, we'll calculate the moment of inertia (I) for the rectangular section:

$$I = (b * h^3) / 12$$

Where b is the width (4 in) and h is the height (12 in).

$$I = (4 \text{ in} * (12 \text{ in})^3) / 12$$

$I = 6912 \text{ in}^4$ [Note from authors, the calculation mistake occurs here, I should equal 576 in^4]

Now, we'll calculate the section modulus (S):

$$S = I / (h / 2)$$

$$S = 6912 \text{ in}^4 / (12 \text{ in} / 2)$$

$$S = 1152 \text{ in}^3$$

Finally, we'll calculate the maximum flexural stress using the formula:

$$\text{Flexural stress} = (M * S) / I \text{ [Note from authors, Flexural stress} = M / S]$$

Since the maximum moment is in ft, we need to convert it to inches:

$$M_{\text{total}} = 15,000 \text{ lb}\cdot\text{ft} * 12 \text{ in/ft} = 180,000 \text{ lb}\cdot\text{in}$$

$$\text{Flexural stress} = (180,000 \text{ lb}\cdot\text{in}) / 1152 \text{ in}^3$$

$$\text{Flexural stress} \approx 156.25 \text{ psi}$$

So, the maximum flexural stress is most nearly 156.25 psi.

Where to Go from Here: Chatbots as Teaching Assistants and Engineers

There are critical similarities and differences between ChatGPT-4 and a human engineer in terms of taking these exams. For example, neither ChatGPT-4 nor the engineer have access to the internet (unlike Bard). However, the engineer would have access to plenty of resources and codes. In one instance, ChatGPT-4 stated that it does not have access to values embedded in a specific chart and hence could not solve such a question. Other factors, such as lucky guessing, fatigue, and stress, are likely to take a toll on human behavior. Overall, the AI-enabled examination process for both the FE and PE took less than 2 h.

In one instance, ChatGPT-4 was able to solve a PE question and arrive at the exact solution value. The chatbots noted that none of the listed answers were correct, because the listed correct answer was a rounded value (as typically is noted in the PE exam). In another instance, the chatbot was able to correctly derive the complete process and solve a number of PE problems. However, the chatbot either copied one wrong value from a question or used an outdated equation to solve the question, and therefore was not able to correctly answer the question. We anticipate that future ChatGPT-4 and Bard generations will be able to serve as suitable teaching assistants and guides for students and young engineers, as noted

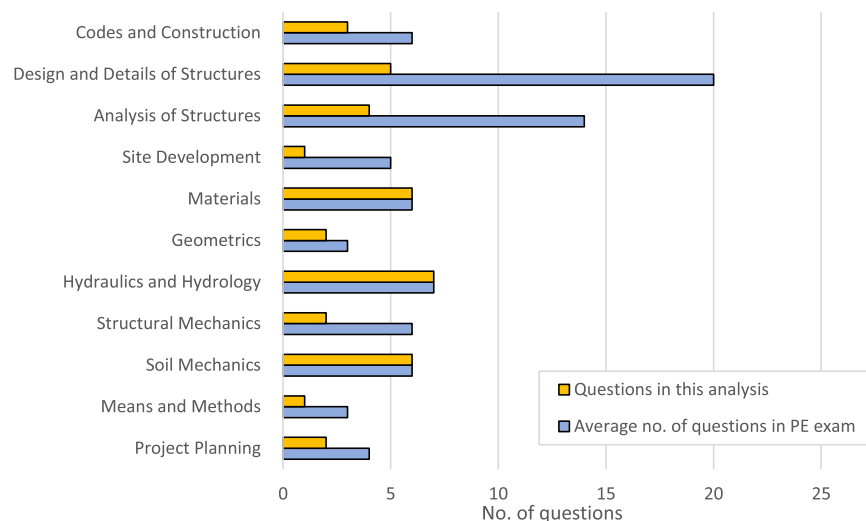


Fig. 3. Number of questions in a typical PE exam and in our analysis.

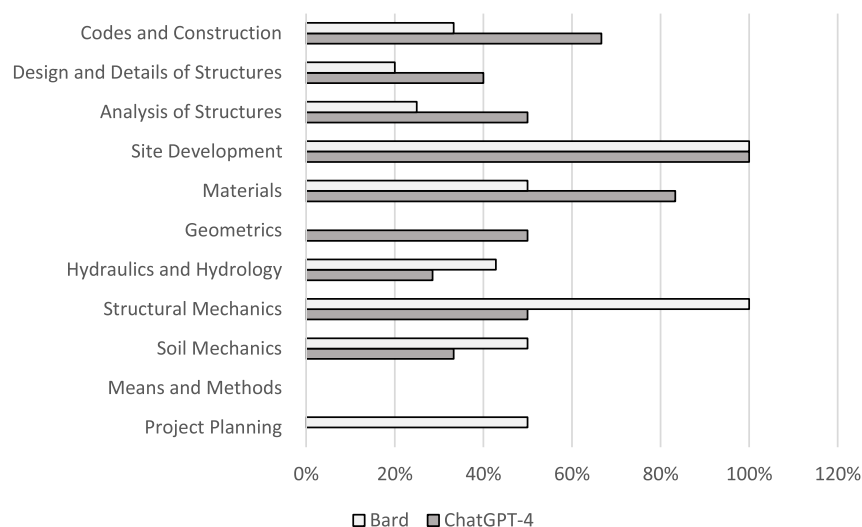


Fig. 4. Results of PE exam analysis.

from the improvements observed in the upgraded Vision edition of ChatGPT-4. The lead author has experimented with using ChatGPT-4 as an example-bank generator to create a variety of solved examples for the courses Statics, Reinforced Concrete Design, and Machine Learning and Causality for Engineers that currently are offered at Clemson University. The lead author reports success and positive feedback (so far) from students attending these courses.

Clearly, ChatGPT-4 and Bard are not without error—they have learned from human and textual processes. Much like human error occurring while entering data into a calculator or computer program, similar simple errors can occur with these artificial intelligence tools. The differentiator is the communicative nature of AI. The detailed step-by-step operations have merit for training young engineers. The interesting challenge is to identify use scenarios. How can you use these tools to help students who struggle to understand the details of the process? Can you give homework problems with solutions worked out and ask the students to verify the accuracy? Can these tools substitute for tutors, and how can we train improvements in responses to avoid the simple errors? Should not we all be concerned about an error that has a magnitude $10\times$ greater or less than the correct response? Can we use the

resulting failure or cost impact of that magnitude of error to make clear the importance of truly understanding disciplinary knowledge? AI is the new pocket calculator, but its accuracy and precision are still to be determined. As a profession, we must embrace the tool and ensure that its use does not negatively impact the health, safety, and welfare of the public.

Data Availability Statement

No data, models, or code were generated or used during the study.

References

- Adamopoulou, E., and L. Moussiades. 2020. "An overview of chatbot technology." In *IFIP advances in information and communication technology*, 373–383. Berlin: Springer. <https://doi.org/10.1007/978-3-030-49186-4>.
- Google. 2023. "Bard FAQ." Accessed February 23, 2023. <https://bard.google.com/faq>.
- Johnston, D. W., N. T. Ahluwalia, and M. B. Gwyn. 2007. "Improving the professional engineering licensure process for construction engineers"

J. Constr. Eng. Manage. 133 (9): 669–677. [https://doi.org/10.1061/\(ASCE\)0733-9364\(2007\)133:9\(669\)](https://doi.org/10.1061/(ASCE)0733-9364(2007)133:9(669)).
NCEES (National Council of Examiners for Engineering and Surveying). 2023a. “Exam prep.” Accessed February 23, 2023. <https://ncees.org/exams/exam-preparation-materials/>.

NCEES (National Council of Examiners for Engineering and Surveying). 2023b. “National Council of Examiners for Engineering and Surveying (NCEES).” Accessed February 23, 2023. <https://ncees.org/>.
OpenAI. 2023. “ChatGPT—Release notes.” Accessed February 23, 2023. <https://help.openai.com/en/articles/6825453-chatgpt-release-notes>.