

Please cite this paper as:

Naser M.Z., Ross B., Ogle J., Kodur V., Hawileeh R., Abdall J., Thai T. (2024). Evaluating the Performance of Artificial Intelligence Chatbots and Large Language Models in the FE and PE Structural Exams. *ASCE Journal of Practice Periodical on Structural Design and Construction*. <https://doi.org/10.1061/PPSCFX.SCENG-1369>.

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

M.Z. Naser^{1,2}, Brandon Ross³, Jennifer Ogle⁴, Venkatesh Kodur⁵, Rami Hawileh⁶, Jamal Abdalla⁶, Huu-Tai Thai⁷

¹School of Civil and Environmental Engineering and Earth Sciences, Clemson University, USA

²Artificial Intelligence Research Institute for Science and Engineering (AIRISE), Clemson University, USA

E-mail: mznaser@clemson.edu, Website: www.mznaser.com

³School of Civil and Environmental Engineering, and Earth Sciences, Clemson University, USA

E-mail: brross2@clemson.edu

⁴School of Civil and Environmental Engineering, and Earth Sciences, Clemson University, USA

E-mail: ogle@clemson.edu

⁵Department of Civil and Environmental Engineering, Michigan State University, USA.

E-mail: kodur@egr.msu.edu

⁶Department of Civil Engineering, American University of Sharjah, UAE.

E-mail: rhaweeleh@aus.edu

⁶Department of Civil Engineering, American University of Sharjah, UAE.

E-mail: jabdalla@aus.edu

⁷Department of Infrastructure Engineering, The University of Melbourne, Australia.

E-mail: tai.thai@unimelb.edu.au

The engineering community has recently witnessed the emergence of chatbot technology with the release of OpenAI *ChatGPT-4* and Google *Bard*. While these chatbots have been reported to perform well and even pass various standardized tests, including medical and law exams, 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 are used to evaluate the chatbots' performance, as commonly present in the FE and PE exams. The chatbots' responses were analyzed based on their relevance, accuracy, and clarity and then compared against 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. We also note that 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* could potentially pass the FE and PE exams. While 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 recently released by OpenAI (2023) and Google (2023). When asked to define what a chatbot is, *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

Please cite this paper as:

Naser M.Z., Ross B., Ogle J., Kodur V., Hawileeh R., Abdall J., Thai T. (2024). Evaluating the Performance of Artificial Intelligence Chatbots and Large Language Models in the FE and PE Structural Exams. *ASCE Journal of Practice Periodical on Structural Design and Construction*. <https://doi.org/10.1061/PPSCFX.SCENG-1369>.

42 underlying algorithms and knowledge base. They are typically used to provide customer support, answer
43 frequently asked questions, assist with various tasks, or offer personalized recommendations.

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

49 **And according to *Bard*,**

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

54 Chatbots are typically programmed using a natural language processing (NLP) system. This allows them to
55 understand and respond to human language. NLP systems are trained on large amounts of data, such as text
56 or transcripts of conversations. This allows them to learn how to understand and respond to human language
57 in a natural way.

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

61 The same chatbots were also asked to define the Fundamentals of Engineering (FE) and Principles
62 and Practice of Engineering (PE) exams. Below are their responses, and as one can see, these
63 responses align well with the formal definitions of these exams, as noted by the National Council
64 of Examiners for Engineering and Surveying (NCEES, 2023a) and others (Johnston et al. 2007).

65 ***ChatGPT-4:***

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

70 • Fundamentals of Engineering (FE) Exam:

71 The FE exam is the first step in the process of becoming a licensed professional engineer. It is
72 designed for recent graduates and students who are close to finishing an undergraduate
73 engineering degree from an EAC/ABET-accredited program. The FE exam is a computer-based test
74 that covers various engineering disciplines, such as civil, mechanical, electrical, and chemical
75 engineering. Passing the FE exam signifies that the candidate has demonstrated a fundamental
76 understanding of engineering principles and is ready to work as an Engineer-in-Training (EIT) or
77 Engineering Intern (EI).

78 • Principles and Practice of Engineering (PE) Exam:

79 The PE exam is the second step in the process of becoming a licensed professional engineer. It is
80 designed for engineers who have gained a minimum of four years of post-college work experience
81 in their chosen engineering discipline. The PE exam tests the candidate's competency.

82 ***Bard:***

Please cite this paper as:

Naser M.Z., Ross B., Ogle J., Kodur V., Hawileeh R., Abdall J., Thai T. (2024). Evaluating the Performance of Artificial Intelligence Chatbots and Large Language Models in the FE and PE Structural Exams. *ASCE Journal of Practice Periodical on Structural Design and Construction*. <https://doi.org/10.1061/PPSCFX.SCENG-1369>.

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

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

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

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

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

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

99 **Approach and findings**

100 The National Council of Examiners for Engineering and Surveying (NCEES) regulates the FE and
101 PE exams (NCEES 2023a). According to the NCEES, the FE exam is a 6-hour exam containing
102 110 questions, and exam takers often have 5 hours and 20 minutes to complete. The FE exam
103 covers a variety of topics spanning 18 areas. These are listed herein, along with the expected
104 number of questions noted by the NCEES FE civil practice exam.

105 On the contrary, the Principles and Practice of Engineering (PE) exam tests for a minimum level
106 of competency in a particular engineering discipline. This exam is designed for engineers who
107 have gained a minimum of four years of post-college work experience in a particular civil
108 engineering discipline. The PE exam includes 80 questions and can be completed in 8 hours. This
109 study uses the Structural Engineering discipline of the PE exam.

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

117 Please note that the actual questions are not listed herein to maintain the NCEES' copyrights.
118 However, a general discussion is provided to describe the outcome of this analysis. The reader is
119 to note that both the FE and PE exams provide some visual-based questions. Some of such
120 questions were simply transcribed to the chatbots in the February 2023 editions of *ChatGPT-4* and
121 *Bard*. We also note that with the release of the Vision edition of *ChatGPT-4* (October 2023), some

Please cite this paper as:

Naser M.Z., Ross B., Ogle J., Kodur V., Hawileeh R., Abdall J., Thai T. (2024). Evaluating the Performance of Artificial Intelligence Chatbots and Large Language Models in the FE and PE Structural Exams. *ASCE Journal of Practice Periodical on Structural Design and Construction*. <https://doi.org/10.1061/PPSCFX.SCENG-1369>.

122 of such questions were revisited to allow the chatbot to visualize and answer such questions.
123 Overall, each chatbot was examined with 79 FE questions and 39 PE questions. While this
124 examination does not conform to actual FE and PE exams, the NCEES still states that (NCEES
125 2023b),

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

- 128
- 129 • *Contain questions from past exams*
 - 130 • *Simulate the format, style, and level of difficulty*
 - 131 • *Provide solutions or feedback on performance in each topic area*

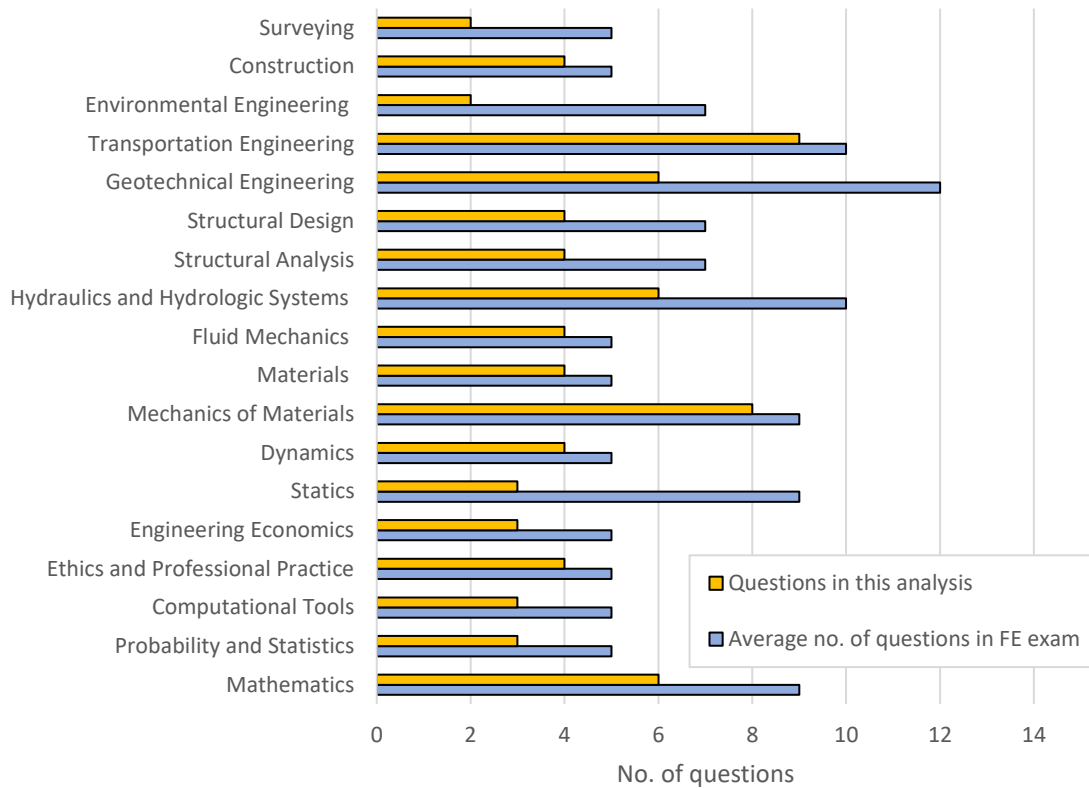
132 In our analysis, we provide some samples of correct and incorrect answers as listed by the chatbots.
133 Some of the values were slightly changed to conceal the actual questions listed by the FE and PE
134 practice exams.

134 FE exam

135 The 18 areas tested at the FE exam are listed in Fig. 1. As one can see, these areas cover general
136 engineering themes. Altogether, each area has five different questions, and a handful of areas
137 provide more than five questions. For example, Fig. 1 shows that Geotechnical Engineering,
138 Hydraulics and Hydraulic Systems, Mathematics, Statics, and Mechanics of Materials have the
139 lion's share. The same figure also shows the number and distribution of questions in a typical FE
140 exam and our analysis. It can be seen that there is good convergence among most of the areas with
141 the exception of Surveying and Environmental Engineering – and hence aid in validating our
142 analysis. The large variation in these two areas stems from the fact that many of the presented
143 questions in NCEES FE practice exam can be transcribed easily (i.e., images or charts).

Please cite this paper as:

Naser M.Z., Ross B., Ogle J., Kodur V., Hawileeh R., Abdall J., Thai T. (2024). Evaluating the Performance of Artificial Intelligence Chatbots and Large Language Models in the FE and PE Structural Exams. *ASCE Journal of Practice Periodical on Structural Design and Construction*. <https://doi.org/10.1061/PPSCFX.SCENG-1369>.



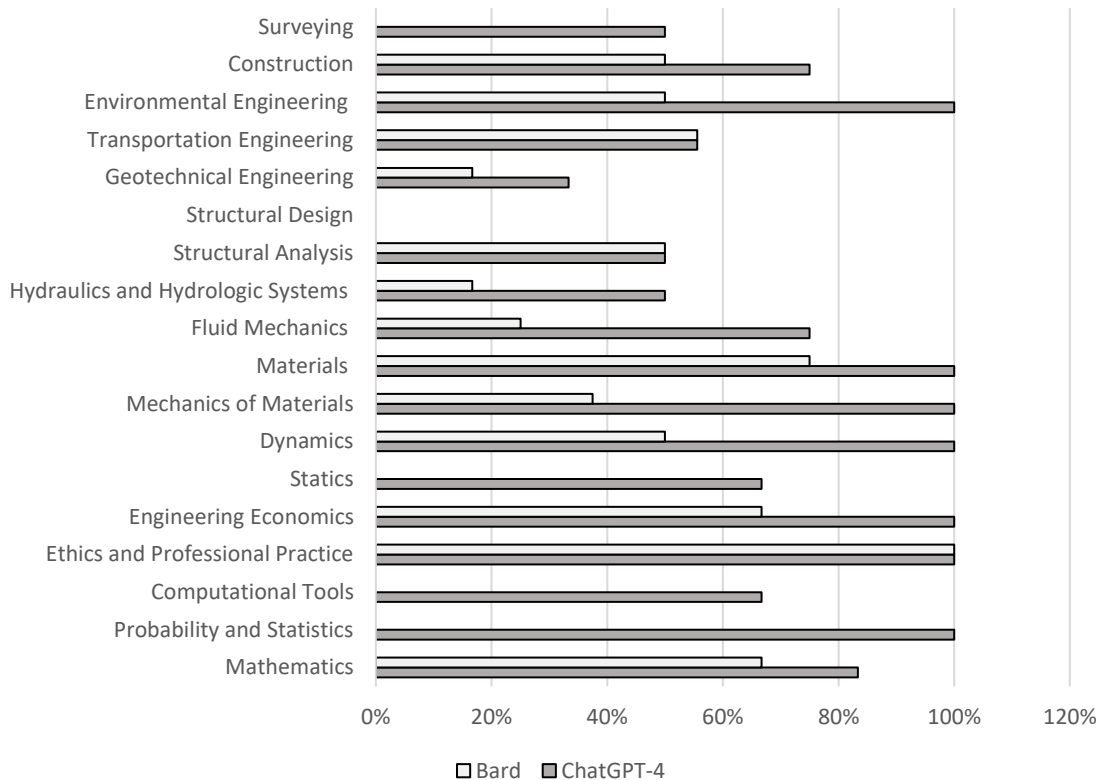
144
145

Fig. 1 Comparison between the number of questions in a typical FE exam and our analysis

146 Figure 2 shows the performance of *ChatGPT-4* and *Bard* across each area. It is quite clear that
147 *ChatGPT-4* outperforms *Bard* in every single entry. In fact, *ChatGPT-4* does not perform below
148 50% except in Structural Design and Geotechnical Engineering areas. On the other hand, *Bard*
149 falls short of the 50% mark consistently and does not answer any correct answers in five areas
150 (Probability and Statistics, Computational Tools, Statics, Structural Design, and Surveying). On a
151 different front, it is interesting to note that both chatbots do not answer any question correctly in
152 the area of Structural Design, but they answer all questions on Ethics and Professional Practice
153 correctly.

Please cite this paper as:

Naser M.Z., Ross B., Ogle J., Kodur V., Hawileeh R., Abdall J., Thai T. (2024). Evaluating the Performance of Artificial Intelligence Chatbots and Large Language Models in the FE and PE Structural Exams. *ASCE Journal of Practice Periodical on Structural Design and Construction*. <https://doi.org/10.1061/PPSCFX.SCENG-1369>.



154
155

Fig. 2 Results of FE exam analysis

156 The findings of our analysis indicate that *ChatGPT-4* and *Bard* scored 70.9% and 39.2% in the FE
157 exam. While the passing score for the FE exam is not commonly announced, however, such a score
158 has been historically believed to be around 60-70%. This implies that the current version of
159 *ChatGPT-4* could potentially pass this exam, and it is likely that future versions of *ChatGPT-4* to
160 pass a complete and standard version of this exam as well. This rationale stems from the fact that
161 6 of the wrongly answered questions by *ChatGPT-4* were close to the correct answers listed by the
162 practice exams – yet these were marked as wrong answers. For instance, in one question,
163 *ChatGPT-4* laid out the complete process with correct formulae and logic but had a calculation
164 mistake that skewed the final answer. In another example, *ChatGPT-4* did not convert the units
165 properly. At the moment, it is quite interesting to note that each question addressed by *ChatGPT-4*
166 was presented with complete thoughts and step-by-step processes and equations. Below are
167 samples of two questions derived from those seen in the FE practice exams.

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

170 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
171 the function with respect to x over the given interval.

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

Please cite this paper as:

Naser M.Z., Ross B., Ogle J., Kodur V., Hawileeh R., Abdall J., Thai T. (2024). Evaluating the Performance of Artificial Intelligence Chatbots and Large Language Models in the FE and PE Structural Exams. *ASCE Journal of Practice Periodical on Structural Design and Construction*. <https://doi.org/10.1061/PPSCFX.SCENG-1369>.

173 $\int (5x^3 + 1) dx$ from 4 to 9

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

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

177 Now, evaluate the antiderivative at the given limits:

178 $[(5/4)(9^4) + 9] - [(5/4)(4^4) + 4]$

179 Calculating the values:

180 $(5/4)(6561) + 9 - [(5/4)(256) + 4] \approx 8190.25 - 324 \approx 7866.25$

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

182 **Q.2 The magnitude (N) of the resultant of the three coplanar forces, A, B, and C, where has an angle of**
183 **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**
184 **1.15 N is most nearly:**

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

188
189 Force A:

190 Magnitude: 2 N

191 Angle: 115 degrees

192

193 Force B:

194 Magnitude: 4.13 N

195 Angle: 90 degrees

196

197 Force C:

198 Magnitude: 1.15 N

199 Angle: 5 degrees

200

201 Force A components:

202 $A_x = 2 * \cos(115^\circ) \approx -1.81$ N

203 $A_y = 2 * \sin(115^\circ) \approx 1.91$ N

204

205 Force B components:

206 $B_x = 4.13 * \cos(90^\circ) = 0$ N

207 $B_y = 4.13 * \sin(90^\circ) = 4.13$ N

208

209 Force C components:

210 $C_x = 1.15 * \cos(5^\circ) \approx 1.14$ N

211 $C_y = 1.15 * \sin(5^\circ) \approx 0.10$ N

212 Sum of x-components:

213 $R_x = A_x + B_x + C_x = -1.81 + 0 + 1.14 \approx -0.67$ N

214

215 Sum of y-components:

216 $R_y = A_y + B_y + C_y = 1.91 + 4.13 + 0.10 \approx 6.14$ N

Please cite this paper as:

Naser M.Z., Ross B., Ogle J., Kodur V., Hawileeh R., Abdall J., Thai T. (2024). Evaluating the Performance of Artificial Intelligence Chatbots and Large Language Models in the FE and PE Structural Exams. *ASCE Journal of Practice Periodical on Structural Design and Construction*. <https://doi.org/10.1061/PPSCFX.SCENG-1369>.

217

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

218

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

219

$$N \approx 6.17$$

220

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

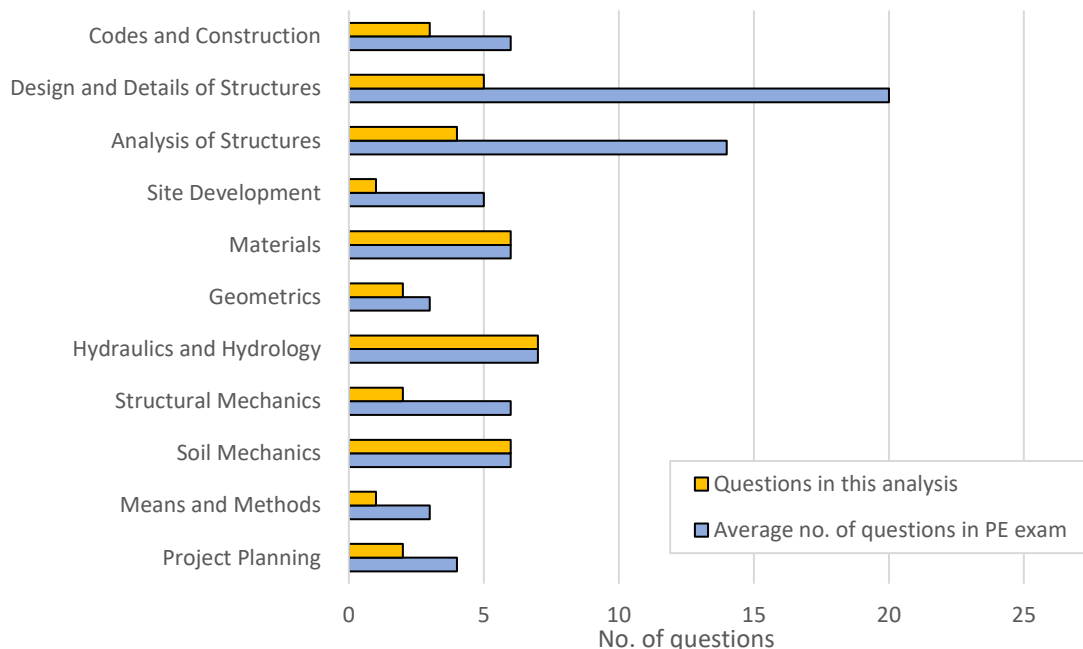
221

222 As one can see, the above solutions are clear and concise.

223 PE exam

224

225 Figure 3 lists the areas covered in a typical PE exam. There are 11 different areas in the Structural-
226 based PE exam. For transparency, the examination, in the case of the PE exam, is much more
227 complicated than that of the FE exam. This is due to the fact that the PE exam contains a larger
228 number of questions with intricate visuals (i.e., design detailing, loading configurations, etc.) that
229 may not be easily transcribed. As such, the number of explored questions in the presented analysis
230 is significantly fewer than those expected in the actual PE exam (see Fig. 3). For completion, we
present the findings of this analysis herein.



231

232

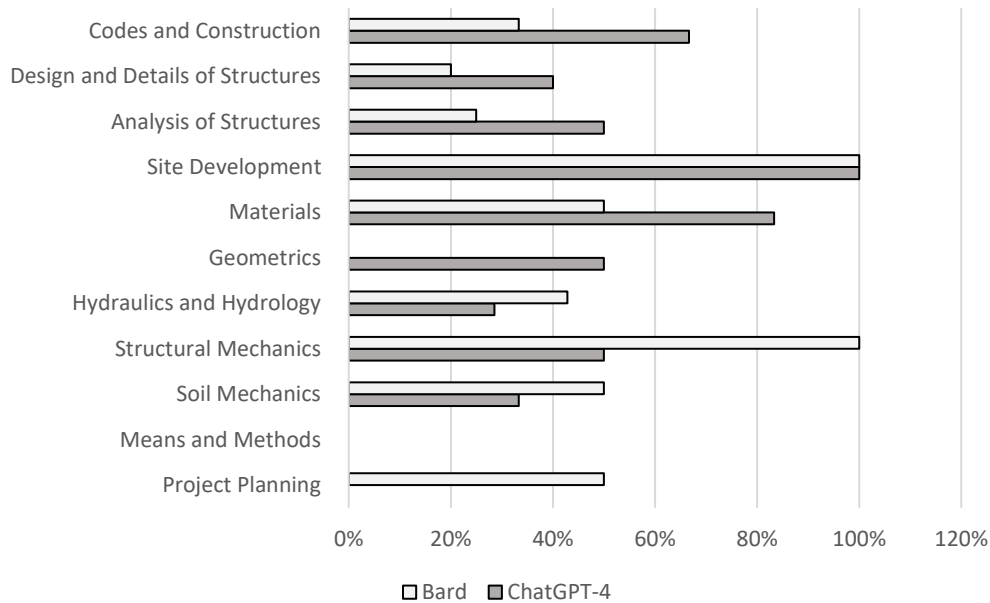
Fig. 3 Comparison between the number of questions in a typical PE exam and our analysis

233

234 Figure 4 breaks down the performance of *ChatGPT-4* and *Bard* across each area. In the case of
235 this exam, it seems that *ChatGPT-4* and *Bard* perform equally (46.2% vs. 41%) and below the
236 50% mark. *ChatGPT-4* performs exceptionally well in the area of Materials and Site Development,
237 while *Bard* performs similarly in Structural Mechanics and Site Development. In terms of Analysis
238 of Structures, Design and Details of Structures, and Codes and Construction *ChatGPT-4* and *Bard*
239 score within 25-67%. Unlike the case of the FE exam, the current version of *ChatGPT-4* and *Bard*
240 are unlikely to pass this exam at the moment. We anticipate that future versions of these chatbots
to be able to pass this exam.

Please cite this paper as:

Naser M.Z., Ross B., Ogle J., Kodur V., Hawileeh R., Abdall J., Thai T. (2024). Evaluating the Performance of Artificial Intelligence Chatbots and Large Language Models in the FE and PE Structural Exams. *ASCE Journal of Practice Periodical on Structural Design and Construction*. <https://doi.org/10.1061/PPSCFX.SCENG-1369>.



241
242

Fig. 4 Results of PE exam analysis

243 A few interesting observations were arrived at while examining the responses received from
244 *ChatGPT-4* and *Bard*. For example, both chatbots seem to be able to understand the solution
245 process behind a traditional design/analysis problem. However, the chatbots struggle to distinguish
246 between uniform and point loads. On one occasion, *Bard* used the elastic deformation equation for
247 a point load acting on a beam to calculate the elastic deflection of a uniform load. On the other
248 hand, *ChatGPT-4* suffered from a simple calculation mistake (see Q. 4 below) that led to a
249 calculation error despite having the process procedure and equations. On a more positive note,
250 both chatbots could cite ACI and ASCE on most occasions properly. Below are samples of two
251 questions derived from those seen in the FE practice exams.

252 **Q. 3 For the following five 28-day average standard-cured concrete strength tests: A, B, C, D, and E with**
253 **a required strength of G ksi, select why the concrete is unsatisfactory:**
254 **(A) Because the average of three consecutive tests is less than G ksi.**
255 **(B) Because Test D did not meet the required strength.**
256 **(C) Because the average of the first three tests is greater than the required strength.**
257 **(D) Because Test A is within 500 psi of the required strength.**

258 According to ACI 318, the concrete is considered satisfactory if both of the following conditions are met:
259

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

264 Let's evaluate the test results:

265 Test A psi

266 Test B psi

Please cite this paper as:

Naser M.Z., Ross B., Ogle J., Kodur V., Hawileeh R., Abdall J., Thai T. (2024). Evaluating the Performance of Artificial Intelligence Chatbots and Large Language Models in the FE and PE Structural Exams. *ASCE Journal of Practice Periodical on Structural Design and Construction*. <https://doi.org/10.1061/PPSCFX.SCENG-1369>.

267 Test C psi
268 Test D psi
269 Test E psi
270 We check the average of three consecutive strength tests.
271 The average of any three consecutive strength tests equals or exceeds the specified compressive strength
272 (f'_c):
273 Average of Tests A, B, and C (satisfies the criterion)
274 Average of Tests B, C, and D (does not satisfy the criterion)
275 Average of Tests C, D, and e (does not satisfy the criterion)
276 • No individual strength test result falls below the specified compressive strength by more than 500
277 psi:
278 • Test C is 375 psi below the required strength (satisfies the criterion)
279 Based on these evaluations, the concrete is unsatisfactory because the average of the last two sets of three
280 consecutive tests does not meet the required strength.
281 Thus, the correct statement is:
282 (A) Because the average of the last three consecutive tests is less than the required strength.

283 **Q. 4 The maximum flexural stress (psi) in a 20 ft long simply supported rectangular member with a width**
284 **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**
285 **the support is most nearly:**
286 To find the maximum flexural stress, we'll first calculate the maximum moment for both the uniform load
287 and point load and then use the flexural stress formula to find the maximum stress.
288
289 Maximum moment due to uniform load:
290 Uniform load, $w = 200 \text{ lb/ft}$
291 Length, $L = 20 \text{ ft}$
292
293 For a simply supported beam with a uniform load, the maximum moment (M_1) occurs at the center of the
294 span:
295 $M_1 = (w * L^2) / 8$
296 $M_1 = (200 \text{ lb/ft} * (20 \text{ ft})^2) / 8$
297 $M_1 = 10,000 \text{ lb*ft}$
298
299 Maximum moment due to point load:
300 Point load, $P = 1000 \text{ lb}$
301 Distance from support, $a = 10 \text{ ft}$
302 Length, $L = 20 \text{ ft}$
303
304 For a simply supported beam with a point load, the maximum moment (M_2) is given by the formula:
305 $M_2 = (P * a * (L - a)) / L$
306 $M_2 = (1000 \text{ lb} * 10 \text{ ft} * (20 \text{ ft} - 10 \text{ ft})) / 20 \text{ ft}$
307 $M_2 = 5,000 \text{ lb*ft}$
308
309 Now, we need to find the maximum combined moment. Since both loads are on the same beam, we'll simply
310 add the two moments to get the maximum combined moment:
311
312 $M_{\text{total}} = M_1 + M_2$
313 $M_{\text{total}} = 10,000 \text{ lbft} + 5,000 \text{ lbft}$
314 $M_{\text{total}} = 15,000 \text{ lb*ft}$

Please cite this paper as:

Naser M.Z., Ross B., Ogle J., Kodur V., Hawileeh R., Abdall J., Thai T. (2024). Evaluating the Performance of Artificial Intelligence Chatbots and Large Language Models in the FE and PE Structural Exams. *ASCE Journal of Practice Periodical on Structural Design and Construction*. <https://doi.org/10.1061/PPSCFX.SCENG-1369>.

315
316 Now, we'll use the flexural stress formula to find the maximum stress:
317
318 Flexural stress = $(M * S) / I$ [Note from authors, Flexural stress = M / S]
319
320 Where:
321 M is the maximum moment
322 S is the section modulus
323 I is the moment of inertia
324
325 First, we'll calculate the moment of First, we'll calculate the moment of inertia (I) for the rectangular section:
326 $I = (b * h^3) / 12$
327 Where b is the width (4 in) and h is the height (12 in).
328 $I = (4 \text{ in} * (12 \text{ in})^3) / 12$
329 $I = 6912 \text{ in}^4$ [Note from authors, the calculation mistake occurs here, I should equal 576 in^4]
330
331 Now, we'll calculate the section modulus (S):
332 $S = I / (h / 2)$
333 $S = 6912 \text{ in}^4 / (12 \text{ in} / 2)$
334 $S = 1152 \text{ in}^3$
335
336 Finally, we'll calculate the maximum flexural stress using the formula:
337 Flexural stress = $(M * S) / I$ [Note from authors, Flexural stress = M / S]
338 Since the maximum moment is in ft, we need to convert it to inches:
339 $M_{\text{total}} = 15,000 \text{ lbft} * 12 \text{ in/ft} = 180,000 \text{ lbin}$
340 Flexural stress = $(180,000 \text{ lb*in}) / 1152 \text{ in}^3$
341 Flexural stress $\approx 156.25 \text{ psi}$
342 So, the maximum flexural stress is most nearly 156.25 psi.

343 **Where to go from here? Chatbots as teaching assistants and engineers!**

344 We would like to point out that there are critical similarities and differences between *ChatGPT-4*
345 and a human engineer in terms of taking the exam. For example, both *ChatGPT-4* and the engineer
346 do not have access to the internet (unlike *Bard*). The engineer, on the other hand, would have
347 access to plenty of resources and codes. In one instance, *ChatGPT-4* stated that it does not have
348 access to values embedded in a specific chart and hence could not solve such a question. Other
349 items, such as lucky guessing, fatigue, and stress, are likely to take a toll on human behavior.
350 Overall, the AI-enabled examination process for both the FE and PE took less than 2 hours.

351 In one instance, this particular chatbot was able to solve a PE question and arrive at the exact
352 solution value. The chatbots noted that none of the listed answers were correct since the listed
353 correct answer was a rounded value (as typically noted in the PE exam). In another instance, the
354 chatbot was able to correctly lay down the complete process and solve a number of PE problem(s).
355 However, the chatbot either copied one wrong value from a question or used an outdated equation
356 to solve the question and, hence, was not able to correctly answer the question(s) at hand. We
357 anticipate that future *ChatGPT-4* and *Bard* generations can serve as suitable teaching assistants
358 and guides for students and young engineers, as noted in the improvements observed in the
359 upgraded Vision edition of *ChatGPT-4*. It is noteworthy that the lead author has experimented

Please cite this paper as:

Naser M.Z., Ross B., Ogle J., Kodur V., Hawileeh R., Abdall J., Thai T. (2024). Evaluating the Performance of Artificial Intelligence Chatbots and Large Language Models in the FE and PE Structural Exams. *ASCE Journal of Practice Periodical on Structural Design and Construction*. <https://doi.org/10.1061/PPSCFX.SCENG-1369>.

360 with using *ChatGPT-4* as an example-bank generator to create a variety of solved examples for
361 the courses of *Statics*, *Reinforced Concrete Design*, and *Machine Learning and Causality for*
362 *Engineers* that are being offered at Clemson University at the moment. The lead author reports
363 success and positive feedback (so far) from students attending such courses.

364 Clearly, *ChatGPT-4* and *Bard* are not without error – they have “learned” from human and textual
365 processes. Much like human error occurring while entering data into a calculator or computer
366 program, similar simple errors can occur with these artificial intelligence tools. The differentiator
367 is the communicative nature of AI. The detailed step-by-step operations have merit for training
368 young engineers. The interesting challenge is to identify use scenarios - how can you use these
369 tools to help students who struggle to understand the details of the process? Can you give
370 homework problems with solutions worked out and ask the students to verify accuracy? Can these
371 tools substitute for tutors, and how can we train improvements in responses to avoid these simple
372 errors? Should not we all be concerned about an error that has a 10× magnitude beyond the correct
373 response? Can we use the resulting failure or cost impact of that magnitude of error to drive home
374 the importance of truly understanding disciplinary knowledge? AI is the new pocket calculator,
375 yet its accuracy and precision are still to be determined. As a profession, we must embrace the tool
376 and ensure that its use does not negatively impact the health, safety, and welfare of the public.

377 **Data Availability**

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

379 **Conflict of Interest**

380 The authors declare no conflict of interest.

381 **References**

- 382 Adamopoulou, E., and Moussiades, L. (2020). “An Overview of Chatbot Technology.” *IFIP*
383 *Advances in Information and Communication Technology*.
- 384 Google. (2023). “Bard FAQ.” <<https://bard.google.com/faq>>.
- 385 Johnston, D. W., Thomas Ahluwalia, N., and Gwyn, M. B. (2007). “Improving the Professional
386 Engineering Licensure Process for Construction Engineers.” *Journal of Construction*
387 *Engineering and Management*.
- 388 NCEES. (2023a). “National Council of Examiners for Engineering and Surveying (NCEES).”
389 <<https://ncees.org/>>.
- 390 NCEES. (2023b). “Exam preparation materials.” <[https://ncees.org/exams/exam-preparation-](https://ncees.org/exams/exam-preparation-materials/)
391 [materials/](https://ncees.org/exams/exam-preparation-materials/)>.
- 392 OpenAI. (2023). “ChatGPT — Release Notes.” <[https://help.openai.com/en/articles/6825453-](https://help.openai.com/en/articles/6825453-chatgpt-release-notes)
393 [chatgpt-release-notes](https://help.openai.com/en/articles/6825453-chatgpt-release-notes)>.

394

395