

Varsity: Can Large Language Models Keep Power Engineering Students in Phase?

Samuel Talkington and Daniel K. Molzahn

Abstract—This paper provides an educational case study regarding our experience in deploying ChatGPT Large Language Models (LLMs) in the Spring 2025 and Fall 2023 offerings of ECE 4320: Power System Analysis & Control at Georgia Tech. As part of course assessments, students were tasked with identifying, explaining, and correcting errors in the ChatGPT outputs corresponding to power factor correction problems. While most students successfully identified the errors in the outputs from the GPT-4 version of ChatGPT used in Fall 2023, students found the errors from the ChatGPT o1 version much more difficult to identify in Spring 2025. As shown in this case study, the role of LLMs in pedagogy, assessment, and learning in power engineering classrooms is an important topic deserving further investigation.

I. INTRODUCTION

In just a few short years, large language models (LLMs) have been ubiquitously deployed in a broad array of fields, including power engineering [1]. The impacts of these models have been rapid and widespread, especially in higher education [2]–[4]. From engineering [4], operations research [5], and algorithms [6], academic communities are asking critical questions about the role of LLMs. These questions are also being raised in power engineering education [7].

The power engineering community has had limited discussion on the impacts of LLMs. Early studies report both promise and peril: while LLMs accelerate ideation, they also confidently generate physically inconsistent or numerically incorrect answers [2], [3]. Electric power systems, as critical infrastructure, may be particularly vulnerable to the risks posed by LLMs, and these risks are especially poignant in the context of power engineering education. If the next generation of power engineers is trained in a way that validates or even promotes frequent errors in units, sign conventions, or phasor arithmetic, this can propagate directly into megawatt-scale mistakes.

Recognizing that LLMs are likely here to stay in the classroom [2], there is an urgent need for community discussion on governance and best practices for their use. Prior work emphasizes that teachers and learners must develop new competencies (e.g., prompt strategies, fact checking) to harness LLMs responsibly [3], [4]. In this spirit, prompt design and assessment techniques that integrate LLMs into coursework in

a way that reinforces fundamental engineering skills, rather than short-circuiting them, are poised to become essential.

This issue became salient to the authors during our experience teaching the Spring 2025 and Fall 2023 iterations of ECE 4320: Power System Analysis & Control, a first course in power systems at Georgia Tech that is typically taken by senior undergraduate and first-year graduate students. In this course, we experimented with deploying OpenAI’s ChatGPT LLM as part of exam assessments. Specifically, students were tasked with interpreting and correcting errors in the solutions produced by ChatGPT to various power systems problems.

Using power factor correction problems as a case study topic, the contributions of this paper are:

- 1) *Assessment protocol*: Case study examples using LLM outputs in power engineering course assessments, including a grading rubric based on error identification and correction.
- 2) *Empirical data*: A retrospective analysis student success in identifying and correcting LLM output errors.
- 3) *Perspectives and guidelines*: Our findings in Section V suggest an opportunity for power engineering educators to refine their assessment practices by harnessing LLMs to promote critical engineering skills.

The remainder of this paper is organized as follows. Section II surveys related work and the ECE 4320 course. Section III details the problem and prompt design. Section IV explains the solution output, the errors it contains, and corrections. Section V presents empirical results. Section VI concludes the paper.

II. BACKGROUND

A. Related work

Reference [1] maps research questions for LLMs in the electric energy sector, including operation and planning. Although [1] did not investigate domain-specific educational impacts, the work of [7] evaluated GPT-3.5 on economic dispatch homework problems, finding that 58% of its numerical outputs violated basic power balance constraints. Two key gaps in prior literature are:

- 1) *Problem setting*: The prior education task explored in [7], while valuable, was a short, single-step economic dispatch problem. The economic dispatch problem, including the formulation in [7], typically uses a highly simplified model that does not encode the physics that governs the underlying power system. Presently, the literature lacks an understanding of how LLMs understand multielement linear circuits in power engineering assignments.
- 2) *Students in the loop*: We currently know little about whether students can identify and correct LLM mistakes in

Support from NSF AI Institute for Advances in Optimization (AI4OPT) #2112533. The work of S. Talkington is supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1650044. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

The authors are with the School of Electrical and Computer Engineering, Georgia Institute of Technology. Email: {talkington,molzahn}@gatech.edu.

electrical engineering contexts, and which scaffolds help. As LLM workflows become increasingly intertwined in classrooms and engineering practice, it is essential to improve our understanding of students' LLM interactions.

In adjacent communities, [5] explores the use of LLMs in optimization education, covering the computation of dual linear programs. Reference [6] covers the performance of ChatGPT o1 for a proof-based analysis of algorithms course, finding that it performed at a passing undergraduate level.

B. Overview of the course and the Georgia Tech experience

At Georgia Tech, ECE 4320 is typically one of the final courses taken by undergraduate ECE students who choose the *Electric Energy Systems Thread*, which is a structured curriculum for students who wish to specialize in electric power systems or power electronics. For many such students, the ECE 4320 classroom is the last time they will be exposed to formal power system concepts in classroom settings.

This course introduces students to key power engineering concepts, such as the power flow equations, the optimal power flow problem, unit commitment, and state estimation. The course thus lays the foundation for a career in power engineering and introduces core concepts for graduate education and research in power systems. The foundational nature of this course means that understanding LLMs' impact is essential.

III. A CASE STUDY: POWER FACTOR CORRECTION

In the Spring 2025 iteration of ECE 4320, both students and ChatGPT version o1 were presented with a power factor correction problem based on the single-phase, 60 Hz circuit in Fig. 1. We focus on this problem as a case study on LLMs in power engineering education. Providing a comparison from an earlier version of ChatGPT (version GPT-4), the Fall 2023 iteration of ECE 4320 posed a related power factor correction problem that is described in the appendix.

A. Problem formulation

In Fig. 1, the active power consumed by the load, P_d , is a constant value of 50 W. The reactive power consumed by the load, Q_d , continuously varies between a minimum value of -36 VAR and a maximum value of 60 VAR. As shown in the circuit above, the voltage source has an RMS value of 100 V.

The students were instructed to act as circuit designers—they could choose (fixed) values for both capacitive reactance X_l and inductive reactance X_c , as well as the switching configurations for S as a function of the changing reactive power demand of the load. Within this problem setting, payments to the utility company that owns the voltage source V_s were modeled as being based on the smallest power factor of the power $P_s + jQ_s$ supplied across the range of reactive demands Q_d .

Thus, the goal of the problem statement was to choose X_l , X_c , and a strategy for configuring the switch S to make the smallest power factor of $P_s + jQ_s$ as large as possible (that is, as close to 1.0, either lagging or leading) as the load's reactive power demand Q_d varies within the range from -36 Var (supplying 36 VAR) to 60 VAR (consuming 60 VAR).

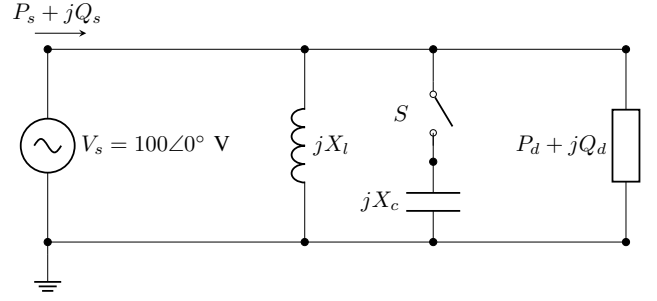


Fig. 1: The circuit presented to ChatGPT o1 and ECE 4320 students as the context for the power factor correction problem.

B. Power factor correction prompt

The ChatGPT o1 model was prompted to design a power factor correction scheme for the circuit in Fig. 1.

Prompt

The following problem for a university-level power systems engineering course explores power factor correction. Consider a balanced, single-phase, 60 Hz circuit with an AC voltage source with RMS value of $V_{\text{rms}} = 100$ V connected to the parallel combination of a voltage source, a switched capacitor, and a load impedance with a complex power consumption of $P_d + jQ_d$ VA. The amount of complex power supplied by the voltage source is denoted as $P_s + jQ_s$ VA. This switched capacitor bank consists of an inductor with reactance X_l in parallel with a capacitor with reactance X_c connected through switch S . The inductor with reactance X_l is always connected in the circuit (not switchable). The active power consumed by the load, P_d , is a constant value of 50 W. The reactive power consumed by the load, Q_d , continuously varies between a minimum value of -36 VAR and a maximum value of 60 VAR. Your task in this problem is to determine the (fixed) values of the reactances X_l and X_c in Ohms as well as a switching strategy for the switch S . Choose these so that the power factor of the power supplied by the voltage source ($P_s + jQ_s$) is as close to unity (1.0) as possible in terms of its magnitude (leading or lagging) for any value of the load's reactive power demand Q_d within the range from -36 VAR (supplying 36 VAR) to 60 VAR (consuming 60 VAR).

IV. LLM OUTPUT, ERRORS, AND CORRECTIONS

When presented with the power factor correction prompt in Section III-B, ChatGPT o1 output an incorrect solution, which we describe below in Section IV-A. Given the ChatGPT o1 output, students were tasked with identifying all errors, explaining the physical principle behind why it was incorrect, and ultimately providing a correct solution.

A. ChatGPT's solution

We next document the *incorrect* output from ChatGPT o1 to the prompt in Section III-B. Embedded in the lengthy solution proposed by ChatGPT, there are three locations where either critical engineering errors or fundamental misunderstandings of power systems were introduced. Each of these locations are demarcated with ①, ②, and ③, respectively.

Response

Analysis and Solution Strategy

We have a single-phase, 60 Hz source of $V_{\text{rms}} = 100$ V supplying a load whose real power is fixed at $P_d = 50$ W, and whose reactive power Q_d can vary over the range $-36 \text{ VAR} \leq Q_d \leq +60 \text{ VAR}$.

(Here, $Q_d < 0$ means the load is supplying reactive power, i.e. effectively capacitive/leading, and $Q_d > 0$ means the load is consuming reactive power, i.e. inductive/lagging.)

We wish to add a parallel branch of a **fixed** inductor (always connected) plus a **switchable** capacitor such that, no matter the value of Q_d in that range, the net reactive power drawn from the 100 V source is as close to zero as possible (i.e., power factor near unity). Denote:

- The inductor's reactance by X_L .
- The capacitor's reactance by X_C .
- The switch S either **opens** (removes the capacitor) or **closes** (places the capacitor in parallel).

Hence, depending on S , the net reactive power seen by the source is:

- 1) **Switch Open** (inductor only): $Q_s = Q_d + Q_L$ where $Q_L = \frac{V^2}{X_L} > 0$ (inductive consumption).
- 2) **Switch Closed** (inductor + capacitor): $Q_s = Q_d + Q_L + Q_C$, where $Q_C = -\frac{V^2}{X_C}$ (capacitive supply).

We want to choose X_L , X_C , and a switching "rule" so that $|Q_s| = |Q_d + Q_L + (\text{possibly } Q_C)|$ is as small as possible for **all** $Q_d \in [-36, +60]$ VAR.

1. Match extremes at unity power factor.

A common approach is to ensure that at the **two extreme** values of the load's reactive power, $Q_d = -36$ VAR (most leading) and $Q_d = +60$ VAR (most lagging), the source's net reactive power is zero in one of the two switch states. ①

- 1) **At $Q_d = -36$ VAR (load is supplying 36 VAR):**

We would like to be able to correct $Q_s \approx 0$ **without** the capacitor, i.e. with the switch open.

$$0 = Q_s = Q_d + Q_L = -36 + \frac{100^2}{X_L}.$$

Solving for X_L :

$$\frac{10,000}{X_L} = 36 \Rightarrow X_L = \frac{10,000}{36} \approx 277.78 \Omega.$$

Thus the (always-connected) inductor's reactance is chosen to be $X_L \approx 278 \Omega$.

- 2) **At $Q_d = +60$ VAR (load is consuming 60 VAR):**

We want to correct $Q_s \approx 0$ **with** the capacitor switched in. Then

$$0 = Q_s = Q_d + Q_L + Q_C = 60 + \frac{10,000}{X_L} - \frac{10,000}{X_C}.$$

From the above we have $\frac{10,000}{X_L} = 36$. Hence

$$60 + 36 - \frac{10,000}{X_C} = 0 \Rightarrow \frac{10,000}{X_C} = 96 \\ \Rightarrow X_C = \frac{10,000}{96} \approx 104.17 \Omega.$$

Therefore, $X_L \approx 278 \Omega$, $X_C \approx 104 \Omega$.

2. Reactive power vs. Q_d under each switch state

With the above reactances:

- 1) **Switch Open:** $Q_s = Q_d + 36$.
 - At $Q_d = -36$, indeed $Q_s = 0$.
 - At $Q_d = 0$, $Q_s = +36$ VAR.
 - At $Q_d = +60$, $Q_s = +96$ VAR.
- 2) **Switch Closed:** $Q_s = Q_d + 36 - 96 = Q_d - 60$.
 - At $Q_d = -36$, indeed $Q_s = -96$ VAR.
 - At $Q_d = 0$, $Q_s = -60$ VAR.
 - At $Q_d = +60$, $Q_s = 0$.

Graphically:

- **Open:** $Q_s = Q_d + 36$, which is a straight line intercepting zero at $Q_d = -36$.
- **Closed:** $Q_s = Q_d - 60$, intercepting zero at $Q_d = +60$.

3. Best switching threshold

We only have **two** discrete states (capacitor either in or out). At any intermediate Q_d , we cannot make $Q_s = 0$ exactly with a single fixed capacitor. Instead, we choose a **threshold** Q_d^* for switching so as to keep $|Q_s|$ as small as possible overall.

- When **Switch Open**, $Q_s = Q_d + 36$.
- When **Switch Closed**, $Q_s = Q_d - 60$.

We pick the state that yields the smaller magnitude $|Q_s|$. In other words:

- 1) Compare $|Q_d + 36|$ vs. $|Q_d - 60|$.
- 2) The "crossover" is where $|Q_d + 36| = |Q_d - 60|$.
 - For $Q_d \geq 0$, we can drop absolute values and check

$$Q_d + 36 = 60 - Q_d \Rightarrow 2Q_d = 24 \Rightarrow Q_d = 12 \text{ VAR}.$$

- Indeed for negative Q_d , the lines do not intersect in that region, but the best boundary for the entire range ends up around $Q_d = 12$ anyway.

Hence the switching "rule of thumb" is:

- If $Q_d < 12$ VAR, use **Switch Open** (only inductor).
- If $Q_d > 12$ VAR, use **Switch Closed** (inductor + capacitor).

At $Q_d = 12$, either choice gives $|Q_s| = 48$ VAR. For values below 12, the open-switch magnitude is smaller; for values above 12, the closed-switch magnitude is smaller.

This simple 2-state approach places the net source reactive power Q_s within at most ± 60 or ± 96 in the extremes ②, but it **exactly** hits zero at $Q_d = -36$ (switch open) and $Q_d = +60$ (switch closed). One cannot do better with only a single fixed capacitor and a single fixed inductor. ③

4. Final numeric answers

- 1) **Inductor reactance**

$$X_L = \frac{V^2}{36} = \frac{10,000}{36} \approx 277.8 \Omega$$

At $f = 60$ Hz, $X_L = 2\pi fL \Rightarrow L = \frac{X_L}{2\pi f} \approx \frac{278}{2\pi \times 60} \approx 0.74$ H.

- 2) **Capacitor reactance**

$$X_C = \frac{V^2}{96} = \frac{10,000}{96} \approx 104.2 \Omega$$

At $f = 60$ Hz, $X_C = \frac{1}{2\pi fC} \Rightarrow C = \frac{1}{2\pi fX_C} \approx \frac{1}{2\pi \times 60 \times 104} \approx 25.5 \mu\text{F}$.

- 3) **Switching strategy**

- **Switch Open** when $Q_d < 12$ VAR.
- **Switch Closed** when $Q_d > 12$ VAR.
- (Exactly at $Q_d = 12$, either state yields $|Q_s| = 48$ VAR, so one might choose either.)

This ensures:

- Perfect cancellation ($Q_s = 0$) at $Q_d = -36$ VAR (inductor only).
- Perfect cancellation ($Q_s = 0$) at $Q_d = +60$ VAR (inductor + capacitor).
- For intermediate Q_d , the net reactive power from the source remains as small as possible given just these two discrete capacitor states.

B. Analysis of ChatGPT's approach to power factor correction

ChatGPT's approach to this problem is somewhat reasonable but fundamentally incorrect. The idea of using the inductor's reactive power to "shift" the range of the reactive power supplied by the source and then correct the resulting power factor using a capacitor makes sense overall and is indeed the right intuition. Moreover, given ChatGPT's assertions about how to approach the problem, the numeric calculations it provides are mostly correct.

1) *Range of the reactive power compensation scheme:* The output provided by ChatGPT breaks down at the point labeled by ① above. Here, ChatGPT decides to fix the supply's reactive power outputs to zero (unity power factor) at the extreme load demands, $Q_d = -36$ VAR and $Q_d = 60$ VAR. This is not the best possible approach to maximizing the smallest power factor, since it does not take advantage of the still large power factor that would occur immediately above and below the extreme values. In other words, with this strategy, the power factor would still be near unity for $Q_d = -37$ VAR and for $Q_d = 61$

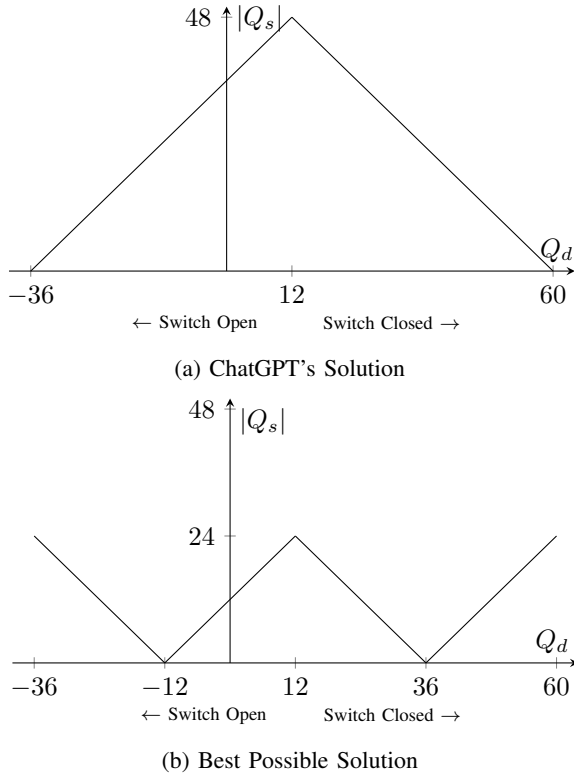


Fig. 2: Visualization of ChatGPT's output and the best solution.

VAR. This suggests that the solution could be improved by inward movement of the points where the supply provides zero reactive power so that these unity power factor points are within the range of Q_d as opposed to being at the extremes.

One can visualize this by considering the magnitude of the reactive power supplied or absorbed by the source, $|Q_s|$, as these unity power factor points change. When the capacitance and inductance values are chosen so that the unity power factor points are at the extremes of $Q_d = -36$ VAR and $Q_d = 60$ VAR, as in the ChatGPT output, $|Q_s|$ corresponds to the plot shown in Fig. 2a, which has a maximum value of $|Q_s| = 48$ VAR. Conversely, moving these points towards the middle of the range would yield a smaller maximum value of $|Q_s|$.

The plot in Fig. 2b has four line segments. The key insight is that the maximum smallest power factor will occur when each of these line segments has the same maximum height. Based on this intuition, the best possible solution evenly spaces the four line segments over the range of Q_d , as shown in Fig. 2b.

In other words, the best possible solution places the points of unity power factor at one-fourth of the range of Q_d , that is, $Q_d = (60 - (-36))\frac{1}{4} + (-36) = \frac{96}{4} - 36 = -12$ VAR and three-fourths of the range of Q_d , that is, $Q_d = (60 - (-36))\frac{3}{4} + (-36) = 72 - 36 = 36$ VAR, with the resulting maximum value of $|Q_s|$ being 24 VAR which is reached at $Q_d = -36, 12$, and 60 VAR. This curve results from switching the capacitor into the circuit at $Q_d = 12$ VAR.

2) *Corrected reactance values:* We next need to choose appropriate values for X_L and X_C . To select X_L , we observe that the power factor should be unity (and thus Q_s should equal zero) when $Q_d = -12$ VAR. Since the switch is open at this point, only the inductor is in the circuit, so we have

$Q_s = Q_d + \frac{V^2}{X_L}$. Plugging in the values for these quantities yields $0 = -12 + \frac{100^2}{X_L}$. Solving gives $X_L = 833.3 \Omega$.

To compute X_C , we observe that the power factor should also be unity (and thus Q_s should equal zero) when $Q_d = 36$ VAR. Since the switch is closed at this point, both the capacitor and inductor are in the circuit, and we thus have $Q_s = Q_d + \frac{V^2}{X_L} + \frac{V^2}{X_C}$. Substituting the values computed so far, we have $0 = 36 + \frac{100^2}{833.3} + \frac{100^2}{X_C}$. Solving gives $X_C = -208.3 \Omega$.

The smallest power factor occurs when $|Q_s|$ is at its maximum value of 24 VAR, which happens when $Q_d = -36, 12$, and 60 VAR. The value for this power factor at these points is given by $\frac{P_s}{\sqrt{P_s^2 + Q_s^2}}$. By conservation of power $P_s = P_d = 50$ W, so we have $\frac{50}{\sqrt{50^2 + 24^2}} = 0.9015$. The voltage source's power factor is leading when $Q_s > 0$ (source supplies reactive power), which occurs for values of $-12 < Q_d \leq 12$ and $36 < Q_d \leq 60$, and lagging when $Q_s < 0$ (source consumes reactive power), which occurs for $-36 \leq Q_d < -12$ and $12 < Q_d < 36$. This corrected solution is recorded in Tables I and II.

TABLE I: Corrected reactances

X_L	833.3 Ω	X_C	-208.3 Ω
-------	----------------	-------	-----------------

TABLE II: Corrected reactive power compensation

Switch Status	Range of Q_d	Smallest Source Power Factor
Open	$-36 \leq Q_d \leq 12$ VAR	0.9015 (lagging when $Q_d = -36$ VAR, leading when $Q_d = 12$ VAR)
Closed	$12 < Q_d \leq 60$ VAR	0.9015 (lagging when $Q_d = 12$ VAR, leading when $Q_d = 60$ VAR)

ChatGPT makes two other errors in its solution. At the point labeled by ② above, the numeric value is incorrect. With the X_L and X_C values and the switching strategy proposed by ChatGPT, the values of Q_s will range between 0 and 48, not ± 60 or ± 96 as indicated in the ChatGPT solution.

Finally, the statement that “one cannot do better with only a single fixed capacitor and a single fixed inductor” at the point labeled by ③ is meaningless as we have a *switched* capacitor.

V. DATA AND DISCUSSION REGARDING STUDENT SUCCESS

As a retrospective analysis, we now describe the success of ECE 4320 students at analyzing errors in the ChatGPT output.

A. Spring 2025: Student success in identifying errors in the output of ChatGPT o1

Students had varying levels of success in identifying ChatGPT's errors in the power factor correction problem from Section III-A. The grading rubric results summarized in Table III show the number of students who identified three, two, one, or zero of ChatGPT's errors.

While a plurality of students identified all three ChatGPT errors, a substantial majority (66.33%) did not detect all errors. The outcome that was the most concerning, in our view, was that approximately a quarter of the class did not detect *any* of the errors. Anecdotally, one student mentioned first solving the problem themselves without looking at the ChatGPT output, getting the same answer as that provided by ChatGPT (known to be incorrect), and then being unsure how to proceed.

TABLE III: Distribution of class success in identifying power engineering mistakes made by ChatGPT o1 ($n = 30$).

# of Errors Identified	Pct. of Students	# of Students
3/3	36.66%	11
2/3	16.66%	5
1/3	23.33%	7
0/3	23.33%	7

TABLE IV: Distribution of class success in correcting the power engineering mistakes made by ChatGPT o1 ($n = 30$).

Concepts corrected	Pct. of Students	# of Students
Q_d , X_L , X_C , and power factor	27%	8
X_L , X_C , and power factors	24%	7
Q_d only	10%	3
Partially, more than half	20%	6
Partially, less than half	14%	4
None	7%	2

B. Student success in correcting ChatGPT's errors

The second part of the question asked students to fix the ChatGPT errors by providing the correct solution. As summarized by the grading rubric results in Table IV, these outcomes also varied. A marginal majority of the class ($\approx 51\%$) successfully corrected the key errors, up to minor issues in the range of Q_d or the reactances. Conversely, about a quarter of the class failed to fully correct any of the errors. While roughly consistent with the number of students that did not identify any of the ChatGPT errors in ChatGPT, note that these were not all the same students.

C. Contrasting with Fall 2023: GPT-4

In the Fall 2023 semester, we used a predecessor to o1, ChatGPT version GPT-4, for a similar power factor correction problem with variations in the circuit parameters. As discussed in the appendix, the ChatGPT GPT-4 output in Fall 2023 included nearly a dozen significant errors, many of which were much more obvious than the relatively subtle logical error from ChatGPT o1 in Spring 2025. Accordingly, students in the Fall 2023 semester performed significantly better at error identification, with 21 of 28 students (75%) finding all or nearly all errors. However, students still found error correction to be challenging, with 18 of 28 students (64%) failing to provide a fully correct solution.

D. Discussion and Limitations

These results suggest that logical errors made by recent versions of LLMs may often be conceptually similar to the errors made by students. Thus, educators may find it difficult to determine whether a student's erroneous solution is simply copied from an LLM or instead indicates a student's misunderstanding in independent work. Moreover, given the rapid quality improvement between ChatGPT version GPT-4 in Fall 2023 and o1 in Spring 2025, educators should keep up to date with evolving LLM capabilities and typical errors.

Understanding the underlying reasons for the similarity between student mistakes and LLM output errors could give insights into LLM characteristics with respect to power engineering. Since we independently created the power factor problems presented in this paper, these problems could not be

part of any LLM training datasets and thus the LLMs are not simply regurgitating training data containing erroneous student solutions. This suggests a deeper justification for this similarity.

We also note several limitations with findings in this paper. The results presented in Section V are retrospective in nature; we only report our experiences in selected assignments for this specific course. The results shown in these tables do not indicate or imply a distribution of the ability of the more general power engineering student population to identify or correct LLM errors. We also do not further describe the precise methodologies used by the students to protect student privacy. This motivates further studies on LLMs that consider additional power engineering courses and different types of problems.

VI. CONCLUSION

As LLMs advance, their role in higher education requires careful examination. This is particularly the case in fields that impact real-world systems, such as electric power engineering. This study assessed the performance of ChatGPT in an undergraduate power system engineering course. The results show that many students struggled to identify errors in the ChatGPT outputs. The share of the class that was unable to identify any of ChatGPT's errors was larger than the share that was able to solve the entire problem correctly themselves. These findings highlight the need for robust assessment strategies and AI-aware grading policies in power engineering education.

REFERENCES

- [1] S. Majumder, L. Dong, F. Doudi, Y. Cai, C. Tian, D. Kalathil, K. Ding, A. A. Thatte, N. Li, and L. Xie, "Exploring the Capabilities and Limitations of Large Language Models in the Electric Energy Sector," *Joule*, vol. 8, pp. 1544–1549, June 2024.
- [2] S. Milano, J. A. McGrane, and S. Leonelli, "Large Language Models Challenge the Future of Higher Education," *Nature Machine Intelligence*, vol. 5, pp. 333–334, Apr. 2023.
- [3] E. Kasneci, K. Sessler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günemann, E. Hüllermeier, S. Krusche, G. Kutyniok, T. Michaeli, C. Nerdel, J. Pfeffer, O. Poquet, M. Sailer, A. Schmidt, T. Seidel, M. Stadler, J. Weller, J. Kuhn, and G. Kasneci, "ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education," *Learning and Individual Differences*, vol. 103, p. 102274, Apr. 2023.
- [4] K. D. Wang, E. Burkholder, C. Wieman, S. Salehi, and N. Haber, "Examining the Potential and Pitfalls of ChatGPT in Science and Engineering Problem-Solving," *Frontiers in Education*, vol. 8, Jan. 2024.
- [5] M. Klamkin, A. Deza, S. Cheng, H. Zhao, and P. V. Hentenryck, "DualSchool: How Reliable are LLMs for Optimization Education?," *arXiv:2505.21775*, 2025.
- [6] M. Ding, R. Kyng, F. Solda, and W. Yuan, "Assessing GPT Performance in a Proof-Based University-Level Course Under Blind Grading," *arXiv:2505.13664*, 2025.
- [7] A. Hickey, C. O. Faoláin, and P. Cuffe, "Large Language Models in Power Engineering Education: A Case Study on Solving Optimal Dispatch Coursework Problems," in *21st International Conference on Information Technology Based Higher Education and Training (ITHET)*, 2024.

APPENDIX

CONTRASTING WITH FALL 2023: THE PITFALLS OF GPT-4

In the Fall 2023 semester, we also experimented with the GPT-4 model on a similar power factor correction problem, using the circuit shown in Fig. 3. The problem setting presented to students and GPT-4 is shown in the prompt below.

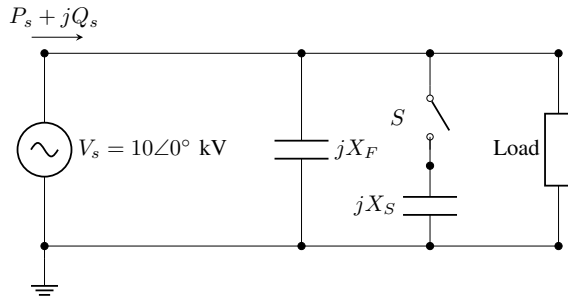


Fig. 3: The circuit presented to the Fall 2023 cohort of ECE 4320 students and ChatGPT GPT-4.

Prompt

The following problem for a university-level power systems engineering course explores power factor correction. Consider a balanced, single-phase, 60 Hz circuit with a 10 kV AC voltage source connected to the parallel combination of a load, a capacitor, and a switched capacitor. The load consumes different amounts of power over time. In the morning, the load consumes complex power of $10 + j5$ MVA. In the afternoon, the load consumes apparent power of 40 MVA at a power factor of 0.8 lagging. In the evening, the load consumes real power of 10 MW at a power factor of unity (1.0). The fixed capacitor is always connected in parallel with the load and the voltage sources, while the switched capacitor can be switched into or out of the circuit. The capacitors' reactances do not change across time periods.

Consider the power factor of the power supplied by the voltage source. Your task for this problem is to choose 1) the best values of capacitive reactances for both the fixed capacitor and the switched capacitor and 2) the switching configurations for the switched capacitor during each time period (morning, afternoon, and evening) that maximize the smallest power factor across the day. In other words, choose the capacitors' reactances and the switching configurations such that the smallest power factor during any time period is as large as possible (close to 1, either leading or lagging). What capacitive reactances would you select, what switching configurations would you choose during each time period, and what are the corresponding power factors? Indicate whether these power factors are leading or lagging.

Response

To maximize the smallest power factor across the day, we need to determine the capacitive reactances and switching configurations for each time period. Let's analyze each time period separately: **Morning:** The load consumes complex power of $10 + j5$ MVA. To maximize the smallest power factor, we want the power factor to be close to 1 (leading or lagging). Since the load's power factor is not specified ①, we can choose a leading power factor for simplicity. Let's aim for a power factor of 0.9 leading.

We can calculate the apparent power (S) using the complex power ($S = P + jQ$). Given that the complex power is $10 + j5$ MVA, the apparent power is 10 MVA. ② Since the voltage is 10 kV, the load impedance is $Z = V^2/S = (10^4)^2/10^7 = 10 \Omega$. ③

Let's select the fixed capacitor's reactance ($X_{c, \text{fixed}}$) such that it compensates the load's reactive power (Q) and provides the desired leading power factor. The reactance can be calculated as $X_{c, \text{fixed}} = -Q/(2\pi f)$, where f is the frequency.

$$X_{c, \text{fixed}} = -5 \text{ MVAR} / (2\pi * 60 \text{ Hz}) = -13.2 \Omega \text{ ④}$$

Since the reactance should be capacitive, we can use $X_{c, \text{fixed}} = 13.2 \Omega$. ⑤

Since the switched capacitor does not need to be used during the morning period (since the fixed capacitor already provides the desired power factor), we do not need to consider the switching configuration.

Afternoon: The load consumes apparent power of 40 MVA at a power factor of 0.8 lagging. To maximize the smallest power factor, we need to bring this power factor closer to 1 (leading or lagging).

Response

Let's aim for a power factor of 0.9 lagging. ⑥ The load's real power (P) can be calculated as $P = S * \text{power factor} = 40 \text{ MVA} * 0.8 = 32 \text{ MW}$. The reactive power (Q) can be determined using $Q = \sqrt{S^2 - P^2} = \sqrt{(40 \text{ MVA})^2 - (32 \text{ MW})^2} = 24 \text{ MVAR}$.

Using the load impedance formula $Z = V^2/S$, we can find the load impedance as $Z = (10^4)^2/40 \text{ MVA} = 2500 \Omega$. ⑦

To compensate for the reactive power and adjust the power factor to 0.9 lagging, we need to find the required capacitive reactance. The reactive power can be calculated as $Q = V^2/X_c$, where X_c is the capacitive reactance.

$$X_c = V^2/Q = (10^4)^2/24 \text{ MVAR} = 4167 \Omega \text{ ⑧}$$

Since we already have the fixed capacitor with reactance $X_{c, \text{fixed}} = 13.2 \Omega$, we need to find the switched capacitor's reactance ($X_{c, \text{switched}}$) such that the total capacitive reactance is equal to the required $X_c - X_{c, \text{switched}}$.

$$X_{c, \text{switched}} = X_c - X_{c, \text{fixed}} = 4167 \Omega - 13.2 \Omega = 4154 \Omega \text{ ⑨}$$

The switching configuration for the switched capacitor during the afternoon period is to connect it in parallel with the fixed capacitor and the load.

Evening: The load consumes real power of 10 MW at a power factor of unity (1.0). Since the power factor is already 1.0, there is no need for power factor correction during this period. Therefore, we don't need to consider capacitors or switching configurations for the evening period. ⑩

Summary: To maximize the smallest power factor across the day, the chosen values for the capacitive reactances and switching configurations are as follows:

Morning: Fixed capacitor: $X_{c, \text{fixed}} = 13.2 \Omega$ (connected in parallel with the load and voltage source) Switched capacitor: Not used

Afternoon: Fixed capacitor: $X_{c, \text{fixed}} = 13.2 \Omega$ (connected in parallel with the load and voltage source) Switched capacitor: $X_{c, \text{fixed}} = 4154 \Omega$ (connected in parallel with the load, fixed capacitor, and voltage source)

Evening: No power factor correction is required.

The corresponding power factors for each time period are:

Morning: 0.9 leading power factor Afternoon: 0.9 lagging power factor Evening: 1.0 (unity power factor) ⑪

In contrast with o1, the GPT-4 response exhibited errors of a much more fundamental character, and in significantly greater number. Below, we summarize these errors in the order in which they are marked in the above response.

- 1) The load's power factor is not given during the morning time period, but we can infer this from the complex power consumption of $10 + j5$ MVA. Specifically, the power factor is $10/\sqrt{10^2 + 5^2} = 0.8944$ lagging. There is no need to arbitrarily choose a power factor for the load during this or any other time period.
- 2) With a complex power of $10 + j5$ MVA, the apparent power is $\sqrt{10^2 + 5^2} = 11.1803$ MVA.
- 3) The formula $Z = V^2/S$ is not correct. The correct expression is $Z = |V|^2/S^*$. The load impedance is $Z = (10 \times 10^3)^2 / (10 \times 10^6 + j5 \times 10^6)^* = 8 + j4 \Omega$.
- 4) The reactive power consumed by a shunt capacitance is $Q = -|V|^2/(X_c)$, so the capacitive reactance that would supply a 5 MVAR is $X_{c, \text{fixed}} = -|V|^2/Q = -(10 \times 10^3)^2 / 5 \times 10^6 = -20 \Omega$, not -13.2Ω . Additionally, we do not necessarily want the fixed capacitor to supply 5 MVAR. While this will lead to unity power factor in the morning, this much reactive power will lead to an oversupply of reactive power during the evening.
- 5) A capacitor has a negative reactance, so $X_{c, \text{fixed}}$ should be a negative value.
- 6) The assertion that we should aim for a power factor of 0.9 lagging is not correct. We can choose values that result in a higher power factor during all time periods.
- 7) Same as in 4), the formula $Z = V^2/S$ is not correct (and the value is also wrong even if the formula had been correct). The load impedance during the afternoon should be $Z = (10 \times 10^3)^2 / (32 \times 10^6 + j24 \times 10^6)^* = 2 + j1.5 \Omega$.
- 8) This expression for the desired capacitive reactance has both a sign error and the value of 4167Ω is not the output of this expression (off by three orders of magnitude). The value should be $(10 \times 10^3)^2 / -24 \times 10^6 = -4.1667 \Omega$.
- 9) The switched capacitor's reactance is based on the difference between the fixed capacitor's reactance and the desired reactance. However, it is not correct to directly subtract the reactances since the capacitors are connected in parallel.
- 10) This analysis of the evening load demand neglects the fact that the fixed capacitor is supplying reactive power and thus changing the power factor away from unity. The choice of the fixed capacitor's reactance should be cognizant of the resulting deviation in the power factor during the evening period.
- 11) The values in the summary are incorrect due to the errors earlier in the solution.