Solving Undergraduate Power Systems Homework Assignments with ChatGPT

Us and Paul Cuffe, *Senior Member, IEEE*

***Abstract*—Large Language Models (LLMs) have exploded into discourse lately, and their ability to complete complex educational tasks has been demonstrated again and again. Here we seek to understand the vulnerability of assessment archetypes in power system education to these models. This is done by completing a selection of common optimisation problems in power system engineering courses to reveal the threat to assessment integrity.**

I. Introduction

Language has been one of the main fields of interest for Artificial Intelligence since well before the first computers even existed. Stories abound since Antiquity of speaking automatons, such as the Brazen Heads of medieval alchemists [1]. In the early 20th century, Turing proposed his famous "Imitation Game". This posited that artificial intelligence should be measured by the ability of machines to fool humans via language [2]. Early exploration of language processing also led to the first program that could solve educational problems in 1964 with STUDENT, which was designed to solve high school maths problems [3].

Modern LLMs are a type of neural network. However, while the first artificial neuron dates to 1958 [4], the fundamental technology that holds the secret to LLM's success dates to 2017 with the "transformer" [5]. Based on understandings of human attention, these transformers "pay attention" to how key concepts interact, and are agnostic to the distance between these [6]. Via this mechanism, they can learn context and how they change word meaning [5]. For example, they can identify that "gravity" could refer to seriousness if the closest noun was "voice". The improvements in the ability of transformer models scales with size, leading to the best models having an astronomical amount of both parameters and training data - hence the term "Large" [7].

LLMs are thus systems that are based on the latest cognitive science understandings of human attention, and are trained by humans to give human-like answers [8]. This allows them to identify important concepts and mimic human responses - provided they have been given enough data on the relevant concepts [9]. However, it is important to note that these models are "stochastic parrots"; they have no underlying understanding [7]. Instead, their responses are generated based on probabilities that were learnt through a combination of summarisation, human feedback and unsupervised reinforcement learning [8]. Nonetheless, the power of contemporary LLMs have proven formidable.

The intersection of language modelling and education which started with STUDENT [3] has continued with these models. The ability of LLMs to pass fundamental benchmark assessments has been widely published recently [10]–[14]. Their abilities have even proven capable of fooling assessors - passing the educational "Imitation Game" [2], [15]. This poses obvious concerns for the integrity of assessments used across all disciplines. However, while much has been written on their ability to code [16], write essays [13] and pass multiple choice exams [10], very little yet has been written about their potential to pass higher-order engineering problems. We seek to close this gap, and identify how dangerous LLMs are to current Power Systems Engineering education.

In this paper, we examine how LLMs perform when faced with problems typically posed to students undertaking courses in power system engineering. Specifically, we gauge ChatGPT's ability to solve optimisation problems, ranging from simple graphical, two-variable problems to more complex economic dispatch problems involving the use of an optimisation solver.

– and P. Cuffe (paul.cuffe@ucd.ie) are with the School of Electrical and Electronic Engineering, University College Dublin.

II. Methodology

This work involved exposing ChatGPT to optimisation assessment tasks that are typically completed by students undertaking courses in power systems engineering. The complexity of these tasks ranges from the more simple graphical solution of a two-variable optimisation problem to a more complex economic dispatch problem. For these investigations, ChatGPT Plus was used to have access to GPT-4 with the Wolfram plug-in.

A. Simple Optimisation Problem

For the graphical solution of a simple optimisation problem, the following problem was formulated for solution by ChatGPT.

$$\underbrace{\text{minimise}}_{x,y} \quad 4x + 7y \tag{1}$$

$$x + y \leq 10 \tag{2}$$

$$-x + 2y \leq -3 \tag{3}$$

$$5x + 3y \leq 42 \tag{4}$$

$$x \geq 3 \tag{5}$$

$$y \geq 1 \tag{6}$$

$$x, y \in \mathbb{R} \tag{7}$$

While the solution of an optimisation problem such as that shown in (1)-(7) is not entirely trivial, solvers such as GAMS, Julia and Pyomo have been developed to solve such problems with relative ease. Learning different syntax for different solvers is one obstacle students overcome during the learning process here.

A technique which may be used to instil an understanding of the mechanics of solving optimisation problems is to illustrate the solution graphically. By visualising how the value of the objective function changes as the values of x and y vary, a student may understand what the *optimal* solution means. Accordingly, the ability of the Wolfram plug-in in ChatGPT Plus to produce a graph of the optimal solution will be examined. Figure 1 shows the prompt that was provided to ChatGPT.

The prompt in Figure 1 contains a clear request that a student may ask of a LLM. The syntax of the optimisation problem itself is chosen to help the tool to "read" the problem. As students are familiar with such syntax from writing reports, this is a reasonable adjustment to the prompt.

```
Solve the following optimisation problem given the
objective function and constraint set as follows, where
x and y are real numbers:

\underbrace{\text{minimise}}_{x,y}4x+7y
x+y\leq10
-x+2y\leq-3
5x+3y\leq42
x\geq3
y\geq1

Produce a graph showing x on the x-axis and y on the y-
axis and showing that the solution obtained is in fact
optimal. In the plots, include the constraints. Also
report the value of the objective function at the
optimal solution.
```

Fig. 1. Prompt provided to ChatGPT for solution of optimisation problem in (1) - (7).

B. *Economic Dispatch Problem*

Economic dispatch problems represent a realistic application of the theory of optimisation. By virtue of the real-world nature of such problems, they are more complex than the simple optimisation problem outlined earlier and require solution using a solver. To that end, ChatGPT is asked to consider the IEEE 14-bus system which includes nine generators (G1 - G9), each of which has a quadratic cost curve denoted by $C_i = a_i P_i^2 + b_i P_i + c_i$, where P_i is the output of each generator. Each generator also has maximum and minimum outputs, $P_{\max,i}$ and $P_{\min,i}$, respectively. Demand is defined as a constant value at a given point in time of 1495.2 MW. The quadratic cost curve coefficients and output constraints are detailed in Table I below.

TABLE I.

GENERATOR PARAMETERS FOR IEEE 14-BUS SYSTEM

Unit	$P_{\min,i}$	$P_{\max,i}$	a_i	b_i	c_i
G1	2.40	12.00	0.025	25.5	24.4
G2	4.00	20.00	0.012	37.6	117.8
G3	0.00	50.00	0.000	0.5	0.0
G4	15.20	76.00	0.009	13.3	81.1
G5	25.00	100.00	0.006	18.0	217.9
G6	54.24	155.00	0.005	10.7	142.7
G7	68.95	197.00	0.003	23.0	259.1
G8	140.00	350.00	0.002	10.9	177.1
G9	100.00	400.00	0.002	7.5	311.9

ChatGPT is asked to complete the following tasks for the economic dispatch problem.

1) Write GAMS code which can be used to solve the problem.
2) Formulate the problem as described in terms of an objective function and a constraint set.
3) From the formulation above, create a Lagrangian function and explain how it may be used to find the optimal solution of the optimisation problem.

The exact prompt supplied to ChatGPT may be viewed in ***APPENDIX***.

III. Results

A. *Simple Optimisation Problem*

To verify the solution reported by ChatGPT, the problem is first solved using Pyomo yielding an answer of $(x_{\text{opt}}, y_{\text{opt}}) = (5, 1)$ for the values of x and y at the optimal solution. The value of the objective function is found to be 27.

ChatGPT returns the following answer to the prompt it receives - "*The optimal solution to the given optimization problem is $x = 5$ and $y = 1$. The value of the objective function at this optimal solution is* 27."

ChatGPT also produced a graph using the Wolfram plug-in to visualise the solution to the optimisation problem. This graph is shown in Figure 2.

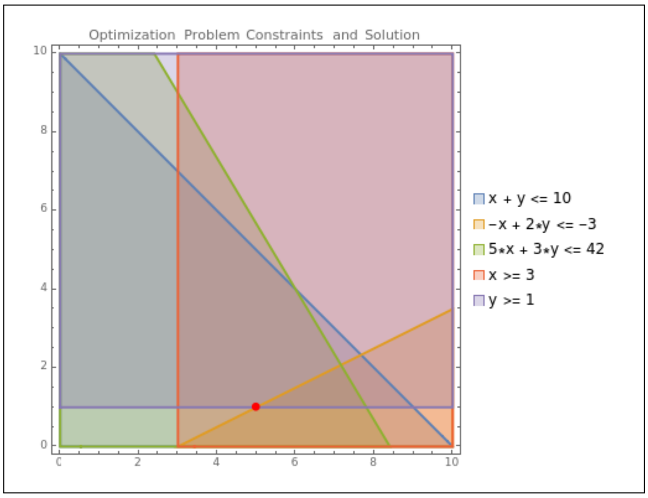

Fig. 2. Graph produced by Wolfram plug-in with ChatGPT Plus showing the values of x and y at the optimal solution at the constraints acting on the solution space.

Despite the obvious lack of axis labels and poor image quality, this graph is accurately depicts the values of the decision variables at the optimal solution.

B. *Economic Dispatch Problem*

Recall that ChatGPT was asked to perform three tasks pertaining to the economic dispatch problem - write GAMS code, formulate the problem as an objective function and constraint set and translate the problem into a Lagrangian function.

ChatGPT successfully wrote a short script using the correct syntax which may be used in GAMS to solve the optimisation problem. This script may be seen in ***APPENDIX***.

ChatGPT then successfully formulated the economic dispatch problem as an optimisation problem with an objective function and constraints as follows.

$$\text{minimise} \quad \text{Cost} = \sum_{i=\text{G1}}^{\text{G9}} (a_i P_i^2 + b_i P_i + c_i) \tag{8}$$

$$\sum_{i=\text{G1}}^{\text{G9}} P_i = \text{Demand} \tag{9}$$

$$P_{\min,i} \leq P_i \leq P_{\max,i} \quad \forall i \tag{10}$$

Finally, ChatGPT converted the optimisation problem described above into a Lagrangian function, shown in (11) below.

$$L(P,\gamma,\mu,\nu) = \sum_{i=\text{G1}}^{\text{G9}} (a_i P_i^2 + b_i P_i + c_i) + \gamma\left(\sum_{i=\text{G1}}^{\text{G9}} (P_i - \text{Demand})\right) - \sum_{i=\text{G1}}^{\text{G9}} \left(\mu_i (P_i - P_{\min,i})\right) + \sum_{i=\text{G1}}^{\text{G9}} \left(\nu_i (P_{\max,i} - P_i)\right) \tag{11}$$

Further, ChatGPT described how the Lagrangian function may be used to solve the economic dispatch problem, stating "To solve the problem by hand using the Lagrangian function, you would set the derivative of the Lagrangian with respect to each variable (the P_i, γ, μ_i and ν_i) equal to zero and solve the resulting system of equations. This gives you the values of the variables that minimise the objective function subject to the constraints". It goes on to say that μ_i and ν_i must be non-negative for the constraints to be satisfied.

IV. Conclusions

We started by identifying that LLMs have exploded in their ability in recent years. Language-processing Artificial Intelligences have long been used to perform educational tasks [3], and the increase in power has massively expanded their potential for use or misuse in this field. Educational institutes have been waking up to this danger, with the list of assessments shown to be vulnerable to these models growing day-by-day [10]–[14]. This work has examined the particular abilities of a LLM to complete higher order tasks typically encountered by undergraduate students undertaking courses in power system engineering. The methodology used in these investigations systematically shows how LLMs may fare against increasingly more abstract and applied concepts faced by students.

The power of recent advances in tools associated with LLMs is best demonstrated in Section III-A. Here, not only was ChatGPT able to solve an optimisation problem, albeit a relatively simple one, but it was also able to produce a graph showing the values of the decision variables at the optimal solution. Where such a task might previously have been used by educators as documentary evidence of a student's effort, the same deliverable can now be achieved with minimal student effort or understanding of the mechanics underpinning the solution of such problems.

This raises obvious concerns for the problem in Section III-B. Where a student does not understand a simple theoretical example, this poses a risk to the ability of the student to demonstrate their own competence in more applied, "real-world" problems. These results show that ChatGPT can write GAMS code which can be pasted into the solver to execute the solution of the economic dispatch problem, formulate the problem mathematically and translate the mathematical formulation into a Lagrangian function.

The recent explosion of LLMs has been heralded from many quarters as a useful tool to make certain difficult tasks easier. However, it also raises important questions surrounding the integrity of assessments. The ability of such models to complete multiple choice questions and lower order assessment tasks has been widely published. This paper demonstrates that the assumption that more niche disciplines and tasks are "safe" is not valid. This is particularly concerning when the work of LLMs is difficult to distinguish from the honest work of a student.

Does this matter? Some might say that if the black box gives the correct answer, let it do all the hard work. A principle of academic integrity is that one is assessed and awarded qualifications based on *one's own* ability and to submit work that is *one's own* - not that of a black box or whatever is inside it.

References

[1] K. LaGrandeur, "The talking brass head as a symbol of dangerous knowledge in friar bacon and in alphonsus, king of aragon," 1999.

[2] A. M. Turing, *Computing machinery and intelligence*. Springer, 2009.

[3] D. Bobrow *et al.*, "Natural language input for a computer problem solving system," 1964.

[4] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain.," *Psychological review*, vol. 65, no. 6, p. 386, 1958.

[5] A. Vaswani *et al.*, "Attention is all you need," in *Advances in Neural Information Processing Systems*, I. Guyon *et al.*, Eds., vol. 30, Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

[6] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *WIREs Data Mining and Knowledge Discovery*, vol. 8, no. 4, e1253, 2018. doi: https://doi.org/10.1002/widm.1253. eprint: https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1253. [Online]. Available: https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/widm.1253.

[7] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the dangers of stochastic parrots: Can language models be too big?" In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, 2021, pp. 610–623.

[8] N. Stiennon *et al.*, "Learning to summarize with human feedback," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., vol. 33, Curran Associates, Inc., 2020, pp. 3008–3021. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2020/file/1f89885d556929e98d3ef9b86448f951-Paper.pdf.

[9] J. Rudolph, S. Tan, and S. Tan, "Chatgpt: Bullshit spewer or the end of traditional assessments in higher education?" *Journal of Applied Learning and Teaching*, vol. 6, no. 1, 2023.

[10] L. Passby, N. Jenko, and A. Wernham, "Performance of ChatGPT on dermatology Specialty Certificate Examination multiple choice questions," *Clinical and Experimental Dermatology*, llad197, Jun. 2023, issn: 1365-2230. doi: 10.1093/ced/llad197. eprint: https://academic.oup.com/ced/advance-article-pdf/doi/10.1093/ced/llad197/50506550/llad197.pdf. [Online]. Available: https://doi.org/10.1093/ced/llad197.

[11] P. Giannos and O. Delardas, "Performance of chatgpt on uk standardized admission tests: Insights from the bmat, tmua, lnat, and tsa examinations," *JMIR Med Educ*, vol. 9, e47737, Apr. 2023, issn: 2369-3762. doi: 10.2196/47737. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/37099373.

[12] R. Bhayana, S. Krishna, and R. R. Bleakney, "Performance of chatgpt on a radiology board-style examination: Insights into current strengths and limitations," *Radiology*, vol. 307, no. 5, e230582, 2023, PMID: 37191485. doi: 10.1148/radiol.230582. eprint: https://doi.org/10.1148/radiol.230582. [Online]. Available: https://doi.org/10.1148/radiol.230582.

[13] D. M. Katz, M. J. Bommarito, S. Gao, and P. Arredondo, "Gpt-4 passes the bar exam," *Available at SSRN 4389233*, 2023.

[14] V. Pursnani, Y. Sermet, and I. Demir, "Performance of chatgpt on the us fundamentals of engineering exam: Comprehensive assessment of proficiency and potential implications for professional environmental engineering practice," *arXiv preprint arXiv:2304.12198*, 2023.

[15] M. Khalil and E. Er, "Will chatgpt get you caught? rethinking of plagiarism detection," *arXiv preprint arXiv:2302.04335*, 2023.

[16] B. Qureshi, "Exploring the use of chatgpt as a tool for learning and assessment in undergraduate computer science curriculum: Opportunities and challenges," *arXiv preprint arXiv:2304.11214*, 2023.