Assessing the Capabilities of Modern Large Language Models in Completing Power System Optimization Coursework

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

The advent of Large Language Models (LLMs) has captured the interest of students and experts in power systems engineering alike. However, this interest has been tempered with nascent fears of the dangers of misuse and academic dishonesty [1]. Ever since the early 20th century and the publishing of Turing's famous "Imitation Game" [2], AI language models have expressly sought to be able to pass as humans. While tools to solve educational problems date to 1964 [3], authorities in academic institutions worry that students may leverage recent advances in LLMs to complete and submit assessments in a dishonest fashion [4]–[8]. This leads us to examine the capabilities of these models - what power system engineering assessments can these models pass?

Chatbots and neural networks have been in existence for some time - the first artificial neuron dates to 1958 [9], and chatbots to 1966 [10]. What differentiates LLMs from previous chatbots is the "transformer" [11]. The transformer enables a LLM to "pay attention" to the context in which words are used [11]–[13]. The linguistic ability of these transformer models scales with size - resulting in models having an astronomical amount of parameters and training data, hence the term "Large'

LLMs generate responses based on the probability that they will be correct - a stochastic process which tries to emulate the processes of the human brain [13], [14]. In the model used in this letter's work, ChatGPT, these probabilities were learned through through a combination of summarisation, human feedback and unsupervised reinforcement learning [12]. Once it satisfied its human trainers, it was given massive amounts of data to absorb [12], [15]. Despite the fact that contemporary language models are "stochastic parrots", with no built-in language understanding [13], their linguistic and technical abilities have prompted widespread interest [5], [7], [16]–[18].

The intersection of language modelling and education which started with STUDENT [3] has continued with these models. While much has been written on their ability to code [19], write essays [20] and pass multiple choice exams [21], fewer publications consider the abilities of LLMs to complete higher-order engineering problems. We seek to close this gap by identifying the capabilities of LLMs in this particular domain. In this letter, 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

To go beyond a qualitative sentiment-based analysis of the threat to assessment integrity posed by LLMs, it is useful to see first-hand how such tools fare at completing tasks typically completed by students. This involves exposing a LLM to assessments that are easily recognisbale within a discipline as being archetypal of those completed by students of that discipline.

III. Case Study

Optimisation was identified as an area with particular relevance to students of power system engineering [22]. A LLM was exposed to an exemplar assignment combining theoretical, applied, general and specific aspects of optimisation which such students typically encounter in their coursework.

Released in late 2022, ChatGPT has "has demonstrated state-of-the-art performance" on a wide range of tasks and its abilities are continually advancing [20]. Of particular interest is the ability of plug-ins to assist with more complex tasks. For this reason, ChatGPT was the chosen LLM for this investigation.

The exemplar assignment shown in Figure 1 contains a highly theoretical and relatively simple linear programming problem accompanied by a question asking students to create a graph to visualise the constraints and optimal solution. Such a task is designed to ensure that students understand the mechanics of optimisation. Further, students are asked to complete an economic dispatch of a system of generators subject to maximum and minimum output constraints and unique cost curves. This task requires students to translate a "real-world" problem into a mathematical formulation, demonstrate an understanding of how such a problem may be solved using the method of Lagrange multipliers and use GAMS to obtain the solution to the problem.

The prompt in Figure 2 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.

Similarly, for (c) in Figure 1, the LLM was supplied a detailed prompt which explains the minimum and maximum output limits and the nature of the cost curves, supplied in the form of a table. The prompt can be found in full in the Appendix. The LLM is asked to complete the following tasks for the economic dispatch problem.

Optimisation Assignment

a) Find the optimal value of the following linear programming problem.

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

$$x+y \leq 10$$
$$-x+2y \leq -3$$
$$5x+3y \leq 42$$
$$x \geq 3$$
$$y \geq 1$$
$$x, y \in R$$

b) Illustrate the solution space of the problem in (a), showing the feasible region and the location of the value of x and y at the optimal solution.

c) A power system consists of nine generators, each with cost curves of the form $C_i(P_i) = a_iP_i^2 + b_iP_i + c_i$, where P_i is the output of unit i (in MW). Each generator has a maximum and minimum output. These parameters are given below. Assume all generators are online. At a given time, the system demand is 1495.2 MW. Determine the optimal dispatch for the power system which minimises fuel costs.

- Formulate the problem mathematically with an objective function and constraints.
- Determine the Lagrangian function for the optimisation problem and explain how it may be used to find the optimal solution.
- Determine the optimal dispatch using GAMS.

Unit	P_{min} (MW)	P_{max} (MW)	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

Fig. 1. Exemplar optimisation assignment featuring questions ranging from theoretical to application and general and specific tasks [23].

```
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. 2. Prompt provided to ChatGPT for solution of optimisation problem in (a) and (b) in Figure 1.

1) Formulate the problem as described in terms of an objective function and a constraint set.
2) Determine the corresponding Lagragian function for the problem and describe how it may be used to arrive at the optimal solution.
3) Write GAMS code which may be used to solve the economic dispatch problem.

IV. Results

A. Linear Programming Problem

To verify the solution reported by the LLM, 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 optimal value of the objective function is found to be 27.

The response from the LLM is shown in Figure 3 below.

> 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.

Fig. 3. LLM response to prompt requesting solution to linear programming problem.

B. Graphical Solution of Linear Programming Problem

As requested, the LLM also produced a graph using the Wolfram plug-in to visualise the solution to the optimisation problem. The graph shows $(x_{\text{opt}}, y_{\text{opt}})$ at the optimal solution and how the optimal solution is constrained within the solution space. This graph is shown in Figure 4.

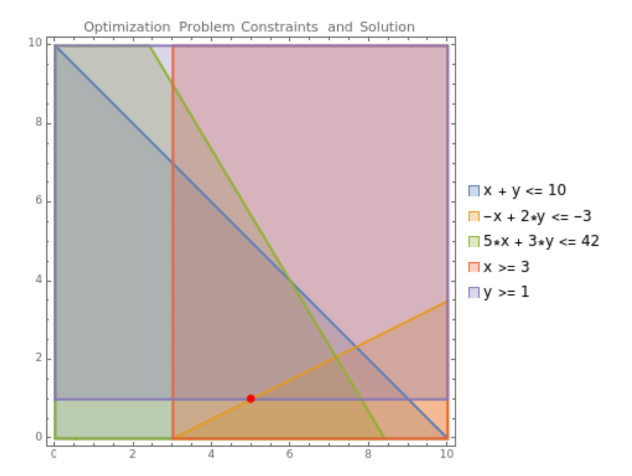

Fig. 4. Graph produced by LLM showing the values of x_{opt} and y_{opt} at the optimal solution and the constraints acting on the solution space.

C. Economic Dispatch Problem

Recall that the LLM was asked to perform three tasks pertaining to question (c). First, the LLM was able to translate the problem from a textual description to mathematical statements. Furthermore, these mathematical statements were at least partially formatted, as shown in Figure 5.

Second, the LLM returned the Lagrangian function for the optimisation problem it formulated as shown in Figure 5, again formatted mathematically. An explanation was also provided as to how the Lagriangian function may be used to arrive at the optimal solution. This explanation is also shown in Figure 6.

$$\min_{P_i} \sum_{i=1}^{9} C_i = \sum_{i=1}^{9} (a_i * P_i^2 + b_i * P_i + c_i)$$

$$\sum_{i=1}^{9} P_i = 1495.2$$

$$P_{\text{min},i} \leq P_i \leq P_{\text{max},i}$$

Fig. 5. Formatted optimisation equations returned by the LLM for the economic dispatch problem.

$$L(P_i, \lambda, \mu_i, \nu_i) = \sum_{i=1}^{9}(a_i * P_i^2 + b_i * P_i + c_i) + \lambda(\sum_{i=1}^{9} P_i - 1495.2) - \sum_{i=1}^{9}\mu_i(P_i - P_{\min,i}) + \sum_{i=1}^{9}\nu_i(P_i - P_{\max,i})$$

The optimal solution is found by setting the derivative of the Lagrangian function with respect to P_i, λ, μ_i, and ν_i to zero and solving the resulting system of equations. This gives the optimal power output of each generator and the optimal values of the Lagrange multipliers.

Fig. 6. Lagrange function for the linear programming problem and explanation as to how it should be used provided by the LLM.

Finally, the LLM wrote code which may be copied and used to find the solution to the economic dispatch problem using GAMS. This code is shown in Figure 7.

```
Set i /G1*G9/;

Alias (i,j);

Parameters
    a(i) /G1 0.025, G2 0.012, G3 0, G4 0.009, G5 0.006, G6 0.005, G7 0.003, G8 0.002,
          G9 0.00211/
    b(i) /G1 25.5, G2 37.6, G3 0.5, G4 13.3, G5 8, G6 10.7, G7 23, G8 10.9, G9 7.5/
    c(i) /G1 24.4, G2 117.8, G3 0, G4 81.1, G5 217.9, G6 142.7, G7 259.1, G8 177.1,
          G9 311.9/
    Pmin(i) /G1 2.4, G2 4, G3 0, G4 15.2, G5 25, G6 54.24, G7 68.95, G8 140, G9 100/
    Pmax(i) /G1 12, G2 20, G3 50, G4 76, G5 100, G6 155, G7 197, G8 350, G9 400/
    demand /1495.2/;

Variables
    P(i)
    cost;

Positive Variable P;

Equations
    obj
    power_balance
    lower_limit(i)
    upper_limit(i);

obj..
    cost =e= sum(i, a(i)*P(i)*P(i) + b(i)*P(i) + c(i));

power_balance..
    sum(i, P(i)) =e= demand;

lower_limit(i)..
    P(i) =g= Pmin(i);

upper_limit(i)..
    P(i) =l= Pmax(i);

Model dispatch /all/;

Solve dispatch using NLP minimizing cost;
```

Fig. 7. GAMS code produced by the LLM.

V. Conclusions

Cases where LLMs have been able to produce answers that can pass assessments across different disciplines have been widely published [5]–[8], [21]. As these tools develop, focus should rightly be placed on their ability to complete complex tasks in niche areas of particular disciplines and the implications for assessment integrity. The work presented in this letter shows how a modern LLM already has the capability to complete higher-order assessment tasks within the discipline of power system engineering.

The response to question (a) shows that the LLM can solve a linear programming problem. The response to question (b) is more compelling as it demonstrates the ability of a LLM to produce a graph to document *its* understanding of the task at hand. This raises significant concerns where such a task may previously have been used to document students' understanding of the problem; it is now conceivable that a student may be able to earn credit from the work of a LLM while not understanding the task at hand.

The LLM's response to question (c) displays its higher-order-thinking ability. A textual description of a power system is converted to a mathematical formulation with ease. Further, this is converted to a Lagrangian function with a description of how this function may be used for optimisation. The ability of the LLM to write GAMS code is also surprising as GAMS is not as widely used as other optimisation solvers.

The rapid growth in the ability of LLMs has been heralded by many as a useful tool for completing difficult tasks. In education, the use of LLMs as tutors could significantly improve outcomes for students. However, it is important that the output of LLMs is not plagiarised in order to earn academic credit. For educators, it is important to instil in students an understanding of LLMs may be used responsibly and irresponsibly and to safeguard the integrity of assessments they deliver.

VI. Appendix

The interaction with the LLM may be viewed here: (LINK).

References

[1] M. R. King and ChatGPT, "A conversation on artificial intelligence, chatbots, and plagiarism in higher education," *Cellular and Molecular Bioengineering*, vol. 16, no. 1, pp. 1–2, 2023.

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

[5] 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.

[6] 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. DOI: 10.1148/radiol.230582.

[7] D. M. Katz, M. J. Bommarito, S. Gao, and P. Arredondo, "Gpt-4 passes the bar exam," 2023. DOI: 10.2139/ssrn.4389233.

[8] 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.

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

[10] J. Weizenbaum, "Eliza—a computer program for the study of natural language communication between man and machine," *Commun. ACM*, vol. 9, no. 1, pp. 36–45, Jan. 1966, ISSN: 0001-0782. DOI: 10.1145/365153.365168. [Online]. Available: https://doi.org/10.1145/365153.365168.

[11] 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.

[12] 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.

[13] 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.

[14] M. F. Bonner and R. A. Epstein, "Object representations in the human brain reflect the co-occurrence statistics of vision and language," *Nature communications*, vol. 12, no. 1, p. 4081, 2021.

[15] 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.

[16] A. Shoufan, "Exploring students' perceptions of chatgpt: Thematic analysis and follow-up survey," *IEEE Access*, 2023.

[17] B. Gordijn and H. t. Have, "Chatgpt: Evolution or revolution?" *Medicine, Health Care and Philosophy*, vol. 26, no. 1, pp. 1–2, 2023.

[18] M. Javaid, A. Haleem, and R. P. Singh, "Chatgpt for healthcare services: An emerging stage for an innovative perspective," *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, vol. 3, no. 1, p. 100 105, 2023.

[19] 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.

[20] E. Kasneci *et al.*, "Chatgpt for good? on opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, p. 102 274, 2023.

[21] 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.

[22] I. J. Pérez-Arriaga, "Session 7: Generation wholesale markets - basic economic functions," in MIT OpenCourseWare, Cambridge MA, 2010. [Online]. Available: https://ocw.mit.edu/courses/ids-505j-engineering-economics-and-regulation-of-the-electric-power-sector-spring-2010/resources/mitesd_934s10_lec_07/.

[23] T. G. Hlalele, R. M. Naidoo, J. Zhang, and R. C. Bansal, "Dynamic economic dispatch with maximal renewable penetration under renewable obligation," *IEEE Access*, vol. 8, pp. 38 794–38 808, 2020.