

Teaching Maxwell Equations with LLM Assistance

Abstract. Large Language Models (LLMs) have been gaining tremendous popularity since early 2023 with the release of the GPT-4 system. The new artificial intelligence methods have succeeded in many fields of activity, including education. This article analyses the applicability of LLMs in supporting the teaching of the "Electromagnetic Fields" course in Electrical Engineering studies. Examples of use in various teaching activities are presented. Experiments conducted using four commercial platforms have shown that the responses generated by the GPT-4 model are clearly of the highest quality. The advantages but also risks associated with the use of the proposed methods were discussed.

Streszczenie. Duże modele językowe (ang. LLM) zysują ogromną popularność od początku 2023 roku wraz z wydaniem systemu GPT-4. Nowe metody sztucznej inteligencji odniosły sukces w wielu dziedzinach, w tym w edukacji. W niniejszym artykule przeanalizowano możliwość zastosowania LLM we wspieraniu nauczania przedmiotu "Podstawy elektromagnetyzmu" na studiach na kierunku Elektrotechnika. Eksperymenty przeprowadzone z wykorzystaniem czterech platform komercyjnych pokazały, że odpowiedzi generowane przez model GPT-4 są zdecydowanie najwyższej jakości. Omówiono zalety, ale i zagrożenia związane z wykorzystaniem proponowanych metod. (**Uczenie równań Maxwella przy wykorzystaniu asystenta LLM**)

Keywords: AI, LLM, Engineering Teaching

Słowa kluczowe: sztuczna inteligencja, nauczanie inżynierów

Introduction

Electromagnetism, a fundamental pillar of physics and electrical engineering, is often perceived as one of the more challenging subjects for students to grasp [1]. The abstract nature of the concepts, coupled with the mathematical complexity, can pose significant hurdles in the learning process. Traditionally, teaching assistants have played a crucial role in bridging this gap, providing personalized guidance and facilitating a deeper understanding of the subject matter [2].

In recent years, the landscape of education has been transformed by the increasing role of artificial intelligence (AI) methods in teaching [3]. AI-powered tools, such as intelligent tutoring systems and adaptive learning platforms, have shown promise in enhancing student learning outcomes and personalizing education [4]. These advancements have been further propelled by the breakthrough associated with the surge in popularity of Large Language Models (LLMs) like GPT-3 and GPT-4, which have demonstrated remarkable capabilities in generating human-like text [5].

This paper focuses on the application of AI, particularly LLMs, in teaching the course "Electromagnetic Fields" in undergraduate electrical engineering studies. The analyses carried out relate to a topic that is being actively developed. Given the rapid pace of advancements in this field, significant scientific publications are emerging weekly. The status of the field, as described in this paper, is accurate as of the end of June 2023. The aim of the author of this paper is to contribute to the discussion [6] on the opportunities, but also the risks, associated with the use of artificial intelligence in the academic educational process.

Course - Electromagnetic Fields

The course "Electromagnetic fields" is one of the core subjects of the electrical engineering degree programme at Warsaw University of Technology. It consists of three complementary parts: Lecture (30h), where the theory of Maxwell Equations is presented; Tutorials(15h), where problem-solving skills are practised; and Laboratory (15h), where real experiments are being carried out. The learning outcomes are verified as follows: quizzes during exercises account for 20% of the total grade, reports from experiments contribute another 20%, and the final examination constitutes the remaining 60%.

The academic lecturer who is leading the course is supported by Teaching Assistants (TA). They often serve as stu-

dents' primary point of contact, providing personalized guidance and clarification on course content. TAs are also responsible for conducting tutorial sessions and laboratory exercises. These sessions allow students to apply theoretical knowledge in a practical context, reinforcing their understanding of the subject. Finally, TAs play a crucial role in the assessment of students. They are typically responsible for grading assignments, quizzes, and laboratory reports.

Students often encounter several key difficulties when studying electromagnetism. The abstract nature of electromagnetic concepts can be challenging to visualize and comprehend. The mathematical complexity of solving Maxwell's equations can be daunting, particularly for those who struggle with advanced calculus. Lastly, applying these theoretical concepts to real-world problems often poses a significant hurdle, as students may find it difficult to translate their theoretical knowledge into practical solutions. On the other hand, this area of knowledge has been stable for more than 100 years and is widely described in the literature. This provides a great deal of support for general language models.

As will be discussed, LLM tools can provide personalized learning experiences, identify gaps in students' understanding, and tailor their teaching strategies accordingly. Moreover, AI-TA can automate administrative tasks, thereby allowing real teachers to focus more on pedagogical human-to-human activities.

Prompts for LLM

The results received by the models highly depend on the prompt. Writing an effective prompt for large language models like GPT-4 is both an art and a science. A well-crafted prompt can guide the model to generate highly relevant responses. The practice teaches the following principles:

- Precision - The more specific and detailed the prompt, the more likely the model will produce a response that aligns with the desired outcome.
- Provide context - Including relevant background information in the prompt can help the model generate a more contextually appropriate response.
- Request format - Mention in the prompt what formatting you are expecting and how long the answer should be.
- Iterative refinement - experiment with your prompts; even small changes could give much improvement.

Current transformer-based language models are based on large-scale processing (like GPT-4 has approx. 1.8 tril-

lion parameters across 120 layers of artificial neural network). They can generate remarkably coherent and contextually relevant text over extended passages. This massive scale allows the model to exhibit a deep understanding of a wide array of topics, making it a powerful tool for various applications, including but not limited to content creation, programming help, and even teaching assistance. However, these skills and knowledge are generic in time and space. Therefore, effective use of this tool requires that information about the context of a given query is communicated to the model. The lack of context is due to the fact that current language models lack memory but also because they have not been trained on restricted materials.

Prompt: Act like a university teacher on the "Electromagnetic fields" course in electrical engineering undergraduate studies.

Output: (...)

Prompt Engineering

Prompt engineering refers to the practice of crafting effective prompts to guide the responses of large language models. It is crucial, as the input prompt heavily influences the output quality.

It has been shown that effective prompt engineering allows to significant increase the quality of teaching experience for computer science students [7]. The study reveals that engineered prompts containing explicit algorithmic hints effectively aid students in solving programming problems, fostering deeper understanding and translating abstract concepts into concrete computational steps.

The paper [8] discusses the use of deep neural networks and prompt engineering for automatic essay scoring. The authors found that engineering the prompts to contain explicit algorithmic hints was effective across almost all categories of problems.

Some key techniques of prompt engineering:

- **Zero-shot:** This technique involves training a model on a task without providing any examples of that task during training. The model is expected to generalize its learning from other tasks to perform the new task.
- **Few-shot:** The model provides a few examples of the task during training in this approach. The model uses these examples to understand the task and generate responses.
- **Chain-of-thought:** This technique provides a series of prompts that guide the model through a specific line of reasoning or a sequence of steps. This can be particularly useful for complex tasks requiring multiple steps [9].
- **Three-of-thoughts:** This technique is a variation of chain-of-thought prompting, where the model is guided through a three-step process to generate a response [10].

OpenAI released a set of GPT best practices¹ such as, splitting complex tasks into simpler subtasks, or giving model time to "think". The suggested methods are astonishingly similar to the advice that any person experienced in solving difficult problems would give.

Extending Knowledge of Model

The prompt-engineering techniques described above are based on the fact that all the chat content is passed to

¹<https://platform.openai.com/docs/guides/gpt-best-practices/>

the model in a single query. However, every LLM system has limitations on the query size (currently up to 32k words). This blocks the possibility of entering large contexts, such as academic textbooks, directly in the query. However, there are two main techniques to manage this issue: fine-tuning and Retrieval-Augmented Generation.

The fine-tuning process [11] starts with a pre-trained model that has been trained on a large corpus of text, such as the entirety of Wikipedia or a large subset of the internet. This phase allows the model to learn various language patterns, structures, and semantics. The pre-trained model is then trained further on a smaller, task-specific dataset. This dataset is typically composed of examples that are directly relevant to the task the model will be performing. For example, if the task is to generate responses to student questions in an electromagnetism course, the fine-tuning dataset might consist of pairs of student questions and expert responses from previous years. This leads to significantly improved performance on the task, as the model is better able to generate responses that are contextually appropriate and relevant [12].

The second approach is the Retrieval-Augmented Generation (RAG) [13], where segments of knowledge source are used to extend the model's knowledge beyond its training data. For example, by performing operations on the embeddings, the model can infer relationships between words that were not explicitly present in the training data [14]. Using a measure of embeddings, the system can efficiently find the relevant fragment in a large set of source data and then use it to build the correct context given in the prompt.

It should also be noted that the fine-tuning process is much more computationally complex than using embeddings. This carries a higher financial cost associated with this technique. In addition, RAG reduces data privacy risks, as the embeddings vector can be computed on the client side.

Possible applications

The LLM technology has a high potential to support teaching. Several examples inspired by the 'Electromagnetic Fields' course will be discussed and demonstrated. However, they could be easily transferred to any other university course.

The experiments presented in this paper were performed using commercial systems from OpenAI, Google and Microsoft. However, it is important to recognise the rapid development of publicly open models such as LLaMA [15] and its derivatives. They are expected to reach comparable quality in the coming months while simultaneously providing the possibility to control the data fully. Recent studies suggest that even small-size models could achieve good performance [16].

A1) Organisational information - The LLM has the ability to answer questions asked in natural language and relating to the content of specific documents. In the case of a university course, these are the course regulations, which describe all organisational rules, resource materials and rules for knowledge verification and final assessment. It is, therefore, sufficient to communicate the course regulations in the context of a conversation, and our virtual assistant will know how to answer questions. It is advisable to introduce a restriction for questions that do not have an answer in the regulations. For example: *If the document does not contain the information needed to answer this question, write: "There is nothing about this in the course rules. Ask the lecturer."*

Clearly written course rules mean these tasks are normally completed without error. Sample question:

Prompt: How many points do I need from tutorials to pass the course?

Output: To pass the course, you need to accumulate a minimum of 151 points in total from all course components, including lectures, tutorials, and labs.

(...)

So, while it's important to aim for a high score in your tutorials, you should also focus on the lectures and labs to ensure you accumulate enough points to pass the course.

A2) Short quizzes

Short quizzes during the lecture are breakers that allow students to check their understanding of the content being taught. Their results also serve as a suggestion to the lecturer as to which element needs to be covered more thoroughly.

The tested AI assistant worked excellently when creating short and simple screening questions. Sample below:

Prompt: Prepare a quiz with three multiple-choice questions for lecture "Electromagnetic induction".

Output: 1. What is Faraday's Law of Induction primarily concerned with?

- a) The creation of a magnetic field by an electric current
 - b) The creation of an electric field by a changing magnetic field
 - c) The creation of a magnetic field by a static electric field
- (...)

A3) Individual explanation

While studying, students can lose their train of thought and have surprisingly simple questions. Asking them to the AI Assistant does not raise reputational concerns.

The implementation of such tasks should be evaluated as excellent. The ability to repeatedly explain the same issue seems particularly valuable. Two examples are given below.

Prompt: What is D?

Output: In the context of electromagnetism, \vec{D} is known as the electric displacement field or simply the electric displacement. It's a vector field that is related to the electric field \vec{E} and the polarization density \vec{P} of a medium.

The relationship between \vec{D} , \vec{E} , and \vec{P} is given by the equation:

$$\vec{D} = \epsilon_0 \vec{E} + \vec{P}$$

(...)

Prompt: What is dA in Gauss Law?

Output: In Gauss's Law, the term $d\vec{A}$ represents a differential area element on the Gaussian surface. This is a small, infinitesimally tiny piece of the surface over which we are integrating. The "d" signifies a differential or small element, and "A" stands for area.

The vector notation \vec{A} indicates that this area element has a direction associated with it. By convention, for a closed surface, this direction is taken to be the outward normal to the surface at that point.

(...)

A4) Individual tasks

The model can consider the history of questions from the students and the explanations they have received. It is

capable, therefore, of preparing individual questions to check the student's progress.

This is a fairly straightforward task for the LLM, where the main challenge seems to be the consistent storage of information about the student's state of knowledge between multiple learning sessions.

Prompt: Students asked you a question: "What is dA in Gauss Law?" You answered with a detailed explanation of the theory and practice of area integration. What question would you ask to check student understanding of this topic?

Output: Consider a point charge Q located at the centre of a spherical Gaussian surface of radius r . The electric field \vec{E} due to the point charge is directed radially outward and its magnitude is given by $\vec{E} = k_e \frac{Q}{r^2}$, where k_e is Coulomb's constant.

Using Gauss's Law, can you calculate the total electric flux through the spherical surface?

A5) Exam problems

The model's linguistic prowess and broad knowledge allow it to prepare any number of problems testing students' skills. It thus becomes possible to easily generate unique question sets of reproducible quality and a controlled level of difficulty.

The experiments carried out have shown that the model can prepare sets of tasks of varying difficulty. The problem remains the appropriate calibration of questions between subjects and the proper valuation of their difficulty. In this case, detailed control by the human lecturer is recommended.

Prompt: Formulate two problems to check student skills on the topic "Forces in a magnetic field". The first should be simple (5 points), and the second difficult (10 points).

Output: Problem 1 (3 points):

A proton is moving with a speed of 3×10^6 m/s in a uniform magnetic field of 0.2 T. The angle between the velocity of the proton and the magnetic field is 90 degrees.

Calculate the force experienced by the proton due to the magnetic field.

Note: The charge of a proton is 1.6×10^{-19} C.

Problem 2 (6 points):

A rectangular loop of wire with dimensions 0.1 m by 0.2 m carries a current of 55 A. The loop is placed in a uniform magnetic field of 0.5 T such that the plane of the loop makes an angle of 30 degrees with the magnetic field.

- Calculate the magnetic force on each side of the loop.

- Determine the net magnetic force on the loop.

- What is the torque acting on the loop due to the magnetic field?

A6) Checking solutions

Assessing tests and answering simple questions are not a problem for the language model. Therefore, it has been decided to test its performance in a more difficult task, which consisted of a problem that required combining knowledge from different branches of electromagnetism and checking open-text written answers that contained several mistakes.

This time, the answers provided by LLM are of a decidedly lower quality. Many repetitions of the experiment were performed to observe the variability. The model had difficulty spotting errors, even errors as obvious as the 'dot product' mistake. Consequently, the scoring of student responses ranged from 4 to 8 points. Such a large discrepancy is definitely not acceptable when assessing real student work. One example below:

Prompt: Students were asked a question: Calculate value of the Poynting vector inside the planar capacitor, which has capacitance C , connected to the constant voltage source U . Then student answered: Poynting vector is a measure of energy flow defined as a dot product between electric and magnetic field. $S = E \cdot H$. In capacitor electric field can be calculate as $E=U/C$. Capacitors are not conducting, so there is no current and also no magnetic field, $H = 0$. Since H is zero, so Poynting vector is also zero.
Check above answer, and grade it in the scale 0-10 points.

Output: ((various quality answers scored from 4 to 8))

The high variability of answers is probably due to the too vague definition of the problem for the LLM. Detailed guidelines for evaluating answers should help considerably.

Experiments with commercial models

For each of the TA activities described above, three tasks were prepared to test the quality of the model's output. Together with the full context, these were then submitted to commercial platforms based on LLM models: OpenAI GPT-4, OpenAI GPT-3.5, Microsoft Bing, Google Bard. Experiments were conducted in June, July and August 2023.

| | GPT4 | GPT3.5 | Bing | Bard |
|--------------|-------|--------|-------|-------|
| A1) Info | 5,5,5 | 3,3,4 | 2,2,2 | 2,1,2 |
| A2) Quiz | 4,5,4 | 5,4,4 | 2,2,3 | 2,3,1 |
| A3) Expl. | 5,5,5 | 5,5,4 | 2,4,3 | 2,4,1 |
| A4) Tasks | 5,5,5 | 3,3,2 | 3,3,2 | 4,1,1 |
| A5) Exams | 5,4,5 | 5,5,4 | 3,2,4 | 2,2,4 |
| A6) Checking | 4,3,5 | 3,3,4 | 2,1,2 | 2,3,3 |

Table 1. Evaluation results for LLM assistant for different groups of tasks (scale from 1 - very bad; to 5 - excellent)

Results of the assessment are presented in Table 1. It should be stressed that the output of LLM is not deterministic, and the systems are constantly evolving, so repeating the experiment never produces an identical result. For this reason, the evaluation of the generated response text is rather qualitative. The assessment was done by the author of this article using a typical student rating scale (1-5).

Conclusions

Experiments have shown that creating a virtual academic assistant for the Electromagnetic Fields course is an entirely achievable goal. Most of the tasks do not present difficulties for the most advanced platforms like GPT-4. However, unconditional trust would be a mistake. Even if a model would have perfect knowledge and skills, mistakes due to misunderstanding will always remain. Therefore, supervision and verification of the work are necessary, just as it is for humans acting as assistants. By building the context, using prompt engineering techniques and adding external expertise, the quality of the model can be significantly improved.

The main challenge now seems to be creating an interface that allows voice and drawing communication with the model. This is not an easy task, but the vision of a virtually unlimited number of assistants always ready to support students individually can raise academic teaching to a whole new level.

However, it must be remembered that an academic career is based on the gradual acquisition of teaching experience while carrying out one's research. Therefore, eliminating teaching human assistants introduces the risk of losing continuity in forming university staff. Unless we introduce AI students, who would constitute a virtual teaching-learning en-

vironment for human assistants.

Acknowledgment

This paper has been prepared with the assistance of various AI-supported tools, such as chat Chat-GPT, Grammarly, Deepl, Overleaf, Bard, Bing, Science (OpenAI plugin), Wolfram Alpha and others.

Authors: dr hab. inż. Bartosz Sawicki, Institute of Theory of Electrical Engineering, Measurement and Information Systems, Faculty of Electrical Engineering, Warsaw University of Technology, ul. Koszykowa 75, 00-662 Warszawa, Poland, email: bartosz.sawicki@pw.edu.pl

REFERENCES

- [1] J. L. Docktor and K. Heller, "Students' conceptual understanding and attitudes in mechanics: A comparison of undergraduate students in introductory physics and engineering mechanics courses," *Physical Review Physics Education Research*, vol. 12, no. 1, p. 010130, 2016.
- [2] S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth, "Active learning increases student performance in science, engineering, and mathematics," *Proceedings of the National Academy of Sciences*, vol. 111, no. 23, pp. 8410–8415, 2014.
- [3] L. Zhang, Y. Wang, and Z. Liu, "AI in education: Recent trends and future directions," *Journal of AI and Education*, vol. 1, no. 1, pp. 1–20, 2023.
- [4] I. Roll, V. Aleven, B. M. McLaren, and K. R. Koedinger, "Evaluation of an intelligent tutoring system in mathematics," *Educational Technology & Society*, vol. 19, no. 2, pp. 77–92, 2016.
- [5] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, "Language models are few-shot learners," *arXiv preprint arXiv:2005.14165*, 2020.
- [6] A. Tamkin, M. Brundage, J. Clark, and D. Ganguli, "Understanding the capabilities, limitations, and societal impact of large language models," 2021.
- [7] P. Denny, V. Kumar, and N. Giacaman, "Conversing with copilot: Exploring prompt engineering for solving cs1 problems using natural language," 2022.
- [8] F. Li, X. Xi, Z. Cui, D. Li, and W. Zeng, "Automatic essay scoring method based on multi-scale features," *Applied Sciences*, vol. 13, no. 11, p. 6775, 2023.
- [9] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou *et al.*, "Chain-of-thought prompting elicits reasoning in large language models," *Advances in Neural Information Processing Systems*, vol. 35, pp. 24 824–24 837, 2022.
- [10] S. Yao, D. Yu, J. Zhao, I. Shafraan, T. L. Griffiths, Y. Cao, and K. Narasimhan, "Tree of thoughts: Deliberate problem solving with large language models," *arXiv preprint arXiv:2305.10601*, 2023.
- [11] J. Howard and S. Ruder, "Universal language model fine-tuning for text classification," *arXiv preprint arXiv:1801.06146*, 2018.
- [12] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2019.
- [13] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel *et al.*, "Retrieval-augmented generation for knowledge-intensive nlp tasks," *Advances in Neural Information Processing Systems*, vol. 33, pp. 9459–9474, 2020.
- [14] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *arXiv preprint arXiv:1301.3781*, 2013.
- [15] H. T. *et al.*, "Llama 2: Open foundation and fine-tuned chat models," 2023.
- [16] Z. Liang, W. Yu, T. Rajpurohit, P. Clark, X. Zhang, and A. Kaylan, "Let gpt be a math tutor: Teaching math word problem solvers with customized exercise generation," 2023.