BOARD #129: AI as a Teaching Assistant: Aiding Engineering Students Beyond Office Hours

Mr. Ernest Wang, University of California, Davis

Ernest Wang is a current undergraduate student in Electrical and Biomedical Engineering at the University of California, Davis. He is interested in the application of commercially available LLM models in helping engineering undergraduates with their studies. His other research interests include microfluidics and bioelectronics.

Harry Zhang, University of California, Davis

Harry contributed to this research through the practical development and testing of the AI Teaching Assistant. He played a key role in implementing the AI TA and gathering the data used to evaluate its performance in answering Electrical Engineering student questions, as presented in this paper.

Prof. Paul J. Hurst, University of California, Davis
Dr. Yubei Chen, University of California, Davis

Yubei Chen is an assistant professor from the ECE department at UC Davis. He has worked with Professor Yann LeCun at Meta AI and NYU Center for Data Science as a postdoctoral researcher. Yubei received his MS/PhD in Electrical Engineering and Computer Sciences and MA in Mathematics at UC Berkeley under Professor Bruno Olshausen. His research interests span multiple aspects of representation learning. He explores the intersection of computational neuroscience and deep unsupervised learning, with the goal of improving our understanding of the computational principles governing unsupervised representation learning in both brains and machines, and reshaping our insights into natural signal statistics. He is a recipient of the NSF graduate fellowship, and ICLR Outstanding Paper Honorable Mention Award. You can learn more about him at https://yubeichen.com.

Kenneth Dyer, Microsoft Corporation

AI as a Teaching Assistant:
Aiding Engineering Students Beyond Office Hours

1) Introduction

Large language models (LLMs) are a class of generative artificial intelligence that excel at generating natural language responses to user queries/demands. LLMs have seen an explosion of both interest and applications in recent years. From writing fictional works to synthesizing functional code, LLMs have demonstrated versatility and effectiveness in written language-based tasks [1, 2]. We are currently at a crossroads of sorts, akin to the release and popularization of search engines, where we do not know the limits of this technology, but we believe it has great potential [3, 4, 5]. Students have begun to take advantage of this technology [6], with many turning to popular LLMs if they are stuck on a homework problem or have a class-related question. Also, there has been interest in integrating LLM technology in classes or class projects. LLMs have demonstrated promising performance in code-based tasks. Thus, papers have been published about using LLMs in code-centric classes [6, 7, 8]. Other subjects where LLMs have been frequently used in higher education are social sciences, business and management, and STEM [20]. We are interested in the application of LLMs in an engineering course. In this paper, since the LLM is answering questions like a teaching assistant (TA) does during office hours, we will refer to it as the AI TA. An AI TA could be useful to students who a) have a conflict with normal office hours or b) are uncomfortable asking questions in office hours or c) are doing homework late at night, when the class instructor and TA are not available [18]. The goal is not to replace or eliminate TA's or professor's office hours as the AI TA has limitations and certainly cannot answer important questions unrelated to the class material (e.g. questions like 'What follow-up class should I take next quarter?' or 'Is Company X good to work for?' or 'What do I need to do to pass the class?'). The AI TA, if extremely effective, could perhaps allow a reduction in the number of TAs in a large course, to help alleviate budget shortfalls at a university, or to allow more students to enroll in a course than the assigned TAs could normally handle. Also, the AI TA could be useful to class auditors or working engineers who are trying to learn on their own and do not have access to any office hours.

2) Contributions

This paper explores the ability of some current LLMs to act as an AI TA and answer student questions related to an electrical engineering course: Microelectronic Circuits. Our three major goals are:

- RG1) Develop a framework for evaluating LLM's answers to student questions
- RG2) Evaluate implementations of LLM chatbots using the proposed framework
- RG3) Assess the potential of LLMs in regards to answering engineering questions

3) Methods

3.1 Course Background

We are investigating the ability of LLM chatbots to act like a teaching assistant (TA) for a required Electrical Engineering (EE) course. Analog circuit design is one of the more challenging and complex topics within the EE curriculum. We selected a first course on analog transistor circuits (a junior-level course at our university) for our work. The prerequisites for this course are courses on semiconductor physics and linear circuit analysis; thus most students enrolled are either juniors or seniors.

In this course on circuits, topics covered are semiconductor physics, Bipolar Junction Transistor (BJT) amplifier circuits, Metal-Oxide Semiconductor (MOS) amplifier circuits, frequency response, and feedback.

In the remainder of the paper, we will refer to the LLM chatbot as the AI TA.

3.2 The Evaluation Framework

When asked a question, answers that are accurate, concise, digestible, and course-related are most desirable, as they will be helpful to a student. To evaluate the performance of an AI TA, we need to ask realistic questions and then evaluate the answers.

To evaluate each answer, we focused on three specific aspects that would be important to a student: 1) how accurate is the answer, 2) how understandable (or digestible) is the answer, 3) how well the AI TA points the student to (or references) textbook material that provided all or part of the answer. Addressing point number one just requires grading the answer. A good answer should be correct and use the variables and terminology associated with the course. Evaluating the digestibility requires evaluating the clarity and conciseness of the answer, while looking for clear well-written sentences, a logical flow, and good formatting of the response.

For the final aspect, we looked at the references in the answer, be it equation numbers, figure numbers, example numbers, page numbers or section numbers that the AI TA generated. Each aspect - Accuracy, Digestibility, or References - of each answer is graded and scored from 0 to 5. An overview of the scoring is in Table 1.

	0	3	5	Score
Accuracy	Response is completely inaccurate and/or grossly out of context of course material	Is mostly correct, and/or with some context not relevant to the course material	Response is fully accurate, and in context with course material	/5
Digestibility	Incoherent, requires multiple readings	Understandable with a careful read	Clear, logical flow that guides the user step-by-step	/5
References	No references provided or incorrect sources	Some wrong sources/ some sources missing	Provides correct sources	/5
Total				/15

Table 1: Scoring Guidelines for Grading Answers

3.3 Setting up the AI TA

For this study, we used the latest OpenAI and Anthropic models: ChatGPT 4o [9] and Claude 3.5 Sonnet [10]. These two models were selected based on previous benchmark results for reasoning and mathematics [11, 12]. In some of our experiments, the AI LLMs are used 'as is' or 'off the shelf' - no training and no instructions. In other experiments, the AI LLMs are trained, i.e. instructions are fed into the AI program along with one or more chapters of the course textbook. In evaluating the ability of LLM chatbots to act like a very good TA, we sought to investigate how the AI TA performs for different amounts of instruction / training prior to asking questions.

In our tests, all chatbots are configured with temperature set to 0.3 and a maximum token size of 2048.

3.3.1 Training

A key factor in LLM performance for a given application is how the LLM is instructed to accomplish the desired task [13]. Therefore, it is important to develop a good set of instructions (for training) to assure the AI TA performs well. We implemented prompt engineering best practices [14, 15] to produce a standard set of instructions. The instructions that we used are given in the Appendix.

3.3.2 "Zero-shot" Configuration:

Zero-shot: In this case, the LLM produces answers to questions with no instruction or training from us. This is an "As Is" or "Off the Shelf" use of an AI chatbot. It is what a student would experience if going to an AI website and asking the chatbot questions. We are interested in zero-shot performance to understand the ability of existing, untrained LLM chatbots to act as an AI TA, and this case can act as a base-line for comparison with the following cases that incorporate training of the AI TA.

In the zero-shot case, we simply submitted the questions through the LLM's Application Programming Interface (API).

3.3.3 "Few-shot" Configuration:

Few-shot: In this case, the AI TA is trained prior to being asked questions. For the training, we feed instructions (which are in the Appendix) to the AI TA along with the one related chapter of the textbook, in the form of a PDF file. After feeding a chapter into the AI TA, questions related to that chapter are asked of the AI TA as in the zero-shot case. (The chapters used in this project are described in Section 3.4.)

3.3.4 "RAG" Configuration:

RAG or *Retrieval-Augmented Generation*: In this case, the AI TA is provided the same instructions (which are in the Appendix) as in the few-shot case. However, here all the chapters of the textbook are first vectorized (using Abacus [16]), and then fed into the chatbot. This vectorization is beneficial as it makes the textbook easier for the LLM to process [19]. The RAG configuration could lead to a very good AI TA, because the AI TA has access to all the textbook material in a desirable form.

3.4 Choosing Questions for the AI TA

To test the LLM chatbot (the AI TA), we decided to focus on the material in two chapters in a popular transistor-circuit textbook: Chapter 2 on semiconductor physics and Chapter 7 on single-stage MOS amplifiers. These two chapters were chosen because they are very different. The chapter on semiconductor physics contains many facts and equations related to doping, semiconductors, electron/hole movement, and diodes. Questions asked to the AI TA about this chapter allow evaluation of an LLM's ability to extract information or draw correct conclusions from facts and equations primarily, or from the relatively simple figures in this chapter.

The chapter on MOS amplifier stages contains more complex and challenging material. This material is typically brand new for students, involves many approximations, and typically generates many office-hour questions. Learning this material requires learning how transistors work in four basic amplifier stages, how the transistors are biased, how to compute gain and

input/output resistance, how and when to use approximations, and how to work with and create small-signal models for MOS transistors. Also, the figures are mainly circuit schematics, which might be difficult for the AI TA to interpret. Questions related to this chapter will help us evaluate an LLM's ability to deal with a complex engineering topic. The AI TA must handle complicated equations and circuit analysis concepts.

3.4.1 Questions

Questions used for testing the AI TA are related to the two chapters mentioned above, and they should help us evaluate and compare the performance of the two LLMs in the three cases (zero shot, few shot, RAG). To this end, we generated over two dozen questions related to the two chapters. The questions fall into several categories, which are as follows.

1) Straightforward Questions

   Answers to these questions can be found entirely within the text in the book, and they could be answered by any student with a good memory after reading the book. The goal of these questions is to verify the ability of LLM models to answer easy and straightforward questions. Example questions are: "What is the doping element used to produce n-type silicon?" and "Why does the common source stage have more gain than a similar common source stage with degeneration?"

2) Extrapolation

   These questions require extrapolation of information in the book to generate an answer. The goal is to evaluate the ability of LLM models to synthesize answers from a collection of relevant information. For example, we could ask the LLM: "Why is the current through a diode dependent on the cross-sectional area?" or "Under what conditions is the current through a diode exactly zero?"

3) Ambiguous or confusing questions

   A desirable feature of any TA, including our AI TA, is to help a student with a question, even when the question is vague, poorly worded, or even incorrect. Such questions test the ability of the AI TA to deal with students who do not understand the material well enough to ask a good question. An example is: "Would you please explain to me how the cascade gain stage works?" (the correct phrase is *Cascode* Gain Stage, not Cascade Gain Stage).

4) Material not textbook and not part of the class

   Questions beyond the course content are asked during office hours and we want to see how the AI TA handles such questions. An example of such a question: "Can you explain

to me how a vacuum tube works? I know my guitar amp uses tubes." This is a reasonable question that many professors and some TAs could answer. The AI TA should realize the question falls outside the scope of the class, and ideally answer the question but add a disclaimer that the answer is not from the textbook as the question falls outside the course content. (See the instructions in the Appendix.)

3.4.2 Challenging Questions:

In contrast to the material on semiconductor physics, questions on the MOS amplifier circuits may have multiple possible correct answers - the answer from the AI TA would be generated by considering all possible correct answers and choosing the best one. Or by giving some or all possible answers. Thus, this chapter presents a significant challenge to the AI TA and will push the LLM chatbot to its limit. An example question is "How could I build a simple circuit with a voltage gain with magnitude 10?" The answer could refer to a common source, a common gate or a common source with degeneration stage. An extremely thorough answer might include all three stages.

3.4.3 Homework Questions:

An important part of an engineering student's coursework is doing homework problems to learn the material and gain hands-on experience. Questions about homework problems, both before they are due and after they are due but before an exam, are very common in office hours. With this in mind, we selected various homework problems from the textbook to test the AI TA.

3.5 Grading Answers from AI TA

To grade the answers generated by each version of the AI TA to our 31 test questions, two of the authors (a student who took the course recently and a professor who teaches circuit design) graded all the answers generated by the different implementations of the AI TA, using the scoring guidelines in Table 1. Scores range from 0 (terrible) to 5 (excellent). In addition, we recorded comments on answers, when it seemed appropriate. A third author submitted the questions to the different versions of the AI TA and stored each answer in a file and made the answers available to the graders. The file names were randomly selected numbers so we did not know what version of the AI TA generated each answer.

4) Results

4.1 Overview

The results for the two chapters are presented separately below, and then they are combined at the end of this section.

4.2 Straightforward Questions

4.2.1 Quantitative Data

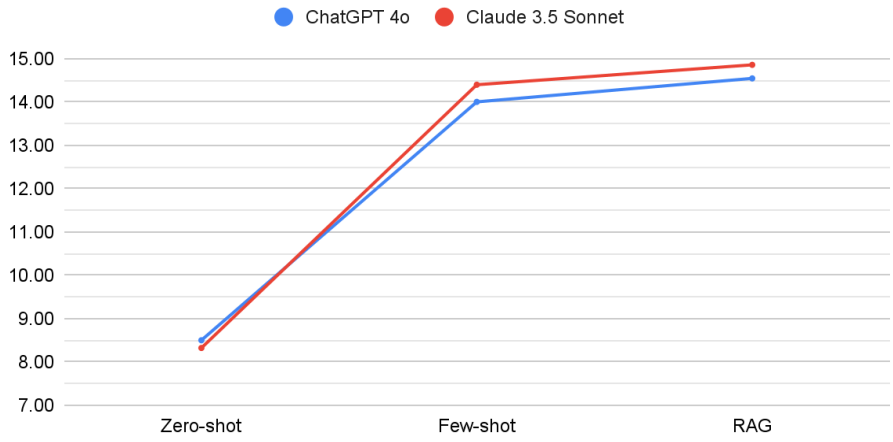


Figure 1: Total Score vs. Implementation - Chapter 2

	Zero-shot		**Few-shot**		**RAG**	
	ChatGPT 4o	Claude 3.5	ChatGPT 4o	Claude 3.5	ChatGPT 4o	Claude 3.5
Accuracy	4.21	3.99	4.77	4.90	4.91	5.00
Digestibility	4.29	4.33	4.63	5.00	4.91	5.00
Reference	0	0	4.60	4.50	4.70	4.85
Sum = Total Score	**8.50**	**8.32**	**14.00**	**14.40**	**14.54**	**14.85**

Table 2: Average Scores for Answers for Chapter 2 - Semiconductor Physics

For the Chapter 2 questions, the average scores for the different implementations of AI TA are given in Table 2 and also plotted in Figure 1. Note that since the zero-shot case had no training/instruction, it could not give information as to where in the book it found useful information, so the reference scores are all 0 for this case. As seen in Figure 1, there is an improvement in performance moving to the right and the RAG Claude 3.5 version has the best score. More specifically, the accuracy score improved by 16% between zero shot and RAG implementations of ChatGPT 4o and improved by 25% between zero shot and RAG implementations of Claude 3.5. The digestibility score improved by 14% and 15% between zero shot and RAG implementations of ChatGPT 4o and Claude 3.5, respectively. The reference score jumped from 0 to 4.6 and 4.5 for ChatGPT 4o and Claude 3.5, respectively, and it improved 2% and 7% between few shot and RAG implementations of ChatGPT 4o and Claude 3.5,

respectively. Looking at the overall performance trends, there is a marked improvement as the LLM is fed more course information.

4.2.2 Qualitative Data

Along with each numerical score, the grader can leave comment(s) on the response as a supplement to the score. In the comments, the grader can either critique or complement the answer. Across the evaluations, several key themes emerge from the comments. Some negative comments mention lack of clarity and correctness of equations in the answers, with specific issues being messy or inconsistent formatting and terminology (e.g., “volume” instead of “area”). Graders frequently noted responses that failed to handle a poorly phrased question or questions that contained incorrect phrasing or assumptions. Positive comments were made about equations, wording of explanations, and alignment with the textbook material. However, discrepancies were also noted in the references, as sometimes sections were mislabeled or misnumbered. Some suggestions for improvement included more clear equations (e.g., using dots or "×" for multiplication), clearer phrasing in answers, and correcting inaccuracies in the student’s question when this occurs.

4.3 Challenging Questions

4.3.1 Quantitative Data

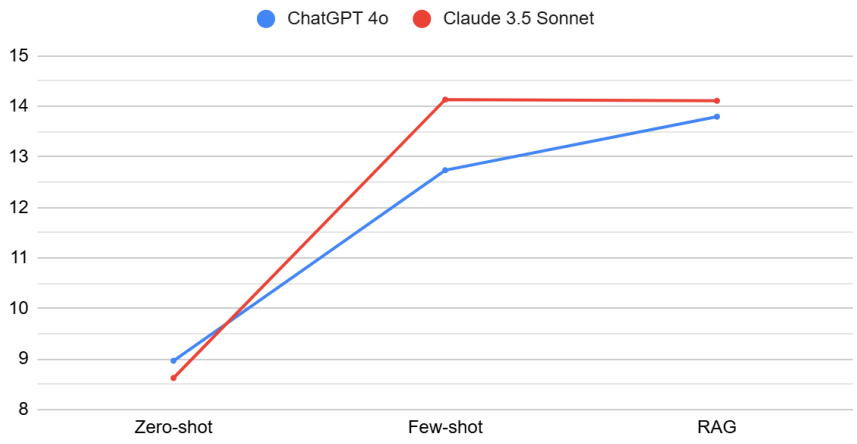


Figure 2: Total Score vs. Implementation - Chapter 7

For the Chapter 7 questions, the average scores for the different AI TA implementations are in Table 3 and also plotted in Figure 2. Note again that the zero-shot case had no training/instruction, so the reference scores are all 0 for this case. As seen in Figure 2, there is an improvement in performance moving to the right and the RAG Claude 3.5 version has the best score.

	Zero-shot		Few-shot		RAG	
	ChatGPT 4o	Claude 3.5	ChatGPT 4o	Claude 3.5	ChatGPT 4o	Claude 3.5
Accuracy	4.44	4.16	4.05	4.62	4.64	4.72
Digestibility	4.51	4.46	4.18	4.54	4.59	4.67
Reference	0	0	4.50	4.97	4.55	4.51
Sum = Total score	8.96	8.62	12.73	14.13	13.79	14.11

Table 3: Average Scores for Answers for Chapter 7 - MOS Amplifier Stages

There is a general trend of improved or steady performance moving to the right, but by an insignificant margin in the few-shot Claude implementation is the best. From a practical standpoint, few-shot and RAG Claude implementations performed the same, and Claude outperforms ChatGPT.

As shown in Table 3, the accuracy score improved by 4% between zero shot and RAG implementations of ChatGPT 4o, and it improved by 13% between zero shot and RAG implementations of Claude 3.5. The digestibility score improved by 1% and 4% between zero shot and RAG implementations of ChatGPT 4o and Claude 3.5, respectively. The reference scores were 4.50 and 4.97 for ChatGPT 4o and Claude 3.5, respectively, for the few-shot case. Interestingly, RAG outperformed ChatGPT 4o in Accuracy and Digestibility, but RAG's reference score was either virtually unchanged or slightly worse than few-shot's scores.

4.3.2 Qualitative Data

Several key issues showed up in the graders' comments. The most common comment is that the AI TA can't "see" the figures mentioned in a question or homework problem. There are multiple versions of 'not seeing': 1) the AI TA states that it is unable to view or analyze the figure, 2) the AI TA sees wrong component values or names in a figure or a schematic in a figure or 3) the AI TA answers the questions for a different circuit than the one in the figure. Another issue is the presence of minor errors in the answer. Examples of minor errors are a) labeling an AC bypass capacitor as a coupling capacitor, b) the equation is correct but the associated text is not completely clear, c) the answer should include a restatement of the question because the student's question contains a false statement or phrase. These small errors don't completely invalidate the answer, but they are undesirable. Graders also noted that some references are not ideal. For example, referencing Section 2.4 when Section 2.4.2 would be better and more helpful. Also, minor problems with equation formatting and terminology were mentioned multiple times.

4.4 Homework Questions

Questions about homework problems are something that the AI TA will be asked but may be tricky to deal with. With new added instructions: "help the student but do not solve the problem or do any calculations", we submitted questions to the two RAG implementations. The questions all read like this: "Please help me to do textbook Problem 2.3".

For Chapter 2 questions, which ask the student to use formulas from the book to compute answers, the answers/results were excellent for Claude (average accuracy = 5.0, digestibility = 4.6, and reference = 5.0) and good from ChatGPT (average accuracy = 3.8, digestibility = 3.6, and reference = 5.0). Unfortunately, for Chapter 7 questions, which typically ask the student to do analysis or design calculations for a specific circuit in a figure (which is a circuit schematic in all cases), the answers/results were not good. Both Claude (average accuracy = 1.6, digestibility = 2.2, and reference = 3.2) and ChatGPT (average accuracy = 2.8, digestibility 3.4 and reference = 4.0) were not consistently helpful. The major problems were a) in some cases, the AI TA used the wrong figure so the AI TA was no help, b) in some cases, it didn't properly digest the figure [for example, 1) it thought there was a resistor loading the output, but there was not, and 2) it thought a transistor was connected as a current source but it was not] so the AI TA was not helpful, or c) in a few cases, the AI TA solved the problem completely, plugging in numbers, going against specific instructions not to do that.

These preliminary results show that the AI TA needs improvement to handle the important task of helping students with homework problems.

5) Discussion & Summary

Looking at the scores in Tables 2 and 3, there is a trend of increasing performance as the AI TA receives more training and information. For basic questions related to the semiconductor physics material, the AI TA performed close to perfect (a few minor issues) when trained as a RAG system. For the more difficult questions related to the MOS amplifier stages, the AI TA demonstrated some improvement in accuracy and digestibility. Overall, the AI TA performed comparatively worse in questions related to MOS amplifier stages.

Figure 3 plots the average of the Total Scores in Tables 2 and 3 vs. implementation. Looking at this plot, the Claude 3.5 Sonnet RAG implementation is slightly better than the few-shot Claude case. Based on our data, Claude 3.5 Sonnet RAG is recommended for creating an AI TA.

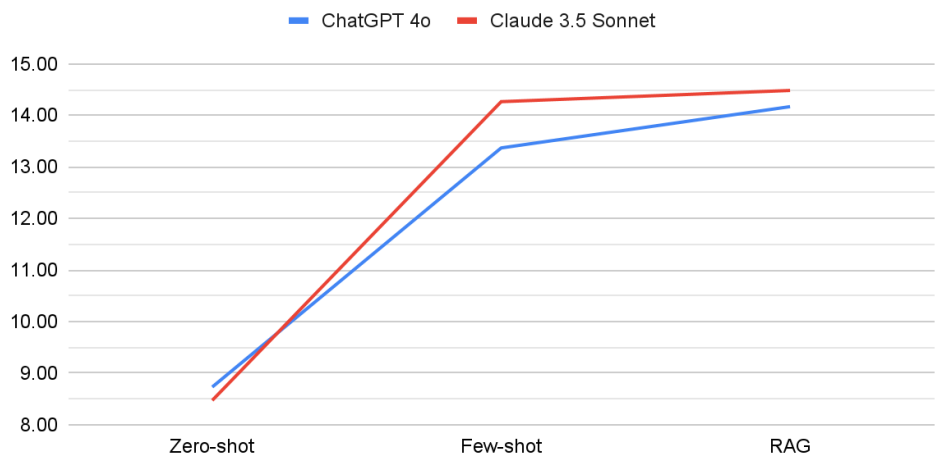


Figure 3: Average Total Score vs. Implementation - Average of Total Scores from Both Chapters

5.1 Accuracy

Overall, we see an average improvement of 14.5% in accuracy between our zero-shot implementation and RAG implementation, with the minimum improvement being 4%. This trend demonstrates potential for increasing accuracy further by increasing the information provided to the LLM model. For straightforward questions, we found an average 20.5% improvement in accuracy between the zero-shot and RAG implementations. For challenging questions, we found an 8.5% increase in accuracy.

The general trend of improving performance as more information is provided is promising, as more material can be added to the RAG to increase the AI TA's knowledge base. From qualitative data on MOS circuit questions, it is clear that there are some small errors that arise when questions are long or require long, complicated answers that include many equations.

5.2 Digestibility

Digestibility follows a similar trend to accuracy with an average 8.5% improvement between zero-shot and RAG implementations. For straightforward questions, we found an average 14.5% improvement in the digestibility score, and for nuanced questions, there was an average 2.5% improvement in the digestibility score. We attribute the improvement in digestibility primarily to the instructions given to the AI TA in the few-shot and RAG cases. In the zero-shot experiment, no instructions were given to the LLM. In the RAG and few-shot experiments, the instructions ask the LLM model to produce an answer in a digestible form. Improved instructions could further improve the digestibility score.

5.3 References

In the zero-shot cases, the AI TA cannot give references (i.e., pages or sections or equations or figures that were the basis of the answer) in the book as it was not fed the textbook; thus no references were present in the answers. This is the reason the reference scores for all zero-shot cases are zero. In the other two cases, all or part of the textbook was provided during training of the AI TA, and it was told that the sources of each answer should be included somewhere in the answer. Thus the focus here is a comparison of the reference scores from the few-shot and RAG experiments, and there is a small improvement in these scores. We found that the RAG version of AI TA was able to better cite the sources of answers.

For many students and some professors, having an AI TA that only points the student to the locations in the book to find answers may be good enough or desirable. One advantage of this is that it forces the student to a) own or have access to the book and b) read at least parts of the book.

5.4 Limitations and Advantages

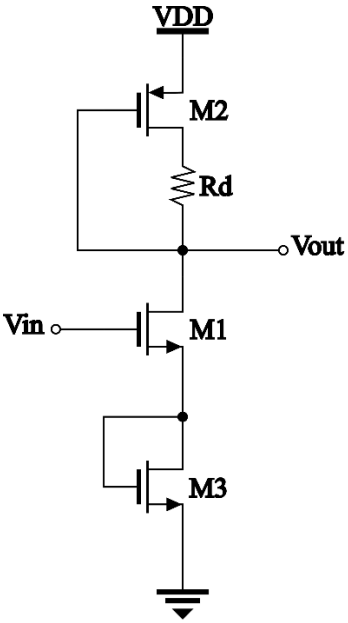
One limitation of this study is the reliance on two graders as a way to evaluate the different versions of the AI TA. While we are confident that we were careful, accurate, and unbiased, perhaps more graders or even automated grading would be desirable. Another concern may be the limited topics and the limited number of questions that were used to test the AI TA implementations. We tried to pick some easy and some hard questions, some easy and some difficult material, and different types of questions, as described earlier in Section 3.4.

A major issue that we verified after recording the data in Section 4.4 is that the AI TA does not seem to accurately digest the transistor-circuit schematics in the book. For example, when fed a PDF file containing a schematic drawing of a common-source amplifier (transistor M1) with cascoded load (transistors M3A and M3B), AI TA answered all question about that circuit incorrectly. It answered as if the circuit was a cascode transistor M2 above common-source M1 (apparently it read the figure caption and then found information about a different but similarly named circuit), and its answers repeatedly mentioned M2 (which is not in the figure) and referred to the wrong circuit topology. Since schematics are such an important part of the course (probably well over half the figures in the book are circuit schematics), this is a huge problem. We tried a fix: we wrote a SPICE netlist for this amplifier with cascoded load. (SPICE [17] is a widely used program for circuit simulation. A SPICE netlist is a complete description of a circuit, using only text.) After feeding the SPICE netlist into a newly released LLM, it understood the circuit and its elements, and it correctly answered questions about the circuit.

After further testing, we believe that providing the AI TA with a SPICE netlist for each schematic is a good and effective option. However, generating a SPICE netlist for every schematic in a circuits textbook would be a big job. Fortunately, recent work [21] has

demonstrated that software can automatically convert a circuit schematic into a SPICE netlist. Also, hopefully future, improved LLMs will be able to correctly interpret a circuit schematic.

An example circuit schematic and its SPICE netlist are shown in Figure 4.




```
VDD vdd 0 dc 1.8
Vin in 0 dc vdc ac 1
RD rm vout rd
M1 vout in s1 0 nch
M2 rm vout vdd vdd pch
M3 s1 s1 0 0 nch
.model nch nmos (kp=200u vto=0.4 lambda=0)
.model pch pmos (kp=100u vto=-0.4 lambda=0)
```

a) b)

Figure 4. a) Schematic of a three-transistor amplifier. b) SPICE netlist for the circuit in (a).

Another concern is that the AI TA has not been able to produce a good or useful circuit schematic, when asked about an amplifier stage. That would be helpful when trying to explain how a circuit works to a student. Drawing a schematic is something that a professor or TA can do easily and would do. Pointing to a schematic, its nodes and circuit elements while talking is the best way to explain the operation of a circuit.

Some advantages of the AI TA that were not mentioned in the opening paragraphs are the following. First, if a student is uncomfortable asking a human (professor or TA) a question (perhaps thinking the question is embarrassing or that the professor is not approachable, or for whatever reason), asking the AI TA should not be uncomfortable or intimidating. Also, the AI TA, which is running computer software, should not discriminate in any way; it is an equal-opportunity helper.

Another advantage of the AI TA is that it can provide its answers in many languages, if requested to do so.

Finally, an AI TA could be helpful to students who didn't understand material from previous classes. For example, asking a professor in a senior engineering class to explain differentiation or complex numbers or statistics – topics that were covered in lower-level classes and are assumed to be understood – might be something a student would not do as it might give the professor a bad impression. However, a student could easily ask the AI TA anything and hopefully get good responses and learn material that was taught in previous courses.

5.5 Comments by Faculty and Students

A limited survey of faculty about the AI TA produced many comments, most notably: 'this sounds interesting', 'I would like to try it', 'aren't you enabling lazy students', 'this might reduce office hour visits which provide useful feedback', 'this might encourage students to not read the textbook', 'sounds interesting', 'this work is a very interesting endeavor', and 'referring students to the book for answers is a good option', 'I hope this doesn't lead to reduced TA support'.

We also performed a small survey of engineering students and found that an overwhelming number of students are using LLMs, mostly ChatGPT 4o. In our informal survey, we found three different categories of LLM usage: 1) Students asked an LLM for help with a homework problem, only looking for a hint so they could start solving the problem on their own. 2) Students used an LLM to solve a problem that they had already solved themselves, to verify their solution. 3) Students (lazy or short-on-time or dishonest?) asked an LLM to solve a homework problem, then copied the solution and turned that in for grading.

6) Future Work

Our results demonstrate great potential for an AI TA. However, additional work may lead to a better and more useful tool for students. One possibility is to fine-tune the AI TA by providing student feedback to answers. Another opportunity for improvement is to find a way for the LLM to see and understand textbook figures and schematics. Another area for future work: presumably the AI TA performance would improve when fed more information, which for example could include lecture slides, other textbooks, homework solutions, exam solutions and more examples. Also, we would like to investigate ways to improve the AI TA's ability to process and understand figures, including circuit schematics.

We are also interested in having students use this tool in an offering of our Microelectronic Circuit course. We are considering, as part of our future work, asking students to use one or more versions of the AI TA in a future offering of the Microelectronic Circuits course and then survey the students to collect feedback and comments.

Another potential use of an AI TA would be to help a student or TA in lab. Sometimes a circuit is built in lab but doesn't work, and inevitably the TA is called on to help debug the nonfunctional circuit. Or sometimes the lab equipment is confusing or malfunctioning. If an AI

TA could be trained for each lab and be able to help debug common problems, that would be useful. Certainly testing the AI TAs with more questions is desirable.

In addition to developing a good AI TA, there are many other issues. A few examples: How will each university, college or department decide to employ or restrict AI software? How to deal with students using AI to do their homework? What will book companies want in return for their books being fed into LLMs?

7) Conclusion

We created an evaluation framework and used that framework to assess three different implementations of two state-of-art LLMs that act as an AI TA for a course on transistor circuits. By grading answers to questions posed to the various AI TAs, we measured the performance of the different versions. We focused on material covering semiconductor physics and MOS amplifier stages.

We found that LLMs scale well, improving when more training and information is provided. Our results demonstrate great potential for an AI TA, and the RAG version does a very good job. As LLMs improve, the AI TA should also improve.

Appendix:

Instructions for training the few Shot and RAG Implementations prior to questions being asked are as follows:

You are a teaching assistant for a class in microelectronic circuit design, using the textbook content that is surrounded by triple quotes. Using information strictly from the textbook, you must provide concise and simple-to-understand answers to student questions. For any question that the book does not cover, you can answer the question to the best of your ability, but you MUST add a disclaimer that the information is not found in the textbook. You are not allowed to perform any numerical calculations, all mathematical answers must be done using the standard variable notations found in the textbook.

The topics you must know are as follows:

- Diode Models and Circuits: This section serves three purposes: (1) make you comfortable with the PN junction as a nonlinear device, (2) cover basic diode circuits e.g., rectifiers and limiters and (3) introduce the concept of linearizing a nonlinear model to simplify the analysis.
- Bipolar Circuits: Following a bottom-up approach, this section establishes critical concepts such as input and output impedances, biasing, and small-signal analysis followed by bipolar amplifier topologies.
- Physics of MOS Transistors: A short introduction to the MOSFET as a voltage-controlled current source and as a variable resistor and deriving its characteristics.
- MOS Amplifiers: Drawing extensively upon bipolar amplifiers, this section deals with MOS amplifiers but at a faster pace.
- Frequency Response: basic concepts such as Bode's rules and association of poles with nodes, this section introduces the frequency response of amplifiers and uses Miller's Theorem to analyze the frequency response of the basic common-emitter/common-source topology.
- Feedback: This section covers (1) the general structure of negative-feedback amplifiers and advantages of negative feedback, (2) different feedback topologies, (3) an intuitive and insightful approach for the analysis of practical feedback voltage amplifier circuit, and (4) why and how negative feedback amplifiers can be unstable.

Follow the following format for problem-solving questions:

1) Summary: summarize the problem and give a brief insight on how to proceed
2) Steps: Explain the solution step by step and explain the insight/logic behind each step. You are not allowed to perform mathematical computations, all steps must be in terms of standard variables found in the textbook
3) References: Give an overview of the solution and all the specific locations in the textbook where you got the references used to come to your solution path

Follow the following format for conceptual questions:

1) Summary: summarize the question and give a brief insight on how to proceed
2) Break down: Conceptually break down the problem into smaller topics and necessary prior knowledge
3) Synthesis: Build a comprehensive explanation using the previous section
4) References: Give an overview of the explanation and all the specific locations in the textbook where you got the references used to answer the question(s) asked

For any question that information cannot be found in the textbook, use the information you are trained on, add a disclaimer that the information is not in the textbook, and provide insight on what topics the student should research to gain their own understanding.

References:

[1] S. Bubeck, *et al.*, "Sparks of Artificial General Intelligence: Early experiments with GPT-4," *arXiv preprint arXiv:2303.12712*, Mar. 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2303.12712

[2] D. Reeping and A. Shah, "Board 50: Work in Progress: A Systematic Review of Embedding Large Language Models in Engineering and Computing Education," *ASEE Conf.*, Portland, June 2024. Available: https://peer.asee.org/47047

[3] L. J. Wiese and A. J. Magana, "A Department's Syllabi Review for LLM Considerations Prior to University-standard Guidance," *ASEE Conf.*, Portland, June 2024. Available: https://peer.asee.org/46436

[4] Y. LeCun, "A Path Towards Autonomous Machine Intelligence Version 0.9.2," OpenReview, 2022. Available: https://openreview.net/forum?id=BZ5a1r-kVsf

[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436-444, May 2015, doi: 10.1038/nature14539

[6] K. Baltaci, M. Herrmann, and A. Turkmen, "Integrating Artificial Intelligence into Electrical Engineering Education: A Paradigm Shift in Teaching and Learning," *ASEE Conf.*, Portland, Jun. 2024. [Online]. Available: https://peer.asee.org/47644

[7] Y. Hicke, A. Agarwal, Q. Ma, and P. Denny, "AI-TA: Towards an Intelligent Question-Answer Teaching Assistant using Open-Source LLMs," *arXiv preprint arXiv:2311.02775*, Nov. 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2311.02775

[8] Y. Ai, M. Baveja, A. Girdhar, M. O'Dell, and A. Deorio, "A Custom Generative AI Chatbot as a Course Resource," Aug. 2024, doi: https://doi.org/10.18260/1-2--46433.

[9] OpenAI, "Hello GPT-4o," *Openai.com*, 2024. https://openai.com/index/hello-gpt-4o/, ver: gpt-4o-2024-08-06

[10] "Introducing Claude 3.5 Sonnet," *www.anthropic.com*. https://www.anthropic.com/news/claude-3-5-sonnet, ver: claude-3-5-sonnet-20241022

[11] openai, "GitHub - openai/simple-evals," *GitHub*, 2024. https://github.com/openai/simple-evals?tab=readme-ov-file#benchmark-results

[12] W. L. Chiang *et al.*, "Chatbot Arena: An Open Platform for Evaluating LLMs by Human Preference," *arXiv preprint arXiv:2403.04132*, Mar. 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2403.04132 (note: for LLM comparisons https://huggingface.co/spaces/lmarena-ai/chatbot-arena-leaderboard)

[13] B. Chen, Z. Zhang, N. Langrené, and S. Zhu, “Unleashing the potential of prompt engineering in Large Language Models: a comprehensive review,” *arXiv.org*, Oct. 2023. https://arxiv.org/abs/2310.14735

[14] “Prompt Engineering Guide – Nextra,” *www.promptingguide.ai*. https://www.promptingguide.ai/

[15] “OpenAI Platform,” *Openai.com*, 2024. https://platform.openai.com/docs/guides/prompt-engineering

[16] “Abacus.AI - The world’s first AI assisted end-to-end data science and MLOps platform,” *Abacus.AI*, 2024. https://abacus.ai

[17] A. Vladimirescu, *The SPICE Book*, Wiley, 1994

[18] S. Abdul-Wahab, N. Salem, and S. Fadlallah, “Students’ reluctance to attend office hours: Reasons and suggested solutions,” *Journal of Educational and Psychological Studies*, vol. 13, p. 715, Oct. 2019.

[19] P. Lewis, *et al.*, "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," *Conference on Neural Information Processing Systems* (NeurIPS), Vancouver, 2020. https://doi.org/10.48550/arXiv.2005.11401

[20] R. Hadi Mogavi, et al., "ChatGPT in education: A blessing or a curse? A qualitative study exploring early adopters' utilization and perceptions," *Computers in Human Behavior: Artificial Humans*, vol. 2, no. 1, 2024.

[21] J. Bhandari, et al., Auto-SPICE: Leveraging LLMs for Dataset Creation via Automated SPICE Netlist Extraction from Analog Circuit Diagrams, *paper under review*, https://arxiv.org/pdf/2411.14299v1