



Incorporation of ChatGPT and Other Large Language Models into a Graduate Level Computational Bioengineering Course

Michael R. King¹ · Adam M. Abdulrahman^{1,2} · Mark I. Petrovic^{1,2} · Patricia L. Poley¹ · Sarah P. Hall¹ · Surat Kulapatana^{1,3} · Zachary E. Lamantia¹

Published online: 7 February 2024

© The Author(s), under exclusive licence to Biomedical Engineering Society 2024

Abstract

The remarkable capabilities of generative artificial intelligence and large language models (LLMs) such as ChatGPT have delighted users around the world. Educators have regarded these tools as either a cause for great concern, an opportunity to educate students on cutting-edge technology, or often some combination of the two. Throughout the Fall 2023 semester, we explored the use of ChatGPT (and Bard, among other LLMs) in a graduate level numerical and statistical methods course for PhD-level bioengineers. In this article we share examples of this ChatGPT content, our observations on what worked best in our course, and speculate on how bioengineering students may be best served by this technology in the future.

Keywords Large language model · ChatGPT · Bard · Higher education · MATLAB

Introduction

In little over 1 year, ChatGPT and related tools have taken the world by storm, with their ability to rapidly generate text, computer programs and/or images on a wide range of subject matter, all in response to plain language text prompts entered by the user. ChatGPT has been shown to have the ability to generate entire articles on a specific topic, such as the role of AI in medicine [1], raising questions related to plagiarism, AI-assisted student cheating [2], and the very nature of crediting authorship for AI-generated text [3]. One of the more useful aspects of ChatGPT is its ability to generate functional computer programs in languages such as MATLAB or Python to achieve a specified task [4], suggesting that it could enhance the education and practice of engineering if employed properly.

One of us (M.R.K.) has taught different versions of a numerical and statistical methods course to biomedical

engineering students for over 20 years, using a textbook written by King and Mody [5] that implements various techniques in the MATLAB programming language. To update the most recent offering of the BME 7410: Quantitative Methods in Biomedical Engineering course at Vanderbilt University in Fall 2023 and elevate its contemporary relevance and interest, we incorporated the use of ChatGPT and similar LLMs in the 7 homework assignments of the course. Enrollment in Fall 2023 was 21 graduate students (mostly PhD students), with one student dropping the course prior to the end of the semester. Overall, this new content was well received by the students, and students completed 100% of the LLM homework components. In this article, we provide some details on the different ways in which LLMs were incorporated and share observations on what worked best and what didn't. We conclude with some predictions on new ways in which LLMs may find utility in biomedical engineering education and practice as these tools continue to rapidly advance in their capabilities (Fig. 1).

✉ Michael R. King
mike.king@vanderbilt.edu; mike.king@rice.edu

¹ Department of Biomedical Engineering, Vanderbilt University, Nashville, TN 37235, USA

² Medical Scientist Training Program, Vanderbilt University School of Medicine, Nashville, TN, USA

³ Department of Physiology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand

Methods

The BME 7410: Quantitative Methods in Biomedical Engineering course includes 7 homework assignments, assigned roughly biweekly on a different theme. Each

Fig. 1 Four realizations of “a bioengineering graduate student comparing the performance of ChatGPT, Google Bard, and Microsoft Bing chatbot in their ability to teach and perform computational methods,” as envisioned by the DALL-E 3 AI image generator via the Microsoft Bing chatbot app for iPhone



assignment involves 1–3 problems taken directly from the King and Mody textbook [5]. The homework themes in Fall 2023 were as follows:

- Homework #1: Linear Regression
- Homework #2: Probability, Error Propagation, and Linear Regression Error
- Homework #3: Nonlinear Root Finding
- Homework #4: Nonlinear Data Regression
- Homework #5: Numerical Quadrature
- Homework #6: ODE Integration
- Homework #7: Statistical Tests

The following guidelines were provided with the first homework assignment of the semester:

ChatGPT ground rules for my class:

Always clearly disclose which parts are LLM generated. Never pass off LLM output as your own work, or leave it ambiguous to the reader (me).

I ask you to critique ChatGPT answers, rigorously (with math, and applying course concepts) whenever possible. On each homework assignment, you solve one of the problems yourself, and then ask ChatGPT

to critique your work (then comment on it) and then for the other problem, switch roles. Ask ChatGPT to solve and you critique its output. Sometimes I will tell you which mode each problem is, sometimes it will be your choice.

•
•
•

You are encouraged to use ChatGPT, Microsoft Bing chatbot, and/or Google Bard chatbot. There are other LLMs out there, some are built on older versions (GPT-3.0 for instance) and less capable.

Have fun with it! My hope is that we learn more, not less, than my pre-GPT classes. I am certainly excited about the inclusion of AI in our course, and I hope you are too.

- MK

Results

In this section, we summarize the student responses to the specific LLM task for each of the 7 assignments.

HW#1:**Large Language Model component of Homework #1 (worth 4 out of 14 points):**

For the first problem, first ask ChatGPT or other LLM to solve it, and then you critique the LLM solution and discuss whether you agree with it (or not), based on the concepts we have discussed in class and read about in the textbook.

For the second problem, first solve the problem yourself, and then ask ChatGPT or other LLM to critique your solution and offer any corrections or improvements that it can think of... and then discuss the similarities and differences of the two approaches.

One immediate observation while grading these assignments is that different users can obtain different results, even when using the same LLM. This could be due to subtle differences in prompt wording, occasional user error, inherent randomness in ChatGPT, or the near daily updates implemented by OpenAI. Usually, ChatGPT is willing to perform mathematical calculations but occasionally claims that it cannot. When prompted to write computer code or perform multistep calculations of any kind, ChatGPT will default to the Python language. This is less relevant to our course which is based in MATLAB, however, and when prompted to program in MATLAB ChatGPT will happily comply despite lacking the capacity to run or test MATLAB code in the user environment. Sometimes its MATLAB code contains errors or attempts to call undefined functions—errors easily identifiable when tested in a separate instantiation of MATLAB. On this first homework assignment, BME 7410 students greatly favored the default (free) GPT-3.5 version of ChatGPT, as opposed to the premium GPT-4, Microsoft Bing chatbot (powered by GPT-4), or Google Bard (powered by the LaMDA algorithm, distinct from GPT). ChatGPT clearly excels in producing non-specific advice about how to solve basic forms of numerical or statistical calculations. One student submitted an interesting interactive session attempting to “force” GPT to do actual mathematical operations when the user suspected that it was not, and just guessing answers instead (and apparently, it was). Another enterprising student asked GPT to create new problems that would illustrate a contradiction when comparing 2 mathematical models via multiple error metrics!

HW#2:

ChatGPT/LLM portion of homework: This homework is longer than most, with 3 problems instead of 2, so just ask ChatGPT (or other LLM) to solve/resolve one of the three problems, then report and interpret these results. Any of the three problems, your choice!

One student provided a nice ChatGPT solution of a simple confidence interval calculation; ChatGPT failed to use the more appropriate t-distribution for a small ($n=4$) sample

size and the student recognized this. Interestingly, the opposite scenario was also submitted, where GPT correctly used a t-distribution, but the student felt that the z-distribution should have been used. The amusement of students with these tools began to show in interactive back-and-forth sessions, with one student setting the stage with prompts such as “I am currently in the woods and don’t have a calculator” followed by “okay I used an abacus.”

HW#3:

Large language model portion: For this homework, ask ChatGPT or other LLM to create a well-defined scalar root-finding problem for you to solve, and then solve it using one of the three methods we have discussed in class. For fun, you may ask GPT to try and make the problem relevant to biomedical engineering (which is more challenging, in my opinion). Briefly discuss whether you think this is a well-designed problem.

One student asked a less common LLM, Claude 2, to create a nonlinear root-finding problem, but decided that the problem was too easy since it could be solved algebraically without the use of a nonlinear algorithm. Interestingly, a second student obtained a very similar problem from ChatGPT. Another student had ChatGPT create a simple root-finding problem, outline a solution approach, and even calculate the answer (in Python). One student asked ChatGPT to create a biomedical problem involving the sum of two exponentials (a classic nonlinear example discussed in class that resists simple logarithmic linearization), however the student noted that GPT’s suggested initial guess was not a good one. Another student obtained a wealth of extra drug dosing information from ChatGPT; as an instructor it is gratifying to see students demonstrate curiosity about the background subject matter. A student asked ChatGPT to solve a newly generated abstract cubic problem using two different methods, which it promptly did. Interestingly, a student found that GPT seemingly won’t use its best initial guess, unless specifically prompted to do so. Finally, one student used Microsoft Bing chatbot to create a biomedical root-finding problem and it generated a good one, providing all of the necessary parameter values as well. The student also determined that the problem had a well-defined solution. Success!

HW#4:

Large language model portion: Okay this one will be fun... solve both of the above problems yourself, and then, for whichever solution you feel most confident about, purposely insert an error in your solution. Then feed your “incorrect” solution to the LLM and ask it to find the error! This may give us new insights into

whether LLMs can be useful for debugging engineering solutions... get creative with the engineered errors!

One student created a nice, subtle error which GPT-4 easily found. GPT-4 offered various opinions about the performance of the built-in MATLAB function `fminsearch`, convincing the user of the value of ChatGPT for engineering debugging. Another student created a good error to plant in their code, which ChatGPT found and then rewrote the code to fix the error without even being prompted to. The student then created a second, more subtle error, which ChatGPT was also able to find! Something about the nature of this particular LLM assignment elicited extra effort from the students beyond what was explicitly requested. A student planted multiple errors at once, which ChatGPT correctly identified, and so the student doubled down and created even more errors to find! ChatGPT successfully found these additional errors as well. One student noted in their assignment that even over the course of a month or two, they have seen a noticeable improvement of ChatGPT's ability to debug MATLAB code, from finding only one error at a time to greater performance in its error finding ability. A student planted several errors at once, found that ChatGPT found all but one of the errors. However, after further prompting, GPT was able to locate the final error and revise the code appropriately. One student was impressed that ChatGPT provided an itemized list of corrections and concluded that it is a valuable debugging tool. A student preferred ChatGPT's version of the code over their own and asked for recommendations of initial conditions to try. This was not a required part of the problem, but useful nonetheless.

In terms of subtle errors to plant, students were generally impressed with ChatGPT's ability to find errors in their programs. One student changed a few of the concentration data values to negative values, and surprisingly ChatGPT caught this and deemed it not quite right. Amusingly, one student noted that ChatGPT found their intentional error, and also found an unintentional error! Another student found that ChatGPT was unable to find their planted error the first time, but successfully located it after a second attempt. A student mismatched the data file lengths, which ChatGPT was able to flag as incorrect. One student tried to "fool" ChatGPT by feeding it a MATLAB script with zero errors... but it did not take the bait. Instead, it suggested that the user try different initial guesses, which is a perfectly valid suggestion for nonlinear algorithms. This student then planted a couple of incorrect data points, which ChatGPT was able to find. They concluded that it helps to tell ChatGPT how many errors there are (if known). A student determined that ChatGPT can find that most subtle of programming errors, the sign error. The student proceeded to ask GPT's opinions on the rest of the assignment as well. One student noted that ChatGPT will find and report errors even when not explicitly

asked to do so, and also found that it did not seem to rely on explanatory comments provided by the programmer when finding MATLAB errors. Interestingly, one student tried the Microsoft Bing chatbot, and determined that unlike OpenAI's ChatGPT, Bing does have the capability to run MATLAB code on its own.

HW#5:

LLM portion: Ask ChatGPT or other LLM to solve one of the two problems above, without specifying the method to be used. Compare it's answer to yours, and discuss how they are different.

Multiple students utilized the newly developed image prompt capability of ChatGPT to feed the original problem statement as input and found that the LLM was able to solve it by writing a new MATLAB code. One student, wise to ChatGPT's ways, introduced this bit of prompt engineering to various tasks: "without using Python". By this point in the course, some students had taken to submitting long interactive sessions with their LLM of choice, packed with new insights about the course material.

HW#6:

LLM portion: This time, let's test the ability of ChatGPT or Google Bard to interpret images. If you do not have access to the ChatGPT mobile app, you can try the latest version of Bard which allows for uploaded images (bard.google.com/chat). After you have solved the two problems above, export one of your Matlab plots of the integrated solution as a jpeg file, suitable for upload to one of these LLMs. Make sure you label your figure well, and then see what ChatGPT/Bard has to say about the figure. Is there any insight or useful commentary it provides?

Students were generally impressed with the amount of information that ChatGPT and Bard are able to extract from a submitted image. Additional prompting was able to push ChatGPT further with its interpretation, for instance commenting on the steady-state behavior of a time-series graph. Bard provided some impressive feedback on a graph plotting an ordinary differential equation solution, somehow understanding it to be a plot of monovalent ligand binding to cell surface receptors vs. time, explaining the behavior at different times, and even describing how to extract kinetic parameters from the curve. Perhaps these surprising insights were facilitated by the internet-enabled functionality of Bard. One student received a very informative response on the receptor binding plot, and ChatGPT placed these numbers into real-world perspective, opining that "the time to reach near-equilibrium seems quite fast." Another student gave Bard a few hints, such as the method of integration used, along with

their input image. One student fed Bard a three-part figure, which the LLM was able to interpret, and Bard provided some interesting background information on the history of environmental policy (the ODE example in question was related to the simulation of toxic chemical concentrations in the atmosphere), and even expressed some opinions on the crisis of climate change! Another student submitted figures to Bard and it suggested ways that the appearance of the figures could be improved, by changing font sizes, axis labels, and legends. Such aesthetic feedback was less commonly encountered. One student provided ChatGPT with an image of a plot and asked if the figure seemed consistent with the governing equations of the problem, and the LLM (wrongly) suggested that the two didn't match. On the other hand, another student fed Bard a figure labeled "ODE figure" and somehow Bard was able to deduce the precise receptor-ligand binding ODE equation. In that session Bard was also able to carry out a stability analysis of the ODE, a task that BME 7410 students are expected to master. Several students correctly recognized that ChatGPT and Bard tended to extrapolate beyond the content of images and provide related information and/or links to fairly relevant topics. One student took a deep dive to explore Bard's ability to accurately read and recognize the content of image input, with the LLM sometimes misreading graphs or even hilariously misidentifying the subject of an image.

HW#7:

LLM portion: For this final homework assignment, there are no restrictions. Think of something creative but relevant to do with the LLM(s) of your choice. Spread your wings and fly!

One student asked ChatGPT to design a lesson plan to teach "Understanding Two-Sample t-Tests with Unequal Variances" and it did a surprisingly good job with this, breaking down how many minutes to devote to each section and what materials would be needed for the lecture. Another student compared the responses of ChatGPT and Google Bard when asked for advice on how to manipulate a data set to obtain a significant result on a statistical test and evaluated the ethical responses both LLMs provided. A student explored the image generation capabilities of ChatGPT, asking it to illustrate a multistep nanofabrication protocol, and then had some additional fun by asking it to design some jewelry. Another student fed the course syllabus to ChatGPT and asked how its services could be more useful for the course... an exercise that could have been quite useful before the start of the semester! A homework assignment and final exam were created by ChatGPT, together with exam solution. Some students asked ChatGPT or Bard to create statistical testing examples complete with answer key, as this was the final unit covered in the course and the topic of homework

#7. Multiple students scanned their own homework assignments and fed the images to Bard, requesting feedback from the LLM. Bard impressed with its ability to decipher various handwriting examples.

Synergistic LLM Content and Activities

The instructor shared various preprints and journal publications on a range of LLM-related topics [6–9]. Throughout the semester, interesting tweets and memes about ChatGPT were also shared in lecture. Some real time, in-class demonstrations of LLM tools were carried out as new capabilities became available (e.g., the image-based prompts of ChatGPT Plus for iPhone). In November 2023, the Bioengineering Division of the American Society of Mechanical Engineers invited M.R.K. to give a webinar presentation on the topic of "Generative AI in Publishing," as part of their Webinars on Generative Artificial Intelligence series. Through a bit of fortuitous scheduling, their regular webinar time precisely overlapped with the BME 7410 lecture time and day, and so we broadcast the webinar live from the classroom. The second presentation in this webinar was on the topic of Generative AI in Research, and quite relevant to the overarching goals of the course. Students in attendance were encouraged to log in to the Zoom webinar via their personal laptops as well, to participate in the Q&A session for both speakers using the chat function of Zoom. Many of the students did just that, making for an interactive and lively experience.

Anonymous Course Evaluation Feedback

There were plenty of non-LLM aspects of the course for students to provide constructive criticism on, however the LLM activities garnered very little criticism in the end-of-semester anonymous surveys. One student commented "LLM homework is interesting and fun," while another student commented "I wonder if ChatGPT's Data Analysis will be solidified soon. It would be cool to see the implementation of how GPT's Data Analysis could be used." Despite all this interesting new content, at least one student remained unconvinced: "I'm not sure the ChatGPT component is needed."

Conclusions

Overall, it seems that the graduate students in BME 7410 enjoyed the LLM component of the course and found it worthwhile. It should be pointed out that no material was removed to make room for this new material, thus representing an expansion of the topics typically covered in this long-running course. There can be no doubt that every student and the instructor gained an extensive new exposure to LLMs

during this course and increased their understanding of the current capabilities and limitations of these tools. Several of us noted that the abilities and behaviors of these LLMs seemed to change from September–December 2023, highlighting the dynamic nature of this emerging field. One interesting idea for future use of LLMs in engineering courses would be to introduce a final project based on LLMs, which would provide for a more in-depth, immersive experience.

Acknowledgements The authors thank other students in the BME 7410 class whose creative use of LLMs deserved mention: Grace Adams, Emily Berestesky, Alexandria Carter, Dara Chanthavisay, Taryn Dunigan, Paolo Francisco, Anna Kittel, Malachy Newman, Aviv Preminger, Drew Ray, and Xuerong Zhang.

References

1. King, M. R. The future of AI in medicine: a perspective from a chatbot. *Ann. Biomed. Eng.* 51:291–295, 2023. <https://doi.org/10.1007/s10439-022-03121-w>.
2. King, M. R. A conversation on artificial intelligence, Chatbots, and plagiarism in higher education. *Cell Mol. Bioeng.* 16:1–2, 2023. <https://doi.org/10.1007/s12195-022-00754-8>.
3. King, M. R. A place for large language models in scientific publishing, apart from credited authorship. *Cell Mol. Bioeng.* 16:95–98, 2023. <https://doi.org/10.1007/s12195-023-00765-z>.
4. King, M. Can GPT-4 formulate and test a novel hypothesis? Yes and no. *TechRxiv.*, 2023. <https://doi.org/10.36227/techrxiv.22517278.v1>.
5. King, M. R., and N. A. Mody. Numerical and Statistical Methods for Bioengineering: Applications in MATLAB. Cambridge: Cambridge University Press, 2010.
6. King, M. GPT-4 aligns with the New Liberal Party, while other large language models refuse to answer political questions. *engRxiv.* <https://doi.org/10.31224/2974>.
7. King, M. Administration of the text-based portions of a general IQ test to five different large language models. *TechRxiv*, 2023.
8. King, M. R. Bing chatbot formulating and testing novel hypotheses in real-time: How slime, chocolate, and Nobel prizes reveal the power and limits of artificial intelligence. *engRxiv.*, 2023. <https://doi.org/10.31224/2937>.
9. King, M. R. Can Bard, Google's experimental Chatbot based on the LaMDA large language model, help to analyze the gender and racial diversity of authors in your cited scientific references? *Cell Mol. Bioeng.* 16:175–179, 2023. <https://doi.org/10.1007/s12195-023-00761-3>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.