ChatGP-Me?

Elias Levy and Bennett Landman
Vanderbilt University, Nashville, Tennessee, United States

Since stepping into the role as Editor in Chief of the *Journal of Medical Imaging* (JMI), I have encountered amazingly creative uses of generative AI ("GenAI," including large language models (LLMs), AI imaging, and multimodal models) that change the way I think about information, probability, and learnable functions. We are clearly in the midst of a transformation of both scientific and creative processes (Is it art?). At my home institution, we have wholeheartedly embraced this discussion, updated our honor code, and partnered with Coursera to offer dozens of GenAI classes beyond our physical campus. Yet, as my daughter starts her first year of college, she is reviewing her syllabi to find exactly the opposite stance, where GenAI is broadly prohibited.

In JMI, I have received papers obviously written by GenAI, and I subscribe to retraction watch reports highlighting what happens when academics lose focus on the ultimate importance of integrity. On a weekly basis, JMI receives submissions that do not appear to have not followed the long-established scientific process (which waste substantial staff and editor time for no public benefit). Yet, it is always easiest to point fingers "over-there" and imagine that the line between right and wrong is crystal clear. Unfortunately, with new technologies, this can be a difficult distinction. For SPIE and JMI, an absolute imperative in the use of GenAI is disclosure.

To provide content for the discussion, I partnered with Elias Levy (co-author and the person who did the vast majority of the work for this editorial) to construct a "ChatGP-me": a chatbot based on my published writings. Then, Elias worked with our lab members in the Medical-image Analysis and Statistical Interpretation (MASI) Lab to evaluate how the GenAI technologies available in 2024 (without specific coding) could impact scientific processes. A synopsis of interaction with Elias's system is included in the Supplementary Material. Briefly, we explore two AI tools and aim to demonstrate what happens when using custom LLMs in five research lab workflows.

The first of these tools is NVIDIA's ChatRTX which allows for secure training and tuning of LLMs locally on relatively affordable graphics cards. ChatRTX can be set up to run on a local area network for use within an organization or deployed on the web for public use. There are various models available for use with the software, including Gemma, Llama2, and Mistral, each with their own strengths and weaknesses. In ChatRTX, tuning is done through the addition of a "dataset" folder and it takes roughly 5 minutes to fine-tune the model with a 1 GB dataset (on an NVIDIA RTX 4090 card). The cost for using this software is almost entirely upfront, with the main cost being the device to run the software itself.

The other tool we looked at is Vanderbilt's Amplify GenAI. The stated mission of the Amplify GenAI team is "to help make enterprise GenAI more open, cost-effective, and accessible." The costs for this software are based on token usage rather than a per-user periodical subscription payment model, with data being stored in the cloud. Amplify GenAI offers a range of pre-trained models including OpenAI's GPT-4o, Anthropic, Google, etc., with a variety of additional features such as assistant creation, prompt templates, workspaces, and sharing within organizations.

Both tools use retrieval augmented generation (RAG) technology, are open-source, and were designed with data privacy in mind. The data going into these models and the prompting used in chats have a very strong effect on the quality of responses. It is also important to acknowledge that at no point is AI meant to replace scientific researchers, but rather to help researchers complete their work with a higher minimum level of quality and do it more efficiently. Through this, we want to help develop a more consistent and safe experience for researchers who wish to use AI in their workflows.

In Quarter 3 2024, Amplify was further along the development process and yielded higher quality responses, so we used it to qualitatively analyze the effectiveness of using custom LLMs in the following lab workflows:

Onboarding: The user in this case is somebody joining a research lab, assumed to have a background in coding and some interest in scientific research. This is characterized within our lab, because we are a moderately sized lab with approximately five new people per semester.

Literature review: The next use case is for assistance with literature reviews, the first step before actually conducting research. The user for this case would be a researcher looking to start a project.

Topic discovery: This use case is for researchers to identify gaps in the literature and is meant to decrease the amount of time it takes for researchers to find a new project.

Pre-submission review: AI will not replace human review, but can be used to quickly evaluate papers for clarity, grammatical errors, and good writing style.

Idea synthesis: The purpose of using AI for idea synthesis is to help researchers ideate and get them out of writer's block.

In Amplify GenAI, an "assistant" is a customizable bot that will respond from a chosen pre-trained model in the platform. Amplify GenAI allows for instructions to help the model understand the content of prompts and how to respond to them. It also allows for the addition of data sources, such as internal documents that models can reference outside of their training. This enables the models to give meaningful information that is relevant to the researcher. These assistants can also be easily shared within an organization using the platform.

In all examples, we used approximately 1 GB of data consisting of 302 of my (BL) refereed journal articles. We decided to only use our own articles for several reasons. First, we wanted to avoid any possible copyright concerns as we only used author-copies of manuscripts in processes internal to our lab. Second, we wanted to be able to ask AI to be critical of our papers for a use case like AI-assisted review, and we felt that it would be overly harsh to ask AI to critique the work of others if we were not willing to looking at our own work through AI's critical lens.

While the integration of AI tools in scientific research offers substantial benefits, it is important to acknowledge and address the ethical implications and potential risks associated with their use, especially in academic environments.

1 Benefits

1. Enhanced efficiency: AI tools can streamline repetitive and time-consuming tasks, such as literature reviews, peer review, and data synthesis, allowing researchers to focus on innovative and complex issues. This increased efficiency can accelerate the pace of scientific discovery.

   - See Examples 2, 3, and 5 in the Supplemental Material.

2. Improved accuracy: Automated assistants can help minimize human errors, such as grammatical mistakes, citation inaccuracies, and inconsistencies in data analysis, thereby improving the overall quality of research outputs.

   - Note: Citation inaccuracy check requires that the LLM can read the source.
   - See Example 4 in the Supplemental Material.

2 Potential Risks

1. Accidental plagiarism: Given that AI models can generate text by synthesizing existing content, there is a possibility that researchers may inadvertently incorporate uncited material into their work. This could lead to intellectual property violations and damage the credibility of the research.

   - See Example 6 in the Supplemental Material.

2. Data privacy: Although tools like ChatRTX and Amplify GenAI are designed with data privacy in mind, the design is not foolproof. Unauthorized access or data breaches in the cloud could compromise proprietary information.

3. Over-reliance on AI: While AI can enhance research, an over-reliance on these tools could lead to complacency in critical thinking and analytical skills. Researchers might become too dependent on AI-generated insights, potentially overlooking novel ideas or alternative approaches that the AI might not consider.

   - See Example 6 in the Supplemental Material.

4. Bias and fairness: AI models are trained on large datasets that may contain inherent biases. This is mitigated slightly through customization of contextual datasets (for example, the assistant datasets), but does not solve the problem. If these biases are not adequately addressed, the outputs generated by AI tools could perpetuate or exacerbate existing biases in research. This is particularly critical in fields like healthcare, where biased data could influence treatment recommendations and outcomes.

   - See Example 6 in the Supplemental Material.

5. Misuse of AI tools: There is a risk that some researchers might misuse AI tools by allowing them to perform tasks that they are not well-suited for or by relying on AI-generated content without proper scrutiny. For example, using AI to write entire research papers, or without significant human oversight, could lead to a decline in the quality, originality, and integrity of scientific outputs.

   - See Example 6 in the Supplemental Material.

3 Concerns from MASI Lab Members

We asked members of the MASI Lab how they used, expected to use, and any strategic concerns associated with LLMs in their scientific writing. Overall, there was agreement on the usefulness of the tool in helping draft and edit their writing, but there were also strong concerns regarding accidental plagiarism.

4 Final Notes

If one asks ChatGPT to write a paper, it will do so immediately and without question. However, it will perform badly as it would likely plagiarize and respond with very broad statements that mean virtually nothing. As one collaborator put it, "There are a lot of words here." Said another way, the paper may be grammatically perfect, but have no substance. The final case presented in the Supplemental Material is an example where we gave ChatGPT 4o the title of a paper that we previously wrote and submitted to SPIE Medical Imaging 2024 and asked it to write the paper for us as a comparison-contrast exercise. We would ask the reader to appreciate how realistic the text appears, while the content is strictly false. Please note that we DID NOT submit this paper to SPIE Medical Imaging 2024.

Unfortunately, the line between grammar and clarity versus hallucinations and plagiarism is not crystal clear with current models. For example, in the pre-submission review, the instruction "If the draft is in outline format, flesh it out into paragraph format" might seem innocuous. This could be intended as simply reformatting text from one style into another. However, it can lead the GenAI to have free reign over content synthesis and open the author to areas of ethical risk and technical inaccuracy. We strongly advise against synthesis tasks for any content that is intended for publication. Authors need to carefully review all text to ensure that the ideas are both their own and also that they ideas combine together to form the overall intended logical argument.

5 Conclusion

GenAI gives us the capability to vastly improve the way in which we conduct research. It can increase efficiency by streamlining tedious tasks, minimizing errors in manuscripts, and maintaining a consistent minimum quality level. That being said, the use of LLMs in an academic setting brings its fair share of risks. From accidental plagiarism to over-reliance on AI and misuse of the technology already occurring, we need to make sure that the research community uses these tools in an intentional way so that we don't denigrate the quality and integrity of our work. With consistent oversight and scrutiny around the way we use these tools, we can find more efficient workflows and increase the overall productivity of research labs and facilities.

We look forward to working with SPIE to implement tools such as these to serve our authors and readers. Yet, we are ever aware of the potential harm (both intentional and unintentional) that they may cause. We will keep an open dialog with our community as we navigate these exciting times.

Best wishes,

Bennett Landman

JMI Editor-in-Chief, with upmost appreciation for Elias Levy and the MASI Lab community