AMERICAN MEDICAL WRITERS ASSOCIATION AMWA JOURNAL
V39 N2 / 2024

THEME ARTICLE

Take the Leap! Steps to Integrate AI Into Your Work

Jenni Pickett, PhD,[1] and Mandy Pennington, BS, MWC[2] / [1]Whitsell Innovations, Inc, Chapel Hill, NC;
[2]Freelance Medical Editor, Downingtown, PA

ABSTRACT

The capability of artificial intelligence (AI) is rapidly increasing and is now sitting on the threshold of the medical writing field. This article presents an AI integration framework that breaks down adoption of this new technology into manageable steps that ensure an informed and thorough approach. Using this framework, individuals and corporations can leverage the benefits of this evolving technology while minimizing risks.

The first step, AI literacy, provides a foundation for informed decision making and appropriate expectations for AI capabilities. This knowledge inspires creative exploration of which use cases would be a suitable application of AI tools. Once the scope of potential uses is defined, risks can be assessed, including incorrect content generation, data leakage, and bias. AI tools can then be evaluated to find tools that can both satisfy the use cases and mitigate critical threats. The final step is to integrate the tools transparently with appropriate guardrails. Then the cycle begins again as AI technology evolves and new applications become possible.

As medical writers are ushered further into the AI era, clear and consistent advocacy for a synergy point between the efficiency of AI and the experience, ability, and humanity of a medical writer will maximize the impact of these innovative models.

The introduction of publicly available generative artificial intelligence (AI) tools has marked a significant turning point in integrating AI into medical communication. Mass-market releases of this new technology started with the launch of ChatGPT by OpenAI in November 2022,[1] which was quickly followed by other significant large language model (LLM) chatbots like Claude by Anthropic and Bard by Google. Generative AI demonstrated remarkable reasoning capabilities that previously required human medical writing expertise such as turning an unformatted data table into a summary paragraph that includes correct comparisons between groups. An upgrade to GPT-4 in late 2023[2] was the first of a wave of large multimodal models (LMMs) that could process and produce images and audio in addition to text.

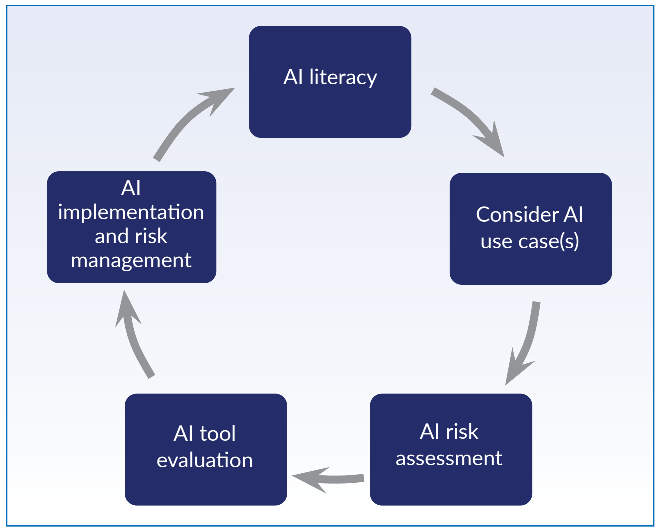

Figure 1. AI integration framework. AI, artificial intelligence.

This period of technological novelty brings with it a wave of AI anxiety among professionals, stemming from fears of job displacement and the reluctance to adapt to changes AI could bring to our daily work. Our psychological response to dramatic change mimics the stages of grief; we start with shock, denial, and anger, then progress to depression due to feelings of overwhelm and inadequacy. Despite these concerns, it's crucial to recognize the potential of AI to optimize drug development processes, reducing the time and expense of bringing new drugs to patients. Exponential growth in the number of AI-enabled drugs and devices may make it necessary for medical writers to adopt generative AI tools to keep up with the workload and do our part to bring treatments to patients faster. This hope for the future can bring us to the upside of the grief response curve, inspiring us to experiment with AI tools, increase our AI literacy, and eventually integrate AI into our work.

In the "Take the Leap! Steps to Integrate AI Into Your Work" presentation at AMWA's 2023 Medical Writing and

ISSN 2163-5315

Communication Conference in Baltimore, Maryland, a comprehensive framework for integrating AI was introduced that is relevant for leaders, employees, and freelancers (Figure 1). As medical communicators, it is up to us to set the foundation for a future in which the efficiency of generative AI is seamlessly blended with the expertise of medical communicators, harnessing the optimum capabilities of both.

AI LITERACY

Successful application of LLM technology to medical writing hinges on our clear understanding of its benefits and risks. LLMs are a subset of AI, specifically within the field of generative AI. LLMs are created through machine learning, specifically deep learning using neural network architecture.

Machine learning is a way to build a computer program that is distinct from typical programming. It uses different hardware designed to perform many computational steps in parallel. Instead of having a programmer define each step for the computer to execute, a data scientist or machine learning engineer sets up a model for the computer to train itself based on provided data sets.[3]

LLMs such as GPT-4 from OpenAI are trained on massive amounts of data, allowing them to comprehend inputs and generate human-like responses. GPT stands for generative pretrained transformer, meaning a machine learning model that generates unique outputs, accesses a broad subset of human knowledge from its pretraining, and understands complex user requests using transformer technology. LLMs can perform tasks like writing, translating languages, coding, creating images, data analysis, and more.

An important step in AI literacy for medical communicators is understanding the difference between LLMs and other AI technologies. Many medical writing and editing AI tools are largely expert systems. Expert systems resemble the structure and consistency you get with highly detailed templates; they are deterministic and built with extensive human knowledge, which leads to predictable outputs. However, they require structured inputs and are less adaptable to varying tasks.

Conversely, LLMs resemble the flexibility and variability you get with simpler, open-ended templates. They are probabilistic, generating outputs based on likelihoods and patterns learned from their vast data sets. This nature makes them more adaptable and flexible but also introduces inconsistency in output and the need for human oversight for accuracy and context.

Writers may interact with LLMs in a standalone chatbot (like ChatGPT), as a functionality in software (like Copilot by Microsoft), or in an internet browser (like Chrome). LLM chatbots, whether standalone or within an application like Word, typically have an interface consisting of a blank box, which leaves the determination of what tasks are suitable and reliable entirely up to the user. AI literacy training to understand what tasks are appropriate for LLMs enables medical writers to leverage these tools effectively, allowing for productivity enhancements while maintaining the high standards of accuracy and context sensitivity crucial in the field.

Understanding how LLMs work helps users predict appropriate tasks. There are fundamental differences in how humans and LLMs write content. Human writers research, understand context, and cite specific sources. They bring a unique perspective and a depth of understanding to their writing, albeit with the possibility of errors and gaps, especially when dealing with unfamiliar topics. LLMs, on the other hand, do not read in the traditional sense. Instead, they are trained on text data broken down into tokens (words or parts of words). LLMs generate text based on patterns learned from their training data, predicting the next token in a sequence. This process can lead to innovative content generation, but it lacks the depth of understanding and context that human writers and editors bring. Moreover, LLMs often cannot trace back to specific sources and might create fake citations or inaccurate content, particularly on topics not well-represented in their training data.

Retrieval-augmented generation (RAG) technology, introduced to mass-market LLM tools in late 2023, assists LLMs by adding a retrieval step before generation. This retrieval step pulls out snippets relevant to the user request from writer-provided or Internet-based sources, which are then used and cited in the generation step.[4]

CONSIDERING AI USE CASES

Armed with a general sense of how LLMs work, users can think of routine challenges in their work that LLMs could help with. However, avoid the feeling that an LLM can solve everything (AI solutionism) when other tools, such as expert systems or regular software would do a better job. Ideal use cases are those that can benefit from an LLM's unique capabilities.

Before evaluating use cases, users should understand that information shared with mass-market LLMs may be used to train a model owned by another entity. AI etiquette requires requesting permission before using someone’s nonpublic content in any AI system and compliance with AI use policies (employer, client, publisher, etc). Because LLMs can provide inaccurate information, use case outputs should be externally verifiable.

LLMs and LMMs can augment users by taking on simple tasks, assist users step-by-step, and amplify users by expanding their skill set (box on next page).

LLM Use Case "A-List"

Augment: delegate specific tasks to the LLM and review the outcome.

Assist: collaborate with the LLM—the model helps write, edit, and create content.

Amplify: the LLM provides new capabilities, such as coding, creating a data visualization, or teaching a new concept.

"Augment" Use Case Examples

- Formatting lists of abbreviations: fixing capitalization and spelling errors with awareness of proper nouns.
- Formatting references: aligning to a provided example style.
- Converting images to text: transform photos of handwriting, slides, or scanned documents to editable text.

"Assist" Use Case Examples

- Preparing slide scripts: generating a draft of a presentation script for a slide based on the slide title and key points.
- Creating images: convert text prompts to pictures for use in presentations or social media.
- Providing technical support: taking users step-by-step through common computer issues (if the LLM recommends entering an admin password or editing registry files, wait for a human to help).

"Amplify" Use Case Examples

- Creating a PubMed search string: converting a text request into Boolean operators.
- Writing macros for Microsoft Office: translating user requests into Visual Basic for Applications and taking the user through the steps to run the program.
- Performing basic data analysis: parsing large data sets and providing charts to visualize the data (users should be mindful that LLMs do not clean data automatically).

AI RISK ASSESSMENT

Before using LLM-based tools, understanding potential risks and planning how to mitigate them is paramount. Establishing a risk profile will help identify tools that fit that profile. Risks related to LLMs can be due to training data limitations, human factors, functional limitations, and/or implementation challenges (Figure 2).

An LLM's knowledge is rooted in its training data. Gaps or weaknesses in training data can result in hallucinations (false information invented by the LLM), outdated outputs, or biased responses.

Examples of Training Data Limitations

- Data set does not include recent information (ie, after the training data cutoff), unless connected to internet browsing capability.
- Data set includes outdated practices and language; LLM is unaware which practices are now preferred or required.
- Data set is missing valuable context because it does not include nondigitized content (eg, conference presentations), content behind a paywall (eg, journal articles), or content in an inaccessible format (eg, regulatory guidances in PDF form).

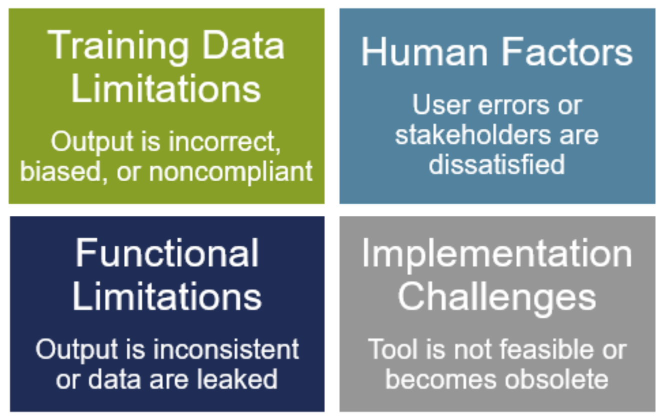

Figure 2. Examples of risk categorization for LLM tools.

LLMs are very different from any technology previously available, which can introduce risks from human users.

Examples of Human Factor Limitations

- Automation bias, the assumption that machine-generated content is accurate.
- Distrust, leading to loss of interest from readers, attrition of employees, or client dissatisfaction.
- Providing the model with an incorrect or outdated source.
- Model damage from poisoned training data or manipulative prompts (prompt injection).

The machine learning process is the root of some risks related to the way LLMs function. LLMs are probabilistic systems that cannot be predicted or entirely understood.

Examples of Functional Limitations

- Opaque decision-making process.
- Inconsistency of output, even with identical prompts.

- Not able to cite sources (unless equipped with RAG technology).
- Can leak your inputs into other users' outputs (eg, proprietary data, protected health information).

Implementing AI systems presents challenges that should be considered as part of a risk assessment.

Examples of Implementation Challengess

- Cost (including employee time and opportunity cost).
- Obsolescence as AI technology quickly advances.
- Finding legal, nonproprietary training data.

Risk evaluations should be captured together with planned risk mitigation steps in a risk management plan. The National Institute of Standards and Technology has created an AI Risk Management Framework and Playbook as a resource to complete this process.[5]

AI TOOL EVALUATION

With a solid understanding of applicable use case(s) and a defined risk profile, it is time to select an appropriate tool to meet both criteria. When evaluating a tool, it is very important to understand which model the tool uses and what, if any, modifications have been made to the model – a tool based on an earlier LLM may be a lot less capable than the more recent models. For an expert system with LLM features, it is important to understand what features of the tool are based on the LLM and what features are based on more deterministic programming so you can determine if the tool matches your risk profile.

Because machine learning involves various degrees of learning, it can be helpful to think of LLM capabilities in layers (Figure 3):

- The foundation model, like GPT-4 in ChatGPT, is similar to the college education of a medical writer, providing a broad base of knowledge.
- Fine-tuning the model with additional specialty data sets like clinical study reports is similar to the specialized knowledge gained by a medical writer in a graduate or certificate program.
- Providing the model with access to your data is like on-the-job training.
- Finally, a collection of proven prompts in a prompt library mimics the efficiency gained with work experience.

When delegating a task to a beginning medical writer, you would provide more detail, instructions, and follow up than you would with an experienced medical writer. The

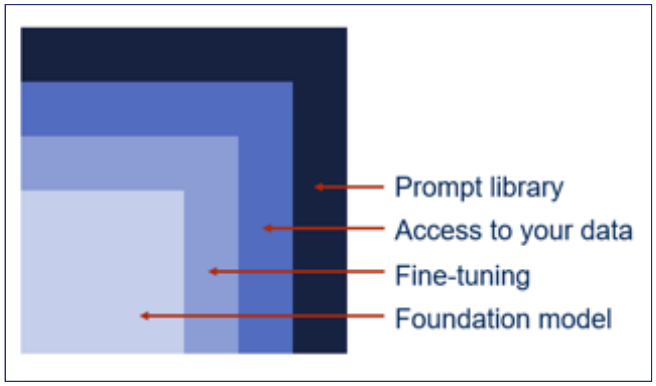

Figure 3. Potential layers of capability of an LLM tool. LLM, large language model.

same logic applies to LLM tools. If the tool has limited layers of capability, your prompt needs a lot of context and specific instruction, and your output may require substantial revision. If your tool has multiple layers of capability, your prompt can be more straightforward and the quality of the LLM output will require fewer edits.

AI IMPLEMENTATION AND RISK MANAGEMENT

After choosing your use case(s), evaluating risks, and assessing and choosing an AI tool, the last step in the AI integration cycle is to integrate the tool into your work or organization. AI transparency is critical before, during, and after implementation. It is vital to address reservations that stakeholders may have and set guardrails for AI use.

Clearly communicate to partners and users what the AI tool is capable of, who will be authorized to use it, when it is appropriate to use AI, why this change is being made, and how to use it appropriately. AI policies to record and share these principles are becoming a common business practice. An AI policy can provide rigid guardrails to protect against the risks identified in your risk management framework.

Building a prompt library of successful, reliable use cases provides a set of flexible guardrails to further improve quality and reduce risk. The PLANTS acronym is a good starting point for building a prompt: persona, length, audience, nuance, type, and style guide (Figure 4).

Improve the probability of successful implementation by starting with 1 or 2 use cases that an LLM can consistently make easier. This feeling of productivity and success can spark more interest in trying other use cases. As you explore new use cases, record not only what works, but what does not work so that those failed prompts can be potentially revised or revisited as LLM capability improves. Sometimes a failed prompt can become successful when adding one or two examples inside the prompt (also known as one-shot or two-shot prompting).

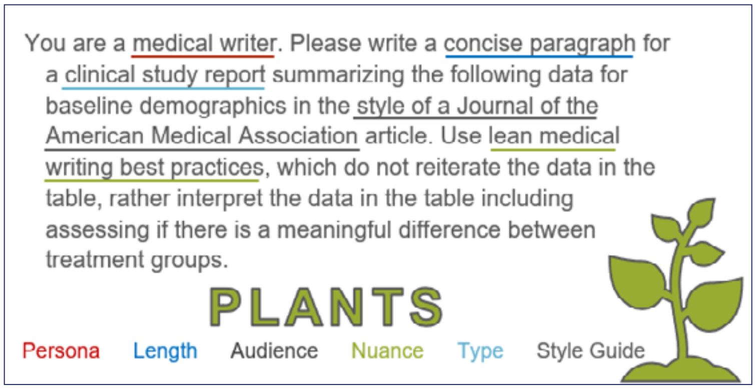

Figure 4. Using the PLANTS method to construct an LLM prompt. LLM, large language model.

A VISION FOR THE FUTURE

Medical writers must have involvement in defining the optimal balance between human effort and AI assistance. At one extreme, staying with the status quo of 100% human effort in drug development means continuing to struggle to accelerate time to market and rising development costs, and potentially falling behind competitors. On the other extreme, replacing entire medical writing functions with AI also presents risks. Health authorities would reject applications after finding missing submission elements, fake references, and hallucinations. Teams would be left with a void of document leadership to break down tasks, set timelines, critically evaluate sources, and gain consensus.

Now is the time for medical writers to define and advocate for a synergy point that combines the expertise of medical writers with the efficiency of AI. The key is to be strategic, integrating AI where it adds value and ensuring that the core responsibilities of medical writing remain grounded in human expertise.

Acknowledgments

The authors acknowledge Cathi Harmon for her contributions to the presentation that preceded this manuscript. We also thank Sean Whitsell, Mary Ellis Bogden, Pam Fioritto, and Ann Winter-Vann for their review of the manuscript.

Author declaration and disclosures: *The authors note no commercial associations that may pose a conflict of interest in relation to this article. The authors discuss multiple AI tools but this article is not intended as an endorsement of any product.*

Author contact: *jenni.pickett@whitsellinnovations.com*

References

1. OpenAI. Introducing ChatGPT. OpenAI. Published November 30, 2022. Accessed January 21, 2024. https://openai.com/blog/chatgpt
2. OpenAI. ChatGPT can now see, hear, and speak. OpenAI. Published September 25, 2023. Accessed January 21, 2024. https://openai.com/blog/chatgpt-can-now-see-hear-and-speak
3. Ochotny N, Byram JK. The use of artificial intelligence and machine learning in clinical research and health care. *AMWA J*. 2023;38(1):30-33.
4. Martineau K. What is retrieval-augmented generation? IBM. Published August 22, 2023. Accessed January 21, 2024. https://research.ibm.com/blog/retrieval-augmented-generation-RAG
5. Trustworthy & Responsible AI Resource Center. Knowledge Base. National Institute of Standards and Technology. Published January 26, 2023. Accessed January 21, 2024. https://airc.nist.gov/Home