ChatGPT and generative AI: What physicians should consider

ChatGPT is a commercial natural language processing tool known as a large language model (LLM). LLMs are a type of augmented intelligence (AI) that can recognize, summarize, translate, predict, and generate text and other content based on knowledge gained from large datasets. ChatGPT and other generative AI tools are designed to be conversational in nature by allowing users to ask questions or request content. These AI tools then generate a plain English, easy-to-understand response to users' queries.

LLMs and generative AI tools are used across many industries and are finding some uses in health care. Health care uses include assistance with administrative functions, such as generating office notes, responding to documentation requests and generating patient messages. Additionally, there has been increasing discussion about clinical applications of LLMs, including use as clinical decision support to provide differential diagnoses and assist in treatment planning. LLMs may also be useful in creating patient education materials.

Limitations of LLMs in health care

While LLM tools show tremendous promise to make a significant contribution to health care in the future, physicians considering using ChatGPT or other LLM tools in a clinical setting or direct patient care should exercise caution and be aware of the technology's current limitations, including:

- **Risk of incorrect and falsified responses:** ChatGPT, one of the most commonly used LLM tools currently available, has a tendency to "hallucinate" when it cannot find an answer to a user query. In this case, the term "hallucinate" is generally used to describe instances in which LLMs may confidently falsify information in order to provide a complete answer to a user's query. ChatGPT in particular has been prone to both hallucinations and other demonstrable errors and omissions. Due to this tendency, clinicians should exercise caution in relying on the results provided without independent verification.

- **Training dataset limitations:** ChatGPT's training data set, while exceptionally large, is not necessarily current. The most recently available information may not be part of the current iteration of the tool and therefore query results may not reflect the most recent clinical guidelines or research. Users should also understand that the sources from which it generates responses may not always be easily verifiable.

- **Lack of knowledge-based reasoning:** Despite the eloquent and conversational tone of their responses, ChatGPT and other LLMs do not actually possess knowledge or perform cognitive reasoning. These tools simply scan available information for what they feel are relevant items and generate a response from these findings. This lack of knowledge and reasoning means it cannot understand or distinguish fact from fiction, right from wrong, etc.

- **LLMs are not currently regulated:** There is no current regulatory structure specifically applicable to these types of AI, no ethical framework, and no widely applicable development standards to ensure their accuracy and performance. ChatGPT and other LLMs are not currently regulated by any health care agency, including the FDA, so the accuracy and veracity of its responses cannot be ensured.

- **Patient privacy and cybersecurity concerns:** Physicians and patients alike should be aware that entering patient personal health information into ChatGPT or other LLM tools may raise patient privacy concerns. There is little transparency or control over the ultimate use of the information entered into a LLM query and the tools could be susceptible to cybersecurity threats. Moreover, LLMs may not be considered "covered entities" or "business associates" under HIPAA. Entering patients' medical information into an LLM could violate compliance with federal or state regulations—resulting in fines or other penalties.

- **Risk of bias, discrimination, and promoting stereotypes:** Given how LLMs are trained and function, they run the risk of potentially incorporating existing bias, returning discriminatory results and promoting stereotypes embedded within training data. Because these types of models cannot necessarily independently identify bias or discrimination or understand when they are perpetuating stereotypes, physicians should be aware of the risk of biased results and results that are discriminatory in their impact on patients from historically marginalized or minoritized communities.

- **Liability may arise from use:** Physicians using ChatGPT or other LLMs at this juncture should be cautious that their reliance on these tools for clinical uses could potentially lead to liability for any resulting patient harm.

The outlook of LLMs in health care

Increasingly, LLM-based tools are being embedded within electronic health record (EHR) systems. Initially, it does not appear as though these tools will be involved in direct patient care but will likely help physicians generate patient messages and evaluate trends using EHR data. However, use of LLMs across EHRs is rapidly evolving. Physicians should take responsibility to understand when and how their EHR systems use these tools and be prepared to answer patient questions. New federal regulations are on the horizon that would require EHR vendors to promote LLM transparency and provide physicians with information about the validity, fairness, and effectiveness of LLM tools.

LLMs, such as ChatGPT and others, are showing significant promise to help minimize the administrative burdens of health care and may ultimately assist physicians in the practice medicine. However, physicians should understand that new technologies present new risks, particularly if they are using LLMs in direct patient care. Physicians need to understand these risks and carefully weigh the benefits of engaging with these new tools prior to integrating them in their workflow.