Application of ChatGPT in Routine Diagnostic Pathology: Promises, Pitfalls, and Potential Future Directions

Casey Schukow, DO, Steven Christopher Smith, MD, PhD,† Eric Landgrebe, ME,‡ Surya Parasuraman, BS,§ Olaleke Oluwasegun Folaranmi, MBBS,‖ Gladell P. Paner, MD,¶ and Mahul B. Amin, MD#***

Abstract: Large Language Models are forms of artificial intelligence that use deep learning algorithms to decipher large amounts of text and exhibit strong capabilities like question answering and translation. Recently, an influx of Large Language Models has emerged in the medical and academic discussion, given their potential widespread application to improve patient care and provider workflow. One application that has gained notable recognition in the literature is ChatGPT, which is a natural language processing "chatbot" technology developed by the artificial intelligence development software company OpenAI. It learns from large amounts of text data to generate automated responses to inquiries in seconds. In health care and academia, chatbot systems like ChatGPT have gained much recognition recently, given their potential to become functional, reliable virtual assistants. However, much research is required to determine the accuracy, validity, and ethical concerns of the integration of ChatGPT and other chatbots into everyday practice. One such field where little information and research on the matter currently exists is pathology. Herein, we present a literature review of pertinent articles regarding the current status and understanding of ChatGPT and its potential application in routine diagnostic pathology. In this review, we address the promises, possible pitfalls, and future potential of this application. We provide examples of actual conversations conducted with the chatbot technology that mimic hypothetical but practical diagnostic pathology scenarios that may be encountered in routine clinical practice. On the basis of this experience, we observe that ChatGPT and other chatbots already have a remarkable ability to distill and summarize, within seconds, vast amounts of publicly available data and information to assist in laying a foundation of knowledge on a specific topic. We emphasize that, at this time, any use of such knowledge at the patient care level in clinical medicine must be carefully vetted through established sources of medical information and expertise. We suggest and anticipate that with the ever-expanding knowledge base required to reliably practice personalized, precision anatomic pathology, improved technologies like future versions of ChatGPT (and other chatbots) enabled by expanded access to reliable, diverse data, might serve as a key ally to the diagnostician. Such technology has real potential to further empower the time-honored paradigm of histopathologic diagnoses based on the integrative cognitive assessment of clinical, gross, and microscopic findings and ancillary immunohistochemical and molecular studies at a time of exploding biomedical knowledge.

Key Words: ChatGPT, pathology, AI, digital pathology, NLP, chatbot, machine learning, anatomic pathology, immunohistochemistry, tumor boards

(*Adv Anat Pathol* 2024;31:15–21)

There has recently been an explosion of progress in the capability and adoption of Large Language Models (LLMs), which are forms of artificial intelligence (AI) models that have been trained on large amounts of text data using deep learning algorithms and exhibit strong emergent capabilities including question answering, translation, and computer programming.[1,2] There has been a wave ofproduct releases built around these LLMs, including PaLM API[3] and Bard[4] from Google, New Bing[5] from Microsoft, Claude[6] from Anthropic, and ChatGPT[7] from OpenAI. In addition to product launches, there has been astonishing progress in the underlying models powering these products, with recent announcements of PaLM 2[8] from Google, LLaMa[9] from Meta, and GPT-4[10] from OpenAI, which have much stronger capabilities than the previous generations of models. Moreover, there have been incredible advances in open-sourced models, including the release of Meta's LLaMa, open-assistant,[11]OpenLLaMA,[12] StableLM,[13] and many others. Improvements in Parameter Efficient Fine-Tuning methods have also been noted,[14] which make it dramatically easier for anyone to give new knowledge and capabilities to these and other open-sourced models.

As recent scholarship regarding potential applications of ChatGPT and similar chatbot technologies (or, simply, "chatbots") in various facets of health care has grown,[15] its relevance in pathology has begun to be explored. Recent literature, albeit very limited, has begun to pave the path toward roles which chatbots like ChatGPT may play in the field. However, little to no scholarship to date has illustrated the actual application of chatbots in routine diagnostic pathology. In this review, we target ChatGPT (built on the GPT-3.5 base model) without the use of plugins or other accessories as a paradigm regarding this aforementioned implementation to pathology. This model meets the desired

From the *Department of Pathology, Corewell Health, Royal Oak, MI; †Department of Pathology, Virginia Commonwealth University School of Medicine, Massey Cancer Center, and Richmond Veterans Affairs Medical Center, Richmond, VA; ‡Cornell University College of Engineering, Ithaca, NY; §Department of Biological Sciences, University of Toledo, Toledo, OH; ‖Department of Anatomic Pathology, University of Ilorin Teaching Hospital, Ilorin, Nigeria; ¶Departments of Pathology and Surgery (Urology), University of Chicago, Chicago, IL; #Department of Pathology and Laboratory Medicine, University of Tennessee Health Science Center, Memphis, TN; and **Department of Urology, University of Southern California Keck School of Medicine, Los Angeles, CA.

The authors have no conflicts of interest to disclose.

Reprints: Casey Schukow, DO, Department of Pathology Corewell Health, 3601 W. 13 Mile Rd Royal Oak, MI 48073 (e-mail: casey.schukow@gmail.com).

All figures can be viewed online in color at www.anatomicpathology.com.

Supplemental Digital Content is available for this article. Direct URL citations are provided in the HTML and PDF versions of this article on the journal's website, www.anatomicpathology.com.

criteria of (1) having strong performance, (2) being freely accessible for public use, and (3) not being augmented with an internet search or other external knowledge, which allows us to directly probe the knowledge of the model and not any database it may be searching over.

Furthermore, we hope to address the promises, pitfalls, and possible future directions of ChatGPT's utility in diagnostic anatomic pathology by way of providing examples of question-and-answer scenarios, illustrating the remarkable capabilities of current technology. Along the way, relevant literature will be referenced, and a brief discussion on the limitations of our methodologies will be offered.

What Is ChatGPT?

ChatGPT (Generative Pretrained Transformer) is an AI tool developed by OpenAI (Parent company: OpenAI, LLC).[16] It learns a statistical model ("Artificial Neural Network") from large amounts of text data and generates creative responses to inquiries. These inquiries simulate human conversation. Recently, ChatGPT has been gaining recognition as the "next big thing"[17] in both health care[18] and academia.[19] Kung et al[20] pitted ChatGPT's previous model, GPT-3, against 305 publicly available US Medical Licensing Exam (USMLE) practice Step exam question prompts (step 1: n = 93, step 2: CK: n = 99, and step 3: n = 113) to observe how its responses would hold up. They found that "ChatGPT performed at > 50% accuracy across all exams, [while] exceeding 60% in most analyses."[20]

According to the USMLE website, "examinees typically must answer approximately 60% of items correctly to achieve a passing score" on any of the USMLE Step exams from year to year.[21] Thus, ChatGPT appears to have overcome the passing threshold of one of the most difficult series of credentialing examinations in the United States. Since then, additional progress has been made with Google's Med-PaLM 2,[22] a fine-tuned LLM that consistently performed at an "expert" doctor level on medical exam prompts, scoring 85% on USMLE-style questions. GPT-4 excelled at a variety of similar human benchmarks, including scoring at estimated 75th percentile in the Medical Knowledge Self-Assessment Program.[10]

This application has both free and paid tiers that offer access to an automated AI natural language processing chatbot system intended to help save users time in their daily routines. It can be thought of as a "virtual assistant."[23] To use the free version of this publicly available platform, users sign up to create an account via this link: https://chat.openai.com/auth/login.[24]

After creating an account, a user can subsequently type his or her questions/inquiries into text-designated areas using a variety of formats (such as spreadsheets, large-text summaries, etc.) Once the user prompts the software (ie, hits the "ENTER" button on a keyboard), ChatGPT will generate an automated response, and the user can ask follow-up questions, like a real conversation. These "conversations" (or chats) then become saved in the user's profile history, where they can be accessed again later. Although uncertainties exist regarding whether the free version of will exist forever, paid versions (of ChatGPT and other chatbots, frankly) may be of minimal concern to the public as long as they are affordable and of great value.

How ChatGPT Works

ChatGPT is a product that allows conversing with (ie, is powered by) a LLM[25] built on the GPT series of Transformer Artificial Neural Network[26] models from OpenAI. These GPT models start with no understanding of language and learn to "autocomplete." or predict the next piece of text over an extremely large set of documents and webpages on the internet. As the model gets increasingly accurate at predicting what text is likely to come next, it develops an increasingly rich understanding of language and develops strong performance across a broad range of tasks, including question answering, translation, and computer programming.[1,2] The result of this training is a "foundation model," [26] or a statistical model of human language, that distills a large amount of knowledge from the internet at the time of its training.[27] ChatGPT is a chatbot built on top of these foundation models with additional training on feedback from humans to improve its truthfulness, reduce toxicity, and more reliably follow instructions.[28] The end result is a chatbot that one can converse with on a broad range of topics.[29]

Technical Limitations of ChatGPT

As a result of its technical design, the version of ChatGPT we explore: (1) has no knowledge of events beyond the time of its training (early 2022),[7] (2) cannot access external data that are not directly provided by the user in the conversation (Although a range of other products including new Bing, Google's Bard, and ChatGPT plugins allow the model to search the internet directly[30]) and (3) may "hallucinate" plausible-sounding text that is not actually true. These models can also repeat common misconceptions or replicate undesirable biases in their training data.[31] As such, it is crucial that humans vet the reliability of these chatbot outputs; this would be particularly relevant to our discussion of its potential use in routine diagnostic pathology when any aspect of the chat output may be used for patient care. In addition, the specific way of asking questions ("prompts") to these models can be very important. For example, asking models to "think step-by-step" can reduce errors on mathematical reasoning tasks by nearly 75% and make the results more easily interpretable by humans.[32] In this work, we have not optimized our question format for ChatGPT, but it is plausible that reliability and interpretability could be improved by more carefully questioning the model or asking it to refine its outputs in light of errors we find in its initial answers.

Promises, Pitfalls, and Potential Future Directions of Application in Routine Diagnostic Pathology

According to a recent review by Sallam,[33] potential benefits of ChatGPT in health care include (1) improved scientific writing; (2) elevated research equity, versatility, and efficiency; (3) improved day-to-day practice (eg, cost savings, streamlined workflow, patient health literacy); and (4) enhanced education (eg, personalized learning). This potential can even be expanded to public health, where ChatGPT may be able to support patients and communities in making informed health decisions.[34] However, while ChatGPT and other chatbots may figure prominently in the future of health care and medical writing, limitations exist that must be considered before widespread, mainstream use.

Before Sallam, Biswas,[35] and Kitamura[36] each individually noted some of these concerns. First, AI use in writing raises ethical questions regarding authorship, accountability, and authenticity. Second, legal issues, such as copyright infringement, health care regulatory

compliance, and frameworks, for ensuring the privacy, quality, and validity of documentation in patients' health records must be considered. Third, as these automated language technologies require large amounts of previously reported data to operate their responses, this may lead to a lack of innovation, accuracy, and, ultimately, bias, unless newer data (eg, medical knowledge) is included in the training or human feedback is used to improve model performance. (It is expected by many within the machine learning (ML) community that these models will continue to get much better at a rapid pace. For example, jumps between models in the GPT series (ie, GPT-1 to GPT-2 to GPT-3 to GPT-4) have all been accompanied by massive qualitative jumps in capability.) In fact, some studies have shown that AI may overlook certain social or cultural aspects of health care (eg, the social determinants of health).[37] Last, but not the least, it is apparent that AI natural language processing–generated medical references must be made transparent and clear to readers whenever they are used.

Thus, what are some pitfalls regarding incorporating ChatGPT and other AI-automated processing systems into routine diagnostic practice? In March 2023, Nakagawa et al[38] addressed this question by simply asking ChatGPT via the prompt: "Pathology and AI: What could possibly go wrong?" Its responses were in line with the concerns and findings of Sallam, Biswas, and Kitamura. ChatGPT acknowledges that current limitations and possible pitfalls include (1) lack of data diversity, (2) perpetuating pre-existing data biases, (3) lack of diagnosis understanding/interpretation, (4) protecting patient health data from unauthorized access, (5) compliance with health care regulations, and (6) validation and accreditation of routine AI usage by regulatory bodies.

These concerns are all valid, as Nakagawa and colleagues acknowledge that there currently exists much uncertainty regarding the impact of AI on pathology. Some of these concerns may also arise from the "black box" theory of AI, in which ML software like ChatGPT have complex decision-making processes, which are not fully understood by humans.[39,40] This complexity often leads to a lack of trust by users, which can be ameliorated by improved ML algorithms trained with increased biological explanation and clinical experience.[41]

However, despite these hurdles, Nakagawa and colleagues also state that if AI technologies are widely accepted, they may be able to help improve efficiency and provide support for pathologists. As Nakagawa and colleagues appropriately identified, humans are best suited for helping to establish new ideas, approaches, and diagnostics and will always need to review AI-generated content for approval. But as workloads increase and AI-generated content become more prevalent, accurate, and efficient, systems such as ChatGPT will become essential aids. ChatGPT's accuracy can already be seen when solving higher-level pathology problems, as determined by a cross-sectional study from Sinha et al[42] in February 2023. In training institutions, ChatGPT and other chatbots may be able to help attending teach trainees how to write brief, properly worded, and accurate pathology reports that completely cover all aspects of diagnoses. Rather than patients asking "Dr. Google"[43] (ie, searching health information online) to find out more about their diagnoses from their pathology reports, too, ChatGPT and other chatbots

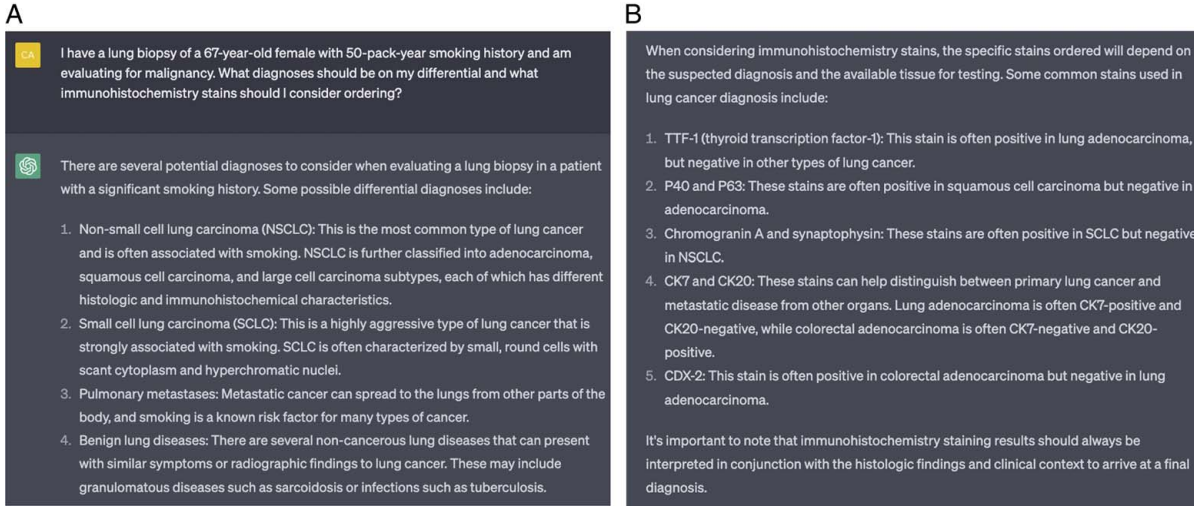

FIGURE 1. A, The first author (C.S., represented by the "CA" to designate the first 2 letters of his first name in the upper left-hand corner) entered an inquiry into ChatGPT: "I have a lung biopsy of a 67-year-old female with 50-pack-year smoking history and evaluating for malignancy. What diagnoses should be on my differential and what immunohistochemistry stains should I consider ordering?" ChatGPT's responses (see B for the remainder) is highlighted starting next to its respective logo underneath the inquiry (ie, There are several potential diagnoses…). Differentials ChatGPT considered included non–small cell lung carcinomas (eg, adenocarcinoma, squamous cell carcinoma, and large cell carcinoma), small cell lung carcinoma, pulmonary metastases, and benign lung diseases (eg, granulomas). The version of ChatGPT is from March 23, 2023. The version of ChatGPT is from March 23, 2023. Time to fully generated response (A and B) after inquiry entry: ~48.30 seconds (Supplementary File 1 regarding recording response times and capturing figures). B, The second part of ChatGPT's response (beginning with When considering…) to the initial inquiry by the first author (C.S., see A). IHC stains mentioned include TTF-1, P40, P63, Chromogranin A, CK7, CK20, CDX-2. ChatGPT ends its response by including a note regarding the importance of correlating immunohistochemistry stains with "histologic findings and clinical context to arrive at a final diagnosis." The version of ChatGPT is from March 23, 2023. Time to fully generated response (A and B) after inquiry entry: ~48.30 seconds (Supplementary File 1 regarding recording response times and capturing figures).

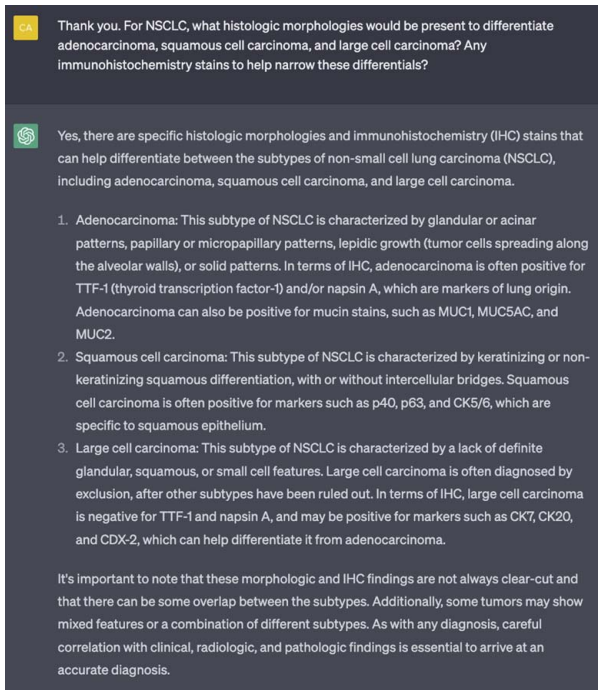

FIGURE 2. The first author (CS) followed up the initial inquiry with another question regarding histologic morphologies and pertinent IHC stains for several of the lung biopsy differentials mentioned in Figure 1A. ChatGPT provided morphologic features of adenocarcinoma, squamous cell carcinoma, and large cell carcinoma. Further, IHC stains include napsin A, MUC1, MUC5AC, MUC2, and CK5/6. The version of ChatGPT is from March 23, 2023. Time to fully generated response after inquiry entry: ~56.67 seconds (Supplementary File 1 regarding recording response times and capturing figures). Please see this image in color online.

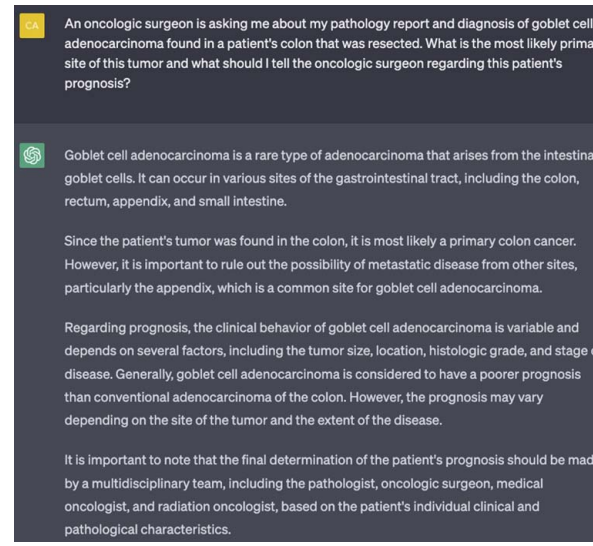

FIGURE 3. The first author (CS) proposed a separate, hypothetical scenario to ChatGPT regarding a rare cancer (goblet cell adenocarcinoma, colon) for discussion at a tumor board (ie, with an oncologic surgeon). ChatGPT provided a brief overview of this cancer, where its most likely primary site may have been (ie, appendix), and gave a general statement of its prognosis (ie, poor). The version of ChatGPT is from March 23, 2023. The version of ChatGPT is from March 23, 2023. Time to fully generated response after inquiry entry: ~17.74 seconds (Supplementary File 1 regarding recording response times and capturing figures). Please see this image in color online.

may provide to be a more appropriate outlet when patients are not able to access their own doctors due to provider time constraints, etc.

We also note that despite the generally superior performance[44] of models like the GPT series, patient data privacy considerations may make the use of such models for diagnosis impossible. Currently, chat histories are by default, stored and usable for training by OpenAI, which creates a serious privacy risk if patient data were to be input. On the contrary, if this feature could be theoretically optimized to accommodate electronic medical records (EMRs), chatbots like ChatGPT may be able to swiftly search and summarize relevant patient history, saving much time for pathologists (eg, prior diagnoses with tumor grades). Regardless, for this and other reasons, it may be preferable to use open-sourced models that can be run locally to avoid leaking confidential information and have more direct control over the model's behavior until widespread EMRs (including via remote access) can be safely, succinctly, and accurately navigated by chatbots.

Although more studies are required to better assess the spectrum of ChatGPT's application in routine diagnostic pathology, the findings mentioned above from Sinha and colleagues are promising. We ("the authors") asked ChatGPT several hypothetical questions for this review to further evaluate its potential in the field. Questions were generated by the first author (C.S.) after creating a free ChatGPT account and were collected in a supplemental file that can be seen linked to this review (Supplemental Digital Content 1, http://links.lww.com/PAP/A33). No plugins or other accessories to the platform were incorporated. Subsequent ChatGPT responses regarding immunohistochemistry (IHC) stain sensitivities, assistance with tumor boards, prognostic information, tumor grading, and creating differential diagnoses were all explored. A representative example of a question and response is presented in Figures 1–3, while the full text and further examples are included in Supplementary File 1, Supplemental Digital Content 1, http://links.lww.com/PAP/A33, at the end of this review.

On the basis of ChatGPT's responses to the questions relating to these very routine types of diagnostic pathology queries, the authors, which include several surgical pathologists in different career stages at different institutions (including an author at an international institution, OOF), brought different perspectives to its responses (S.C.S., G.P.P., and M.B.A.). The authors' thoughts on the promise of this type of technology, pitfalls, including concerns revealed by ChatGPT's responses, and ideas regarding the potential of chatbot technology to impact the future of diagnostic pathology, writ large, are summarized in Tables 1–3. The application of ChatGPT and other chatbot technologies in pathology is just beginning to unfold in terms of promises (Table 1), pitfalls (Table 2), and potential future applications (Table 3). In Table 3, specifically, we summarize such potential mainly with largely unanswered (to date) rhetorical questions as to where the field may be heading with ChatGPT and other chatbot technologies.

With the rapid advance of AI systems comes significant potential but also a need for prudence. At the time of this review's submission, a more advanced model of ChatGPT

Copyright © 2023 Wolters Kluwer Health, Inc. All rights reserved.

TABLE 1. Promises of ChatGPT in Diagnostic Pathology

1. Ability to rapidly summarize publicly available knowledge in response to a specific question in a very specialized medical setting (eg, IHC stains in diagnostic histopathology, BerEP4 for thyroid lesions).
2. Ability to be asked to expand response in further detail with reference to a response already rendered by the system (eg, additional IHC stains for thyroid/lung lesions).
3. Ability to address subspecialty-specific questions, such as an unusual tumor subtype, with specific considerations, in an era of very subspecialized diagnostic practice (eg, goblet cell adenocarcinoma).
4. Ability to generate initial differential diagnoses and recommend first steps to consider for testing (eg, mixed dermal inflammatory infiltrate).
5. Ability to tailor responses to very specific diagnostic scenarios (eg, reactive vs. mesothelioma), or modify or prioritize responses based on the user supplying additional clinical data (eg, specifying young age, sex, and family history).
6. Ability to access and summarize even rare knowledge (eg, tumors with clear cell-papillary pattern arising in VHL syndrome).

IHC indicates immunohistochemistry; VHL, von Hippel-Lindau.

(GPT-4) has been released,[10] demonstrating the speed by which AI-generated software is continually updating. Many factors can influence how useful information can be extracted from AI, including how: (1) questions are framed, (2) users express emotions, and (3) the program implements responses. A successor to ChatGPT's GPT-3.5 model,[10] GPT-4 has been shown to be valuable in creating patient medical records from patient-physician transcripts, answering USMLE and other standardized/licensing exam problems, and providing physician consultation advice.[45] Another major advantage of this and other subsequent versions (ie, GPT-4.5) can also analyze images which is more relevant to pathology. GPT-4, however, is not routinely available to all users (as of April 2023). To obtain it, subscribers must pay a fee of $20 (US dollars, or USD) per month to obtain "ChatGPT Plus" accounts, and there are currently user waitlist delays with usage limits as the software upgrades to accompany this demand.[46]

TABLE 2. Pitfalls of ChatGPT in Diagnostic Pathology

1. Inability to distinguish or dismiss questions that are non sequitur or largely irrelevant (eg, grading of melanoma).
2. Potentially limited access to cross-referencing labels, synonyms, and technical details (eg, IHC stains).
3. Responses may, infrequently, include explicitly incorrect responses, rendered with an authoritative tone, with no explanation (ie, "hallucinations"; eg, describing mucinous tubular and spindle cell carcinoma with *TFE3* translocation)
4. Lack of primary source references for responses.
5. Responses may not express an appropriate degree of uncertainty or incomplete degree of published knowledge (eg, histologic range of *ALK*-translocated RCC)
6. Responses may be sourced in "backward-looking" databases and not employ new technology or reflect evolving understanding and nomenclature (eg, clear cell-papillary RCC rather than "tumor" as in under updated WHO classifications).
7. Unrestricted patient access to extended diagnostic and prognostic discussion in the era of immediate release of pathologic diagnoses.
8. Inability of the algorithm to address "local" differences in available or validated resources for diagnosis or idiosyncrasies of distinct patient populations.

IHC indicates immunohistochemistry; WHO, World Health Organization.

TABLE 3. Potential Future Directions of ChatGPT in Diagnostic Pathology

How will ChatGPT impact the future of published medical literature?
1. Need for a new framework on publication/authorship ethics in a new age of AI-sourced digital composition.
2. Should primary research/scholarship be written differently with anticipation of impending direct use by AI algorithms?
3. Will trends toward open access increase or decrease in response to AI availability and direct use of all publicly available open-source literature?
4. Will publication types such as general reviews become less important than primary scholarship or opinion given availability of AI-based summaries?
5. Will publication impact assessment need to account for AI access?

How will ChatGPT impact the future of training residents and fellows?
1. Will fund of knowledge become less important than learning resource utilization?
2. Need for training in critical assessment of output of AI-sourced tools and commentary for future users?
3. Will availability of AI-assisted approaches encourage or discourage further subspecialization in diagnostic pathology?
4. Can ChatGPT be leveraged to teach or mentor (eg, how to write pathology reports)?
5. Will future trainees and attendings save time searching EMRs with chatbot-assisted technologies without breaching patient privacy and confidentiality?

How will ChatGPT impact the future of interdisciplinary medical practice?
1. Will availability of AI-sourced commentary change the frequency, and nature of consultation between practices?
2. Will consultation with pathologists become less valuable to clinicians considering resources like ChatGPT?
3. Will ChatGPT be able to help alleviate patients from addressing "Dr. Google" when they have inquiries regarding their pathology reports when they are not able to ask their own doctors?

AI indicates artificial intelligence; EMRs, electronic medical records.

Limitations of This Review

A more in-depth review and analysis of all available models with further research assessing practical applications is a promising area for future work. Although we admit that this may be a major limitation of our presenting review (ie, only focusing on 1 chatbot model as opposed to the many listed earlier in the introduction), the intention of this article is to provide a conversational primer regarding the application of chatbots in routine diagnostic pathology while using ChatGPT as a focal point for reference, discussion, and deliberation. In other words, we recognize that ChatGPT is one application powered by LLMs, and even though there has been an "explosion" of LLMs and other chatbots at this point, this article is not intended to be a comprehensive review of all LLMs and their present utilities as addressed in the medical literature.

Albeit, we agree that more rigorous, scientific studies (such as blinded patient-controlled or provider-controlled trials, and quality control analyses when accessing EMRs) would help better evaluate the utility of ChatGPT and other chatbots in routine diagnostic pathology. In present, our

entered prompts and evaluation of them merely introduce this utility with anecdotal but well-thought-out comments that may hopefully drive future research. Furthermore, we fail to address the economic impact of ChatGPT and, more broadly, LLM creation in the health care environment (counterpoint: this is out of the scope of this review but is definitely a topic that should be addressed in subsequent studies).

Regardless, the continued evolution and explosion of these foundation model-based technologies make a review such as this difficult to fully address as, no matter when we publish this article, it will be out of date. In the end, this is the truth about how fast the chatbot technology is moving.

CONCLUSIONS

Although more studies are needed to explore the nuances of how to properly enter prompts, ensure response reliability, and derive and validate more specific use cases, there is great potential for AI-generated chatbots like ChatGPT in health care and even routine diagnostic pathology. The potential to bring to one's fingertips, within an extraordinarily short period of time, a useful distillate collating vast amounts of information from diverse sources, as would otherwise be humanly impossible to scan and review, represents a "game-changing" next phase for AI. Although this is true of potential applications in many fields, we would argue that it cannot be more salient than in medicine, where the "doubling time" of medical knowledge has been estimated at around 73 days on average.[47]

We strenuously advocate for prudence in any use of ChatGPT outputs for clinical patient care at this point, since we have only begun to understand this technology and do not know, yet, what we do not know. Thus, the output of any AI chatbot currently available should be vetted through traditionally accepted sources of medical information. Still, we are not certain if we could imagine a more impressive and promising demonstration of the potential for AI. In an era of practice where we must be more productive in less time and yet integrate an ever-greater complexity of data into diagnostic decisions and reporting, we cannot help but wish for an artificially intelligent ally at the scope.

REFERENCES

1. Brown T, Mann B, Ryder N, et al. Language models are few-shot learners. *Adv Neural Inf Process Syst*. 2020;33:1877–1901.
2. Chen M, Tworek J, Jun H, et al Evaluating large language models trained on code. arXiv preprint arXiv:2107.03374. 2021 Jul 7.
3. Generative AI. for Developers. Google LLC. PaLM API. Accessed May 24, 2023 https://developers.generativeai.google/products/palm?utm_source=sem&utm_medium=cpc&utm_campaign=us-palm-mkr-bkws-exa&utm_content=rsa-palm&gclid=CjwKCAjw9pGjBhB-EiwAa5jl3L2RSCfv7qrRzWRt7XtR1-g4r7ycDex6gMvkAPy2qXB-P_LEZL2gEBoC3z8QAvD_BwE
4. Pichai S. An important next step on our AI journey. Google. 2023. Accessed May 24, 2023https://blog.google/technology/ai/bard-google-ai-search-updates/
5. Introducing the new Bing. Microsoft Bing. Accessed May 24, 2023. https://www.bing.com/new
6. Introducing Claude. Anthropic. Accessed May 24, 2023. https://www.anthropic.com/index/introducing-claude
7. Introducing ChatGPT. OpenAI. Accessed May 24, 2023. https://openai.com/blog/chatgpt
8. Ghahramani Z. Introducing PaLM 2. Google. Accessed May 24, 2023. https://blog.google/technology/ai/google-palm-2-ai-large-language-model/
9. Meta Platforms, Inc. Introducing LLaMA: A foundational, 65-billion-parameter language model. Accessed May 24, 2023.https://ai.facebook.com/blog/large-language-model-llama-meta-ai/
10. GPT-4. OpenAI. Accessed May 24, 2023. https://openai.com/research/gpt-4
11. Open Assistant. Accessed May 24, 2023. https://open-assistant.io/
12. Geng X, Liu H. OpenLLaMA: An Open Reproduction of LLaMA. GitHub. Accessed May 24, 2023. https://github.com/openlm-research/open_llama
13. Stability AI Launches the First of its StableLM Suite of Language Models. Stability AIAccessed May 24, 2023. https://stability.ai/blog/stability-ai-launches-the-first-of-its-stablelm-suite-of-language-models
14. Mangrulkar S, Gugger S, Debut L, et al. PEFT: State-of-the-art Parameter-Efficient Fine-Tuning methods. GitHub. Accessed May 24, 2023. https://github.com/huggingface/peft
15. Asch DA. An Interview with ChatGPT about Healthcare. *NEJM Catalyst*. 2023;4:1–8.
16. About Open AI. Accessed April 8, 2023. https://openai.com/about/
17. Blumptre P. ChatGPT. Health care's next big thing? Medical Economics. Accessed April 8, 2023. https://www.medicaleconomics.com/view/chatgpt-health-care-s-next-big-thing-
18. O'Connor S. ChatGPT. Open artificial intelligence platforms in nursing education: tools for academic progress or abuse? [published correction appears in Nurse Educ Pract. 2023 Feb;67:103572]. *Nurse Educ Pract*. 2023;66:103537.
19. Stokel-Walker C. AI bot ChatGPT writes smart essays—should professors worry? *Nature*. 2022. doi:10.1038/d41586-022-04397-7 [published online ahead of print, 2022 Dec 9].
20. Kung TH, Cheatham M, ChatGPT, et al. Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models. *medRxiv* [Preprint]. 2022.12.19.22283643.
21. Scoring & Score Reporting. USMLE.org. Accessed April 8, 2023. https://www.usmle.org/bulletin-information/scoring-and-score-reporting#:~:text=Notice%20of%20such%20review%20and,to%20achieve%20a%20passing%20score
22. Matias Y, Corrado G Our latest health AI research updates. Google. 2023. https://blog.google/technology/health/ai-llm-medpalm-research-thecheckup/
23. Envisioning the Healthcare Landscape with ChatGPT. NYMC News and Events. Accessed April 8, 2023. https://www.nymc.edu/news-and-events/news-archives/envisioning-the-healthcare-landscape-with-chatgpt.php#:~:text=The%20Potential%20Benefits%20of%20ChatGPT%20in%20Healthcare&text=A%20significant%20application%20of%20ChatGPT,topics%2C%20including%20symptoms%20and%20treatments
24. ChatGPT login. Accessed April 8, 2023. https://chat.openai.com/auth/login
25. Harrer S. Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine. *EBioMedicine*. 2023;90:104512.
26. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *Adv Neural Inf Process Syst*. 2017;30:1–11.
27. Bommasani R, Hudson DA, Adeli E, et al On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258. 2021 Aug 16.
28. Adamopoulou E, Moussiades L. An overview of chatbot technology. *Artificial Intelligence Applications and Innovations*. 2020;584:373–83.
29. Li J, Dada A, Kleesiek J, et al ChatGPT in healthcare: a taxonomy and systematic review. nedRxiv. 2023 [Pre-print].
30. How Does ChatGPT Really Work? *New York Times*. Accessed April 17, 2023. https://www.nytimes.com/2023/03/28/technology/ai-chatbots-chatgpt-bing-bard-llm.html
31. Lin S, Hilton J, Evans O Truthfulqa: Measuring how models mimic human falsehoods. arXiv preprint arXiv:2109.07958. 2021 Sep 8.
32. Kojima T, Gu SS, Reid M, et al Large language models are zero-shot reasoners. arXiv preprint arXiv:2205.11916. 2022 May 24.

33. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare (Basel)*. 2023;11:887.
34. Biswas SS. Role of Chat GPT in Public Health. *Ann Biomed Eng*. 2023;51:868–869. [published online ahead of print, 2023 Mar 15].
35. Biswas S. ChatGPT and the future of medical writing. *Radiology*. 2023;307:223312.
36. Kitamura FC. ChatGPT is shaping the future of medical writing but still requires human judgment. *Radiology*. 2023;307: 230171.
37. Delgado J, de Manuel A, Parra I, et al. Bias in algorithms of AI systems developed for COVID-19: a scoping review. *J Bioeth Inq*. 2022;19:407–419.
38. Nakagawa K, Moukheiber L, Celi LA, et al. AI in pathology: what could possibly go wrong? *Semin Diagn Pathol*. 2023;40: 100–108.
39. Rouse M Black Box AI techopedia.com. Accessed April 8, 2023. https://www.techopedia.com/definition/34940/black-box-ai#:~: text=Black%20box%20AI%20is%20any,of%20explainable% 20AI%20(XAI)
40. Savage N. Breaking into the black box of artificial intelligence. *Nature*. 2022. doi:10.1038/d41586-022-00858-1. [published online ahead of print, 2022 Mar 29].
41. Poon AIF, Sung JJY. Opening the black box of AI-medicine. *J Gastroenterol Hepatol*. 2021;36:581–584.
42. Sinha RK, Deb Roy A, Kumar N, et al. Applicability of ChatGPT in assisting to solve higher order problems in pathology. *Cureus*. 2023;15:e35237.
43. Van Riel N, Auwerx K, Debbaut P, et al. The effect of Dr Google on doctor-patient encounters in primary care: a quantitative, observational, cross-sectional study. *BJGP Open*. 2017;1:bjgpopen17X100833.
44. LMSYS Org. Chatbot Arena Leaderboard Updates (Week 2). LMSYS Org. Accessed May 24, 2023. https://lmsys.org/blog/ 2023-05-10-leaderboard/
45. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. *N Engl J Med*. 2023;388:1233–1239.
46. How can I access GPT-4? OpenAI. Accessed April 12, 2023. https://help.openai.com/en/articles/7102672-how-can-i-access-gpt-4
47. Densen P. Challenges and opportunities facing medical education. *Trans Am Clin Climatol Assoc*. 2011;122:48–58.