Contents lists available at ScienceDirect

Academic Pathology 11/1 (2024) 100101

Academic Pathology

journal homepage: www.journals.elsevier.com/academic-pathology

Review Article

Establishing priorities for implementation of large language models in pathology and laboratory medicine

Simone Arvisais-Anhalt, MD [a], Steven L. Gonias, MD, PhD [b,*], Sara G. Murray, MD, MAS [c,**]

[a] Department of Laboratory Medicine, University of California San Francisco, San Francisco, CA, USA
[b] Department of Pathology, University of California San Diego, La Jolla, CA, USA
[c] Department of Medicine, University of California San Francisco, San Francisco, CA, USA

A B S T R A C T

Artificial intelligence and machine learning have numerous applications in pathology and laboratory medicine. The release of ChatGPT prompted speculation regarding the potentially transformative role of large-language models (LLMs) in academic pathology, laboratory medicine, and pathology education. Because of the potential to improve LLMs over the upcoming years, pathology and laboratory medicine clinicians are encouraged to embrace this technology, identify pathways by which LLMs may support our missions in education, clinical practice, and research, participate in the refinement of AI modalities, and design user-friendly interfaces that integrate these tools into our most important workflows. Challenges regarding the use of LLMs, which have already received considerable attention in a general sense, are also reviewed herein within the context of the pathology field and are important to consider as LLM applications are identified and operationalized.

Keywords: Artificial intelligence, GPT, Large language models (LLMs)

Introduction

With the release of ChatGPT (OpenAI, San Francisco, CA, https://openai.com/), Artificial Intelligence (AI) and Large Language Models (LLMs) have emerged as potentially transformative tools for the economically efficient delivery of clinical care. Although "hope" has been blended with "hype," extensive development of LLMs in healthcare is already underway. High impact applications of LLMs include approaches that may mitigate time-consuming tasks for patient-facing clinicians, including note writing and responses to patient messages.[1] In Pathology, AI has received attention because it may improve the speed and accuracy of diagnosis and allow for more helpful prognostication.[2] In order for academic pathologists to most fully benefit from new technologies like generative AI, we must envision the entire scope of its possible applications for each of our major missions, including education, clinical practice, and research.

GPT stands for "Generative Pre-trained Transformer" and refers to a family of LLMs, developed by OpenAI, in which deep neural networks are used to generate text in response to prompts. While GPT may have been the first LLM to gain international attention, as the model that fueled the popular ChatGPT interface, it is part of a larger trend in many technology companies in which generative AI is being used to create content. LLMs are based on the transformer architecture,[3] allowing them to process text input in parallel and model relationships between words. LLMs such as GPT are pre-trained on large amounts of textual data, such as web pages or books, and then fine-tuned on specific tasks such as question answering, language translation, and text classification. This makes them particularly effective at generating fluent and persuasive text in response to prompts. A modified version GPT powers ChatGPT, which has been available to the public as an online tool since late 2022, with subsequent improvements including GPT-4.

ChatGPT has emerged as a disruptive force, illuminating the potential for generative AI to facilitate radical changes in healthcare and science. In clinical medicine, the widely used electronic health record vendor Epic has already released GPT functionality for testing in clinical care, such as drafted responses to patient messages. In research, these tools are capable of contributing significantly to scientific writing and journals have established editorial policies regarding the use of LLMs. The *Science* family of journals prohibits the use of AI without explicit permission, while *Nature* and the Springer Nature family of journals developed a more moderate approach, requiring disclosure of LLM utilization in the Methods, Acknowledgements, or in the Introduction of an article.[4,5] Given the strong performance of ChatGPT in writing introductions and background sections for scientific articles and grant applications,[6,7] it is reasonable to assume that LLMs will be utilized liberally by the scientific and medical communities, perhaps beyond what is formally disclosed in articles and grant applications.

* Corresponding author. Department of Pathology, UCSD School of Medicine, 9500 Gilman Drive, La Jolla, CA, 92093, USA.
** Corresponding author. Department of Medicine, University of California, San Francisco, 521 Parnassus Avenue, San Francisco, CA, 94117, USA.
E-mail addresses: sgonias@health.ucsd.edu (S.L. Gonias), sara.murray@ucsf.edu (S.G. Murray).

https://doi.org/10.1016/j.acpath.2023.100101
Received 14 March 2023; Received in revised form 2 October 2023; Accepted 29 October 2023, Available online xxxx
eISSN2374-2895/

Within a few months after GPT-3.5 was made available to the public through ChatGPT, Microsoft released a more capable version of the LLM incorporated into its search engine, Bing, providing evidence that the function of internet search engines is undergoing rapid transformation. We are currently seeing rapid evolution of generative AI offerings that will be used in healthcare, with cloud computing platforms such as Azure (https://azure.microsoft.com/en-us/products/cognitive-services/openai-service/) now offering HIPAA-compliant use of GPT and Amazon Web Services (AWS) offering competing LLMs through Amazon Bedrock (https://aws.amazon.com/bedrock/). These tools will allow healthcare institutions to experiment with and develop LLM solutions with the proper security in place to ensure patient privacy. Thus, in the near future, the current functionality of ChatGPT will most likely appear limited compared with what will be available.

Table 1 provides a theoretical list of common activities in pathology and laboratory medicine that may be assisted or improved by LLMs. The elements in Table 1 were generated by ChatGPT in February 2023, in response to the question, "What are the applications of large language modules in pathology and laboratory medicine?"[8] While some brainstorming on the part of academic pathologists would likely augment this table with many additional excellent use cases, the LLM produced a reasonable first draft. Will AI modalities such as ChatGPT improve the effectiveness and efficiency of the practice of academic pathology and laboratory medicine?

The applications of AI in pathology were reviewed previously in an article in this Journal.[9] The previous article discussed the complex science of AI, different types of machine learning, and how practicing pathologists may begin to embrace this important field. Herein, we take a different approach, stimulated by the extremely user-friendly nature of LLM applications like ChatGPT. While there is a distinction between becoming a user of a novel technology and using the technology well, becoming a user of this AI application requires essentially no training and no AI expertise. Thus, experts in pathology and laboratory medicine are in a position to influence development of LLMs and their applications so they best serve the needs of our profession and healthcare globally. In this article, we select for discussion three examples relevant to the practice of academic pathology and laboratory medicine, which we view as "low hanging fruit" for potential applications of LLMs in our field.

LLMs may be powerful assets in pathology education

For academic pathologists, the availability of easy-to-use AI tools to assist in the development of presentations and lectures may represent an outstanding step forward. Recent reviews have addressed applications of AI in medical education.[10,11] The novel aspect of LLM applications like ChatGPT in pathology education is their ease of utilization. As an exercise, we will focus on the challenge presented to a first-year assistant professor in a pathology and laboratory medicine department who is asked to contribute to the department's education mission. Because our theoretical faculty member is a hematopathologist, the topic will be anemia. To address his/her challenge, the assistant professor opens ChatGPT and enters the query: "Provide an outline for a medical school lecture on causes of anemia." The response is shown in Table 2.[12]

There are multiple advantages to the output, when compared with an outline generated by a human, independently of the LLM. First, the required outline is completed rapidly. The output provides a starting point upon which the faculty member may build and improve. The lecture may include topics that would not have been considered if the faculty member developed the lecture without collaboration from the LLM. Finally, and perhaps most importantly, human oversight and management are "built in" when LLMs are used in medical lecture outline development because the LLM is merely creating a draft and any inaccuracies may be corrected by an attentive lecturer as he/she expands the content. The importance of human review must be stressed because LLM applications like ChatGPT may be prone to "hallucination," a phenomenon in which the LLM includes factually inaccurate information in the generated output.[13] Human review mitigates this problem before information is shared with less experienced learners.

LLMs may provide helpful annotation to laboratory and pathology reports in patient portals such as MyChart

The Information Blocking Provision of the 21st Century Cures Act (Cures Acts), which mandated nearly immediate release of electronic health information (EHI), including hospital laboratory test results, through means such as patient portals, has resulted in a major shift in the traditional healthcare delivery paradigm.[14] While the healthcare community is largely in favor of releasing laboratory results to patients and understands the benefits of patients having easy access to and being in control of their health information,[15] notable concerns have been raised.[16–18] An important concern is the anxiety that a patient may experience when receiving hospital laboratory results that have not been pre-screened and interpreted by a patient-facing physician. Use of an LLM application represents a potentially important strategy for providing helpful annotation to abnormal laboratory results in MyChart and other patient portals so that the patient has a better understanding of the information delivered.

The importance of annotation of results in patient portals may indeed increase with widespread automatic release of anatomic pathology reports. From the patient's perspective, pathology reports may be even more difficult to understand than abnormal laboratory test results. Field-specific jargon is essential in final pathology reports to best inform diagnosis and treatment plans; however, the same language may be incomprehensible or even misleading to patients who are not familiar with a particular disease entity. To test whether LLMs may facilitate translation of anatomic pathology reports in patient access portals, we uploaded a hypothetical hematopathology report into ChatGPT-4 and asked the LLM to provide a summary and patient education materials at a 6th grade reading level (Table 3).[19] The presented report is an analysis of bone marrow, which shows evidence of a plasma cell neoplasm, which may be complicated by overlying reactive polytypic bone marrow plasmacytosis. Pathologists understand how the detailed comments in the

Table 1
Potential applications of LLMs in Pathology and laboratory medicine[a].

1. Medical report generation: Large language models can be used to generate medical reports for pathology and laboratory medicine. These reports can be generated automatically by the model, saving time and reducing the workload for pathologists and laboratory professionals.
2. Data analysis: Large language models can be used to analyze large volumes of clinical data, such as electronic health records, to identify patterns and trends that can help with disease diagnosis, treatment planning, and outcomes analysis.
3. Clinical decision support: Large language models can provide decision support to clinicians, helping them to make more accurate and informed decisions about patient care. For example, a large language model can analyze patient symptoms and medical history to suggest potential diagnoses or treatment options.
4. Patient education: Large language models can be used to develop chatbots or virtual assistants that can assist patients in getting information about their medical condition, treatment options, and post-treatment care.
5. Quality control: Large language models can be used to monitor laboratory testing processes and identify potential sources of error or quality issues. This can help improve the accuracy and reliability of laboratory results.
6. Medical literature analysis: Large language models can be used to analyze medical literature and extract relevant information for clinical decision-making. This can help medical professionals stay up-to-date with the latest research and treatment options.

[a] Output provided by ChatGPT.[7]

Table 2
ChatGPT outlines a medical school lecture on the causes of anemia[a].

1. Introduction: A. Definition of anemia B. Importance of understanding the causes of anemia C. Overview of the lecture
2. Overview of anemia: A. Red blood cells and their role B. Hemoglobin and its importance C. Symptoms and signs of anemia
3. Classification of anemia: A. Microcytic hypochromic anemia B. Normocytic normochromic anemia C. Macrocytic anemia
4. Causes of microcytic hypochromic anemia: A. Iron deficiency anemia B. Anemia of chronic disease C. Thalassemia D. Sideroblastic anemia
5. Causes of normocytic normochromic anemia: A. Chronic kidney disease B. Hemolysis C. Blood loss D. Bone marrow failure E. Hemoglobinopathies
6. Causes of macrocytic anemia: A. Folate deficiency B. Vitamin B12 deficiency C. Myelodysplastic syndromes D. Alcoholism
7. Conclusion: A. Summary of the lecture B. Importance of early diagnosis and management C. Future directions for research

This lecture can be supplemented with case studies and visual aids such as diagrams, images and slides to further illustrate the various causes of anemia.

[a] Output provided by ChatGPT.[12]

Table 3
ChatGPT assists in interpretation of a complex pathology report by providing an analysis at a 6th grade reading level.[19]

Hypothetical report - FINAL DIAGNOSIS: Peripheral blood smear and bone marrow, right posterior iliac crest, aspirate smears, clot, and core biopsy: Likely plasma cell neoplasm, ~15–25 % plasma cells of total marrow cellularity; progressive trilineage hematopoiesis, dysmegakaryopoiesis, blasts <5 %; see comment.

Comment: The bone marrow aspirate demonstrates preserved progressive trilineage hematopoiesis, scattered atypical megakaryocytes, and increased plasma cells accounting for ~20 % of the intact nucleated cells by manual aspirate count. The core biopsy is normocellular (40–50 %) but contains focal loose aggregates of atypical plasma cells highlighted by CD138 immunostain (~15–25 %) without definitive evidence of light chain restriction by in situ hybridization or immunostain. Flow cytometry shows an abnormal plasma cell population with aberrant co-expression of CD56 and CD117 in only a subset of the plasma cells. It is not possible by staining to determine whether these cells express kappa, lambda, or a mixture of both. Although the overall morphologic and immunophenotypic findings suggest an underlying plasma cell neoplasm, the process may be partially masked by reactive bone marrow plasmacytosis, or, possibly two clonal populations (biclonal gammopathy). Serum protein electrophoresis is recommended. Clinical correlation and correlation with pending cytogenetics studies also is recommended.

Output From GPT4:

Diagnosis: The doctors found that there is a good chance you have a disease called a plasma cell neoplasm. This is a type of illness where a specific kind of cell in your bone marrow, called a plasma cell, grows more than it should. The doctors found about 15–25 % of the cells in your bone marrow are these plasma cells.

They also noticed some other changes in your bone marrow, like different growth patterns of cells that make blood and platelets. However, they didn't find a large number of very young, developing cells, called blasts, which can sometimes point to certain types of blood cancers.

The doctors used special tests (like staining and flow cytometry) to learn more about the plasma cells in your bone marrow. Some of the plasma cells were behaving unusually and expressed certain markers, but the tests could not confirm a few important details.

Because of this, the doctors recommend more tests. They suggest a test to check the proteins in your blood and another one to look at the genetic makeup of your cells.

Patient education materials:

It can be helpful to know a little more about what's happening inside your body. Your bone marrow, which is found inside your bones, is like a factory for your blood cells. This includes red blood cells (that carry oxygen), white blood cells (that fight infections), and platelets (that help blood clot).

In your case, there might be too many plasma cells. Plasma cells are a type of white blood cell that makes antibodies to help your body fight off infections. But if they grow too much, they can cause problems. This condition is known as a plasma cell neoplasm, and the most common type is called multiple myeloma.

There are many different tests that doctors can use to understand what's happening in your bone marrow. These include special stains, flow cytometry, genetic tests, and blood protein tests. Each test gives a little more information, so the doctors can put together the whole picture and find the best way to help you.

It's important to remember that this report is just one step in figuring out what's going on. Your doctors are working hard to get a clear understanding so they can give you the best care. If you have questions or concerns, don't hesitate to talk to your healthcare team. They're there to help!

bone marrow analysis report may be used together with other test results to constitute the required diagnostic work-up for this patient.[20,21] The interpretation of the report by ChatGPT is noteworthy in that it filters details of importance to the practicing pathologist and oncologist but still provides the patient with essential information he/she needs.

Of course, integrating AI tools into clinical practice requires substantial thought and is not without significant risk. Extensive fine-tuning and validation are necessary to ensure that LLMs deliver appropriate patient summaries through patient access portals such as Mychart. At least initially, LLM-derived summaries will need to be reviewed by a clinician prior to release to patients. Thus, choosing the highest-value use cases for LLM-derived summaries and inserting the LLM tool into the correct place in the workflow is critical to ensure that clinician-review is a seamless process. The most appropriate time for physician review of LLM-generated statements may be when the clinician is reviewing the initial report, prior to signing it for release, but institution-specific workflows and governance will need to determine the appropriate place for this review to occur. Even though physician review may result in minor delays, LLM-generated statements for patient portals still may be very helpful to the patient and an excellent way to save physician time. The highest-value use cases for LLMs in clinical care will be determined as we start testing solutions and studying their impacts.

LLMs may improve efficiency in hospital laboratory and pathology workflows

Perhaps one of the most exciting opportunities associated with LLMs in medicine is improvement of workflows. Pathologists, like other physicians, typically choose their specialty area in order to positively impact patient health. A frustrating realization involves the percent effort that must be dedicated to activities that are only indirectly related to patient care. Much attention has been paid to this problem in the recent literature and its relation to clinician burnout. The clinician burnout literature focuses mainly on patient-facing clinicians, the effects of the tools they use, such as the electronic health record (EHR), and administrative activities, such as acquiring prior authorizations.[22,23] The amount of time patient-facing physicians spend with patients compared with other activities has been directly examined and is concerning.[24] Other studies have focused on the amount of time patient-facing clinicians spend on case documentation after hours.[25] Factors that influence the well-being of pathologists and laboratory medicine clinicians have received less attention. A logical assumption related to findings from the clinician burnout literature would be that activities such as case report documentation and interactions with the EHR and Laboratory Information Systems may negatively impact pathologist well-being. Any opportunity that allows pathologists to spend more time formulating diagnoses and less time doing chart review or documenting their observations and conclusions would improve efficiency and, in our opinion, pathologist well-being.

A significant opportunity for LLM utilization may be to simplify workflows on the autopsy service. Leveraging the ability of LLMs to provide excellent summaries of abundant and complex data, LLMs may be used to rapidly review and summarize lengthy patient charts prior to performing autopsies. These summaries would augment other activities applied by the pathologist to focus the autopsy examination so that the results are as useful as possible. However, once again, before autopsy pathologists can develop a level of trust regarding clinical history summaries provided by LLMs and thus, decrease personal effort dedicated to

direct chart review, extensive validation and LLM refinement will be necessary. Comparing clinical history summaries with those generated by pathology residents or by autopsy pathology attendings may constitute an excellent and informative clinical trial.

The ability of the LLM to abstract and reformat text into any desired template may be deployed in concert with real time transcription during the gross examination and during histology slide reviews. Similar applications for clinical care are already on the market from companies like Nuance (https://www.nuance.com/index.html) and Ambience (https://www.ambiencehealthcare.com/), but are not yet developed for pathologists. The LLM may also be used to track elements of the case from the initial chart review through basic histology, immunohistochemistry, and other ancillary tests, which may be indicated. For example, when working with an LLM, as you feed it text through dictation or writing you might also ask specific summarizing questions of the content you have provided to it such as, "What elements of this autopsy template am I missing?" or, "What stains I have ordered?" In this manner, an LLM application would function as a "partner" for the autopsy pathologist.

Ensuring safe and ethical use of LLMs in pathology

When LLMs present non-factual information in a convincing manner, the process is typically referred to as "hallucination." It is generally assumed that this problem will be attenuated as LLMs become more sophisticated. While GPT-4 makes mistakes, it is also capable of catching them, either with the assistance of a human or another LLM.[1] For now, however, human oversight will remain critical as we refine these tools to transform our daily work.[26]

Risks imposed by LLMs extend beyond the phenomenon of hallucination and include many of the same risks seen in non-generative AI. Because ChatGPT was trained using the internet, biases that are present on the internet may be translated into its function and output. For example, it has been established in the literature that LLMs may suffer from biases related to gender and race.[27] In addition, biases across machine learning modalities may be related to missing data, lack of sample size, and bias in the underlying data, all of which contribute to flaws in the output of the algorithm.[28] Bias and errors are magnified in LLMs if the date of training is remote, resulting in outdated responses. Tools such as ChatGPT do not currently provide references for its output, although future versions may and this is an active area of research in the LLM community.[29]

Caution regarding biases in LLM output is equally important in scientific studies and in healthcare. The FDA regulatory landscape is continually evolving as it and other regulatory bodies aim to ensure that AI tools are trustworthy in clinical medicine. At the time this article was written, the FDA has approved six AI/ML-enabled medical devices in clinical chemistry, fifteen in hematology, six in microbiology, and four in anatomic pathology.[30] A more comprehensive approach to ensuring we develop ethical and robust tools will be needed as generative AI enters clinical practice in fields such as pathology and laboratory medicine. Likewise, as we develop generative AI tools for use in pathology and laboratory medicine, we will need to pay close attention to the evolving regulatory landscape to ensure we are in compliance. Thus, we think it is important for thought leaders in pathology and laboratory medicine to acquire an understanding of LLMs, their potential applications in academic pathology, and their limitations. This will allow the field to determine priorities for improving and deploying these AI modalities and advance our missions in clinical practice, research, and education.

Conclusions

LLMs are poised to revolutionize how text is processed across many industries. Pathology and laboratory medicine are fields that may be significantly enhanced by LLM applications. While numerous risks have been identified regarding the application of LLMs to the practice of medicine, these problems can be attenuated through additional LLM training, oversight by trained clinicians, and robust validation and quality monitoring. Given our substantial experience in validation and quality monitoring of diagnostic testing, we are well poised to be early adopters of these novel tools and help develop the standards for safe use of generative AI. Academic pathologists, pathology departments, and pathology professional societies are encouraged to embrace this technology now to guide its development and safe and ethical deployment so we may realize its full potential and improve our function as academic pathologists.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Funding

This article was written without extramural support of any kind.

References

1. 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. doi:10.1056/NEJMsr2214184.
2. Försch S, Klauschen F, Hufnagl P, Roth W. Artificial intelligence in pathology. *Dtsch Arztebl Int*. 2021;118(12):199–204. doi:10.3238/arztebl.m2021.0011.
3. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *ArXiv*. 2017. doi:10.48550/arXiv.1706.03762.
4. Thorp HH. ChatGPT is fun, but not an author. *Science*. 2023;379:313. doi:10.1126/science.adg7879.
5. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature*. 2023;613(7945):612. doi:10.1038/d41586-023-00191-1.
6. Stokel-Walker C, Van Noorden R. What ChatGPT and generative AI mean for science. *Nature*. 2023;614:214–216. doi:10.1038/d41586-023-00340-6.
7. van Dis EAM, Bollen J, Zuidema W. , van Rooij R, and Bockting CL. ChatGPT: five priorities for research. *Nature*. 2023;614:224–226. doi:10.1038/d41586-023-00288-7.
8. Text generated by ChatGPT in response to "[What are the applications of large language modules in pathology and laboratory medicine?]". [February, 2023]. OpenAI.
9. Rashidi HH, Tran NK, Betts EV, Howell LP, Green R. Artificial intelligence and machine learning in pathology: the present landscape of supervised methods. *Acad Pathol*. 2019;6, 2374289519873088. doi:10.1177/2374289519873088.
10. Wartman SA, Combs CD. Medical education must move from the information age to the age of artificial intelligence. *Acad Med*. 2018;93:1107–1109. doi:10.1097/ACM.0000000000002044.
11. Rakha EA, Toss M, Shiino S, et al. Current and future applications of artificial intelligence in pathology: a clinical perspective. *J Clin Pathol*. 2021;74:409–414. doi:10.1136/jclinpath-2020-206908.
12. Text generated by ChatGPT in response to "[Provide an outline for a medical school lecture on causes of anemia]". [February, 2023]. OpenAI.
13. Ali R, Tang OY, Connolly ID, et al. Performance of ChatGPT, GPT-4, and google bard on a neurosurgery oral boards preparation question bank. *Neurosurgery*. 2023. doi:10.1227/neu.0000000000002551 (online ahead of print).
14. Federal Register Volume 85, Issue 85 (May 1, 2020). edited by Services DoHaH Office of the Federal Register, National Archives and Records Administration, p. 25642-25961.
15. Arvisais-Anhalt S, Ratanawongsa N, Sadasivaiah S. Laboratory results release to patients under the 21st Century Cures Act: the eight stakeholders who should care. *Appl Clin Inf*. 2023;14(1):45–53. doi:10.1055/a-1990-5157.
16. Arvisais-Anhalt S, Lau M, Lehmann CU, et al. The 21st Century Cures Act and multiuser electronic health record access: potential pitfalls of information release. *J Med Internet Res*. 2022;24(2), e34085. doi:10.2196/34085.
17. Gerber DE. 21st Century Cures Act: implementation without understanding implication? *JCO Oncol Pract*. 2022;18(2):85–87. doi:10.1200/OP.21.00436.
18. Winget M, Haji-Sheikhi F, Brown-Johnson C, et al. Electronic release of pathology and radiology results to patients: opinions and experiences of oncologists. *J Oncol Pract*. 2016;12(8):e792–e799. doi:10.1200/JOP.2016.011098.
19. Text generated by ChatGPT in response to a hypothetical hematopathology case report [Please summarize the following bone marrow aspirate report at the 6th grade reading level and include patient education materials] [July, 2023]. OpenAI.
20. Cowan AJ, Green DJ, Kwok M, et al. Diagnosis and management of multiple myeloma: a review. *JAMA*. 2022;327:464–477. doi:10.1001/jama.2022.0003.
21. Robillard N, Wuilleme S, Moreau P, Bene MC. Immunophenotype of normal and myelomatous plasma-cell subsets. *Front Immunol*. 2014;5:137. doi:10.3389/fimmu.2014.00137.

22. Kroth P, Morioka-Douglas N, Veres S, et al. Association of Electronic Health Record design and use factors with clinician stress and burnout. *JAMA Netw Open*. 2019;2, e199609. doi:10.1001/jamanetworkopen.
23. West CP, Dyrbye LN, Shanafelt TD. Physician burnout: contributors, consequences and solutions. *J Intern Med*. 2018;283(6):516–529. doi:10.1111/joim.12752.
24. Toscano F, O'Donnell E, Broderick JE, et al. How physicians spend their work time: an ecological momentary assessment. *J Gen Intern Med*. 2020;35(11):3166–3172. doi: 10.1007/s11606-020-06087-4.
25. Apathy NC, Rotenstein L, Bates DW, Holmgren AJ. Documentation dynamics: note composition, burden, and physician efficiency. *Health Serv Res*. 2023;58(3):674–685. doi:10.1111/1475-6773.14097.
26. Harrer S. Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine. *Lancet*. 2023;90:104512. doi:10.1016/j.ebiom.2023.104512.
27. Liang PP, Wu C, Morency L-P, Salakhutdinov R. Towards understanding and mitigating social biases in Language Models. *ArXiv*. 2021. doi:10.48550/arXiv.2106.13219.
28. Gianfrancesco MA, Tamang S, Yazdany J, et al. Potential biases in machine learning algorithms using electronic health record data. *JAMA Intern Med*. 2018;178(11): 1544–1547. doi:10.1001/jamainternmed.2018.3763.
29. Gao T, Yen H, Yu J, Chen D. Enabling Large Language Models to generate text with citations. *ArXiv*. 2023. doi:10.48550/arXiv.2305.14627.
30. Artificial intellence and machine learning (AI/ML) – enabled medical devices. U.S. FOOD & DRUG ADMINISTRATION. 10/05/2022. Accessed September 29, 2023. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices.