




Commentary

ChatGPT: Increasing accessibility for natural language processing in healthcare quality measurement

Julie Tsu-Yu Wu MD, PhD^{1,2}, Erica S. Shenoy MD, PhD^{3,4} , Evan P. Carey PhD^{5,6} , Gil Alterovitz PhD^{4,5},
Michael J. Kim MD^{5,7} and Westyn Branch-Elliman MD, MMSc^{4,5,8} 

¹Department of Medicine, Palo Alto VA Healthcare System, Palo Alto, California, ²Stanford University School of Medicine, Palo Alto, California, ³Massachusetts General Hospital and Mass General Brigham Boston, Massachusetts, ⁴Harvard Medical School, Boston, Massachusetts, ⁵Veterans' Affairs National Artificial Intelligence Institute, Washington, DC, ⁶Department of Biostatistics and Informatics, University of Colorado School of Public Health, Denver, Colorado, ⁷Irvine School of Medicine, University of California–Irvine, Irvine California and ⁸Section of Infectious Diseases, Department of Medicine, VA Boston Healthcare System, West Roxbury, Massachusetts

Artificial intelligence is being increasingly looked to for its potential to automate healthcare quality measurement, such as healthcare-associated infections (HAIs). However, a key limitation in automation has been the lack of universal and reliable algorithms for the measurement of nonautomatable, unstructured data elements required to identify HAIs based on surveillance definitions, such as clinical symptoms. The introduction of generative pretrained transformer chat-bots (eg, ChatGPT) has the potential to advance HAI surveillance due to its facile language-processing capabilities.

In this issue of *Infection Control and Healthcare Epidemiology*, Perret and Schmidt explore the potential for ChatGPT to support HAI surveillance activities for facilities with limited information technology (IT) resources to support automated detection.¹ Although standardized definitions exist for HAIs, local data collection and recording practices vary, leading to differences in how these definitions are interpreted. As a future application, Perret and Schmidt's report raises questions about how this technological advancement could be leveraged to expand infection surveillance algorithms. This expansion would involve adapting the tool to local practices with minimal intervention, promoting accessibility and consistency in application of definitions for HAI surveillance.

What is ChatGPT and what makes it revolutionary?

ChatGPT represents a paradigm shift for artificial intelligence in its generative capacity. The platform's ability to generate high-quality responses to human input has generated substantial interest. This generative component implies both an ability to understand a wide variety of natural language and the ability to follow human-generated prompt instructions correctly. In its capacity to parse human language, ChatGPT represents a refinement of prior large language models (LLMs).² LLMs are deep-learning models that are pretrained on text data available from billions of books, articles, and conversations across the internet. To fine tune the model for specific tasks, the model can leverage the knowledge obtained in

pretraining to reduce the number of human-provided labels required to achieve task-specific accuracy. This process reduces the amount of time necessary for manual chart review—a time-consuming activity in infection prevention control (IPC).

Although LLM technology is not itself new, ChatGPT made LLMs accessible to a larger audience. ChatGPT is one of the largest LLMs (including >175 billion parameters) and was pretrained on a diverse training corpus.² It was specifically trained for use in conversational tasks, which, combined with an easy-to-use, web-based interface enables humans and LLMs to engage through the familiar medium of human language. This low barrier to use enabled those who were previously unfamiliar with programming to interact with and use ChatGPT. The diversity of ChatGPT training material enables it to offer insights on a wide variety of topics, and its conversational ability enables it to fine tune those answers in response to human-generated prompts. This ability is enabling subject-matter experts, such as clinicians, to use LLMs for their needs with minimal programming or technical expertise.³

ChatGPT: The next revolution in HAI surveillance?

HAI surveillance represents a key opportunity for revolution by ChatGPT. HAI are preventable, costly, and linked to hospital reimbursement through the CMS Hospital-Acquired Condition Reduction Program. Surveillance and reporting are a major focus of IPC programs.⁴ As discussed in detail by Shenoy et al,⁵ some HAI surveillance, such as that for ventilator-associated events, is almost entirely automatable using structured data elements that do not require advanced AI-based solutions. Other types of surveillance, such as that for catheter-associated urinary tract infections (CAUTIs), are “mostly” automatable, meaning that some elements of the definition can be programmed electronically but others, particularly clinical symptoms, require some element of human review for measurement.

Attempts to use simple, natural-language processing (NLP)-based strategies for CAUTI surveillance have failed, mostly due to the complexity of documentation practice and the need for more advanced informatics to “read” the clinical notes.⁶ Earlier NLP algorithms relied heavily on task-specific training data, making them sensitive to local documentation practices with limited potential for scale-up and spread.⁶ In contrast, ChatGPT's

Corresponding author: Westyn Branch-Elliman; Email: Westyn.Branch-Elliman@va.gov

Cite this article: Wu JT-Y, Shenoy ES, Carey EP, Alterovitz G, Kim MJ, Branch-Elliman W. ChatGPT: Increasing accessibility for natural language processing in healthcare quality measurement. *Infect Control Hosp Epidemiol* 2024. 45: 9–10, doi: 10.1017/ice.2023.236



extensive pretraining data may reduce the hyperlocal adaptation that has hampered prior efforts.

To demonstrate feasibility of ChatGPT as a tool to reduce chart review burden, Perret and Schmid¹ applied ChatGPT to automated detection of CAUTI to demonstrate how it may facilitate detection without requiring extensive training data or programming expertise. They trained a ChatGPT model on a relatively small cohort of ‘synthetic’ patient data (data that are not derived from actual patients) to identify CAUTI cases confirmed by physician review. They tested their approach on 2 data sets: one with structured elements easily extractable without NLP, and another extended data set that included structured data and clinical symptoms typically found in unstructured physician notes. Following 18 rounds of training with 2–25 queries each, ChatGPT performed effectively on both data sets, achieving a sensitivity of 91%, specificity of 95%, positive predictive value of 83%, and negative predictive value of 97% for the extended data set. Although other CAUTI surveillance algorithms might achieve similar performance, this valuable proof-of-concept study shows that IPC can apply ChatGPT in a healthcare setting without requiring specialized technical expertise or IT resources.

Moving from the theoretical to the real: Policy and regulatory challenges

Perret and Schmid’s article¹ serves as a valuable case study demonstrating early use of ChatGPT for surveillance.¹ However, it also underscores the existing limitations and prerequisites that must be addressed before deployment. Notably, ensuring access to reliable and pertinent training data emerges as a key challenge. One cannot use real, protected patient data for ChatGPT’s training because it might expose sensitive information. To circumvent this issue, Perret and Schmid¹ employ fabricated patient data processed into an Excel spreadsheet (Microsoft, Redmond, WA), which differs from real-world clinical notes. Clinical notes can pose challenges, such as historical data copy forward, inconsistencies within the same note, variable spellings and abbreviations, among other real-world implementation barriers. Thus, how their model performance will translate to actual clinical notes remains unknown; theoretically, however, ChatGPT should be able to learn how to read clinical documentation despite these challenges with real-world documentation. However, beyond concerns about learning how to identify the true signal from the noise in the electronic health record, implementing this technology will necessitate strategies for secure data management and local deployment of ChatGPT within a HIPAA-protected environment. Numerous LLMs akin to ChatGPT that can be deployed locally have been released; however, the optimal approach remains uncertain.²

Beyond the complexities of accessing pertinent real-world data, policy challenges are also an important consideration. With the introduction of LLMs such as ChatGPT in healthcare, legal and regulatory frameworks must adapt. LLMs are generative and thus prone to hallucination, defined as creating answers that sound credible but do not have a reliable basis. This situation creates new liability challenges and need for quality control. Nevertheless, with its extensive user base exceeding 100 million and growing, clinical providers are likely already utilizing ChatGPT for tasks, even without awaiting regulatory guidance.⁷ However, users should be aware of potential risks and develop plans for addressing them.

ChatGPT and HAI surveillance: What next?

Despite the caveats listed above, the availability of LLMs that can read (or at the very least, screen) clinical notes represents a potential revolution in HAI surveillance and day-to-day IPC practice. Not only could such technology reduce the human resources required to conduct HAI surveillance, but ChatGPT’s reduced need for location-specific training data also allows broader healthcare applications, potentially standardizing surveillance practices and workflows across facilities and improving interfacility comparison. ChatGPT’s broad availability benefits resource-constrained systems, enabling quality improvement even in less-resourced settings.

The use of ChatGPT to support surveillance activities is attractive because it may free up IPC resources to work on other important HAI prevention activities. However, despite the promise of reducing IPC resources spent reviewing charts, the introduction of advanced technologies like ChatGPT may shift workload without reducing it. Upfront planning for algorithm maintenance and review needs to be built into any system. Artificial intelligence systems will require ongoing maintenance, quality assurance, and re-evaluation to ensure that output remains accurate and trustworthy.

Are we on the undergoing a revolution in HAI surveillance? Not yet. But the research by Perret and Schmidt represents an important step in moving the conversation forward.

Acknowledgments. The views expressed are those of the authors and do not necessarily represent those of the US Department of Veterans’ Affairs or the US federal government.

Financial support. No financial support was provided relevant to this article.

Competing interests. WBE reports research funding from the VA Health Services Research and Development Service. All other authors report no conflicts of interest relevant to this article.

References

- Perret J, Schmid A. Application of OpenAI GPT-4 for the retrospective detection of catheter-associated urinary tract infections in a fictitious and curated patient data set. *Infect Control Hosp Epidemiol* 2023;44:xxx–xxx. doi: [10.1017/ice.2023.189](https://doi.org/10.1017/ice.2023.189)
- Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med* 2023;29:1930–1940.
- Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front Artif Intell* 2023;6:1169595.
- Hospital-acquired conditions. Centers for Medicare and Medicaid Services website. <https://www.cms.gov/medicare/payment/fee-for-service-providers/hospital-aquired-conditions-hac/hospital-aquired-conditions>. Updated September 6, 2023. Accessed October 16, 2023.
- Shenoy ES, Branch-Elliman W. Automating surveillance for healthcare-associated infections: rationale and current realities (part I/III). *Antimicrob Steward Healthc Epidemiol* 2023;3:e25.
- Branch-Elliman W, Strymish J, Kudesia V, Rosen AK, Gupta K. Natural language processing for real-time catheter-associated urinary tract infection surveillance: results of a pilot implementation trial. *Infect Control Hosp Epidemiol* 2015;36:1004–1010.
- Hu K, Hu K. ChatGPT sets record for fastest-growing user base—analyst note. Reuters website. <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>. Published February 2, 2023. Accessed September 5, 2023.