EDITORIALS

Check for updates

The Artificially Intelligent Teacher: Applying Natural Language Processing to Critical Care Education

Varuna Prakash, M.D., M.H.Sc.[1], Dominique Piquette, M.D., Ph.D.[1,2], and Andre Carlos Kajdacsy-Balla Amaral, M.D.[1,2]

[1]Interdepartmental Division of Critical Care Medicine, University of Toronto, Toronto, Ontario, Canada and [2]Department of Critical Care Medicine, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada

The intensive care unit (ICU) can be a difficult environment for medical education. The work is fast-paced, the workload is high, patients' courses are complex, and their deteriorations are rapid. Educators in this setting are tasked with balancing the often-competing goals of clinical efficiency and effective teaching (1, 2). These barriers also contribute to difficulties in assessing trainee performance. The oral case presentation, in which trainees must review a vast amount of data from a patient's history, physical exam, and investigations and then apply clinical reasoning to synthesize an appropriate plan of care, represents a valuable tool for assessing what and how our trainees are learning. Case presentations allow educators to gauge clinical understanding, teach clinical reasoning, correct misperceptions, and form impressions regarding learner entrustment (3, 4). In an ideal world, learners would receive feedback on their presentations in real-time, enabling them to monitor their own progress and motivating them toward specific outcomes (3, 5). However, this is often not possible because of time pressures in our busy postpandemic clinical environments, not to mention other barriers, such as discomfort with engaging in difficult conversations, lack of training in providing feedback, and lack of direct observation of trainees' tasks (5). Given these constraints, innovative tools to facilitate feedback for case presentations would be a welcome addition to our teaching environment. These tools should provide timely and consistent feedback without increasing the staffing workload.

In this issue of *ATS Scholar*, King and colleagues describe a novel proof-of-concept method of evaluating ICU case presentations by using natural language processing (NLP) techniques (6). Put simply, NLP is a way of using computers to extract and infer meaning from natural-language text and speech (7). Because natural language is complex, dynamic, and often ambiguous, simple

The views expressed in this article do not communicate an official position of the University of Toronto or Sunnybrook Health Sciences Centre.

ATS Scholar Vol 3, Iss 4, pp 505–508, 2022

DOI: 10.34197/ats-scholar.2022-0114ED

rule-based methods of processing it are insufficient. Rather, newer artificial intelligence and machine learning algorithms are the engines that power this type of processing by allowing computers to learn, decipher, and analyze vast quantities of data. NLP is ubiquitous in consumer life, with uses ranging from predictive text on our smartphones and emails to automated voice assistants such as Siri and Alexa. We have only begun exploring uses of NLP in healthcare; so far, its applications include predicting admissions of patients in the emergency department, identifying patients suitable for clinical trials, mining medical records to predict treatment efficacy and adverse events, and analyzing the quality and comprehensiveness of clinical documentation, to name a few (8–11).

King and colleagues applied NLP to the ICU medical education context by orally recording ICU trainees' case presentations for daily rounds and comparing them to a staff physician's recording that was used as a reference standard. They employed NLP to analyze the content (i.e., the clinical concepts discussed by the learner) and style (e.g., descriptiveness, verbosity, and density of concepts) of each oral presentation. The trainees' recording was then compared with the attending physician's for the same case. In doing so, they could identify how much trainees' case presentations diverged from the attending's in content and style. King and colleagues term this method "automated comparative assessment"; in other words, their results provide a snapshot of how far away trainee presentations are from the reference standard and in which dimensions they differ. The authors envision a future in which this method is used to provide trainees with rapid feedback about their presentations, thus providing more timely, frequent, and consistent assessments of performance than is feasible for educators alone to undertake.

King and colleagues acknowledge some limitations in their work, many of which are rich opportunities for further exploration. First, this is a proof-of-concept study that required trainees to prerecord simulated patient presentations. The transition to real-time bedside NLP, while balancing privacy, regulatory, and technological constraints, will require further work. Second, as clinical information can be communicated with infinite stylistic variations, defining a standard is somewhat elusive. Using more words or more descriptive phrasing does not necessarily translate to a better presentation, and neither does the seniority of the presenter. The authors suggest having a reference standard consisting of multiple senior physicians' presentations or a consensus standard as a solution to this challenge. We further suggest comparing oral presentations to other forms of communication (e.g., written clinical notes on the same patient) to enrich feedback quality and create a more comprehensive understanding of trainees' skills.

NLP's strengths are in reliably processing vast amounts of data without incurring the human limitations of cognitive overload, fatigue, habituation, and distractibility. In other words, the real power of NLP tools is in helping educators gather more frequent and reliable data points regarding trainee performance. NLP can therefore augment but not replace traditional forms of feedback. Any implementation of these tools must be enriched by a discussion with an educator to properly guide learners on the basis of sociocultural factors, institutional culture, and their own learning trajectories (12).

Finally, we must remember that NLP, like all tools, can be prone to unintended consequences. We must especially be wary of unintentionally perpetuating bias: when we teach our algorithms to learn, we may also subject them to the human foibles of stereotyping and propagating racist, sexist, or discriminatory ideas (13). As we devise more applications for NLP in health care, we must take care to build intentional countermeasures to mitigate these biases.

King and colleagues' proof-of-concept work inspires us to envision usages beyond oral case presentations in which NLP can be applied. Rather than creating use cases for existing technology, it is helpful to start by understanding the human factors and limitations that we face in our work environment and then ask how NLP can help us solve these problems. NLP may improve the use of electronic medical records to allow more efficient and less cumbersome documentation (14), predict and alert medical incidents (9, 15), and intelligently cue providers with useful additions to problem lists (9). For example, NLP could highlight inconsistencies within a problem plan or alert us to potential conflicts with a consultant's plan (e.g., if a specific antibiotic is being considered that may lower seizure threshold in a patient with epilepsy or if anticoagulation is recommended in a patient with recent bleeding risk). Here, too, we must proceed cautiously: like all new clinical tools, we must ensure that artificial intelligence and NLP solutions are rigorously usability-tested in the real clinical environment to avoid falling into the common pitfalls of habituation and alert fatigue.

We live in a world in which artificial intelligence and NLP are deeply integrated into our consumer lives, but we are only beginning to scratch the surface of their uses in health care. We commend King and colleagues for their foray into innovative uses of these technologies and hope their work paves the way for more potential applications in the medical education sphere and beyond.

Author disclosures are available with the text of this article at www.atsjournals.org.

REFERENCES

1. Joyce MF, Berg S, Bittner EA. Practical strategies for increasing efficiency and effectiveness in critical care education. *World J Crit Care Med* 2017;6:1–12.

2. Robillard N, Vincelette C. Barriers to workplace-based learning in the intensive care unit: a qualitative description of internal medicine residents' and intensivists' perspectives. *Med Teach* 2022;44: 629–635.

3. Melvin L, Cavalcanti RB. The oral case presentation: a key tool for assessment and teaching in competency-based medical education. *JAMA* 2016;316:2187–2188.

4. Landreville JM, Cheung WJ, Hamelin A, Frank JR. Entrustment checkpoint: clinical supervisors' perceptions of the emergency department oral case presentation. *Teach Learn Med* 2019;31:250–257.

5. Burgess A, van Diggele C, Roberts C, Mellis C. Feedback in the clinical setting. *BMC Med Educ* 2020;20:460.

6. King AJ, Kahn JM, Brant EB, Cooper GF, Mowery DL. Initial development of an automated platform for assessing trainee performance on case presentations. *ATS Scholar* 2022;3:548–560.

7. Chen M, Decary M. Artificial intelligence in healthcare: an essential guide for health leaders. *Healthc Manage Forum* 2020;33:10–18.

8. Zhang X, Kim J, Patzer RE, Pitts SR, Patzer A, Schrager JD. Prediction of emergency department hospital admission based on natural language processing and neural networks. *Methods Inf Med* 2017;56:377–389.

9. Locke S, Bashall A, Al-Adely S, Moore J, Wilson A, Kitchen GB. Natural language processing in medicine: a review. *Trends Anaesth Crit Care* 2021;38:4–9.

10. Weissman GE, Harhay MO, Lugo RM, Fuchs BD, Halpern SD, Mikkelsen ME. Natural language processing to assess documentation of features of critical illness in discharge documents of acute respiratory distress syndrome survivors. *Ann Am Thorac Soc* 2016;13:1538–1545.

11. Davoudi A, Tissot H, Doucette A, Gabriel PE, Parikh R, Mowery DL, *et al.* Using natural language processing to classify serious illness communication with oncology patients. *AMIA Annu Symp Proc* 2022;2022:168–177.

12. Ramani S, Könings KD, Ginsburg S, van der Vleuten CPM. Feedback redefined: principles and practice. *J Gen Intern Med* 2019;34:744–749.

13. Hovy D, Prabhumoye S. Five sources of bias in natural language processing. *Lang Linguist Compass* 2021;15:e12432.

14. Kaufman DR, Sheehan B, Stetson P, Bhatt AR, Field AI, Patel C, *et al.* Natural language processing-enabled and conventional data capture methods for input to electronic health records: a comparative usability study. *JMIR Med Inform* 2016;4:e35.

15. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, *et al.* A guide to deep learning in healthcare. *Nat Med* 2019;25:24–29.