Discourage Paint-by-Numbers Presentations

Aaron D. Baugh, M.D.* and Lekshmi Santhosh, M.D., M.A.Ed.

Division of Pulmonary/Critical Care Medicine, University of California San Francisco, San Francisco, California

ORCID IDs: 0000-0002-9527-691X (A.D.B.); 0000-0002-9897-3462 (L.S.)

To the Editor:

We read with interest the recent article by King and colleagues on the use of natural language processing (NLP) as an assessment tool for oral patient presentations (1). Although the use of artificial intelligence and machine learning has greatly impacted medicine, no tool is better than the application asked of it. Careful deliberation is needed to understand the potential risks to the learning environment of applying NLP to oral presentations.

Oral presentations simultaneously serve a number of rhetorical roles: establishing shared knowledge for decision making, an opportunity to practice integrative clinical reasoning, and case-based review of evidence-based medicine. The relative importance of these purposes shifts with the contextual listening audience and other context. It is unclear to us how using a small collection of model presentations could capture this dynamism. Furthermore, evidence demonstrates learners' persistent tendency to overlook these multiple objectives in pursuit of the singular goal of demonstrating mastery to their assessor (2). Would this tendency be increased or decreased by the presence of a scoring rubric that assessed only for similarity to a static, gold-standard presentation?

There is potentially broad disservice to the educational goals of patient presentation. One major weakness consistently identified in instructor feedback is a lack of contextualization that leads students to learn the wrong lessons (3). NLP-based rubrics would preserve this quality. Although artificial intelligence can high-light how the learner's work was different, it is silent on why the attending made different choices. Those struggling to improve are as adrift in deriving important principles from these differences as they are in more conventional feedback models. Through these patterns, the use of NLP-based feedback can encourage harmful cognitive schema, such as overconfidence.

NLP-based feedback would create an artificial but academically consequential benchmark: the idea that one single form of case presentation is most correct. We are concerned about the implicit impact

---

*Corresponding author (e-mail: aaron.baugh@ucsf.edu).

of such messaging. Overconfidence is already an important cause of medical errors. Qualitative studies highlight how the culture of medicine contributes to this phenomenon by promoting showmanship (4). This form of perfectionism is closely related to low uncertainty tolerance, which can inhibit skills development, is associated with reduced likelihood of working in underserved communities, and can encourage low-value care (5). Scoring for similarity to a presenter with greater medical knowledge and personal experience may suppress the opportunity to acknowledge and build tolerance for uncertainty. The endorsement of a single "perfect" presentation would seem to double down on this suite of maladaptive traits.

Far from narrowing only the conceptions of learners, this can also foreclose institutional notions of diversity and acceptability. The authors correctly mention the potential harms to learners who do not speak English as their primary language or who are from backgrounds underrepresented in medicine (1). One obvious concern is the ability of NLP to appropriately interpret a speaker's accent and cadence. The more subtle point worth emphasis is that language is an expression of culture, and each favors certain patterns in diction or grammatical construction. There is a potential for arbitrary penalization if learners are scored for deviance from the model presentation's single cultural framework because these deviances might not impact clinical care. Such a grading scheme might either encourage imposter syndrome through the negative feedback learners receive (6) or require learners to expend mental effort of mimicry that would be better applied to clinical reasoning.

What learners absorb is not limited to only those lessons we intend. Although there are many strengths in the proposed work, we should also grapple with the many unintentional risks to overconfidence, equity, and the learning process. In education, just as in practicing clinical medicine, we must above all strive to do no harm.

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

REFERENCES

1. 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.

2. Lingard L, Schryer C, Garwood K, Spafford M. 'Talking the talk': school and workplace genre tension in clerkship case presentations. *Med Educ* 2003;37:612–620.

3. Haber RJ, Lingard LA. Learning oral presentation skills: a rhetorical analysis with pedagogical and professional implications. *J Gen Intern Med* 2001;16:308–314.

4. LaDonna KA, Ginsburg S, Watling C. "Rising to the level of your incompetence": what physicians' self-assessment of their performance reveals about the imposter syndrome in medicine. *Acad Med* 2018;93:763–768.

5. Santhosh L, Chou CL, Connor DM. Diagnostic uncertainty: from education to communication. *Diagnosis (Berl)* 2019;6:121–126.

6. Lingard L, Cristancho S, Hennel EK, St-Onge C, van Braak M. When English clashes with other languages: insights and cautions from the Writer's Craft series. *Perspect Med Educ* 2021;10:347–351.