Journal of the American Medical Informatics Association, **30(7)**, 2023, 1225–1226
https://doi.org/10.1093/jamia/ocad091
Editorial

Editorial

AI in health: keeping the human in the loop

Public discourse about artificial intelligence (AI) and generative AI, in particular, is ubiquitous. AI has been a focus of research in biomedical and health informatics since its inception and in publications the *Journal of American Medical Informatics Association* since its inaugural issue that included a threaded bibliography on medical diagnostic decision support systems by Dr. Randy Miller (a future JAMIA Editor-in-Chief).[1] In that same issue and apropos of the title of this editorial, Dr. Ted Shortliffe's provocative editorial was entitled "Dehumanization of patient care—are computers the problem or the solution?"[2] This month I highlight 5 papers focused on AI that provide key lessons about the importance of keeping the human in the loop.

Lyell et al[3] examined real-world safety problems involving machine learning (ML)-enabled medical devices by analyzing safety events reported to the US Food and Drug Administration's Manufacturer and Use Facility Device Experience program. Using an existing framework for safety problems with health information technology, they identified whether a reported problem was due to the ML device or its use, and key contributors to the problem. They also classified the consequences of events. The majority of the 266 safety events were associated with ML devices that primarily used image-based data as compared to signal-based data. Ninety-three percent of problems involved the ML device with 82% related to data acquisition and <10% to algorithm errors. Sixteen percent of the events resulted in harm. Use problems (7%) were 4 times more likely than device problems to cause harm. This study highlights the need to approach ML device safety from a whole-system perspective including user interactions with devices rather than focusing only on the algorithm.

ChatGPT has stimulated much debate among the public as well as in our field of biomedical and health informatics about the use of large language models and generative AI. Liu et al[4] compared the utility of ChatGPT ($n = 37$) versus human-generated ($n = 29$) suggestions for improving 7 clinical decision support (CDS) alerts. Five clinicians rated the suggestions on usefulness, acceptance, relevance, understanding, workflow, bias, inversion, and redundancy. Nine of the 37 (24%) recommendations generated by ChatGPT were among the 20 highest rated suggestions and clinicians perceived them as offering unique perspectives. The study findings suggest that ChatGPT could complement but not replace human reviewers in optimizing the CDS alert logic.

March 11, 2023 marked the end of the federal public health emergency for COVID-19 in the United States. However, public health considerations around Long COVID remain. With the goal to de-black-box ML-based phenotype algorithms for Long COVID, the Case Study in this issue demonstrates the transferability of ML-based phenotype algorithms from one electronic health record (EHR) to another.[5] As part of the National Institutes of Health Researching COVID to Enhance Recovery (RECOVER) Initiative, researchers with the National COVID Cohort Collaborative (N3C) developed and trained a ML-based phenotype to identify patients in the N3C EHR repository who were likely to have Long COVID. Subsequently, researchers from the *All of Us* study partnered with N3C to reproduce the output of the N3C model using *All of Us* EHR repository data and N3C's open-source code. Through this process, the researchers generated key lessons to de-black-box phenotyping algorithms, prevent unnecessary work, and promote open science. While one lesson focuses on technology—leverage open-source code and a common data model (in this example the Observational Health Outcomes Partnership [OMOP] model) that is shared among all participants—the other 2 lessons emphasize the human component: convene small teams of methods and programming experts from each participating group to meet regularly during the translation and testing process; and provide well-written, stepwise instructions and document code with sufficient detail about assumptions, data cleaning steps, and derived variables.

A tutorial by Wang et al[6] provides an overview of top-down and bottom-up AI approaches for application to EHR data to assist researchers in deciding which approaches are suitable for their research. The tutorial includes a review of building block concepts including ML, classical ML algorithms, deep learning neural networks, and techniques useful in low-resource settings (eg, transfer learning). The authors categorize major AI computational methods commonly applied in the healthcare domain into top-down and bottom-up paradigms based on the data and annotation needed, training technique, and function of the resulting algorithm, and characterize the strengths and limitations of each paradigm. Lastly, they propose a decision tree to help researchers decide whether to pursue a top-down or bottom-up approach, a foundational first step in the use of AI.

van der Vegt et al[7] conducted a systematic review of studies reporting clinically deployed AI-based sepsis prediction algorithms in the adult hospital setting; this included appraisal of methodological quality, deployment and evaluation methods, and outcomes. As a strategy for validating the SALIENT implementation framework, the authors also identified contextual factors that influenced implementation of the AI-based sepsis prediction algorithms in the studies and mapped these factors to stages of the framework (I = problem definition; II = retrospective development; III =Silent trial; IV = Pilot trial; V = Large trial/Roll-out; Post=Post deployment study).

Received: 10 May 2023. **Editorial Decision:** 10 May 2023. **Accepted:** 10 May 2023

The 30 studies in the review represented implementation of 8 different AI-based sepsis prediction algorithms with 5 studies reporting significantly decreased mortality post-implementation. The mapping to the SALIENT implementation framework clarified where and when barriers ($n = 14$), enablers ($n = 26$), and key decision points ($n = 22$) arose during the implementation process. The findings suggest that the SALIENT implementation framework contributes to overcoming gaps in human guidance for implementing AI-based algorithms for sepsis and may be applicable to non-sepsis algorithms.

To advance application of AI in health that results in safe, high-quality, and equitable care and improved quality of life, it is essential to focus not only on the technical quality of the AI but also on the critical factors related to implementation and use.

Suzanne Bakken*

School of Nursing, Department of Biomedical Informatics, and Data Science Institute, Columbia University, New York, New York, USA

*Corresponding Author: Suzanne Bakken, PhD, RN, FAAN, FACMI, FIAHSI, School of Nursing, Department of Biomedical Informatics, and Data Science Institute, Columbia University, 630 W. 168th Street, New York, NY 10032, USA; sbh22@cumc.columbia.edu

REFERENCES

1. Miller RA. Medical diagnostic decision support systems—past, present, and future: a threaded bibliography and brief commentary [published correction appears in J Am Med Inform Assoc 1994 Mar-Apr;1(2):160]. *J Am Med Inform Assoc* 1994; 1 (1): 8–27.
2. Shortliffe EH. Dehumanization of patient care—are computers the problem or the solution? *J Am Med Inform Assoc* 1994; 1 (1): 76–8.
3. Lyell D, Wang Y, Coiera E, Magrabi F. More than algorithms: an analysis of safety events involving ML-enabled medical devices reported to the FDA. *J Am Med Inform Assoc* 2023; 30 (7).
4. Liu S, Wright AP, Patterson BL, *et al.* Using AI-generated suggestions from ChatGPT to optimize clinical decision support. *J Am Med Inform Assoc* 2023; 30 (7).
5. Pfaff ER, Girven AT, Crosskey M, *et al.*; on behalf of the N3C and RECOVER Consortia. De-black-boxing health AI: demonstrating reproducible machine learning computable phenotypes using the N3C-RECOVER Long COVID model in the All of Us data repository. *J Am Med Inform Assoc* 2023; 30 (7).
6. Wang M, Sushil M, Miao BY, Butte AJ. Bottom-up and top-down paradigms of artificial research approaches to healthcare data science using growing real-word big data. *J Am Med Inform Assoc* 2023; 30 (7).
7. van der Vegt AH, Scott IA, Dermawan K, Schnetler RJ, Kalke VR, Lane PJ. Deployment of machine learning algorithms to predict sepsis: systematic review and application of the SALIENT Clinical AI Implementation Framework. *J Am Med Inform Assoc* 2023; 30 (7).