



Safety first: historical lessons on the responsible implementation of artificial intelligence

Travis J. Miles[^], Michael T. Guinn, Ravi K. Ghanta

Michael E. DeBakey Department of Surgery, Baylor College of Medicine, Houston, TX, USA

Correspondence to: Ravi K. Ghanta, MD. Michael E. DeBakey Department of Surgery, Baylor College of Medicine, 7200 Cambridge Street Floor 7, Houston, TX 77030, USA. Email: ravi.ghanta@bcm.edu.

Response to: Wang Z, Wang L. Safety Regulation of Machine Learning in Cardiac Surgery. J Thorac Dis 2024. doi: 10.21037/jtd-24-990. [Epub ahead of print].

Submitted Jul 22, 2024. Accepted for publication Aug 15, 2024. Published online Aug 26, 2024.

doi: 10.21037/jtd-24-1172

View this article at: <https://dx.doi.org/10.21037/jtd-24-1172>

It is our opinion that the safe implementation of artificial intelligence (AI) in medicine should center on two key concepts: (I) standardizing the knowledge, methods, and rigor with which it is developed and (II) ensuring the standards expected of such tools are held to the same thresholds expected of novel treatments or devices. History warns us of the risks associated with the rash implementation of innovative technologies, including perspectives arising from the work of the pioneering mathematician Alan Turing (1912–1954).

It is difficult to overstate the impact of Turing on the field of AI. His scientific work laid the foundation for major breakthroughs in computer science and machine learning (1,2). However, it is Turing's work as a cryptographer in Operation Ultra during World War II that illustrates an important lesson regarding the responsible implementation of AI.

Turing played a leading role in solving the Enigma code, a sophisticated cipher code used by the German military to encrypt their communications (3). Undoubtedly, successful decryption of Enigma played a substantial role in the eventual Allied victory. However, overconfidence in the novel technology led to major setbacks for the Allies later in the war (4). For example, Allied leadership ignored conventional reconnaissance forecasting a major enemy offensive when it wasn't corroborated by Ultra intercepts. The intelligence failure resulting from this overreliance on Ultra resulted in the subsequent campaign, known as the Battle of the Bulge, to be one of the costliest in American

military history.

History teaches us that whenever innovative technology is implemented in high-stakes decision making, safeguards must be enacted to account for unforeseen failures. We appreciate the insightful commentary by Wang *et al.*, who thoughtfully emphasize the safety considerations regarding AI-enabled decision support tools in clinical medicine (5). AI has the potential to improve surgical care. However, a balance must be struck between technological innovation and responsible implementation. AI-enabled decision support tools should supplement, not detract from, sound clinical judgement. In an effort to develop a framework for evaluating and implementing AI in surgery, the American College of Surgeons Committee on Health Information Technology summarized the current state of AI-enabled decision support tools in surgery (6). Their review revealed several concerning patterns with regards to rigor and reporting in the AI literature. As we have discussed, many of these studies rely on small samples of single center experiences, thereby producing models at substantial risk of over-fitting and thereby limiting generalizability. Additionally, proper data handling and model development are crucial to ensure AI research is safe and reliable. While work in this area is ongoing, increased attention is being given to standardizing the reporting of AI in medicine (7). First and foremost, standards must be in place to assure the science is of the utmost quality before any AI-enabled decision support tool is implemented clinically.

[^] ORCID: 0009-0008-2687-2380.

Often described as an “AI Chasm”, there is a significant gap between high performing machine learning algorithms in-silico and clinically meaningful decision support tools (8). If AI is to be successfully translated clinically, the safety and suitability of these tools must be rigorously assessed as would any other intervention or diagnostic test in healthcare. Recognizing this, the bioinformatics community has developed a comprehensive framework for evaluating AI-enabled decision support tools inspired by the clinical trials used to regulate approval of drugs and devices (9). Subjecting potential AI tools to rigorous assessment in this manner in conjunction with the appropriate oversight of relevant regulatory agencies provides a logical first step to ensure the safe and effective implementation of AI in healthcare.

When considering the future role of AI in medicine, a “safety first” approach is prudent. We should heed the historical lessons regarding the risks associated with overreliance on novel technological innovations. As surgeons, we must consider the potential innovations of AI while preserving sound clinical judgement and prioritizing patient safety.

Acknowledgments

Funding: This work was supported by the National Institutes of Health (NIH), National Heart Lung and Blood Institute (T32HL139430 to T.J.M. and M.T.G.).

Footnote

Provenance and Peer Review: This article was commissioned by the editorial office, *Journal of Thoracic Disease*. The article did not undergo external peer review.

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://jtd.amegroups.com/article/view/10.21037/jtd-24-1172/coif>). T.J.M. and M.T.G. are the post-doctoral research fellows participating in the Baylor College of Medicine T32 Research training program in cardiovascular surgery funded through the National Institutes of Health National Heart Lung and Blood Institute (No. T32HL139430) (received as salary support). The other author has no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are

appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Turing AM. On Computable Numbers, with an Application to the Entscheidungsproblem. Proceedings of the London Mathematical Society 1937;42:230-65.
2. Turing AM. Computing Machinery and Intelligence. Mind 1950;59:433-60.
3. Randell B. A Turing Enigma. In: Koutny M, Ulidowski I, editors. CONCUR 2012 – Concurrency Theory. Berlin, Heidelberg: Springer; 2012:23-36.
4. Deutsch HC. The Historical Impact of Revealing the Ultra Secret. The US Army War College Quarterly, Parameters 1977. doi:10.55540/0031-1723.1102.
5. Wang Z, Wang L. Safety Regulation of Machine Learning in Cardiac Surgery. J Thorac Dis 2024. doi: 10.21037/jtd-24-990. [Epub ahead of print].
6. Loftus TJ, Altieri MS, Balch JA, et al. Artificial Intelligence-enabled Decision Support in Surgery: State-of-the-art and Future Directions. Ann Surg 2023;278:51-8.
7. Kolbinger FR, Veldhuizen GP, Zhu J, et al. Reporting guidelines in medical artificial intelligence: a systematic review and meta-analysis. Commun Med (Lond) 2024;4:71.
8. Keane PA, Topol EJ. With an eye to AI and autonomous diagnosis. NPJ Digit Med 2018;1:40.
9. Park Y, Jackson GP, Foreman MA, et al. Evaluating artificial intelligence in medicine: phases of clinical research. JAMIA Open 2020;3:326-31.

Cite this article as: Miles TJ, Guinn MT, Ghanta RK. Safety first: historical lessons on the responsible implementation of artificial intelligence. J Thorac Dis 2024;16(8):5492-5493. doi: 10.21037/jtd-24-1172