NEJM AI

NEJM AI 2024;1(1)
DOI: 10.1056/AIe2300197

EDITORIAL

Injecting Artificial Intelligence into Medicine

Isaac S. Kohane, M.D., Ph.D.[1]

Received: October 17, 2023; Accepted: October 19, 2023; Published: December 11, 2023

Abstract

Introducing a new journal from NEJM Group: *NEJM AI*, a platform aimed at informing readers and guiding the responsible development of artificial intelligence (AI) to enhance the quality of health care. This editorial reflects on AI's historical milestones and current capabilities, particularly large language models, and the imperative for their rigorous clinical evaluation. It discusses the crucial need for AI in medicine to undergo the same level of scrutiny as any clinical intervention, predominantly through randomized controlled trials, despite the challenges posed by the technology's complexity. Looking forward, *NEJM AI* commits to fostering a multidisciplinary discourse and the development of a transparent, patient-centered approach to AI in health care, with an emphasis on the critical importance of diverse and accessible datasets. The editorial concludes by envisioning a future where *NEJM AI* not only informs but actively shapes the ethical integration of AI into a health care system that respects patient autonomy and upholds the highest standards of care.

Welcome to a collaborative journey as we launch *NEJM AI*. Whether you are a reader or author, we hope that you share our primary goal: augmenting the capabilities of clinicians, patients, and their larger community using the latest entrant to our ecosystem — AI — to deliver safe and effective health care to the highest of our collective standards. We approach this goal with optimism yet bear a constant awareness that misuse or careless implementation of these technologies will precipitate systemic harms to all parties in the health care system, most importantly, patients.

The spotlight on AI has been intensifying since its breakthrough in image recognition in 2012,[1] achieving performance rivaling human capability across various medical disciplines. By 2020, as revealed by a PubMed query, clinical evaluations of AI and machine learning had transformed from mere topics into a burgeoning field, that year alone boasting over 300 articles appraising these technologies. Of course, the aspiration to augment clinicians with AI, whether for safety or to overcome our cognitive limitations, is hardly new and

The author affiliation is listed at the end of the article.

Dr. Kohane can be contacted at Isaac_Kohane@hms.harvard.edu or at Harvard Medical School, Department of Biomedical Informatics, 10 Shattuck Street, Boston, MA 02115.

dates back to at least the 1950s.[2] In 1970, in the pages of the *New England Journal of Medicine*, the noted endocrinologist William Schwarz wrote that "computing science will probably exert its major effects by augmenting and, in some cases, largely replacing the intellectual functions of the physician …[and will influence] in a fundamental fashion the problems of both physician manpower and quality of medical care, it will also inevitably exact important social costs — psychologic, organizational, legal, economic and technical."[3] Fifty years later, the development of large online datasets, including medical texts and health records, the soaring performance of graphical processing units, and advances in neural network architectures gave ample reason for optimism that Schwarz's forecast might be realized in this century.[4]

Enter large language models (LLMs), which catapulted onto the stage and into the lives of patients and doctors in the fall of 2022, fueled by a social media frenzy with stories spanning from the transformative to the cautionary.[5] A retrospective glance to 1812, when the *New England Journal of Medicine and Surgery* made its debut, reminds us that randomized trials were then an unimaginable 130 years away. Yet, 208 years later, in 2020, of the aforementioned clinical AI evaluations published, less than 1% were prospective, randomized controlled trials. This makes clear our urgent mandate: In addition to needed technical advances, AI must meet the same bar for clinical evidence that is expected from other clinical interventions. For a given AI tool to be used, evidence that it will perform in a safe and effective manner must be demonstrated, preferably using randomized controlled trials designed to test the tool against an established standard.

Randomized controlled trials with LLMs will not be easy. The breadth of these programs' capabilities and unknowns about what data they have already "seen" makes their evaluation on narrowly defined tasks somewhat artificial and not entirely reflective of their usage by clinicians or patients. Necessarily, ensuring that pluripotent AI programs are "clinical grade" and safe will require a multifaceted, collaborative approach among clinicians, patient advocacy groups, and a myriad of stakeholders. Transparency will be the principal tool required to win the trust of our readers, regardless of the form this evaluation takes.

Evaluation is but one dimension of AI's future in health care. Questions loom large: Who will steer the innovation ship and toward which horizon? Our various article types, from Perspectives to Policy Corner, will serve as a vibrant forum, dissecting these questions and more, with a commitment to spotlighting not just successes but also informative failures and cautionary tales. Moreover, in the world of medical AI, well-annotated, representative, diverse, and freely available datasets arguably stand among the most precious resources, enabling investigators to assess the mettle of AI programs in realistic, representative tasks. Thus, articles unveiling new datasets, benchmarks, and innovative, reproducible protocols will not merely be welcomed but celebrated (see ai.nejm.org for a full list of article types).

So, who do we envision contributing to the pages of *NEJM AI*? Clinicians? AI researchers? In our eyes, the most impactful articles will blossom from the fertile ground of multidisciplinary teams, reflecting the vibrance at the intersection of computer science, clinician–patient dynamics, and biomedical research. Our editorial board, a meld of pioneers from the realms of AI, life sciences, ethics, and policy, is genuinely thrilled to engage with our community to find how best to navigate these complex, uncharted waters with precision, curiosity, and integrity.

Fast-forward a decade, and should you inquire about *NEJM AI*'s journey and its measure of success, many responses might serve. However, the paramount answer will hinge on this essential query: Did we help guide our health care system through the maelstrom of rapidly evolving technology, ensuring that patients are afforded an unprecedented standard of health care, all while safeguarding the autonomy and dignity they (i.e., we) inherently deserve? Although second in importance, answering the following query affirmatively is dear to all of us: Did we effectively advocate for the design and implementation of AI so that the practice of medicine is as emotionally and intellectually rewarding as some earnest medical school applications seem to anticipate?

NEJM AI's activities will extend considerably beyond the publication of a journal. We already host a popular podcast, *NEJM AI Grand Rounds*, a newsletter, and two to three seminars per year in a hybrid in-person/webcast format. We will organize international meetings with a specific thematic focus as the need arises. For example, we organized the Symposium for Responsible AI for Social and Ethical Healthcare (RAISE) that met at the end of October, and the white paper will be shared on our website and those of collaborating journals. Although the advent of AI in medicine is momentous, we are only at the very

beginning. I invite those of you who share our enthusiasm to engage with us in these activities as well as within the pages of our journal.

Disclosures

Author disclosures are available at ai.nejm.org.

Author Affiliation

[1] Harvard Medical School, Department of Biomedical Informatics, Boston

References

1. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. NeurIPS Proceedings. 2012 (https://papers.nips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html).
2. Haward LRC. The robot anaesthetist; an introduction to the automatic control of anaesthesia by means of an electro-encephalographic intermediary. Med World 1952;76:624-626.
3. Schwartz WB. Medicine and the computer. The promise and problems of change. N Engl J Med 1970;283:1257-1264. DOI: 10.1056/NEJM197012032832305.
4. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019;380:1347–58. DOI: 10.1056/NEJMra1814259.
5. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med 2023;388:1233-1239. DOI: 10.1056/NEJMsr2214184.