Laying an equitable data foundation for foundation models

Anurag Agrawal

Trivedi School of Biosciences, Ashoka University, Sonipat, Haryana, 131029, India

This year has been exciting for med-tech futurists. Every month brings news of a new barrier being broken by adaptable foundation artificial intelligence (AI) models trained on a vast quantity of data at scale.[1] These include generative pre-trained transformer (GPT) chatbots trained on massive amounts of text data (large language models, LLM), such as ChatGPT (OpenAI, San Francisco, CA, USA). ChatGPT has now been successfully used for writing a paper or grant, collating and presenting data, getting pass-scores on medical licensing exam questions, or even satisfactorily answering patients' queries.[2–7] In some comparisons, such as an evaluation of responses by physicians or ChatGPT to questions posted by patients to websites, the jury favoured the chatbot.[6] Each of these applications represents in-context learning, a feature of foundation models, through which they can carry out general tasks that they were never explicitly trained for. The vision is now shifting from incremental improvements in tools designed for specialized medical tasks, to applications that are generalist rather than specialist.[1] Given enough data, this seems plausible.

Much remains to be done, of course. Nuanced examination reveals many teething troubles, importantly 'hallucinations' and misinformation.[8] This is not specific to ChatGPT but rather seems to be a property of LLMs, of which there are many, at various stages of development, including BioGPT (Massachusetts Institute of Technology, Boston, MA, USA), LaMDA (Google, Mountainview, CA, USA), Sparrow (Deepmind AI, London, UK). While LLM chatbots provide facile answers to questions, they continue to lack human cognitive understanding or nuanced ability to sift information. This, along with the known tendency to 'hallucinate' and create fake information is a potentially major problem when it comes to medical applications.

The solution may lie in better architectures and data. I expect the choice of data used for pre-training and fact-checking to matter. BioGPT is being trained on PubMed articles for biomedical applications, hopefully increasing reliability. Given a theoretical potential for continued 'hallucination' which might even become more subtle, it is likely there will be need for additional post-processing using highly-reliable tools and data-sources. I am therefore concerned about inequity in available biomedical data across global regions, with most data representing white males living in high-income nations. In medicine, there are many questions that have different correct answers depending on who you are and where you live. How would we lay the correct data foundation for foundational AI, to recognize these differences?

The South East Asian Region (SEARO-WHO, World Health Organization) contains a quarter of the world population, over twenty major languages spoken by at least a million people, and many traditional medical systems. It is poorly represented in global medical literature. There is a need to increase the body of trusted medical literature from underrepresented regions, so as to foster research agenda, clinical practice, and health policy, but through an appropriate lens. This is where quality journals with a regional focus, such as *The Lancet Regional Health – Southeast Asia*, a part of The Lancet Regional Health initiative, have an important role, alongside curated regional knowledge repositories maintained by the WHO.[9]

Much more would be needed, of course, including generation and capture of trusted information outside the boundaries of standard medical databases. Language itself will be an issue, based on the Indian experience where only a few languages had adequate digital data for training LLM. Towards this end, the Indian government has recently launched Bhashini, a national public digital platform for languages, to ensure adequate representation.[10] This will surely be just one of many necessary steps across the region ensuring more equitable knowledge representation and minimization of bias in the source data for artificial intelligence (AI) initiatives. Unbiased analysis of biased data remains biased, and that too in ways with unforeseeable discriminatory consequences.[11]

I compliment *The Lancet Regional Health – Southeast Asia* on its first and very successful year of adding to the body of knowledge that advocates change in, or illuminates, clinical practice and health policy in Southeast Asia. This is more than promoting academic equity and showcasing regional research – it is part of a new foundation for building trusted knowledge sources and solutions in the age of AI.

Contributors
Anurag Agrawal as sole author, all views personal.

E-mail address: anurag.agrawal@ashoka.edu.in.

Correspondence

References
1 Moor M, Banerjee O, Abad ZSH, et al. Foundation models for generalist medical artificial intelligence. *Nature*. 2023;616:259–265.
2 Giannos P, Delardas O. Performance of ChatGPT on UK standardized admission tests: insights from the BMAT, TMUA, LNAT, and TSA examinations. *JMIR Med Edu*. 2023;9:e47737.
3 Else H. Abstracts written by ChatGPT fool scientists. *Nature*. 2023;613:423.
4 Gao CA, Howard FM, Markov NS, et al. Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *NPJ Digit Med*. 2023;6:75.
5 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:ocad072.
6 Ayers JW, Poliak A, Dredze M, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Intern Med*. 2023. https://doi.org/10.1001/jamainternmed.2023.1838. published online April 28.
7 Mihalache A, Popovic MM, Muni RH. Performance of an artificial intelligence chatbot in ophthalmic knowledge assessment. *JAMA Ophthalmol*. 2023. https://doi.org/10.1001/jamaophthalmol.2023.1144. published online April 27.
8 *Hallucinations could blunt ChatGPT's success*. IEEE Spectrum; 2023. https://spectrum.ieee.org/ai-hallucination#toggle-gdpr. Accessed April 30, 2023.
9 WHO Library. https://www.who.int/library; 2023. Accessed April 30, 2023.
10 *About Bhashini*. Bhashini; 2023. https://www.bhashini.gov.in/. Accessed April 30, 2023.
11 Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366:447–453.