The promise of large language models in health care

The past few months have seen a rapid acceleration of innovation in the field of generative artificial intelligence with large language models (LLMs), capturing academic, media, and public attention. GPT-3, BERT, and ChatGPT are examples of LLMs that have been trained on swathes of the internet to produce impressive responses to human queries. The ability of these generalised LLMs to write fiction, develop computer code, and speculate on the future has prompted recognition that we are approaching artificial general intelligence for the first time.[1]

It has already been proposed that LLMs, such as ChatGPT, could have applications in the field of health care due to the large volumes of free-text information available for training models. It has been thought that models could help with the writing of discharge letters by summarising a patient's hospital stay after reading their medical records. It has also been proposed that LLMs could be used in medical research, which is supported by the empirical example of GatorTron, an LLM trained on more than 90 billion words of text from electronic health records.[2] The same model was tested on its ability to answer medical questions in a natural language. These capabilities attract speculation on the use of LLMs as a medical triage service or an application to provide plain English answers to medical questions from the public. However, it must be emphasised that such technologies in their current form have the potential to incite harm, an effect that has been well described in the field of mental health, where natural language processing algorithms have shown biases relating to religion, race, gender, nationality, sexuality, and age.[3]

Although already attracting widespread speculation about their future potential, LLMs are still in their infancy. Current limitations include the amount of computer power required for LLMs to function, which is often costly. LLMs also require huge volumes of data to be trained effectively, which might dictate that the sharing of data between institutions could be required to train algorithms. Such data sharing presents a unique challenge in the field of health care, as strict data privacy laws and institutional data protection agreements contribute to data silos. Solutions are emerging, such as the use of federated learning or synthetic data to allow cross-border data sharing.[4] Although these techniques have proved successful in other domains, their applicability to LLMs remains uncertain. However uncertain the future of this research is does not make it impossible, nor far away.

AnmA has roles with the UK National Institute for Health Research, Health Data Research UK, NHS England & Improvement, and Moorfields Eye Hospital. AnaA declares no competing interests.

*Anmol Arora, Ananya Arora
aa957@cam.ac.uk

School of Clinical Medicine, University of Cambridge, Cambridge CB2 0SP, UK (AnmA, AnaA)

1 Lauret J. GPT-3: the first artificial general intelligence? July 22, 2020. https://towardsdatascience.com/gpt-3-the-first-artificial-general-intelligence-b8d9b38557a1 (accessed Jan 14, 2023).
2 Yang X, Chen A, PourNejatian N, et al. A large language model for electronic health records. *Npj Digit Med* 2022; **5:** 1–9.
3 Straw I, Callison-Burch C. Artificial intelligence in mental health and the biases of language based models. *PLoS One* 2020; **15:** e0240376.
4 Arora A, Arora A. Generative adversarial networks and synthetic patient data: current challenges and future perspectives. *Future Healthc J* 2022; **9:** 190–93.

Insurance pools' merging in China needs careful design

Chen Xinxin and colleagues' call for healthy ageing in China is encouraging.[1] To address equity concerns and mobility difficulties, the policy recommendation of "establishing a unified national insurance system that encompasses all citizens, regardless of occupation, resident status, or place of residence"[1] is justifiable. However, it deserves further considerations.

Notably, insurance pools' merging could mismatch administrative authorities and expenditure responsibilities at different levels of the government, creating huge moral hazards. After merging, despite the upward transferal of some expenditure management authority (eg, benefit package formulation), the local governments would still have the supervision authority (eg, payment inspection and cost control). However, local governments no longer need to take the expenditure responsibility, which is instead borne by the higher level of the Government in charge of the insurance funds. Due to the information asymmetry, the higher level of the Government cannot perfectly monitor the behaviours of the local governments, nor can it directly inspect the medical institutions to detect fraud. In this case, the local governments have strong incentives to benefit medical institutions and insured people under their administration, with the tendency to relax regulations around cost control and overuse insurance funds.

Thus, the merging of health insurance pools might need careful research. Two policy options can be beneficial to improve the current system. First, it might be necessary to develop a harmonised benefit package for local insurance pools based on the economic reality and cost-effectiveness analyses. Second, China can consider adopting the population-based risk adjustment method for cross-subsidisation among insurance pools.[2]

I declare no competing interests.

Yian Fang
yianfang@bjmu.edu.cn

Department of Health Policy and Management, School of Public Health, Peking University, Beijing 100191, China