Machine learning

https://doi.org/10.1038/s41591-024-03257-3

# Improving primary healthcare with generative AI

Winnie Yip



Studies in China show how large language models can improve primary healthcare systems, but equitably scaling this technology will require attention to rural, low-resource settings and the companion policies that support its implementation.

CREDIT: RAPEEPONG PUTTAKUMWONG / MOMENT / GETTY

With a rapidly aging population and rising levels of non-communicable diseases, China has recognized the importance of building a strong primary healthcare system to ensure affordable, equitable and effective healthcare that is also financially sustainable. A major goal of the country's Healthy China 2030 policy is to build a vertically integrated healthcare delivery system anchored on strong primary care that prioritizes prevention and people[1]. However, progress has been slow. A key bottleneck is a lack of qualified health professionals who are competent and trusted by the general population[2,3].

There is widespread enthusiasm for the proposition that digital technologies, especially artificial intelligence (AI), offer solutions for improving the clinical competency of primary healthcare providers in China. Software companies have introduced a plethora of products that have been installed in healthcare facilities countrywide. However, most of these have not gone through rigorous scientific validation for effectiveness. Also, many products adopt first-generation AI tools that are primarily clinical decision support systems that computerize existing clinical guidelines. Primary healthcare providers, especially those who are less qualified, have not found them helpful. In this issue of *Nature Medicine*, Li et al.[4] and Wan et al.[5] describe the successful integration of large language models (LLMs) into primary care settings, to enhance healthcare providers' capabilities and improve patient care experiences.

Both studies report on the development and retrospective or in silico validation of the models, followed by prospective clinical evaluation. In their randomized, controlled trial, Li et al.[4] show that primary care providers (PCPs) assisted by DeepDR-LLM — a diabetes screening model that integrates a LLM with image-based deep learning — effectively identified referable diabetic retinopathy at a higher rate than unassisted PCPs and helped facilitate effective self-management of the condition by patients. To help enhance efficiency and reduce burn-out among receptionist nurses, Wan et al.[5] developed a site-specific prompt engineering chatbot (SSPEC) that answered patients' questions more efficiently and empathetically than receptionist nurses did. Furthermore, the authors' two-arm, randomized study found that nurses aided by SSPEC produced higher patient satisfaction scores than those of unaided nurses.

These innovations and rigorous evaluations are encouraging and much needed. In addition to demonstrating clinical competency, both studies also demonstrate the ability of the LLMs to enhance the perceived empathy of patient–provider interactions. In the SSPEC model of Wan et al.[5], the technology's courteous, three-step approach to answering patient questions demonstrated a level of engagement that was difficult for real-life receptionist nurses to match[5]. In the study by Li et al.[4], a post-deployment evaluation revealed that most patients favored recommendations from PCPs assisted by the DeepDR-LLM technology over those received from unassisted PCPs, with 73.92% of the recommendations provided by the technology-assisted PCPs considered "empathetic" or "very empathetic"[4]. Because empathetic communication is essential for building trust between providers and patients, these findings indicate the potential of LLMs to strengthen public confidence in primary care.

Many of the vertically integrated delivery systems that have emerged in recent years in China have been led by hospitals, and close to 80% of total health resources are spent on hospitals — leaving primary care services under-resourced[6]. Among primary care settings, those with the greatest need for enhanced primary care competencies are resource-poor rural areas and community health centers in the central and western regions. However, many existing studies are based in high-resource settings, including the two in this issue. The study by Wan et al.[5] draws on data from two major urban medical centers and the findings of Li et al.[4] are based in part on data from the Huadong Sanitorium in Shanghai, a well-resourced hospital that primarily serves retired government officials. The extent to which these study results are generalizable — and the extent to which these technologies are adoptable in under-resourced settings in China — is unclear and requires more research.

In China, the majority of primary healthcare facilities are not staffed by physicians with general practitioner (GP) qualifications, who must receive 5 years of general (undergraduate) medical education plus 3 years of specialist training. Provider training and qualifications are highly variable across geographic areas, with most GPs practicing in urban areas and more economically developed regions in the east[7]. Qualified GPs generally choose not to practice in rural areas or lower-income regions because of lower pay, sparser promotion opportunities and less-attractive living conditions in general. The PCPs in the study by Li et al.[4] are probably qualified GPs. To address the shortage of GPs in rural and lower-income areas, China should consider staffing primary healthcare facilities with nurses and community health workers. Whether the LLM models developed by Li et al.[4] and Wan et al.[5] could enhance healthcare delivery by these workers will also require further investigation. Otherwise, technological advancement may further exacerbate inequity.

Technology is not a panacea for improving primary care — that requires a systemic approach. First, unless providers are properly incentivized (financially and non-financially) to improve quality of care, it is unlikely that they will be motivated to spend time learning a new technology. At present, providers in China are financially rewarded for the volume — rather than quality — of care provided. Quality-specific financial compensation and promotion criteria need to be introduced to enhance the adoption of quality-improving technology.

Second, for non-communicable diseases such as diabetes and hypertension, poor patient awareness is a major contributor to poor health outcomes. Among patients with diabetes who were surveyed in 2018, for example, only 36.7% were aware of their condition, and only 50.1% of those receiving treatment were able to effectively control their hemoglobin A1c levels[8]. Similarly, a 2023 study found that only 45.6% of people with hypertension were aware of their condition, and only 26.8% of those receiving treatment had controlled their blood pressure[9]. Effective primary care requires community outreach; focusing solely on improving the quality of care delivered at hospitals and clinics overlooks the majority of patients who remain undiagnosed.

Third, there need to be clear policies on how to finance technologies, from development to adoption. In particular, what roles should the government and the private sector have? For technologies that benefit low-resource settings and are not profitable for commercial entities to scale, the government needs to have a core financing role. Finally, regulation to ensure the safety and efficacy of technology is critical. The rapid changes and complexity of technologies such as LLMs makes their regulation and approval particularly challenging[10].

Strengthening primary healthcare is a global priority, and the studies by Li et al.[4] and Wan et al.[5] demonstrate how generative AI can help achieve this goal. However, expanding this technology's impact will require more rigorous research in low-resource settings, as well as systemic policy dialogue and decision-making.

**Winnie Yip**
Department of Global Health and Population, Harvard T. H. Chan School of Public Health, Boston, MA, USA.
e-mail: wyip@hsph.harvard.edu

Published online: 18 September 2024

## References

1. Tan, X., Liu, X. & Shao, H. *Value Health Reg. Issues* **12**, 112–114 (2017).
2. Li, X. et al. *Lancet* **395**, 1802–1812 (2020).
3. Zhang, P. et al. *Health Policy Plan.* **30**, 1243–1250 (2015).
4. Li, J. et al. *Nat. Med.* https://doi.org/10.1038/s41591-024-03139-8 (2024).
5. Wan, P. et al. *Nat. Med.* https://doi.org/10.1038/s41591-024-03148-7 (2024).
6. Yip, W. et al. *Lancet* **394**, 1192–1204 (2019).
7. State Council. *Guidelines for Establishing General Practitioners* (State Council of the People's Republic of China, 2011).
8. Wang, L. et al. *J. Am. Med. Assoc.* **326**, 2498–2506 (2021).
9. Zhou, T. et al. *Lancet Glob. Health* **11**, e83–e94 (2023).
10. Blumenthal, D. & Patel, B. *NEJM AI* https://doi.org/10.1056/AIpc2400545 (2024).

## Competing interests

The author declares no competing interests.