Large language models to support community health workers

David Levine [1]

Abstract

The Challenge: Across the globe, 2 billion underserved individuals rely on the services of over 2 million community health workers (CHWs) for primary healthcare. These CHWs, predominantly women from low-resource backgrounds, bridge the gap in healthcare access in their communities. However, with only weeks of training and limited resources, the care quality they deliver is often inconsistent.

Our Solution: Enter ChatCHW - a tool that marries large language models like ChatGPT with local clinical guidelines. This smartphone app empowers CHWs to provide consistent and excellent care. ChatCHW's features include:

- **Expertise at Your Fingertips**: Powered by advanced large language models like ChatGPT or Google's Bard, ChatCHW brings medical expertise to every CHW.
- **Localized and Accurate**: Large language models can hallucinate wrong (but confident) answers. To ensure accuracy, ChatCHW relies solely on inputs from the CHW agency:
    - local clinical guidelines or CHW training manuals.
    - real-time data on disease prevalence and
    - A list of the equipment and supplies a CHW carries.
- **User-Friendly Design**: With multi-language support, ChatCHW communicates seamlessly with CHWs. Its intuitive design allows CHWs to adjust literacy levels and the use of medical terminology. ChatCHW also turns most input into a simple touch (e.g., for Yes or No), and supports voice input for longer answers.
- **Integration with EMRs**: If CHWs have access to electronic medical records (EMRs), ChatCHW streamlines data entry and retrieval, enhancing patient care.
- **Implementation is Easy**: Setting up ChatCHW in a new region is hassle-free. Just upload local guidelines or training manuals, and the system is ready to go. It is easy to update the database of guidelines when a region has new pathogens (e.g., COVID), diagnostics or treatments.

How It Works: When a patient arrives, the CHW speaks or types in the symptoms the patient presents with. ChatCHW consults the clinical guidelines and suggests diagnostic questions. As the visit progresses, the app recommends tests to perform, a diagnosis, and treatment (sometimes including referral to higher-level care). Visual aids ensure CHWs understand each step. For the organization as a whole, ChatCHW offers regional insights through a dashboard, highlighting trends and potential healthcare concerns such as a new disease outbreak.

[1] Haas School of Business, University of California, Berkeley, USA Levine@berkeley.edu
Competing interest statement: I have no competing interests.

Keywords: Artificial intelligence, large language models, retrieval-augmented generation, community health worker (CHW)

Globally, over two million community health workers (CHWs) are the primary healthcare providers for nearly two billion people living in poverty. This article explores how the integration of large language models (LLMs), when enhanced with localized clinical guidelines, training content, disease incidence data, available supplies, and accessible via mobile apps on CHWs' devices, could significantly improve the healthcare services these workers offer.

The problem

CHWs provide health education and treat common ailments in their community. CHWs in low- and middle-income countries (LMICs) nations are typically women with an 8th to 10th grade education who live in poor communities. Most CHWs receive very low pay and receive only a few weeks' training.[2]

On the one hand, CHWs appear on average to improve health outcomes[3], and the very best programs are highly effective.[4] On the other hand, low training and few resources implies the quality of care CHWs provide is often low.[5]

For example, we know that even when treated by medical doctors, patients in poor nations often receive antibiotics instead of ORS for uncomplicated diarrhea (where antibiotics are not helpful,

---

[2] Olaniran A, Banke-Thomas A, Bar-Zeev S, Madaj B (2022) Not knowing enough, not having enough, not feeling wanted: Challenges of community health workers providing maternal and newborn services in Africa and Asia. *PLoS ONE* 17(9): e0274110. https://doi.org/10.1371/journal.pone.0274110

[3] Swider SM. Outcome effectiveness of community health workers: an integrative literature review. *Public Health Nur*s. 2002 Jan-Feb;19(1):11-20. doi: 10.1046/j.1525-1446.2002.19003.x. PMID: 11841678.
2012
Henry Perry and Rose Zulliger. "How Effective are Community Health Workers?" Departments of International Health and Health, Behavior and Society, Johns Hopkins Bloomberg School of Public Health, 2012.
https://www.childhealthtaskforce.org/sites/default/files/2019-07/How%20Effective%20are%20CHWs_Evidence%20Summary%20Condensed%28JHSPH%2C%202012%29.pdf

[4] Björkman Nykvist, Martina, Andrea Guariso, Jakob Svensson, and David Yanagizawa-Drott. "Reducing child mortality in the last mile: A randomized social entrepreneurship intervention in uganda." (2017).

[5] Olaniran, et al., *op cit.*
More generally, Das and coauthors document low quality care in clinics; CHWs have even less training and more limited supplies.
Das, Jishnu, Alaka Holla, Veena Das, Manoj Mohanan, Diana Tabak, and Brian Chan. "In urban and rural India, a standardized patient study showed low levels of provider training and huge quality gaps." *Health affairs* 31, no. 12 (2012): 2774-2784.
Das, Jishnu, Alaka Holla, Aakash Mohpal, and Karthik Muralidharan. "Quality and accountability in health care delivery: audit-study evidence from primary care in India." *American Economic Review* 106, no. 12 (2016): 3765-3799.

while ORS is very valuable), but do not always receive antibiotics for strep throat or for pneumonia[6] .

CHWs have limited medical supplies. For diagnosis, CHWs typically are supposed to have a rapid strep test, thermometer, and a weighing scale for newborns. For treatment she is supposed to have items such as paracetamol (branded as Tylenol in the US), some first aid supplies, ORS and zinc, contraceptives, and (in some programs) antibiotics. ASHAs frequently run out of supplies and often are not able to refill at the clinic.

CHWs typically focus on maternal and child health. For example, they treat diarrhea with ORS (and sometimes zinc). They typically distinguish whether a sore throat is just a cold or is strep that should be treated with antibiotics, whether a cough is a cold virus or likely to be pneumonia that should be treated with antibiotics, etc. They refer serious cases to higher levels of care. In some settings (such as India) they also provide antibiotics for likely bacterial pneumonia (the most common killer of children). At the same time, CHWs are increasingly involved in screening for and treating noncommunicable diseases such as high blood pressure and diabetes.

Many CHW programs already provide smartphones or tablets. These devices largely track activities.[7] The largest of these products, OpenSRP, is an open source digital health platform to support CHWs. It has over 150 million patients registered in 14 nations as of June 2023 (and growing rapidly).[8] OpenSRP also provides guidance on specific types of care (e.g., antenatal care). India is also developing a digital tool for its million plus ASHAs (India's name of CHWs).

Can artificial intelligence help?

Even without specialized training, by 2023 LLMs such as GPT4 had already passed medical exams and shown good diagnostic abilities.[9] However, LLMs are not appropriate for deployment to CHWs:

- LLMs can give inappropriate advice
    - LLMs are trained on the entire Internet. Thus, their recommendations need not accord with local clinical guidelines — or even use the appropriate word for a local medicine (e.g., Is the local name paracetamol vs. acetaminophen?).
    - LLMs have an end date to training. When a new danger arises (e.g., COVID), the LLM won't immediately know about any changes in guidelines.
    - LLMs sometimes produce "hallucinations" —fabricated responses that often sound authoritative and convincing.
- LLMs lack local data
    - LLMs do not know the diagnostic tools (e.g., a thermometer) and medicines a CHW is supposed to carry. They know even less about stockouts with this specific CHW.

---

[6] Wagner, Zachary, John Bosco Asiimwe, William H. Dow, and David I. Levine. "The role of price and convenience in use of oral rehydration salts to treat child diarrhea: A cluster randomized trial in Uganda." *PLoS medicine* 16, no. 1 (2019): e1002734.

[7] E.g., Emmanuel G, Emmanuel AWR (2018) A Mobile Application System for Community Health Workers: A Review. *Glob J Res Rev* Vol.5 No.2:11. DOI: 10.21767/2393-8854.100040

[8] OpenSRP website. https://opensrp.io Last accessed April 1, 2024.

[9] Savage, T., Nayak, A., Gallo, R. et al. Diagnostic reasoning prompts reveal the potential for large language model interpretability in medicine. *npj Digit. Med*. 7, 20 (2024). https://doi.org/10.1038/s41746-024-01010-1

    - LLMs cannot access electronic medical records
    - LLMs do not know local base rates for problems in a region, such as whether this region has high rates of malaria this month
- LLMs have limited communication
    - LLMs (in 2024) communicate largely via textual input and output, while most CHWs in the field will have a difficult time typing responses to questions.
    - A medical visit often requires communication after the visit, such as an after-visit summary and reminders to complete medicine doses. LLMs lack those capabilities.
- LLMs are not validated:
    - An NGO or Ministry of Health will not recommend an AI assistant until it has been tested thoroughly.
- LLMs do not learn from experience
    - If multiple patients in a region have an illness (e.g., malarai), it is important to update the base rates for that disease both to improve diagnosis and to address root causes of the outbreak. LLMs do not cumulate queries to update regional data.

ChatCHW: A proposed solution to these challenges

We propose ChatCHW - a tool that marries large language models like ChatGPT with local clinical guidelines and a smartphone app.

To avoid hallucinations and ensure its advice fits local guidelines, ChatCHW uses only clinical guidelines and training materials (not the entire Internet) in forming its responses. To ensure its guidelines are up to date, users can easily add new guidelines when needed; for example, when a new pathogen such as COVID arises.

To ensure its recommendations fit local conditions, ChatCHW uses real-time data on disease prevalence and on this CHW's supplies.

If the CHW has access to an electronic medical record (EMR), ChatCHW will use the EMR to reduce the number of questions the CHW asks and will update the EMR at the end of the session.

The ChatCHW app simplifies the completion of most questions, by presenting them as checkboxes. For longer answers, the app accepts spoken input. Each CHW can adjust ChatCHW's language, level of literacy, and use of medical jargon.

ChatCHW will also use phone capabilities to create outputs, such as sending opt-in reminders to CHWs or patients.

Using ChatCHW

To implement ChatCHW in a new region, a Ministry of Health or NGO uploads local guidelines or training manuals, and the system is ready to start the validation process. and runs a set-up program. The program creates a database of guidelines. Importantly, if a new

pathogen (e.g., COVID), diagnostic, or treatment arises, the implementer just loads updated guidelines. The implementer also enters the supplies that CHWs carry and current rates of infectious diseases such as malaria.

When a patient arrives, the CHW enters a brief written or spoken description of the patient and their symptoms into the app. The app sends this query to the database, which finds the relevant pages of the clinical guidelines. The app incorporates these relevant pages, along with data on local conditions provided by local health officials, into the prompt for the large language model. The prompt instructs the LLM to use only these guidelines and data; if the answer is not there, then reply “I don’t know.”

Using only the guidelines and local data provided, the LLM initially recommends questions for the CHW to ask the patient. The app formats these questions into checkboxes and gets the CHW’s responses The app sends the CHW’s responses, along with an updated prompt, to the LLM. The LLM then suggests additional questions or physical exams. Once exams are completed, the app prompts the LLM to propose a diagnosis and recommend a treatment plan, which may involve referring the patient to a higher level of care. If the CHW requests, ChatCHW will use images of videos to show how to perform an exam, present the relevant excerpts from the clinical guidelines, and explain the reasoning of a diagnosis.

If the patient opts in, ChatCHW drafts after-visit summaries. After approval by the CHW, ChatCHW delivers the summary to each patient, including information on care, when to see a doctor, and how to prevent recurrence (e.g., teaching key events for handwashing with soap after a diarrhea case). Patients can choose to receive these instructions as a voice message as well as SMS or WhatsApp message.

When appropriate, ChatCHW also delivers reminders to patients (e..g, to complete a dose of antibiotics) and to CHWs (e.g., for a follow-up visit).

Building in learning

ChatCHW generates a detailed record of each interaction. The backend of the system will aggregate all interactions into a dashboard for district, regional and national healthcare systems. This dashboard will help identify patterns of stockouts and progress remediating them.

A backend LLM will also help identify new patterns of symptoms. These analyses will help identify new outbreaks such as cholera or malaria. Patterns in non-response to treatment can help identify new forms of antibiotic and antiviral resistance.

Importantly for global health, unexpected patterns of symptoms can also help identify novel pathogens. Spotting new pathogens early can create essential time to learn about it and create appropriate containment policies, vaccines or cures.

Concerns

There are many ways that artificial intelligence in medicine can fail to improve health. These challenges are amplified in the low-resource settings that ChatCHW is designed for.

Some challenges for low-resource settings are familiar, such as low Internet connectivity in some poor regions. Fortunately, most people (even in poor nations) have decent mobile phone connectivity. Moreover, it should be possible to run powerful LLMs on smartphones within a few years.[10]

Many programs for smartphones in poor nations run into hardware problems such as breakage or a lack of charge.[11]

Other challenges are familiar for any introduction of information technology in medicine, such as resistance by care providers.[12] ChatCHW may not be effective if it makes CHWs work slower or more onerous. It is important to design the system with CHWs so that the tool does not slow CHWs down. It is important that CHWs perceive ChatCHW as a tool to empower them, not control them. Finally, communities must perceive ChatCHW as a high-tech tool so that CHWs gain status for their expertise.

More generally, ChatCHW development should follow all WHO guidelines for artificial intelligence in healthcare[13]. These include principles of transparency, data protection and privacy, and post-release auditing and impact assessments. Many poor nations will need to update their regulatory framework to cover AI assistants in healthcare.

Even when relying on local clinical guidelines or training materials, the system could still give erroneous advice. For example, the ASHA program of community healthcare workers in India has been around for decades. Thus, they have multiple volumes of training materials that cover core topics such as diarrhea and respiratory infections.[14] The system will need extra instructions to ensure it uses the most recent guidelines for a set of symptoms.

In addition, LLMs were originally designed for reading in text, but some clinical guidelines use tables and/or flowcharts. LLMs' capabilities are improving for reading in tables and flowcharts.[15] Nevertheless, it is crucial that an implementer tests the quality of the LLM's comprehension of such complex guidelines.

---

[10] Erik van Klinken. "Apple shows how LLMs can be run on smartphones." *Techzine.* December 22, 2023 https://www.techzine.eu/news/applications/114701/apple-shows-how-llms-can-be-run-on-smartphones/

[11] Kruse C, Betancourt J, Ortiz S, Valdes Luna SM, Bamrah IK, Segovia N. Barriers to the Use of Mobile Health in Improving Health Outcomes in Developing Countries: Systematic Review. *J Med Internet Res.* 2019 Oct 9;21(10):e13263. doi: 10.2196/13263. PMID: 31593543; PMCID: PMC6811771.

[12] Safi S, Thiessen T, Schmailzl KJ. Acceptance and Resistance of New Digital Technologies in Medicine: Qualitative Study. *JMIR Res Protoc*. 2018 Dec 4;7(12):e11072. doi: 10.2196/11072. PMID: 30514693; PMCID: PMC6299231.

[13] WHO, *Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models.* 18 January 2024 https://www.who.int/publications/i/item/9789240084759

[14] See ASHA training manuals at https://www.nhm.gov.in/index1.php?lang=1&level=3&sublinkid=184&lid=257

[15] For example, Peng Li, Yeye He, Dror Yashar, Weiwei Cui, Song Ge, Haidong Zhang, Danielle Rifinski Fainman, Dongmei Zhang, Surajit Chaudhuri. "Table-GPT: Table-tuned GPT for Diverse Table Tasks." 13 Oct 2023. arXiv:2310.09263

ChatCHW as part of suite of tools

Integrating phone sensors can add new capabilities to CHWs. For example, pneumonia is the most common cause of death for children in poor nations. A phone's microphone can act as a stethoscope to diagnose pneumonia[16]. The microphone should also be able to diagnose some serious heart problems (such as acute rheumatic carditis) that are resolvable even in many low-resource settings.

Anemia is pervasive in poor nations, often affecting half or more of the population[17] (xx). A smartphone camera can take a photo of fingernails, which is sufficient to diagnose anemia with high validity[18]. The camera can also diagnose some skin cancers[19]. Both anemia and many early stage skin cancers can be treated inexpensively in most low-resource settings.

More generally, a smartphone with ChatCHW can accept data from any bluetooth enabled medical device.

Validating ChatCHW

There are many stages to moving from idea to product that an NGO or Ministry of Health could deploy at scale. ChatCHW will come with protocols to validate its quality in different settings.

These protocols will use qualitative and quantitative methods to measure if the system provides appropriate questions, assessments, diagnoses and treatments; operational data on lag times, etc.; and feedback from CHWs and patients.

Typical stages of testing include:

[16] Hongxing Luo, Pablo Lamata, Salomé Bazin, Thea Bautista, Natsuki Barclay, Mehrdad Shahmohammadi, Jolijn M Lubrecht, Tammo Delhaas, Frits W Prinzen, Smartphone as an electronic stethoscope: factors influencing heart sound quality, *European Heart Journal - Digital Health*, Volume 3, Issue 3, September 2022, Pages 473–480, https://doi.org/10.1093/ehjdh/ztac044

[17] Yarlini Balarajan, Usha Ramakrishnan, Emre Özaltin, Anuraj H Shankar, SV Subramanian. " Anaemia in low-income and middle-income countries." Vol. 378, Issue 9809, P2123-2135. Aug. 2, 2011 DOI:https://doi.org/10.1016/S0140-6736(10)62304-5

[18] Mannino RG, Myers DR, Tyburski EA, et al. Smartphone app for non-invasive detection of anemia using only patient-sourced photos. *Nat Commun* 2018; 9: 1–10.

[19] Goyal, Manu, Thomas Knackstedt, Shaofeng Yan, and Saeed Hassanpour. "Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities." *Computers in Biology and Medicine* 127 (2020): 104065.

1. Design ChatCHW in cooperation with CHWs, their supervisors and medical directors. Study CHW needs.[20] Share prototypes of the interface to CHWs and others to ensure the system meets their needs.[21]
2. Test ChatCHW using a suite of hundreds of simulated patients.
3. Have CHWs use the app, with researchers posing as simulated patients.
4. A survey experiment can use vignettes of patients, where each CHW will answer a random set of vignettes without using ChatCHW, and then a different random set with ChatCHW.[22] The order of using ChatCHW or not can be randomized. This survey will probably overestimate CHW accuracy when not using ChatCHW, as the CHWs know we are observing their (simulated) care. Nevertheless, assuming that knowledge gaps are important impediments to quality care, ChatCHW should further increase CHW accuracy.
5. Doctor-observed usage: A medical doctor can observe CHWs using ChatCHW with patients to ensure the system provides high quality recommendations.
6. ChatCHW probably cannot be deployed to an entire healthcare system at once. Thus, a randomized or quasi-experimental deployment can compare health outcomes in regions with early deployment to comparison regions that do not have ChatCHW. This evaluation can survey both CHWs and patients. This roll-out should also include focus group discussions and key informant interviews with both CHWs and patients.
7. It is important to ensure ChatCHW meets its goals once it is at scale. ChatCHW will le CHWs give feedback after each patient visit.
8. Deployments should also include both CHWs and medical professionals in an ongoing advisory board.

These studies and feedback mechanisms will help those employing ChatCHW understand how to operate ChatCHW (e.g., training needs, hardware challenges, and so forth); how to integrate ChatCHW with existing digital platforms for CHWs; whether ChatCHW improves accuracy in diagnosis and decision making; and the potential gains in cost effectiveness and health that would support adoption by government and NGOs.

A pathway to scale

As noted above, many CHWs already use smartphone software. It is conceptually straightforward for these CHW programs to add ChatCHW as a module in their existing software. For example, OpenSRP has hand-coded a small number of WHO guidelines as modules of their software.[23] ChatCHW takes a complementary approach that can fill in the many guidelines OpenSRP has not yet coded. If OpenSRP adopted ChatCHW, the new software

[20] For example, see the excellent study “It cannot do all of my work”: Community Health Worker Perceptions of AI-Enabled Mobile Health Applications in Rural India
Chinasa T. Okolo. Srujana Kamata, Nicola Dell, Aditya Vashistha
CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing SystemsMay 2021 Article No.: 701Pages 1–20. https://doi.org/10.1145/3411764.3445420

[21] We have shown mock-ups of ChatCHW interface consistent with the text described above to a convenience sample of 14 CHWs and medical doctors in India and Africa. Results are extremely preliminary, but encouraging. All reviewers thought the system would be helpful.

[22] “How To Evaluate Your Community Health Worker Training Program” https://chwtraining.org/training-points-you-should-be-evaluating/ Last accessed April 1, 2024.

[23] OpenSRP website. https://opensrp.io Last accessed April 1, 2024.

could immediately support over 150 million patients. India could adopt a version of this software to reach over a million CHWs (ASHAs), serving roughly a billion patients.

It might cost $20 to $30 million dollars per year to provide ChatCHW and smartphones to all ASHAs in India for five years. It is too early to know the effectiveness of ChatCHW. Assume, for example, ChatCHW lowers mortality from child diarrhea and pneumonia in India by 5%. In that case, ChatCHW would save roughly 11,000 lives a year.[24] These assumptions imply a cost of under $3000 for a life saved, or less than $65 to save a disability-adjusted life year. Even if effectiveness is only half this level and costs were double, ChatGPT would be one of the most cost-effective health interventions.

ChatCHW may also receive support from wealthy nations due to two important global public goods it provides. First, as noted above, it can help identify new outbreaks and novel pathogens. Given the recent history of Ebola, COVID, Zika, novel pathogens are a major concern for donor nations. Second, the typical child in a low- or middle-income nation receives 25 doses of antibiotics in their first 5 years. Many of these doses are not necessary, and this overuse of antibiotics can lead to antibiotic resistance.[25] As with novel pathogens, antibiotic resistance is a major concern for rich nations.

Conclusion

This article outlines a combination of a large language model, retrieval-augmented generation based on local clinical guidelines or training materials, and a front-end app on a community health worker's phone. The goal of ChatCHW is to empower CHW's with an AI assistant while using local knowledge and avoiding hallucinations.

While there is no operable system at this time (Spring 2024), this innovation has great potential to save lives in a highly cost-effective manner.

Statement: During the preparation of this work the author used GP4 to edit some passages. After using this tool, the author reviewed and edited the content as needed and take full responsibility for the content of the publication

[24] *Pneumonia & Diarrhea Progress Report 2020* - Johns Hopkins Bloomberg and International Vaccine Access Project. https://publichealth.jhu.edu/sites/default/files/2024-02/ivacpdpr2020ax.pdf.

[25] Fink, Günther et al. "Antibiotic exposure among children younger than 5 years in low-income and middle-income countries: a cross-sectional study of nationally representative facility-based and household-based surveys." *The Lancet Infectious Diseases*, Volume 20, Issue 2, 179 - 187