EXECUTIVE SUMMARY

Generative AI (GenAI) has the potential to improve health and healthcare in low- and middle-income countries (LMICs). Where is GenAI currently being used and what are the greatest successes? How can we realize greater impact and unlock the full potential of GenAI, both for behavior change and broader healthcare applications?

To help answer these questions, from August to December 2024 we conducted an extensive review including two roundtable events, in-depth interviews with dozens of people who are actively working on applications of GenAI to health and healthcare in LMICs (including academics, health system leaders, implementers, and funders), and a quantitative survey with over 100 respondents. Additionally, we reviewed 14 GenAI accelerator programs for health that have collectively supported over 250 projects worldwide.

This white paper has a specific focus on the use of GenAI tools to drive health-related behavior change (HBC). Our scoping analysis, framework and key recommendations are inclusive of a range of health use cases, to contextualise HBC interventions within the wider ecosystem, and facilitate broadly applicable learnings.

Here is what we found.

Where is GenAI being currently used, and what are the greatest successes?

Use cases typically centered around applying large language models (LLMs) to health and healthcare-related tasks related to summarization, classification, extraction, translation, and/or conversation (please see definitions in Table 1). Use cases typically fell in one of three categories: direct-to-consumer, direct-to-provider, or system-level.

Some examples include:

Direct-to-Consumer

Offering personalized counseling on sensitive topics (e.g., HIV testing, sexual and reproductive health) via conversational LLM-based agents. In some cases, using improved voice capabilities of LLMs to better engage consumers, especially in low-literacy settings.

Direct-to-Provider

Providing better support for healthcare worker-to-consumer communication in traditional helpdesk workflows, including triaging and routing incoming questions, providing personalized suggested responses for healthcare workers, and live translation between languages.

System-Level

Generating early-warning alerts for potential emerging pandemics by analyzing large amounts of unstructured data from diverse sources, such as health records, news articles, social media and climate data.

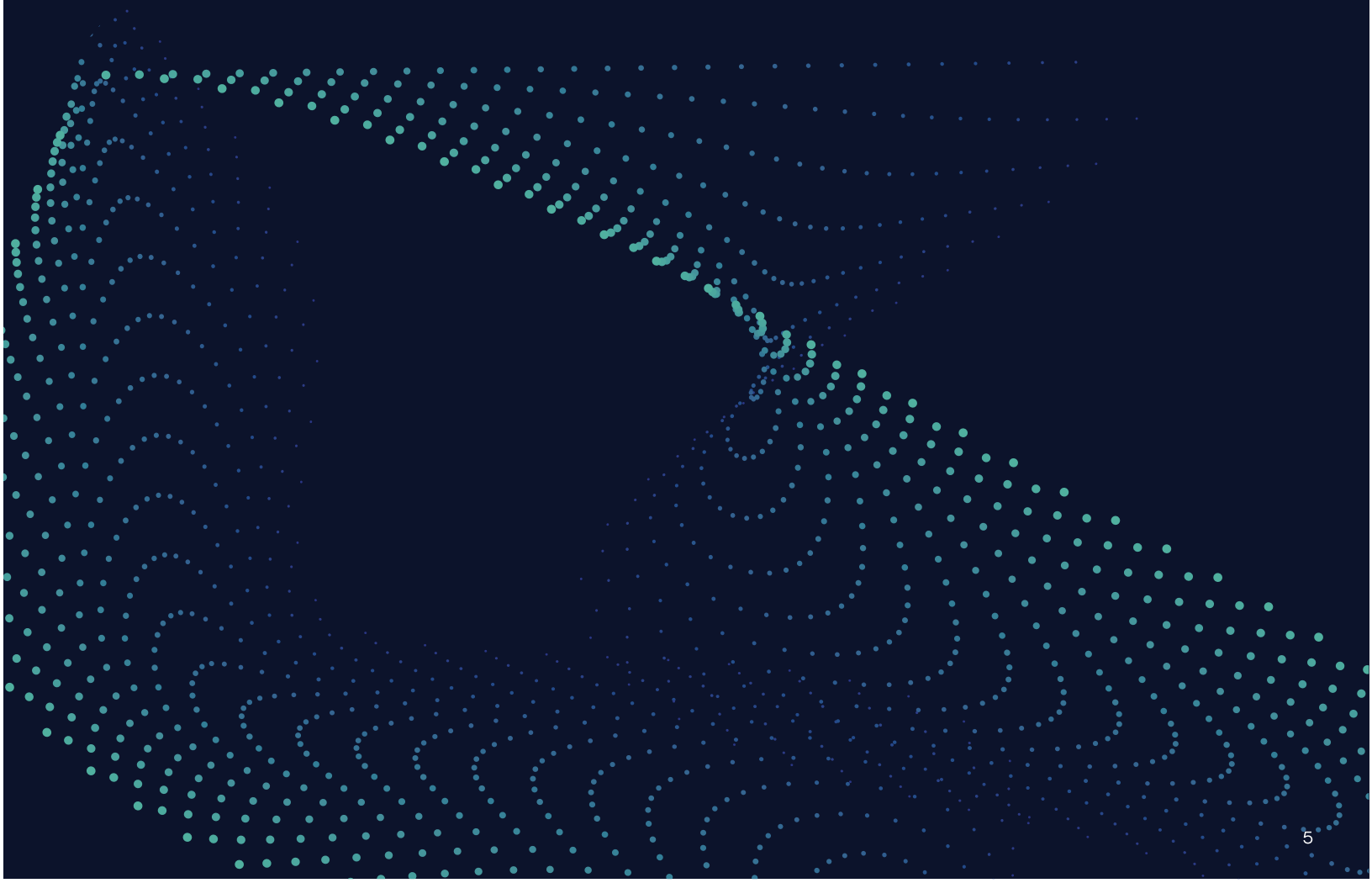
EXECUTIVE SUMMARY

We provide quantitative summaries of projects from the GenAI accelerator programs, encompassing a broad range of health use cases, as well as survey respondents' perspectives on priority use cases and health areas, and key factors and barriers for successful implementation of GenAI health interventions. Additionally, we profile five case studies of GenAI deployments with a specific focus on health-related behavior change in LMICs.

In terms of scale of deployment, we found many projects in the "pilot phase", including some that are deployed to over 10,000 monthly users. As of late 2024, we found only one application reaching scale (to 100,000 or more monthly users) of GenAI in health-related behavior change for LMICs, detailed in our included case studies. Several pilots had promising preliminary data on cost-effective impact, including health worker efficiency gains, and all are planning further evaluation in 2025 with a move to greater scaling. Pilots conducted as part of a broader scaled system (for example, an existing helpdesk workflow with millions of total users that is now testing integrating GenAI for efficiency improvements) have a more predictable path to fast scaling.

Given the nascency of the field, the relatively small scale of existing projects and limited evaluation data is unsurprising, but highlights the need for sustained focus to realize greater health impact.

While this review represents the most comprehensive analysis of GenAI in health-related behavior change to date, it is not exhaustive. Our findings focus on deployments funded by major GenAI accelerator programs and insights from expert interviews. While we believe this provides a representative snapshot of the current landscape, we recognize that some applications may not be captured. Additionally, this paper has not directly explored the perspectives of end users. We welcome input on additional large-scale deployments and user-centered insights to build collective knowledge in this evolving space.



HOW CAN WE REALIZE GREATER IMPACT AND UNLOCK THE FULL POTENTIAL OF GENAI?

Share learnings

Stakeholders wanted to learn more from others' experiences; this is especially important given how quickly technology and applications are evolving. Specific needs included: (a) understanding of the types of tasks LLMs are well suited to; their weaknesses; and strategies to address; and (b) summaries of specific successes, with concrete case studies reporting on comparable outcome metrics.

Strategies to address:

a. Produce practical guidance on how to identify LLM applications while mitigating risks and then pilot/validate/scale them. A regular update process will be required given technical capabilities are changing quickly.

b. Utilize consistent outcome metrics to describe scale of projects (such as monthly active users, total users and retention of users) and specificity regarding the type of AI system being used (for example deterministic vs. generative) to facilitate meaningful comparisons and benchmarking.

c. Establish a regular process to identify successes and disseminate learnings, including case studies.

Focus on actionable measurement

Stakeholders wanted better ways of measuring benefits, costs, and risks, in ways that provide rigorous but also timely data. For example, funders cannot wait 3 years for results of a randomized controlled trial to guide annual investment decisions, but we still need scientifically valid ways to measure success to inform implementation decisions in the interim. Establishing a clear evidence base will also be essential for supporting government decisions to implement successful applications at a national scale.

Strategies to address:

a. Establish standards for measurement and best practice, with concrete examples.

b. Identify opportunities for implementer partnership with academics on measurement. Since many projects are in pilot phase (and some are starting to scale), there is a time-sensitive opportunity for accelerative partnerships.

Improve language & localization

Experts noted that the quality of models varies by language, by medium (with voice particularly important for low-literacy settings) and by use case (e.g., health-specific contexts).The fact that large language models are not trained on or fluent in local languages was the most commonly selected barrier to using GenAI in healthcare settings in LMICs in our quantitative survey. We highlight the importance of identifying and closing gaps in quality as a key next step.

Strategies to address:

a. Establish standardized measures to evaluate model performance across different languages and specific health contexts to ensure consistent quality.

b. Curate high-quality datasets for underserved languages, including region-specific dialects, culturally relevant health information, and voice data for low-literacy populations.

Improve technical capacity & shared infrastructure

Experts noted that technical capabilities of GenAI implementers varied dramatically; similarly some funders and health system leaders identified gaps in their own knowledge that, if addressed, would allow them to make more impactful funding and procurement decisions. They also noted that some technical barriers (e.g., language models) likely would be better addressed centrally vs. in a fragmented way.

Strategies to address:

a. Identify elements of technical infrastructure that should be shared, and establish ways to centralize these efforts.

b. Provide technical capacity and consulting expertise to health system leaders, funders, and implementers.

Improve digital & basic health infrastructure

Throughout our research, the risk of inadvertently perpetuating the digital divide emerged as a key concern: no matter how advanced AI models and datasets become, their potential to effect behavior change is wasted if the people who need them most cannot access the necessary digital or physical infrastructure (for example, lack of stable internet connection or access to healthcare facilities recommended by GenAI chatbots).

Strategies to address:

a. Prioritize investment in basic healthcare infrastructure alongside digital interventions.

b. Consider whether GenAI is the highest impact way of addressing your use case, taking into account existing basic healthcare and digital infrastructure.

c. Evaluate an organization's digital readiness before deploying AI tools to avoid avoidable costly failures, and first focus funds on ensuring digital readiness where needed.