## Reviewer Criteria that we need to answer

* **Faculty merit:** Faculty is accomplished in research, community engagement, and open source contributions, with potential to contribute to responsible innovation.
* **Research merit**: Faculty's proposed research is aligned with Google Research interests, innovative, and likely to have a significant impact on the field.
* **Proposal quality**: The research proposal is clear, focused, and well-organized, and it demonstrates the team's ability to successfully execute the research and achieve a significant impact.
* **AI ethics principles**: The research proposal strongly aligns with [Google's AI Principles](https://ai.google/principles/).

## Past Google Work in this area:

[Emerging practices for Society-Centered AI](https://research.google/blog/emerging-practices-for-society-centered-ai/)

[5 products and features that make the digital world more accessible](https://blog.google/outreach-initiatives/accessibility/global-accessibility-awareness-day-google-product-update/)

### A very related application Google has funded: **Focusing on society’s health needs**

<https://nhsrcindia.org/sites/default/files/2024-01/Kilkari%20and%20Mobile%20Academy%20Evaluation%20Report%202022-23.pdf>

<https://armman.org/kilkari/>

Access to timely maternal health information can save lives globally: [every two minutes a woman dies during pregnancy or childbirth](https://www.who.int/news/item/23-02-2023-a-woman-dies-every-two-minutes-due-to-pregnancy-or-childbirth--un-agencies) and [1 in 26 children die before reaching age five](https://data.unicef.org/topic/child-survival/under-five-mortality/). In rural India, the education of expectant and new mothers around key health issues pertaining to pregnancy and infancy required scalable, low-cost technology solutions. Together with [ARMMAN](https://armman.org/), Google Research supported [a program](https://research.google/blog/using-ml-to-boost-engagement-with-a-maternal-and-child-health-program-in-india/) that uses mobile messaging and machine learning (ML) algorithms to predict when women might benefit from receiving interventions (i.e., targeted preventative care information) and encourages them to engage with the [mMitra](https://armman.org/mmitra/) free voice call program. Within a year, the mMitra program has shown a 17% increase in infants with tripled birth weight and a 36% increase in women understanding the importance of taking iron tablets during pregnancy. Over 175K mothers and growing have been reached through this automated solution, which public health workers use to improve the quality of information delivery.

These efforts have been successful in improving health due to the close collective partnership among the community and those building the AI technology. We have adopted this same approach via collaborations with caregivers to address a variety of medical needs. Some examples include: the use of the [Automated Retinal Disease Assessment](https://health.google/caregivers/arda/) (ARDA) to [help screen for diabetic retinopathy](https://blog.google/technology/health/5-myths-about-medical-ai-debunked/) in 250,000 patients in clinics around the world; our partnership with [iCAD](https://www.icadmed.com/) to bring our [mammography](https://blog.google/technology/ai/icad-partnership-breast-cancer-screening/) AI models to clinical settings to aid in breast cancer detection; and the development of [Med-PaLM 2](https://sites.research.google/med-palm/), a medical large language model that is now being [tested with Cloud partners](https://cloud.google.com/blog/topics/healthcare-life-sciences/sharing-google-med-palm-2-medical-large-language-model) to help doctors provide better patient care.

**Proposals should specifically cover one or more of the following topics:**

* Applications of AI with beneficial impact to underserved communities and/or that address societal-level challenges (e.g., applications to improve healthcare reach for underserved communities; upskilling for reemployment) [CR: Tagging this topic UNDERSERVED for later]
* Methods, tools and frameworks to promote stakeholder participation and alignment of AI to societal and/or community-specific needs (e.g., participatory AI or data governance frameworks; pluralistic AI alignment approaches; approaches for evaluating/mitigating AI harms) [CR: ALIGN]
* Studies of the touchpoints and integrations of AI and society (e.g., studies of current or emerging societal, economic, or cultural impacts of AI; studies of awareness, adoption, and attitudes towards AI) [CR: INTEGRATE]

**Prior Proposal:** [**MIT Solve Application**](https://docs.google.com/document/d/1KwHyqofp5EO0LjI02JPJGd27MUEu1qXGWaXBnQ0K97E/edit?usp=sharing)SM to edit in material from here.

Building Values-Aligned GenAI Systems to Enhance Last-Mile Maternal Healthcare Access in Rural India

**Introduction**

India experiences over 60,000 births daily and endures a maternal mortality rate of 100 per 100,000 births, resulting in four maternal deaths per hour during childbirth. Enhancing healthcare systems and improving digital health literacy through generative artificial intelligence (GenAI) can significantly benefit underserved communities by acknowledging cultural and psychosocial nuances, thereby enabling human-centric design of such systems.

Last-mile maternal healthcare in India is a multi-level, human-dependent system that relies heavily on Accredited Social Health Activists (ASHAs), Anganwadi workers, and Auxiliary Nurses and Midwives (ANMs) . These frontline workers deliver care to at-risk-women under various national programs (ASHs are recruited by the state under National Health Mission, Anganwadi workers are nationally recruited care workers under the Integrated Child Development Scheme, ANMs are trained to serve at public health institutions to support medical caregivers). Despite the existence of mobile app pilots like the initially well-received mSakhi application, scalability challenges remain. Other initiatives, such as chatbots for menstrual health, antenatal and postnatal care, and nutrition, have also been piloted. Notable examples include the voice-based pre-recorded calling system, mMitra, and the voice + messaging system, Kilkari, co-developed by Google and the national nonprofit ARMMAN. These systems aim to drive social and behavioral change by disseminating verified information through interactive mediums.

However, a significant limitation of these systems is their lack of personalization and inability to offer advice for scenarios not pre-coded into the system. Conversations with state program officers and private funders revealed that existing systems often need to be redesigned even to target different individuals within the same household. For instance, in rural Jalgaon, shared mobile phones mean that information delivered through calls, messages, and WhatsApp is often mediated by male family members, affecting its relevance and efficacy for the actual beneficiaries.

To address these challenges and improve healthcare information delivery, we propose employing user personas to deliver relevant information that supports maternal care in households. Our goal is to design and integrate a GenAI-based system into an evidence-based digital literacy intervention, demonstrating how to adapt healthcare tools to advance maternal care program delivery. This initiative aims to enhance awareness among local women about their maternal care responsibilities, childcare obligations, and eligibility for support programs throughout their pregnancy.

Objectives

1. Develop a Values-Aligned GenAI System: Create an AI-driven platform tailored to the cultural and psychosocial context of rural India, capable of delivering personalized maternal healthcare information.
2. Enhance Digital Health Literacy: Implement a digital literacy program to improve understanding and usage of the AI system among local women and healthcare workers.
3. Support Frontline Healthcare Workers: Provide ASHAs, Anganwadi workers, and ANMs with advanced tools to better serve at-risk women and improve maternal healthcare outcomes.
4. Drive Behavioral Change: Use personalized, relevant information to encourage social and behavioral changes that improve maternal and child health indicators in underserved communities.

Methodology

1. User-Centric Design: Conduct extensive field research to understand the needs, preferences, and cultural contexts of the target population. Develop user personas and scenarios to guide system design.
2. AI Development: Utilize generative AI techniques to create a system capable of providing personalized healthcare advice. Train the AI on a diverse dataset, including local language nuances and cultural specifics.
3. Integration and Pilot Testing: Integrate the AI system with existing healthcare infrastructure and pilot it in selected rural areas. Collect feedback from users and healthcare workers to refine the system.
4. Digital Literacy Program: Develop and implement a training program for local women and healthcare workers to enhance digital health literacy and effective use of the AI system.
5. Evaluation and Scaling: Monitor and evaluate the impact of the system on healthcare outcomes. Use data-driven insights to scale the system to other regions and healthcare domains.

Expected Outcomes

• Improved maternal healthcare access and quality in rural India.

• Increased digital health literacy among local women and healthcare workers.

• Enhanced support for frontline healthcare workers through AI tools.

• Demonstrable social and behavioral changes leading to better maternal and child health indicators.

Conclusion

This project aims to leverage GenAI to create a culturally and contextually relevant system that enhances maternal healthcare access in rural India. By addressing the limitations of existing systems and focusing on user-centric design and digital literacy, we aspire to drive meaningful improvements in healthcare outcomes for underserved communities.

Budget and Timeline

Budget:

• AI Development: $X,000

• Field Research and User-Centric Design: $X,000

• Integration and Pilot Testing: $X,000

• Digital Literacy Program: $X,000

• Monitoring and Evaluation: $X,000

Timeline:

• Month 1-3: Field research and user-centric design.

• Month 4-6: AI development and initial testing.

• Month 7-9: Integration with existing systems and pilot testing.

• Month 10-12: Implementation of digital literacy program.

• Month 13-15: Monitoring, evaluation, and refinement.

• Month 16-18: Scaling and broader implementation.

Team and Expertise

• Principal Investigator: [Your Name] - Expertise in xxx

• Co-Investigators: [Names] - Expertise in AI and healthcare system, public health, digital literacy, and community engagement. Behavioral Change, Underserved communities, Consumer Psychology, Human Judgment and Decision Making, Computational Social Psychology

• Technical Team: AI developers, data scientists, and user experience designers.

• Field Team: Local healthcare workers and community organizers.

Contact Information

• Principal Investigator: [Your Name]

• Institution: [Your Institution]

• Department: [Your Department]

• Email: [Your Email]

• Phone: [Your Phone Number]

By focusing on the unique cultural and psychosocial contexts of rural India, our proposal aims to create a sustainable, scalable solution that significantly improves maternal healthcare outcomes for underserved communities.

Table of Contents

1. The Problem
   1. Maternal healthcare system [must define the type of care and interaction paradigm explicitly, e.g. digital health literacy] providing care to mothers is ineffective in recurring services delivery at a consistent pace. [UNDERSERVED]
   2. Local level programs are challenging to track by state lawmakers. [INTEGRATE]
   3. Systems need to be redeveloped from scratch for adapting to multifaceted user contexts, sometimes even to target a different person in the same house, using the same device, and the same medium of communication. [ALIGN]
2. Existing Solutions and Gaps
   1. NGOs working with govts. can collect data but do not have the capacity to invest in user experience with the systems they deploy, resulting in a subpar effect on the beneficiaries and lack of impact from programs.
   2. The reality is a single phone is used in many households. The husband has the phone for a majority of the day. Phone is also shared with mother-in-law, children, and then the wife.
   3. Providing cultural- and individual-context-sensitive information is hard to scale (if it requires a retrain/RAG-deploy cycle every time new data from a different context becomes available). The system can also catastropically forget relevant information across redeployments.
      1. Segmentation: whom should information be tailored to reach, and what requirements does it need to satisfy.
      2. Privacy and Trust: people may be embarrassed to utilize the system and we need to understand if we can protect the privacy of the user.
         1. Potential solution: request users to self-delete messages if this is a shared device. Send a message at the end of their chat, perhaps?
         2. GDPR standpoint: what counts as “personally identifiable” is a range of information. If data is going to be input into a RAG system, that data needs to be anonymized; it cannot rely on PII for generating new responses in the future. Informed consent is critical!
            1. If we do check all these boxes, we can potentially use this data for other users as well.
   4. Evaluating whether the provided information is culturally-sensitive to the specific socio-economic stratum of society and individually appropriate is an open problem.
      1. What are the variables that define cultural sensitivity?
      2. How do we evaluate whether it is culturally sensitive?
         1. One way to think about it is binary evaluation: is a response tailored to a segment vs. not. Then further evaluate if it changes how that segment uses the product for those who receive tailored responses vs. those who don’t. One way to evaluate it can be rating at the end of a conversation.
         2. What dimensions is cultural sensitivity measured along?
      3. Language is correlated with culture, so there is a domain shift.
   5. **Lacking co-development for local populations.**
3. Solution
   1. Co-development of the platform needs to be clearly underscored!
      1. We have built prior solutions for this population
   2. Digital literacy platform that enables:
      1. disseminating information and collecting behavioral data through chatbots for end users [UNDERSERVED]
      2. identifying the social and individual contextual factors in which the interaction is embedded, towards generating appropriate responses for the entire family in a single AI system [INTEGRATE]
      3. human-AI coupling for lifelong social learning through interactions to adapt to unseen contexts in an already scarce-data regime [ALIGN].
4. Research Design:
   1. Building personas of users for chatbots to respond to and remember.
   2. Building query engine that identifies correlated questions and answers them consistently for a persona.
   3. Requirements gathering for the types of information delivery relevant for different user contexts (e.g. adolescents, teachers, parents, healthcare workers)
   4. Developing reliable evaluation criteria for evaluating cultural and individual appropriateness of responses and evaluating potential risks at the societal (e.g. stigma reinforcement) and individual (e.g. isolation, self-confidence issues) levels
   5. Collecting real-world data in a privacy sensitive manner towards developing lifelong social learning systems
   6. Personalizing the answering of questions on culturally sensitive matters. [I think actually achieving this within a year will be quite ambitious, although maybe we can have initial results from the collected data in the previous point.]
   7. ??
5. Study Size and Partner Background
   1. Aadhar Sanstha details and history of work for underserved populations in Jalgaon. Example of the other application we built for them to support healthcare data digitization in remote regions without internet access.
   2. What is the study size we want to propose? Do we want a power calculation included for effect size possible to identify? [I think we should, especially if we are trying to also validate some evaluation metrics through the pilots]
6. Adversarial Threats and Mitigations
   1. What are the risks involved in building out this system and how might users game it? What about breach of trust and data leakage? [Also call-back to evaluating risk at the societal and individual levels, scoped within the specific socio-economic stratum. See point 4d. where I’ve added some examples.]
   2. How do we mitigate these risks?
7. Ethics and Transparency
   1. Separate section not needed perhaps?
   2. Fold this into prior sections if necessary.

NM’s Feedback:

1. How do we define success?

Writing it up:

1. The Problem
   1. Past Sakhi Proposals.
   2. [MIT 100K Accelerate Proposal](https://docs.google.com/document/d/1IVmrRMHZ1TPg3VcD5K1-cLhlEE2ixLZ8lFxRg-U-XeY/edit)
   3. [MIT PKG Innovation Fellowship](https://docs.google.com/document/d/1bS_nv5EOgNAtdKCEBkoGTc-BAg2G2d3lfD-6Uhg57Eg/edit)
   4. [LLAMA Impact Grant Proposal](https://docs.google.com/document/d/1uG6PVf_i1gsYeLgjJm6st7il2SgbddXfQxadu0QHN6k/edit?usp=sharing)
2. Existing Solutions and Gaps
3. Solution
   1. Past Sakhi Proposals.
   2. [MIT 100K Accelerate Proposal](https://docs.google.com/document/d/1IVmrRMHZ1TPg3VcD5K1-cLhlEE2ixLZ8lFxRg-U-XeY/edit)
   3. [MIT PKG Innovation Fellowship](https://docs.google.com/document/d/1bS_nv5EOgNAtdKCEBkoGTc-BAg2G2d3lfD-6Uhg57Eg/edit)
   4. [LLAMA Impact Grant Proposal](https://docs.google.com/document/d/1uG6PVf_i1gsYeLgjJm6st7il2SgbddXfQxadu0QHN6k/edit?usp=sharing)
4. Research Design:
   1. Swapneel worked on this: [[Spreeha Fdn] One Fact Foundation - Services Agreement (One Fact receiving services).docx](https://docs.google.com/document/d/1_UodF66oUHPh-LLVx2p9TOzFGEVrla1n/edit#bookmark=id.yxzp695yx705)
   2. 2 groups:
      1. Treated (get app) - Baseline and Endline surveys
      2. Control (no app) - Baseline and Endline surveys
   3. NM Feedback:
      1. Randomization: Randomly assign the people you are recruiting to the treatment condition. Mention [fact-sheet](https://rchiips.org/nfhs/FCTS/MH/MH_FactSheet_499_Jalgaon.pdf) used for appropriate randomization.
      2. Confirm that the randomization scheme is unbiased.
         1. If randomized well, then baselines across treated and control should be the same.
      3. When running lab experiments, ensure that there is no spillover.
         1. People cannot have knowledge of other conditions and others who are in that condition.
         2. In the field, this will have to be carefully ensured (how will you stop people from talking about the app? Do we care about people finding out or not finding out about the app?)
         3. Location-based segregation into different conditions? Blocked randomization?
            1. Areas still need to be randomly assigned so that overall we don’t end up with only rural areas in one condition, urban areas in the other condition. What are the most important variables that would matter in this context?

Education, income level, family housing.

* + - * 1. Show that randomization is done uniformly on all these variables that could affect the treatment.
      1. e.g. NM helped run experiments to randomize across offices; Toronto treated, Vancouver untreated, etc.
    1. What does the app do?
       1. Takes input about their background and personal information (name, role in the household, and education level, age, gender, household income / easier to get no. of cows, horses, proxies to measure income / rooms in their house / whether you have a toilet)
       2. Goal: improve maternal mortality rate
       3. Theory of change: improving family literacy about antenatal care will reduce risks of mortality at childbirth. But why?
       4. Maternal deaths ← Preventable illnesses (anemia, syphilis, etc.) ← Lack of literacy ← App usage
          1. This is a different paper than saying Literacy ← App usage
       5. Assessment of antenatal care:
          1. Blood Pressure
          2. Urine testing
          3. Severe anemia
       6. But are these self-report-able even?
       7. Do we have these metrics available?
       8. **Are there any ways to assess ANC levels?**
       9. **If it is about literacy we want to assess not just whether people have *read* things but actually *learned* things?**
       10. **Do we also need an app for the control group – just without the messages. That app would just be a reminder app where they can report things (journaling app).**
       11. Allows them to ask thematic questions based on antenatal care:
           1. Family Health (focus on Wellbeing)
           2. Personal Health (focus on Anemia)
           3. Financial Support (focus on govt. welfare schemes)
       12. Generates contextually relevant responses based on user persona identified.

1. Study Size and Partner Background
2. Adversarial Threats and Mitigations
3. Ethics and Transparency (separate section optional)
4. Budget: 100K

Transatlantic collaboration with impact in India

12.5k per pilot \* 3 pilots = 37.5k

5k \* 2 travel costs = 10k

6k hosting and serving costs

10k research engineer

* SimPPL hires a research engineer

7k - EXTRA / BUFFER

8k - 10% indirect costs

3 sites, 1 organization

Diverse socioeconomic strata

1. Timeline

Is this here now a second attempt of the previous write-up and then later a third? That is very confusing. Can we only have on live-version we are editing?

Deploying Expert-guided AI Systems that Continuously Adapt to Consumer Behavior

India invests close to USD 4 billion a year in the National Health Mission in order to support women and child development schemes, a significant portion of which is intended to support a cadre of last-mile health workers, called the Accredited Social Health Activists (ASHAs), that deliver care directly to pregnant women in underserved communities and assist them in receiving medical checkups.

State-wide partnerships are designed to allow nonprofits to lead technological interventions to support the work of ASHA workers and advance health outcomes for underserved communities in rural India. Co-PI Raman is a Board Member at an organization called [SimPPL](https://simppl.org) that has worked with a nonprofit to deliver technological tools that digitize healthcare survey data collection without relying on internet access for 116 care workers serving 16 blocks in the Jalgaon district of Maharashtra. PI Mazar is an advisor to SimPPL Founder and postdoc, Dr. Swapneel Mehta, who will serve as a Research Engineer in this project.

Similar to many other rural areas, [Jalgaon](https://rchiips.org/nfhs/FCTS/MH/MH_FactSheet_499_Jalgaon.pdf) has a high prevalence of anemia (50%), and conversations reveal a lack of adequate antenatal care awareness among females aged 15-49.

We are interested in developing a digital literacy intervention including a behavioral reporting system that evaluates the effects of a multilingual WhatsApp chatbot trained to provide personalized guidance on dealing with maternal health issues.

[Previous mobile-based interventions](https://nhsrcindia.org/sites/default/files/2024-01/Kilkari%20and%20Mobile%20Academy%20Evaluation%20Report%202022-23.pdf) to advance awareness of best practices for maternal and child health have relied on voice recorded phone calls as reminders to improve antenatal care but suffer a number of challenges including the lack of continuous monitoring, evaluation, flexibility of the messaging medium, and most importantly, the limited personalization to the needs of the beneficiaries, as laid out in [national reports](https://nhsrcindia.org/sites/default/files/2024-01/Kilkari%20and%20Mobile%20Academy%20Evaluation%20Report%202022-23.pdf).

Our proposal arises from a past collaboration with the additional Municipal Commissioner of Jalgaon and Addl. CEO of the National Health Authority, IAS Dr. [Praveen Gedam](https://en.wikipedia.org/wiki/Praveen_Gedam), who is known for his strategies to advance technological adoption in this region.

The democratization and public release of large language models (LLMs) has accelerated the creation of specialized tools to serve as human decision support systems that have access to vast troves of knowledge across the internet [CITE Attention is all you Need, GPT-4].

Such systems have comprehensively demonstrated that they are able to generate answers to queries posed by users coming from different walks of life, across a range of tasks, surpassing human performance on a number of them [CITE LLM Benchmarks].

Due to this ability, they have been integrated into a number of chatbots, especially where cost-effectiveness is of paramount importance and they can provide immediate answers where human labor may have a higher cost as well as turnaround time.

Healthcare is one of the areas that is seeing a rapid adoption of such artificial intelligence (AI) chatbots in order to address basic patient queries.

Most often, such consumer-facing AI systems are designed in a manner that specializes a generalist base model by ‘adapting’ it to a set of tasks including potentially retraining it to perform better on the target task while retaining the insights gained from a broader semantic understanding of the knowledge space.

One goal of such a specialization is to limit the amount of incorrect information produced when responding to a user query, termed colloquially as ‘hallucination’.

For example, when answering queries about maternal health, a model trained on medical question-answering (MedQA) tasks will likely be more accurate than a model trained on data from the social media platform, Reddit [CITE MedQA].

However, the adoption of such technology is limited by the inability of most AI systems to adapt to the user that generates the query; for example, a medical chatbot gives similar answers with to patients that have vastly different health indicators even though their risk and requirement profiles may be different, since the information provided for RAG is the same in both cases.

Furthermore, the chatbot will be focused on the topics that have already been represented in the training dataset, with a limited ability to deliver personalized suggestions to the users; this creates a disparity against marginalized communities whose healthcare issues are seldom represented in mainstream datasets.

When deployed for inference tasks, LLM-driven systems are ‘frozen in time’; without retraining; even as tasks change, the model’s knowledge remains the same, and causes issues that we attribute to distribution shift, or the test time dataset being significantly different from the train time dataset.

For example, a model trained on Indian data until 2021 will be unable to answer questions about women and child development schemes launched in India in 2023.

The naive solution to such problems is to continuously retrain systems; more efficient solutions may be to retrain only following significant drops in metrics that evaluate performance on a sample from real time data streams; but ultimately such solutions still require expensive retraining on a fraction of new data which is not only expensive but also wasteful (it was months until a new version of chatGPT from OpenAI was released that could update its knowledge to a more recent ‘cutoff date’ until which it was trained).

Researchers evolved another ad hoc solution to fix this that conditionally generates responses based on retrieved information from a predetermined knowledge base creating the popular retrieval augmented generation (RAG) chatbot.

Given their ability to limit hallucination of information in answers, RAG LLM systems have rapidly become an industry standard to deploy consumer-facing LLM services.

They are employed in chatbots, automated calling, customer service functions, auto-responders, AI ‘agents’, enterprise data search, and other areas.

Another solution is to train model adapters that specialize a base model–say MedQA–on a number of different tasks with a set of weights learned per-task that can be applied only when answering queries relevant to that task–for example antenatal care (ANC) [CITE adapters].

Both adapters and RAG chatbots still suffer the limitation of generating ‘one-sided’ answers based purely on the knowledge provided to them, without personalizing the response to the end user.

This is a root cause of the lack of ‘cultural sensitivity’ and may be a reason for poor performance on ‘sociotechnical metrics’ by most chatbots.

In India, WhatsApp used on phones is accessible to women only at certain times in the day given that there is often a single phone in the household.

This is where Chirag adds in a section about human personas and LLMs adapting their response to the persona, requiring a change in AI systems and motivates continual learning.

Given the relatively rapid adoption of consumer-facing AI systems and their inability to adapt to human personas…

### Text from Chirag (incorporates text from Swapneel)

Notes from CR : So from the conversation over call, I find the two core research questions to be:

•⁠ ⁠how do we design a reliable framework for monitoring where the needle is wrt healthcare access in these communities

•⁠ ⁠how do we co-develop these AI systems along with the communities (here my ML opinion is we should be looking at lifelong learning and adaptation, I can expand on this technically)

—- Actual usable text:

**Grassroot Challenges and Issues.** The unique social dynamics governing technology adoption in the target underserved socio-economic sections of Indian society pose particular challenges for developing AI systems that can meaningfully drive digital healthcare literacy and behavior change. For instance, most families in rural Jalgaon use a shared mobile phone. Consequently, information delivered through calling, messaging, and WhatsApp usually passes first through the male members of the family to the others. In other cases, interaction time for different family members is rationed during specific times of the day. In the past 8 months working with a nonprofit, the Aadhar Bahuddeshiya Sanstha, and collaborating with the former Jalgaon municipal commissioner and Addl. CEO of the National Health Authority of India, we have identified the challenges for state-wide programs administered by nonprofits, supporting healthcare workers in the region. State program officers confirm a large number of them lack effective transparency and monitoring required to improve multi-stakeholder engagement

In conversations with multiple state program officers and private funders [cite], we learned that most systems have to be redesigned from the ground up even to target a different person in the same house, using the same device, and the same medium of communication. This gives rise to several challenges before an effective AI system can be designed and developed:

* *Identifying key user personas and the contextual information needed beyond facts*  
  An AI-based solution needs to support various user types, including, for instance, adolescents, parents, teachers, and ASHA workers. Each persona typically requires different contextual information: adolescents need sensitive health information, parents need guidance on support, teachers need structured educational materials, and ASHA workers need insights into community-specific social dynamics and stigma. Studies have shown that health information tailored to specific demographics enhances effectiveness (Glanz et al., 2008).
* *Lack of personalized adaptation for each individual within the user segment.*

Personalizing health information to meet the specific needs and contexts of individuals is crucial. Current systems often fail to dynamically adjust to an individual's unique circumstances and preferences. This is especially vital in shared usage contexts where multiple family members share a device to cater to each user's specific requirements*.*

* *Privacy and Trust*
  + Privacy Concerns: Ensuring data privacy in a context where multiple family members use a single device is inherently challenging. Systems must secure personal health information while offering granular privacy controls (Dove et al., 2017).
  + Building Trust: This involves transparent communication about data usage, benefits and limitations of AI recommendations. While trust can be built through community engagement and co-design processes involving the end-users (Gillespie et al., 2018), this is challenging in our target setting. Given the intermediary role of male family members, establishing trust requires engaging all family members in understanding and valuing the AI system's benefits and truly being informed when providing consent.
* *Reliable evaluation of personal-, social-, and cultural-sensitivity of generated responses.* 
  + Cultural Sensitivity: AI systems must generate responses sensitive to the user’s cultural and social context. This involves understanding and integrating local beliefs, practices, and languages into the chatbot’s responses (George et al., 2018). In underserved socio-economic sections of society where cultural norms strongly influence behavior, AI systems must respect and incorporate these norms to be effective.
  + Evaluation Metrics: Developing robust metrics to evaluate the sensitivity and appropriateness of AI responses in these settings is critical. Broad techniques have been proposed for this, including user feedback, qualitative assessments, and evaluating adherence to cultural norms (Kelly et al., 2019). However, adapting these to the stigmas and social dynamics, and mixed-device usage contexts guiding our target demographic remains an open challenge.

**Proposed Research and Outcomes.**

[Swapneel is going to discuss this with Nina. Explain the mixed-methods design including interviews and qualitative analysis and how this leads into the evidence-driven interventions goes here]

* Interviews and Qualitative Analysis: Conducting in-depth interviews with diverse stakeholders, including adolescents, parents, teachers, and ASHA workers. The qualitative data will help understand the nuanced needs and barriers faced by different personas.
* Evidence-Driven Interventions: Using insights from qualitative analysis to design targeted interventions. These will be pilot-tested, and iterative feedback will be incorporated to refine the interventions. Quantitative measures will assess the impact on health literacy, behavior change, and system usability.

**Long-term Outlook: Towards Agents with Lifelong Socially-Adaptive Capabilities.**

The insights gained from this project is meant to inform the subsequent development of AI agents with lifelong, socially-adaptive capabilities. Concretely, such agents would need to adapt to users over time. To be effective in dynamic and diverse user environments, the deployed AI systems must continuously adapt to changing user contexts. As users and communities engage with the system, the user needs, preferences, social norms evolve continuously to adopt the technology. Continuously adapting to these in deployment ensures the AI remains relevant and effective in providing personalized healthcare guidance over time. Research highlights the necessity of such adaptability for the long-term success of AI in health interventions (Topol, 2019).

Beyond the individual AI agent, the broader technological framework needs to incorporate all stakeholders so that the humans and AI agents can continually co-evolve. The qualitative analysis in this project through interviews to identify the relevant stakeholders and personas will serve to inform the logistics of deploying locally viable feedback loops and participatory design processes so that AI systems can evolve in ways that reflect the cultural and social contexts of the communities they serve.

REFERENCES

* Glanz, K., Rimer, B. K., & Viswanath, K. (2008). Health Behavior and Health Education: Theory, Research, and Practice. Jossey-Bass.
* Dove, E. S., et al. (2017). Beyond individual consent in precision health research. Nature Genetics.
* Gillespie, N., et al. (2018). Trust in AI: A Framework for Ethical Governance. IEEE Transactions on Technology and Society.
* George, A., et al. (2018). Cultural Sensitivity in AI: Framework and Methods. Journal of Cross-Cultural Psychology.
* Kelly, R., et al. (2019). Evaluating the Sensitivity of AI Responses in Healthcare. International Journal of Medical Informatics.
* Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine.
* El Ayadi, Alison M., et al. "Preliminary impact of an mHealth education and social support intervention on maternal health knowledge and outcomes among postpartum mothers in Punjab, India." Research Square (2023).
* Choudhury, Avishek, Onur Asan, and Murari M. Choudhury. "Mobile health technology to improve maternal health awareness in tribal populations: mobile for mothers." Journal of the American Medical Informatics Association 28.11 (2021): 2467-2474.
* Murthy, Nirmala, et al. "Effects of an mHealth voice message service (mMitra) on maternal health knowledge and practices of low-income women in India: findings from a pseudo-randomized controlled trial." BMC public health 20 (2020): 1-10.

**Title:** Improving Prenatal Health Literacy in Underserved Populations Using Personalized Messaging

**Abstract:** According to govt. Reports (NHSRC, 2023). We employ a WhatsApp chatbot to deliver control messages based on local learning modules created for ASHA workers under the National Rural Health Mission, and treated messages that are personalized to the context of the end user to evaluate the differences in adoption of the system between treated and control groups. The learning modules train ASHAs to deliver timely and relevant information manually to each of the beneficiaries that they are responsible for supporting. We measure adoption and behavioral change for the treated and control groups indicating the efficacy of the intervention. The system will employ LLMs trained using continual learning techniques that incorporate ASHA worker guidance and feedback to improve their relevance to the end user persona.

**Research Contributions:** There are a number of areas we focus on when building an artificial intelligence (AI) based solution:

Our research focuses on advancing health literacy in India by personalizing the delivery of mobile messages to urban and rural families. Our work seeks to advance health literacy to reduce the risk of preventable maternal deaths, improving existing mobile health literacy programs (El Ayadi et. al, 2023; Choudhury et. al, 2021). among our target population with the potential to scale into a state-wide partnership within six months of program completion. Building on work that shows how digital literacy interventions contribute to significant improvements in health outcomes, our work answers the question, **Does personalization of messaging improve the impact of mobile-based interventions to improve health literacy among rural and urban villages in Jalgaon, India**? We investigate whether **input from end users and care workers can be applied to adapt mobile messaging to increase the relevance of responses provided to beneficiaries**. We conduct a randomized controlled trial utilizing our existing WhatsApp-based chatbot to deliver a novel digital literacy intervention to end users and investigate how it can improve existing nationwide mobile-based interventions (Murthy et. al, 2020).It allows us to:

**Identifying key user personas and the contextual information beyond facts:** Support various user types, for instance, based on stages of pregnancy, health risks, age, gender, socioeconomic status, education level, and profession, each of whom requires different contextual information. Studies have shown that health information tailored to specific demographics enhances effectiveness (Glanz et al., 2008). **Personalize the adaptation for each individual within the user segment:** Personalize health information to meet the specific needs and contexts of individuals. Current systems often fail to dynamically adjust to an individual's unique circumstances and preferences. This is especially vital in shared usage contexts where multiple family members share a device to cater to each user's specific requirements*.*

**Improve Privacy and Trust:** Privacy Concerns: Ensuring data privacy in a context where multiple family members use a single device is inherently challenging. Systems must secure personal health information while offering granular privacy controls (Dove et al., 2017).

**Building Trust**: This involves transparent communication about data usage, benefits and limitations of AI recommendations. While trust can be built through community engagement and co-design processes involving the end-users (Gillespie et al., 2018), this is challenging in our target setting. Given the intermediary role of male family members, establishing trust requires engaging all family members in understanding and valuing the AI system's benefits and truly being informed when providing consent.

**Current Challenges:** As one of the countries with the highest prevalence of maternal deaths India invests over USD 4 billion annually into healthcare services and national welfare programs to create an expansive system of last-mile care delivery at the village level. With a focus on health literacy improvement and care provision, trained accredited social health activists (ASHAs) offer frontline support maternal and child health programs at the rural and urban levels through personal visits to their beneficiaries. States actively engage in partnerships with nonprofits to launch mobile applications complementary to in-person programs, and a National Health Systems Resource Center (NHSRC) [report](https://nhsrcindia.org/sites/default/files/2024-01/Kilkari%20and%20Mobile%20Academy%20Evaluation%20Report%202022-23.pdf) discusses challenges with multi-state, mobile-based, awareness-focused interventions. It underscores the lack of long-term impact achieved at local and national levels, attributing them in part to generic content affecting compliance, lack of adaptability, and hyper focused design that limits their adaptability. Most families in rural Jalgaon use a shared mobile phone. Consequently, information delivered through calling, messaging, and WhatsApp tends to pass first through the male members of the family to the others. In other cases, interaction time for different family members is rationed during specific times of the day. We further verified the challenges with over 30 interviewees at the national, state, district, and block levels to confirm they all faced the **lack of personalized responses.** Programs are created specific to a single end user while mobile devices are often shared by family members for whom the messaging feels generic and irrelevant. This results in noncompliance, reduced trust, and dropoffs.

**Adapting Interventions**

**Interviews and Qualitative Analysis**: We will conduct in-depth interviews with diverse stakeholders, including adolescents, parents, teachers, and ASHA workers. The qualitative data will help understand the nuanced needs and barriers faced by different personas.

**Evidence-Driven Interventions**: We will use insights from qualitative analysis to design targeted interventions such as the consenting use of end user data to target interventions to persona types. These will be pilot-tested, and iterative feedback will be incorporated to refine the interventions. Quantitative assessments will compare the group-wise differences in adoption, health literacy, and participant dropoff following the cessation of incentives provided to end users.

**Baseline and Endline Surveys**

We will randomize the allocation of 1000 study participants into a single-arm treatment and control group, where WhatsApp message personalization is applied only to the treated group while others receive existing messaging around maternal care drawn from past learning modules including [Mobile Academy](https://armman.org/mobile-academy/). Personalization is based on data collected from interviews and qualitative analysis that will be used to adapt the LLM responses generated to the beneficiary persona without retraining the LLM. We will provide 120 days of continuous access to the system along with usage-based financial incentives to improve compliance by promoting app usage for both groups. We plan to conduct baseline (pretreatment) and endline surveys for both groups that receive access to the system in order to account for any sampling biases, especially those based on location, socioeconomic status, The Aadhar Bahuddeshiya Sanstha, our partner in ongoing healthcare projects, has established significant health and monitoring infrastructure locally, operating a network of 116 care workers across 16 blocks and hundreds of villages in Jalgaon. We have supported care workers in the region to digitize their offline data collection in remote regions that eliminates the downstream data entry task of up to 17 handwritten pages per family visited.

**Long-term Outlook: Towards Agents with Lifelong Socially-Adaptive Capabilities.**

The insights gained from this project are meant to inform the subsequent development of AI agents with lifelong, socially-adaptive capabilities. To be effective in dynamic and diverse user environments, the deployed AI agents must continuously adapt to changing user contexts. As users and communities engage with the system, the user needs, preferences, social norms evolve continuously to adopt the technology. Continuously adapting to these in deployment ensures the AI remains relevant and effective in providing personalized healthcare guidance over time. Research highlights the necessity of such adaptability for the long-term success of AI in health interventions (Topol, 2019). Beyond the individual AI agent, our broader technological framework aims to incorporate all stakeholders so that the humans and AI agents can continually co-evolve. The qualitative analysis in this project through interviews to identify the relevant stakeholders and personas will serve to inform the logistics of deploying locally viable feedback loops and participatory design processes so that AI systems can evolve in ways that reflect the cultural and social contexts of the communities they serve.

**Budget (Total = USD 100,000)**

| **Name** | **Description** | **Cost** | **Units** | **Total** |
| --- | --- | --- | --- | --- |
| Nonprofit Resources and Staff per Pilot Study | Cost to deploy and monitor pilot experiment for 350 users per site | 14,500 | 3 | 43,500 |
| Travel to Site | Includes 2-way airfare and local transportation to reach site | 5,000 | 2 | 10,000 |
| Cloud Computing | Host servers and obtain LLM-generated responses | 8,000 | 1 | 8,000 |
| Hardware | Purchase local mobile phones for testing chatbot performance | 1,000 | 2 | 2,000 |
| Research Engineer | System development and observability infrastructure | 19,500 | 1 | 19,500 |
| Project Lead | Project implementation and reporting | 17,000 | 1 | 17,000 |

**Timeline**

| **Activity** | | | | Q4 2024 | | | Q1 2025 | | | Q2 2025 | | | Q3 2025 | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Recruit Participants | | | |  |  |  |  |  |  |  |  |  |  |  |  |
| Test WhatsApp Chatbot Prototype | | | |  |  |  |  |  |  |  |  |  |  |  |  |
| Conduct and Analyze Baseline Surveys | | | |  |  |  |  |  |  |  |  |  |  |  |  |
| Implement Adaptive Messages Based on Previous Analysis | | | |  |  |  |  |  |  |  |  |  |  |  |  |
| Deploy Chatbot Intervention (120 days) | | | |  |  |  |  |  |  |  |  |  |  |  |  |
| Conduct and Analyze Endline Surveys | | | |  |  |  |  |  |  |  |  |  |  |  |  |
| Synthesize Report | | | |  |  |  |  |  |  |  |  |  |  |  |  |

REFERENCES

National Health Systems Resource Centre (NHSRC), “Evaluation of Kilkari and Mobile Academy Project”, <https://nhsrcindia.org/sites/default/files/2024-01/Kilkari%20and%20Mobile%20Academy%20Evaluation%20Report%202022-23.pdf>

Glanz, K., Rimer, B. K., & Viswanath, K. (2008). Health Behavior and Health Education: Theory, Research, and Practice. Jossey-Bass.

Dove, E. S., et al. (2017). Beyond individual consent in precision health research. Nature Genetics.

Gillespie, N., et al. (2018). Trust in AI: A Framework for Ethical Governance. IEEE Transactions on Technology and Society.

George, A., et al. (2018). Cultural Sensitivity in AI: Framework and Methods. Journal of Cross-Cultural Psychology.

Kelly, R., et al. (2019). Evaluating the Sensitivity of AI Responses in Healthcare. International Journal of Medical Informatics.

Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine.

El Ayadi, Alison M., et al. "Preliminary impact of an mHealth education and social support intervention on maternal health knowledge and outcomes among postpartum mothers in Punjab, India." Research Square (2023).

Choudhury, Avishek, Onur Asan, and Murari M. Choudhury. "Mobile health technology to improve maternal health awareness in tribal populations: mobile for mothers." Journal of the American Medical Informatics Association 28.11 (2021): 2467-2474.

Murthy, Nirmala, et al. "Effects of an mHealth voice message service (mMitra) on maternal health knowledge and practices of low-income women in India: findings from a pseudo-randomized controlled trial." BMC public health 20 (2020): 1-10.

[Digital interventions improve maternal health outcomes in india - Manyata For Mothers](https://www.manyataformothers.org/blog/digital-interventions-improve-maternal-health-outcomes-in-india/)

[Mobile health technology to improve maternal health awareness in tribal populations: mobile for mothers - PMC](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8510310/)

[Preliminary impact of an mHealth education and social support intervention on maternal health knowledge and outcomes among postpartum mothers in Punjab, India](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10775402/)

**Comments:**

**NM**

1. The scope is really broad, setting this up will take a lot of time!
2. Criticism of other programs re: “needle moves back”, is unjustified. Those programs could be like alarm clocks, where the removal of the intervention is perfectly fine to let the needle move back.
3. Part I: Personalization; Part II: Use the personalized system!
   1. Limit the part that discusses cultural sensitivity
   2. Focus on immediate experiment “deploy the personalized system to test its real-world utility” before jumping topics to a new area
   3. How easy is it to use this system?
   4. Has the system caused any issues within your family?
   5. Compare personalized vs. non-personalized.
   6. Do they trust it more, use it more, more likely to recommend it to their friends?
   7. Even when you stop paying them an incentive to use it, they drop off on the non-personalized arm but personalized arm is continued to be used
   8. THEN we can say this is increasing engagement, and continues to have an impact
   9. NOW we can deploy it, compare it with others and then have mothers use it for 9 months and then measure and monitor the health indicators. **BUT this is not in scope of the current experiment**.
   10. Then we can discuss important questions, spillover, healthcare workers, sharing, etc.
   11. Even without this, we have a sufficiently expansive grant proposal!
       1. “We want to do things right, and before even getting to the step where we are deploying chatbots to improve welfare, we want to test that this system materially improves their experience compared to un-personalized systems”
       2. Motivation: we believe personalization makes a difference in these tools for welfare, engagement, satisfaction, and trust. Can we get people excited about personalized systems?
       3. The ethical way to deploy AI systems is to incorporate local communities at the first phase!

* + 1. Nina got a healthcare-focused Pioneering Ideas grant from a foundation; the entire purpose is to use random forests over existing data from an experiment to assess which treatment would have been the best to assign to each person to increase vaccination uptake. The grant was just for the validation phase and was written such to show that validation stage is important to have evidence to then find collaborator for second phase which would be to do a personalized treatment assignment experiment. So, we asked for money for phase I motivating it by saying that it will help to secure interested parties for stage II, and it worked. The granting agency thought it is worthhile to fund such an effort. Point it: Phase

I should be enough!

We have developed and deployed a chatbot in Bangladesh for addressing menstrual health and hygiene among young girls and women in rural and urban slums in Dhaka and Chittagong. The system has received support from MIT's IDEAS Social Innovation Challenge through August 2024 to run pilots as well as their Delta V educational accelerator program. The conversations we have had with over 35 unique officers and nonprofit representatives have made it clear that there are systemic issues with existing technologies that our solution fixes. For this work, we propose a novel, maternal health chatbot but we have significant prior experience building such systems.

We have developed and deployed a chatbot in Bangladesh for addressing menstrual health and hygiene among young girls and women in rural and urban slums in Dhaka and Chittagong. The system has received support from MIT's IDEAS Social Innovation Challenge through August 2024 to run pilots as well as their Delta V educational accelerator program. The conversations we have had with over 35 unique officers and nonprofit representatives have made it clear that there are systemic issues with existing technologies that our solution fixes. For this work, we propose a novel, maternal health chatbot but we have significant prior experience building such systems. Furthermore, PI Nina Mazar has engaged in such projects before. She won a Robert Wood Johnson Foundation, Pioneering Ideas, $260,000, 9 months, as the sole PI in order to implement personalized treatments on a pre-collected dataset..

This project has received interest for two state-wide pilots that will be launched if we can win the grant funds to support it. We have letters of support to indicate commitment from nonprofits that want to partner with us and we have raised funds to launch other pilots before. There is a lot of work to be done and we are the ones to do it, but we really need the funding and support. We are actively applying to other avenues to secure funding for this work. IRB approvals have been received for data analysis from the chatbot and have a turnaround time of ~4 weeks.