



SPECIAL TOPIC ARTICLE

Expanding impact of mobile health programs: SAHELI for maternal and child care

Shresth Verma¹ | Gargi Singh¹ | Aditya Mate^{1,2} | Paritosh Verma^{1,3} |
 Sruthi Gorantla^{1,4} | Neha Madhiwalla⁵ | Aparna Hegde⁵ | Divy Thakkar¹ |
 Manish Jain¹ | Milind Tambe¹ | Aparna Taneja¹

¹Google Research India, Bangalore, India

²Harvard University, Cambridge, Massachusetts, USA

³Purdue University, West Lafayette, Indiana, USA

⁴Indian Institute of Science, Bangalore, India

⁵ARMMAN, Mumbai, India

Correspondence

Aparna Taneja, Google Research India, Bangalore, India.

Email: aparnataneja@google.com

Shresth Verma and Gargi Singh contributed equally to this study.

Aditya Mate, Paritosh Verma, and Sruthi Gorantla had done the work during an internship at Google Research India.

Abstract

Underserved communities face critical health challenges due to lack of access to timely and reliable information. Nongovernmental organizations are leveraging the widespread use of cellphones to combat these healthcare challenges and spread preventative awareness. The health workers at these organizations reach out individually to beneficiaries; however, such programs still suffer from declining engagement. We have deployed SAHELI, a system to efficiently utilize the limited availability of health workers for improving maternal and child health in India. SAHELI uses the Restless Multi-armed Bandit (RMAB) framework to identify beneficiaries for outreach. It is the *first deployed application* for RMABs in public health, and is already *in continuous use* by our partner NGO, ARMMAN. We have already reached ~ 130K beneficiaries with SAHELI, and are on track to serve one million beneficiaries by the end of 2023. This scale and impact has been achieved through multiple innovations in the RMAB model and its development, in preparation of real world data, and in deployment practices; and through careful consideration of responsible AI practices. Specifically, in this paper, we describe our approach to learn from past data to improve the performance of SAHELI's RMAB model, the real-world challenges faced during deployment and adoption of SAHELI, and the end-to-end pipeline.

INTRODUCTION

Mobile health (mHealth) programs, that leverage the widespread use of cellphones, are a crucial resource for bridging information inequities for underserved and marginalized communities in the global south (Tshikomanana and Ramukumba 2022; Gupta et al. 2022), especially in areas such as public health and social services where access to authoritative information is unevenly distributed.

Many nongovernmental organizations (NGOs) periodically send automated voice messages to improve health outcomes of beneficiaries. However, in spite of high adoption, adherence is a key challenge in public health information programs ARMMAN (2019), Jakob et al. (2022), Eysenbach (2005), Meyerowitz-Katz et al. (2020). NGOs often employ live service calls made by health workers to boost engagement via encouragement or through logistic changes requested by beneficiaries. However, given the

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *AI Magazine* published by John Wiley & Sons Ltd on behalf of Association for the Advancement of Artificial Intelligence.



States Covered in India	19
Partner NGOs	40
Partner Hospitals	97
Health Workers Trained	235K
Beneficiaries	27.2M

Scale of ARMMAN

FIGURE 1 A beneficiary receiving preventive health information.

comparatively large number of potential beneficiaries, it is important to maximally utilize the limited availability of health workers, and thus it is crucial to identify the best recipients for such service calls (see Figure 1).

While AI models can help health workers in optimizing their service calls, deploying these models in the context of mHealth programs for underserved communities presents unique challenges. First, available data are sparse and skewed due to small numbers of service calls. Second, NGOs are constrained by a very limited compute budget. Third, responsible deployment of the AI models is particularly important in such settings.

In this paper, we show how we address these research challenges in our deployed AI model—a deployed Restless Multi-Armed Bandits (RMAB) model for public health—together with our NGO partner ARMMAN (2008) to help improve the quality of service of their mHealth program focusing on maternal and child care. India suffers from high maternal and neonatal mortality rates Meh et al. (2022), World Health Organization (WHO) (2020), and ARMMAN (2008) runs one of the largest mHealth programs in this domain in India. Our system, SAHELI (System for Allocating Healthcare-resources Efficiently given Limited Interventions), is the result of deep partnership of an interdisciplinary team of researchers. SAHELI (meaning “female friend” in Hindi) is designed to assist, rather than substitute, health workers in their normal workflow. The following are the key contributions of deployed SAHELI: SAHELI includes the first deployed application of RMABs for public health, and it is continuously in use by our partner NGO ARMMAN; a key novelty of the deployment is that it both predicts RMAB model parameters and computes optimal policies; in contrast with most past research that has focused on computing optimal policies. To that end, we provide an improved and robust machine learning prediction framework by performing model selection and evaluation of real-world RMAB systems; we deployed SAHELI on cloud infrastructure with an emphasis on frugality throughout the end-to-end pipeline given the resource constraints of the NGO partner; we present

responsible AI practices to address ethical considerations for deploying an AI system for impact in underserved communities, particularly in this nonwestern context.

SAHELI has been developed as a platform, with the ability to be scaled to more NGOs in more domains. Our source code and data dictionary are available on Github¹.

RELATED WORK

While several works in the healthcare domain have studied patient adherence for diseases like HIV Tuldrà et al. (1999), cardiac problems Son et al. (2010), Corotto et al. (2013), and tuberculosis Killian et al. (2019), Pilote et al. (1996), these largely focus on building machine learning classifiers to predict future adherence to prescribed medication. With such models, the pool of beneficiaries flagged as “high-risk” can itself be very large. Furthermore, the one-shot predictions of these models fail to capture the sequential decision making aspect of the problem. Other approaches that consider sequential decision making challenges, such as Pollack et al. (2002), Liao et al. (2020), Brisimi et al. (2018) adopt reinforcement learning techniques to build personalized health monitors that can send timely notifications or activity suggestions to users. However, these models assume that notifications can be sent at will, and as such, do not address the challenge of limited service call resources.

Alternatively, RMABs have seen significant theoretical investigation, motivated by resource allocation challenges, such as in antipoaching patrols Qian et al. (2016), multichannel communication Liu and Zhao (2010), sensor monitoring and machine maintenance tasks Glazebrook, Ruiz-Hernandez, and Kirkbride (2006). While they provide important contributions, none of these works have seen a real-world deployment, and most have not been field tested.

Key reasons for the lack of RMAB deployment are their significant computational and data requirements. For example, just the optimization problem of computing the optimal allocation π , while assuming that the transition parameters \mathcal{P} are available is already known to be PSPACE-hard Papadimitriou and Tsitsiklis (1999). Furthermore, in the real world, these transition parameters are not just unknown but also hard to infer for real beneficiaries enrolling with ARMMAN and other similar health programs, as they come with no historical transition data. Despite such difficulties, our work is the first to deploy RMABs in tackling a real-world maternal healthcare task via frugal design choices discussed below.

PROBLEM INTRODUCTION

ARMMAN is a nongovernmental nonprofit organization based in India, focused on improving maternal and child health outcomes among underserved and underprivileged communities ARMMAN (2008). Their flagship program, “mMitra,” is a mHealth service that aims to leverage the extensive cellphone penetration in India to send out critical preventive health information to expectant or new mothers via automated voice messages. A large fraction (~90%) of mothers in the mMitra program are below the World Bank international poverty line World Bank (2020). Despite the acute economic disadvantages faced by these mothers, such automated voice messages prove to be a feasible mode of information dissemination at scale, thanks to the wide accessibility of low-cost phones.

After enrollment into the mMitra mHealth program, beneficiaries receive 1–2-min voice messages with health information according to beneficiary’s gestational age or age of the infant. Unfortunately, despite the proven effectiveness of this information program in improving maternal health outcomes, ARMMAN often sees dwindling engagement rates among beneficiaries, including frequent dropouts. Around 22% of beneficiaries dropout of the program after just 3 months. To counter this issue, ARMMAN leverages health workers that place live service calls (phone calls) to a limited number beneficiaries on a weekly basis to encourage beneficiaries’ participation, address requests/complaints, and attempt to prevent engagement drops. This limitation in number of service calls raises the key question of deciding which beneficiaries to pick for live service call in order to improve engagement rates among the beneficiaries.

RESTLESS MULTI-ARMED BANDITS (RMAB)

The RMABs model was first introduced by Whittle (1988) to address limited resource allocation problems, but has not received much attention in terms of real-world deployments. An RMAB consists of a set of N arms, where each arm is associated with a two-action MDP Puterman (2014). An MDP $\{S, \mathcal{A}, r, P\}$ consists of a set of states S , a set of actions \mathcal{A} , a reward function $r : S \times \mathcal{A} \times S \mapsto \mathbb{R}$, and a transition function P , where $P_{s,s'}^\alpha$ is the probability of transitioning from state s to s' when action α is chosen. The reward function in our setup is given as $r(s, \alpha, s') = s'$. An MDP policy $\pi : S \mapsto \mathcal{A}$ maps to the choice of action to take at each state. The long-term discounted reward for a policy π , starting from state $s_0 = s$ is defined as $R_\gamma^\pi(s) = E[\sum_{t=0}^{\infty} \gamma^t r(s_{t+1}) | s_0 = s]$ where $s_{t+1} \sim P_{s_t, s_{t+1}}^{\pi(s_t)}$ and

$\gamma \in [0, 1)$ is the discount factor. The total reward in the RMAB is defined as the sum of the total rewards accrued by individual arms of the RMAB.

In the setup we consider, each arm of the RMAB models a beneficiary enrolled with ARMMAN, who can be in one of two states $S = \{0, 1\}$, corresponding to “Not Engaging (NE)” and “Engaging (E),” respectively. Engagement in our setup was defined in consultation with the subject matter experts at ARMMAN: we define a beneficiary as engaged when she listens to at least one call in a week for more than 30 s. The action space for each arm consists of two actions, $\mathcal{A} = \{0, 1\}$, where 1(0), typically called the active (passive) action, refers to selecting (not selecting) the beneficiary for the live service call. Beneficiaries may transition from say their E state to NE state from 1 week to the next week based on their transition probabilities defined on passive or active actions. The planner’s goal is to select actions on arms, that is, deliver live service calls, so as to maximize the total reward, which is the number of beneficiaries in the engaged state, accrued by the RMAB. However, the budget constraint demands that the planner can choose no more than k arms ($k \ll N$) for the active action at any given timestep, that is, no more than k live service calls per week.

The dominant technique for solving RMABs uses the Whittle Index heuristic Whittle (1988), which is shown to have asymptotic optimality under some conditions Weber and Weiss (1990), and to provide excellent performance in practice Qian et al. (2016). Whittle indexes are formulated using the idea of passive subsidy, and informally rank arms so as to choose the top k , based on how attractive it is for a planner to activate each arm. For computing Whittle index, we use binary search algorithm from Qian et al. (2016)

Previous study

Our previous study conducted in April 2021 Mate et al. (2022) is the first to present real-world service quality improvement using RMABs in the context of mMitra program. This study tested an RMAB-based policy against two baselines of interest, and showed RMAB outperforming its competitors. The study spanned 7 weeks and included 23,003 real-world beneficiaries who were distributed in three groups corresponding to the RMAB policy, round robin (RR), and current standard of care (CSOC). Whereas RR corresponds to a non-AI heuristic for systematically calling beneficiaries, CSOC did not call any individuals. The results from this pilot study are shown in Table 1.

The pilot results demonstrated that the RMAB method cuts ~30% of the beneficiary engagement drops experienced by the other groups. Furthermore, whereas RMAB

TABLE 1 RMABs demonstrate statistically significant superior performance when compared against other non-AI approaches, namely current standard of care (CSOC) and round robin (RR), as showed by Mate et al. (2022).

Improvements	RMAB over CSOC	RMAB over RR	RR over CSOC
% Reduction in total beneficiary engagement drops	32.0%	28.3%	5.2%
<i>p</i> -value estimate	0.044	0.098	0.740

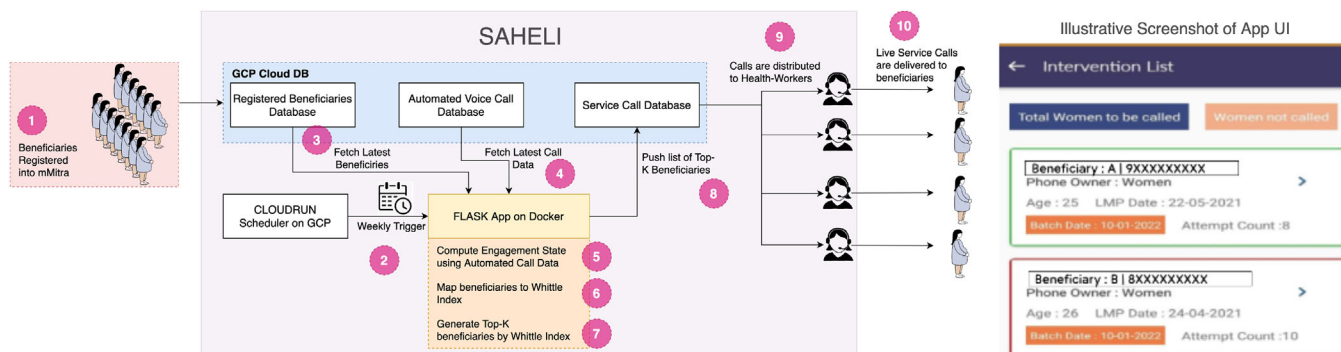


FIGURE 2 Pipeline of deployed system. Beneficiary information on app UI is available only to the health worker in charge.

achieves statistically significant improvement against CSOC ($p < 0.05$) and RR ($p < 0.1$), RR fails to achieve any statistically significant improvement over CSOC (see Mate et al. 2022 for more details). This key result forms the basis of relying on RMAB-based strategy over other non-AI strategies as a basis of SAHELI. In this paper, we describe the journey from this initial study to the final deployment. Whereas we use the same overall RMAB learning and optimization approach, we made multiple changes to provide significant enhancements that reduce data anomalies and improve computational performance of this RMAB-based strategy. Additionally, our deployed cloud application now automates the data exchange process with the NGO's systems while requiring minimal compute resources to be feasibly handled by the NGO. We now describe the end-to-end SAHELI system.

DEPLOYING SAHELI

We now introduce SAHELI and its architecture. We begin by discussing the different components, and follow that up with the description of the AI pipeline. We then discuss the frugal design choices—both in modeling and infrastructure—that were required to finalize the deployment.

System architecture

We first describe all the interactions within SAHELI's ecosystem (refer Figure 2). The health workers in the field

periodically register beneficiaries through door-to-door visits or at the hospitals (step 1). The socio-demographic data such as age, language, income range, as well as information on gestational age is then entered into the database maintained by ARMMAN (step 3). Automated voice messages tailored to the beneficiaries' gestational age are sent with the help of a telecommunication provider (step 4). The meta-data of the outcome such as duration of the call, failure reason, and so forth, is also pushed to ARMMAN's database. As beneficiaries' engagement with the voice messages diminishes over time, live service calls are made by ARMMAN to encourage beneficiaries to engage with the program (step 10). However, due to limited resources on the NGO's side, only a limited number of live service calls can be made each week. The AI pipeline predicts which beneficiaries would benefit most from receiving a service call in any given week. This list of beneficiaries is then generated at the start of each week and distributed across health workers in an automated fashion as shown on Figure 2 in steps through 2–9.

The AI pipeline (described in the next section) for a dynamically growing population is deployed on infrastructure hosted on Google Cloud Platform (GCP). The AI pipeline is wrapped as an application using Flask, which is containerized using Docker. The docker image is created to contain the requisite code scripts for the AI pipeline with apt environment requirements. Our default GCP container settings are to use 6 vCPUs and 16GiB memory. A weekly scheduler job on GCP triggers the Flask application, which then generates the list of beneficiaries.

Step 8 in Figure 2 shows the generation of the list of beneficiaries that should be intervened in the given week

using the AI pipeline. This list is ingested in ARMMAN's cloud databases, which serve as the back-end of a client mobile application (screenshot provided in Figure 2) used by the health workers. This client application randomly distributes the list of scheduled service calls among health workers based on their weekly availability. An illustrative screenshot (not real beneficiary) is also shown in Figure 2. The health worker sees a list of beneficiaries that he/she can call, along with certain features like number of call attempts. They can also click on a particular beneficiary and see more information about the beneficiary and past calls with them (not shown). The calls are made through the week with a maximum of three call attempts to the same beneficiary. All the beneficiaries in the generated list receive the aforesaid service calls. The model is currently providing services to beneficiaries enrolling at an average rate of 20K beneficiaries per month with a budget of 1000 calls per week.

SAHELI streamlines the entire deployment workflow in a singular pipeline, and automates its orchestration and execution, making this process computationally efficient, cost-effective, and easy to debug. As more beneficiaries get enrolled periodically, the beneficiary cohort in the application can now be updated automatically.

Health workers can then make the calls (step 10 in Figure 2) to these beneficiaries motivating them to listen to the voice messages and address any logistic issues (e.g., time slots, language of communication, and others) that might be affecting their engagement. As we show later in the paper, motivating the beneficiaries is a key to driving adherence. However, it bears repeating that given the limited availability of the health workers, they can only make a limited number of calls. In our AI pipeline, we focused on identifying the right set of beneficiaries to call, and not on automating the contents of the service call. *This is a key design choice in SAHELI*: we, thus, complement the human-to-human engagement between the health worker and the beneficiary, and together they contribute towards aiding a particular beneficiary and driving higher engagement with the mHealth program. This model of working together with the health workers embodies ARMMAN's core "tech plus touch" philosophy ARMMAN (2008) and is essential to our successful outcomes.

Pipeline description

This section describes the modules in the AI pipeline for both the offline model training and the online model execution. The offline model creation begins with the processing of the training data (i.e., historic data from past mHealth studies), clustering of processed data, and the RMAB modeling per cluster. The transition

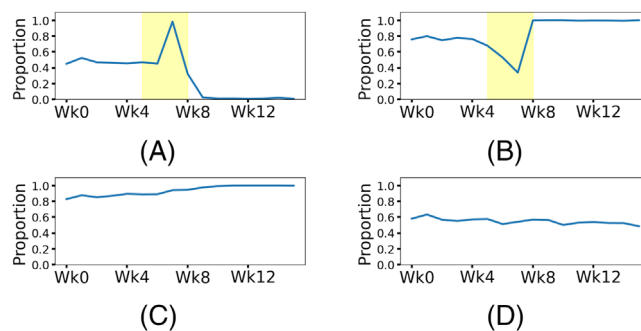


FIGURE 3 Figures (A) and (B) show anomalous engagement behavior while figures (C) and (D) are genuine behaviors. The y-axis shows the proportion of cluster-population in engaging state.

probabilities and the Whittle indexes are then learned per cluster. Additionally, a mapping from socio-demographic features of a beneficiary to a cluster is also learned offline. This mapping is used to treat a new beneficiary during model execution—transition probabilities and Whittle index values for the new beneficiary are given by the corresponding values of the beneficiary's mapped cluster. These individual modules are now described. For data privacy reasons, the data pipeline only uses anonymized data and no personally identifiable information (PII) is made available to the AI models.

Data processing

We train the model on a dataset obtained from historic data collected by ARMMAN, consisting of demographic features and listenership patterns. However, during the pre-deployment trials, we observed some anomalous engagement behaviors—the engagement behavior for some beneficiaries was extremely spiky and unexpected. Figure 3A,B shows two such anomalous groups with a clear peak and dip contrasted with groups having genuine engagement behavior. Upon investigation, we found that this spiky behavior resulted from unanticipated real-world events like network outages.

We detect and exclude such anomalies from SAHELI's data training pipeline. We first group beneficiaries based on their passive transition probabilities. For grouped beneficiaries, we then obtain a running mean of their engagement over time where the mean is calculated over a window of 3 weeks. We filter out all groups with more than 20% change in running mean engagement within a week. Figure 3C,D shows two groups that do not exhibit anomalous behavior and is maintained in the data pipeline.

Additionally, further discussions with ARMMAN pointed out long-term engagement issues in some

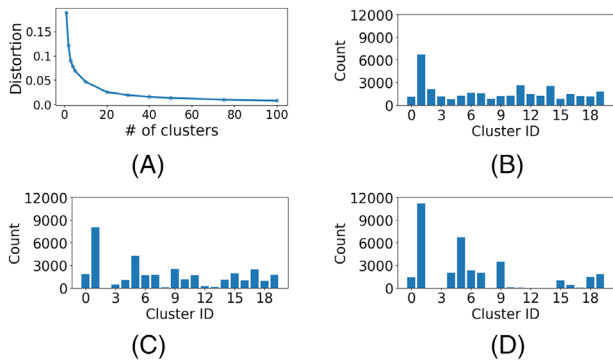


FIGURE 4 Figure (A) shows elbow plot with distortion for varying number of clusters. Figures (B)–(D) show the distribution of predicted clusters using the Feature Only (FO), Feature and Warm-up (FW), and Warm-up Only (WO) mapping functions.

beneficiaries, such as the registration of a wrong or out-of-service phone number, or the beneficiary not being pregnant. Live service calls in these cases are not productive. Thus, as a preprocessing step, we do not consider beneficiaries who have not listened to any automated voice calls in the past 6 weeks.

Clustering

We face a data scarcity and skew challenge in our domain. Specifically, our training dataset comprises of beneficiaries from our own past studies where intervention data are available for only a limited set of these beneficiaries. Thus, to define the parameters of the RMAB model, we cluster beneficiaries as an effective way of addressing data scarcity. We cluster the beneficiaries per their transition behaviors for passive actions using *k-means* clustering. We obtain transition probabilities for each of these clusters by aggregating their transitions as a whole.

However, the optimal number of clusters is a design choice not readily addressed by *k-means*. We experimented with the number of clusters ranging from 1 to 100, and looked at the *distortion* metric. Distortion is the sum of squared distances of each point from its corresponding centroid, where smaller distortion implies better clustering. We plot the distortion values for multiple number of clusters and find 20 to be the ideal choice using elbow-method. The results are shown in Figure 4A where the x-axis is the number of clusters and the y-axis is the distortion value. This clustering approach has the added advantage of offering computational frugality.

Mapping features to clusters

When a new beneficiary enrolls into the system, the system only knows about their demographic data. We,

therefore, need to learn a mapping of a beneficiary's socio-demographic features to clusters, to enable inferring transition probabilities and Whittle indexes for newly enrolled beneficiaries (step 6 in Figure 2). We experimented with different mapping functions to identify the best one: Features Only (FO) mapping—beneficiaries' socio-demographic features only; Warm-up Only (WO) mapping—transition probabilities computed from warm-up period (first 6 weeks post enrollment); and lastly Feature and Warm-up (FW) mapping—using a combination of the above two.

We compute mean absolute error between predicted and ground truth passive transition probabilities as a performance metric and found them as [0.40, 0.37, 0.38] for FO, FW, and WO strategies, respectively. In addition to MAE, we plot the distribution of beneficiaries predicted in different clusters (refer Figure 4B–D). Having a sparse cluster distribution is undesirable since large clusters lowers the granularity of Whittle index planning. As an extreme example, if all beneficiaries are mapped to a single cluster, they would all have the same transition probability and thus the same Whittle indexes. Since the cluster size is now much larger than the number of arms to be pulled, the beneficiaries within that cluster would be chosen randomly for receiving service calls, which would degrade the performance.

Thus, to ensure equitable cluster distribution, we computed Entropy and Gini index values for the predicted distribution of number of beneficiaries per cluster. Entropy values came out to be [2.81, 2.56, 2.04] for FO, FW, and WO, respectively, and Gini indexes were [0.29, 0.48, 0.57]. Given the error similarities for the three strategies, and higher entropy/lower Gini index implies more equitable clusters, we chose FO as our strategy.

RMAB modeling and Whittle index computation

These transition probabilities per cluster are used to compute Whittle indexes for all beneficiaries, similar to Mate et al. (2022), that is, computing $2 \times k$ unique indexes where k is the number of clusters. There are two Whittle indexes per cluster as beneficiaries may be in the engaging or nonengaging states. Whittle index indicates the benefit of performing an active action on a beneficiary: higher Whittle indexes are chosen to receive service calls (step 7 in Figure 2). By mapping beneficiaries to clusters, the Whittle indexes can be precomputed per cluster at the beginning of the deployment, thus providing a frugal solution ideal for large scale deployment with minimal resources.

Frequency of repeated live service calls

We initially enforced a frequency restriction that required ensuring no beneficiary be called more than once in $\eta + 1$ weeks (we set $\eta = 3$). Algorithmically, we implement this by appending η sets of dummy “sleeping states” to the state space that we force the beneficiaries to transition through each time they are called. This augmentation yields a state space of size $2\eta + 2$ and a transition matrix of size $(2\eta + 2) \times (2\eta + 2)$. However, our pilot tests reveal that repeat calls made within just $\eta = 3$ weeks are less effective. For instance, we observed that 30% of “Nonengaging” beneficiaries converted to “Engaging” due to the first service call; however, this number drops to 20% for repeat calls made just 3 weeks later. To address this issue, along with the subject matter experts at ARMMAN, we increased the sleeping period, η , to 12 weeks.

Frugality of system design

Successful deployments of AI systems like SAHELI in social good settings require conscious focus on frugality across the system design. This frugality means reducing both the direct costs (e.g., number of calls) and indirect costs (e.g., computational requirements) on our NGO partners. Here are some design choices in SAHELI that have led to frugality in its operations:

1. Clustering of beneficiaries allows us to compute transition probabilities and Whittle indexes at a cluster level as opposed at the beneficiary level. Since we use 20 clusters for thousands of beneficiaries, it provides a significant scale-up in performance, while simultaneously reducing data demands for learning RMAB model parameters.
2. As described above, we updated the “sleeping states” parameter η to 12. However, this increases the Whittle index computation time sharply, owing to a bulky transition matrix of size 26×26 . With frugality in mind, we use the insight that a sleeping constraint with large η can be approximated as a permanent sleeping constraint, akin to setting η to $+\infty$, for the purposes of index computation. This is because in index computation, the contribution of reward terms appearing after η timesteps is discounted by a factor of γ^η ($\gamma < 1$), which precipitously diminishes to zero. This simplification compresses the transition matrix to 4×4 , and unlocks a $25\times$ speedup in index computation, as shown in Figure 5.
3. Lastly, multiple frugal design choices were made in the orchestration of cloud infrastructure. Specifically, we run our services on-demand using a task sched-

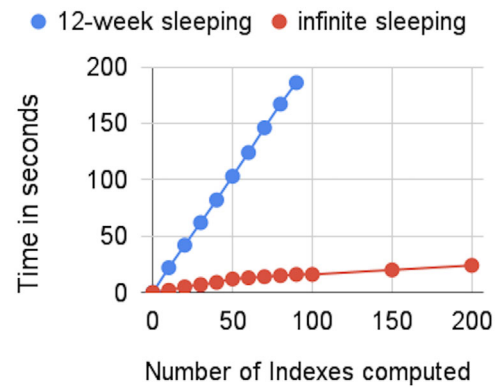


FIGURE 5 Index computation is significantly faster with the infinite sleeping approximation.

uler on default container settings of 6 vCPUs and 16GiB memory.

APPLICATION USE AND PAYOFF

We now discuss the impact of SAHELI on both the beneficiaries as well as the AI community in more detail. SAHELI is deployed and in continuous use at ARMMAN. It has already reached 130K beneficiaries, and is on track to reach one million beneficiaries by the end of 2023. We provide a summary of impact from SAHELI in Table 2.

Engagement results

In order to evaluate the impact of live service calls through SAHELI, we study the engagement behavior of a cohort of 5000 beneficiaries for 12 weeks, registered between February 2022 and April 2022. Additionally, we create a holdout set of beneficiaries registered in the same time period but are not given any live service calls (we obtained ethical approvals before our studies; see section Responsible AI practices for further discussion). We make sure that both the SAHELI and holdout groups have equal number of beneficiaries, equal number engaging beneficiaries at the start of experiment, and similar socio-demographic features.

Figure 6A shows how many engagements did not occur in the holdout group that occurred in the SAHELI group, aggregated cumulatively across months. It demonstrates that the SAHELI group received significant benefit with an additional 328 engagements over the holdout group cumulatively at the end of three months. We also measured the difference in terms of time spent listening to mMitra voice calls. More time spent implies more content exposure, as well as better adherence with the mHealth program. In particular, by the end of month 3, the SAHELI group had

TABLE 2 A summary of impact from SAHELI.

Impact from SAHELI	
Beneficiaries served	130K
In continuous use since	April 2022
Relative engagement drops prevented over holdout group ^a	30.5%
Additional average per beneficiary content exposure over holdout group ^a	12 s
Relative increase in content exposure over holdout group ^a	46.4%
For bottom 25 percentile of listeners, additional average per beneficiary content exposure over holdout group ^a	39 s
For bottom 25 percentile of listeners, relative increase in content exposure over holdout group ^a	130%

^aResults from a sample of 5000 beneficiaries.

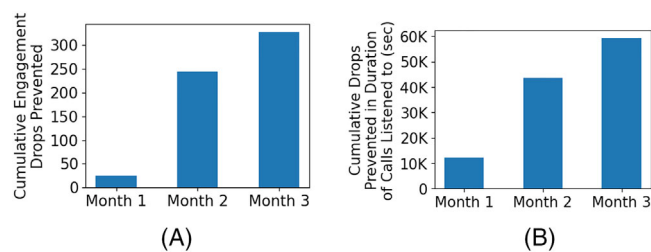


FIGURE 6 (A) Prevention in drop in engagement (cumulative). (B) Increased time spent listening to calls (cumulative).

listened to 59,336 s (~ 12 s per beneficiary, but please see analysis below) *more of content than the holdout group* (Figure 6B). Similar to Mate et al. (2022), we define the relative improvement in listenership metric over the holdout group as

$$\% \text{ improvement} = \frac{\Delta \text{listenership (SAHELI, holdout)}}{\text{listenership in holdout}} \quad (1)$$

As the holdout group has 1075 drops in engagements and 127,711 s drop in duration of calls listened to over 3 months, SAHELI prevented **drop in engagements by 30.5%** with an **additional content exposure of 46.4%** in comparison to the holdout group. This analysis demonstrates SAHELI's success in achieving our core objectives of improving information dissemination.

Who is benefitted from SAHELI?

In order to determine the characteristics of beneficiary who gain the most from SAHELI, we divide the 5000 beneficiaries in our cohort based on two criterion

- (1) Listenership prior to the start of study
- (2) Gestational age at the time of enrollment

First, we consider the listenership of beneficiaries 1 month prior to start of live service calls delivered through

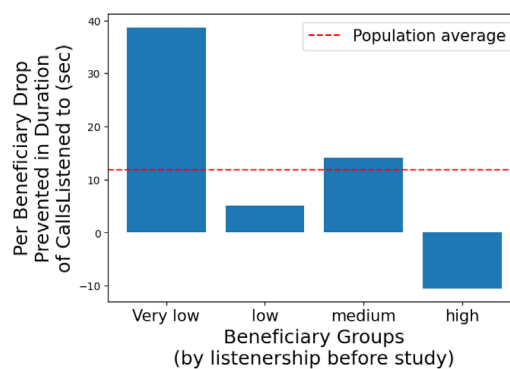


FIGURE 7 Increased time spent listening to calls (over 3 months). The metric is shown for beneficiaries belonging to *very low, low, medium, and high* quartiles of listenership before the start of study.

SAHELI. In this time period, we calculate the mean duration of calls listened to every week. Based on this metric, we divide the 5000 strong cohort into quartiles of listenership—*very low, low, medium, and high*. These quartiles thus characterize the initial behavior of beneficiaries. Next, we repeat the same steps for the holdout population which does not receive any service calls. Finally, we plot how many more seconds of mMitra content is listened by every beneficiary in the quartiles in SAHELI group as compared to the same quartiles in the holdout group (Figure 7).

While the population average increase in content listenership is ~ 12 s, beneficiaries with different listenership profiles before being exposed to SAHELI show very distinctive behaviors. Specifically, the *very low* quartile of beneficiaries gains the most in SAHELI, with 39-s additional content listenership over the holdout group. In absolute terms, the *very low* quartile in holdout group has per beneficiary 30 s of increase in duration of calls listened to over 3 months while the SAHELI group has per beneficiary 69 s of increase in duration of calls listened in the same time period for the same quartile. Thus, using Equation (1), we note that in relative terms, the *very low* quartile

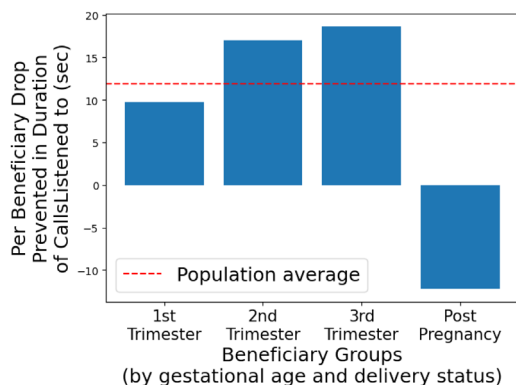


FIGURE 8 Increased time spent listening to calls (over 3 months). The metric is shown for pregnant mothers in their 1st, 2nd, and 3rd trimesters of pregnancy and for beneficiaries who have already delivered.

has 130% additional content exposure in comparison to the holdout group (see Figure 7).

For the second criterion, we consider the gestational age of beneficiaries and their delivery status at the time of enrollment. For pregnant women, we use the gestational age at the time of enrollment to calculate their pregnancy trimester. Similar to Figure 7, in Figure 8, we plot for every gestational age bucket, how many additional seconds of mMitra calls are listened by every beneficiaries in the SAHELI group as compared to the holdout group. Specifically, we observe that beneficiaries close to the delivery date (higher trimester) have greater benefit from being in the SAHELI group.

Impact of live service calls

We performed a qualitative study to understand human-AI collaboration due to the AI system. We conducted a total of 24 interviews, two focus group discussions, and approximately 90 h of observation. We found that healthcare workers engaged positively with targeted predictions through the AI system that integrated into their day-day workflows seamlessly. It helped them improve the engagement of beneficiaries, provided an opportunity to support them in their care journeys and understand their needs.

We also investigated the reasons for why live service calls helped improve engagement with ARMMAN's mMitra mHealthprogram from the perspective of the beneficiary. Specifically, we conducted a follow-up study with a sample of beneficiaries who were given live service calls 1 year ago. We could successfully reach out to 306 beneficiaries, out of which 134 recalled the details of the service call from a year ago. Table 3 shows the responses to our follow-up

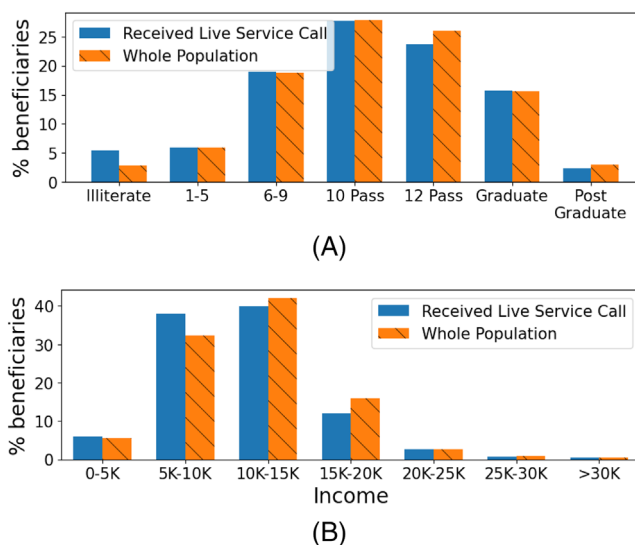


FIGURE 9 Distribution of (A) education (highest education received) and (B) income (monthly family income in Indian Rupees) across cohort that received service call and the whole population.

study by these 134 beneficiaries. Particularly, 50.75% beneficiaries engaged more with mMitra calls after getting more information about the program. The service calls also helped improve listenership by making logistical updates such as updating delivery date (9.7%), changing time slot of receiving the call (8.21%), or updating the phone number (2.99%).

Fairness of the RMAB model

Model fairness in nonwestern contexts has not received much attention in the literature Sambasivan et al. (2021). Responsible AI principles of the Government of India's NITI AAYOG (2021) for example, requires nondiscrimination based on sensitive markers like caste and religion. These sensitive data are specifically not collected by ARMMAN for mMitra, thereby, making it inaccessible to SAHELI's AI models. We worked with public health and field experts to evaluate other indicators such as education, and income levels that signify markers of socio-economic marginalization. ARMMAN's goals for SAHELI are to favor beneficiaries of lower income and lower education levels for service calls. We conducted a post hoc analysis of the deployment to evaluate if SAHELI indeed met such preferences.

Figure 9A shows the distribution of beneficiaries aggregated across SAHELI's enrollments split into different education levels in India. We compare those who were chosen for live service calls by SAHELI versus the enrolled population. The x-axis portrays the education levels; for

**TABLE 3** Follow-up study responses.

Did the call help you to listen to the mMitra calls more regularly?	# of Beneficiaries	% of Beneficiaries
Yes, after getting more information about mMitra, I am listening to the calls more regularly	68 (in 134)	50.75%
Not really	30	22.39%
Yes, after updating my delivery date, I was able to get the right information	13	9.7%
Yes, after changing time slot, I am able to listen to the calls more regularly	11	8.21%
Have not asked my wife	4	2.99%
Yes, after changing the number, I am able to listen to the calls more regularly	4	2.99%
Any other	4	2.99%

instance grade 1–5 represents primary school, grade 6–9 middle school, 10th pass junior high, and 12th pass represents senior high school. The y-axis is the percentage of beneficiaries per education category. For instance, SAHELICalls 5.5% of beneficiaries who had very little formal education or were illiterate, whereas this group was 2.8% of the overall enrolled population.

We did a similar analysis split by income as depicted in Figure 9B. The x-axis contains buckets of average monthly income of the beneficiary household in Indian Rupees, and the y-axis denotes the percentage of beneficiaries in that income category. As an example, the category “5–10K” contains around 30% of the beneficiaries in the population, and almost 40% of the beneficiaries who received a service call.

Both these figures show that SAHELI favors the beneficiaries in the “illiterate” education category and in the “5–10K” income category. This distribution is in line with ARMMAN’s goals—SAHELI favors beneficiaries of lower income and lower education levels for service calls.

Enabling new research

From identifying the right problem to solve, to creating an AI solution, testing it in pilot, iterating on learnings and finally, establishing an end-to-end integrated system, we made our journey to this deployment. With this work, we provide other AI researchers an important case study to take an AI model from the lab out on the field. In our pursuit of deployment of SAHELI, we uncover several research challenges, for example, we overcame the challenges of data scarcity and frugal design. Our work hopefully inspires additional research in robust and computationally efficient approaches for RMABs and other AI applications for mHealth. Further, we are improving SAHELI by incorporating recent advances in Decision Focused Learning for RMAB problems Wang et al. (2023), which opens up new direction of real-world large-scale applications of Decision Focused Learning.

RESPONSIBLE AI PRACTICES

We recognize the responsibility associated with deploying real-world AI systems that impacts underserved communities. In our approach, we have iteratively designed, developed, and deployed the system in constant coordination with an interdisciplinary team comprised of ARMMAN’s field staff, social work researchers, public health researchers, and ethical experts. Along with seeking ethical approvals through review boards at Google and ARMMAN, we have taken additional steps to constantly monitor and mitigate the risks associated with SAHELI by abiding with AI principles at Google (2018) as well as key policy making bodies in India such as the NITI AAYOG (2021). Our success draws attention to the practices around responsible AI including ethics, fairness, and accountability in the nonwestern context Sambasivan et al. (2021) where SAHELI is deployed. We now discuss three of the core Responsible AI principles that impacted the design of SAHELI.

Socially beneficial: The intent of this work is to bring the power of AI in service to some of the most marginalized communities in the global south. The challenges faced by our team were limited resources in every dimension—limited data on the beneficiaries, limited compute available to the NGO, and limited health workers to make the outreach calls. Thus, we had to develop new algorithms that were not data hungry, and were bounded in their computational requirements. To that affect, SAHELI is the first large-scale deployment of RMABs for public health.

Avoid reinforcing unfair bias: As discussed in the previous section, we have undertaken extensive analysis to study model’s fair treatment of beneficiaries.

Incorporate privacy design principles: We take significant measures to ensure that participant consent is understood and recorded in a language of the community’s choice at each stage of the program. Data stewardship resides in the hands of the NGO, and only the NGO is allowed to share data. This dataset will never be used by Google for any commercial purposes. In this dataset, sensitive features such as caste and religion are never collected

and stored. SAHELI's data pipeline only uses anonymized data and no PII is made available to the AI models. Lastly, domain experts at ARMMAN have been deeply involved in the development and testing of SAHELI and have provided continuous input and oversight in data interpretation, data consumption and model design.

LESSONS LEARNED

Over the course of 1 year of our experiments moving from Pilot study to Deployment, we learned several lessons along the way. Most importantly, we learned that even a successful pilot study cannot be translated as-is in to a full-scale deployment, and that several considerations are critical for wide-scale adoption of AI tools and scaling up of impact.

Selecting the right problem: There are multitude of problems that require to be solved to address the needs of the underserved communities. In our interactions with ARMMAN, *we realized that we could create the most impact with our techniques by improving the selection of the right beneficiaries for manual intervention*, as opposed to automating the communication with the beneficiary. Our choice of problem is consistent with the “tech plus touch” philosophy of ARMMAN (2008), and ensures that we complement the human expertise of the health worker. This way, each chosen beneficiary continued to have a one-on-one interaction with a health worker, while simultaneously improving the overall engagement with the mHealth program.

Immersion into the real-world problem: We learned that immersing in the working of a NGO and public health infrastructure is critical in understanding the context of the problem. The authors went on multiple field visits to understand the stakeholders involved in the mMitra's workflow. The health workers interact with the beneficiaries across multiple mHealth programs, and thus can speak to the needs and behaviors of the beneficiaries. For instance, upon interacting with these health workers, we understood how telecom outages lead to more anomalous and incomplete data than we had anticipated. We also understood the decreased value in utility of calling the same beneficiary again shortly after a previous call. *These field visits forced us to re-evaluate our assumptions, and led to better data processing and modeling choices*, as discussed in the earlier sections. For instance, after these discussions, we incorporated a new anomaly detection mechanism in our data pipeline, and impacted our choice of horizon (η) in our RMAB model.

Fairness of AI models: AI algorithms and datasets can reflect, reinforce, or reduce unfair biases. It is imperative on AI designers to seek to avoid unfair impacts on people,

particularly on underserved and marginalized communities. As discussed in the section on Responsible AI practices, we worked with public health and field experts to demonstrate fairness of our approach. As we mentioned before, 94% of our potential beneficiary population are below WHO's poverty index. *Studying multiple socio-demographic attributes was essential to evaluate fairness of our approach*. We worked closely with ethics experts, the ARMMAN's ethics team, and Google's ethics teams and extensively validated the fairness of our models.

End-to-end integration testing: In addition to the lessons learned on problem selection and model development, we also ran into several issues in our end-to-end integrated pipeline. On one occasion, we saw poor results because the data schema had evolved in the data storage pipeline at ARMMAN. *Testing of our application required our NGO partner to be equally involved in the validation of SAHELI's outputs —as domain experts, they are better equipped to identify counter-intuitive behaviors*. Our experiences uncovering issues in the end-to-end pipeline led to improved communication practices, better documentation, and tighter test goals. Social good applications like SAHELI has real-world consequences for beneficiaries in underserved communities, and it is critical that there be a real partnership for testing and integration.

CONCLUSION

In this paper, we presented SAHELI, the first ever deployment of RMABs in the public health domain for allocation of limited resources. SAHELI is built on an improved and robust framework that both predicts RMAB parameters and computes optimal policies for it, in contrast with most past research that has only focused on computing optimal policies. It has been built with careful design choices inspired by close interactions with all stakeholders. It incorporates numerous lessons learned by embedding ourselves in the real-world domain. SAHELI has been deployed on cloud infrastructure with an emphasis on frugality, and has reached out to 130K beneficiaries so far and aims to reach one million by 2023. Furthermore, in this paper, we also discuss the importance of responsible AI practices in deploying AI systems at scale, especially in the social domain. This work serves as an important case study for AI researchers and NGO communities alike to take ML models from the lab and deploy them in the field.

ACKNOWLEDGMENTS

We thank Lovish Madaan and Aakriti Kumar for their insights and feedback through the project, Harshit Khaitan, Anirudh GP, and Hasan Rafiq for their support in deployment on GCP, and Manoj Sandur Karnick for aid

with program management. We are also very grateful to Suresh Chaudhary, Sonali Nandalaskar, and the entire mMitra team from ARMMAN for their continuous support since pilot in making this journey to deployment possible.

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict.

ORCID

Shresth Verma  <https://orcid.org/0000-0003-0370-5471>

ENDNOTE

¹<https://github.com/armman-projects/SAHELI>

REFERENCES

- ARMMAN. 2008. "About ARMMAN." Accessed: August 12, 2022. <https://armman.org/about-us>.
- ARMMAN. 2019. "Assessing the Impact of Mobile-based Intervention on Health Literacy among Pregnant Women in Urban India." Accessed: August 12, 2022. <https://armman.org/wp-content/uploads/2019/09/Sion-Study-Abstract.pdf>.
- Brisimi, T. S., R. Chen, T. Mela, A. Olshevsky, I. C. Paschalidis, and W. Shi. 2018. "Federated Learning of Predictive Models from Federated Electronic Health Records." *International Journal of Medical Informatics* 112: 59–67.
- Corotto, P. S., M. M. McCarey, S. Adams, P. Khazanie, and D. J. Whellan. 2013. "Heart Failure Patient Adherence: Epidemiology, Cause, and Treatment." *Heart Failure Clinics* 9(1): 49–58.
- Eysenbach, G. 2005. "The Law of Attrition." *Journal of Medical Internet Research*, 7(1): e11.
- Glazebrook, K., D. Ruiz-Hernandez, and C. Kirkbride. 2006. "Some Indexable Families of Restless Bandit Problems." *Advances in Applied Probability* 38: 643–72.
- Google. 2018. "Artificial Intelligence at Google: Our Principles." Accessed: August 12, 2022. <https://ai.google/principles>.
- Gupta, K., S. Roy, R. Poonia, S. R. Nayak, R. Kumar, K. Alzahrani, M. Alnefaie, and F. Al-Wesabi. 2022. "Evaluating the Usability of mHealth Applications on Type 2 Diabetes Mellitus Using Various MCDM Models." *Healthcare* 10: 4.
- Jakob, R., S. Harperink, A. M. Rudolf, E. Fleisch, S. Haug, J. L. Mair, A. Salamanca-Sanabria, and T. Kowatsch. 2022. "Factors Influencing Adherence to mHealth Apps for Prevention or Management of Noncommunicable Diseases: Systematic Review." *Journal of Medical Internet Research* 24(5): e35371.
- Killian, J. A., B. Wilder, A. Sharma, V. Choudhary, B. Dilkina, and M. Tambe. 2019. "Learning to Prescribe Interventions for Tuberculosis Patients Using Digital Adherence Data." In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*.
- Liao, P., K. Greenewald, P. Klasnja, and S. Murphy. 2020. "Personalized Heartsteps: A Reinforcement Learning Algorithm for Optimizing Physical Activity." In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4(1): 1–22.
- Liu, K., and Q. Zhao. 2010. "Indexability of Restless Bandit Problems and Optimality of Whittle Index for Dynamic Multichannel Access." *IEEE Transactions on Information Theory* 56(11): 5547–67.
- Mate, A., L. Madaan, A. Taneja, N. Madhiwalla, S. Verma, G. Singh, A. Hegde, P. Varakantham, and M. Tambe. 2022. "Field Study in Deploying Restless Multi-Armed Bandits: Assisting Non-Profits in Improving Maternal and Child Health." *Proceedings of AAAI* 36(11): 12017–25.
- Meh, C., A. Sharma, U. Ram, S. Fadel, N. Correa, J. Snelgrove, P. Shah, R. Begum, M. Shah, T. Hana, H. Fu, L. Raveendran, B. Mishra, and P. Jha. 2022. "Trends in Maternal Mortality in India Over Two Decades in Nationally-Representative Surveys." *BJOG: An International Journal of Obstetrics & Gynaecology* 129: 550–61.
- Meyerowitz-Katz, G., S. Ravi, L. Arnolda, X. Feng, G. Maberly, and T. Astell-Burt. 2020. "Rates of Attrition and Dropout in App-Based Interventions for Chronic Disease: Systematic Review and Meta-Analysis." *Journal of Medical Internet Research* 22(9): e20283.
- NITI AAYOG. 2021. "Responsible AI: Approach Document for India." Accessed: August 12, 2022. <https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-20202021.pdf>.
- Papadimitriou, C. H., and J. N. Tsitsiklis. 1999. "The Complexity of Optimal Queuing Network Control." *Mathematics of Operations Research* 24(2): 293–305.
- Pilote, L., J. P. Tulskey, A. R. Zolopa, J. A. Hahn, G. F. Schecter, and A. R. Moss. 1996. "Tuberculosis Prophylaxis in the Homeless: A Trial to Improve Adherence to Referral." *Archives of Internal Medicine* 156(2): 161–65.
- Pollack, M. E., L. Brown, D. Colbry, C. Orosz, B. Peintner, S. Ramakrishnan, S. Engberg, et al. 2002. "Pearl: A Mobile Robotic Assistant for the Elderly." In *AAAI Workshop on Automation as Eldercare, AAAI, 2002*, volume 2002. Edmonton, Alberta, Canada.
- Puterman, M. L. 2014. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York, NY: John Wiley & Sons.
- Qian, Y., C. Zhang, B. Krishnamachari, and M. Tambe. 2016. "Restless Poachers: Handling Exploration-Exploitation Tradeoffs in Security Domains." In *AAMAS*.
- Sambasivan, N., E. Arnesen, B. Hutchinson, T. Doshi, and V. Prabhakaran. 2021. "Re-imagining Algorithmic Fairness in India and Beyond." *CoRR*, abs/2101.09995.
- Son, Y.-J., H.-G. Kim, E.-H. Kim, S. Choi, and S.-K. Lee. 2010. "Application of Support Vector Machine for Prediction of Medication Adherence in Heart Failure Patients." *Healthcare Informatics Research* 16(4): 253–59.
- Tshikomana, R. S., and M. M. Ramukumba. 2022. "Implementation of mHealth Applications in Community-Based Health Care: Insights from Ward-Based Outreach Teams in South Africa." *PLOS ONE* 17(1): 1–15.
- Tuldrà, A., M. J. Ferrer, C. R. Fumaz, R. Bayés, R. Paredes, D. M. Burger, and B. Clotet. 1999. "Monitoring Adherence to HIV Therapy." *Archives of Internal Medicine* 159(12): 1376–77.
- Wang, K., S. Verma, A. Mate, S. Shah, A. Taneja, N. Madhiwalla, A. Hegde, and M. Tambe. 2023. "Scalable Decision-Focused Learning in Restless Multi-Armed Bandits with Application to Maternal and Child Health." *Proceedings of the AAAI Conference on Artificial Intelligence* 37(10): 12138–46. <https://doi.org/10.1609/aaai.v37i10.26431>
- Weber, R. R., and G. Weiss. 1990. "On an Index Policy for Restless Bandits." *Journal of Applied Probability* 27(3): 637–48.
- Whittle, P. 1988. "Restless Bandits: Activity Allocation in a Changing World." *Journal of Applied Probability* 25(A): 287–98.

World Bank. 2020. *Poverty and Shared Prosperity 2020: Reversals of Fortune*. Washington, D.C.: The World Bank.

World Health Organization (WHO). 2020. "Newborns: Improving Survival and Well-Being." Accessed: August 12, 2022. <https://www.who.int/news-room/fact-sheets/detail/newborns-reducing-mortality#:~:text=Neonates,in%20child%20survival%20since%201990.>

How to cite this article: Verma, S., G. Singh, A. Mate, P. Verma, S. Gorantla, N. Madhiwalla, A. Hegde, D. Thakkar, M. Jain, M. Tambe, and A. Taneja. 2023. "Expanding impact of mobile health programs: SAHELI for maternal and child care." *AI Magazine* 44: 363–376. <https://doi.org/10.1002/aaai.12126>

AUTHOR BIOGRAPHIES

Shresth Verma is a Pre-doctoral Researcher at Google Research India in the Multi-Agent Systems for Social Impact group and an incoming PhD student at Harvard University Center for Research on Computation and Society. Previously, he was a Data Scientist at United Health Group where he developed healthcare risk models for millions of beneficiaries. His research interests include bandits, reinforcement learning, and multi-agent systems with applications in public health and conservation.

Gargi Singh is currently a Pre-doctoral Researcher at Google Research India in the AI for Social Good team. She has a Bachelors and Masters in Computer Science and Engineering from the Indian Institute of Technology, Kanpur, India. She completed her Masters thesis under the supervision of Dr Ashutosh Modi in textual fine-grained emotion prediction. She received the best student paper award in ACII 2021 and the innovative application award in IAAI 2023. Her research interests encompass different areas in applied artificial intelligence.

Aditya Mate is an incoming Applied Scientist at Microsoft New England. He has a Ph.D. in Computer Science from Harvard University advised by Prof. Milind Tambe. Previously, he completed his B.Tech and M.Tech from the Indian Institute of Technology, Bombay (IIT Bombay). Aditya's research explores techniques from Sequential Decision Making, Machine Learning, and Causal Inference for solving impactful real-world problems, especially in public health.

His work on improving maternal healthcare has been recognized by the "IAAI 2023 Innovative Application Award" and he also won the second prize for his talk at the "INFORM 2022 Doing Good With Good OR Competition." His work has won the best paper awards at NeurIPS 2021 and NeurIPS 2020 workshops on machine learning for public health. His work has also been featured in the Nature Asia newsletter, Harvard newsletter and has appeared on the Google AI blog. Aditya is also fortunate to have previously been a research intern at IBM T.J. Watson Research Center and a Research Intern at Google Research (thrice).

Paritosh Verma is a PhD student in the Department of Computer Science at Purdue University. He is advised by Prof. Alex Psomas. His research focuses on problems at the interface of computer science and mathematical economics, particularly on algorithmic fairness in resource allocation settings and mechanism design.

Sruthi Gorantala is a PhD student at the Indian Institute of Science Bangalore. Her research interests are algorithms, optimization, and fairness. Her research is supported by a Google PhD Fellowship.

Neha Madhiwalla is a Research Director at ARM-MAN. She has a PhD in Social Sciences and has been involved in public health research for over 20 years. Her areas of expertise include maternal and child health, health systems, medical education, and training and bioethics. She is presently involved in research on mHealth, exploring the use to telemedicine for the rehabilitation of children with malnutrition; human-computer interactions in mhealth programs for maternal and child health; using AI to improve the effectiveness of large-scale digital health communication programs. She has worked at the Centre for Enquiry in Health and Allied Themes (CEHAT) and Centre for Studies in Ethics and Rights, both affiliated to the Anusandhan Trust, between 1987 and 2013. She worked as a Research Scientist at the Ramalingaswami Centre for Equity and Social Determinants of Health, Bangalore in 2016 and 2017. She has been a member of the research ethics committee at BARC Hospital, Mumbai since 2016. She has had a parallel career in development practice. She founded Chehak Trust, an organization which has been involved in developing community-based models for education, adolescent health, and disability rehabilitation. Chehak Trust was awarded the Edelweiss-Kubera Award for Innovation



in Education in 2010. She received the MacArthur Fund for Leadership Fellowship (2002–2005) to develop a comprehensive adolescent sexual and reproductive health program for adolescent girls.

Aparna Hegde is a Stanford and Cleveland Clinic trained, internationally renowned Urogynecologist. She is the Founder of NGO ARMMAN, which creates scalable mHealth-based programs to impact maternal and child health in 20 states of India (reach: 36 million women and their children and 318,000 health workers). Dr Hegde is an Associate Professor (Hon) and head of the upcoming Department of Urogynecology at Cama Hospital, Mumbai, one of India's first University-based Centers of Excellence in the field with a fellowship program. Dr Hegde is an accomplished Researcher, an NIH grantee, and a pioneer in 2D/3D pelvic floor ultrasound. She is the Chair of FIUGA, the foundation arm of IUGA (International Urogynecology Association), and the Chair of the IUGA Publications Committee. Dr Hegde is a Skoll Awardee (2020), Senior Ashoka Fellow (2021) and was listed by Fortune Magazine as one of the World's 50 Greatest Leaders in 2021. She has also won the Women Transforming India award by Niti Aayog in 2022.

Divy Thakkar is a Staff Program Manager at Google Research where he leads Strategy and Operations for several critical functions and Google's academic research strategy and investments in India, South-east and South Asia. Divy is an HCI researcher with interests in examining Human–AI interactions in low-resource and/or high-stakes environments. He co-leads efforts towards shaping a world-class AI for Social Good team impacting lives at-scale.

Manish Jain is a member of Google Research India where he is leading an effort on understand, model and leverage propagation networks of information. He has also contributed towards research on digital health using reinforcement learning and decision theory. A

few years prior to joining Google, he was the founding CTO of Avata Intelligence. He managed the startup for 8 years before it was acquired by Procore Technologies in early 2020. At the time of its acquisition, Avata's core product was a question–answering chatbot over corporate documents (built for employees of the corporation). He also holds a PhD in Computer Science, focusing on large-scale optimization, game theory and graph-based algorithms.

Milind Tambe is a Principal Scientist and Director for “AI for Social Good” at Google Research; concurrently, he is Gordon McKay Professor of Computer Science and Director of Center for Research in Computation and Society at Harvard University. He is a recipient of the IJCAI John McCarthy Award, AAI Feigenbaum Prize, AAI Robert S. Engelmere Memorial Lecture Award, AAMAS ACM Autonomous Agents Research Award, INFORMS Wagner prize for excellence in Operations Research practice and MORS Rist Prize. He is a fellow of AAI and ACM. For his work on AI and public safety, he has received Columbus Fellowship Foundation Homeland security award and commendations and certificates of appreciation from the US Coast Guard, the Federal Air Marshals Service and airport police at the city of Los Angeles.

Aparna Taneja is a Researcher at the Multi Agents Systems for Social Impact team in Google Research India. She collaborates with several NGO's and academic partners in the fields of public health and conservation, and her primary focus is the collaboration with ARMMAN, an NGO focused on improving maternal and child health outcomes in India. Aparna joined Google in 2015 and worked with the Google Maps team in Zurich to improve search quality for Maps. She received her PhD in Computer Science at ETH Zurich under the supervision of Prof. Marc Pollefeys. She then pursued a postdoc at Disney Research Zurich. She received her Bachelors and Masters in Computer Science from IIT Delhi in 2006.