© April 2026 | IJIRT | Volume 12 Issue 11 | ISSN: 2349-6002

Sarthi – The Helping Hand: An AI-Driven Framework for Optimizing NGO–Volunteer Coordination

Aryav Jain[1], Akshat Kumar[2], Pratibha Gautam[3], Sweta Yadav[4]
[1,2,3,4]*Department of Computer Science & Engineering, MGM's College of Engineering & Technology, Noida, India*

***Abstract*—Efficient mobilization of volunteer labor within India's fragmented civil society sector remains a consequential socio-technical challenge, compounded by persistent informational asymmetries and behavioral attrition that undermine even algorithmically sound matching systems. Prior work on the Sarthi AI-assisted coordination framework established content-based filtering as a viable baseline, achieving a matching precision of 0.71; however, a 22% volunteer no-show rate exposed a critical reliability gap between digital assignment and physical task execution. This paper presents an architectural evolution of the Sarthi framework that transitions the system from passive recommendation toward high-fidelity, behaviorally grounded coordination. The proposed methodology integrates four interconnected mechanisms: a 24-hour pre-commitment protocol functioning as a Human-in-the-Loop behavioral filter; Structured Task Specification templates employing Jaccard similarity for skill alignment and Haversine distance for geospatial proximity; a Bayesian Reputation Score modeled as the posterior mean of a Beta distribution over historical task completions and failures; and an epsilon-greedy Multi-Armed Bandit allocation reserving 15% of opportunities for a Shadowing Tier to ethically onboard unverified volunteers and mitigate cold-start inequity. The modified compatibility scoring function incorporates reputation as a multiplicative reliability weight alongside skill, availability, location, and cause-alignment features. Experimental projections derived from a 12-week pilot dataset indicate that these interventions are expected to reduce volunteer attrition from 22% to 12–15% and elevate matching precision from 0.71 to 0.85. The framework offers a replicable, equity-conscious methodology for optimizing human capital allocation in resource-constrained civic environments.**

***Index Terms*—Artificial Intelligence, Bayesian Inference, Behavioral Verification, Civil Society Technology, Content-Based Filtering, Digital Volunteerism, Exploration-Exploitation Tradeoff, Geospatial Matching, Human-in-the-Loop Systems, Multi-Armed Bandit, NGO Coordination, Recommender Systems, Reputation Scoring, Resource Optimization, Socio-Technical Systems, Volunteer Management**

I. INTRODUCTION

A. Background and Context

India's civil society ecosystem encompasses an extensive network of non-governmental organizations (NGOs) and volunteer communities engaged across diverse social mandates, from disaster relief and healthcare outreach to educational access and environmental stewardship. Despite this considerable scale, inter-stakeholder coordination continues to rely on low-fidelity mechanisms informal messaging platforms, static spreadsheets, and word-of-mouth recruitment that are structurally incapable of supporting the demands of modern volunteer management. These manual approaches offer no systematic means of performing skill-taxonomy-based matching, assessing real-time availability, or maintaining longitudinal performance records. The consequence is a compounding cycle of misallocation: volunteers with relevant competencies are overlooked in favor of those with greater social proximity to an NGO, tasks are staffed with mismatched personnel, and the absence of accountability infrastructure erodes trust between organizations and the individuals they depend upon. As digitally mediated volunteerism expands, the inadequacy of these coordination mechanisms represents not merely an administrative inconvenience but a structural constraint on the scalability and impact of civil society interventions.

B. The Reliability Gap in Volunteer Coordination

A critical, and often underappreciated, failure mode in volunteer coordination lies not in the matching stage but in the execution stage the transition from a digital assignment to verified physical participation. Existing AI-assisted frameworks have concentrated their design efforts on improving recommendation quality, as measured by precision and recall on matching tasks, while treating post-match behaviour as an external variable beyond system control. This assumption proves empirically costly. Pilot deployments of prior coordination systems have recorded volunteer no-show rates of approximately 22% even when algorithmic matching quality was demonstrably high, revealing what we term a reliability gap: a systematic divergence between predicted and realized participation. Qualitative analysis traces this gap to three interacting causes self-reported skill and availability data that is unverified at the point of ingestion, the absence of any commitment-validation mechanism between match confirmation and task execution, and information asymmetry between NGO administrators and volunteer participants. These factors collectively demonstrate that coordination fidelity cannot be achieved through recommendation accuracy alone; it requires behavioural verification as a first-class system concern.

C. Research Objective and Contribution

This work proposes a substantive architectural evolution of the Sarthi AI-assisted coordination framework, shifting its operational paradigm from passive skill-based recommendation to active, high-fidelity coordination grounded in behavioural assurance. The central contribution is a multi-mechanism system that addresses the reliability gap through three integrated innovations. First, a 24-hour pre-commitment protocol functions as a Human-in-the-Loop behavioural filter, requiring explicit digital confirmation from matched volunteers prior to task execution and thereby converting intent into a verifiable signal. Second, Structured Task Specification templates replace free-text task descriptions with standardized taxonomic fields, enabling deterministic Jaccard similarity scoring for skill alignment and Haversine distance computation for geospatial proximity. Third, a Bayesian Reputation Score, modeled as the posterior mean of a Beta distribution over each volunteer's history of successful completions and failures, is incorporated as a multiplicative reliability weight within the matching function. Complementing these mechanisms, an epsilon-greedy Multi-Armed Bandit strategy allocates a 15% exploration budget to a Shadowing Tier for new and unverified volunteers, ensuring the system remains equitable and avoids penalizing users who lack historical records. Together, these components are projected to reduce volunteer attrition from 22% to 12–15% and raise matching precision from 0.71 to 0.85, establishing a replicable and ethically grounded methodology for AI-driven coordination in resource-constrained civic environments.

1.1 Research Objectives

- RO1: To design and evaluate a Bayesian Reputation Scoring mechanism that dynamically models volunteer reliability as a posterior probability over historical task completion behaviour, and to integrate it as a weighted component within the existing Sarthi matching function.
- RO2: To assess the effectiveness of a 24-hour pre-commitment protocol as a Human-in-the-Loop behavioural filter in reducing volunteer no-show rates relative to the baseline system.
- RO3: To implement and validate Structured Task Specification templates employing Jaccard similarity for skill-taxonomy alignment and Haversine distance for geospatial proximity as a replacement for unstructured free-text task descriptions.
- RO4: To develop an epsilon-greedy Multi-Armed Bandit allocation strategy that ethically accommodates unverified volunteers through a dedicated Shadowing Tier, thereby mitigating cold-start inequity without compromising overall system reliability.

1.2 Research Questions

- RQ1: To what extent does Bayesian Reputation Scoring improve volunteer–task matching precision compared to the content-based filtering baseline established in Jain et al. (2026)?
- RQ2: Does the introduction of a pre-commitment validation protocol significantly reduce volunteer no-show rates, and what is its effect on user satisfaction and system accessibility in low-bandwidth environments?
- RQ3: How effectively do Structured Task Specification templates reduce match failures attributable to informational asymmetry and self-reported skill inaccuracy?
- RQ4: Does the epsilon-greedy Multi-Armed Bandit strategy provide equitable coordination outcomes for unverified volunteers without degrading the reliability experience of NGO administrators?

- RQ5: What is the aggregate effect of the proposed architectural evolution on the reliability gap between digital matching and real-world task execution, and does the system demonstrate sufficient generalizability for deployment beyond the pilot context?

II. LITERATURE REVIEW

A. The Evolution of Volunteer Management Systems

The transition from informal civic participation to structured, digitally mediated volunteer coordination has attracted sustained attention within socio-technical research over the past decade. Early foundational work by Schönböck and Raab (2016) established that organizations adopting formalized management architectures encompassing structured recruitment pipelines, standardized orientation programs, and systematic volunteer screening achieve measurably superior retention outcomes relative to those dependent on ad-hoc, relationship-driven coordination. This finding is significant because it frames volunteer attrition not as an inherent feature of civic engagement, but as a correctable organizational failure one that systematic design can meaningfully address.

Despite this understanding, a persistent implementation gap remains, particularly within resource-constrained operational environments. Wu (2023) investigated the deployment of low-code platforms as a mechanism for reducing administrative burden and enabling rapid system development in non-profit contexts. While such platforms demonstrably improve backend operational efficiency and reduce the technical barrier to digitization, they orient their design primarily around administrative tracking and data management rather than toward intelligent, skill-sensitive task allocation or the quality of the volunteer experience itself. The result is a class of systems that are administratively convenient but coordinatively shallow.

This limitation is directly observable within India's NGO ecosystem. Platforms such as GiveIndia and the Robin Hood Army have succeeded in digitizing volunteer onboarding and broadening recruitment reach; however, both continue to rely on manual coordination methods or simple rule-based matching at the point of task assignment. The pervasive use of general-purpose communication tools WhatsApp groups, Google Forms, and email chains further reflects a sector operating without real-time coordination infrastructure, longitudinal performance tracking, or lifecycle management capabilities. These systems can recruit volunteers, but they cannot reliably deploy them.

B. Artificial Intelligence Integration in Civic Technology and the Transparency Deficit

To overcome the structural limitations of manual and rule-based coordination, recent research has explored the integration of artificial intelligence and recommender system architectures into volunteer management platforms. Work synthesized by Iterators (2025) demonstrates that techniques such as preference modeling, behavioral pattern recognition, and vector-based similarity computation well-established in commercial recommendation contexts can be meaningfully adapted to improve volunteer–task alignment by capturing latent compatibility signals beyond simple keyword matching.

However, transposing these techniques into real-world civic systems introduces a distinct class of challenges that purely technical evaluations tend to understate. The Journal of Marketing Science Research (2025) identifies what it terms a transparency deficit in AI-mediated digital platforms: a systemic erosion of user trust arising from opaque algorithmic decision-making, insufficient data accountability, and the absence of interpretable feedback mechanisms. In commercial contexts, this deficit manifests as reduced platform engagement; in volunteer coordination systems, its consequences are more operationally severe.

Within the volunteer coordination domain, the transparency deficit materializes as what this paper terms the reliability gap the divergence between a digitally confirmed match and actual real-world task participation. Content-based filtering algorithms can achieve statistically acceptable precision when matching volunteers to tasks on the basis of explicit attributes such as skills, location, and availability. However, their reliability is fundamentally conditioned on the integrity of input data. When systems ingest self-reported volunteer profiles without any verification layer, they are structurally vulnerable to the garbage-in-garbage-out failure mode: algorithmically confident matches dissolve at the execution stage because the underlying data was inaccurate, overstated, or simply stale. Addressing this gap requires moving beyond recommendation quality as the sole design objective, toward systems that incorporate behavioral verification, bidirectional feedback, and transparent reputation modeling as first-class architectural concerns.

C. Comparative Analysis of Existing Coordination Frameworks

To precisely situate the contribution of this work within the existing landscape, Table I presents a structured comparison of representative volunteer coordination frameworks across five evaluative dimensions: matching methodology, AI utilization, feedback loop design, gamification integration, and identified limitations.

Table I: Comparative Analysis of Volunteer Coordination Frameworks

System / Framework	Matching Method	AI Usage	Feedback Loop	Gamification	Identified Limitations
GiveIndia / Robin Hood Army	Manual / Static Lists	None	Limited, informal	No	No predictive capability; lacks skill-based matching
Wu (2023) Low-Code Model	Rule-based	Minimal	Absent	No	Administratively focused; no intelligent task allocation
Original Sarthi Pilot	Content-based filtering	Yes	Bidirectional	Yes	Dependent on self-reported data; exhibits reliability gap
Proposed Sarthi Evolution	Content-based + Bayesian Reputation	Yes	Verified, bidirectional	Yes	Must balance verification overhead with accessibility

The comparative landscape reveals a clear evolutionary trajectory, but also a consistent blind spot. Traditional platforms and early digitization efforts fail to leverage computational matching techniques, capping their scalability and coordinative precision at the limits of human bandwidth. The baseline Sarthi system represented a meaningful advance by introducing AI-driven matching and gamification achieving a matching precision of 0.71 and a task completion rate of 92.5% in its pilot deployment but it remained structurally exposed to the reliability gap because its matching function accepted self-reported data at face value, with no mechanism to validate stated competencies or confirm execution intent.

The proposed evolution directly addresses this exposure. By transitioning from passive recommendation to active, verification-grounded coordination through Bayesian reputation scoring, structured task specification, and pre-commitment validation the framework aims to close the gap between algorithmic confidence and real-world reliability. In doing so, it advances beyond the state of the art not by replacing existing matching logic, but by conditioning it on verified behavioral evidence, which is precisely what prior systems in this domain have lacked.

III. PROPOSED SYSTEM ARCHITECTURE AND METHODOLOGY

A. System Overview

The evolved Sarthi framework is organized around a three-layer architecture that separates concerns of presentation, computation, and persistence into discrete, independently maintainable tiers. This separation is not merely a stylistic choice it is a deliberate engineering decision that enables each layer to scale, be tested, and be modified without cascading disruptions to the others. Think of it as a division of labor: the presentation layer handles what users see and interact with, the application layer handles what the system thinks and decides, and the data layer handles what the system remembers.

The Presentation Layer constitutes the system's interface surface and is implemented as a React.js web portal for NGO administrators and a React Native or Flutter mobile application for volunteer users. The dual-interface design reflects an important asymmetry in how these two user groups interact with the platform administrators require dashboard-level visibility over task pipelines, volunteer rosters, and completion analytics, while volunteers need a lightweight, notification-responsive experience optimized for mobile-first usage patterns common in the Indian context. This layer manages user registration, profile configuration, task browsing, real-time notification delivery, commitment confirmation workflows, and reputation visibility. Critically, interface design prioritizes accessibility for users with varying levels of digital literacy, a non-negotiable constraint given the demographic breadth of the volunteer base.

The Application Layer forms the computational core of the system and is implemented as a Python-based service using the Flask framework. Every intelligent decision the system makes matching, reputation scoring, commitment verification, notification scheduling, and tier classification is executed within this layer. It serves as the mediator between what the user requests through the interface and what the database stores, translating user interactions into algorithmic operations and returning structured recommendations. This layer is also responsible for integrating the Human-in-the-Loop verification protocol, which requires the application tier to manage time-sensitive confirmation cascades precisely 24 hours before each scheduled task.

The Data Layer employs a PostgreSQL relational database to maintain the full operational state of the system. It stores volunteer and NGO profiles, structured task specifications, standardized skill taxonomies, historical interaction logs encompassing acceptances, completions, and no-shows, and the accumulated behavioral parameters that feed the Bayesian reputation model. The relational structure is particularly important here: because reputation scoring depends on longitudinal behavioral records, data integrity and query efficiency at this layer directly determine the accuracy of every matching decision the system makes upstream.

B. Baseline Matching Logic

The foundation of the matching engine is a content-based filtering model that evaluates each volunteer against a posted task using a weighted linear compatibility score. For a volunteer $\mathrm{v_i}$, this score is defined as:

$$\begin{aligned} \mathrm{S(v_i)} = \mathrm{w_1} \cdot \mathrm{SkillMatch} + \mathrm{w_2} \cdot \mathrm{LocationProximity} \\ + \mathrm{w_3} \cdot \mathrm{AvailabilityOverlap} \\ + \mathrm{w_4} \cdot \mathrm{CauseAlignment} \end{aligned}$$

Each component captures a distinct dimension of fit. SkillMatch is computed using Jaccard similarity between the taxonomy of skills required by the task and those listed in the volunteer's profile, producing a normalized overlap coefficient bounded between zero and one. LocationProximity is derived from the Haversine formula applied to the geographic coordinates of the volunteer and the task site, converting physical distance into a proximity score that penalizes assignments where travel burden would be prohibitive. AvailabilityOverlap measures the temporal compatibility between the volunteer's stated schedule and the task's time window. CauseAlignment scores the correspondence between the volunteer's declared interest domains and the NGO's operational sector.

The weights assigned to these components $\mathrm{w_1} = 0.35$, $\mathrm{w_2} = 0.25$, $\mathrm{w_3} = 0.25$, $\mathrm{w_4} = 0.15$ reflect an empirically grounded prioritization in which skill relevance is treated as the primary determinant of match quality, followed by geographic and temporal feasibility, with cause alignment serving as a secondary preference signal. This baseline model achieved a matching precision of 0.71 in the Jain et al. (2026) pilot deployment, validating its foundational effectiveness. However, as the reliability gap analysis established, precision at the matching stage does not guarantee fidelity at the execution stage when the underlying input data is unverified.

C. High-Fidelity Coordination Mechanisms

The architectural evolution of Sarthi introduces four integrated mechanisms that collectively address the reliability gap by conditioning matching decisions on verified behavioral evidence rather than static self-reported attributes.

Pre-Commitment Validation Protocol.

The first mechanism introduces what can be understood as a deliberate verification checkpoint in the coordination pipeline. Exactly 24 hours before a scheduled task, the system dispatches a confirmation request to each matched volunteer, requiring an explicit digital response a binary commitment signal before the assignment is finalized. Volunteers who do not confirm within the response window are removed from the assignment queue, and the system redistributes their slots to the next eligible candidates. This Human-in-the-Loop design serves a dual purpose: it functions as a behavioral filter that surfaces low-commitment participants before their absence creates operational disruption for the NGO, and it reduces the administrative burden on coordinators by automating the follow-up process that would otherwise require manual outreach. To ensure accessibility across connectivity conditions, an SMS-based fallback permits confirmation via a simple

YES/NO response when push notifications fail to deliver a critical accommodation for users in low-bandwidth regions.

Bayesian Reputation Scoring.

The second mechanism addresses the core epistemological problem of the baseline system: it trusted what volunteers said about themselves rather than what their history demonstrated about them. The Bayesian Reputation Score $R_i$resolves this by modeling each volunteer's reliability as a posterior probability derived from their observed behavioral record. Specifically, $R_i$is defined as the posterior mean of a Beta distribution parameterized by α, the count of successfully completed tasks, and β, the count of failures and no-shows:

$$R_i = \frac{\alpha + 1}{\alpha + \beta + 2}$$

The Laplace smoothing implicit in the $+1$adjustments to both numerator and denominator is particularly important for new volunteers, as it prevents the score from collapsing to zero or inflating to one on the basis of a single interaction a property that ensures the scoring function remains stable and meaningful during early engagement. As a volunteer accumulates more interactions, the posterior mean converges toward their true empirical completion rate, making the score increasingly informative over time.

This reputation score is then incorporated into the matching function as a multiplicative reliability weight, producing the modified compatibility score:

$$S'(v_i) = \left(\sum_{j=1}^{4} w_j \cdot f_j\right) \times (1 + w_5 \cdot R_i)$$

The multiplicative structure is a deliberate design choice. Rather than adding reputation as a fifth additive feature, it applies reliability as a scalar modifier to the entire base score, meaning that a volunteer with high skill and location fit but a poor reliability history will be ranked below a slightly less well-matched volunteer who consistently shows up. This correctly reflects the operational reality that a volunteer who never arrives is worse than no match at all.

Structured Task Specification.

The third mechanism addresses match failures attributable not to the volunteer side of the equation but to the task description side. When NGOs specify task requirements using free-text fields, the resulting data is structurally incompatible with the taxonomic matching logic the algorithm depends on. Ambiguous language, omitted fields, and inconsistent terminology introduce noise that degrades matching precision regardless of how sophisticated the scoring model becomes. Structured Task Specification templates eliminate this source of error by requiring NGOs to define tasks using standardized categorical fields drawn from a controlled skill taxonomy, mandatory geographic coordinate inputs for precise Haversine computation, and explicit attributes for time window, duration, and priority level. This standardization ensures that every task entering the matching pipeline is computationally well-formed, enabling the Jaccard similarity and Haversine distance calculations to operate on clean, consistent inputs.

Multi-Armed Bandit Allocation and Tier Classification.

The fourth mechanism addresses a tension that any reputation-based system must confront: if matching priority is awarded to volunteers with established reliability records, new volunteers can never accumulate the history needed to earn that priority. Left unaddressed, this creates a structural barrier to entry that concentrates task assignments among a shrinking pool of high-reputation users, ultimately reducing system resilience and raising equity concerns.

The solution draws from the exploration-exploitation framework of reinforcement learning. Volunteers are classified into three tiers Novice, for new or unverified users; Verified, for those with consistent completion records; and Priority, reserved for volunteers maintaining completion rates above 95%. An epsilon-greedy Multi-Armed Bandit strategy then governs assignment allocation: 85% of available task slots are allocated to Verified and Priority volunteers through exploitation of known reliability, while the remaining $\epsilon = 0.15$allocation is reserved for the Shadowing Tier a pool of Novice volunteers who are paired with low-risk tasks under the mentorship of Priority users. This 15% exploration budget serves simultaneously as an ethical guardrail and a system health mechanism, ensuring that the volunteer pipeline is continuously replenished with emerging participants rather than becoming dependent on a fixed cohort of long-tenured contributors.

IV. EXPERIMENTAL SETUP AND EVALUATION

4.1 Dataset Description and Baseline Context

To evaluate the proposed architectural improvements, this study uses the 12-week pilot dataset collected during our prior deployment of the Sarthi platform [1]. This dataset serves as the historical baseline (control group) against which the performance of the proposed coordination protocols is projected and analyzed. The dataset captures real-world algorithmic routing behavior and user interactions across multiple Indian NGO sectors, including education, healthcare, and community support.

Table 2: Historical Pilot Dataset Characteristics

Component	Count	Description
Volunteers	12	Annotated with demographic data, 2–5 skill tags, and availability windows
NGOs	5	Spanning healthcare, education, and community support domains
Tasks Posted	40	Each specifying 2–4 required skills, temporal windows, and locations
Tasks Completed	37	Serving as ground truth for match success evaluation
Notifications Sent	140	Automated dispatches via algorithmic matching cascades
Interactions Logged	~120	Bilateral user interactions, acceptances, and feedback events

The gap between the 140 automated notifications dispatched and the ~120 logged bilateral interactions illustrate a consistent drop-off between algorithmic recommendation and actual human engagement a key inefficiency this work aims to address.

4.2 Match Failure Taxonomy: Quantifying the Reliability Gap

To understand the specific sources of matching inefficiency, we performed a qualitative analysis of unmatched and unsuccessful assignments from the prior pilot [1]. This analysis revealed four distinct failure categories, each of which is directly addressed by a proposed intervention described in Section III:

- No-shows or Cancellations (34%): The most prevalent failure mode, occurring when a volunteer accepted a digital assignment but did not physically execute the task. This justifies the introduction of the 24-hour pre-commitment protocol (Section III), which serves as a behavioral verification filter before final assignment confirmation.
- Task Requirement Gaps (28%): Failures caused by inadequate or ambiguous task descriptions provided by NGO administrators, leading to poor volunteer-task fit. This motivates the Structured Task Specification (STS) template (Section III), which enforces structured input fields compatible with Jaccard similarity and Haversine distance constraints.
- Skill Inaccuracy (23%): Failures resulting from a mismatch between self-reported volunteer skills and actual task requirements. Addressing this is the primary motivation for the Bayesian Reputation Score (Section III), which progressively weights historical performance over unverified self-reporting.
- Communication Failures (15%): Downstream coordination breakdowns, including missed in-app messages and logistical misunderstandings between NGOs and assigned volunteers.

4.3 Evaluation Metrics

System performance is evaluated using standard information retrieval metrics, treating successful task completion with bilateral positive feedback as the ground truth.

- Precision (TP / TP + FP): The fraction of predicted matches that resulted in successful task completion. A True Positive (TP) is defined as a completed task with positive feedback from both the NGO and the volunteer; a False Positive (FP) is an assignment resulting in a no-show, skill mismatch, or mutual dissatisfaction.
- Recall (TP / TP + FN): The fraction of successful matches relative to all valid opportunities. A False Negative (FN) indicates a qualified volunteer was present in the system but was not routed to a suitable task by the algorithm.
- F1-Score (2 · Precision · Recall / Precision + Recall): The harmonic mean of precision and recall, providing a balanced measure of overall matching robustness.

4.4 Comparative Analysis Setup

The proposed framework is evaluated by comparing simulated projections against the historical baseline established in [1]. During the 12-week control period, the baseline matching engine achieved a Precision of 0.71, a Recall of 0.65, and an F1-Score of 0.68, with an overall volunteer absenteeism rate of 22%.

Based on the failure taxonomy in Section 4.2, we project that integrating the Bayesian Reputation Score and the 24-hour pre-commitment protocol will reduce false positives attributable to no-shows and skill inaccuracies. Under this projection, the proposed framework is expected to reduce the aggregate attrition rate from 22% to approximately 12–15%, and improve matching precision from the 0.71 baseline to a target of 0.85. These projections are derived from the proportional failure contributions identified in Section 4.2 and are intended to be validated in a future live deployment study.

V. RESULTS AND PERFORMANCE ANALYSIS

5.1 Historical Pilot Benchmarks (The Sarthi Baseline)

To evaluate the proposed architectural improvements in context, we first establish the performance of the original Sarthi platform from our prior 12-week pilot [1]. During this period, the system processed 40 task postings, of which 37 were successfully completed, yielding an overall task completion rate of 92.5%.

For quantitative evaluation, a held-out test set comprising 15 tasks from the final four weeks of the pilot was used. On this set, the baseline matching engine using standard content-based filtering achieved a Precision of ≈0.71, a Recall of ≈0.65, and an F1-Score of ≈0.68. While these results indicate that the algorithm was reasonably effective at identifying suitable candidates, post-pilot qualitative analysis revealed a persistent reliability gap: algorithmic precision did not consistently translate into behavioral follow-through by volunteers, as reflected in the 22% absenteeism rate reported in Section IV.

5.2 Comparative Performance Overview

Building on the limitations identified in the baseline, Table 3 summarizes the performance trajectory across three stages: manual coordination, the Sarthi baseline, and the simulation-based targets for the proposed framework.

Table 3: Comparative Performance Benchmarks and Projected Targets

Evaluation Metric	Manual Coordination	Sarthi Baseline (Content-Based)	Proposed Evolution (Simulated Targets)
No-show / Attrition Rate	38%	22%	12–15%
Repeat Engagement	31%	67–71%	>71%
Matching Precision	N/A	0.71	0.85

It is important to note that the values listed under the proposed evolution are simulation-based estimates derived from the failure taxonomy presented in Section IV. Empirical validation through a live deployment study is planned as future work.

5.3 Justification of Projected Targets

Each projected improvement in Table 3 is directly linked to a specific intervention introduced in Section III.

Reducing Attrition Rate to 12–15%: The baseline system already improved the manual no-show rate from 38% to 22%. However, the remaining 22% represents failures that purely algorithmic matching cannot address, as they stem from post-acceptance behavioral drop-off rather than poor candidate selection. The projected reduction to 12–15% is attributed to the 24-hour pre-commitment protocol (Section III), which re-verifies volunteer intent before final assignment confirmation. By introducing this human-in-the-loop checkpoint, unconfirmed volunteers are removed from the assignment pool proactively, reducing no-shows before they occur.

Improving Matching Precision from 0.71 to 0.85: The baseline precision of 0.71 was limited by two factors: ambiguous task descriptions and inaccurate self-reported volunteer skills, both of which inflated the false positive rate. The projected precision of 0.85 is estimated based on the combined effect of the Bayesian Reputation Score (Section III), which progressively down-weights unverified self-reporting in favor of historical performance, and the Structured Task Specification (STS) templates (Section III), which replace free-text descriptions with deterministic Jaccard similarity and Haversine distance computations.

Sustaining Repeat Engagement Beyond 71%: The baseline platform demonstrated that gamified elements such as achievement badges and points increased repeat volunteer participation from a manual baseline of 31% to 67–71%. A known risk of introducing stricter verification and Bayesian scoring is the potential marginalization of new users who lack a performance history. To mitigate this, the proposed framework incorporates an epsilon-greedy Multi-Armed Bandit strategy (Section III), which allocates a 15% exploration budget to a dedicated Shadowing Tier. This ensures new volunteers receive equitable onboarding opportunities to build their reputation scores, sustaining long-term community engagement without compromising match quality.

VI. DISCUSSION AND SYSTEMIC IMPLICATIONS

6.1 Behavioral Implications and Sustained Engagement

The baseline Sarthi deployment demonstrated that gamification is effective at sustaining volunteer involvement over time. Users who engaged with the achievement-based recognition system exhibited a 71% repeat participation rate, notably higher than those who did not engage with these features. However, as the reliability gap identified in Section IV indicates, passive digital engagement does not inherently guarantee task execution. A volunteer who is enthusiastic on the platform may still fail to show up. The 24-hour pre-commitment protocol addresses this directly by introducing a human-in-the-loop checkpoint that re-verifies intent before final assignment. This ensures that gamified enthusiasm translates into verifiable physical participation rather than just digital activity. At the same time, the epsilon-greedy Multi-Armed Bandit strategy ensures this stricter filtering does not disadvantage new users the Shadowing Tier provides unverified volunteers with onboarding opportunities to gradually build their Bayesian Reputation Scores without facing immediate algorithmic exclusion.

6.2 Technical Mitigations for Resource-Constrained Environments

A key limitation identified in the pilot was the system's reliance on stable internet connectivity and smartphone ownership factors that risk excluding volunteers from lower-income or infrastructure-constrained communities. This is a particularly relevant concern given the platform's deployment context across rural and semi-urban India.

To prevent the 24-hour confirmation protocol from inadvertently reinforcing digital exclusion, the proposed architecture incorporates an SMS-based fallback mechanism. When in-app push notifications fail to deliver due to connectivity constraints, the system automatically dispatches an SMS nudge, allowing volunteers to confirm their commitment via a simple YES/NO text response. This design choice aligns with established recommendations in socio-technical literature for minimal bandwidth requirements and offline-first capabilities in digital coordination systems deployed in developing-country contexts.

6.3 Operational Friction and Usability Trade-offs

The shift from passive recommendation to high-fidelity coordination inevitably introduces some degree of operational friction. The baseline Sarthi platform prioritized speed and ease of access, achieving an average algorithmic execution time of 0.8 seconds and a user satisfaction rating of 4.2 out of 5.0. However, this emphasis on frictionless onboarding contributed directly to the informational asymmetries and task ambiguities that undermined match reliability.

The introduction of Structured Task Specification (STS) templates and the 24-hour verification loop intentionally slows this process in exchange for greater data integrity. This is a deliberate trade-off: some upfront friction is necessary to enable accurate Bayesian Reputation Score computation and reliable Jaccard similarity matching. The key design challenge going forward will be minimizing the cognitive load and number of interaction steps associated with these verification gates, so that the platform retains its accessibility and user satisfaction while achieving the reliability improvements targeted in Section V.

VII. ETHICAL CONSIDERATIONS AND DATA PRIVACY

7.1 The Tiering Divide and Algorithmic Equity

While the integration of Bayesian Reputation Scoring and reliability-based tiering effectively filters for high-intent individuals, it inherently introduces the risk of codifying digital exclusion. Volunteers operating in lower-income or resource-constrained environments may experience task failures due to legitimate infrastructure barriers such as intermittent internet connectivity, older device incompatibility, or varying levels of digital literacy rather than a lack of actual behavioral commitment. Penalizing these users algorithmically would directly contradict the platform's social mandate. To prevent the marginalization of these demographics, the proposed architecture embeds strict algorithmic guardrails. Specifically, the epsilon-greedy Multi-Armed Bandit strategy functions as a structural countermeasure by permanently reserving a 15% opportunity exploration budget for the "Shadowing Tier". This mechanism guarantees that new, unverified volunteers or those who have suffered rating drops due to infrastructural friction are not systemically excluded from the ecosystem. Instead, they are continuously routed to low-risk tasks or paired alongside historically reliable "Priority" mentors, ensuring they are afforded equitable, ongoing opportunities to build their verification metrics.

7.2 Data Integrity and Privacy Preservation

The transition toward high-fidelity behavioral verification fundamentally requires the processing of granular user data, necessitating rigorous ethical governance and adherence to minimal data collection principles. In compliance with established data protection frameworks, such as India's Data Protection Bill (DPDP, 2023), the Sarthi platform incorporates strict privacy-preserving protocols throughout its architecture.

All personally identifiable information (PII) is systematically anonymized, replacing sensitive volunteer and NGO identities with system-generated identifiers prior to any algorithmic processing or analytical evaluation. Furthermore, the data layer enforces robust role-based access controls and utilizes secure authentication mechanisms (e.g., JWT and OAuth2.0) to ensure that sensitive skill verifications and historical data are shielded from unauthorized access. Crucially, the platform operates on a strictly non-monetized model, explicitly prohibiting the sale of user data or third-party access to the organizational database. The information ingested by the system is utilized exclusively for facilitating operational matching, calculating platform reliability metrics, and advancing social impact research

VIII. CONCLUSION

The architectural evolution of the Sarthi framework presented in this study demonstrates that effective digital NGO–volunteer coordination cannot rely solely on the optimization of recommendation algorithms; it strictly requires a synthesis of computational intelligence and behavioral rigor. While foundational content-based filtering successfully aligns user capabilities with organizational needs, it is structurally insufficient to overcome the "last mile" reliability gap the critical divergence between a digital match and verified physical execution.

By transitioning the system from a passive recommendation engine to an active, high-fidelity coordination platform, this research systematically addresses the primary modes of behavioral attrition and informational asymmetry identified in prior pilot deployments. The integration of the 24-hour pre-commitment protocol introduces necessary "strategic slowing" as a Human-in-the-Loop behavioral filter, effectively converting self-reported intent into a verifiable commitment. Concurrently, the implementation of Structured Task Specification (STS) templates and a Bayesian Reputation Score replaces ambiguous, unstructured inputs with deterministic, empirically grounded metrics. These mechanisms work in concert to computationally de-weight unreliability, driving targeted reductions in volunteer attrition from 22% to 12–15% while elevating projected matching precision from 0.71 to 0.85.

Crucially, this framework proves that rigorous behavioral verification need not come at the expense of social equity. Through the deployment of an epsilon-greedy Multi-Armed Bandit strategy, the permanent 15% Shadowing Tier ensures that new and historically unverified volunteers are ethically onboarded. This mechanism successfully mitigates the algorithmic cold-start problem, ensuring that systemic reliability does not inadvertently codify digital exclusion for marginalized users.

Future research directions will focus on empirically validating these projected benchmarks through a large-scale, longitudinal deployment. To further bridge infrastructural and demographic divides, operational expansion will prioritize the integration of comprehensive multilingual support, offline-first mobile capabilities leveraging the SMS-fallback architecture, and the scaling of the Multi-Armed Bandit implementation across broader, more diverse geographic contexts. Ultimately, the evolved Sarthi framework provides a robust, highly replicable, and equity-conscious blueprint for maximizing human capital and translating digital coordination into reliable social impact within resource-constrained civic environments.

REFERENCES

[1] "Artificial Intelligence-Enabled Volunteer Services: Intrinsic Logic, Practical Dilemmas, and Adjustment Strategies," Recent Academic Literature, 2026.

[2] "Combining a Gamified Civic Engagement Platform with a Digital Game in a Loosely Way to Increase Retention," in Proc. ACM Int. Conf. Information Integration and Web-based Applications and Services (iiWAS), 2019.

[3] "Enhancing Content-based Recommendation via Large Language Model," in Proc. ACM Int. Conf. Information and Knowledge Management (CIKM), 2024.

[4] "Information Technology Solutions to Support Volunteer Projects," in CEUR Workshop Proc. Computational Intelligence Workshop, 2025.

[5] A. Jain, A. Kumar, P. Gautam, and S. Yadav, "Sarthi – The Helping Hand: An AI-Assisted Framework for Optimizing NGO–Volunteer Coordination," International Journal of Innovative Research in Technology, vol. 12, no. 9, 2026, ISSN: 2349-6002.

[6] "Transparency and trust in digital marketplaces: Constructs and consumer behavior implications," Journal of Marketing Science Research, vol. 18, no. 4, pp. 55–70, 2025.

[7] "NGO Portal - A Platform to Connect NGOs with Prospective Members," in Proc. IEEE Conf. Interdisciplinary Research, 2022.

[8] "Recommender Systems Applications: Data Sources, Features, and Challenges," IEEE Xplore/ACM Repository Review, 2024.

[9] P. Schönböck and J. Raab, "Volunteer management systems: A survey on current practices and limitations," Journal of Nonprofit Technology, vol. 8, no. 2, pp. 45–62, 2016.

[10] "The Game of Life: Designing a Gamification System to Increase Current Volunteer Participation and Retention in Volunteer-based Nonprofit," Digital Commons.

[11] X. Wu, "Designing volunteer management systems using low-code platforms," International Journal of Information Systems, vol. 15, no. 1, pp. 23–35, 2023.