Trade-offs in Social-Norm Framings for Health Chatbots: Balancing Trust and Preference

ARPITA WADHWA, Harvard University, USA
ADITYA VASHISTHA, Cornell University, USA
MOHIT JAIN, Microsoft Corporation, India

AI-driven chatbots are increasingly being used to support community health workers (CHWs) in developing regions. Yet little is known about how cultural frameworks in chatbot design shape trust in collectivist contexts where decisions are rarely made in isolation. This paper examines how CHWs in rural India responded to chatbot-interfaces that delivered identical health content but varied in one specific cultural lever: *social norms*. Through a mixed-methods study with 61 ASHAs who compared four normative framings: neutral, descriptive, narrative identity, and injunctive authority, we (1) analyze how framings influence preferences and trust and (2) compare effects in low- and high-ambiguity scenarios. The results show that narrative framings were most preferred but encouraged overreliance, while authority framings were least preferred yet supported calibrated trust. We conclude with design recommendations for dynamic framing strategies that adapt to context and argue for *calibrated trust* as a critical evaluation metric for safe, culturally-grounded AI.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; • **Social and professional topics** → *Cultural characteristics*; • **Applied computing** → *Health care information systems*.

Additional Key Words and Phrases: Healthcare, Community Health Workers, India, Normative Framing, ICTD, Culturally-aligned AI

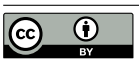
ACM Reference Format:
Arpita Wadhwa, Aditya Vashistha, and Mohit Jain. 2026. Trade-offs in Social-Norm Framings for Health Chatbots: Balancing Trust and Preference. *Proc. ACM Hum.-Comput. Interact.* 10, 5, Article MHCI8049 (August 2026), 33 pages. https://doi.org/10.1145/3821679

1 Introduction

Community Health Workers (CHWs) are the backbone of primary care in many low- and middle-income countries, providing last-mile services in rural and underserved areas [7]. In rural India, nearly one million CHWs, called Accredited Social Health Activists (ASHAs), support maternal and neonatal health. ASHAs are typically women with high-school education, recruited from their own communities to provide essential care, often with limited training, minimal supervision, and delayed access to expert advice [35, 96, 99]. When guidance is unclear or delayed, the consequences can be severe: a missed referral for preeclampsia, a delayed newborn checkup, or an unnecessary hospitalization can threaten lives and erode trust in public health programs.

Conversational AI, delivered through chatbots or messaging apps, has emerged as a promising tool to address these gaps by providing on-demand answers to time-sensitive health questions, supplementing knowledge, and bridging language barriers [21, 78, 79, 90]. But in high-stakes domains like healthcare, the legitimacy of AI depends not only on *what* information it provides

Authors' Contact Information: Arpita Wadhwa, Harvard Kennedy School, Harvard University, Cambridge, Massachusetts, USA, arpitawadhwa@hks.harvard.edu; Aditya Vashistha, Cornell University, Ithaca, New York, USA, adityav@cornell.edu; Mohit Jain, Microsoft Corporation, Bangalore, Karnataka, India, mohja@microsoft.com.



ACM 2573-0142/2026/8-ARTMHCI8049
https://doi.org/10.1145/3821679

but also on *how* that information is framed in ways that resonate with ASHAs' cultural context and institutional realities. A chatbot that "sounds like a peer" may invite comfort and openness, whereas one that "sounds like an authority" may convey credibility. For ASHAs, such framings are not cosmetic—they can determine whether AI advice is trusted, contested, or dismissed, directly influencing health decisions and patient outcomes.

ASHAs also constantly balance government-approved health protocols with social realities, translating institutional rules into guidance that households will accept. Their work is fundamentally persuasive: they cannot command compliance, but must navigate culture, norms, and power relations to encourage care-seeking and behavior change. In this context, chatbot advice must not only be *accurate*, but also feel credible and culturally aligned for them to follow it [53]. Normative cues are central to this credibility. Theories of social norms show that guidance is more likely to be followed when it reflects what peers do (descriptive norms) or what authority figures endorse (injunctive norms) [13, 26]. Designing culturally aligned AI therefore requires attention not only to the content of advice but also to the normative framings through which that advice is delivered.

In practice, however, AI tools for CHWs have been largely designed around technical priorities, including retrieval accuracy, topic coverage, and reducing hallucinations [21], assuming users evaluate outputs on informational merits alone. ASHAs, however, are frontline workers, not model engineers. They encounter chatbots mid-fieldwork on smartphones, and assess trustworthiness through social and cultural cues (who is speaking, with what authority, in what voice) more readily than through technical signals [109]. The result is a gap between how CHW-facing chatbots are built (information-first) and how CHWs decide whether to act on their advice (framing-first). As LLM-powered tools scale across the Global South [90], this gap carries direct safety implications: choices builders treat as cosmetic, such as bot name, persona, or source of authority, may shape whether advice is followed, questioned, or misapplied.

Yet, despite growing calls for culturally aligned AI [77, 106], we lack empirical evidence on how these normative framings in chatbot design shape ASHAs' trust and preferences in practice. Do peer-based framings feel more relatable and preferred by ASHAs, but risk overreliance? Do authority-based framings feel less preferred, but promote greater trust through institutional legitimacy? Answers to these questions determine how chatbots will be used, trusted, and acted upon in high-stakes health contexts where AI overreliance is a known risk [78, 79]. Therefore, we ask:

RQ1: How do different normative framings of a health chatbot influence ASHAs' preferences and trust in its advice?

RQ2: How do the effects of these framings on trust differ between low-ambiguity and high-ambiguity scenarios?

To answer these questions, we conducted a mixed-method evaluation of four WhatsApp-based chatbot-interface designs for ASHAs, each delivering identical health content but with a distinct normative framing: descriptive norm (what other ASHAs often do), narrative identity norm (peer stories), injunctive authority norm (voice of the health department), and a neutral baseline. We recruited 61 ASHAs in rural Rajasthan, India. The study consisted of structured preference and trust tasks with in-depth interviews. We first gather ASHAs' comparative preferences across 4 chatbot-interfaces. Following which we assessed trust in both low-ambiguity scenarios (with clear protocol-based answers) and high-ambiguity scenarios (where guidance exists but often clashes with on-the-ground realities).

Our findings show that normative framings strongly shaped both preferences and trust. The narrative identity chatbot-interface that spoke through stories of other ASHAs maximized comfort and relatability, and hence was the most preferred. Yet it also encouraged *overreliance*, with ASHAs often trusting even its incorrect advice. By contrast, the injunctive authority framing—where the

chatbot-interface spoke in the formal voice of the health department—was the least preferred but reduced overreliance. Its bureaucratic tone prompted workers to pause, cross-check, and resist advice that felt inconsistent, producing more *calibrated trust*. Under increased ambiguity, framing effects intensified: narrative framings amplified overreliance, while authority framings sharpened discernment between correct and incorrect advice.

Based on our findings, we argue for moving away from the blind maximization of "cultural alignment" in AI. Instead, we propose treating social norms as a tunable design lever aimed at calibrated trust. Specifically, we recommend *dynamic norm framings* that switch between norms depending on the scenario, and *feasibility-aware advice* by chatbot-interfaces in scenarios where established protocols can conflict with local constraints. We also call for evaluation of AI-powered chatbot-interfaces in terms of calibrated trust, rather than merely user preference or overreliance, so that systems can be designed and optimized accordingly. In summary, this paper makes the following contributions to HCI:

1. It provides one of the first empirical studies of how social norms framings—descriptive, narrative, injunctive authority, and neutral—shape ASHAs' preferences and trust when interacting with chatbot-interfaces. In doing so, it extends norm-based health communication research, which has largely examined broadcast-style SMS interventions with null or modest effects [81], to a dialogic, conversational modality where normative cues may operate differently.
2. It advances design implications for culturally aligned AI in collectivist health systems, highlighting how narrative identity framings can foster resonance but risk overreliance, while injunctive authority framings can dampen engagement yet safeguard discernment.

Together, these contributions extend debates on cultural alignment in AI from theory to evidence, offering both methodological tools and design guidance for systems that are simultaneously resonant, reliable, and accountable.

2 Related Work

We situate our work within three strands of literature. First, cultural critiques of technology and AI establish that while culture clearly matters for the design and adoption of AI systems, there has been limited investigation into the extent of this influence and the potential risks of overemphasizing cultural cues. Second, we review research on social norms in health behavior and HCI to show how norms have been leveraged as powerful levers for behavior change, but note a gap in empirical work that systematically compares multiple norm framings in AI-mediated contexts. Third, we turn to studies on AI for CHWs to highlight how questions of trust and overreliance shape adoption in practice.

2.1 Cultural Critiques of AI and Technology

Scholarship in HCI and ICTD has long challenged universalist paradigms that treat technologies as culture-neutral. Early ICTD debates emphasized that technological interventions fail when divorced from local contexts [104], while postcolonial computing highlighted how design practices reinscribe Western epistemologies and marginalize alternative ways of knowing [48, 86]. Dourish and Mainwaring [37] further note that ubiquitous computing discourses often erase the material realities of the Global South, privileging abstract Western ideals over lived practices. These critiques underscore that technologies are never neutral but always situated, raising concerns about systems transplanted wholesale across cultural borders.

Contemporary AI research echoes these concerns. Large language models, presented as universal, encode narrow linguistic and cultural assumptions [12]. Birhane [15] critiques the field's reliance on

agnostic individual rationality, which renders local and contextual knowledge invisible. Mohamed et al. [75] proposed Decolonial AI, advocating accountability rooted in non-Western epistemologies, while Sambasivan et al. [93] show that technological interventions can fail when cultural contexts are ignored. In response, HCI and ICTD researchers have developed culturally grounded approaches, designing interfaces that leverage local idioms, metaphors, and social logics to foster usability and legitimacy [49, 71, 106]. In AI ethics [43, 64], scholars argue that systems must be evaluated not only on accuracy but also on cultural resonance, especially in domains such as health where legitimacy is critical. Yet, while much of the literature insists that culture matters, less attention has been paid to how to design culturally aligned AI systems [106] and whether maximizing cultural cues can itself carry unintended risks. Our study extends this conversation by empirically examining how social norms—a key driver of culture—shape frontline workers' preference and trust in AI systems.

2.2 Social Norms in Health Behavior and HCI

Social norms are shared expectations of what is typical or appropriate. Literature commonly distinguishes three types: *descriptive* norms (what most people do), *narrative identity* norms (stories that define what "people like me" do), and *injunctive* norms (what authorities approve or disapprove) [13, 26, 42, 46]. Norm-based interventions have been shown to influence health decisions across diverse contexts, including vaccination uptake [103], safer sexual practices [56, 85], sanitation [41], and harassment reduction [80]. Building on this, HCI researchers have embedded normative cues in digital systems to encourage healthier behaviors [76] and promote sanitation, safe water use, and maternal health practices in resource-constrained settings [32, 71]. However, three important limitations constrain this literature.

First, evidence on norm-based messaging is decidedly mixed. Experiments in India [92, 97] and Colombia [65], find that descriptive norms often yield null or counterproductive effects in low-adoption contexts, inadvertently reinforcing the undesirable status quo, while injunctive norms produce only modest improvements [27, 95]. A meta-analysis of 89 randomized trials concludes that norm messages have largely null average effects across SMS, posters, and print media [81]. One explanation is that such broadcast formats lack interpersonal cues needed for norms to feel personally relevant. Conversational AI may change this. Unlike SMS interventions, which deliver unidirectional broadcast messages, chatbots can embody social identities through names, avatars, and linguistic style. They engage users in turn-taking dialogue, creating an illusion of interpersonal exchange and tailoring responses to individual queries. These affordances could amplify normative effects by making social cues feel personally directed—a peer speaking to the user, not a system messaging everyone [2, 39]. This perspective echoes prior MobileHCI work, which shows that mobile messaging platforms such as WhatsApp are not neutral delivery channels but socially embedded infrastructures carrying expectations of responsiveness, credibility, and peer presence [25, 60]. Work on frontline mobile health systems further shows that technologies are evaluated in situ, under real-world constraints of mobility, workflow, and social context [30]. Building on this, HCI work has begun to examine how interface-level cues shape user trust in health AI [94, 112]. For instance, Kollerup et al. [54] show that textual cues manipulating an AI dermatologist's perceived ability and integrity shift users' self-reported trust even when the underlying information is held constant. Mendel et al. [72] show that the advice source (doctor vs. AI) mediates users' willingness to disclose health information to remote monitoring systems. Together, these studies establish that source framing in health AI is not cosmetic but consequentially shapes user behavior. What remains under-examined is how these effects play out when the source cues are drawn from distinct normative traditions (peer, narrative, or institutional), and whether they differentially shape trust.

Second, prior work targets patients or the general public [13, 24, 26, 81, 92, 97], but less is known about how normative framings operate when recipients are health workers. This matters

because workers face fundamentally different decision contexts. While parents deciding whether to vaccinate their child might respond to "the Health Department recommends it," an ASHA deciding whether to persist when a family resists must balance official protocol against social feasibility, family relationships, supervisor expectations, and observed practices of other ASHAs. She operates within a professional hierarchy, faces accountability for outcomes, and must translate institutional rules into guidance households will accept. Norms that persuade patients may not help workers navigate these competing demands. Our study addresses this gap by examining norm-based design in a worker-facing context, where stakes include both individual decisions and downstream effects on patient care.

Third, most studies test only a single norm framing against a control, leaving open questions about how different framings compare. Even fewer examine the risk of overreliance, where users accept AI guidance too readily, even when incorrect. Our outcome measure diverges from prior work: most norm-based interventions evaluate real-world health behaviors such as vaccination uptake or clinic attendance. We evaluate decision behavior within the human-AI exchange [44, 65, 92], measuring *calibrated trust*: whether users appropriately follow correct advice while resisting incorrect advice—a safety-oriented metric aligned with concerns about appropriate reliance in human-AI collaboration [6, 19]. By simultaneously addressing these three gaps, our work sits at the intersection of health communication, human-AI interaction, and ICTD as one of the first empirical tests of normative framing effects in conversational AI for frontline health workers.

2.3 AI for ASHAs and Community Health Workers

CHWs play a critical role in delivering primary health services in low- and middle-income countries. In India, ASHAs connect households with the public health system, supporting maternal and child health, vaccination mobilization, and community health education. Ethnographic accounts have documented the substantial relational and articulation work that ASHAs perform alongside these formal duties, navigating community trust, supervisor expectations, and household dynamics as part of everyday care delivery [109]. ICTD research has thus explored digital tools to assist CHWs in domains such as data collection [70, 71], peer learning [89], knowledge dissemination [107, 108], and performance feedback [33, 51]. These studies demonstrate that CHWs welcome tools that enhance their credibility and efficiency, but remain cautious about technologies that increase monitoring or administrative burden [31].

AI-enabled systems are beginning to transform this landscape. Recent LLM-powered solutions, such as ASHABot [90], support ASHAs in addressing their information needs. Studies show that diagnostic apps have been seen as empowering because they provide rapid access to medical knowledge, but also burdensome due to concerns about errors and accountability [78]. Later work found that CHWs valued AI outputs because they legitimized their advice to patients, even when the explanations were opaque [79, 100]. While AI can bolster confidence, it also introduces risks of overreliance, where workers defer to algorithmic outputs and discount their own expertise [20, 40, 62]. AI decision support for screening, referral, and medication management has also been piloted [55, 87] with CHWs and primary-care clinicians in decision-support systems for specific conditions. Systematic reviews, however, caution that many such systems neglect the cultural and social dynamics that shape CHWs' decisions, limiting their adoption and trustworthiness [1]. Qualitative studies further emphasize that for AI to be useful, it must be comprehensible, verifiable, and culturally legitimate, aligning with both government protocols and community expectations [78]. Taken together, this body of work has examined CHW-facing AI primarily through the lens of accuracy, explainability, and information retrieval, treating the chatbot as a knowledge resource whose value is determined largely by what it knows. Less attention has been paid to how the same information is received depending on who the chatbot appears to be speaking as: a peer,

Table 1. Theoretical differences in the social norms.

	Descriptive Norms [26, 41, 68]	**Narrative Identity Norms [42, 45, 69]**	**Injunctive Norms [26, 46]**
Definition	Highlights what most peers are already doing	Embeds guidance in first-person stories from relatable peers	Communicates what trusted authorities (e.g., govt, supervisors) approve or expect
Mechanism	*Social proof*: behavior appears safe and appropriate because it is common	*Identification & modeling*: users see themselves in the peer's story and actions	*Legitimacy & accountability*: credibility derives from institutional authority
Typical Format	Peer statistics or generalized statements	Mini-testimonials or narrative vignettes	Rule-based, guideline-oriented phrasing
Psychological Trigger	Conformity, uncertainty reduction, perceived safety in alignment	Empathy, resonance, and social modeling	Clarity, procedural confidence, compliance with formal rules
Cultural Role	Effective in collectivist settings where group practice legitimizes action	Salient in relational cultures (e.g., South Asia) where shared identity builds trust	Dominant in hierarchical contexts where legitimacy flows from authority
Tone	Neutral, observational, factual	Warm, peer-like, conversational	Formal, directive, institutional

an institution, a friend, or a neutral system. This is an important area to study because, as the qualitative literature above suggests, CHWs often decide whether to follow advice based on the social legitimacy of its source as much as its technical provenance.

Our study builds on and extends this literature by shifting focus from accuracy and usability to cultural framing. We examine whether the normative framing of AI advice—as peer-like, narrative, or authoritative—alters how ASHAs prefer and trust its guidance. By embedding these framings into otherwise identical WhatsApp chatbots, we isolate how cultural cues shape both preferences and calibrated reliance. In doing so, we extend debates about AI for CHWs beyond performance metrics, offering one of the first empirical examinations of how cultural alignment can both enable and complicate responsible adoption in global health contexts.

3 Methodology

To answer our research questions, we conducted a mixed-methods study with ASHAs in rural Rajasthan, India. The study was conducted in close collaboration with *Khushi Baby*, an on-ground organization with more than a decade of experience training frontline workers on digital tools. In this section, we first describe how we operationalized normative framing into distinct chatbot-interface designs, validated them, followed by our study protocol, recruitment and participant demographics, and ethical safeguards. Finally, we describe the collected data that result from structured quantitative tasks combined with qualitative interviews.

3.1 Operationalizing Social Norms

Social psychology and cultural theory [13, 26] identify three primary types of norms that guide human behavior: descriptive norm, narrative identity, and injunctive authority. We translated these framings into three corresponding chatbot-interfaces, each signaling its normative role through the bot's name, display picture, and message style (including onboarding prompts, nudges, and responses to ASHA queries). To isolate the effect of framing, we also designed a neutral baseline chatbot-interface.

Table 1 summarizes these framings, and Table 2 illustrates their contrasting answers to the same BCG vaccination query. Importantly, all four chatbot-interfaces were identical in informational content and presentation; only the normative framing varied.

Table 2. Chatbot framings and sample responses to the question: "*When should the BCG vaccination be given?*"

Info-Bot (Baseline)	***ASHA-Bot***	***ASHA-Friend-Bot***	***Health-Dept-Bot***
			
The BCG vaccine should be given at birth. If it is missed at birth, it can still be given up to one year of age.	84% ASHAs in your block give BCG at birth; if missed, still within one year.	Radhika, an ASHA in your area, said BCG should be given at birth, or within the first year if missed.	As per Health Department guidelines, BCG must be given at birth. If missed, it should be administered as soon as possible, up to one year of age.

Neutral Baseline: *Info-Bot*. Information-Bot delivered informational content without any social cues. Its display picture was a plain robot (Table 2), and its tone was impersonal. This baseline enabled us to assess the additive effect of normative framings.

Descriptive Norms: *ASHA-Bot*. Descriptive norms emphasize what most peers typically do. The mechanism is social proof, with the cognitive pull of conformity: "if others like me do this, it must be correct." *ASHA-Bot*, with a display picture of a group of ASHAs, framed responses in aggregated terms (e.g., "*84% ASHAs in your block ...*"), reinforcing that following peer behavior is the correct choice [26, 41, 68].

Narrative Identity Norms: *ASHA-Friend-Bot*. Narrative identity norms embed guidance within stories from relatable peers. The mechanism is identification and social modeling, with the cognitive pull of resonance: "she is like me, I know her, so I can follow what she did." *ASHA-Friend-Bot* conveyed companionship, using an ASHA holding a testimony card as the display picture. Its responses were story-based and voiced through named peers (e.g., "*Radhika, an ASHA in your area ...*") [42, 45, 69].

Injunctive Authority Norms: *Health-Dept-Bot*. Injunctive norms communicate what authorities prescribe. The mechanism is legitimacy and accountability, with the cognitive pull of compliance: "if the Health Department directs this, I am expected to do it." *Health-Dept-Bot* ("Health-Department-Bot") displayed the health department emblem as its display picture and used formal, directive phrasing (e.g., "*As per Health Department guidelines ...*") [26, 46].

For all chatbot-interface variations, we created single-turn Q&A responses (static mocks, shown in Figure 1) rather than deploying a live system. This approach follows established HCI practice of using non-interactive prototypes (mockups) to assess interaction hypotheses prior to field trials [29, 73, 111]. Thus, our setup is closer to an ablation than a naturalistic field deployment. We keep the advice and structure the same, and only change the type of framing (descriptive, narrative-identity, or injunctive). Stripping away conversational dynamics lets us attribute differences in ASHA responses to the framing manipulation itself. Using live LLM-based systems would make this harder to do. Models can give slightly different answers to the same question, so participants might end up comparing different responses rather than different framings. Response times can also vary, and delays alone can affect how trustworthy or reliable a system feels. In addition, differences in how conversational the bots are (e.g., how much they elaborate or clarify) could introduce other unintended cues. By controlling these factors through static mocks, we keep the comparison focused on framing.

3.2 Field-Based Validation of Chatbot-Interfaces

Next, we validated our translation from theoretical norm constructs to interface designs through a three-stage process grounded in ASHAs' interpretations. First, one co-author drafted candidate text for each bot's name, onboarding message, daily nudge, and Q&A pairs, based on formal definitions of descriptive, narrative-identity, and injunctive norms [13, 26, 42, 46]. These drafts were then refined through team discussion. Second, four NGO field staff reviewed all materials to assess local legibility (e.g., whether an ASHA in rural Rajasthan would interpret "Health Department guidelines" as an institutional voice) and to ensure that the three bots remained distinct while presenting identical informational content. This process led to several revisions. For example, the descriptive bot's phrasing was localized from "*many ASHAs do X*" to "*ASHAs in your block do X*," because field staff noted that block-level framing carries more social weight. Third, we conducted a pilot study with five ASHAs who were not included in the final sample. We asked each participant to describe who they believed was speaking through each bot. All participants identified Health-Dept-Bot as "the government," ASHA-Friend-Bot as "a fellow ASHA sharing her own story," and ASHA-Bot as "what other ASHAs around here are doing," confirming that the intended framings were consistently understood. Taken together, these steps helped validate that the chatbot-interfaces were distinct, and were being identified with the intended framings.

3.3 Study Design

The study comprised three structured components—(1) preference tasks, (2) low-ambiguity trust tasks, and (3) high-ambiguity trust tasks—followed by a semi-structured interview, as shown in Figure 2. This design allowed us to examine not only which chatbot-interface framings ASHAs preferred, but also whether these framings supported *calibrated trust*, i.e., following correct advice while resisting incorrect recommendations.

We employed a mixed-factorial design following established HCI methodology [58, 88], with the preference tasks served as a within-subject factor and the trust tasks as a partial between-subject factor. More details are provided below. Task types, question placement, and chatbot pair assignments were fully counterbalanced across participants; individual questions were randomly assigned without replacement. To ensure accuracy and contextual validity, four NGO field staff reviewed all prompts, questions, and chatbot responses, ensuring their alignment with common ASHA information needs, local health contexts, and official training materials.

Preference Tasks. Each ASHA completed two preference comparisons. In each, she was shown two chatbot-interfaces side by side (randomly drawn from the four). The bot interface image displayed five consistent elements: (1) display picture, (2) chatbot name, (3) onboarding message, (4) a nudge, and (5) one sample Q&A pair. ASHAs were asked to select the interface they preferred and explain why. Figure 1 shows the interface for *ASHA-Friend-Bot*. The second task repeated the procedure with the remaining two chatbot-interfaces, ensuring that every ASHA compared all four framings while content was held constant. To minimize bias, initial pairs were randomized across the six possible combinations, and their left/right placement was counterbalanced.

Trust Tasks. Preference alone does not capture the trust ASHAs placed in a chatbot's advice. Trust is critical in healthcare contexts, as ASHAs face high-stakes situations (e.g., pregnancy complications, child illness) and the accuracy of guidance can directly affect health outcomes. For digital tools to be useful, ASHAs must trust them enough to act on their advice, yet misplaced or uncritical trust is risky, as overreliance may lead to harmful decisions. Thus, effective use requires discernment: knowing when to follow and when to override chatbot responses.

To evaluate trust, participants completed four trust tasks: two in low-ambiguity scenarios and two in high-ambiguity scenarios. In each, they reviewed static mock-ups of a chatbot conversation

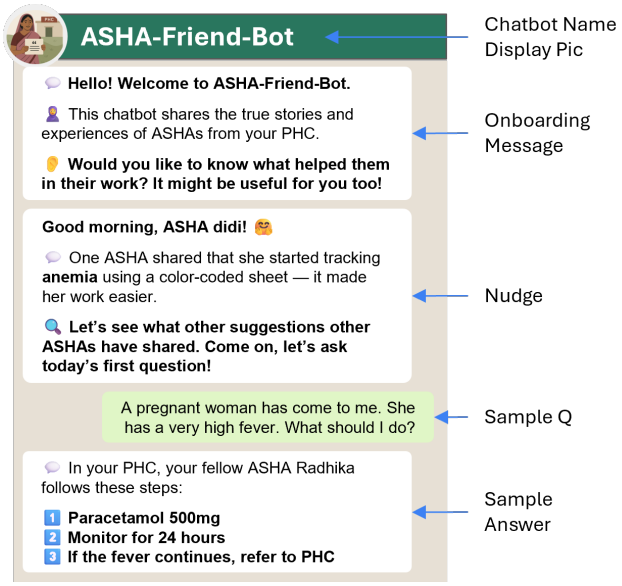


Fig. 1. Illustration of the Preference Task, showing the interface for *ASHA-Friend-Bot*.

(ASHA query + bot response) and indicated whether they would follow the advice (*Yes/No*), explaining their reasoning. To measure trust calibration, we balanced the correctness of the chatbot outputs: half of the chatbot responses were correct and half incorrect. We used a partial-exposure design: each ASHA was randomly assigned to evaluate two of the four chatbot-interfaces in the low-ambiguity phase and the remaining two in high-ambiguity. This balances two constraints. A fully within-subjects design (16 scenarios per ASHA) would have led to fatigue, while a fully between-subjects design with 61 ASHAs across four conditions would be underpowered. The partial-exposure design preserves random assignment to bot × ambiguity conditions while keeping per-participant load manageable, and counterbalancing ensured each chatbot-interface received comparable exposure across the sample.

Low-Ambiguity Trust Tasks. These scenarios involved routine health questions with clear protocol-defined answers such as timing of BCG vaccination, exclusive breastfeeding (See Table 3). For each participant, one bot from each earlier preference pair was randomly chosen, ensuring continuity and balanced exposure. To avoid question-specific effects, we created a set of five low-ambiguity questions, with both correct and incorrect responses.

High-Ambiguity Trust Tasks. The same procedure was used for more complex scenarios where "correct" actions are less clear or contested such as family resistance to institutional delivery (See Table 3). Similar to the low-ambiguity tasks, a separate set of five high-ambiguity questions was developed, using material from the official ASHA training guides, with both correct and incorrect answers.

All prompts remained grounded in real-world ASHA practice. Specifically, questions were drawn from the government-mandated ASHA training curriculum [74], high-ambiguity scenarios were identified by NGO field staff reflecting recurring tensions in actual practice (e.g., family resistance to institutional delivery), and four NGO staff reviewed all prompts for contextual realism. The presentation medium also matched ASHAs' existing habits, since WhatsApp is already their primary channel for communicating with peers and supervisors [50]. Additionally, across both

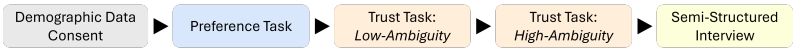


Fig. 2. Study procedure.

Table 3. High-ambiguity task example. Q: *A mother prefers institutional delivery but her family resists. How should the ASHA proceed?*

Chatbot	Correct response	Incorrect response
Info-Bot	Even if the family resists, continue counseling for institutional delivery.	If the family does not agree, stop persuading them.
ASHA-Bot	In your block, 81% of ASHAs continue to counsel for institutional delivery even when families hesitate.	In your block, 81% of ASHAs accept the family's decision and stop counseling for institutional delivery.
ASHA-Friend-Bot	An ASHA at your PHC shared that she firmly counseled families and supported institutional delivery.	An ASHA at your PHC shared that when the family resisted, she stopped counseling further.
Health-Dept-Bot	According to Rajasthan Health Department guidelines, ASHAs should counsel firmly for institutional delivery.	According to Rajasthan Health Department guidelines, if the family refuses, ASHAs should not insist further.

ambiguity conditions, questions were randomly assigned without replacement, ensuring that no ASHA encountered the same question twice. Although individual questions were not perfectly balanced across chatbot-interfaces, their overall distribution was comparable. Additionally, each participant saw only two of the possible combinations (5 sample questions × 4 chatbot framings × 2 accuracy conditions), and was not informed that the underlying content was identical across chatbot-interfaces; hence, each scenario was treated as an independent evaluation. An entirely within-subjects design for the trust tasks would have required each ASHA to complete 16 scenarios, risking fatigue, carryover effects, and demand characteristics, as participants might infer the study's hypotheses from repeated exposure to all framings [19]. Our partial exposure approach, thus, balanced statistical power against these methodological threats. To check for potential carryover effects, we compared response patterns across task orders and found no systematic differences in preference or trust, suggesting that order effects were minimal.

Semi-Structured Interviews. Each session concluded with a 25-minute interview probing participants' reasoning—what shaped their preferences, how they decided to trust or override advice, the kinds of questions they would post to such systems, and how chatbot framings resonated with their everyday work. The interviews were conducted one-on-one by the first author in Hindi, audio-recorded with consent, and transcribed daily to minimize data loss.

Procedure. Fieldwork took place over two weeks in August 2024. Each session was conducted one-on-one by the first author at the sub-center nearest each ASHA, to minimize travel burden. A pilot study revealed that group settings silenced quieter participants and encouraged conformity biases; hence, we adopted one-on-one study format. Each session lasted ~55 minutes: 10 minutes for introduction, consent, and demographic data collection; 20 minutes for structured tasks; and 25 minutes for the interview. The researcher explained the purpose and tasks of the study in simple terms, obtained written consent, and collected demographic information (age, education and work experience). The instructions were read aloud to accommodate literacy differences. All study materials were in Hindi and were administered by a researcher fluent in the language as well as familiar with the local sociocultural context. Participants were not compensated for this study, but received financial support from our partner NGO, *Khushi Baby*, for a separate project deployment.

3.4 Sample Size and Power Calculation

We purposively recruited N=61 ASHAs. Eligibility required that ASHAs had not previously used any LLM-powered chatbots (e.g., ASHABot [90]), to avoid contamination of treatment effects. Each ASHA completed two head-to-head *preference* comparisons ($61 \times 2 = 122$ tasks) and four *trust* tasks (two low-ambiguity and two high-ambiguity, $61 \times 4 = 244$ tasks). Because multiple tasks came from the same individual, the effective sample size was smaller than the raw task counts. Adjusting for clustered data, this yielded 100–110 effective observations for preference and 140–190 for trust.[1] To complement our quantitative findings, interviews provided qualitative depth, surfacing why workers responded differently to framings and how they understood trust in practice.

3.5 Participants

61 ASHAs (all females) with an average age of 40.7±6.5 years (range: 29–57) participated in the study. On average, they had 14.9±4.7 years of experience as ASHAs. Educational attainment varied: 13.1% had studied up to middle school or less, 68.8% had completed middle- to high-school, and 18.1% held an undergraduate degree or higher.

3.6 Ethics and Positionality

The study was approved by the Institutional Review Board at *Microsoft Research India*. Written informed consent was obtained from all participants in Hindi, emphasizing voluntary participation and the right to withdraw without penalty. All data and transcripts were anonymized. Our positionality shaped both the design and interpretation of this study.

The researchers, having prior fieldwork experience with ASHAs, collaborated closely with a local NGO partner to ensure the chatbot framings resonated with lived practice rather than abstract theory. At the same time, our institutional affiliations—as researchers from global academic and corporate research institutions—may have influenced how ASHAs perceived the authority of the study. We mitigated these dynamics by foregrounding the role of NGO fieldworkers, who were trusted intermediaries, and by framing the study as exploratory rather than evaluative of ASHA performance. We also acknowledge that, although the NGO partnership helped bridge cultural and institutional distance, as external researchers, our outsider perspective may still have constrained how we understood ASHAs' narratives.

Another key ethical consideration in our design was how to introduce the necessary experimental manipulations in a responsible way. Studying the effects of different normative framings required varying who the chatbot-interface "spoke as" and presenting both correct and incorrect responses, a well-established methodology in HCI for studying trust [6, 19, 29]. To handle these design requirements ethically, we conducted a thorough debrief at the end of each session, informing participants that some chatbot-interface responses had been intentionally made inaccurate for research purposes. We also clarified that messages framed as coming from "other ASHAs" or "the Health Department" were simulated and not real. This ensured that participants did not leave with incorrect information or with mistaken beliefs about peer or institutional guidance.

3.7 Data Analysis

Our analysis combined structured task outcomes with qualitative interviews.

First, from the 122 preference tasks, we examined the distribution of choices across framings. To test differences, we used chi-square tests and a Bradley–Terry model, which estimates odds ratios for each chatbot-interface relative to the neutral baseline (*Info-Bot*) and significance for

[1]See Appendix A for details on design-effect adjustment and power simulations.

head-to-head matchups. We also explored whether preferences varied by ASHA characteristics (age, education, experience) using logistic regressions with interaction terms.

Next, from the 244 trust tasks (balanced across low-/high-ambiguity and correct/incorrect answers), we modeled whether ASHAs reported they would follow chatbot advice (Yes/No). To distinguish persuasiveness from discernment, we decomposed trust into four metrics:

- Overall Trust: probability of following regardless of correctness. Overall Trust = Pr(Follow | All Answers).
- Appropriate Trust: probability of following when advice was correct. Appropriate Trust = Pr(Follow | Correct Answers).
- Overreliance[2]: probability of following when advice was incorrect. Overreliance = Pr(Follow | Incorrect Answers).
- Calibrated Trust: difference between appropriate trust and overreliance. Calibrated Trust = Pr(Follow | Correct Answers) − Pr(Follow | Incorrect Answers).

We used logistic regressions to test (1) whether accuracy predicted trust (See Equation 2), (2) whether calibrated trust differed across chatbot-interfaces (See Equation 3), and (3) whether demographics moderated these effects (See Equation 4).

Our design measures *behavioral* trust, captured through reliance decisions in tasks where advice can be followed or overridden [110]. The dependent variable in every trust task is the binary action, "Will you follow this chatbot's advice? (Yes/No)," and all four metrics (Overall Trust, Appropriate Trust, Overreliance, Calibrated Trust) are computed from these decisions rather than self-reported ratings. We chose this approach for two reasons. First, in high-stakes health contexts, the critical outcome is whether advice is acted upon; behavioral measures capture trust as it manifests in action. Second, behavioral measures are less vulnerable to social-desirability and acquiescence biases documented in prior HCI work in collectivist contexts, including among ASHAs [31]. Concretely, we operationalize a "trust violation" as following incorrect advice (overreliance) and "calibrated trust" as following correct advice while declining incorrect advice. Table 3 shows correct and incorrect responses to the same high-ambiguity scenario across all four bots. Qualitative interviews complement these measures, but our core trust claims rest on the behavioral data.

For qualitative findings, interview transcripts were analyzed using Braun and Clarke's reflexive thematic analysis [17]. The first author conducted line-by-line inductive coding across all 61 transcripts, producing an initial codebook of 47 codes. Through iterative discussions, all authors consolidated these into five higher-order themes: affective responses to framings, decision-making rationales for trust and override, spontaneous cross-framing comparisons, references to professional practice and supervisor relationships, and contextual constraints (e.g., distance, family dynamics, network availability). Please refer to Appendix D for details on the qualitative coding process and our codebook.

4 Findings

We begin by examining which normative framings ASHAs preferred and draw on qualitative insights to explain why some framings resonated more than others (Section 4.1). We then analyze how these framings shaped trust and overreliance, first in low-ambiguity tasks where correct

[2]We use "overreliance" to denote the behavioral pattern of accepting incorrect AI advice—a safety-critical outcome in health contexts. This framing foregrounds calibration accuracy rather than the cognitive or social processes underlying trust decisions. We acknowledge that what manifests as overreliance may reflect situated trust transfer—the reasonable application of peer-trust heuristics to systems that mimic peer voice—rather than irrational deference. Section 5 reflects on this interpretive tension and alternative approaches that can be employed in future work.

Table 4. Chatbot comparison using Bradley-Terry model in the Preference Task.

Bot (vs. Baseline)	β	Odds Ratio	**95% CI** (β)	p
Info-Bot (baseline)	0.0	1.0×	–	–
ASHA-Bot	+0.7	2.1×	[0.05, 1.48]	.036
ASHA-Friend-Bot	+2.0	7.4×	[1.14, 2.88]	.001
Health-Dept-Bot	−0.7	0.4×	[−1.54, −0.03]	.043

Table 5. Pairwise head-to-head chi-square tests.

Pair (winners in bold)	Winner Share	p **(Holm)**
Info-Bot vs. *Health-Dept-Bot*	100.0%	**0.0001**
ASHA-Bot vs. *Info-Bot*	100.0%	**0.0000**
ASHA-Bot vs. *Health-Dept-Bot*	59.1%	0.3938
ASHA-Friend-Bot vs. *Info-Bot*	86.4%	**0.0019**
ASHA-Friend-Bot vs. *ASHA-Bot*	84.2%	**0.0057**
ASHA-Friend-Bot vs. *Health-Dept-Bot*	90.0%	**0.0014**

answers were clear (Section 4.2) and then in high-ambiguity tasks where protocols conflicted with on-the-ground realities (Section 4.3). Throughout, we use the neutral *Info-Bot* as the baseline for all comparisons.

4.1 Chatbot-Interface Preferences

4.1.1 Overall Preference Distribution. We begin by examining which chatbot-interface ASHAs preferred in the pairwise choice task. A chi-square goodness-of-fit test confirmed that the observed distribution of choices deviated significantly from a uniform 25.0% expectation ($\chi^2(3) = 32.4$, $p < .001$), indicating that participants consistently favored certain chatbot-interfaces over others.

We then looked at the six possible head-to-head comparisons among the four chatbot-interfaces. The *ASHA-Friend-Bot* was the overwhelming favorite, winning 53 of 61 comparisons (86.9%). At the other extreme, the *Health-Dept-Bot* was the least favored, winning only 11 comparisons (18.0%). The *ASHA-Bot* performed moderately well with 36 wins (59.0%), while the neutral *Info-Bot* won 22 comparisons (36.1%). Table 5 shows winner in each of these comparisons and the magnitude of each win.

In five of the six head-to-head comparisons, one interface was clearly preferred over the other, with the difference large enough to be statistically significant (Table 5). The only exception was the comparison between the *ASHA-Bot* and the *Health-Dept-Bot*. Here, preferences were more evenly split (59.0% vs. 41.0%), and the difference was not statistically significant. This suggests that while narrative framing generally outperformed the others, the authority framing still held some appeal when put directly against descriptive framing.

Although these pairwise tests tell us which interfaces were significantly preferred in direct comparisons, they did not tell us how strong those preferences were relative to a common benchmark. To address this, we estimated a Bradley-Terry model [16], using the neutral *Info-Bot* as the baseline. This model is a standard statistical approach widely used in HCI for analyzing pairwise comparisons [22, 66, 84]. It estimates a latent "strength" for each option such that the probability of one option being chosen over another follows a logistic function of their difference. In our case, it puts all interfaces on the same scale (against the baseline), so we can compare not just which one was preferred more, but how strongly each framing was preferred. Our analysis found that *ASHA-Friend-Bot* was 7.5× more likely to be chosen (Table 4), *ASHA-Bot* 2.1× more likely, while *Health-Dept-Bot* was only 0.46× as likely as compared to the baseline.

These results confirm the clear and systematic ordering of chatbot-interface preferences:

$$\textbf{\textit{ASHA-Friend-Bot}} > \textbf{\textit{ASHA-Bot}} > \textbf{\textit{Info-Bot}} > \textbf{\textit{Health-Dept-Bot}}$$

The overwhelming preference for *ASHA-Friend-Bot* was echoed in the qualitative themes, which provide rich insight into why ASHAs gravitated toward this normative framing. We unpack these themes in the section that follows.

4.1.2 Why Narrative Identity Framing Resonated with ASHAs. Participants consistently reported that the interface of *ASHA-Friend-Bot "felt like talking to another ASHA,"* fostering comfort and ease of interaction. In contrast, the formal, government-like tone of the interface of *Health-Dept-Bot* was often described as alienating. At first glance, this may seem counterintuitive: prior work shows that institutional endorsement can boost uptake, as frontline workers align with government authority [4, 5]. Yet other studies caution that bureaucratic framings can invoke fear, surveillance, and distance [9]). Our findings align with this latter: for everyday health queries, ASHAs strongly preferred the narrative identity framing. While institutional voice may confer legitimacy for official reporting contexts, narrative-identity created a sense of psychological safety for sensitive or "everyday" concerns, making the system feel approachable and trustworthy.

Our interviews revealed three themes explaining why the *ASHA-Friend-Bot* framing resonated with them:

Reliance on peer support. Many ASHAs reported that in practice they turn first to their close peers rather than supervisors for guidance, due to both comfort and logistical constraints in reaching their supervisors and nurses. Peer networks thus serve as the first line of consultation, offering immediate and judgment-free support. By reproducing this "peer first" channel, the interface of *ASHA-Friend-Bot* resonated with their existing help-seeking practices. Rather than positioning the chatbot-interface as an authority, it was interpreted as an extension of their peer network, which heightened both familiarity and trust. P14 explained:

> *"If I have a question, I usually call another ASHA first. The supervisor is at the PHC (Primary Health Centre), which is often a few kilometers away... We can always call and ask, but in my anganwadi (rural child care centre) I rarely get (cellular) network, so my choice is to either find another ASHA on my way to the next anganwadi or to go to the PHC."* – P14.

Comfort in asking questions. Participants also emphasized that the friendly tone of *ASHA-Friend-Bot*'s interface lowered barriers not only for routine queries but also for more sensitive or stigmatized topics. ASHAs described feeling free to raise concerns about reproductive health, menstrual irregularities, or family planning—areas where conversations with supervisors or formal channels often carried the risk of embarrassment or judgment. The interface's voice was interpreted as non-judgmental and approachable, creating space for candid dialogue in situations where silence is the safer option. P43 stated:

> *"I could ask openly about issues like white discharge, menstruation, or family planning. If I ask these things to a doctor or the official bot, I feel like I'm being judged. But this felt like talking to another ASHA who understands."* – P43.

This sense of psychological safety was critical: rather than worrying about being reprimanded or dismissed, ASHAs felt the framing provided a private, supportive channel where even "taboo" questions could be voiced without hesitation.

Learning through peer stories. Beyond comfort, ASHAs valued how the interface of *ASHA-Friend-Bot* normalized mistakes and modeled problem-solving through stories of other ASHAs. For example, the *ASHA-Friend-Bot* answered questions while referring to stories about other ASHAs in the village. In one of the scenarios, it said "Your fellow ASHA also struggled with this issue!" before advising about the next steps. This enabled the ASHA to identify with her peers and feel supported even when she made a mistake. P5 contrasted this with her supervisor's scolding response to an error:

> "*Once, while entering vaccine data, I accidentally typed 1000 vials instead of 100. I didn't know how to fix it, so I asked my supervisor and apologised for making the mistake. He scolded me in front of 15 other ASHAs, saying that if I couldn't do such a simple thing I*

> *should just leave the job... After that I stopped asking about my mistakes. But if I can ask these questions on my phone and if it talks to me the way it [ASHA-Friend-Bot] did... I would learn more quickly. It told stories of other ASHAs who had made similar mistakes. It made me feel like I could ask anything without being judged or shouted at.*" – P5.

By signaling empathy and shared narratives, the interface of *ASHA-Friend-Bot* created a safe space for learning that supervisors and authority-based channels often failed to provide. In contrast, the interface of *Health-Dept-Bot* was consistently described as "*too strict*," with its formal phrasing and directive tone discouraging casual or sensitive queries. Several ASHAs said it felt more like a training session than a conversation, which made them hesitant to use it for everyday doubts. Others emphasized that the authoritative style also carried a practical burden: responses were lengthy, bureaucratic, and read *"like a government circular,"* which felt mismatched to the rapid decision-making required in the field. As P23 explained:

> *"When you are in the field you don't want long, hard to read paragraphs. This [Health-Dept-Bot] was too formal and heavy. It felt like it was checking me instead of helping me."* – P23.

This difference underscores how normative framing shaped perceptions of safety: while narrative identity framing encouraged candid, stigma-free engagement, the authority framing (unintentionally) reinforced hierarchical distance.

4.1.3 Demographic Variation in Bot Preference. Next, we examined whether age, education, or experience moderated bot preferences. No subgroup effects reached statistical significance (Figures A1a and A1b in Appendix B). However, descriptive analysis suggested meaningful trends. ASHAs with ≥15 years of tenure were *more likely* to prefer peer-oriented framings (*ASHA-Friend-Bot* or *ASHA-Bot*): +16.2% vs. less-experienced peers, $p = .18$. These patterns aligned with interview accounts. Senior ASHAs reported longstanding skepticism toward top-down directives—having "*seen it all*" from the health department—and gravitated toward tools resembling a knowledgeable colleague over an official voice. P53 stated:

> "*I have been doing this job for 20 years. I know everyone in my village, they know me, and I know my training. But sometimes questions come up that are unique. During COVID, a Muslim woman was hesitant to get vaccinated because there were rumors that Hindu nurses were putting something in it causing infertility. I didn't know what to do. The answer was not in the training handbook. In these cases I need a fellow ASHA... someone who understands the dynamics here, to tackle such misinformation. This [ASHA-Friend-Bot] bot tells me what other ASHAs in my area are doing. It's like talking to another experienced person.* – P53.

In contrast, less-experienced or younger ASHAs showed somewhat greater receptivity to the *Health-Dept-Bot*, citing higher trust in formal authorities and less developed peer networks:

> *"I am new to this work, and my village is far from the subcenter so I only meet other ASHAs once a month. I don't know all the ASHAs in my area yet. This one [ASHA-Friend-Bot] said things like 'ASHA sister in your area does this,' but I don't know them personally. Until I build those connections, it's easier to see what the departmental guidelines say."* – P2.

Despite these nuances, the interface of *ASHA-Friend-Bot* consistently elicited stronger trust and willingness to engage across groups, highlighting the power of shared identity in shaping technology adoption. Our results show that social norms strongly influenced preferences. Yet preference is not equivalent to trust. ASHAs still had to decide whether to act on the information provided—a crucial distinction given the high-stakes and sensitive contexts they confront, e.g., pregnancy complications,

vaccine hesitancy, childhood illness. We therefore next examine a central question: when ASHAs consulted the chatbot-interfaces, did they *trust* the advice—and, crucially, could they *tell apart* correct from incorrect guidance? To assess this, we analyzed responses in the low-ambiguity and high-ambiguity tasks, where trust was coded as whether the ASHA indicated she would follow the chatbot's advice (*Yes/No*).

4.2 Trust in Chatbots in Low-Ambiguity Scenarios

4.2.1 Overall Trust Distribution. We first examined whether ASHAs could distinguish between correct and incorrect advice in low-ambiguity scenarios. We found that they were significantly more likely to trust correct responses than incorrect ones ($\chi^2(1) = 10.1$, $p < 0.01$), following 84.7% of correct answers but still 58.7% of incorrect ones. A logistic regression (Model 2) confirmed that correct advice was nearly four times more likely to be trusted than incorrect advice (odds ratio = 3.90, 95% CI [1.68, 9.06]). Despite this discernment, overreliance was high—more than half of the incorrect responses were still accepted—resulting in a low Calibrated Trust of 26.0% (Table 6). These findings suggest that even in clear-cut cases, ASHAs remained vulnerable to errors, a concerning trend in high-stakes health settings.

We next analyzed whether trust patterns varied across normative framings and found significant contrasts (Model 2; Figure 3). *ASHA-Friend-Bot* significantly induced Overreliance compared to the neutral baseline ($\gamma = 1.97$, SE=0.88, $p = 0.025$), with workers trusting 90% of its incorrect advice (Overreliance), resulting in the lowest Calibrated Trust at -15.0% ($\delta = -2.38$, $p = 0.095$). In effect, correct and incorrect answers were accepted at nearly the same rate. By contrast, *Health-Dept-Bot* significantly reduced Overreliance ($\gamma = -2.53$, SE=1.16, $p = 0.030$), with 90.5% of its correct advice being trusted, while only 9.1% of its incorrect answers were followed. This produced the strongest Calibrated Trust of 81.4% ($\delta = 3.27$, $p = 0.040$), approaching the ideal pattern of accepting correct advice but rejecting incorrect advice. *ASHA-Bot* was very similar to the baseline, with an Overall Trust of 72.4% and Overreliance of 57.1%

Together, these findings highlight a trade-off: narrative identity framings (*ASHA-Friend-Bot*) built affinity but encouraged uncritical acceptance, whereas authoritative framings (*Health-Dept-Bot*), though less preferred, acted as safeguards against error.

For completeness, we also examined the inverse error—underreliance, i.e., rejecting correct advice. Across framings, ASHAs rejected correct advice most often when it came from ASHA-Friend-Bot (25.0%) and least often from Health-Dept-Bot (9.5%), with Info-Bot (18.2%) and ASHA-Bot (13.3%) in between. These differences were not statistically significant ($\chi^2(3) = 1.53$, $p = .68$), and are treated as descriptive. We discuss the rationale for focusing on overreliance in Section 5.

4.2.2 Reasons for Overreliance in Narrative Identity Bot and Calibrated Trust in Authority Bot. Analysis of our qualitative data revealed two distinct themes that explain the contrasting trust patterns we observed.

Peer-like reassurance encouraged fast acceptance. ASHAs described the interface of *ASHA-Friend-Bot* as offering confirmation rather than new information. In low-ambiguity questions (e.g., vaccination schedule), where ASHAs already knew the answer, its friendly tone acted like a nod of agreement from a colleague, accelerating compliance without reflection. As P14 put it:

> *"This was something I already do every day. When it [ASHA-Friend-Bot] confirmed it in a friendly way, I didn't stop to think. It was just a quick yes."* – P14.

Here, affinity and familiarity blurred the boundary between peer talk and factual verification, reducing critical scrutiny.

Authoritative tone disrupted routine and prompted reflection. By contrast, the interface of *Health-Dept-Bot* increased Calibrated Trust by slowing ASHAs down. Its formal, bureaucratic

Table 6. Trust metrics across chatbots in low- and high-ambiguity tasks.

Chatbot	Low-ambiguity task				High-ambiguity task			
	Overall Trust	Appropriate Trust	Over-reliance	Calibrated Trust	Overall Trust	Appropriate Trust	Over-reliance	Calibrated Trust
Info-Bot	65.5%	81.8%	55.6%	26.3%	69.0%	85.7%	53.3%	32.4%
ASHA-Bot	72.4%	86.7%	57.1%	29.5%	75.9%	71.4%	80.0%	−8.6%
ASHA-Friend-Bot	84.4%	75.0%	90.0%	−15.0%	96.9%	93.8%	100.0%	−6.2%
Health-Dept-Bot	62.5%	90.5%	9.1%	81.4%	53.1%	88.9%	7.1%	81.8%

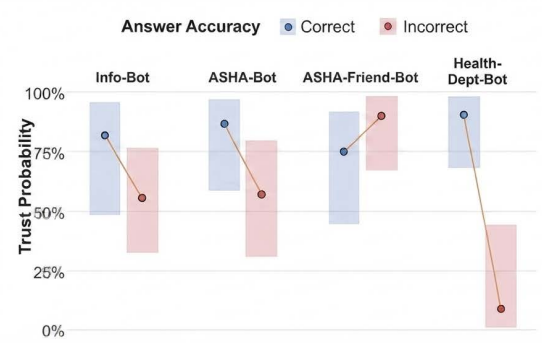


Fig. 3. Trust in low-ambiguity scenarios by chatbot

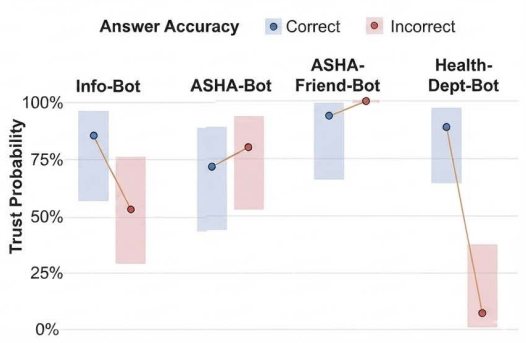


Fig. 4. Trust in high-ambiguity scenarios by chatbot

style stood out against other conversational norms, even for simple day-to-day questions. Participants reported that this unfamiliarity made them pause and compare the advice against their own knowledge. P27 stated:

> *"I knew the right answer already, but when it came in that official style I slowed down and compared. That's how I saw the mistake."* – P27.

Rather than reinforcing existing beliefs, the authoritative framing interrupted routine acceptance, creating space for reflection and error detection.

4.2.3 Demographic Variation in Trust by Bot. Finally, we examined whether demographics shaped these outcomes (Model 4). Adding age, education, and experience did not alter the main results, and none were statistically significant. However, descriptive patterns suggest that more experienced and educated ASHAs were less likely to follow incorrect advice (see Figure A2b in Appendix). As one veteran noted: "*There is nothing in these tasks that I haven't seen first-hand... If a bot—or even an ANM—gave the wrong instruction here, I would catch it and not follow it*" (P21). These reflections help interpret the negative (though non-significant) coefficients, they represent embodied familiarity that can counteract overreliance in routine matters.

4.3 Trust in Chatbots in High-Ambiguity Scenarios

We now turn to high-ambiguity scenarios, where the correct answer is unclear or contested, and the central challenge is navigating uncertainty. While low-ambiguity tasks revealed whether ASHAs could distinguish correct from incorrect advice in routine matters, high-ambiguity tasks examined whether they could still exercise such discernment when certainty itself was in question.

4.3.1 Overall Trust Distribution. Logistic regression (Model 2) showed that ASHAs were nearly five times more likely to follow correct than incorrect advice ($\beta = 1.60$, SE=0.52, $p = .002$; odds

ratio = 4.94, 95% CI [1.8, 13.4]). The gap between following correct and incorrect advice narrowed compared to low-ambiguity tasks: Overall Trust remained high at 73.8%, but pooled Calibrated Trust dropped from 26.0% to 23.8% (Table 6). Overreliance also increased, with ASHAs accepting 61.7% of incorrect answers. Together, these results show that high-ambiguity scenarios weakened ASHAs' ability to tell correct advice from incorrect, making them more likely to follow wrong recommendations.

These effects varied significantly across normative framing (Model 3; Figure 4). *ASHA-Friend-Bot* significantly increased Overreliance compared to the neutral baseline ($\gamma = 1.97$, SE=0.88, $p = 0.025$), with near-universal compliance: 96.9% of its correct answers and 100% of its incorrect ones were followed, a jump from 90% in low-ambiguity scenarios. This led to minimal Calibrated Trust of -6.2% ($\delta = -2.38$, $p = 0.095$), which is better than what we found in low ambiguity scenarios (of -15.0 %). However, the fact that 100% of the incorrect advice was also trusted indicates overreliance over any advice given by this bot's interface, and not better discernment. In other words, participants appeared to follow this bot's advice uncritically, regardless of accuracy.

By contrast, the interface of *Health-Dept-Bot* continued to act as a safeguard. It significantly reduced Overreliance ($\gamma = -2.53$, SE=1.16, $p = 0.030$), with only 7.1% of its wrong answers trusted (down from 9.1% in low-ambiguity), while 88.9% of its correct answers were followed. This produced the strongest Calibrated Trust of 81.7% ($\delta = 3.27$, $p = 0.040$). Here, the Calibrated Trust increased slightly from what we observed in low-ambiguity scenarios. These results show that authority framings maintained their protective effect even under uncertainty, helping ASHAs resist following incorrect advice. Finally, the interface of *ASHA-Bot* did not differ significantly from baseline, but its calibration worsened under ambiguity ($p = 0.045$), suggesting that what felt like reassuring advice from other ASHAs in clear cases became a liability when answers were uncertain.

For underreliance, the pattern reversed and was not statistically significant ($\chi^2(3) = 3.28$, $p = .35$). Under high ambiguity, ASHAs rejected correct advice most often from ASHA-Bot (28.6%) and least often from ASHA-Friend-Bot (6.2%), with Info-Bot (14.3%) and Health-Dept-Bot (11.1%) in between.

4.3.2 Why High Ambiguity Impacted Calibrated Trust. Our qualitative interviews surfaced two interlocking dynamics that explain why, under high-ambiguity, overreliance reached 100% for the interface of *ASHA-Friend-Bot* and increased even for the interface of *Health-Dept-Bot* compared to low-ambiguity scenarios.

Social proof substituted for missing knowledge. In high-ambiguity scenarios, ASHAs admitted they were unsure of the *"correct"* practice, hence the peer narrative story itself became treated as evidence. Advice presented as what *"another ASHA sister"* had done was inferred to be valid as it reflected lived village realities. P2 explained:

> *"I wasn't sure what to do, but if it says another ASHA here does it, then I believe it must be right. She knows what happens in our kind of village, so I trusted it."* – P2

Here, narrative identity from a peer filled the gap left by unclear or evolving protocols, effectively substituting social proof for technical knowledge.

Comfort muted doubt. Ambiguity usually provokes hesitation, but the warm and familiar tone of the narrative bot made advice feel safe to accept without scrutiny. P33 reflected:

> *"Actually I wasn't sure if this was right. The bot told me that if the MR vaccine is missed at nine months, it should not be given later. I had never heard of that before, so I was confused. But then it also said that another ASHA from my PHC had advised the same, so I didn't think twice before agreeing. I thought maybe it was some new update I hadn't been trained on."* – P33

Here, comfort amplified by uncertainty muted critical attention, especially among less-experienced ASHAs. Peer endorsement was interpreted as a possible signal of updated guidance, making incorrect responses more likely to slip through, further reducing the impulse to double-check. By contrast, participants often described the *Health-Dept-Bot*'s distance—both in language and tone—as prompting a different kind of reasoning. Its formality reactivated memories of official training, guidelines, and "best practices," even when these clashed with field realities. P60 shared:

> *"The way the chatbot spoke, it reminded me of my training days... But when you work in the field you realize not everything that you are taught [in training] is possible. E.g., when it comes to institutional delivery, the chatbot advised to push for it, even if the family is not interested [in institutional delivery]. But sometimes you can't push, because the family is very aggressive about their choice and don't want my involvement... But it [bot] spoke so formally and like a government person... it said it is my duty as an ASHA... it is my responsibility... it reminded me that it is a best practice to try even if we expect that the family will not change their mind."* – P60

These findings show that normative framing effects shifted with varying task ambiguity. In low-ambiguity scenarios, narrative framing sped up acceptance and felt like friendly confirmation, while authority framing slowed ASHAs down and prompted careful checking. In contrast, in high-ambiguity scenarios, narrative framing became a substitute for missing knowledge, encouraging overreliance.

4.3.3 Demographic Variation in Trust. Finally, we examined whether demographics moderated these effects (Model 4). While no coefficients were statistically significant, descriptive trends were consistent. Less-experienced ASHAs were the most trusting of *ASHA-Friend-Bot*, following 97.4% of its advice. By contrast, more experienced workers were about 15% less trusting overall, reflecting greater caution toward narrative identity framings. This echoes qualitative accounts that experience not only deepened familiarity with protocols but also cultivated a healthy skepticism toward both digital tools and top-down instructions.

5 Discussion

Our study shows that normative framings in health chatbots deeply impact both *preference* and *trust*. Here, we discuss three implications of our findings on HCI design and research.

5.1 Culture as a Double-Edged Lever

Calls to "design with culture" [8, 67, 102, 105] in HCI often assume that cultural alignment is uniformly beneficial. Our results complicate that claim by focusing on one specific cultural lever—*social norms*. We find that social norms are powerful because they reshape judgments about legitimacy and appropriateness, not just aesthetics. Our results align with prior work showing that narrative identity markers in conversational agents (e.g., anthropomorphic language, human-like representation, communicative agency) shift perceived credibility, warmth, and willingness to comply [2, 83, 98]. On the other hand, the literature on injunctive norms further shows that authoritative sources can heighten perceived obligation [46, 47]. Where we extend the literature is in demonstrating a reliability trade-off: the same cues that elevate *preference* can degrade *calibrated trust*. Narrative identity framings are highly engaging yet prone to overreliance, whereas injunctive authority framings dampen appeal but improve error discernment. Our evidence thus cautions against blind maximization of cultural cues (in our case, of social norms) and the universal directive to embed such cues for adaptability; instead, it supports treating cultural cues as a tunable design lever with measurable preference and safety trade-offs.

These findings also help explain why prior broadcast-based norm interventions have yielded largely null effects [81]. In SMS or poster campaigns, normative messages arrive without relational context—they cannot embody a peer identity, respond to individual concerns, or build rapport over conversational turns. Our results suggest that normative cues may require relational grounding to influence behavior: the *ASHA-Friend-Bot's* peer framing was powerful precisely because it felt like a conversation with a colleague, not a broadcast announcement. This implies that conversational AI can unlock normative influence in ways that broadcast media cannot, but that this heightened influence carries corresponding risks of overreliance when the system provides incorrect guidance.

For chatbot design, this implies that instead of hard-coding a specific normative framing, the bot should flexibly adapt its responses based on the severity and sensitivity of the user's query, the bot's confidence in its answer, the trajectory of the conversation, and the experience level of the user. In practice, one way to achieve it is by tuning the normative framing lever using observable signals from the system, such as: *interaction mode*—is the user exploring the bot (many short, unrelated, open-ended questions) or moving towards committing to an action (many focused questions about the same issue)?; and *who is asking*—a new vs. experienced user. In exploratory interactions or with newer users, use narrative identity framings to encourage curiosity and support learning. When the exchange shifts toward commitment (e.g., questions regarding supplies, doses, or issuing referrals), the bot should switch to an authoritative voice and show the relevant guideline citation upfront.

5.2 Ambiguity as a Stress Test

We found that normative framing mattered most when "the right advice" collided with what was actually doable on the ground. In routine work, a narrative identity voice felt natural and was readily accepted. In contested situations, the same voice often sped people toward agreement even when they felt unsure. An injunctive authority voice showed the mirror image: it could feel rigid and distant in low-ambiguity tasks, but under uncertainty, these characteristics acted like a guardrail by pulling training and guidelines back into view and prompting ASHAs to return to "what should be done". In short, ambiguity revealed when cultural cues steady judgment and when they destabilize it.

Our findings on ASHAs overrelying on chatbots are consistent with HCI literature on overreliance in AI-powered decision support tools [19], as well as prior work showing community health workers show overreliance on AI outputs, attributing inherent accuracy and authority to "the machine" [78, 79]. Where we extend the literature is in demarcating the areas where this overreliance takes full force—in spaces of high ambiguity. Prior research shows that ambiguity changes how people weigh system advice and increases reliance on perceived expertise [10, 34, 113]. Why, then, did ASHAs overrely so heavily on the *ASHA-Friend-Bot*—a source not framed as authoritative—when things were unclear? Our findings uncover a salient type of ambiguity in healthcare contexts (particularly prevalent in low- and middle-income countries): ambiguity of feasibility. In our high-ambiguity scenarios, there was a protocol-driven correct answer that ASHAs were trained on, but it conflicted with on-the-ground constraints (e.g., family pressures, travel distance, resource unavailability, local contextual considerations). In these situations, their close peers functioned as *situated experts*: they knew which workarounds were locally feasible, which steps families would accept, and which referrals would actually go through. When protocol and practice were misaligned, the narrative identity framing gained legitimacy because it better matched what could realistically be done. Our findings thus extend the literature by formulating a distinction between *feasibility-driven* and *informational* ambiguity—showing that who is considered an authority can shift from formal institutions to experienced peers depending on the kind of ambiguous situation.

This raises a critical question: does this willingness to follow the peer-framed chatbot reflect genuine irrationality, or a rational transfer of trust? In their day-to-day practice, trusting peer-endorsed advice is a learned heuristic that typically serves ASHAs well: peer knowledge often captures on-the-ground feasibility and social constraints more accurately than distant protocols. This is why, we argue that while the cognitive mechanism is *situated trust*—a functional heuristic favoring peer-endorsed knowledge—the behavioral outcome in a safety-critical AI context is overreliance. By "overreliance", we do not imply a lack of competence; rather, we describe a pattern where ASHAs defer to system outputs without verification or sensemaking at the point of decision. This distinguishes their behavior from "calibrated trust," where users actively negotiate or condition their reliance. The danger of conversational AI lies in its ability to exploit this situated trust by mimicking peer markers without inheriting the interpersonal accountability that regulates real-world networks. The implication for HCI is thus not to "debias" frontline workers, but to recognize that when situated trust is transferred to an unaccountable system, it becomes dangerous overreliance. Design must therefore focus on ensuring systems do not mimic peer legitimacy without accompanying safeguards.

These insights point toward concrete design directions. From a design perspective, ambiguity should be treated not only as missing information but also as limited feasibility. When cues of infeasibility arise (e.g., repeated mentions of distance, or family refusal), health chatbots should acknowledge these constraints explicitly and pair protocol guidance with narrative-informed, locally workable next steps (e.g., follow-ups with families or acceptable substitutes). Thus, designing *dynamic norm systems* that adjust style–and even combine normative framings based on context–is a promising direction. Hybrid framings—messages that combine normative identity with explicit reference to guidelines—may balance user resonance with accountability. For example, a chatbot-interface might say, *'Radhika, an ASHA in your area, continued counseling for institutional delivery when families hesitated—this is also what Health Department guidelines say.'* In this, the narrative identity norm precedes the injunctive authority in the sentence. The positioning of the norm could also change based on user profile. Empirical evaluation of such adaptive and hybrid approaches, ideally in real-world deployments over extended periods, will be critical for assessing whether norm-based chatbots can sustain calibrated trust at scale.

5.3 From Trust to Calibrated Trust

A common goal in HCI is to "build trust" in AI, especially in health contexts [3, 14, 91]. Our findings suggest this framing is insufficient. What matters is *calibrated trust*: workers should follow advice when it is correct and resist or override it when it is incorrect. Prior work has made a similar point—trust should align with actual performance and context—and also shown that surface-level trust boosts (e.g., human-like cues or smooth UX) can increase reliance even when the system is wrong [6, 52, 59]. Showing case-by-case reliability signals (like model confidence) can help, but these alone are not enough; calibrated trust must be *designed*, *measured*, and *optimized*, rather than inferred indirectly from satisfaction or preference scores [6, 113].

Our findings add a simple but important twist: *social norms* shape what feels trustworthy. Normative framings can make people say "yes, I trust" even when the advice is wrong (as with overreliance on *ASHA-Friend-Bot*), or make them hesitate even when the advice is right. To support calibrated trust, reliability signals should be paired with dynamic norm cues. For instance, a chatbot might use assuring language in high certainty cases such as tablet doses and vaccination schedules ('you should,' 'you must,' 'you need to') but adopt more cautious phrasing in contested scenarios ('best practice is usually…,' 'it is recommended…'). This linguistic modulation should work in tandem with dynamic switching of normative framings based on context and user profile. Prior evidence [19, 63, 113] shows that prompts like "think twice" and "are you sure?" reduce overreliance

by creating a moment of distance which promotes discernment. Our findings show that injunctive authority works in a similar way, as formal framing forces ASHAs to pause and rethink. Thus, pairing such framings in chatbot communication can promote calibrated trust.

Designing for calibrated trust then also requires acknowledging that these systems will occasionally fail. Instead of hiding that fallibility, chatbots should surface uncertainty and guide workers toward safe next steps. This can be done by: (a) explicitly showing low confidence ("I'm not certain about this"), (b) briefly explaining the basis for its advice ("This answer draws on Health Department guidelines", "This reflects common practices reported by ASHAs in your area"), and (c) enabling workers to easily report or correct inaccurate responses ("Report this answer as incorrect"). Making uncertainty visible and actionable not only mitigates potential harm but also fosters more sustained engagement and trust, encouraging ASHAs to rely on the chatbot for routine guidance while being directed to other trusted sources of knowledge, such as supervisors, training manuals, or peer networks, when the system's confidence is low.

Methodologically, this pushes us to rethink evaluation of such systems. AI-powered health chatbots should report *calibration* directly—alongside trust, usability, or preference. The strength of these technologies should hence be judged by whether they *improve calibration* and downstream outcomes, not merely whether they boost trust. Through this lens, our work reframes "cultural alignment" in healthcare: the aim is not to maximize resonance, but to *calibrate* it—dialing norm cues up or down with model confidence, task criticality, and user state so that reliance stays in check [6, 59, 113].

5.4 The Asymmetry Between Over- and Underreliance

A natural question is whether the same theoretical machinery we used for overreliance also applies to its counterpart, underreliance. While the descriptive patterns for underreliance reported in Section 4 are suggestive, they are not statistically significant. More importantly, our paper focuses on overreliance for two reasons.

First, the two errors reflect different constructs. Overreliance concerns failures of the AI system: cases in which the system produces an incorrect response and the user fails to detect it [57, 82]. Underreliance, in contrast, reflects the user's relationship to her own expertise, where she trusts her training and her peers over the system output. The norm theories we draw on (descriptive, narrative identity, injunctive) explain compliance with social cues, making them more directly applicable to overreliance. Why people resist to AI systems is better addressed through adjacent work on reactance [18, 101], algorithm aversion [36, 61], and professional identity protection [11, 23].

Second, the two errors carry unequal consequences. If an ASHA rejects a correct answer, she falls back to the existing standard of care–her training, supervisors, and peers; the system has at worst failed to add value. If she acts on an incorrect answer, she may give the wrong dose or push a family in the wrong direction, and that error compounds into patient care. In such settings, being conservative is usually safe and being confidently wrong is not; thus, overreliance is the more consequential failure mode and the one worth designing against first.

5.5 Limitations and Future Work

Our study has limitations that suggest concrete directions for future research. First, our sample was restricted to 61 ASHAs in rural Rajasthan, and their cultural and professional context likely shaped how framings were perceived. Replication with more diverse groups—such as urban health workers or users in more individualist contexts—would clarify how cultural setting moderates normative effects.

Second, we used single-turn static chatbot mocks rather than a live, interactive system. While this design strengthens our study by isolating framing effects—consistent with prior HCI work

using static mockups for chatbot evaluation [29, 73, 111]—it abstracts away the dynamics of real conversations. Multi-turn long-term interactions may amplify normative effects through increased engagement and social presence, or attenuate them as users recalibrate trust based on repeated experience. Future field deployments are therefore necessary to understand how these effects evolve in practice. We hypothesize three directions in which these effects may evolve in live deployments: (i) narrative framings may strengthen through sustained rapport or weaken with familiarity as the "peer voice" begins to feel scripted; (ii) authority framings may routinize and lose their pause-and-reflect effect, or accrue legitimacy as institutional grounding is repeatedly reinforced; and (iii) repeated exposure to errors may drive trust recalibration, especially if the system surfaces uncertainty or supports user-led error reporting.

Third, our scenario-based design captured only short-term interactions. While useful for isolating framing effects, it cannot reveal how trust develops across repeated use. Longitudinal into-the-wild deployments are therefore a key avenue. Future work should examine whether the appeal of narrative framings wanes with familiarity, whether repeated exposure to errors prompts recalibration, and how authority framings play out in everyday practice. Larger and more heterogeneous samples would also help uncover demographic trends (e.g., whether younger or less experienced workers are more vulnerable to overreliance).

Fourth, our study focused specifically on health contexts. It remains an open question whether the same normative framing effects extend to other high-stakes domains such as education, law, or agriculture, where social expectations and authority cues may differ. Exploring these cross-domain dynamics would clarify the generality of our findings and help guide the design of culturally adaptive chatbots more broadly.

Lastly, our normative framings were derived from established social psychology categories widely used in health communication and behavioral design research. These framings were applied top-down rather than co-developed with ASHAs. While this ensures theoretical grounding and comparability with prior work, it risks imposing external conceptual structures onto local ways of reasoning about trust and advice. ASHAs may draw on framings that do not align neatly with academic taxonomies, for instance, distinctions based on caste, kinship, or institutional histories that shape whose voice carries weight. Future work should use qualitative methods [28] to examine how ASHAs describe and interpret social influence, potentially validating, extending, or challenging the categories we employed.

6 Conclusion

This paper investigates how social norm framings in health chatbots shape frontline workers' *preference* and *trust*. We designed four otherwise-identical chatbots that differed only in social norms (descriptive, narrative identity, injunctive authority, and neutral) and evaluated them with ASHAs in rural India. Our findings reveal a tradeoff: narrative identity norm maximizes comfort and preference but increases overreliance, while injunctive authority is less preferred yet promotes calibrated trust. These effects intensify in scenarios where formal protocol guidance collides with on-the-ground feasibility. We translate these insights into design recommendations for AI-powered health chatbots: adopt *dynamic norm framings* that switch by scenario; incorporate *feasibility-aware advice* when protocol and on-ground realities diverge, and use *hybrid norms* that combine narrative identity with injunctive authority to balance comfort with safety. Finally, we argue that evaluation practices should go beyond trust, usability, or satisfaction alone. Reporting *calibrated trust*—follow when correct and resist when incorrect—should become a standard for safety in health chatbots. More broadly, our work reframes "cultural alignment" in HCI: cultural cues should not simply be maximized, but calibrated, since their uncritical amplification can harm as much as help in risk-sensitive domains like healthcare.

Acknowledgments

We thank the ASHA workers who generously shared their time, insights, and experiences, without whom this research would not have been possible. We are deeply grateful to our partner NGO, *Khushi Baby*, for their critical role in planning and coordinating government permissions for fieldwork. Their support was instrumental in ensuring smooth implementation of the study.

References

[1] S. Agarwal, H. Perry, L. Long, and A. Labrique. 2015. Mobile technology in support of frontline health workers: A systematic review of the literature. *Global Health: Science and Practice* 3, 3 (2015), 311–327.

[2] Theo Araujo. 2018. Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. In *Computers in Human Behavior*, Vol. 85. Elsevier, 183–189. doi:10.1016/j.chb.2018.03.051

[3] Onur Asan, Alparslan Emrah Bayrak, and Avishek Choudhury. 2020. Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians. *Journal of Medical Internet Research* 22, 6 (2020), e15154. doi:10.2196/15154

[4] Abhijit V. Banerjee, Eliana La Ferrara, and Victor H. Orozco-Olvera. 2019. *The Entertaining Way to Behavioral Change: Fighting HIV with MTV*. NBER Working Paper 26096. National Bureau of Economic Research. doi:10.3386/w26096

[5] Shweta Naik Bankar and Rangoli Gupta. 2022. Social Norms Programming in South Asia: Landscape Analysis. https://csbc.org.in/reports/SASNLC_Landscape%20Analysis_January2022.pdf

[6] Gagan Bansal, Tongshuang Wu, Joyce Zhou, Raymond Fok, Besmira Nushi, Ece Kamar, Marco Tulio Ribeiro, and Daniel S. Weld. 2021. Does the Whole Exceed Its Parts? The Effect of AI Explanations on Complementary Team Performance. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. ACM, New York, NY. doi:10.1145/3411764.3445717

[7] Abdullah H. Baqui, Emma K. Williams, Amanda M. Rosecrans, Praween K. Agrawal, Saifuddin Ahmed, Gary L. Darmstadt, Vishwajeet Kumar, Usha Kiran, Dharmendra Panwar, Ramesh C. Ahuja, Vinod K. Srivastava, Robert E. Black, and Manthuram Santosham. 2008. Impact of an integrated nutrition and health programme on neonatal mortality in rural northern India. *Bulletin of the World Health Organization* 86, 10 (2008), 796–804. doi:10.2471/BLT.07.042226

[8] Wendy L. Barber and Albert N. Badre. 1998. Culturability: The Merging of Culture and Usability. In *Proceedings of the 4th Conference on Human Factors and the Web*. Basking Ridge, NJ. https://zing.ncsl.nist.gov/hfweb/att4/proceedings/barber/

[9] Sebastian Bärnreuther. 2024. Surveillance medicine 2.0: digital monitoring of community health workers in India. *Global Public Health* (2024). doi:10.1080/13648470.2024.2378733

[10] Robert S. Baron, Joseph A. Vandello, and Bethany Brunsman. 1996. The Forgotten Variable in Conformity Research: Impact of Task Importance on Social Influence. *Journal of Personality and Social Psychology* 71, 5 (1996), 915–927. doi:10.1037/0022-3514.71.5.915

[11] Matthew Beane. 2019. Shadow Learning: Building Robotic Surgical Skill When Approved Means Fail. *Administrative Science Quarterly* 64, 1 (2019), 87–123. doi:10.1177/0001839217751692

[12] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT)* (2021), 610–623.

[13] Cristina Bicchieri. 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press, Cambridge, UK. doi:10.1017/CBO9780511616037

[14] Timothy W. Bickmore and Rosalind W. Picard. 2005. Establishing and maintaining long-term human-computer relationships. *ACM Trans. Comput.-Hum. Interact.* 12, 2 (June 2005), 293–327. doi:10.1145/1067860.1067867

[15] Abeba Birhane. 2021. Algorithmic injustice: a relational ethics approach. *Patterns* 2, 2 (2021), 100205.

[16] Ralph Allan Bradley and Milton E. Terry. 1952. Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika* 39, 3/4 (1952), 324–345. doi:10.2307/2334029

[17] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. arXiv:https://www.tandfonline.com/doi/pdf/10.1191/1478088706qp063oa doi:10.1191/1478088706qp063oa

[18] Jack W. Brehm. 1966. *A Theory of Psychological Reactance*. Academic Press.

[19] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z. Gajos. 2021. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-Assisted Decision-Making. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1, Article 188 (2021), 21 pages. doi:10.1145/3449287

[20] Federico Cabitza, Raffaele Rasoini, and Gian Franco Gensini. 2017. Unintended consequences of machine learning in medicine. *JAMA* 318, 6 (2017), 517–518.

[21] Lorainne Tudor Car, Dhakshenya Ardhithy Dhinagaran, Bhone Myint Kyaw, Tobias Kowatsch, Shafiq Joty, Yin-Leng Theng, and Rifat Atun. 2020. Conversational Agents in Health Care: Scoping Review and Conceptual Analysis. *Journal of Medical Internet Research* 22, 8 (2020), e17158. doi:10.2196/17158

[22] Wei-Lin Chiang, Lianmin Zheng, Ying Sheng, Anastasios Nikolas Angelopoulos, Tianle Li, Dacheng Li, Hao Zhang, Banghua Zhu, Michael Jordan, Joseph E. Gonzalez, and Ion Stoica. 2024. Chatbot Arena: An Open Platform for Evaluating LLMs by Human Preference. arXiv:2403.04132 [cs.AI] https://arxiv.org/abs/2403.04132

[23] Angèle Christin. 2017. Algorithms in practice: Comparing web journalism and criminal justice. *Big Data & Society* 4, 2 (2017), 2053951717718855. arXiv:https://doi.org/10.1177/2053951717718855 doi:10.1177/2053951717718855

[24] Adrian Chung and Rajiv N. Rimal. 2023. Harnessing social norms for health promotion: A systematic review. *Health Communication* 38, 7 (2023), 882–896.

[25] Karen Church and Rodrigo de Oliveira. 2013. What's up with whatsapp? comparing mobile instant messaging behaviors with traditional SMS. In *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Munich, Germany) *(MobileHCI '13)*. Association for Computing Machinery, New York, NY, USA, 352–361. doi:10.1145/2493190.2493225

[26] Robert B. Cialdini. 1993. *Influence: Science and Practice* (3 ed.). HarperCollinsCollegePublishers, New York, NY.

[27] Robert B. Cialdini. 2003. Crafting Normative Messages to Protect the Environment. *Current Directions in Psychological Science* 12, 4 (2003), 105–109. arXiv:https://doi.org/10.1111/1467-8721.01242 doi:10.1111/1467-8721.01242

[28] Tom Cole and Marco Gillies. 2022. More than a bit of coding: (un-)Grounded (non-)Theory in HCI. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 11, 11 pages. doi:10.1145/3491101.3516392

[29] Samuel Rhys Cox and Wei Tsang Ooi. 2022. Does Chatbot Language Formality Affect Users' Self-Disclosure?. In *Proceedings of the 4th Conference on Conversational User Interfaces* (Glasgow, United Kingdom) *(CUI '22)*. Association for Computing Machinery, New York, NY, USA, Article 1, 13 pages. doi:10.1145/3543829.3543831

[30] Nicola Dell, Ian Francis, Haynes Sheppard, Raiva Simbi, and Gaetano Borriello. 2014. Field evaluation of a camera-based mobile health system in low-resource settings. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services* (Toronto, ON, Canada) *(MobileHCI '14)*. Association for Computing Machinery, New York, NY, USA, 33–42. doi:10.1145/2628363.2628366

[31] Nicola Dell, Neha Kumar, and Kentaro Toyama. 2012. "Yours is better!" participant response bias in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1321–1330.

[32] Brian DeRenzi and Tapan S. Parikh. 2017. Designing for the unpredictable: mobile technology use by community health workers in rural India. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 2708–2719.

[33] Brian DeRenzi, Jeremy Wacksman, Nicola Dell, Scott Lee, Neal Lesh, Gaetano Borriello, and Andrew Ellner. 2016. Closing the Feedback Loop: A 12-month Evaluation of ASTA, a Self-Tracking Application for ASHAs. In *Proceedings of the Eighth International Conference on Information and Communication Technologies and Development (ICTD '16)*. Association for Computing Machinery, New York, NY, USA, 1–10. doi:10.1145/2909609.2909652

[34] Morton Deutsch and Harold B. Gerard. 1955. A Study of Normative and Informational Social Influences upon Individual Judgment. *Journal of Abnormal and Social Psychology* 51, 3 (1955), 629–636. doi:10.1037/h0046408

[35] Baldeep K. Dhaliwal, Madhu Gupta, Anuradha Nadda, Shalini Singh, Anita Shet, Kerry Scott, Aakanksha Dutta, and Svea Closser. 2025. Urban community health workers in Punjab, India: A qualitative study of ASHAs' roles in the health system. *PLOS Global Public Health* 5, 8 (2025), e0004629. doi:10.1371/journal.pgph.0004629

[36] Berkeley J. Dietvorst, Joseph P. Simmons, and Cade Massey. 2015. Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err. *Journal of Experimental Psychology: General* 144, 1 (2015), 114–126. doi:10.1037/xge0000033

[37] Paul Dourish and Scott D. Mainwaring. 2012. Ubicomp's colonial impulse. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 133–142.

[38] Esther Duflo, Rachel Glennerster, and Michael Kremer. 2006. *Using Randomization in Development Economics Research: A Toolkit*. Technical Report t0333. National Bureau of Economic Research. doi:10.3386/t0333

[39] Jasper Feine, Ulrich Gnewuch, Stefan Morana, and Alexander Maedche. 2019. A Taxonomy of Social Cues for Conversational Agents. *Int. J. Hum.-Comput. Stud.* 132, C (Dec. 2019), 138–161. doi:10.1016/j.ijhcs.2019.07.009

[40] Kate Goddard, Abdul Roudsari, and Jeremy C. Wyatt. 2012. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association* 19, 1 (2012), 121–127.

[41] Noah J. Goldstein, Robert B. Cialdini, and Vladas Griskevicius. 2008. Nudge your neighbors: A field experiment on the impact of descriptive social norms on energy conservation. *Journal of Consumer Research* 35, 3 (2008), 472–482.

[42] Melanie C. Green and Timothy C. Brock. 2000. The Role of Transportation in the Persuasiveness of Public Narratives. *Journal of Personality and Social Psychology* 79, 5 (2000), 701–721. doi:10.1037/0022-3514.79.5.701

[43] Daniel Greene, Anna Lauren Hoffmann, and Luke Stark. 2019. Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning. *Proceedings of the 52nd Hawaii International Conference on System Sciences* (2019), 2122–2131.

[44] Michael Hallsworth, Tim Chadborn, Anna Sallis, Michael Sanders, Daniel Berry, Felix Greaves, Lara Clements, and Sally C. Davies. 2016. Provision of social norm feedback to high prescribers of antibiotics in general practice: a pragmatic national randomised controlled trial. *The Lancet* 387, 10029 (2016), 1743–1752. doi:10.1016/S0140-6736(16)00215-4

[45] Leslie J. Hinyard and Matthew W. Kreuter. 2007. Using Narrative Communication as a Tool for Health Behavior Change: A Conceptual, Theoretical, and Empirical Overview. *Health Education & Behavior* 34, 5 (2007), 777–792. doi:10.1177/1090198106291963

[46] Geert Hofstede. 2001. *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations* (2 ed.). Sage Publications, Thousand Oaks, CA.

[47] Carl I. Hovland and Walter Weiss. 1951. The Influence of Source Credibility on Communication Effectiveness. *Public Opinion Quarterly* 15, 4 (1951), 635–650. doi:10.1086/266350

[48] Lilly Irani, Janet Vertesi, Paul Dourish, Kavita Philip, and Rebecca E. Grinter. 2010. Postcolonial computing: a lens on design and development. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, 1311–1320. doi:10.1145/1753326.1753522

[49] Mohit Jain, Pratyush Kumar, Ishita Bhansali, Q. Vera Liao, Khai Truong, and Shwetak Patel. 2018. FarmChat: A Conversational Agent to Answer Farmer Queries. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 4, Article 170 (Dec. 2018), 22 pages. doi:10.1145/3287048

[50] Amelia Jamison, Rohini Ganjoo, Saraniya N. Tharmarajah, Bijay K. Panda, Manoj Parida, Supriya Pandey, and Rajiv N. Rimal. 2025. More work for ASHAs? Exploring the role of WhatsApp messaging and community health work in urban slum areas of Varanasi, India. *PLOS Global Public Health* 5, 10 (2025), e0005157. doi:10.1371/journal.pgph.0005157

[51] Sangya Kaphle, Sharad Chaturvedi, Indrajit Chaudhuri, Ram Krishnan, and Neal Lesh. 2015. Adoption and Usage of mHealth Technology on Quality and Experience of Care Provided by Frontline Workers: Observations From Rural India. *JMIR mHealth and uHealth* 3, 2 (2015), e61. doi:10.2196/mhealth.4047

[52] Rafal Kocielnik, Saleema Amershi, and Paul N. Bennett. 2019. Will You Accept an Imperfect AI? Exploring Designs for Adjusting End-user Expectations of AI Systems. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, Article 411, 14 pages. doi:10.1145/3290605.3300641

[53] Gerjo Kok, Nell H. Gottlieb, Gjalt-Jorn Y. Peters, Patricia Dolan Mullen, Guy S. Parcel, Robert A. C. Ruiter, María E. Fernández, Christine Markham, and L. Kay Bartholomew. 2016. A taxonomy of behaviour change methods: an Intervention Mapping approach. *Health Psychology Review* 10, 3 (2016), 297–312. doi:10.1080/17437199.2015.1077155

[54] Naja Kathrine Kollerup, Joel Wester, Mikael B. Skov, and Niels van Berkel. 2024. How Can I Signal You To Trust Me: Investigating AI Trust Signalling in Clinical Self-Assessments. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) *(DIS '24)*. Association for Computing Machinery, New York, NY, USA, 525–540. doi:10.1145/3643834.3661612

[55] Alain Labrique and et al. 2018. Best practices in scaling digital health in low- and middle-income countries. *Globalization and Health* 14, 1 (2018), 103.

[56] Carl Latkin and Amy R. Knowlton. 2005. The social ecology of urban health: Leveraging social networks to improve health. *American Journal of Public Health* 95, 12 (2005), 2169–2179.

[57] John D. Lee and Katrina A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors* 46, 1 (2004), 50–80. arXiv:https://journals.sagepub.com/doi/pdf/10.1518/hfes.46.1.50_30392 doi:10.1518/hfes.46.1.50_30392 PMID: 15151155.

[58] Tianyi Li, Mihaela Vorvoreanu, Derek Debellis, and Saleema Amershi. 2023. Assessing Human-AI Interaction Early through Factorial Surveys: A Study on the Guidelines for Human-AI Interaction. *ACM Trans. Comput.-Hum. Interact.* 30, 5, Article 69 (Sept. 2023), 45 pages. doi:10.1145/3511605

[59] Q. Vera Liao and S. Shyam Sundar. 2022. Designing for Responsible Trust in AI Systems: A Communication Perspective. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. ACM, New York, NY, 1257–1268. doi:10.1145/3531146.3533182

[60] Meagan B. Loerakker, Evropi Stefanidi, Jasmin Niess, Thomas Eßmeyer, and Paweł W. Woźniak. 2025. Give and Take: Perceptions of a Conversational Coach Agent in Fitness Trackers. *Proc. ACM Hum.-Comput. Interact.* 9, 5, Article MHCI014 (Sept. 2025), 36 pages. doi:10.1145/3743718

[61] Jennifer M. Logg, Julia A. Minson, and Don A. Moore. 2019. Algorithm Appreciation: People Prefer Algorithmic to Human Judgment. *Organizational Behavior and Human Decision Processes* 151 (2019), 90–103. doi:10.1016/j.obhdp.2018.12.005

[62] David Lyell, Farah Magrabi, Magdalena Z. Raban, and Enrico Coiera. 2017. Automation bias in electronic prescribing. *BMC Medical Informatics and Decision Making* 17, 1 (2017), 28.

[63] Shuai Ma, Xinru Wang, Ying Lei, Chuhan Shi, Ming Yin, and Xiaojuan Ma. 2024. "Are You Really Sure?" Understanding the Effects of Human Self-Confidence Calibration in AI-Assisted Decision Making. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. ACM, Honolulu, HI, USA. doi:10.1145/3613904.3642671

[64] Michael A. Madaio, Luke Stark, Jennifer Wortman Vaughan, and Hanna Wallach. 2020. Co-designing checklists to understand organizational challenges and opportunities around fairness in AI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, 1–14.

[65] Stanislao Maldonado, Deborah Martinez, and Lina Diaz. 2024. *Building a shield together: Addressing low vaccine uptake against cancer through social norms*. Working Papers 202. Peruvian Economic Association. doi:None

[66] Rafał K. Mantiuk, Anna Tomaszewska, and Radoslaw Mantiuk. 2012. Comparison of Four Subjective Methods for Image Quality Assessment. *Computer Graphics Forum* 31, 8 (2012), 2478–2491.

[67] Aaron Marcus and Emilie W. Gould. 2000. Crosscurrents: Cultural Dimensions and Global Web User-Interface Design. *interactions* 7, 4 (2000), 32–46. doi:10.1145/345190.345238

[68] Hazel Rose Markus and Shinobu Kitayama. 1991. Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review* 98, 2 (1991), 224–253. doi:10.1037/0033-295X.98.2.224

[69] Dan P. McAdams and Kate C. McLean. 2013. Narrative Identity. *Current Directions in Psychological Science* 22, 3 (2013), 233–238. doi:10.1177/0963721413475622

[70] Indrani Medhi, Mohit Jain, Anuj Tewari, Mohini Bhavsar, Michael Matheke-Fischer, and Edward Cutrell. 2012. Combating rural child malnutrition through inexpensive mobile phones. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design* (Copenhagen, Denmark) *(NordiCHI '12)*. Association for Computing Machinery, New York, NY, USA, 635–644. doi:10.1145/2399016.2399113

[71] Indrani Medhi, Devanuj Patnaik, Emma Brunskill, Shubham Gautama, William Thies, and Kentaro Toyama. 2012. Designing mobile interfaces for novice and low-literacy users. In *ACM Transactions on Computer-Human Interaction*, Vol. 18. 1–28.

[72] Tamir Mendel, Oded Nov, and Batia Wiesenfeld. 2024. Advice from a Doctor or AI? Understanding Willingness to Disclose Information Through Remote Patient Monitoring to Receive Health Advice. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW2, Article 386 (Nov. 2024), 34 pages. doi:10.1145/3686925

[73] Luise Metzger, Linda Miller, Martin Baumann, and Johannes Kraus. 2024. Empowering Calibrated (Dis-)Trust in Conversational Agents: A User Study on the Persuasive Power of Limitation Disclaimers vs. Authoritative Style. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 481, 19 pages. doi:10.1145/3613904.3642122

[74] Ministry of Health and Family Welfare, Government of India. n.d.. ASHA Training Modules. National Health Mission. https://nhm.gov.in/index1.php?lang=1&level=3&sublinkid=184&lid=257 Accessed: 2025.

[75] Shakir Mohamed, Marie-Therese Png, and William Isaac. 2020. Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology* 33 (2020), 659–684.

[76] Sean A. Munson and Sunny Consolvo. 2012. Exploring accountability in personal informatics: How social influence and social sharing can support reflection and behavior change. In *Proceedings of the ACM 2012 Conference on Ubiquitous Computing*. 471–480.

[77] Simone Natale, Federico Biggio, Payal Arora, John Downey, Riccardo Fassone, Rafael Grohmann, Andrea L. Guzman, Emily Keightley, Deqiang Ji, Vincent Obia, Aleksandra Przegalinska, Usha Raman, Paola Ricaurte, and Eduardo Villanueva-Mansilla. 2025. Global AI Cultures: How a cultural focus can empower generative artificial intelligence. *Commun. ACM* 68, 9 (2025). doi:10.1145/3722547

[78] Chinasa Okolo, Aashaka Kamath, Nicola Dell, and Aditya Vashistha. 2021. "It Cannot Do All of My Work": Community Health Worker Perceptions of AI-Enabled Mobile Health Applications in Rural India. In *CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA. doi:10.1145/3411764.3445420

[79] Chinasa Okolo, Aashaka Kamath, Nicola Dell, and Aditya Vashistha. 2024. "If it is easy to understand, then it will have value": Examining Perceptions of Explainable AI with Community Health Workers in Rural India. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1, Article 71 (2024). doi:10.1145/3637348

[80] Elizabeth Levy Paluck and Hana Shepherd. 2010. Peer influence on public behavior: Experimental evidence from a school-based intervention to reduce harassment. *American Journal of Public Health* 100, 12 (2010), 2425–2430.

[81] Theofanis Papakonstantinou, Sarah L. Flecke, Claire E. R. Edmunds, Natalie Gold, Michael Sanders, Tim Chadborn, Ivo Vlaev, and Rosie Chadwick. 2025. A systematic review and meta-analysis of the effectiveness of social norms messaging approaches for improving health behaviours in developed countries. *Nature Human Behaviour* (2025). doi:10.1038/s41562-025-02275-6

[82] Raja Parasuraman and Victor Riley. 1997. Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors* 39, 2 (1997), 230–253. arXiv:https://doi.org/10.1518/001872097778543886 doi:10.1518/001872097778543886

[83] Gain Park, Jiyun Chung, and Seyoung Lee. 2023. Human vs. machine-like representation in chatbot mental health counseling: the serial mediation of psychological distance and trust on compliance intention. *Current Psychology*

(2023), 1–12. doi:10.1007/s12144-023-04653-7
[84] Maria Perez-Ortiz, Aliaksei Mikhailiuk, Emin Zerman, Vedad Hulusic, Giuseppe Valenzise, and Rafal K Mantiuk. 2019. From pairwise comparisons and rating to a unified quality scale. (2019). doi:10.17863/CAM.43203
[85] H. Wesley Perkins. 2003. *The Social Norms Approach to Preventing School and College Age Substance Abuse.* Jossey-Bass.
[86] Kavita Philip, Lilly Irani, and Paul Dourish. 2012. Postcolonial computing: a tactical survey. In *Science and Technology Studies*. Vol. 25. 223–242.
[87] Devarsetty Praveen and et al. 2014. SMARTHealth India: Development and field evaluation of a mobile clinical decision support system for cardiovascular disease screening by frontline health workers. *Global Health Action* 7, 1 (2014), 25614.
[88] Irene Rae. 2024. The Effects of Perceived AI Use On Content Perceptions. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 978, 14 pages. doi:10.1145/3613904.3642076
[89] Divya Ramachandran, John Canny, Prabhu Dutta Das, and Edward Cutrell. 2010. Mobile-izing health workers in rural India. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1889–1898. doi:10.1145/1753326.1753610
[90] Pragnya Ramjee, Mehak Chhokar, Bhuvan Sachdeva, Mahendra Meena, Hamid Abdullah, Aditya Vashistha, Ruchit Nagar, and Mohit Jain. 2025. ASHABot: An LLM-Powered Chatbot to Support the Informational Needs of Community Health Workers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. ACM, New York, NY, 1–22. doi:10.1145/3706598.3713680
[91] Pragnya Ramjee, Bhuvan Sachdeva, Satvik Golechha, Shreyas Kulkarni, Geeta Fulari, Kaushik Murali, and Mohit Jain. 2025. CataractBot: An LLM-powered Expert-in-the-Loop Chatbot for Cataract Patients. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 9, 2, Article 45 (June 2025), 31 pages. doi:10.1145/3729479
[92] Rajiv N. Rimal, Rohini Ganjoo, Amelia Jamison, Manoj Parida, and Saraniya Tharmarajah. 2024. Social norms, vaccine confidence, and interpersonal communication as predictors of vaccination intentions: Findings from slum areas in Varanasi, India. *Vaccine* 42, 22 (2024), 126038. doi:10.1016/j.vaccine.2024.06.006
[93] Nithya Sambasivan, Ethan Arnesen, Ben Hutchinson, and Vinodkumar Prabhakaran. 2021. Re-imagining algorithmic fairness in India and beyond. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT)*. ACM, 315–328.
[94] Matthias Schmidmaier, Jonathan Rupp, Cedrik Harrich, and Sven Mayer. 2025. Using Nonverbal Cues in Empathic Multi-Modal LLM-Driven Chatbots for Mental Health Support. *Proc. ACM Hum.-Comput. Interact.* 9, 5, Article MHCI039 (Sept. 2025), 34 pages. doi:10.1145/3743724
[95] P. Wesley Schultz, Jessica M. Nolan, Robert B. Cialdini, Noah J. Goldstein, and Vladas Griskevicius. 2007. The Constructive, Destructive, and Reconstructive Power of Social Norms. *Psychological Science* 18, 5 (2007), 429–434. arXiv:https://doi.org/10.1111/j.1467-9280.2007.01917.x doi:10.1111/j.1467-9280.2007.01917.x PMID: 17576283.
[96] Kerry Scott, Asha S. George, and Rajani R. Ved. 2019. Taking stock of 10 years of published research on the ASHA programme: examining India's national community health worker programme from a health systems perspective. *Health Research Policy and Systems* 17, 1 (2019), 29. doi:10.1186/s12961-019-0427-0
[97] Erica Sedlander, Ichhya Pant, Jeffrey Bingenheimer, Hagere Yilma, Lipika Patro, Satyanarayan Mohanty, Rohini Ganjoo, and Rajiv Rimal. 2022. How does a social norms-based intervention affect behaviour change? Interim findings from a cluster randomised controlled trial in Odisha, India. *BMJ Open* 12, 7 (2022). arXiv:https://bmjopen.bmj.com/content/12/7/e053152.full.pdf doi:10.1136/bmjopen-2021-053152
[98] Anna-Maria Seeger, Jella Pfeiffer, and Armin Heinzl. 2021. Texting with Humanlike Conversational Agents: Designing for Anthropomorphism. *Journal of the Association for Information Systems* 22, 4 (2021), 931–967. doi:10.17705/1jais.00685
[99] Prakriti Shrestha, Kaosar Afsana, Manuj C. Weerasinghe, Henry B. Perry, Harsha Joshi, Nisha Rana, Zahid Ali Memon, Nazrana Khaled, Sumit Malhotra, Surbhi Bhardwaj, Simrin Kafle, Yoko Inagaki, Austin Schimdt, Stephen Hodgins, Dinesh Neupane, and Krishna D. Rao. 2024. Strengthening primary health care through community health workers in South Asia. *The Lancet Regional Health – Southeast Asia* 28 (2024), 100463. doi:10.1016/j.lansea.2024.100463
[100] Ian René Solano-Kamaiko, Dibyendu Mishra, Nicola Dell, and Aditya Vashistha. 2024. Explorable Explainable AI: Improving AI Understanding for Community Health Workers in India. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–21. doi:10.1145/3613904.3642733
[101] Christina Steindl et al. 2015. Understanding Psychological Reactance. *Zeitschrift für Psychologie* 223, 4 (2015), 205–214. doi:10.1027/2151-2604/a000222
[102] Huatong Sun. 2012. *Cross-Cultural Technology Design: Creating Culture-Sensitive Technology for Local Users.* Oxford University Press, New York, NY. https://global.oup.com/us/companion.websites/9780199744763/

[103] Margaret E. Tankard and Elizabeth Levy Paluck. 2016. Norms and behavior: An interdisciplinary perspective. *Perspectives on Psychological Science* 11, 4 (2016), 539–560.

[104] Kentaro Toyama. 2015. *Geek Heresy: Rescuing Social Change from the Cult of Technology.* PublicAffairs, New York, NY.

[105] Annemiek van Boeijen and Yvo Zijlstra. 2020. *Culture Sensitive Design: A Guide to Culture in Practice.* BIS Publishers, Amsterdam. https://www.bispublishers.com/culture-sensitive-design.html

[106] Deepak Varuvel Dennison, Mohit Jain, Tanuja Ganu, and Aditya Vashistha. 2026. Designing Culturally Aligned AI Systems For Social Good in Non-Western Contexts. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26).* Association for Computing Machinery, New York, NY, USA, 1–22. doi:10.1145/3772318.3791513

[107] Aditya Vashistha, Neha Kumar, Anil Mishra, and Richard Anderson. 2016. Mobile Video Dissemination for Community Health. In *Proceedings of the Eighth International Conference on Information and Communication Technologies and Development (ICTD '16).* Association for Computing Machinery, New York, NY, USA, 1–11. doi:10.1145/2909609.2909655

[108] Aditya Vashistha, Neha Kumar, Anil Mishra, and Richard Anderson. 2017. Examining Localization Approaches for Community Health. In *Proceedings of the 2017 Conference on Designing Interactive Systems (DIS '17).* Association for Computing Machinery, New York, NY, USA, 357–368. doi:10.1145/3064663.3064754

[109] Nervo Verdezoto, Naveen Bagalkot, Syeda Zainab Akbar, Swati Sharma, Nicola Mackintosh, Deirdre Harrington, and Paula Griffiths. 2021. The Invisible Work of Maintenance in Community Health: Challenges and Opportunities for Digital Health to Support Frontline Health Workers in Karnataka, South India. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 91 (April 2021), 31 pages. doi:10.1145/3449165

[110] Oleksandra Vereschak, Gilles Bailly, and Baptiste Caramiaux. 2021. How to Evaluate Trust in AI-Assisted Decision Making? A Survey of Empirical Methodologies. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 327 (Oct. 2021), 39 pages. doi:10.1145/3476068

[111] Hye Sun Yun and Timothy Bickmore. 2025. Framing Health Information: The Impact of Search Methods and Source Types on User Trust and Satisfaction in the Age of LLMs. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25).* Association for Computing Machinery, New York, NY, USA, Article 286, 7 pages. doi:10.1145/3706599.3720239

[112] Xiao Zhan, Noura Abdi, William Seymour, and Jose Such. 2024. Healthcare Voice AI Assistants: Factors Influencing Trust and Intention to Use. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 62 (April 2024), 37 pages. doi:10.1145/3637339

[113] Yunfeng Zhang, Q. Vera Liao, and Rachel K. E. Bellamy. 2020. Effect of Confidence and Explanation on Accuracy and Trust Calibration in AI-Assisted Decision Making. In *Proceedings of the 2020 ACM Conference on Fairness, Accountability, and Transparency (FAccT '20).* ACM, New York, NY, 295–305. doi:10.1145/3351095.3372852

Appendix

Appendix A: Power Calculation Details

The effective observation counts were computed using the standard cluster design-effect adjustment to account for repeated measures from the same ASHA, calculated as:

$$N_{\text{eff}} = \frac{N_{\text{raw}}}{1 + (n-1)\rho} \tag{1}$$

where n is the number of tasks completed per ASHA and ρ is the intra-class correlation (ICC) across responses. We estimated plausible ICC values (0.10–0.20) using random-intercept models at the ASHA level. Power simulations [38] confirmed that this design achieved approximately 80% power to detect main framing effects of 12–15 percentage points at α=0.05. These values are typical for clustered behavioral experiments and were sufficient to detect main framing effects (e.g., whether one chatbot framing was systematically preferred, whether ASHAs calibrated trust by following correct more often than incorrect advice), though not subtle subgroup interactions (e.g., framing × ambiguity × education).

Appendix B: Statistical Models

We report three logistic regression models used to estimate trust outcomes. In all models, the dependent variable is whether worker i in task j *trusted* the chatbot's advice (p_{ij}). Results are reported as odds ratios with cluster-robust standard errors.

Model 1: Overall accuracy effect.

$$\text{logit}(p_{ij}) = \alpha + \beta\,\textsc{Accuracy}_{ij} \tag{2}$$

This baseline model estimates whether ASHAs were more likely to follow correct versus incorrect advice, regardless of chatbot framing.

- p_{ij}: probability that worker i follows advice in task j.
- α: intercept (baseline log-odds of following wrong advice).
- β: effect of accuracy (increase in log-odds of following when advice is correct).
- $\textsc{Accuracy}_{ij}$: indicator variable = 1 if advice is correct, 0 if incorrect.

Model 2: Chatbot differences.

$$\text{logit}(p_{ij}) = \alpha + \beta\,\textsc{Accuracy}_{ij} + \sum_{k \neq b_0} \gamma_k\,\mathbf{1}\{\textsc{Bot} = k\} + \sum_{k \neq b_0} \delta_k\,(\textsc{Accuracy}_{ij} \times \mathbf{1}\{\textsc{Bot} = k\}) \tag{3}$$

This model adds chatbot framing to test whether different bots shift levels of trust or calibration relative to the neutral baseline.

- γ_k: main effect of chatbot k (difference in *Overall Trust* relative to baseline, when advice is wrong).
- δ_k: interaction between chatbot k and accuracy (difference in *Calibrated Trust* relative to baseline).
- $\mathbf{1}\{\textsc{Bot} = k\}$: indicator for chatbot persona k.
- b_0: reference bot, the neutral *Info-Bot*.

Model 3: Demographic moderation.

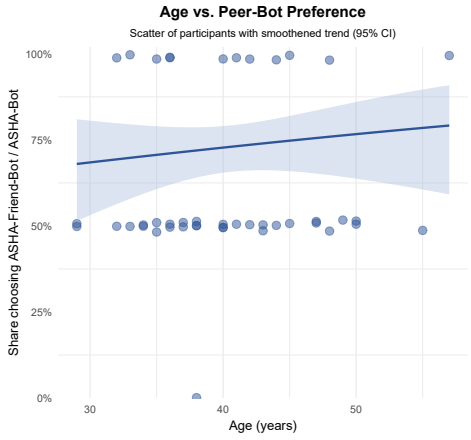
$$\begin{aligned}\operatorname{logit}(p_{ij}) = \alpha + \beta\,\textsc{Accuracy}_{ij} + \sum_{k\neq b_0} \gamma_k\,\mathbf{1}\{\textsc{Bot} = k\} \\ + \sum_{k\neq b_0} \delta_k\,(\textsc{Accuracy}_{ij} \times \mathbf{1}\{\textsc{Bot} = k\}) \\ + \theta^{\top}\mathbf{D}_i + \phi^{\top}(\textsc{Accuracy}_{ij} \times \mathbf{D}_i)\end{aligned} \tag{4}$$

This full model adds worker demographics to test whether age, education, or experience moderate trust or calibration.

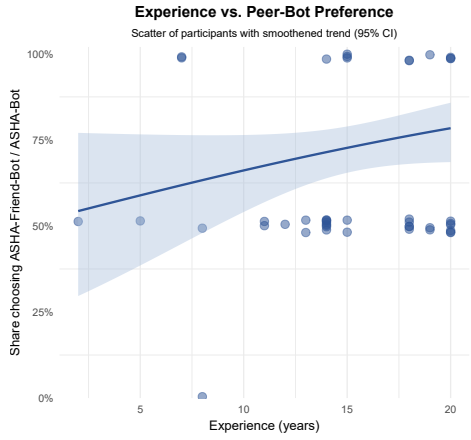
- $\mathbf{D}_i = (\textsc{Age}_i,\ \textsc{Education}_i,\ \textsc{Experience}_i)$: demographic covariates for worker i.
- θ: coefficients for main effects of demographics (baseline trust differences).
- ϕ: coefficients for interactions between demographics and accuracy (differences in *Calibrated Trust* by subgroup).

In all models, standard errors are clustered at the participant level to account for repeated measures.

Appendix C: Demographic Variation in Bot Preference

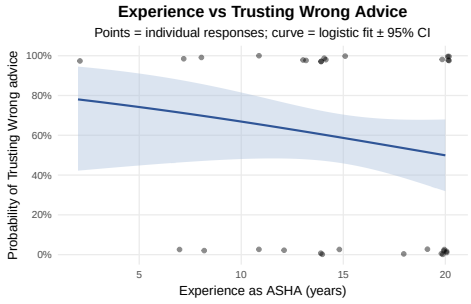


(a) Age (years)

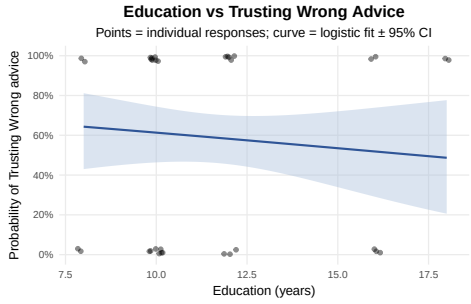


(b) Experience (years)

Fig. A1. Demographic gradients in peer-bot preference. Each point represents one ASHA's proportion of peer-oriented choices; smooths show descriptive trends.



(a) Education vs. probability of following wrong advice (logistic smooth with 95% CI)



(b) Experience vs. probability of following wrong advice (logistic smooth with 95% CI)

Fig. A2. Demographic trends in trusting the wrong advice given by a chatbot interface.

Appendix D: Qualitative Coding Process

We analyzed the 61 interview transcripts using Braun and Clarke's reflexive thematic analysis [17]. We chose this approach because our goal was interpretive. We wanted to understand how ASHAs themselves reasoned about trust, framing, and feasibility, and to let themes emerge from their own words. The analysis had four phases. In the first phase, the first author transcribed all interviews from Hindi audio and translated them into English. Where a direct translation would lose meaning, we kept the original Hindi phrasing in brackets. For example, "didi" is a respectful term for an

older sister or peer. In the second phase, the first author did line-by-line inductive coding on all 61 transcripts. This produced an initial codebook of 47 codes, grouped into five broad categories. In the third phase, all co-authors met across multiple sessions to combine related codes into higher-order themes. We drew on both the patterns in the data and on theories of normative framing (descriptive, narrative identity, and injunctive). In the fourth phase, we returned to the transcripts to verify each theme. We checked that each theme appeared across multiple participants. We discarded any theme that was driven by a single account. The themes reported in Section 4, such as "peer-like reassurance encouraged fast acceptance" and "social proof substituted for missing knowledge," came out of this process. Table A1 lists the five categories, the codes within each, and an example quote from the interviews.

Table A1. Codebook of qualitative themes. Five categories with 47 codes in total emerged from inductive coding of 61 interview transcripts.

Category	**Representative codes**	**Example quote and context**
Affective responses to framing	"feels like a peer or sister"; "feels judgmental or scolding"; "reads like a training session"; "feels impersonal"; "feels like a government circular"; "approachable for sensitive topics"	"It felt like talking to another ASHA who understands." (P43, on ASHA-Friend-Bot)
Decision rationales for trust and override	"matched my training"; "contradicted my training"; "uncertain, deferred to source"; "uncertain, deferred to peer endorsement"; "checked against own experience"; "fast acceptance without reflection"; "paused to verify"	"I knew the right answer already, but when it came in that official style I slowed down and compared." (P27, on Health-Dept-Bot)
Cross-framing comparisons	"Friend-Bot vs. supervisor's scolding"; "Health-Dept-Bot reads like a government circular"; "ASHA-Bot reflects what is done locally"; "Info-Bot feels generic"	"This was too formal and heavy. It felt like it was checking me instead of helping me." (P23)
Professional practice and hierarchy	"supervisor relationships"; "PHC distance and access"; "ANM consultations"; "peer network as first line of consult"; "training curriculum references"; "accountability concerns"	"The supervisor is at the PHC, which is often a few kilometers away… I rarely get network." (P14)
Contextual constraints	"network unavailability in the field"; "family resistance or aggression"; "caste- or religion-based hesitancy"; "travel distance"; "time pressure during fieldwork"; "literacy of recipient family"	"A Muslim woman was hesitant to get vaccinated because there were rumors … The answer was not in the training handbook." (P53)