AmarDoctor: Designing an AI-Driven, Multilingual Digital Health Platform for Underserved Bengali-Speaking Communities

Project Overview

AmarDoctor is an AI-driven, multilingual, voice-interactive digital health platform designed to support primary-care triage, symptom interpretation, and everyday health decision-making for underserved Bengali-speaking populations. I designed and developed AmarDoctor as a response to structural barriers in healthcare access, including language constraints, low health literacy, fragmented medical records, and limited familiarity with conventional digital interfaces.

The platform integrates conversational interfaces, multimodal data processing, and human-centered AI to enable patients, clinicians, and caregivers to interact with health information in ways that align with their abilities, communication styles, and real-world constraints.

My role: End-to-end designer and developer (user research, interaction design, system architecture, implementation, deployment, and evaluation)

Problem Context and Motivation

In Bangladesh and among Bengali-speaking communities more broadly, access to reliable primary-care guidance is often limited by:

1. **Language mismatch:** Most digital health tools operate in English or formal medical Bengali, while patients describe symptoms using colloquial language.
2. **Low health literacy:** Users frequently struggle to interpret symptoms, navigate menus, or understand clinical terminology.
3. **Fragmented medical history:** Medical records are primarily paper-based, handwritten, and easily lost or damaged.
4. **Digital unfamiliarity:** Older adults and low-literacy users often find menu-driven apps cognitively demanding.

Early field observations and informal interviews revealed that existing digital health solutions failed not because of lack of intelligence, but because they did not align with how people *communicate, remember, and make decisions* about their health.

This motivated the central design question: **How can an AI-based health platform adapt to users' linguistic, cognitive, and contextual realities rather than forcing users to adapt to technology?**

User Groups

AmarDoctor was designed with multiple overlapping stakeholders in mind:

1. **Primary users:** Patients with limited health literacy and varying levels of technological familiarity
2. **Secondary users:** Older adults managing chronic conditions
3. **Tertiary users:** Clinicians who require structured, interpretable patient information.

Rather than designing a single "ideal user", the system was intentionally built to support **diverse abilities, preferences, and access constraints**.

Research and User Insights

Key Findings

Through deployment feedback, clinician discussions, and iterative testing, several consistent insights emerged:

- **Colloquial expression dominates symptom reporting**
  Users rarely describe symptoms using textbook terminology. Phrases like " বুক জ্বালা করছে " (burning in chest) or metaphorical expressions were common, making rigid symptom checkers unusable.

- **Menu-based navigation increases cognitive load**
  Older adults and first-time users struggled with hierarchical menus and form-based input, often abandoning tasks mid-way.

- **Medical history recall is unreliable**
  Many users could not recall diagnoses or medications accurately because records were scattered across handwritten prescriptions.

- **Trust depends on interpretability**
  Users and clinicians both expressed discomfort with "black-box" AI recommendations without understandable explanations.

These insights directly shaped both **interaction design** and **system architecture** decisions.

Design Strategy

1. Conversational Interface as the Primary Entry Point

To reduce navigation complexity, I designed a **conversational assistant** that acts as a unified interface for:

1. Symptom reporting
2. Consultations booking
3. Find doctor chambers or facilities like blood bank and hospitals
4. Guidance and next-step suggestions

This shifted interaction from *menu navigation* to *natural dialogue*, aligning better with users' everyday communication patterns.

2. Multilingual and Colloquial Language Support

I implemented a multilingual symptom-assessment and clinical decision support pipeline that:

1. Recognizes colloquial Bengali expressions
2. Translates them into structured clinical representations
3. Maps them to standardized symptom categories
4. Provides provisional diagnostic suggestions with associated confidence scores **exclusively** to physicians, who can verify or override them.

This allowed users to speak naturally without needing medical vocabulary, while still generating clinically meaningful outputs and streamlining physicians' workflows.

3. Accessibility-First Interaction Design

Design decisions were guided by accessibility and inclusivity: 1. Minimal text density 2. Clear visual hierarchy 3. Voice input and output to support low-literacy users 4. Reduced reliance on multi-step workflows.

4. Transparency and Trust

To address concerns around AI opacity: 1. Outputs were framed as *guidance*, not diagnoses 2. Language emphasized user autonomy 3. Explanations were presented in simple, contextual terms.

Core Modules and System Components

1. AI-Driven Symptom Assessment

- Conversational intake using multilingual NLP
- Context-aware follow-up questions
- Primary-care triage recommendations

HCI focus: Aligning conversational flow with how users narrate illness rather than forcing predefined forms.

2. Medical Document Digitization Pipeline

To address fragmented medical histories, I developed a pipeline that: - Extracts text from handwritten prescriptions and reports - Identifies diagnoses, medications, and dates - Converts them into structured digital health records.

HCI focus: Reducing memory burden on patients and improving continuity of care without requiring technical expertise.

3. Dietary Guidance via Multimodal Interaction

Recognizing daily decision fatigue among patients with chronic illnesses, I built a module that: - Identifies food items from images - Provides culturally relevant dietary suggestions - Adapts recommendations to health conditions.

HCI focus: Supporting everyday health decisions through lightweight, context-aware assistance rather than clinical overload.

4. Clinician-Facing Summaries

Clinicians receive structured summaries generated from patient interactions: - Symptom timelines - Extracted medical history - Key risk indicators.

HCI focus: Integrating AI outputs into clinician workflows without increasing cognitive burden.

Prototyping and Implementation

- Iterative prototyping with continuous feedback
- Rapid deployment to real user groups
- Short video demos embedded to demonstrate interaction flows

Technical implementation details are intentionally secondary in this case study; emphasis is placed on **how technical decisions supported human-centered goals**.

Evaluation and Outcomes

Evaluation Approach

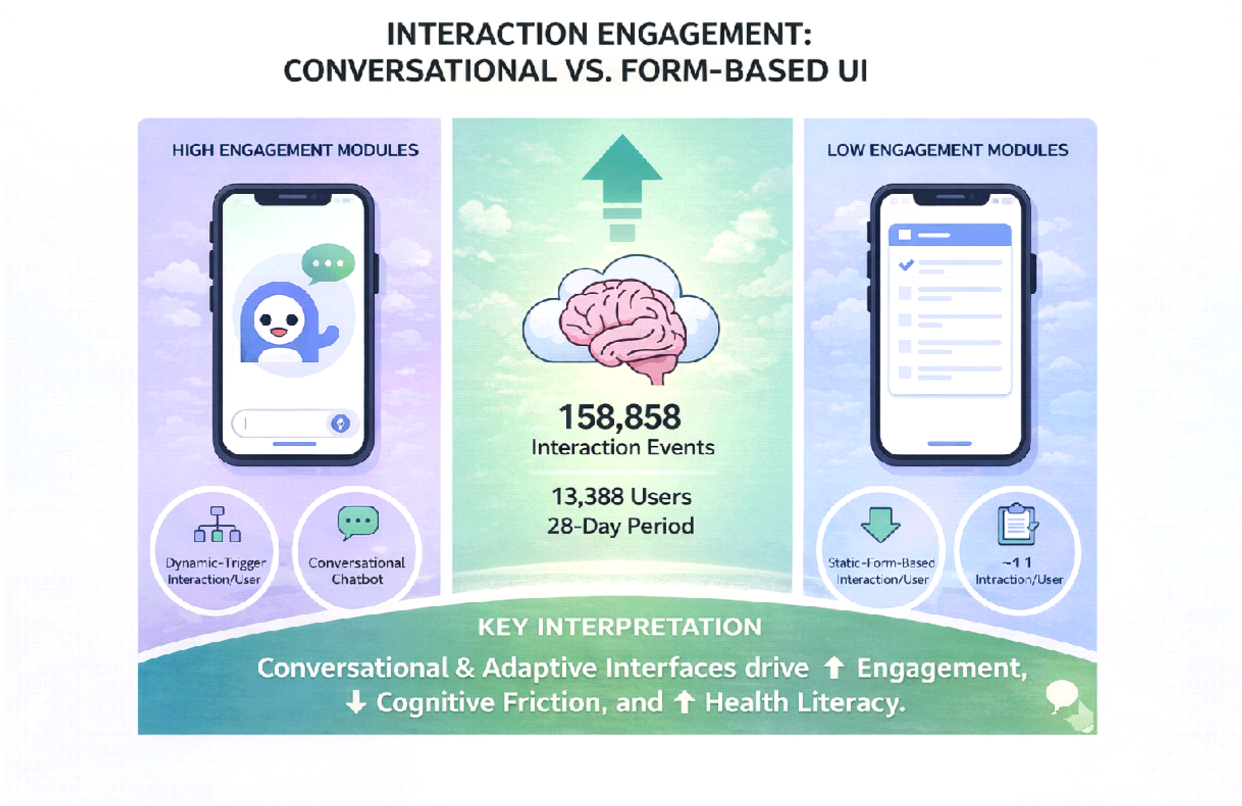
The **AmarDoctor** platform was evaluated using a mixed quantitative approach combining interaction log analysis from real-world deployment and controlled diagnostic validation using clinically constructed vignette cases. This dual strategy was chosen to assess both user interaction behavior and technical reliability, reflecting the human-centered goals of the system.

Usage and Engagement Analysis

Over a **28-day** period, the platform recorded **158,858** interaction events from **13,388** users. Event-level logs were analyzed to understand how users interacted with different system modules and to identify patterns of sustained engagement.

Analysis revealed that conversational and adaptive modules were used more intensively than static, form-based interfaces. The dynamic triage module recorded **15,439 events** from **2,057 users**, with an average of **7.51** interactions per active user, indicating multi-step engagement within a single health inquiry. Similarly, the conversational chatbot module recorded **10,187** events from **1,835 users**, averaging **5.56** interactions per user, suggesting repeated dialogue-based interaction across sessions.

In contrast, traditional static screens, such as consultation history checking, medical record viewing, and appointment booking, showed substantially lower interaction depth, averaging approximately **1.5** interactions per user. These patterns suggest that users primarily engaged with static screens for single-pass tasks, while conversational modules supported exploratory and iterative health decision-making.



Diagnostic Validation

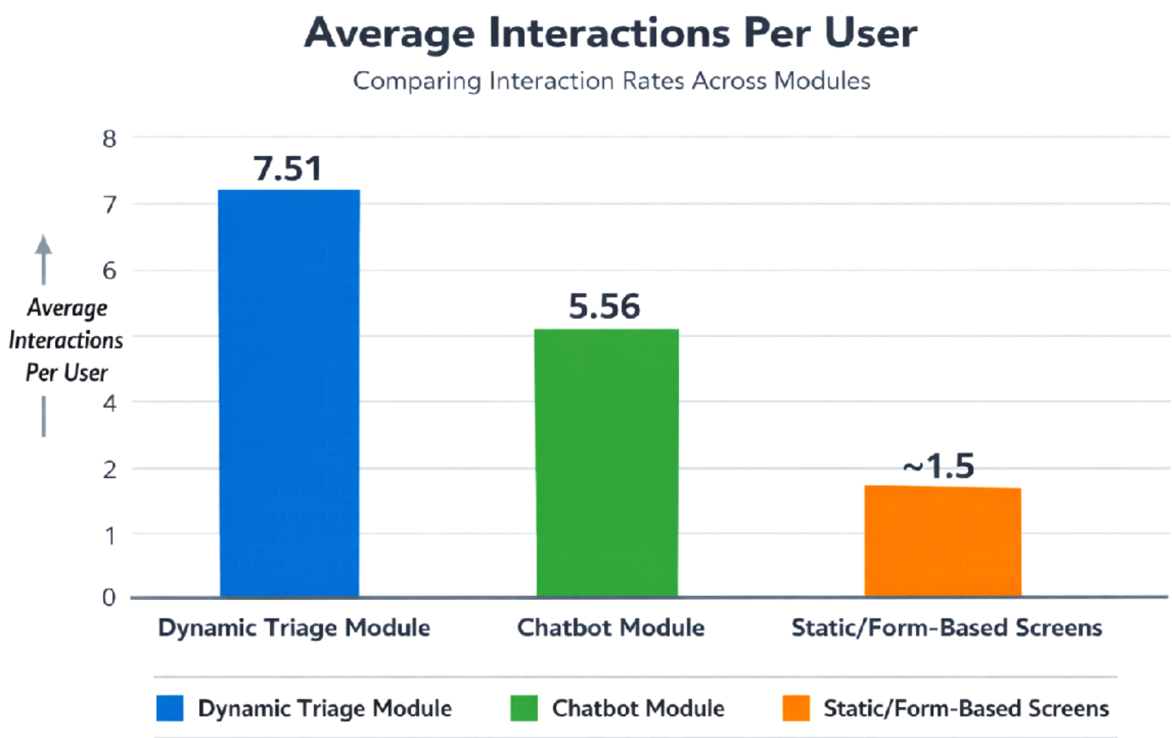
To evaluate the technical reliability of symptom interpretation and diagnostic suggestion, the system was validated against **185** clinically constructed vignette cases representing common primary-care scenarios. Each vignette included structured symptom descriptions and reference diagnoses.

Across these cases, the system achieved an overall diagnostic precision of **81.08%**, indicating consistent alignment between user-described symptoms and structured diagnostic outputs under controlled conditions. This evaluation focused on precision rather than recall to prioritize correctness of suggested conditions in an assistive, non-clinical decision-support context.

Interpretation of Findings

Taken together, these results highlight the complementary roles of interaction design and technical performance in AI-driven healthcare systems. While diagnostic precision establishes a baseline level of system reliability, engagement metrics reveal how interaction models shape real-world adoption and sustained use.

The substantially higher interaction depth observed in conversational modules suggests that interfaces aligned with users' natural language and cognitive models can reduce interaction friction and support more sustained engagement, particularly for users with limited health literacy. These findings reinforce the importance of evaluating AI systems not only by predictive performance, but also by how effectively they integrate into users' everyday decision-making practices.



Observed Impact

- **Deeper engagement through conversational design**
  Users interacted more frequently with the dynamic triage and chatbot modules than with static screens, suggesting that conversational and adaptive interfaces better matched how users think and communicate about health concerns.
- **Reduced interaction friction for low-literacy users**
  Repeated, multi-step use of the triage flow indicates that users were able to navigate health decisions without relying on complex medical terminology or form-based interfaces.
- **Increased trust in AI-assisted guidance**
  Continued use of conversational modules across sessions suggests that users perceived the system as a reliable support tool rather than a one-time information source.

While formal large-scale user studies were beyond scope, iterative deployment provided strong qualitative validation of the design direction.

Limitations and Future Work

Evaluation scope

Engagement metrics were derived from system interaction logs rather than controlled laboratory studies. While these logs capture naturalistic usage patterns at scale, they do not fully explain users' motivations, goals, or reasons for disengagement.

User diversity and context

Although the platform reached a large user base, the evaluation does not yet disaggregate engagement by demographic variables such as age, literacy level, or health condition. As a result, it is difficult to attribute observed interaction patterns to specific user groups or contextual factors.

Clinical validation boundaries

Diagnostic precision was evaluated using clinically constructed vignette cases rather than real-world clinical outcomes. While this approach enables consistent benchmarking, it does not replace prospective clinical validation or physician-in-the-loop assessment.

Interaction quality beyond frequency

Event counts capture how often users interact with system modules, but not the qualitative quality of those interactions. Future studies will incorporate task success measures, error recovery analysis, and user-reported confidence to better assess interaction effectiveness.

Future directions

Planned work includes mixed-methods evaluation combining interaction logs with usability testing, interviews, and participatory design sessions, as well as expanding validation to real-world clinical settings in collaboration with healthcare professionals.

Reflections and Learnings

Working on **AmarDoctor** showed me that building effective health AI systems is not only about making accurate models, but also about designing interactions that people can understand and trust. Even though the system achieved reasonable diagnostic precision, real usage data showed that users spent more time with conversational and adaptive features than with static screens. This helped me see how important interaction design is for encouraging continued use, especially for people with limited health literacy.

Looking at interaction logs changed how I think about evaluation. Instead of treating engagement numbers as secondary, I learned to see repeated interaction as a sign that the system matched how users think and communicate about health. The dynamic triage and chatbot modules, in particular, supported step-by-step exploration rather than one-time information lookup.

At the same time, the project made clear that numbers alone cannot explain the full user experience. While logs showed what users did, they did not explain why users trusted the system or where they felt confused. This experience increased my interest in mixed-methods HCI research, combining usage data with usability testing and interviews. Overall, AmarDoctor strengthened my goal of pursuing graduate study in HCI to design AI systems that are both technically reliable and genuinely helpful in real-world healthcare settings.

In summary, Designing AmarDoctor fundamentally reshaped my understanding of HCI: Accessibility is not an add-on; it must shape system architecture from the start - Conversational interfaces can lower barriers, but only when grounded in users' lived language - Trust in AI systems emerges from transparency, not technical sophistication.

Most importantly, this project reinforced that **AI is only meaningful when it adapts to people, not when people adapt to AI**.

Relevance to HCI Research

AmarDoctor sits at the intersection of: - Human-centered AI - Accessible and adaptive interaction design - Multimodal and conversational interfaces - Health equity and participatory design.

This project motivates my pursuit of formal training in HCI research methods, particularly ethnography, participatory design, and qualitative evaluation, to strengthen future systems with rigorous, publishable research foundations.

Related publication:
AmarDoctor: An AI-Driven, Multilingual, Voice-Interactive Digital Health Application for Primary Care Triage and Patient Management to Bridge the Digital Health Divide for Bengali Speakers (arXiv, 2025)