UNDP accelerator labs

Information Integrity

AI-Assisted pre-bunking of misinformation on natural disasters

Iffat Anjum
Head of Exploration, Accelerator Lab
UNDP Bangladesh

Co-building the Accelerator Labs as a joint venture with:

german cooperation
DEUTSCHE ZUSAMMENARBEIT

QATAR FUND FOR DEVELOPMENT
صندوق قطر للتنمية

UNDP Core Partners

Action Partner:

Cabinet Office

The challenge

- **Misinformation impacting social and political landscapes**

  Mis/disinformation has evolved into a systemic threat, distorting public opinion, eroding trust in institutions, and fuelling social unrest.

- **Social media platforms and private networks**

  The digital landscape used for influence operations and disinformation campaigns, with coordinated narratives targeting government institutions, the military, and vulnerable groups, including women and religious and ethnic minorities. Often reaching Communities through Private Channels/Networks.

- **Hard to debunk once misinformation spreads**

  Once misinformation takes root, it becomes resistant to correction or debunking, even when confronted with factual evidence. False narratives, especially those tied to identity or emotion, tend to ‘stick’ in public memory, often outlasting or overshadowing subsequent debunking.

The solution

- **Pre-bunking approach**

  Using pre-bunking strategy to proactively alert the public of the types of misleading narratives that are likely to circulate in natural disaster, without responding to any specific claim.

- **Harnessing Digital Fluency of the youth**

  Bangladesh has a large base of digitally fluent youth who are active online. However, they lack a credible platform for raising awareness and inoculation about mis/disinformation.

- **Youth as trusted communicators to reach older adults**

  The youth hold trust within their families and communities, and can serve as effective communicators of accurate, accessible contents, leveraging private networks to reach the older adults.

- **Leveraging agentic AI to enhance the solution**
  - Predicts rumors analysing historical trends and real-time discourse analysis
  - User profiling to provide personalization of the tone in which information is delivered
  - Looks for and summarizes authoritative sources of information to explain how a rumor or given piece of information may be misleading

AI Tools:

(a) Multimodal Active Learning
(b) Foundational LLM (Llama, mistral) with Bengali and English

Key functions:

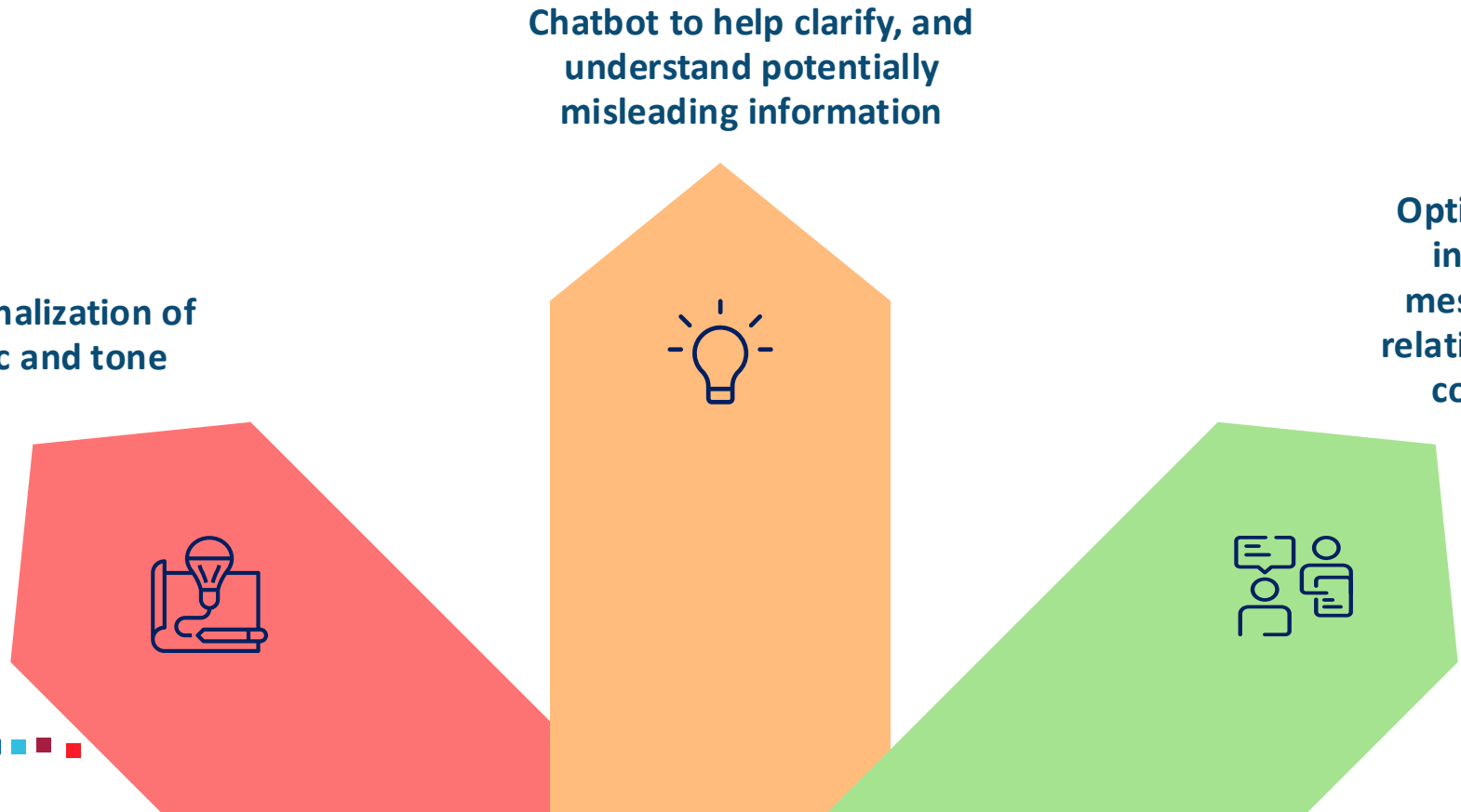
(a) Predict common patterns & techniques of misinformation
(b) Generate pre-bunking messages that are personalized in tone & framing via user profiling.

Prototype Design

A prototype pre-bunking application that provides personalized information on potential rumors and false or misleading narratives about natural disasters

Digitally fluent youth: Primary user

Older adults: Both direct and indirect user



How We Build the Weather-Related Misinformation Dataset

Data Collection

- Weather misinformation collected from Bangladesh + global sources
- Includes storms, cyclones, earthquakes, earthquakes, floods
- Global items reviewed by annotators annotators for Bangladesh relevance relevance

Relevance Annotation

- Annotators mark relevance to Bangladesh context
- If relevant: identify which region(s)
- Outputs geo-tagged rumors (coastal, (coastal, Sylhet, Rangpur, Dhaka, etc.) etc.)

Categorization

- Themes: fake casualty numbers, mega-earthquake rumors, weather-modification conspiracies, embankment collapse claims
- Extracts emotional triggers and linguistic patterns

Output: A geo-tagged misinformation knowledge base linking narratives → regions → virality patterns.

Predicting Region-Specific Misinformation Using Weather Forecasts

Weather Forecast Layer

Real-time district-level weather predictions

Storms, rainfall, heatwaves, cyclones

Historical Rumor Layer

Historic rumor patterns matched to each weather type

Includes frequency, virality, emotional tone

Matching Engine

AI model matches upcoming weather events with past misinformation themes

District-Level Prediction Map

Identifies which area may face which type of rumor

Example: Heavy rain in Sylhet → historical dam-collapse rumors

Personalized Pre-bunking Messaging Using User Personas + LLM

Young digitally fluent users

- Communication preference
- Tone sensitivity
- Language style

Older adults

- Communication preference
- Tone sensitivity
- Language style

Sensitive/anxious users

- Communication preference
- Tone sensitivity
- Language style

LLM Engine

- LLM rewrites core information in different voices
- Tone adaptation: calm, factual, humorous, empathetic
- Ensures cultural + linguistic tuning for Bangladesh

Calm/Factual Version

Heavy rainfall can lead to confusion online. Here's what actually happens...

Youth-Friendly Direct Version

Watch out! These rumors usually appear during heavy rain in your area...

Elder-Friendly Warm Version

If you hear unusual claims during the rain, don't worry. Here's the verified update...

Outcome: Each user receives a tailored, trusted, easy-to-understand inoculation message.

Pilot test

- User test with 19 participants over 3 weeks in Nov-Dec 2025

Demographic info

57.9% Male (11)

42.1% Females (8)

18-25 years of age

Regional representation

- Dhaka: 6 (31.6%)
- Chittagong: 3 (15.8%)
- Mymensingh: 4 (21.1%)
- Rajshahi: 4 (21.1%)
- Barishal: 2 (10.5%)

Key aspects

- Online training and app install
- … notifications/day
- Sharing with older adults in family and community
- Collected feedback on the user interface, key features etc.

System Usability Scale (SUS) Analysis

The System Usability Scale (SUS) is a quick and dirty, reliable tool for measuring the usability of a wide range of products and services. It consists of 10 Likert-scale items, where users rate statements from 1 (Strongly Disagree) to 5 (Strongly Agree).

Average SUS score **87.75/100**, which indicates **Excellent usability**, significantly above the industry benchmark of 68.

	Mean	Max	Median
I think that I would like to use this app frequently.	4.50	5	5.0
I found the app unnecessarily complex.	1.50	3	1.0
I thought the app was easy to use.	4.70	5	5.0
I think that I would need the support of a technical person to use this app.	1.20	2	1.0
I found that the various functions in this app were well-integrated.	4.60	5	5.0
I thought there was too much inconsistency in this app.	2.20	4	2.0
I would imagine that most people would learn to use this app very quickly.	4.80	5	5.0
I found the app very cumbersome to use.	1.90	3	2.0
I felt very confident using the app.	5.00	5	5.0
I needed to learn a lot of things before I could get going with this app.	1.00	1	1.0

Key Insights: Users show strong confidence (mean 5.0), minimal perceived learning curve (mean 1.0), and minor pain points related to inconsistency (mean 2.20) or cumbersomeness (mean 1.90).

Trust, Relevance & Behavior Analysis

5 Likert-scale items measuring trust and behavioral outcomes

Question	Mean	Max	Median
The misinformation prevention shown in the app felt reliable and factual.	3.50	5	4.0
The alerts and updates were timely and relevant to the weather in my location.	4.30	5	5.0
The app helped me identify false or misleading weather information online.	4.50	5	4.5
I shared at least one verified or rebuttal message from the app with others.	3.70	5	3.5
I am more confident about verifying weather-related claims online after using this app.	3.40	5	3.0

Key Insights: High scores on helpfulness (4.5) and timeliness (4.3), but lower confidence gains (3.4) suggest room for educational features. Reliability perception varies (3.5) with some development-stage issues noted.

Most Liked Features

Real-Time Notifications and Location-Based Based Alerts

(11/19 users)

> "Location based alert and rumors scanning"
>
> (Participant 01)

> "It shows which rumours are true and which are false. It also tells me if any disaster is coming to my selected area"
>
> (Participant 09)

AI Chatbot for Queries

(6/19 users)

> "I like the explanation of the chatbot, it explains the rumor rumor from different sources"
>
> (Participant 04)

> "Verifying news by AI"
>
> (Participant 16)

Standout Specifics

Earthquake Alerts

(8/19 users)

> "Recently there was a major earthquake in my area and I got notification about it... I got actual info from here with source"
>
> (Participant 01)

> "For the earthquake notifications i wasn't confused about the magnitudes "
>
> (Participant 11)

Disaster Information and Reliability

(7/19 users)

> "The notification and disasters related information"
>
> (Participant 02)

> " I was clarified that a big earthquake is coming is rumored"
>
> (participant 14)

Earthquake alerts were frequently cited as helpful during recent events, and the sourced information provided reduced user panic, enhancing feelings of safety.

Confusing or Complicated Aspects

A significant majority of users (17/19) reported that they found "nothing" confusing or complicated about the app, indicating an overall positive user overall positive user experience regarding clarity and ease of use.

Minor Concerns (3/19 users)

- Rumor news confusion
- Excessive notifications
- AI blinking/loading issues

> "Sometimes the rumor news felt confusing to me"
>
> (Participant 02)

> "When trying to talk with chatbot, every time it continuously continuously blinks for infinite time"
>
> (Participant 09)

These findings emphasize that most users found the app intuitive, with only minor technical issues and occasional content clarity questions emerging questions emerging from a small subset of participants.

Suggestions for Community Usefulness

1

Affiliation and Expansion (2/19 users)

"By affiliating with the government"

(Participant 09)

2

AI Improvements (5/19 users)

"Make the ai remember previous conversation, and a history tab for the ai"

(Participant 03)

" Ai doesn't give short , cumulative Answers" (participate 15)

3

Increase Awareness and Accessibility (6/19 users)

"Raising awareness about the app to my surroundings"

(Participant 01)

"Launch it and open this app for everyone"

(Participant 11)

Usage Patterns and Behavioral Changes

1

Notification-Driven Usage

(19/19 users)

> "After notifications I entered into the app and trying to understand what's new and chat with AI"
>
> (Participant 01)

2

Verification During Events

(19/19 users)

> "Yes recently after earthquake"
>
> (Participant 03)

All users used it for recent earthquakes or disasters.

3

Behavioral Change

(19/19 users)

> "Yes, it did. Now I use the app instead of random posts online"
>
> (Participant 02)

Note: One user (Participant 10) reported no change.

These patterns highlight a critical shift in user behavior: actively seeking out and relying on verified information from the app instead of engaging with unverified "random posts online" or rumors, especially during critical events.

Least Useful Features

Most users found all features valuable, with only a small subset identifying these as less useful.

Language Conversion (English to Bangla)

(8/19 users)

> "Converting to Bangla"
>
> (Participant 02)

Seen as unnecessary

Other Minor Features

(2/19 users)

Examples: World Map, pre-written AI text (the welcome text on the chatbot)

These findings suggest that "Language Conversion (English to Bangla)" and other minor features could be deprioritized in future updates to future updates to optimize resources for more impactful developments.

Top Improvement Recommendations for National National Rollout

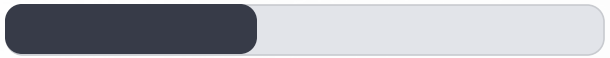
14/19

Accuracy and Timing

(14/19 users)

> "Accuracy"
>
> (Participant 03)

Enhance alert precision



8/19

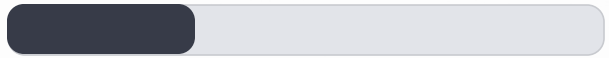
UI and Speed

(8/19 users)

> "UI and Speed "
>
> (Participant 01)

> "UI is very basic and boring. Improve the speed of loading a page"
>
> (Participant 10)

Make more engaging and faster



6/19

AI Enhancements

(6/19 users)

> "A backup system for the chat history with the chatbot"
>
> (Participant 02)

Backup chat history, better responses .

Ai gives long context answer and can't summarize (participant 15)

Recommendations and way forward

Strategic priorities based on user research findings

1

Prioritize Core Strengths

Enhance real-time alerts and AI verification, as these drive satisfaction

2

Address Pain Points

Improve AI reliability, add chat history, refine UI/speed for better engagement

3

Expand Reach

Focus on awareness campaigns and government partnerships to boost community impact

4

Monitor Metrics

Track SUS (target >85) and trust scores post-improvements

accelerator labs

~~Best~~ Next practices for a more sustainable future

Get in touch
iffat.anjum@undp.org
ishtiaque@cs.toronto.edu
jeremy.boy@undp.org

accelatorlabs.undp.org

Co-building the Accelerator Labs as a joint venture with:

Action Partner:

german cooperation
DEUTSCHE ZUSAMMENARBEIT

QATAR FUND FOR DEVELOPMENT
صندوق قطر للتنمية

UNDP Core Partners

Cabinet Office

UNDP Core Partners

Action Partners

MINISTRY OF ECOLOGICAL TRANSITION

giz

Healthy Internet Project

AfriLabs

Center for Music Ecosystems

newcities

Honey Bee Network

nesta

Utrecht University School of Economics

QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

IESE Business School University of Navarra

HYUNDAI

COLUMBIA | SIPA
School of International and Public Affairs

Cabinet Office

MIT MANAGEMENT SLOAN SCHOOL

Value proposition and impact

- Value proposition of the solution
  - Anticipatory Misinformation Shield
  - Localized and Culturally Tuned Messaging
  - Youth-Led Community Dissemination
- Expected impact if this solution were to scale
  - Strengthened Public Resilience and Trust
  - Scalable Civic Infrastructure for Crisis Communication