Localising AI for crisis response

Putting power back in the hands of frontline humanitarians and local communities

AUTHORS
Aleks Berditchevskaia
Kathy Peach
Isabel Stewart

ABOUT NESTA

We are Nesta. The UK's innovation agency for social good. We confront challenges that affect millions of people, from inequality and ill-health to the climate crisis.

We believe that innovation offers more potential now than ever before. We see opportunities to mobilise citizens and influence behaviour. Private and public capital that can be used more creatively. A wealth of data to mine.

And so we draw on these rich resources by bringing together diverse teams. Data scientists, designers and behavioural scientists. Practitioners, academics, entrepreneurs and people with lived experience.

Together, we design, test and scale new solutions to society's biggest problems. We partner with frontline organisations, build new businesses and work to change whole systems. Harnessing the rigour of science and the creativity of design, we work relentlessly to put new ideas to the test.

We'll keep going until we change millions of lives, for the better. Find out more at www.nesta.org.uk

ABOUT NESTA'S CENTRE FOR COLLECTIVE INTELLIGENCE DESIGN

Nesta's Centre for Collective Intelligence Design helps create new ways for communities to use technology to harness their insights, ideas and power to act on the problems that matter to them and create the futures they want. We design tools and projects that allow communities to respond collectively to challenges, and that help public and voluntary sector institutions strengthen trust and collaboration with citizens. We use rigorous research methods to test, learn and evaluate each solution. Our flagship Collective Intelligence Design Playbook helped to define the field and is used by practitioners around the world. We have worked with organisations from the UN to the BBC.

To learn more, see nesta.org.uk/project/centre-collective-intelligence-design or email the team at collective.intelligence@nesta.org.uk

---

AUTHORS

Aleks Berditchevskaia
Kathy Peach
Isabel Stewart

Recommended citation: Berditchevskaia, A., Peach, K., Stewart, I., (2022). *Localising AI for crisis response*. London: Nesta.

If you'd like this publication in an alternative format, such as Braille or large-print, please contact us at: information@nesta.org.uk

Design: Green-Doe Graphic Design Ltd.

Published July 2022
ISBN: 978-1-913095-65-9

PHOTO CREDITS

All workshop photos were obtained with participants' informed consent. Photos taken by Sarita Dhungana, Cathy Essouma, Kathy Peach, and Saurav Poudel.

FURTHER READING AND OTHER OUTPUTS FROM THIS PROJECT

Detailed methodology, results and resources: Nepal (add link when available)

Other outputs from this project include a landscape analysis of Collective crisis intelligence for frontline humanitarian response and a briefing paper on Participatory AI for humanitarian innovation.

Localising AI for crisis response

Putting power back in the hands of frontline humanitarians and local communities

00

Executive summary

Photo: Edouard Tamba, Unsplash

In this project we set out to show that there is an alternative to the current trajectory of artificial intelligence (AI) development in the humanitarian sector. We aimed to put the power of AI in the hands of frontline humanitarian responders – rather than international actors in the Global North. And to address the critiques and risks of humanitarian AI, including bias and lack of transparency, by involving local communities in the design, development and evaluation of this technology.

Our project aimed to design and evaluate new proof-of-concept **Collective Crisis Intelligence** tools. These are tools that combine data from crisis-affected communities with the processing power of AI to improve humanitarian action. It is a subset of the broader field of Collective Intelligence Design. We involved local communities and other stakeholders in the development process using **Participatory AI** methods to try to mitigate the main concerns about humanitarian AI.

The project was a partnership between Nesta's Centre for Collective Intelligence Design (CCID) and Data Analytics Practice (DAP), the Nepal Red Cross and Cameroon Red Cross, IFRC Solferino Academy, and Newcastle University's Open Lab. It was funded by the UK Humanitarian Innovation Hub, and carried out between April 2021 and May 2022.

Between April and September 2021, we researched and published the first ever landscape analysis of Collective Crisis Intelligence for Frontline Humanitarian Response, and a briefing paper on Participatory AI for Humanitarian Innovation. Between August 2021 and May 2022, we developed and evaluated early prototypes of **two new collective crisis intelligence tools**:

- **NFRI-Predict**[1] is a tool that predicts which non-food aid items (NFRI) are most needed by different types of households in different regions of Nepal after a crisis.
- **Report and Respond** is a French language SMS-based tool that allows Red Cross volunteers in Cameroon to check the accuracy of COVID-19 rumours or misinformation they hear from the community while they're in the field, and receive real-time guidance on appropriate responses.

Photo: Lucian Alexe, Unsplash

We found that collective crisis intelligence has the potential to make local humanitarian action more timely and appropriate to local needs.

- In Nepal, Red Cross Society staff rated the NFRI-Predict tool more highly for accuracy (the ability to match the needs of communities) and vulnerability (the ability to meet the needs of the most vulnerable) than the current process for determining NFRI distribution. A majority also thought the tool would make the process of NFRI distribution faster.
- In Cameroon, the results of a comparative evaluation showed that Red Cross volunteers thought the Report and Respond tool would lead to improvements in the timeliness and ability to match community needs when providing responses to misinformation about COVID-19.

We demonstrated that collective crisis intelligence can also transform locally-generated data to drive new forms of (anticipatory) action.

- In Cameroon, we showed that collective crisis intelligence could elevate the utility of existing community feedback data. We repurposed data already held by the IFRC, and used it to prototype a faster way to report, monitor and respond to COVID-19 misinformation – with the goal of containing existing rumours and identifying emerging rumours before they spread.
- In Nepal we demonstrated that applying AI to a new dataset of community NFRI preferences from geographically distinct regions and more than 3000 households could give the Nepal Red Cross important insights into which aid items will be needed in advance of a crisis. This has the potential to enable more anticipatory action, including through the stockpiling and prepositioning of the most essential goods for people in different regions.

We validated that collective crisis intelligence and participatory AI can help increase trust in AI tools, but more research is needed to untangle the factors that were responsible.

- In both countries, the majority of frontline staff and volunteers rated the CCI tools we developed more highly on 'trust' than their current operational processes.

We confirmed that participatory AI can overcome several critiques and limitations of AI – as well as helping to improve model performance.

- Participatory AI activities helped us mitigate some of the risks associated with humanitarian AI: data and model bias, data gaps, lack of transparency and explainability, lack of accountability and the imbalanced power dynamics that dominate AI development. For example:
- In Nepal, we validated and refined the data inputs used to train the AI model – specifically removing Ethnicity as an input – due to local preferences. This resulted in an algorithm of equivalent accuracy that also took into account the concerns of stakeholders and communities on the frontline of crises.
- In Cameroon, involving Red Cross staff in generating a dataset of new rumours to test the model, helped identify a 'blindspot' in the functionality of the AI. As a result, this issue can be addressed to improve the model before deployment.
- In Nepal, collaborative problem framing enabled the Red Cross team to work on an issue that has been a concern for a number of years – feedback from communities that NFRIs were not meeting their needs. Few existing humanitarian AI tools take the specific local problems faced by frontline humanitarians as their starting point, even though this is important for their uptake.
- Participatory AI is useful for navigating the trade-offs between potential harms of AI and real-world needs. Discussions with participants in both countries helped surface the gaps between general critiques of AI, raised by research and developer communities, and the attitudes of local stakeholders. Local actors were permissive of limitations and concerns, being willing to apply the tool beyond its intended scope as long as it was 'good enough' or improved their current process. This highlights a tension between idealised responsible deployment of AI and the pragmatic reality of how these tools may be used.

The barriers and challenges

The main challenges we encountered in our project included the lack of existing good quality data (organisational data and open data), as well as the lack of technical infrastructures, skills and data literacy in local organisations that would more readily enable data-innovation and AI-development. We also found it hard to shift ways of working from traditional operating modes where local organisations are delivery partners towards more active 'co-design'. In addition to this, the novelty and complexity of the topics and technology meant it took longer for our teams to understand each other and establish effective, multi-disciplinary ways of working. Finally, we found that the currently limited methodological and digital toolbox for participatory AI, made the design of appropriate and effective activities difficult.

Recommendations

To grow collective crisis intelligence and participatory AI approaches in the humanitarian sector, R&D efforts need to focus on the following: developing AI methods that can function with sparse or low quality datasets, and digital tools that can work in low-resource settings. The humanitarian sector and funders also need to take a more coordinated approach to filling data gaps and establish guidance to help organisations properly document existing AI models and data sets. Greater investment in technical and community participation skills is needed, along with further rigorous experimentation and codification of participatory AI methods to support more widespread adoption.

Call to action

It is time to upend and localise the development of humanitarian AI. We must move away from top-down initiatives and proprietary systems that risk reinforcing colonial power dynamics, undermining the localisation agenda, and potentially causing more harm to crisis affected communities. At a minimum, humanitarian technology projects should be required to demonstrate how affected communities have been involved in the development and oversight of new tools. But our project has shown that it is possible to build AI with local infrastructure, local data, and local talent, and that it is possible to build AI that responds to local values and priorities. However, much more investment and experimentation is needed to realise a future where locally-developed and owned AI becomes 'business as usual'.

01

Localising AI for the humanitarian sector

Photo: Kunal Parmar, Unsplash

Artificial intelligence (AI) has huge potential to enhance humanitarian action. Recent years have seen multiple applications from predicting vulnerability and risk mapping to automating damage assessment and modelling long-term recovery efforts.[2] Making the most of these opportunities is particularly important in the light of escalating and increasingly complex crises.

Critiques of humanitarian AI

However, current approaches to developing AI-enabled tools for the humanitarian sector risk reinforcing existing power dynamics by, once again, placing power in the hands of international actors at the expense of local organisations and frontline communities.

Too often AI models are developed by large humanitarian institutions and Western companies as black-box systems built on biased datasets, without input or oversight by the groups affected by their use. Even when algorithms are interpretable, interrogating them requires technical expertise, which is lacking on the humanitarian front line.

Box 1: Key critiques of Humanitarian AI

1. **Bias and data gaps**
   The data used to train models is one of the main sources of bias in AI. If datasets are incomplete, unsuitable for the problem being addressed, or not representative, it can lead to unfair decisions and inaccurate model outputs. Other types of bias can emerge if models are developed by teams who have no contextual understanding.
2. **Privacy and security**
   The datasets involved in modelling problems related to humanitarian crises can be very sensitive. They might contain personal information that violate privacy and pose security risks, or haven't been collected with informed consent.
3. **Transparency and explainability**
   The predictions and recommendations made by some AI systems, particularly those that use deep learning algorithms, can be opaque to human decision makers. Even when machine learning algorithms are interpretable, this often requires a high technical expertise and time that frontline humanitarian staff lack.
4. **Hype and power**
   The hype around private sector technology can lead to overstatement of AI capabilities, transfer of unsuitable models to new contexts, or the deployment of untested approaches. As long as actors in the Global North maintain an advantage over local partners in terms of infrastructure, skills and data, there is a risk that AI will consolidate existing power structures in a way that is fundamentally at odds with the localisation agenda.
5. **Accountability**
   Establishing accountability and ensuring clear mechanisms for addressing unfair outcomes is hard when it comes to AI. It's often not clear who is responsible when a decision or analysis is discriminatory in projects that involve a large number of actors.

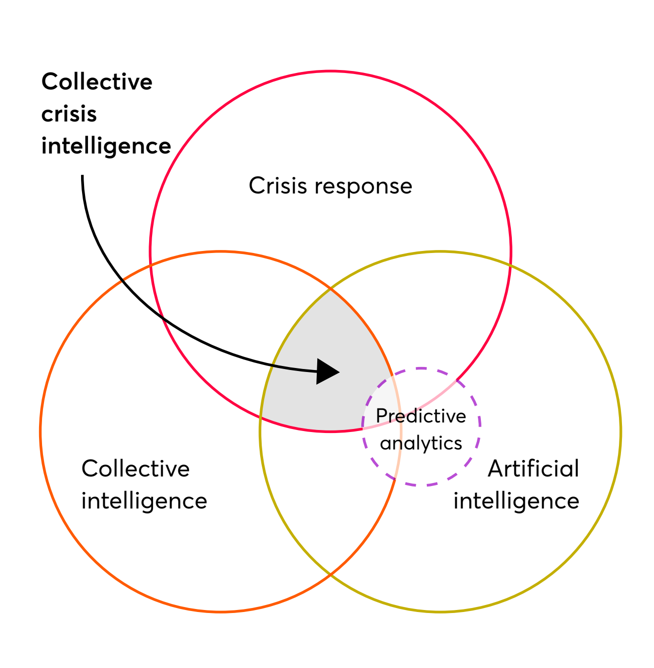
Adapted from the World Bank's Responsible AI for Disaster Risks Management[3] and Nesta's Participatory AI for Humanitarian Innovation.[4]

These critiques are just some of the ways that AI systems pose a risk to humanitarian principles and stand in opposition to the localisation agenda proposed as part of the Grand Bargain. **The localisation of aid cannot be achieved without the localisation of AI**.

Combining frontline collective intelligence and AI for crisis response

This report describes the results of an Accelerated Innovation Collaboration to research, design, and test collective crisis intelligence solutions for community-based and frontline humanitarian action.

Collective crisis intelligence (CCI) describes a new methodological approach for humanitarian action. It combines large scale data generated by and with affected communities or frontline responders with the processing power of AI for more effective crisis mitigation, response or recovery. CCI is a subset of a wider field known as collective intelligence design.[5]



Over a ten-month period, teams from the Red Cross in Cameroon and Nepal worked with teams from Nesta and Newcastle University in the UK, data science and community engagement fellows from Nepal and Cameroon, and the IFRC's Solferino Academy to prototype **two new collective crisis intelligence tools**. The tools address two important problems facing many humanitarian organisations:

1. Improving the distribution of resources after a crisis to match community needs
2. Identifying and acting on misinformation that threatens humanitarian action.

To help mitigate the potential negative impacts of AI (see Box 1), we involved frontline staff and crisis-affected communities in the development of these tools and their evaluation, using **participatory AI**.[6]

Participatory Artificial Intelligence (PAI) is an emerging field of practice that brings participatory methods into the AI development lifecycle. In its broadest sense, it refers to the involvement of a wider range of stakeholders than just technology developers in the creation of an AI system, model, tool or application.

Collective crisis intelligence in combination with participatory AI could help disrupt the current trajectory of AI development to create tools that are grounded in localisation and aligned with humanitarian principles. But, because of their novelty, there are few examples about how to apply these methods in practice and little evidence about their potential impact. This project is the first to combine collective crisis intelligence with participatory AI to generate insights about the added value and challenges of these approaches in the real world.

The overarching aims of this work were to:

1. Design and develop two novel proof-of-concept collective crisis intelligence tools.
2. Evaluate the effectiveness of these tools and the difference they could make to humanitarian operations.
3. Test whether collective crisis intelligence and participatory AI can address some of the concerns around humanitarian AI.
4. Test whether participatory AI can lead to tools that are trusted by frontline users and crisis-affected communities.

Our project took place during a global pandemic, which demanded the attention of our frontline colleagues, shifted our project team interactions online, and prevented us from implementing all the activities we had originally planned. Despite this, we made substantive progress on all four aims.

Throughout, we were mindful of the unequal power dynamics that often dominate technology development. Our aim was to minimise these as much as possible. By working with local partners and frontline humanitarians from concept development to evaluation, recruiting local data science fellows, and openly publishing our code, results, and resources we have tried to demonstrate the viability of a more collaborative, alternative pathway for AI development.

This report provides a summary of what we did, how we did it, and our key findings. It is published alongside two technical methodological reports and other resources; see the Appendix for an overview.

This project was funded by a grant from the UK Humanitarian Innovation Hub (UKHIH). UKHIH is funded by the UK's Foreign, Commonwealth and Development Office (FCDO) and hosted by Elrha – a global humanitarian organisation and the UK's leading independent supporter of humanitarian innovation and research.

02

Nepal: using collective crisis intelligence to predict needs before a crisis strikes

Photo: Alex Gurung, Unsplash

The CCI tool: NFRI-Predict

NFRI-Predict is a new tool that predicts which non-food items are most needed by communities after a crisis. The tool is an early stage prototype that focuses on making predictions about the 'essentialness' of items needed by communities after a flood.

The tool was designed in response to feedback from crisis-affected communities that non-food related aid items (NFRIs) provided by the Red Cross after crises do not always adequately meet their needs.

HOW TO USE THE TOOL

To make a prediction, Red Cross staff enter information about single or multiple households using a simple web-based interface. The tool uses information about household location, demographics and vulnerability indicators such as economic and health status. Users then choose between different visualisations to compare the items needed by different households. The tool allows users to save predictions for future sessions or print their outputs to share the results with colleagues and inform decision making.

Figure 1: A simple overview of the user flow for NFRI-Predict tool

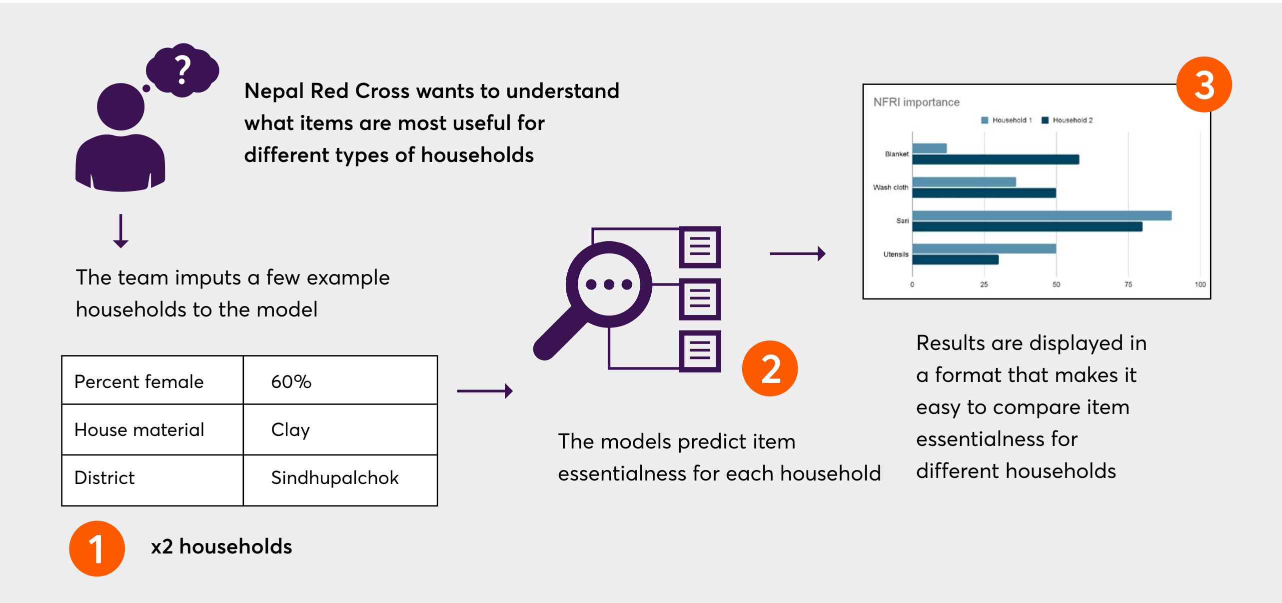
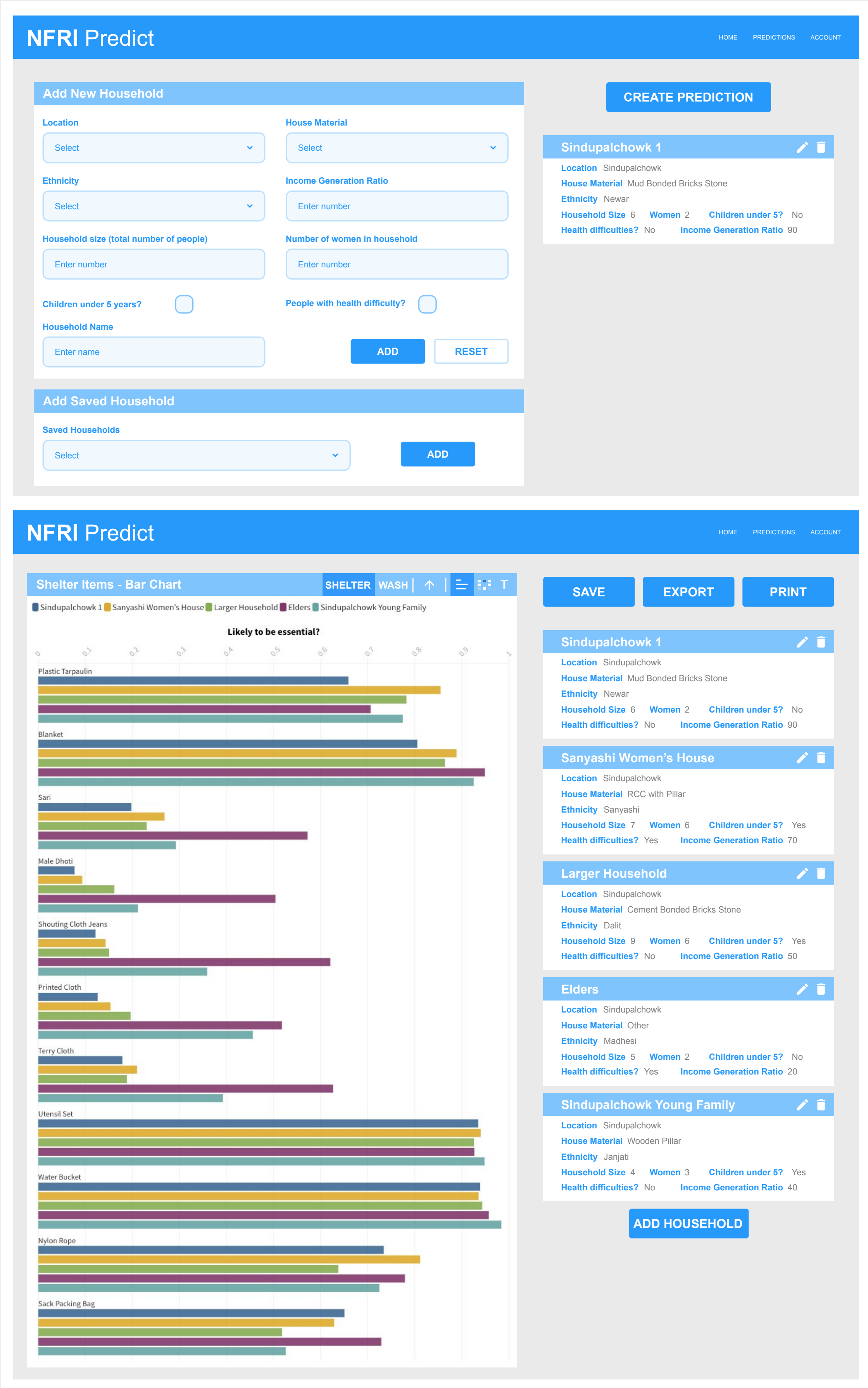


Figure 2: The prototype web interface for NFRI-Predict



Top: shows the input interface. Bottom: shows one of the possible visualisation screens users can choose as an output. The prototype can be downloaded from GitHub.

Items are typically distributed within two standard packages: Shelter (covering housing and clothing items); and Wash (covering health and sanitary items). We created separate classification models for each of these packages.

USE CASES FOR NFRI-PREDICT

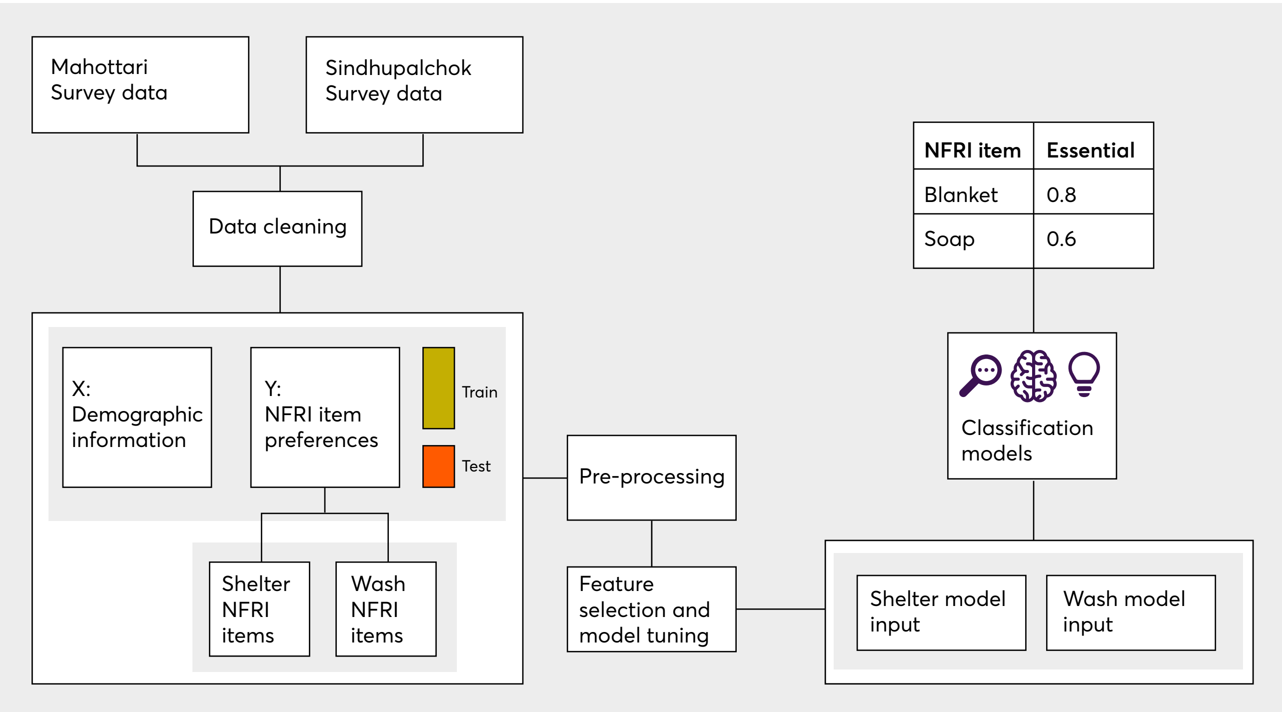
1. **Predicting resource needs in advance of a crisis** can help humanitarian organisations, governments and citizens to prepare better in advance of a crisis by stockpiling necessary supplies in strategic locations.
2. It can also be used to help with **identifying the resources needed by the most vulnerable** in the immediate aftermath of a crisis, before a detailed damage assessment is carried out.
3. If resources are constrained, the tool can be used to **better understand differential needs** to prioritise the allocation of limited resources.

The **primary end users** of the tool are Red Cross team members involved at different stages of disaster planning and response.[7] This includes the central and district level Nepal Red Cross teams involved in rapid assessment, NFRI stockpiling and planning, and local coordination for NFRI distribution after a crisis.

HOW THE TOOL WORKS

The tool combines intelligence from crisis-affected communities and the processing power of AI. It uses a classification algorithm to make predictions about which aid items would be most essential for different types of households in different regions.[8] This model is trained on a dataset of community views, where households were asked to rate which NFRI items[9] are essential after a flood. This new dataset was collected during the project by Red Cross volunteers from over 3,000 households[10] in Sindhupalchok and Mahottari, two geographically distinct regions in Nepal. The survey was the first ever large scale dataset about the NFRI preferences of communities collected by the Nepal Red Cross Society.

Figure 3: Overview of the technical methodology used to develop the classification model(s)



Designing NFRI-Predict using participatory AI

Throughout the tool development process, we designed and delivered a range of participatory AI activities with the following aims:

- Increase the usefulness of the tool for frontline users
- Create high-quality, representative and machine-readable data that is appropriate for the problem being addressed
- Optimise the AI model's performance for diverse values and preferences
- Create appropriate levels of trust and understanding of the tool by all stakeholders
- Ensure all stakeholders have a say in evaluating the performance of the tool.

We identify four potential levels of activities for participatory AI. These draw on the types and levels of participation previously identified by researchers in the citizen science and other participatory research communities, namely consultation, contribution, collaboration and co-creation (or co-design). As participation was included at multiple touchpoints along the AI pipeline, and included activities where the technical team interacted with participants, we categorise the level of participation as **collaboration**.

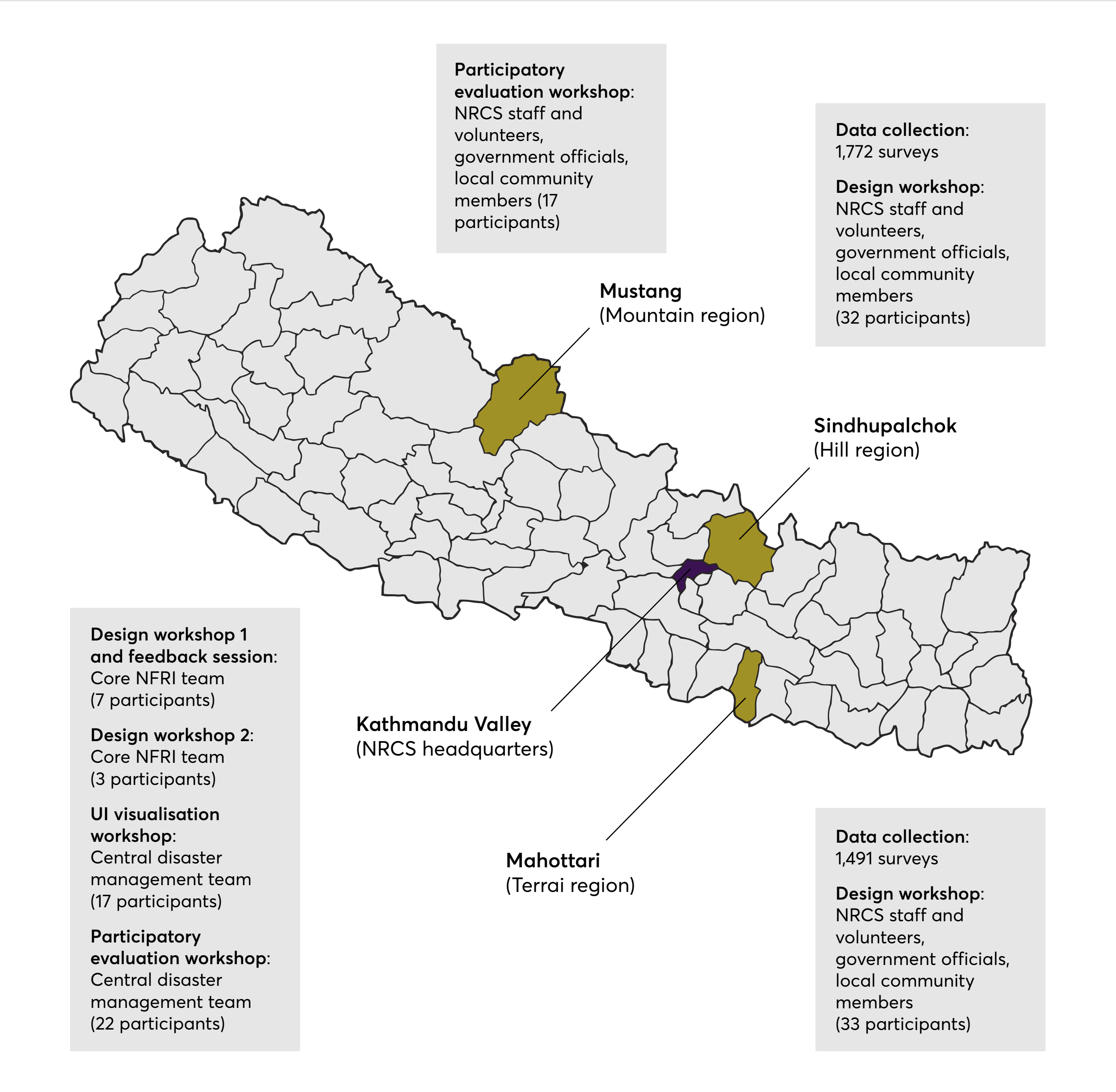
We used a combination of in-person and online workshops in four different locations (see Figure 4). **Four key stakeholder groups** were involved in these activities:[11]

1. Central Nepal Red Cross Society staff, responsible for overall decision making and planning about disaster management and the NFRI process.
2. District level Nepal Red Cross Society staff and volunteers, responsible for local coordination and distribution of NFRI in the aftermath of natural disasters.
3. Local officials including representatives from government, the police and army, involved in local coordination and response in the aftermath of natural disasters.
4. Local community members both with and without previous experience of receiving NFRIs.

Collaborative problem framing and initial tool design took place in August and September 2021. All other tool development and participatory AI activities happened between January-April 2022.

Photo: Saurav Poudel

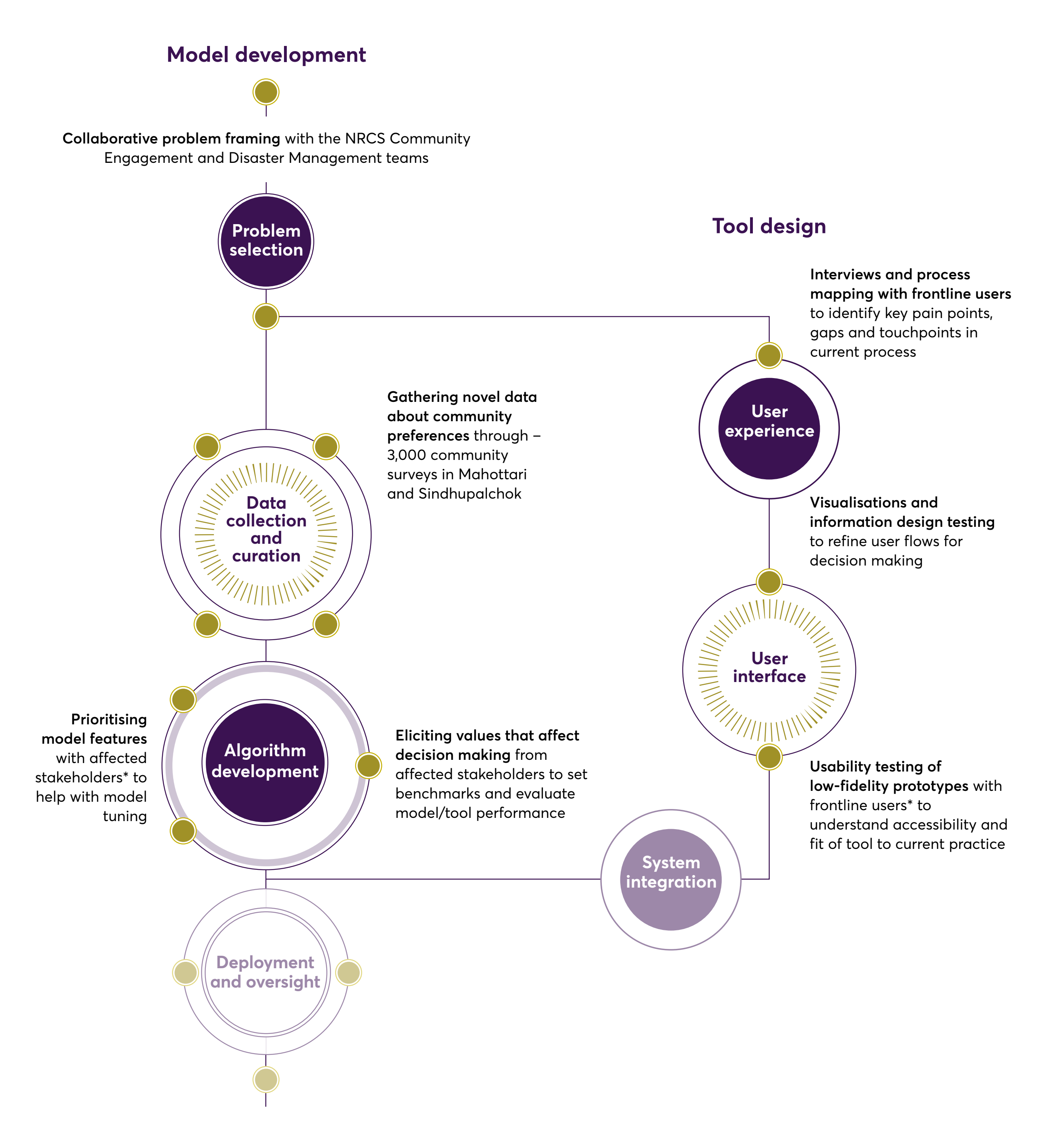
Figure 4: Overview of all participatory AI and collective intelligence (data collection) activities by location and participant groups



UI = user interface. NRCS = Nepal Red Cross Society.

Figure 5 shows an adapted version of Nesta's participatory AI framework with the participatory activities that we tried at different stages of model development and tool design highlighted in bold. Participants from Kathmandu and district workshops helped us prioritise data inputs for training the model, identify relevant criteria for evaluating the tool and tested the usability of interfaces. Workshops were facilitated by the Red Cross community engagement team, a data science fellow, and researchers from Nesta and Newcastle University.

Figure 5: An overview of the participatory AI interventions throughout model development and tool design



Adapted from the framework published in Participatory AI for humanitarian innovation: a briefing paper. *stakeholders included Red Cross staff, local government and officials and individuals from crisis-affected households. System Integration and Deployment and Oversight appear lighter as they were out of scope for the project.

Figure 6: Group work during evaluation workshops

Left: Group feedback about the UI prototype for the NFRI-Predict tool.
Right: A group of local residents in Mustang discuss the potential impacts of the tool.

NFRI-Predict Results

In this section, we provide an overview of how well NFRI-Predict performed in both technical and participatory evaluations, as well as describing to what extent we were able to address the critiques of AI through participatory AI interventions. We conclude with results about the perceived trustworthiness and usefulness of the tool as judged by different stakeholders.

HOW WELL DOES THE TOOL PERFORM ON A TECHNICAL EVALUATION?

Our technical evaluation (Table 1) consisted of measuring the performance of our models on a test dataset. We also tested for model bias to ensure that any limitations of the model were well documented for future users.

Table 1: Key results from the technical evaluation

Key results from the technical evaluation	
Technical performance of the model	**Model performance on test dataset measured using F1 score**[12] • Wash (health and sanitary) items: 0.91 • Shelter (housing and clothing) items: 0.86 **A bias audit**[13] helped us identify a reduced performance of the model for predicting the clothing preferences of two ethnic groups, predominantly found in the Sindhupalchok district.

Complete results from evaluation activities can be found in the detailed Technical Report.

Although the accuracy for our model might appear high, our technical evaluation helped us to surface the limitations of our modelling approach. By comparing different performance measures, we discovered that while the model is very capable at predicting when an item is deemed essential, it struggled to identify when items were not essential. The most common error made by the model was to predict non-essential items as falling into the essential category.[14] We recommend exploring other approaches to modelling to address this limitation and the differences in performance surfaced by the bias audit before the tool is operationalised (see What next?).

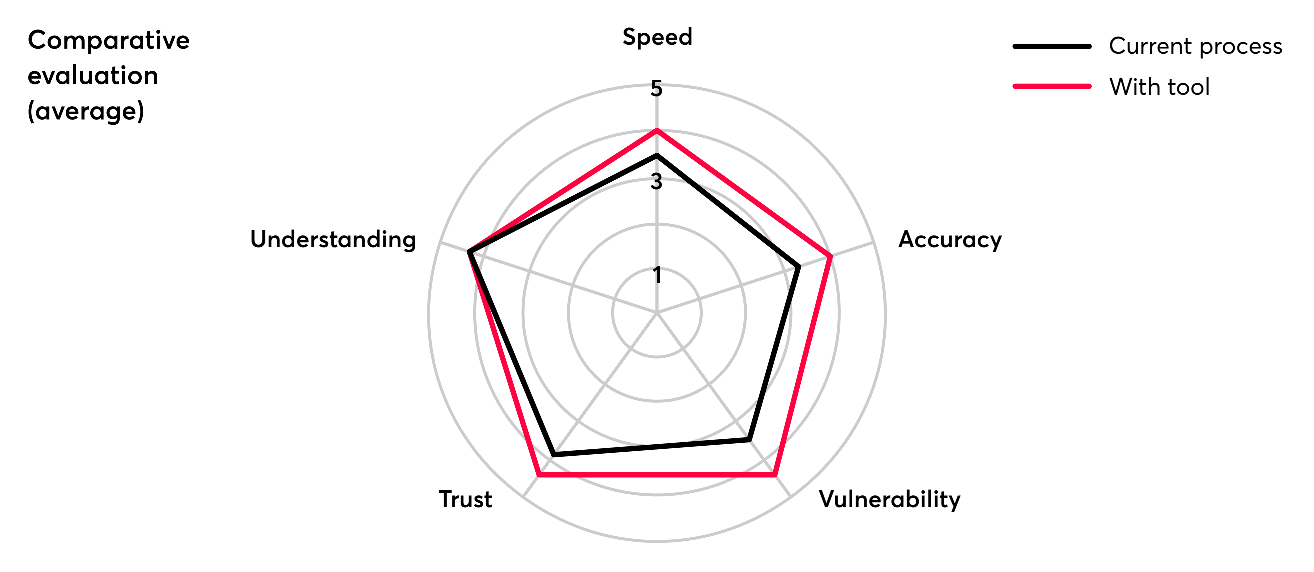
HOW WELL DOES THE TOOL PERFORM IN A PARTICIPATORY STAKEHOLDER EVALUATION?

As part of our participatory AI activities, we involved Nepal Red Cross Society staff in a comparative evaluation. Groups were made up of two to three individuals, with eight groups of 22 participants in total. They reviewed a range of hypothetical scenarios in which they imagined using the tool in comparison to current operational practice and assigned ratings on a scale of 1-5 across five criteria: speed, accuracy, vulnerability, trust and understanding. Table 2 provides a summary of the key results and Figure 7 shows the average rating across the criteria for all groups. Participants thought NFRI-Predict has the potential to increase the speed and accuracy of localised humanitarian response, as well as improving the ability to serve the most vulnerable.

Table 2: A summary of the main results from the comparative evaluation

Key results from comparative evaluation	
Accuracy and speed of crisis response, and ability to meet the needs of the most vulnerable	**Central disaster management team (Kathmandu)** • Six out of eight groups of Nepal Red Cross Society staff gave a higher rating to the tool than the current process for **accuracy** (the ability to match the needs of communities) and **vulnerability** (the ability to meet the needs of the most vulnerable). • Five out of eight groups thought the tool would make the process of NFRI distribution **faster**. **District-level workshop participants (Mustang)** • Red Cross staff at the district level gave a higher rating across criteria of speed, accuracy and vulnerability with the tool in comparison to the current process during evaluation activities. But they were only prepared to use the model if it was retrained with data gathered from their own district.
Phase of crisis response	• Participants identified the highest utility for the tool during the Preparedness phase of crisis management, where they rated the tool more highly than the current process across all criteria.[15]

Figure 7: Comparative evaluation between NFRI-Predict (red) and the current process (black) by Red Cross staff in Kathmandu



Radar chart shows average scores of eight groups (22 participants). Participants were asked to rate their level of agreement with statements (e.g. NFRIs are delivered quickly). 1 = strongly disagree, 3= neutral, 5 = strongly agree.

Figure 8: Red Cross Staff completing a comparative evaluation activity

Photo taken during a participatory evaluation workshop in the Mustang district.

HOW DID THE PARTICIPATORY AI ACTIVITIES INFLUENCE THE TOOL AND HELP US ADDRESS SOME COMMON CRITIQUES OF AI?

Table 3 summarises how participatory AI activities influenced the design and development of the tool as well as demonstrating how participation can start to address some of the main critiques of humanitarian AI.

Table 3: Overview of aims and outcomes of participatory AI activities and their link to some of the main critiques of AI systems

Aims of activity, and main critique being addressed*	Key results from PAI activities
Increase the usefulness of the tool for frontline users *AI critique: Hype and power*	**Collaborative problem framing** with the Nepal Red Cross teams enabled them to develop a tool to address an issue that has been a concern for a number of years – that NFRI packages did not meet community needs. The Nepal Red Cross team had not previously had the mandate or opportunity to address this issue. **Information design, visualisation and usability testing of low fidelity prototypes** with Nepal Red Cross staff influenced both the design of the user-interfaces and the visualisation options. For example, we created visualisations that allow comparisons between households based on user feedback.
Create high-quality, representative and machine-readable data that is appropriate for the problem being addressed *AI critique: Bias and data gaps*	By **generating a novel dataset** about community preferences with representative samples from two of the districts most commonly affected by floods, we were able to model a different type of problem than typically addressed by AI tools in the humanitarian sector.
Optimise the AI model's performance for diverse values and preferences *AI critique: Transparency and Explainability*	As a result of **prioritising model features** with frontline users and crisis-affected communities, we excluded ethnicity as an input feature. Notably, this didn't affect the accuracy of the model and helped us realise the limitations of automated feature selection which had led to over-selection of data inputs for the model.
Create appropriate levels of trust and understanding of the tool by all stakeholders *AI critique: Transparency and Explainability*	Knowing that we would bring together data scientists and stakeholder groups with low levels of technology literacy to interrogate how the model works prompted us to choose a more interpretable model that would be easier to explain. Discussing different scenarios for the tool's use with participants during workshops helped them understand the limits of its functionality and how to interpret the outputs it produced more critically.
Ensure all stakeholders have a say in evaluating the performance of the tool *AI critique: Accountability*	**Eliciting values for decision making** to create evaluation criteria and asking stakeholders to choose between different versions of the model based on their preferences for how it should work meant we were introducing new mechanisms of accountability.

*Relevance of activities to specific AI critiques are the result of post-hoc analysis.

Below are two vignettes which illustrate how the participatory AI activities helped to optimise the AI model to take into account community values and preferences, and to determine the criteria by which the tool was evaluated.

Box 2: Prioritising model features with workshop participants

What we did

We asked Red Cross staff, government officials, and community members in Sindhupalchok and Mahottari to identify the three data inputs they thought were most and least important for predicting NFRI needs. We used two different prioritisation methods: group discussions with voting, and pairwise ranking using the AllOurIdeas platform.

What we found

During district workshops, 66% of all participants ranked ‘Ethnicity’ as one of the least important data inputs using group prioritisation. This was a key point of consensus, and a range of justifications came out during group discussions. Some participants felt that ethnicity would not be useful for making predictions, whereas others felt it shouldn’t be used for ethical reasons.

The impact on model development.

Following the workshops, we trialled the model with and without ethnicity as an input feature and discovered there was a minimal impact on model accuracy (both models had an accuracy of ~81%).[16] This helped us realise the limitations of automated feature selection which had led to over-selection of data inputs for the model. The final version of the model does not use Ethnicity as an input feature.

Figure 9: Red Cross staff and volunteers in Sindhupalchok design workshops

Box 3: Identifying key areas of consensus and difference between stakeholders

What we did

We used focus group discussions and ranking to identify key areas of consensus amongst stakeholders about which values should drive decision-making about NFRIs. These activities were carried out with Red Cross staff, government officials, and community members who previously received NFRIs in Sindhupalchok and Mahottari.

What we found

All stakeholders agreed that it was most important to respond as quickly as possible (**speed**) and address the needs of the most vulnerable (**vulnerability**). Red Cross staff and governmental officials also highlighted **accuracy** as an important aim.

The impact on model development

We used these three priority values as criteria in the comparative evaluation of the tool’s performance versus the current process. This approach helped to provide a more nuanced understanding of the model’s performance using criteria that matter most to local users and affected stakeholders.

Figure 10: Community participants in Mahottari design workshops

HOW TRUSTWORTHY IS THE TOOL AND SHOULD IT BE OPERATIONALISED?

"As beneficiaries and all stakeholders are involved, it gives the true and correct information so we trust it."

Nepal Red Cross staff member (Mustang district)

The trustworthiness of AI systems is a key implementation barrier. We asked participants in evaluation workshops to compare their trust in the current process versus a hypothetical scenario where they could use the tool. We also measured participants' general attitudes towards the idea of computers making decisions using post-workshop questionnaires, as well as specific positive or negative associations towards the tool and its future development. Table 3 provides an overview of the results.

Table 4: Overview of results related to trust and operationalising the NFRI-Predict tool

Overview	Results
Trust	**Tool specific** **Central disaster management team (Kathmandu)** • Seven out of eight evaluation groups gave a higher or equal trust rating to using the tool in comparison to the current process. • 73% of staff felt positive[17] about the tool overall but this fell to 36% at the prospect of making decisions using **only** the outputs of the tool. **District-level workshop participants (Mustang)** • 88% of participants felt positive about the tool overall. • District level Red Cross staff attributed their trust in the tool to knowing that it was developed with the involvement (and data) of communities. **General AI** • 68% and 94% of participants in Kathmandu and Mustang, respectively, thought that we should trust computers to make decisions about NFRIs.[18]
Uptake and usefulness	• 90% of the evaluation participants[19] agreed that it was worth developing NFRI-Predict to full operational capacity. • Five participants (26%) in the Kathmandu workshops reported that they would make changes to their current NFRI process based on their experience during the evaluation activities. The remaining participants didn't know (47%) or didn't plan to make changes (26%).

"...it is about the voice and involvement of local people and technology. The involvement of local people helps to determine the local needs, so the priority must be provided to the ones [tools] which are made by local level consultation and using local technology."

Nepal Red Cross staff member (Mustang district)

What next?

The short timeframe of the project meant we were not able to deploy or test the tools in operational settings. To reach this stage, further work to improve the model could focus on: a) balancing the dataset[20] or; b) collecting new data about community preferences using an alternative methodology. This could include ranking items using a scale to make it easier for the model to distinguish between items' degrees of 'essentialness' and improve it's ability to predict differential needs.

Further priorities would be to expand data collection activities to cover more districts,[21] to scope novel data collection methods that could support regular updates about preferences, and to broaden the scope of the tool to other types of crisis beyond floods. These steps would help increase the tool's utility.

Alongside this, further work is needed to ensure that the tool can be integrated into the Nepal Red Cross Society information management infrastructure[22] and NFRI operational processes. At present, the Nepal Red Cross does not have permanent staff with the technical capabilities required to develop or maintain AI models, or the digital systems infrastructure to do so. This is a major barrier to the application of CCI in local humanitarianism. In the immediate term, however, the outputs from this project will feed into the Nepal Red Cross Society's revision of its standard NFRI packages and manual, due to be approved in Autumn 2022.

WIDER RELEVANCE FOR THE HUMANITARIAN SECTOR

Examples of predictive analytics in the humanitarian sector have historically tended to focus on predicting where a crisis will happen, who will be affected, how big the impact will be, and when it will strike. Our previous research showed **a gap and opportunity for CCI tools to help organisations get smarter at predicting the resources needed** to mitigate or respond to a crisis before it occurs.[23] NFRI-Predict is one example of how data about the needs and preferences of crisis-affected communities can be used to fill this gap and build a more timely and relevant response.

The Nepal Red Cross also identified potential future uses for this tool, including exploring how to use it for more accurate estimations of the value of cash transfers. They are also considering other applications for collective crisis intelligence approaches to help them develop more effective operating models, for example predicting volunteer deployment needs depending on crisis type and demands, or to predict blood donation requirements to better manage supply and demand.

Our experience has shown that factors such as geographic location can have a large effect on NFRI preferences, indicating their importance for determining nuanced and contextual humanitarian responses.[24] We have also seen that predictive models can underperform when faced with minority cases, which does not guarantee acceptable, equitable outcomes in crisis situations. Ensuring the representation of minority populations could be achieved by combining local knowledge with quantitative oversampling strategies. Other information, such as data on the physical geography of a location, may also enable better performance across regions and demographics.

This highlights the need for greater exploration of the most appropriate modelling strategy in the development of CCI tools. A deeper analysis would help determine whether large country-wide models are outperformed by more localised models.

03

Cameroon: using collective crisis intelligence to track and respond to misinformation in real-time

Photo: Edouard Tamba, Unsplash

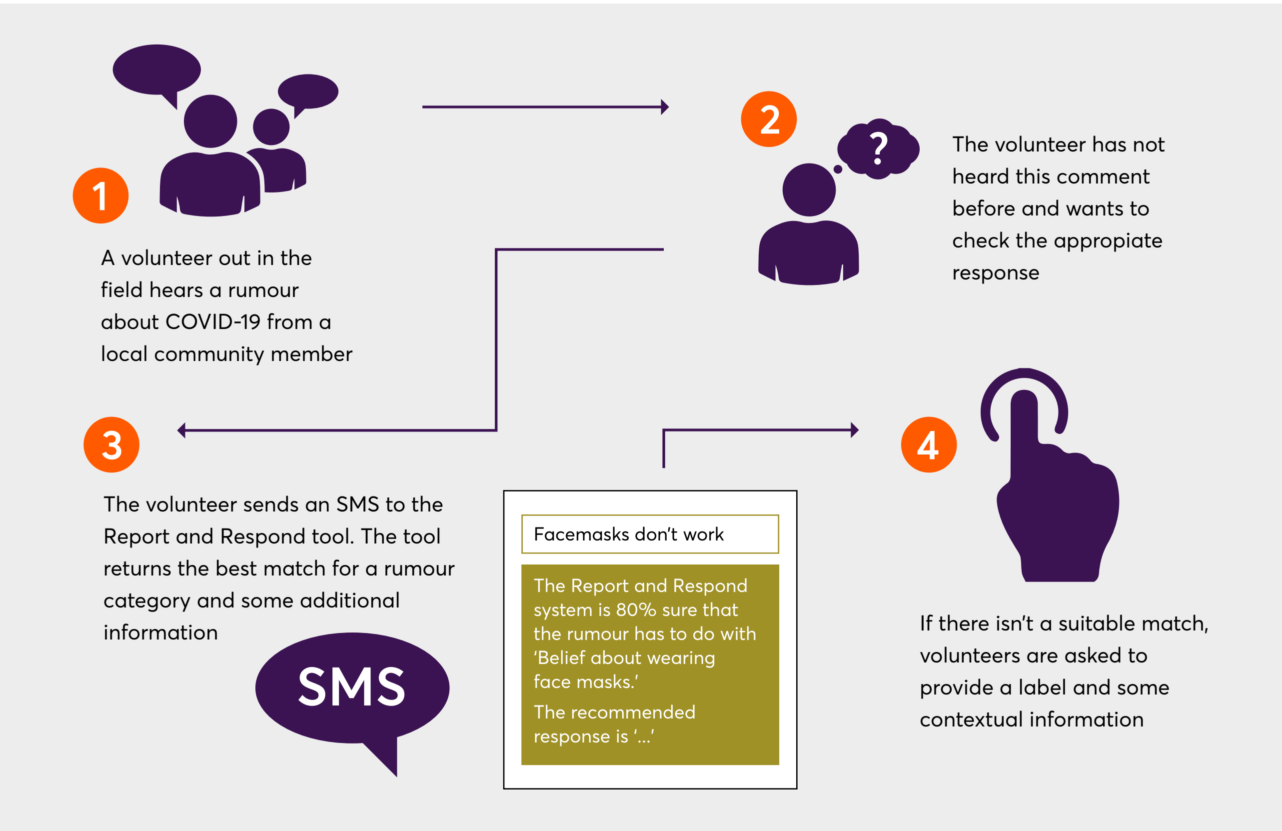
The CCI tool: Report and Respond

Report and Respond is a French language SMS-based tool that allows Red Cross volunteers to check the accuracy of observations they hear from the community while they're in the field, and respond appropriately. The tool is an early-stage prototype that focuses on misinformation about COVID-19.

HOW TO USE THE TOOL

Volunteers send an SMS to check the accuracy of any COVID-19 related community feedback. The tool checks the incoming message against a database of rumour-response pairs, and sends the volunteer an appropriate response in real time (Figure 11). When rumours don't have a match, volunteers are asked to provide a short description and contextual information.

Figure 11: A simple overview of the user flow for rumour classification using Report and Respond



Rumours without a match are split into groups based on similarity. These are regularly reviewed by Red Cross staff and/or volunteers to identify emerging rumour categories that can be added as a label (Figure 12).

Figure 12: A simple overview of the user flow for clustering emerging rumours using Report and Respond

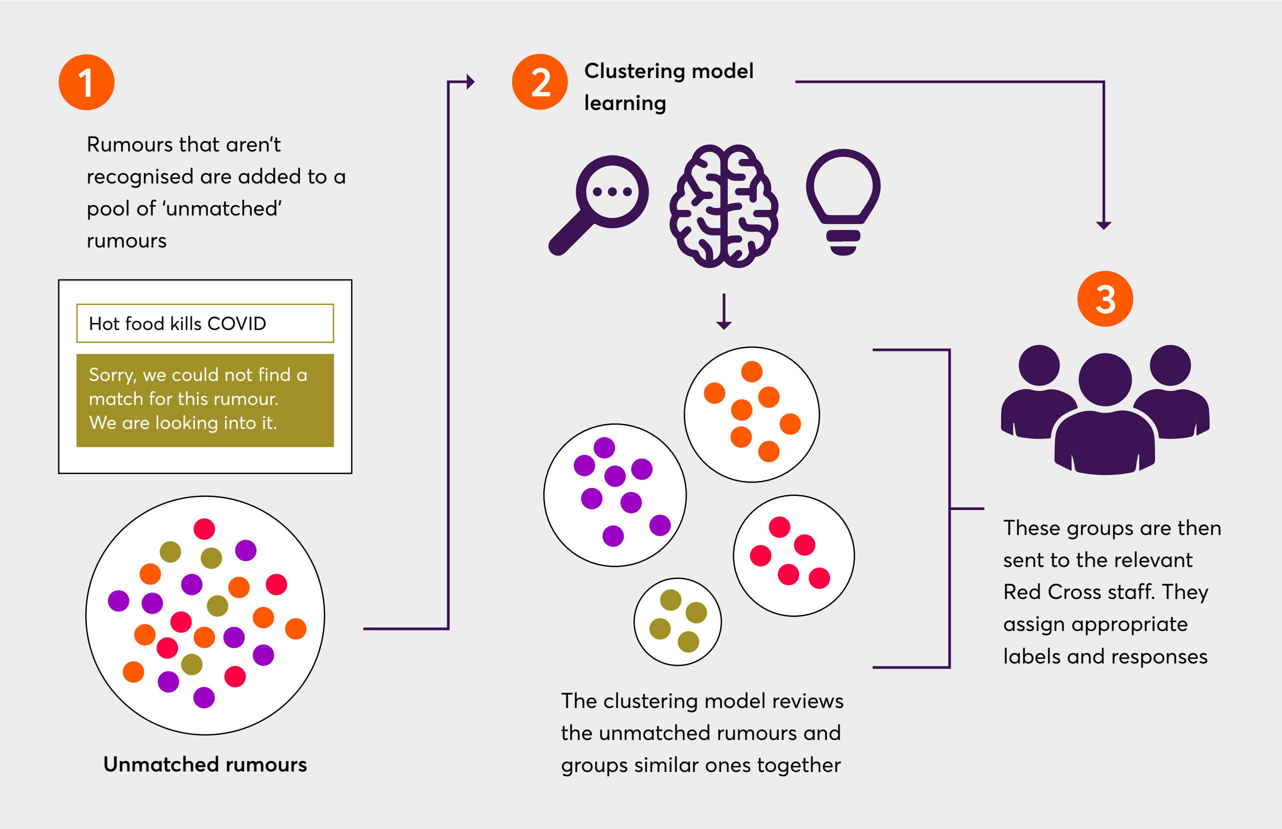
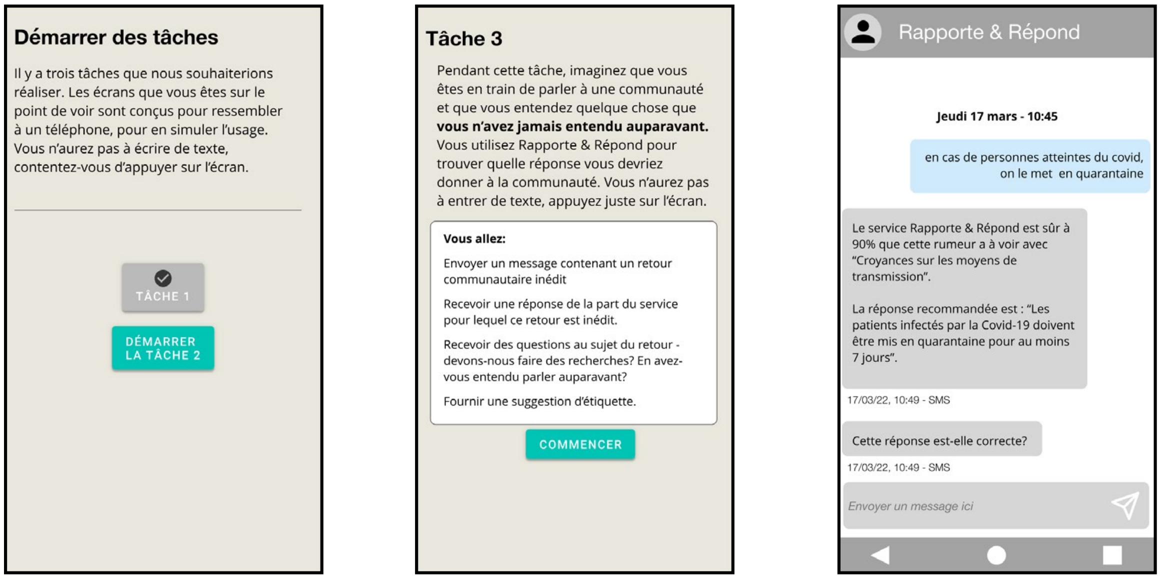


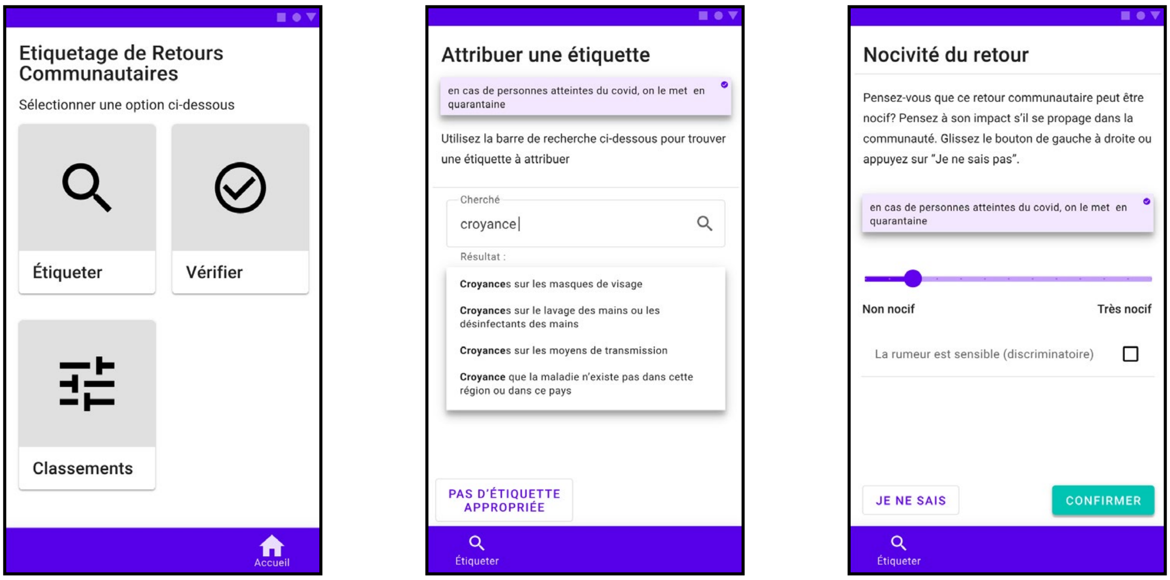
Photo: Cathy Essouma

Figure 13: Screenshots showing the Report and Respond SMS-based prototype for use by Red Cross volunteers in the field



Alongside the SMS-based interface, volunteers and Red Cross staff can access a **standalone labelling interface** (see Figure 14). This application-based interface is used to assign new labels[25] and/or verify unmatched rumours, or to monitor basic statistics about COVID-19 related misinformation.

Figure 14: Screenshots showing the prototype interface



USE CASES FOR REPORT AND RESPOND

1. **Providing reliable and timely information about rumours** can save lives when it comes to health-related misinformation.[26] Report and Respond streamlines this process by providing an immediate response to guide interactions with community members..
2. New categories of rumours or conspiracy theories can emerge at any time during a crisis.[27] The outputs from the clustering model and the Labelling tool can be used for **keeping track of emerging rumours and monitoring changes** in the patterns of misinformation over time.

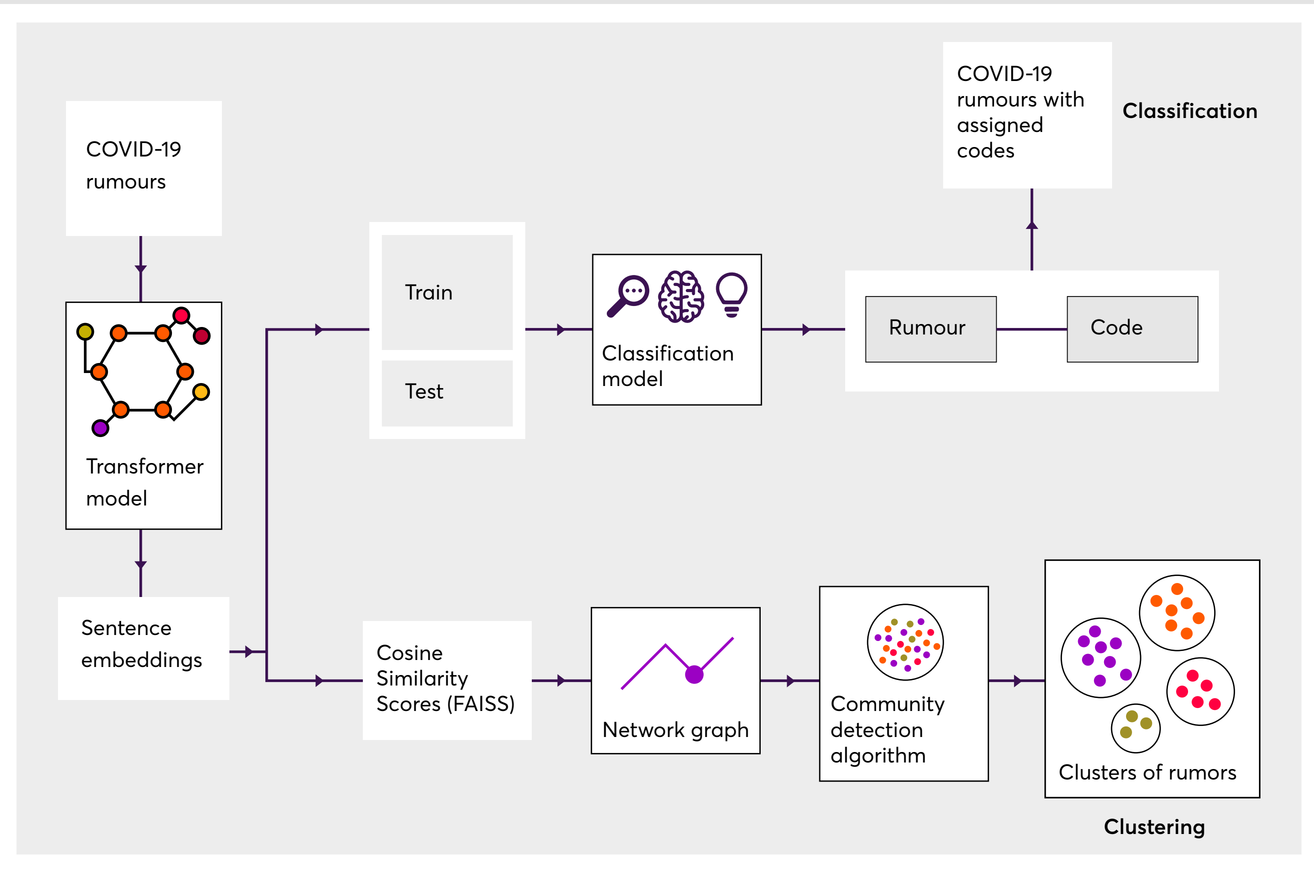
The **intended end users** of the tool are Red Cross volunteers and the Community engagement team of the Cameroon Red Cross. Together, these two groups are responsible for developing key messages about COVID-19 and gathering community feedback in the field, placing them at the frontline of combatting misinformation.

HOW THE TOOL WORKS

This tool is designed to make use of existing and new community intelligence using two different AI models.

The tool uses two different AI models to process incoming rumours. Rumours that fit the existing categories in the database are processed by a classification model created by reusing an existing Red Cross dataset of more than 6000 COVID-19 related rumours, beliefs and observations collected in Cameroon between 2020-21. Unrecognised rumours are processed by a clustering model which groups similar rumours together ready for analysis and labelling by Red Cross staff.

Figure 15: Overview of the technical methodology used to develop the two models



Designing Report and Respond using participatory AI

Throughout the tool development process we designed and delivered participatory AI activities that sought to:

- Increase the usefulness of the tool for frontline users.
- Create high-quality, representative and machine-readable data to reduce the risks of AI model bias.
- Optimise the AI model's performance.
- Create appropriate levels of trust and understanding of the tool by all stakeholders.
- Ensure all stakeholders have a say in evaluating the performance of the tool.

As participation was included at multiple touchpoints along the AI pipeline, and included activities where the technical team interacted with participants, we categorise the level of participation as **collaboration**.

We held in-person workshops in three different locations (see Figure 16 for an overview) involving **two key stakeholder groups**:

1. Red Cross volunteers from two French speaking regions in Cameroon.
2. Red Cross staff working in community engagement and communication.

Collaborative problem framing and initial tool design took place in August and September 2021. All other tool development and participatory AI activities happened between January-April 2022.

Figure 16: Overview of all participatory AI activities by location and participant groups

Tibati and Kribi (in person)

Design workshop:
CRC volunteers
(30 participants)

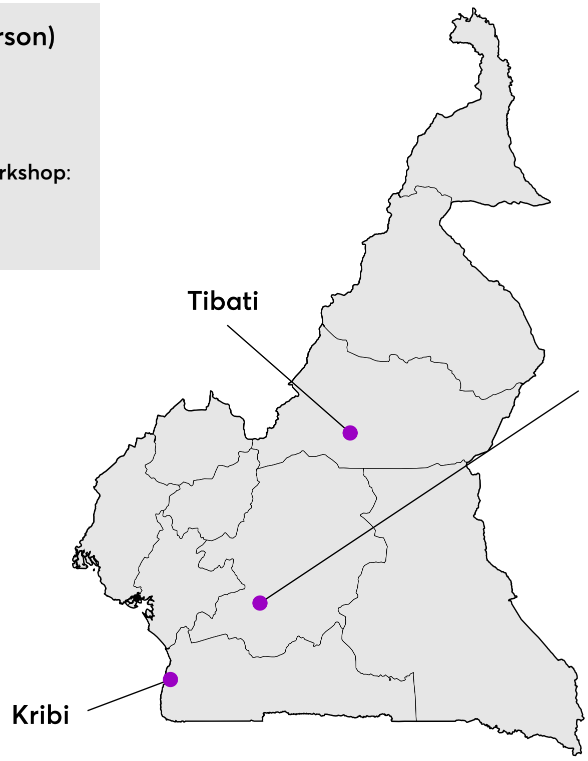
Participatory evaluation workshop:
CRC volunteers
(24 participants)

Yaounde (in person)

Data labelling workshop:
CRC staff
(5 participants)

Classification workshop:
CRC staff
(5 participants)

Responses workshop:
CRC staff
(5 participants)



Remote

Data labelling workshop:
CRC volunteers
(11 participants)

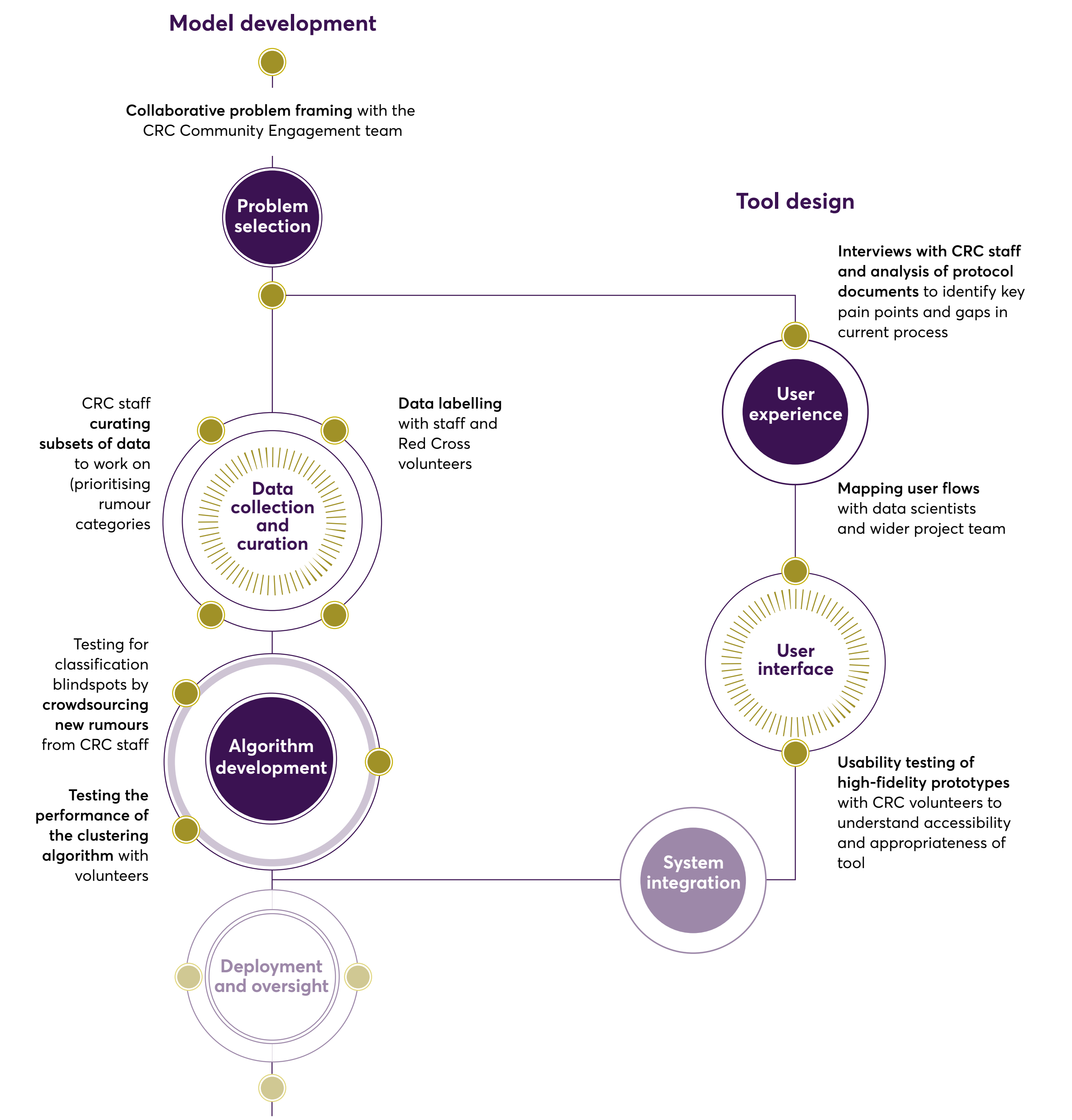
Yaounde
CRC = Cameroon Red Cross headquarters

Participant numbers for workshops with volunteers are shown as the combined total of Kribi and Tibati. Workshops were facilitated by Red Cross staff, a community engagement fellow, a data science fellow, and/or researchers from Nesta.

We identified suitable intervention points during model development and the tool design using an updated version of Nesta's Participatory AI framework. See Figure 17 for an overview of activities we carried out.

Firstly, we worked with Cameroon Red Cross (CRC) staff to curate our dataset, verify labels for the dataset, create new rumours to test the classification algorithm, and create responses for existing rumour categories. With volunteers, we focussed on testing the usability of the interfaces and the reporting process, testing the clustering algorithm, and discussing potential impacts of the tool, including issues related to trust and uptake.

Figure 17: An overview of the participatory AI interventions throughout model development and tool design



System Integration and Deployment and Oversight appear lighter as they were out of scope for the project. CRC = Cameroon Red Cross. Adapted from the framework published in Participatory AI for humanitarian innovation: a briefing paper.

Figure 18: Volunteers in design workshops in Kribi and Tibati

Left: Individual testing of user-interface prototypes for the Report and Respond tool. Right: Group work.

Report and Respond Results

In this section, we provide an overview of how well the Report and Respond tool performed in both technical and participatory evaluations, as well as describing to what extent we were able to address the critiques of AI through participatory activities. We also cover the perceived trustworthiness and usefulness of the tool for frontline stakeholders.

HOW WELL DOES THE TOOL WORK ON A TECHNICAL EVALUATION?

Our technical evaluation consisted of measuring the performance of our models using multiple test datasets. We also tested for model bias to ensure that any limitations of the model were well documented for future users.

Table 5: Key results from the technical evaluation activities

Key results from the technical evaluation	
Technical performance of the model	**Classification model performance measured using F1 score:**[28] • The test dataset from Cameroon: 0.86 • The test dataset from Democratic Republic of Congo: 0.74 • The test dataset generated through crowdsourcing with Red Cross staff: 0.75 **Clustering model performance using:**[29] • The test dataset from Cameroon: 92% A **bias audit**[30] helped us identify a reduced performance of the model for classifying feedback from men across certain categories of rumour.

Complete results from evaluation activities can be found in the detailed Technical Report.

Photo: Edouard Tamba, Unsplash

Although the overall accuracy of the model on the test dataset of rumours from Cameroon appears high, we observed variance across different rumour categories. For example, the model has an accuracy of 91.7% for the rumour category 'Croyance que certaines personnes/institutions gagnent de l'argent à cause de la maladie' ('Belief that some people or institutions are making money from the disease') and only 58% for the category 'Croyance que la maladie existe ou est réelle' ('Belief that the disease exists or is real').

We also tested the model's performance using two additional datasets:

1. A comparable community feedback dataset from the Democratic Republic of Congo (DRC), from the IFRC Go platform.[31]
2. A new rumour dataset created by the Cameroon Red Cross staff.

Although the model shows a similar decrease in accuracy for both datasets, the performance is still relatively high. This suggests that the model can be generalised to other French language countries and contexts in the future. The reduction in accuracy we observed for the test dataset generated by the Red Cross staff[32] highlighted a limitation of our model – that it struggled to assign multiple labels when rumours spanned more than one category.[33] Through discussions of this result with Cameroon Red Cross staff, we identified that the ability to assign multiple labels would be important for future iterations of the model.

HOW WELL DOES THE TOOL PERFORM IN A PARTICIPATORY STAKEHOLDER EVALUATION?

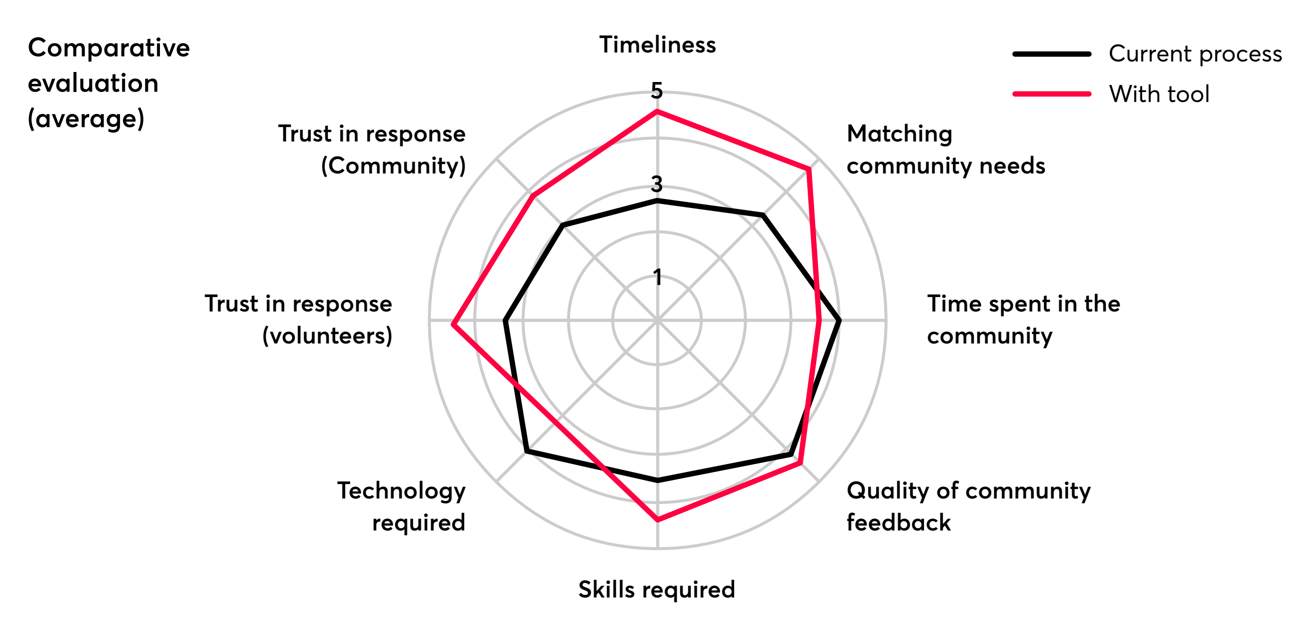
As part of our participatory AI activities, we involved Red Cross volunteers in a comparative evaluation. For this, we asked them to rate a hypothetical scenario of using the tool during their fieldwork in comparison to their usual process for collecting community feedback. The evaluation criteria were inspired in part by the recommendations of the OECD DAC Network,[34] adapted to make them more accessible for the Red Cross volunteers.

The results show that participants thought that the largest improvements would be for the timeliness of the response and the ability to respond to community needs. They anticipated that the tool would give more reliable responses to rumours, and that the benefits of this would be felt by both community members and volunteers. Table 6 provides a summary of the key results. Figure 19 shows the average result for six groups (17 participants in groups of two to four people) made up of Cameroon Red Cross volunteers. Volunteers thought the Report and Respond tool has the potential to increase the speed of localised humanitarian response, as well as improving the ability to meet the needs of communities.

Table 6: A summary of the main results from the comparative evaluation

Key results from comparative evaluation	
Timeliness and efficacy of response, and ability to meet the needs of communities	• Six out of six evaluation groups gave the tool a higher rating for **timeliness** in comparison to the current, paper-based system for reporting rumours heard in the community. • Five out of six evaluation groups thought the tool would help them better address the needs of the communities. • Volunteers saw community members as the main beneficiaries of the tool. They thought the tool would improve their ability to respond to community feedback by making the information they provide more **reliable** and **efficient**.
Skills and technology requirements	• Four out of six groups thought the tool would require the same amount or less technology than the current system, which requires multiple steps and switching between different technologies and formats. • Volunteers recognised that they would require new skills in order to use the system, but they were prepared to undergo the relevant training. They also thought it was important that community members were made aware of the new tool.

Figure 19: Comparative evaluation between Report and Respond (red) and the current process (black) by Red Cross volunteers



Radar chart shows average scores of six groups (17 participants). Participants were asked to assign a rating to statements about the criteria (e.g. how well are you able to meet the needs of the community? 1 = not well, 5 = very well).

Figure 20: Participatory evaluation workshop in Kribi, showing Red Cross volunteers completing comparative evaluation activities

"The community benefits from the tool. It's important because we'll respond rapidly to the needs of the community."

Red Cross Volunteer (Kribi)

HOW DID THE PARTICIPATORY AI ACTIVITIES INFLUENCE THE TOOL AND HELP US ADDRESS SOME COMMON CRITIQUES OF AI?

We tried multiple participatory AI interventions despite the challenges of designing, delivering and interpreting engagement activities with the project team split between multiple time zones and languages. Table 7 provides an overview. These activities showed that participation can start to address some of the critiques of humanitarian AI – and have an impact on AI performance as well as trust or acceptance of the tool by different stakeholders.

Table 7: Overview of aims and outcomes of participatory AI activities and their link to some of the main critiques of AI systems

Aims of activity, and main critique being addressed*	Key results from PAI activities
Increase the usefulness of the tool for frontline users *AI critique: Hype and power*	**Collaborative problem framing** with the Cameroon Red Cross and their volunteers enabled us to identify and work together on an issue that had emerged as a key concern for the team (and the sector) in recent years. **Usability testing of high fidelity prototypes** with Cameroon Red Cross volunteers helped us surface social and technical implementation challenges. We were able to address some of these by setting out the system architecture, data and network requirements in the technical specification we created for the tool.
Create high-quality, representative and machine-readable data to reduce the risks of AI model bias *AI critique: Bias and data gaps*	By curating the dataset and carrying out **data labelling with the Red Cross staff**, we were able to create a higher quality dataset for training the model that was tailored to operational priorities of the Community Engagement team. It also served as a proof of concept to demonstrate that future data labelling of community feedback by the IFRC could be improved by involving country-level Red Cross staff in verification and validation. We also attempted a data labelling activity with Red Cross volunteers which proved unsuccessful but offered lessons about the importance of careful activity design, better tools and facilitation for the success of participatory AI (see Box 4).
Optimise the AI model's performance *AI critique: Transparency and explainability*	**Crowdsourcing new rumours** from staff helped us to surface an issue with our classification model. We found it struggled to assign more than one label to rumours that spanned multiple categories. This will need to be addressed during future iterations of the tool.
Create appropriate levels of trust and understanding of the tool by all stakeholders *AI critique: Transparency and Explainability*	Knowing that we would bring together data scientists and stakeholder groups with low levels of technology literacy to interrogate how the model works prompted us to choose a more interpretable model that would be easier to explain. **Crowdsourcing new rumours** with staff and discussing model biases with volunteers helped them to understand the limitations of the tool and how they could interpret the outputs it produced more critically.
Ensure all stakeholders have a say in evaluating the performance of the tool *AI critique: Accountability*	Involving volunteers in evaluating the model and introducing a volunteer-led approach to **testing the clustering algorithm** meant we were introducing new mechanisms of accountability.

*Relevance of activities to specific AI critiques are the result of post-hoc analysis.

The vignettes that follow provide a detailed description of two participatory AI activities we tried during the project. The first is an example of the importance of user-friendly task design and tools, as well as careful facilitation, for the success of participatory AI. Both vignettes describe how Red Cross staff influenced the development of the classification model.

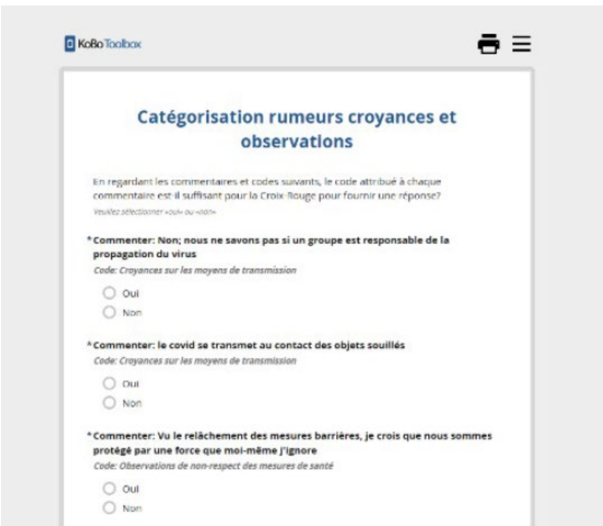
Box 4: Data labelling with Red Cross staff and volunteers

What we did

We designed a data labelling activity for Red Cross staff to help us verify and correct three of the eight rumour categories in the IFRC Go dataset. We created a two-step labelling activity using Microsoft Excel where participants had to: 1) verify the existing label and; 2) suggest an alternative if it was incorrect. Staff reviewed 530 comments in total, with each participant covering 120 rumours. There was overlap between rumours so each one was reviewed by at least three people. The activity was facilitated by the local community engagement and data science fellows, with remote support from the Nesta technical team.

We also created a simpler (one-step) version of the activity for volunteers to cover the remaining codes. We designed the activity in KoBo after failing to find a commercial or open source data labelling platform with a user-friendly mobile interface that allowed flexible task assignment and functioned well in low-resource settings. Eleven volunteers took part remotely on their phones after a short briefing by a member of the Red Cross Community Engagement team.

Figure 21: Example activity for labelling with CRC volunteers using KoBo Toolbox



What we found

Red Cross staff

In total, 130 rumours were classified as incorrectly labelled by two or more participants leaving 400 verified rumours for the three codes. Inspection of a random subset confirmed that the task had been completed correctly by the participants.

Red Cross volunteers

We expected the activity with volunteers to deliver similar results. But random inspection showed that 40 out of 50 samples had been incorrectly labelled (by multiple volunteers). Red Cross staff suggested that poor quality of the results could be due to the lack of variety in the task, or technical difficulties leading to frustration and attention lapses. It's also possible that the volunteers misunderstood the purpose of the activity due to the limited facilitation and their lack of previous experience with similar tasks.

What this means for model development

We used the 400 verified rumours from the staff labelling activity as part of our training data for the classification model. Although the results from the volunteer labelling weren't useful for model development, they helped to highlight the importance of creative task design and good facilitation, as well as the need for better digital tools to support participatory AI activities in low-resource settings.

Box 5: Crowdsourcing new rumours to test the classification model

What we did

We asked frontline staff to generate a novel dataset of rumours to help us test the robustness of our model to more naturalistic rumours. We asked participants to contribute one to three new rumours for each category. Five members of the Cameroon Red Cross Communications team took part in the activity.

What we found

We tested the model on the new crowdsourced dataset. Although the performance of the model remained relatively high (~75%), this was a decrease of more than 10% in comparison to its performance on the original test data.

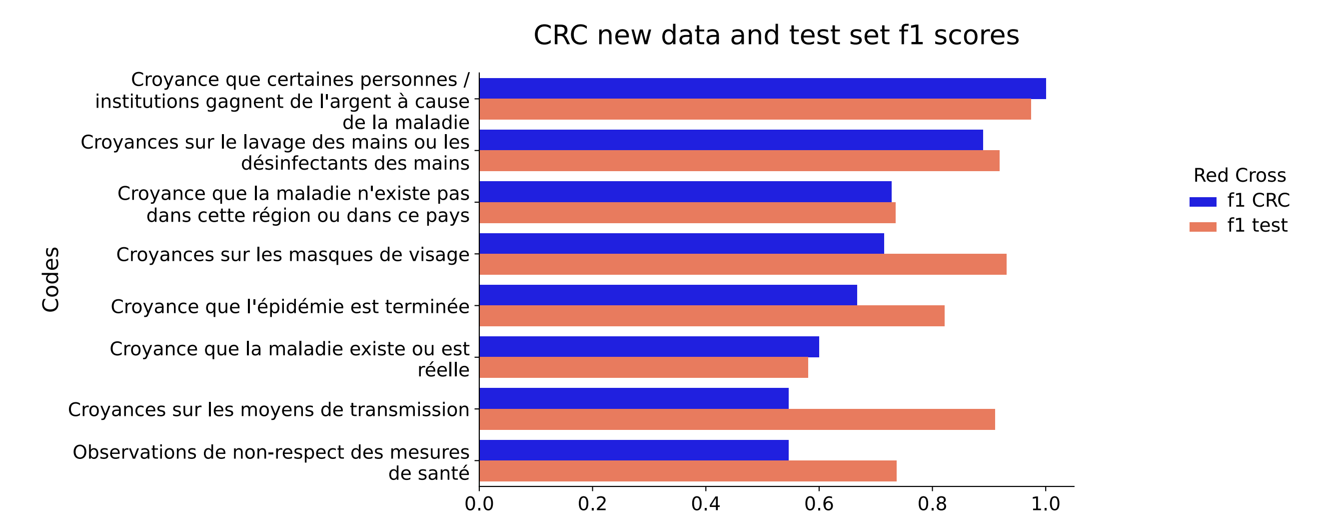
This activity helped us identify a blindspot where the model wasn't performing as expected. Several of the rumours suggested by participants were more naturalistic so they covered more than one category. We found the model struggled to assign multiple labels per rumour, despite being developed with this functionality built in.

Follow up discussions gave the team a chance to discuss how ambiguous rumours and incorrectly assigned labels could impact the model's accuracy. This highlights the value of participatory AI for helping participants to better understand the modelling process.

What this means for model development

This activity highlighted that real-world rumours and community beliefs often span more than one rumour category. Red Cross staff felt it was important for the model to be able to assign multiple labels. For the model to handle these instances in the future, it would need to undergo additional training using community feedback datasets that have more than one label.

Figure 22: Model performance on rumours generated through the crowdsourcing activity with staff (blue) in comparison to the performance on the original test dataset (orange)



HOW TRUSTWORTHY IS THE TOOL AND SHOULD IT BE OPERATIONALISED?

The trustworthiness of AI systems is a key implementation barrier, especially when tools replace well established processes. We used focus group discussions to understand how biases might impact the trustworthiness of the tool. We also measured participants' general attitudes towards the idea of computers making decisions about misinformation using post-workshop questionnaires, as well as specific positive or negative associations towards the tool and its future development. Table 8 provides an overview of the results.

Table 8: Overview of results related to trust and operationalising the Report and Respond tool

Overview	Key results
Trust	**Tool specific** • The majority of volunteers in Cameroon gave a higher or equal trust rating to the tool in comparison to the current process (five out of six groups). See Figure 20. • Overall, 83% of participants in the evaluation felt positive about the tool.[36] • Between the design and evaluation workshops, the proportion of volunteers who felt more positive about the tool increased by ~10%. This was the result of fewer people being neutral (neither positive or negative). **General AI** • All volunteers (n = 24) agreed that we should trust computers to help us understand community feedback,[37] even those who did not fully understand how the tool worked.
Uptake and usefulness	• 96% of volunteers (n = 23) agreed that further investment in the tool was worthwhile.

Photo: Edouard Tamba, Unsplash

What next?

Further work is necessary to make the tool ready for testing in operational settings and full deployment. As a priority: a) the coding system of the Red Cross should be refined on a regular basis to eliminate redundant labels and enable the distinction between closely related topics and; b) the system should allow for multiple labels per rumour so the model can learn to respond appropriately to community feedback that spans multiple categories.

Further work could focus on training the model on the full training dataset from IFRC Go, incorporating all of the additional rumour categories[38] and using rumours from other sources, such as social media platforms. Training models on a variety of data sources would enable triangulation of their performance for detecting rumours. Using additional, timely sources of data would also help to more reliably determine which rumours are emerging at any point in time. The model we have developed is only appropriate for a French language context, however the approach is generalisable. Extending the functionality of the tool to cover other languages would rely on additional data and modelling.[39]

The short timeframe of the project only allowed us to test discrete parts of the workflow in an isolated fashion, outlined in the description of 'How the tool works'. We weren't able to fully develop and test the new internal workflows that would be necessary for the Cameroon Red Cross Society to make the most of this tool. For this, the Community Engagement and Accountability team needs to develop regular processes for: a) reviewing the outputs of the clustering algorithm to keep track of emerging rumour categories; b) labelling and verifying rumours as they come in; c) assigning appropriate responses for new rumour categories and; d) updating responses for existing rumour categories if new guidance emerges.

At present, the Cameroon Red Cross does not have permanent staff with the technical capabilities required to develop AI models, or the digital systems infrastructure to process and maintain datasets. This is a major barrier to the application of CCI in local humanitarianism.

WIDER RELEVANCE FOR THE HUMANITARIAN SECTOR

The scale and spread of misinformation and disinformation is a growing challenge for humanitarian organisations. The COVID-19 pandemic has highlighted the problem, with misinformation helping to exacerbate the crisis and hindering humanitarian response. The Report and Respond tool shows the potential of involving frontline staff and volunteers in the efforts to address this problem. Putting CCI tools in the hands of those in regular contact with communities can help humanitarian organisations address misinformation in a more timely manner, provide locally appropriate responses, and adapt their interventions accordingly through better real-time monitoring.

We hope other initiatives to address misinformation, such as Médecins Sans Frontières' 'MSF-Listen' project[40] and the KatiKati tool being developed by the team behind the Africa's Voices project,[41] may be able to adapt and incorporate elements of our tool into their own operations.

04

What we learned about the benefits of collective crisis intelligence and participatory AI

Photo: Aalok Atreya, Unsplash

Our project was one of the first attempts to test the value of two novel approaches in humanitarian innovation: collective crisis intelligence and participatory AI.

The results presented in earlier sections confirm that collective crisis intelligence approaches can lead to better solutions, and that participatory activities can have a substantive impact on AI performance, trust, and acceptance of the tool by different stakeholders. We summarise the key findings below.

Although promising, they offer only an initial view of the potential of these emerging innovation approaches. Further experiments are needed to demonstrate how they can be applied most effectively to confer maximum benefits to local humanitarian responders and the communities they serve.

Collective crisis intelligence has the potential to make humanitarian action more timely and appropriate to local needs

Collective crisis intelligence tools combine locally-generated data and insights from the people closest to a crisis with the processing power of AI. Our comparative evaluation suggests that these tools have the potential to significantly improve local humanitarian action.

- In Nepal, Red Cross Society staff rated the NFRI-Predict tool more highly for accuracy (the ability to match the needs of communities) and vulnerability (the ability to meet the needs of the most vulnerable) than the current process for determining NFRI distribution. A majority also thought the tool would make the process of NFRI distribution faster.
- In Cameroon the results of the comparative evaluation showed that Red Cross staff and volunteers thought that the collective crisis intelligence Report and Respond tool would lead to significant improvements in the speed of the response and the ability to respond to community needs.

Collective crisis intelligence can enhance the utility of locally-generated data and drive new forms of action

Citizen-generated data and frontline insights are at the heart of collective crisis intelligence tools. Although this often means creating new datasets, tools can also draw on existing data held by humanitarian organisations. AI models can help transform these community datasets into insights that drive locally appropriate decisions and more anticipatory action.

- In Cameroon, we showed that collective crisis intelligence could elevate the value of underutilised community feedback data. We repurposed data already held by the IFRC, and used it to prototype a faster way to report, monitor and respond to COVID-19 misinformation, with the goal of containing it more effectively.
- In Nepal we demonstrated that applying AI to a new dataset of community NFRI preferences from geographically distinct regions and 3000 diverse households could give the Nepal Red Cross important insights into how they can predict which aid items will be needed in advance of a crisis. This has the potential to enable more anticipatory action, including through the stockpiling of the most essential goods for people in different regions.
- The Nepal Red Cross are already considering other ways they could use CCI approaches to develop more anticipatory operating models – predicting volunteer deployment needs depending on crisis type and community needs, predicting blood donation requirements, or more accurately estimating the value of cash transfers.

Collective crisis intelligence and participatory AI can help increase trust in AI tools

The perceived trustworthiness of AI systems is important for their uptake and, ultimately, their social licence to operate. In our project, stakeholders had high levels of trust in the collective crisis intelligence tools we developed, but more work is needed to untangle the factors which had the most influence on building that trust.

- In both countries, the majority of frontline staff and volunteers rated the CCI tools we developed more highly on 'trust' than their current operational processes for tackling those problems.
- In Cameroon we saw an increase in the proportion of volunteers who felt positive (rather than neutral) about the CCI tool in successive workshops.
- In Nepal, knowing that technology had been developed with the involvement of local stakeholders and using local data increased trust in the technology and the likelihood of using it amongst frontline staff, but only if it isn't the only input into decision making.

Participatory AI can overcome several critiques and limitations of AI to improve model performance

The curation of datasets, prioritisation of input features, and other specifications of model design are typically chosen by data scientists in combination with automated methods. We showed that, when activities are well designed, participatory approaches can be a useful input into any of these stages of technical development and can lead to better performing models.

- In Nepal, this meant validating and refining the data inputs used to train the AI model – specifically removing Ethnicity as an input – due to local preferences. This resulted in an algorithm of equivalent accuracy that also took into account the concerns of stakeholders and communities on the frontline of crises.
- In Cameroon, involving Red Cross staff in generating a dataset of new rumours to test the model, helped identify a 'blindspot' in the functionality of the AI. As a result, this issue can be addressed to improve the model before deployment.
- In Cameroon, we demonstrated that it is possible to help improve the quality of community data by designing collective crisis intelligence tools that build validation into the workflow. We also showed that Red Cross staff could successfully complete labelling activities to generate datasets of higher quality data for training an AI model. Ultimately, this should help increase trust in citizen-generated data, helping it to be taken more seriously as a source of intelligence.[42]

Participatory AI helps to surface tensions between the assumptions and standards set by AI gatekeepers versus the pragmatic reality of implementation

Local stakeholders are rarely involved in setting the benchmarks that models are measured against. Likewise, recommendations for how AI tools should and shouldn't be used are often set by technologists based in the Global North who are far from understanding the local context.

- During evaluation activities in both countries, we learned that frontline stakeholders were more pragmatic about using AI across a broader range of circumstances. They preferred to use a tool that was 'good enough' rather than 'perfect', as long as it improved on the current process.
- In Nepal, we identified key areas of consensus and differently held values between stakeholder groups in order to set more appropriate evaluation criteria for the model. Alongside the technical evaluation, this approach can help give a more nuanced understanding of AI performance.

Participatory AI creates opportunities for building and sharing new capabilities among frontline staff and data scientists

Our previous research highlighted that frontline staff, including volunteers, are often underutilised as key stakeholders in CCI innovations. The interdisciplinary environment of collective crisis intelligence projects creates multiple opportunities for upskilling among project team members and participants alike.

- Participatory activities helped to convince data scientists of the importance of involving a wider community of stakeholders in model development, particularly for problem definition, data wrangling and surfacing post-deployment challenges.[43]
- In Cameroon, both staff and volunteers benefited from learning new skills through labelling and design workshops. The activities also motivated the volunteers to play a more active role in the battle against misinformation.
- The project offered multiple opportunities for skills exchange between individuals and partners. For example, the Red Cross valued learning about AI and trying out digital tools for prioritisation and prototyping during community engagement activities. At the same time, data science fellows developed new capabilities in designing and delivering participatory activities.

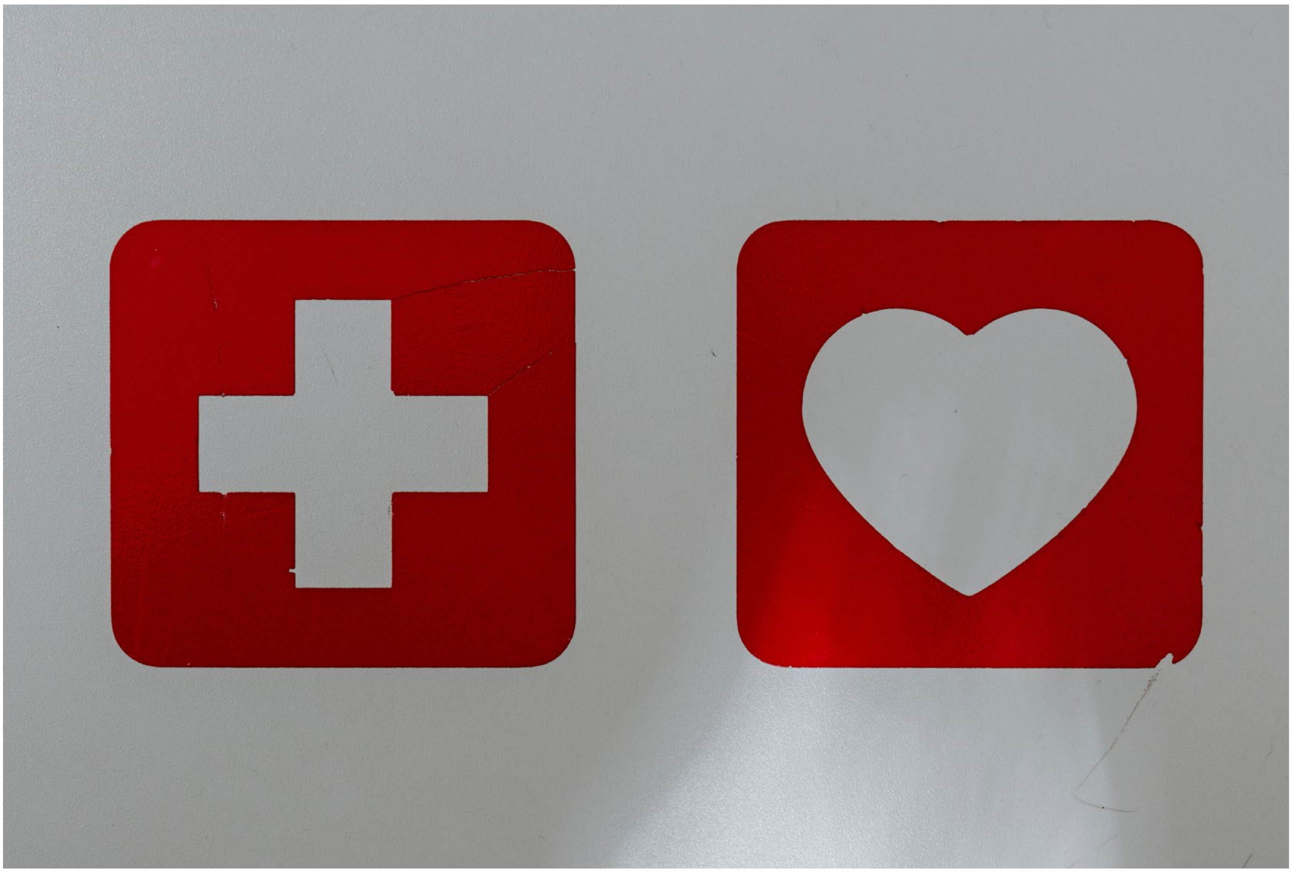


Photo: Claudio Schwarz, Unsplash

05

Barriers, challenges and how to overcome them

Photo: Edouard Tamba, Unsplash

Innovators who want to make use of collective crisis intelligence may encounter several potential technical, organisational and methodological challenges as they develop new tools. Our landscape analysis, 'Collective crisis intelligence for frontline humanitarian action', set out a comprehensive overview of barriers to the application of CCI by humanitarian organisations. Rather than repeating these, we outline the five main barriers we faced during project implementation. Addressing these barriers should be a priority for the sector in order to reap the full benefits of collective crisis intelligence.

Local humanitarian organisations have data, but it's not AI-ready and open datasets can't fill the gaps

Our experience shows the majority of community datasets held by humanitarian organisations are small scale. We found that even larger, labelled datasets can suffer from poor quality labels, particularly when categories are subjective and coding taxonomies are not reviewed regularly. Although we hoped to enrich our models with open data, we faced similar issues of gaps or insufficient granularity to be easily repurposed for AI development.

Unless we invest in levelling the data playing field and develop AI techniques that can function under these constraints, AI will remain inaccessible and inappropriate for the countries most affected by humanitarian crises and local humanitarian responders.

Recommendation

Technologists should develop AI methods that can function in low resource settings, data sparse environments, or with poorly labelled datasets, rather than data hungry supervised methods and increasingly larger scale language models.

Recommendation

The humanitarian sector should also take a more coordinated approach towards filling data gaps. It must invest in building up open datasets that are relevant and useful to local frontline humanitarians, and help ensure they are up-to-date and high quality.

Local organisations are currently ill-equipped to participate in AI development

During our partnership, we were told of multiple organisational datasets and existing models that could be relevant to our work together, but local Red Cross staff could not provide a description of what the data contained or how the models worked due to poor historic documentation. Negotiating access to existing data took many months, while the capabilities and technology infrastructure to collect, analyse and securely manage new data in the long term was limited.

Locally-developed AI that can address local needs will only be possible if international funders and humanitarian organisations invest in locally-grown data science talent and digital infrastructures, and build data capabilities across the board, not just in information management teams.

Recommendation

The humanitarian sector should insist on properly documenting existing AI models and support local organisations to develop and maintain data inventories they can share with collaborators.

Recommendation

Make the shift towards widespread data capability across all departments, including AI literacy, to understand when it can or can't be used. Create local embedded data fellowships, building on the UKHIH model, to bring technical talent into local organisations.

Shifting from 'implementation' partnerships to multi-disciplinary 'co-design' partnerships

It took time for our core project team to transition away from classic (imbalanced) roles in programme delivery towards multi-disciplinary 'co-design', with equal contributions to defining agendas and success. During early stages of the project, misalignment about these different operating modes and lengthy contract negotiations caused delays and frustration for all partners. These challenges were further exacerbated by our inability to meet in person due to COVID-19 travel restrictions.

It's unfair to expect local partners to be able to make the radical shift required without dedicated support to transition to a different way of collaborating. Supporting this transformation should be a sector-wide priority as it moves towards local development and ownership of new AI tools.

Recommendation

Allocate sufficient time dedicated to developing enabling environments for co-design, as well as prioritising alignment on roles and expectations.

Participatory AI requires new community engagement skills, and data scientists who engage with participation

AI can be difficult to explain, especially to stakeholders from very different backgrounds. A key challenge for the field is creating activities that help participants make the conceptual leap so they can contribute meaningfully to participatory AI activities. We had to design most of our activities for explaining AI and engaging affected communities and frontline volunteers from scratch. Delivering the activities required a basic level of AI literacy, as well as flexibility from facilitators. Our experience also highlighted the challenge of translating outputs from participatory AI activities into actionable changes to the model in the absence of a shared vocabulary between community engagement and technical practitioners.

There is still a long way to go before we'll be able to say which participatory AI methods are the most effective. But as the evidence builds, activities need to be codified and made simple for others to use.

Recommendation

Create easy-to-use guides for running participatory AI activities, with a clear role for technical teams and provide specialised training for community engagement practitioners.

Too few digital tools work in low-resource settings

There are relatively few digital tools to support participatory AI activities that function well in low-resource settings. KoBo Toolbox is a rare example, which we had to repurpose for multiple tasks in the absence of viable alternatives. In both countries, we saw the value of demonstrations and visualisations in making AI more tangible. In Cameroon, high-fidelity prototypes helped participants understand how the AI model could support operational tasks in the field. However, we found that existing digital prototyping tools hadn't been developed with low-resource settings in mind. Although interacting with prototypes helped participants connect the concept of AI with their operational reality, their frustration with technical issues sometimes outweighed the benefits.

Our unsuccessful attempts to involve volunteers in data labelling also highlighted the need for new open source labelling tools with user-friendly workflows, high data protection standards, and flexible control over task assignment. In general, participatory AI needs more creative digital tools, especially those that allow communities to enter data in their own words.

Recommendation

Develop adaptable, open-source, and user-friendly digital tools that function in low resource settings to support participatory AI.

06

A call to action

Photo: Michael Romanov, Unsplash

NO AI ABOUT US WITHOUT US[44]

It is time to upend and localise the development of humanitarian AI. We must move away from top-down initiatives and proprietary systems that risk reinforcing colonial power dynamics, undermining the localisation agenda, and potentially cause more harm to crisis affected communities.

The sector must stop seeing crisis affected communities as mere data points, and start to create opportunities for them to actively shape and own the tools that can help them address the crisis they face. This needs a coordinated commitment to a different way of doing things from humanitarian funders, innovators, and researchers across the board. At a minimum, humanitarian technology projects should be required to demonstrate how affected communities have been involved in the development and oversight of new tools.

Our project has shown that it is possible to build AI with local infrastructure, local data, and local talent, and that it is possible to build AI that responds to local values and priorities. But much more investment is needed to realise a future where locally-developed and owned AI becomes 'business as usual'.

Through our Accelerated Innovation Collaboration, we have started to demonstrate the potential for collective crisis intelligence – the combination of intelligence from crisis-affected communities with the predictive power of AI – to create new operating models and more anticipatory and appropriate humanitarian action. But there is much more to discover, and a need for more rigorous experimentation.

In our previous report, 'Collective Crisis Intelligence for Frontline Humanitarian Innovation',[45] we set out 10 key research and development opportunities in this space (Table 9). We are grateful for the support of the UK Humanitarian Innovation Hub, which has allowed us to start to explore some of these opportunities. We call on more humanitarian innovation funders to identify the role they can play in nurturing an ecosystem of actors committed to advancing this emerging field.

Table 9: Ten key research and development opportunities for collective crisis intelligence

Expanding CCI solutions to new users	Applying CCI solutions to new issues in crisis management	Leveraging new technologies in CCI solutions
Develop CCI solutions with and for frontline responders and affected communities	Expand situational awareness to include misinformation and disinformation	Leverage unsupervised or semi-supervised machine learning techniques and increasing availability of open data
Use collective intelligence methods to deepen community participation in crisis management	Predict the resources needed for crisis mitigation, response and recovery	Model the complexity of crises and the network effects of humanitarian actions for better anticipation
	Monitor humanitarian response and recovery efforts	Participatory modelling for improved multi-stakeholder decision making
	Leverage CCI to facilitate distributed intelligent actions for crisis response	Use CI to bridge the gap between human reasoning and AI predictions

Appendix

THE PROJECT OUTPUTS

Over the course of the project we developed and tested discrete parts of the workflows described above. In summary, we created:

Nepal

- A new dataset about NFRI preferences at the household level.
- A new dataset about additional items that communities felt should be included in Red Cross NFRI distribution.
- A classification algorithm to predict the 'essentialness' of Shelter (housing and clothing) items.
- A classification algorithm to predict the 'essentialness' of Wash (health and sanitation) items.
- A prototype of a new desktop-based NFRI-Predict tool interface.

Full model documentation: CCI Nepal. Detailed methodology and results: Technical Report Nepal.

Cameroon

- A classification algorithm for matching rumours to existing COVID-19 rumour categories.
- A clustering algorithm for grouping new rumours into emerging categories.
- A prototype of the SMS-based Report and Respond tool interface.
- A prototype of the mobile-based Labelling tool interface.

Full model documentation: CCI Cameroon. Detailed methodology and results: Technical Report Cameroon.

ACKNOWLEDGEMENTS

We would like to thank all workshop participants in Sindhupalchok, Mustang, Mahottari and Kathmandu (Nepal), and Tibati and Kribi (Cameroon) for their generosity in sharing their time and expertise with us. Special thanks to the District Heads of the Red Cross teams, who helped to organise regional workshops, to the volunteers who helped with data collection in Mahottari and Sindhupalchok, and to all community members who contributed to the generation of a new dataset.

Project team

Nepal Red Cross Society: Rudra Adhikari, Dharma Datta Bidari, Deepak Dawadi, Sarita Dhungana, Tara Gurung, Sakun Joshi, Anupa Koirala, Uddhav Nepal, Janardan Pokharel, Sachin Raut.

Cameroon Red Cross Society: Henshaw Arrey, Guy-Stephane Djob, Hyacinthe Olinga Eloundou, Cathy Essouma, Rodrigue Etono, Fabrice Ewane, Princewill Nkongho, Santana Nnang.

IFRC: Thierry Balloy, Sophie Louise Everest, Laurent Fernandez, Viviane Fluck, Manorama Gautum, Herve Gazeu, Ramya Gopalan, Heather Leson, Patricia Mugenzi, Yves Stephane Ngaleu, Sushama Pandey, Annemarie Poorterman.

Nesta: Aleks Berditchevskaia, Jyldyz Djumalieva, Rob Fuller, Isabel Gill, Nasra Hussein, Pius Ngwa, Lusine Mehrabyan, Kathy Peach, Saurav Poudel, George Richardson, Joana Murta Rosa, Meera Shah, Isabel Stewart, Oli Whittington.

Open Lab, Newcastle University: Simon Bowen, Tom Feltwell, Ian Johnson.

Advisory Group: João Porto de Albuquerque, Maesy Angelina, Nell Gray, Weisi Guo, George Hodge, Ewan Oglethorpe, Mona Sloane, John Twigg.

UKHIH Team: Mark Beagan, Ben Ramalingam, Tonia Thomas.

A special thanks to Sharath Srinivasan and Luke Church from the KatiKati platform who helped us to refine the design of the Report and Respond tool. We're also grateful to Shreyasha Poudel, Robert Soden and Manveer Kalirai for the two workshops that helped us explore the potential unanticipated impacts of the NFRI-Predict tool.

Endnotes

1. NFRI = non-food relief items
2. Responsible AI for Disaster Risk Management. The Open Data for Resilience Initiative and the World Bank. https://www.gfdrr.org/en/publication/responsible-artificial-intelligence-disaster-risk-management Accessed on 28 June 2022
3. Responsible AI for Disaster Risk Management. The Open Data for Resilience Initiative and the World Bank. https://www.gfdrr.org/en/publication/responsible-artificial-intelligence-disaster-risk-management Accessed on 28 June 2022.
4. Berditchevskaia, A., Peach, K., and Malliaraki, E. (2021). Participatory AI for humanitarian innovation: a briefing paper. London: Nesta.
5. Berditchevskaia, A., Peach, K., Gill, I., Whittington, O., Malliaraki, E., and Hussein, N. (2021). Collective crisis intelligence for frontline humanitarian response. London: Nesta
6. Berditchevskaia, A., Peach, K., and Malliaraki, E. (2021). Participatory AI for humanitarian innovation: a briefing paper. London: Nesta.
7. This tool was developed specifically to meet the needs of the Nepal Red Cross, but we believe that it could also be useful for other organisations involved in crisis response.
8. We selected a logistic regression model to optimise for interpretability.
9. 22 NFRI items were included in the survey. These items are typically distributed by the Red Cross within two standard packages: Shelter (covering housing and clothing items); and Wash (covering health and sanitary items).
10. Our sample was balanced between male and female respondents (~50:50) and featured households from both rural and urban locations.
11. Stakeholder mapping was carried out by Nesta and Nepal Red Cross Society/IFRC community engagement staff
12. Measured using the global micro F1-score across all items. The F1 score measures performance through a weighted average of Precision and Recall.
13. We evaluated the performance of our model on subsets of the test datasets based on the following protected characteristics: Ethnicity, Respondent Gender, Location.
14. The error was largely due to an imbalance in our training dataset: many respondents identified the majority of NFRI items as ‘essential’. As the current process followed by the Red Cross deems all items essential, we consider this type of error as causing less harm than other types of errors.
15. We asked groups of participants in Kathmandu to imagine using the tool during Preparedness, Early Warning and Response. All groups identified potential benefits of using the tool in comparison to the current process across all three stages of crisis response. See Technical Report for a detailed breakdown of results.
16. Measured using the F1 micro score.
17. Here we report the combined % of participants who selected ‘a little positive’ or ‘very positive’ in the post-workshop survey.
18. We posed the question in this way to capture broader attitudes towards AI throughout the project. Using this language also helped to ensure that the question was more accessible to all stakeholders.
19. Here we report the combined % of participants from Kathmandu and Mustang workshops.
20. A balanced dataset has the same number of input samples for each output category of interest.
21. An important priority for new data collection is including a district from the mountain region, e.g. Mustang. Sindhupalchok and Mahottari were chosen for their representativeness of other districts in the Hill and Terai regions, respectively. These two regions have historically seen the most flooding.
22. The technical specification published on GitHub alongside this report provides an overview of what is needed.
23. Berditchevskaia, A., Peach, K., Gill, I., Whittington, O., Malliaraki, E., and Hussein, N. (2021). Collective crisis intelligence for frontline humanitarian response. London: Nesta
24. See detailed analysis in Technical Report.
25. We use the terms label, code and category interchangeably.
26. Erina Mahmud. Community engagement to counter misinformation in Rohingya refugee camps https://www.oecd-ilibrary.org/sites/e8f2389a-en/index.html?itemId=/content/component/e8f2389a-en%20%20 Accessed 26June 2022.
27. Islam, M. S., et. al. (2021). COVID-19 vaccine rumors and conspiracy theories: The need for cognitive inoculation against misinformation to improve vaccine adherence. PLOS ONE, 16(5), e0251605. https://doi.org/10.1371/journal.pone.0251605
28. We report accuracy using the global micro F1 score, which gives an average across all labels. The F1 score measures model performance through a weighted average of Precision and Recall.

29. Clustering accuracy is reported using the average homogeneity score. Homogeneity score measures the proportion of comments in a cluster that convey similar meaning. It is assessed using manual inspection of clusters. This was carried out by the local data science fellow and the Red Cross volunteers during the evaluation workshops. See Technical Report for details.

30. The IFRC GO dataset contained incomplete information about the age and location of respondents so we limited our bias audit to gender.

31. https://go.ifrc.org/

32. For three of eight rumour categories, the model showed a drop in accuracy of more than 10% for the crowdsourced dataset.

33. Our classification model is a multi-class model, which should be able to assign multiple labels. In most cases, the model struggles to assign more than one code per data point. This is a limitation of the underlying training dataset, which didn't contain many rumours with multiple codes.

34. https://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm

35. We interpret this as referring to the multiple steps involved in the current system which involve switching between technologies. Volunteers start by photocopying paper to collect community feedback in the field. After collecting feedback in the field they collate all of the pieces of paper and share them with the Red Cross who transfer the feedback into a digital format so it can be sent to the IFRC GO platform. Volunteers referred to the current process as cumbersome, inefficient and costly.

36. Here we report the combined % of participants who selected 'a little positive' or 'very positive' in the post-workshop survey. The remaining participants (17%) reported neutral feelings about the tool (neither positive or negative).

37. We posed the question in this way to capture broader attitudes towards AI throughout the project. Using this language also helped to ensure that the question was more accessible to stakeholders.

38. The model was trained using eight rumour categories that were identified as priorities by the Cameroon Red Cross Society.

39. As a starting point, the model should be retrained using different text pre-processing and transformer models.

40. https://msf-listen.org/

41. Education in Emergencies in Protracted Crisis 2019-2023. https://www.africasvoices.org/wp-content/uploads/2021/04/AVF_FCDO-EiE-Final-Report-26th-February-2021.pdf Accessed 28 June 2022.

42. The Premise tool is an example of an existing tool that makes use of verification mechanisms to ensure data quality. Fiorella Riccobono. Getting Started with Hyperlocal Data Collection. Premise. (https://www.premise.com/blog/getting-started-with-hyperlocal-data-collection/). Accessed 26 June 2022.

43. In particular, problem definition, data prioritisation and helping them to think more robustly about deployment. See: Saurav Poudel. My Nesta Experience. Nepali Wanderer, 12 May 2022. https://nepaliwanderer.com/2022/05/12/my-nesta-experience-%ef%bf%bc/. Accessed: 26 June 2022.

44. Inspired by a participant quote during evaluation activities in Nepal and the widely used slogan Nothing About Us Without Us. See: https://en.wikipedia.org/wiki/Nothing_About_Us_Without_Us. Accessed 26 June 2022.

45. Berditchevskaia, A., Peach, K., Gill, I., Whittington, O., Malliaraki, E., and Hussein, N. (2021). Collective crisis intelligence for frontline humanitarian response. London: Nesta

nesta

58 Victoria Embankment
London EC4Y 0DS

+44 (0)20 7438 2500

information@nesta.org.uk

@nesta_uk

www.facebook.com/nesta.uk

www.nesta.org.uk

ISBN: 978-1-913095-65-9