Participatory AI for humanitarian innovation

A briefing paper

AUTHORS
Aleks Berditchevskaia
Eirini Malliaraki
Kathy Peach

WHAT IS IN THIS BRIEFING PAPER AND WHO IS IT FOR?

This briefing paper is 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. It is part of the initial background research and analysis for a year-long 'Accelerated Innovation Collaboration' to research, design and test collective crisis intelligence (CCI) solutions for community-based and first-line humanitarian responses. The partners in this project are Nesta and the International Federation of Red Cross and Red Crescent Societies (IFRC) Solferino Academy, with additional support from the Alan Turing Institute and the Digital Civics Centre (led by Open Lab at Newcastle University).

The analysis in this paper provides a conceptual framework for the efforts of our partnership to operationalise participatory AI approaches by designing, developing and testing collective crisis intelligence solutions with affected communities and frontline responders in Nepal and Cameroon Red Cross Societies. It accompanies our report: Collective crisis intelligence for frontline humanitarian response.

We are publishing it for the wider humanitarian community and hope it can serve as a guide and inspiration for those responsible for developing and supporting AI solutions in humanitarian contexts. It sets out how participatory approaches for developing AI systems can be employed in the sector based on existing examples. Participatory design for AI development is still an emerging and nascent practice. As it matures it will be important for the sector to introduce additional regulatory measures as well as instilling a culture of responsible data practices and data literacy to ensure that the use of machine intelligence remains appropriate and aligned to core humanitarian principles.

AUTHORS

Aleks Berditchevskaia
Eirini Malliaraki
Kathy Peach

Recommended citation: Berditchevskaia, A., Peach, K., and Malliaraki, E. (2021). *Participatory AI for humanitarian innovation: a briefing paper.* London: Nesta.

Published September 2021

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

ACKNOWLEDGEMENTS

This briefing paper draws on an initial rapid evidence review by Imre Bard. We are also very grateful for the valuable support on this project from our colleagues at Nesta's Centre for Collective Intelligence Design and in the Data Analytics team.

We would like to thank the IFRC Solferino Academy for their partnership on this project, as well as their guidance, insight and feedback on this research and report. At the IFRC Solferino Academy, we would especially like to thank Heather Leson, Ian O'Donnell, Annemarie Poorterman, Laurent Fernandez and Ramya Gopalan.

We are very grateful to the following people who are members of the CCI Advisory Group. George Hodge, Joao Porto De Albuquerque, Maesy Angelina, Nell Gray, John Twigg, Mona Sloane, Ewan Oglethorpe, Dharma Datta Bidari, Ben Ramalingam, Weisi Guo and David Fallows. In particular we would like to thank Joao Porto De Albuquerque and Mona Sloane for additional reviews. We would also like to thank Robert Soden for contributing to the research, and Nicholas Leader at FCDO and Mark Beagan at UKHIH.

The design for this report was created by Mike Green.

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

Participatory AI for humanitarian innovation

A briefing paper

Introduction

In 2016, the World Humanitarian Summit's Grand Bargain alliance committed to a 'participation revolution'. The subsequent emphasis on localisation within the sector has sought to shift the balance of power in favour of crisis-affected communities. Despite this, in most cases, local actors remain far from decision making when it comes to the design and delivery of humanitarian interventions.

By the end of 2020, more than a third of the signatories of the Grand Bargain alliance had missed the target of allocating 25 per cent of funding locally,[1, 2] while a research mapping and prioritisation exercise by humanitarian innovation funder, Elhra, revealed that only 33 per cent of all innovation projects and 57 per cent of all research projects consulted with members of affected communities during the development of an output. For emerging digital innovations, the challenges of community participation can be made worse by low digital and data literacy or failures of technical infrastructure.[3]

At the same time, humanitarian interventions are increasingly supported by data-driven solutions and there is growing interest in predictive or decision-support methods powered by artificial intelligence (AI) technology.[4] As we and others have outlined, these types of approaches can be powerful tools for the humanitarian sector to draw on.[5, 6] At their best they enhance the speed, scale and processing capabilities of anticipation and response to crises. But they also pose new risks to the rights and freedoms of the communities who are the recipients of humanitarian aid and threaten to reinforce the power imbalances that have shaped past humanitarian interventions. In addition, discussions about the ethics of using social media, mobile and other data from vulnerable crisis-affected communities to understand crises as they unfold are still underdeveloped.[7]

We stand at an important moment in time: there is a unique opportunity to unite the participation agenda with the development of new AI-enabled humanitarian interventions drawing on the foundational principles of participatory design of technology. Starting in the 1970s, participatory design of computational systems intended not only to improve the usability and accuracy of emerging technologies but also to provide a new process for democratisation and dialogue in society. By foregrounding discussion of stakeholder values and the conflicting priorities of different groups, resulting tensions were treated as resources of design rather than viewed as an inconvenient threat. These principles are not new to the humanitarian and development sectors – being reflected in the Participatory Learning and Action approaches of the 1990s. However, they have not yet been applied to the way the sector designs or develops AI solutions.

In this working paper, we have mapped out the current approaches to participatory design of AI systems from the nascent field of participatory AI,[8] and explored how these approaches may be adapted to a humanitarian setting to design new 'collective crisis intelligence' (CCI) solutions. We define **collective crisis intelligence** as the combination of methods that gather intelligence from affected communities and frontline responders with AI for more effective crisis mitigation, response or recovery. We hope that by returning to the emancipatory roots of participatory design, these approaches can offer a practical solution to the tensions between localisation, meaningful participation and the increased use of AI across the humanitarian sector.

01

What is participatory AI?

Participatory artificial intelligence or **participatory machine learning** in their broadest sense refer to the involvement of a wider range of stakeholders than just technology developers in the creation of an AI system, model, tool or application. As a field, it sits within the broader category of participatory design of technology.

There is growing interest in using participatory approaches for the design, development and evaluation of AI systems across industry, academia and the public sector. This area of scholarship and practice has emerged from the AI ethics and design research communities in response to growing concerns about the limitations of AI systems and the harms they cause.[9, 10]

In this paper we refer to AI using a range of terms that includes models, algorithms, tools and systems. We define AI as **intelligent technology systems that react to inputs from the environment to make decisions or take actions that can affect that environment**. This includes everything from the back-end algorithms that power Google's search engines and voice assistants on smartphones to AI-powered hardware systems like robots and drones.[11]

Participatory design has been suggested as a way of operationalising the multiple ethics frameworks, guidelines and principles to ensure a more responsible approach to AI.[12, 13] The paradigm of participatory design for technology was originally created to distribute power when emerging computational systems were developed and used. The field was spearheaded in the 1970s by researchers in Scandinavia who used participation to ensure that the resulting tools reflected the needs and values of the people and groups using or impacted by the use of technology.[14]

Participatory AI has drawn on a range of different theoretical framings within the broader category of **participatory design**. These paradigms influence the choices that practitioners make with respect to methods and the overall design of the participatory process:

a. **Human and user-centred design**[15, 16] where solutions are developed to meet the needs of end users in specific contexts.

b. **Value-sensitive design,**[17, 18] that emphasises human values of different affected stakeholder groups throughout the design process.

c. **Human-in-the-loop machine learning,**[19, 20] which focuses on creating accessible user interfaces that enable non-experts to interact with and test machine learning (ML) models.

d. **Participatory modelling,**[21] where computational models of real-world systems are co-created with input from different stakeholder groups to facilitate discussion and planning.

e. **FAccT agenda,**[22] that foregrounds fairness, accountability and transparency as core principles in the design of socio-technical systems.

Currently, using participatory methods to shape the development of AI technology is still far from the norm.[a] Most existing case studies of participatory AI have focused on time-limited engagements, where participants contribute to just one part of the technology development pipeline. We found fewer examples of a more comprehensive approach to participatory design where stakeholders are involved from concept design all the way through to governing the resulting AI system post-deployment.[23] There was also little evidence of real-world examples where participatory approaches helped to negotiate value tensions and competing interests of different stakeholders impacted by AI.[24]

To date, the practice of participatory AI has been driven by academics, who sometimes work in partnership with community organisations, foundations or private companies. Although major technology companies, notably Google and Microsoft, have also started to explore participatory approaches, in most cases, they still 'design for' not 'with' users and affected stakeholders. This often means that the communities affected by the system have little scope to influence the problem selection or determine the process.

We found almost no examples where communities or stakeholder groups who would be directly impacted by the use of an AI system were involved as 'co-creators' at any stage of model development. Our research showed that co-created projects were typically a collaboration between technical teams and the frontline-users of the model. This finding foregrounds the challenge we are likely to face when trying to apply participatory AI methods in the context of CCI – namely the lack of good practice and evidence around the efficacy of different participatory approaches.

Previous work has highlighted the crowd labour dependence of supervised machine learning approaches.[25, 26] We distinguish between the default 'participation' of crowdworkers[27] in the creation of AI systems to serve existing systems of power[28] and participatory AI approaches, where participation is motivated by improving outcomes or usefulness of the tool for frontline users and those impacted by the decisions of the AI in question.[b]

a. Particularly when considering participatory AI projects as a proportion of the total number of AI systems being developed worldwide.

b. The boundary between these modes of engagement is not always clear: participatory design projects with deep community engagement sometimes remunerate community members in recognition of their expertise and other projects may use hybrid models where they engage both local and distributed volunteers.

02

A framework for operationalising participatory AI

Based on a rapid analysis of existing participatory AI case studies and the academic literature, we have proposed a new framework which identifies the multiple opportunities for participatory approaches to inform the design, development and implementation of AI solutions.

The framework highlights the different purposes of participation at each stage of the AI pipeline and suggests methods that may be used to achieve this. Not all opportunities for community participation will be, or will need to be, incorporated for the development of every single AI solution. Some solutions may incorporate participation at just one or two touchpoints in the AI development process, while others may seek a more comprehensive approach to engagement.

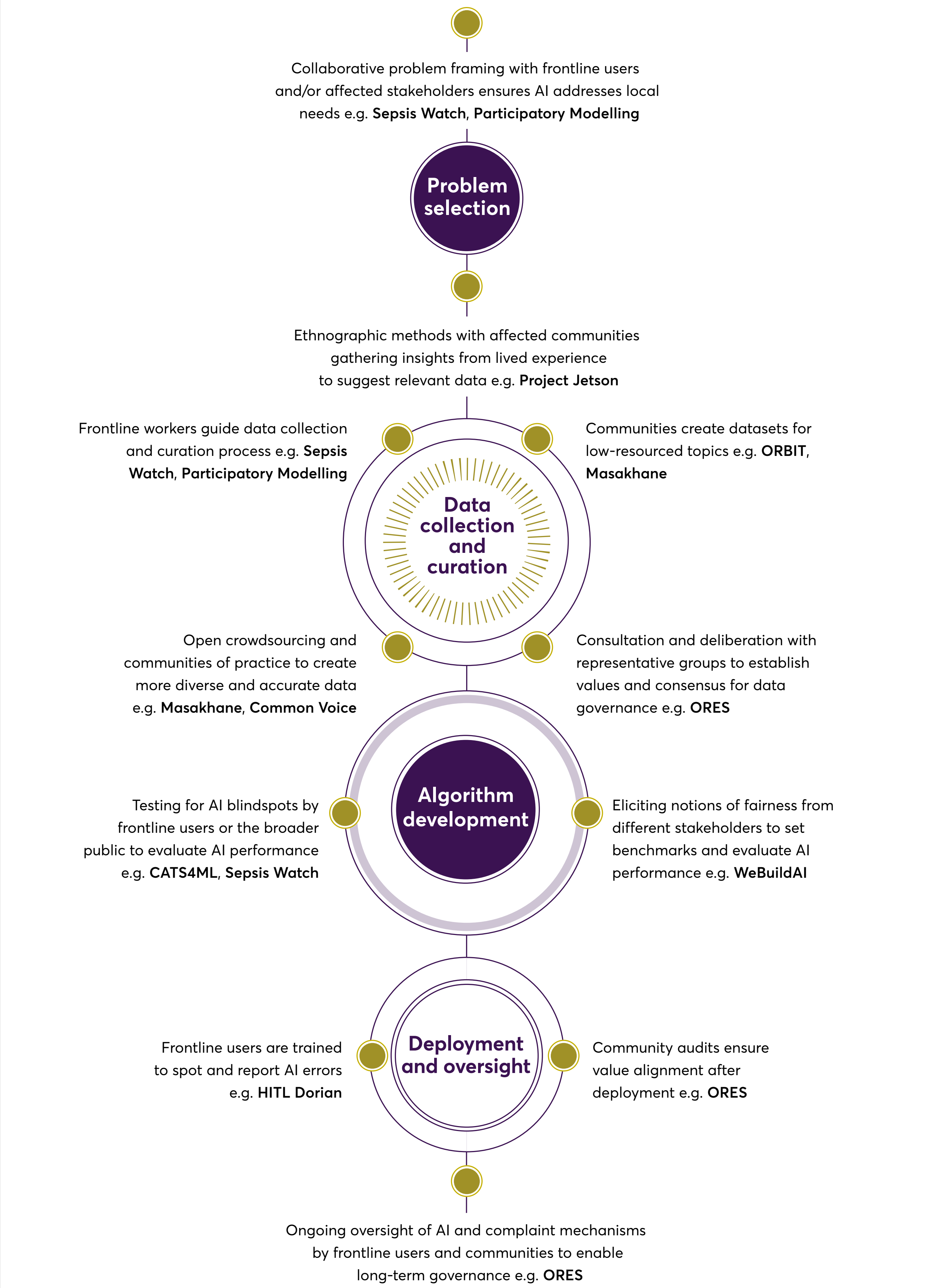
The opportunities suggested (green circles in Figure 1) have been drawn from existing case studies of participatory AI initiatives, which are cited as examples. The framework focuses specifically on the potential participation of frontline users and communities affected by the use of the tool or the problem that it has been developed to solve. It is non-exhaustive. As the field matures we expect the methods used in participatory AI will expand beyond those currently covered by the framework. We look forward to engaging with this emerging practice to help build an evidence base for what works as part of our ongoing research into collective intelligence for better AI.[29]

For now, we hope practitioners can use it as a starting point to help them plan their own participatory approaches in response to the challenges posed by AI technology. We recommend using the model in Figure 1 alongside the classification of the different levels of participation in the pages that follow and the design questions in Box 1 (see Part 4 for detail).

Box 1: Five key design questions

- Who defines the process and what counts as success?
- Whose participation is required?
- What is the intent behind participation?
- How will participants be rewarded?
- What is the process for closing the project?

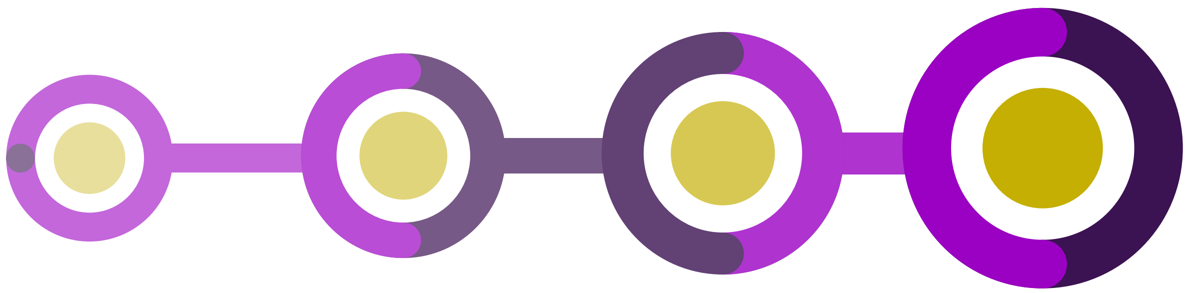
Figure 1: A framework for operationalising participatory AI



Four levels of participatory design of AI

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,[30, 31] namely consultation, contribution, collaboration and co-creation (or co-design). We explain each of these terms below.

Figure 2: Four levels of participatory AI



Consultation

Input occurs outside of the core AI development process and it is not guaranteed that it will impact the design of the AI.

Contribution

Participation is time-limited and involves stakeholders completing tasks that are necessary to AI development. Participants do not interact with each other or the technical team.

Collaboration

Participation has multiple touchpoints along the AI pipeline and includes activities where the technical team interacts with participants.

Co-creation

The most comprehensive form of participation where external stakeholders help to initiate and oversee the project as well as collaborating throughout the AI pipeline.

Consultation

Consultation refers to participation where input occurs outside of the core AI development process. The goals of consultations are typically twofold; some initiatives seek to understand public attitudes for different applications of AI or regulatory policy, while others aim to increase data literacy and raise awareness of the impacts of AI systems. These approaches are more often framed as public engagement exercises rather than participatory design. Frequently used methods include traditional research methods such as focus groups and interviews or surveys and polls. Deliberative approaches have also been used, but these are less frequent and more often in relation to public views of technologies and their regulation, rather than specifically focusing on the design of tools. Deliberation enriches more quantitative consultation methods by offering insight on why people hold certain opinions and the values involved.

Consultation typically has a less direct pathway to technology development but some high-profile initiatives can influence regulators and politicians who shape future technology policy and procure AI for public sector use. For example, the Citizens' Biometrics Council convened by the Ada Lovelace Institute in the UK,[32] where a group of 50 demographically diverse individuals was convened to discuss biometric technology such as facial recognition. Over six months, the group met for a series of workshops where they learnt about and deliberated the use of biometrics. The recommendations made by the group were shared with researchers, regulators and technology developers. Consultations can also be an important stepping stone towards increasing each individual's capacity to participate in designing technology in the future. Data literacy and activism projects that empower individuals and marginalised communities such as Our Data Bodies[33] prioritise building the skills that individuals and community groups need to determine how their data is used through activities and tools, such as their Digital Defense Playbook.[34]

Sometimes, technical teams or organisations use consultation methods like focus groups and interviews as a form of 'user research' in the scoping phase of technology development. Here, consultation helps innovation teams or engineers to better understand the problem the technology is aiming to address. These consultations can have a direct impact on technology development by surfacing important data sources, model features and performance benchmarks to include in a predictive model. In this case participation still occurs outside the core development process and participants may not realise that their contributions will impact the design of an algorithm. For example, during the development of the UN's displacement prediction tool, Project Jetson, the UN innovation team used ethnographic methods to reveal unexpected relevant factors to include in their model. Through observation and interviews with Somali refugee communities about their experiences, they realised that the market price of goats was correlated with patterns of displacement (see Case study 4 in the Appendix). They used this insight as a core input for predicting the likelihood of new arrivals in refugee camps located in Ethiopia, Kenya and Djibouti.[35]

Contribution

Contribution refers to participation that is usually time-limited to one stage of the AI development pipeline and involves external stakeholders completing one of the tasks that is necessary to AI development e.g. data collection, data labelling, validation of model outputs. Common methods include both targeted and open crowdsourcing. In this type of participation, the engineering teams who design the model often develop self-guided training materials for volunteers to develop their skills and ensure that tasks are completed to a consistent standard.

Open crowdsourcing initiatives typically aim to increase the diversity of data or perspectives for analysis and benefit from large scale online participation. Mozilla's Common Voice project[36] is an example where online crowdsourcing is being used to generate the world's most diverse and largest open dataset for training AI voice assistants. When the input from a specific community is requested it is usually to draw on their frontline and contextual expertise in order to improve the accuracy and performance of the model. The AI system developed by the Qatar Computing Research Institute to gather and classify real-time data in the immediate aftermath of a natural disaster is one example where frontline humanitarians help to oversee and correct a model's performance. The model was deployed during Hurricane Dorian in the Bahamas to rapidly assess building damage based on social media images. Thanks to the efforts of a group of local frontline responders, any classification errors made by the model were quickly identified and corrected. The participatory process a novel labelled dataset that could be used to retrain the model for future deployment (see Case study 1).

Collaboration

Collaboration refers to participatory practices with multiple touchpoints along the AI development pipeline and/or where external stakeholders are able to meaningfully contribute to interrogating the model and shaping the features that it uses to make predictions or classifications, even if they were not involved in problem setting. Collaborative participatory AI projects can have higher time and resourcing demands on participants because everyone needs to develop a good understanding of the problem being addressed as well as the technology. Common methods include the use of workshops, collaborative workspaces, shared tools and regular face-to-face or virtual meetings where stakeholders can test models as they are being developed and share best practice.

For example, the WeBuildAI project used a collaborative approach to develop a new matching algorithm for the non-profit 412 Food Rescue which helps to connect and deliver food donations (see Case study 5). The aim of participation was to ensure that the performance of the algorithm was optimised according to the values of different stakeholder groups including employees, community volunteers and recipients of food donations. Participant groups had different priorities for how food should be distributed, with some favouring efficiency and volunteer experience, while others emphasised equity as the most important optimisation criteria. The research team behind WeBuildAI used small group workshops to surface these different preferences and create representative models. Holding these interactive sessions helped the research team to improve the final model, as well as helping different stakeholders to understand how the algorithm made decisions.

Co-creation

Co-creation is the most comprehensive form of stakeholder involvement in AI design and development. Engagement occurs at multiple stages throughout the pipeline and outside it. Multiple stakeholder groups involved discuss their needs, values and priorities, both with respect to the problem space and the technology. In these examples, smaller groups of end-users or frontline experts set the initial problem definition and work closely with developers to iterate on the design of the model, the user interface and even the processes that can support its use post-deployment. This typically involves multiple rounds of iteration between the technical team and a small core group of participants, although a larger community of practice/interest may also be peripherally involved.

These collaborations can also help end-users build their skills and expertise in using AI over the course of the initiative. Occasionally they may result in the creation of governance and oversight structures for the AI system after it is deployed.

Co-creation processes typically require considerable time and resourcing commitments from all stakeholders, which can be difficult to achieve, making them rare. But they also have the broadest potential range of benefits from participation, moving beyond model accuracy towards building trust and feelings of ownership by all parties. Sepsis Watch, an AI-based tool that predicts the likelihood of patients developing sepsis in hospitals, applied a co-creation approach to involve multiple stakeholder groups throughout the design process (see Case study 2). From the outset, the development of the tool was driven by frontline users, in this case doctors and nurses. These frontline workers helped to define the problem, led the curation of data sources used to train the model and tested early versions of the tool. Importantly they also took part in the governance and oversight of the process throughout the two-and-a-half years it took to develop the tool. During co-creation different stakeholders might contribute more or less at different stages of the AI pipeline depending on their interests and skills.

The ORES process for AI development (see Case study 6 in the Appendix) was introduced by the engineering team of Wikipedia to allow their different language communities to co-create AI-enabled content moderation systems. Each time the ORES process has been used, it has been adapted to the needs and values of the volunteer community requesting the model. A smaller group of community members have multiple touchpoints with the engineering team during model development but after the model is deployed, all community members can be involved in the oversight of the AI system. The ORES process allows for ongoing monitoring of the model's performance with community-led auditing or error reporting mechanisms. Importantly, the communities also have the agency to withdraw the model if they are concerned about its performance. This action was taken by the Spanish language Wikipedians when they spotted an error in the model following a crowdsourced audit of the AI. At its best co-creation demonstrates some of the original principles behind participatory design, where communities retain decision making power over both the development and use of new technology.

The intent behind participation

Participatory approaches can serve various objectives

The aim of participation varies between different participatory AI projects. Consultation and contribution models often allow for shorter-term engagement, where participants might provide more diverse and/or contextual data, assign labels to train a model, or validate and contest errors produced by the AI system. The primary intent behind these projects is often to improve the accuracy and performance of a model so optimisation criteria are geared towards speed or accuracy metrics and validating a technical proof-of-concept. This mode of participation is not always tied to particular value commitments, and it's not explicitly oriented towards ensuring that people of different experiences, backgrounds, and expertise are involved and empowered to make decisions. It allows for a more flexible approach to participation, for example drawing on open crowdsourcing to access a larger pool of contributors who all contribute individually.

On the other hand, the projects whose overall intent is about trust, legitimacy and long-term stakeholder buy-in often foreground discussion and seek to establish collective understanding of preferences and values. These types of initiatives follow the collaborative or co-creation approaches to participation with a smaller core group of individuals who participate throughout the development process. These smaller groups have more time to build relationships with the technical teams through regular participatory workshops and can have more influence on the process that is used to develop the tool.

Shifting the power dynamics of AI design and deployment

The four levels describe how participatory design differs in terms of the tasks involved in the design and development of a technology, as well as the resources and time that are needed. But on their surface, they say little about the ability of participation to challenge the distribution of power, trust and agency between the different stakeholders affected or involved in the use of AI.[37] In participatory research, these issues are sometimes framed as the 'depth' of engagement,[38] which is expected to increase between the four levels. For example, our review showed that consultation typically resulted in less influence on technology development while co-creation offered more opportunities to shift power towards affected stakeholders or frontline users. Yet co-creation is rare and requires significant resources and commitment, as evidenced by our case studies where co-creation processes took between two-and-a-half and five years from start to finish.

Still, it is possible for engagement to be meaningful and impactful within any of the four levels,[39] as long as the contributions of participants can influence or shape the process of technology development. When the contributions of those involved are taken seriously, it helps to build trust between the different stakeholder groups. This can ultimately lead to a more equitable distribution of power, reconnecting to the original principles of participatory design. The framework in Figure 1 and levels of participation described above aim to centre communities in decision making about technology. Unfortunately, the label of participation is often used to justify political agendas that are not open to change or as part of extractive technocratic practices with little benefit to the participants.[40] Realising the original democratisation agenda of participatory design will require significant humility from those who currently hold power.

In the next section we have applied the framework to three existing case studies to illustrate different levels of participation: contribution, collaboration and co-design.[c] The cases all follow different participatory processes but ultimately result in outcomes that are useful to frontline users and/or other groups directly impacted by the tool and problem it addresses.

c. Further case studies can be found in the Appendix.

03

Participatory AI in practice

Throughout our research, we identified multiple initiatives that had used different approaches to participatory AI. We completed a detailed case study analysis for six of these to illustrate the differences in intentions, methods and outcomes, as well as practical considerations.

Table 1 gives an overview of all six case studies, including the level of participation used and the intended outcome of the initiative. Three of the case studies are presented below alongside the participatory AI framework: the remaining three case studies can be found in the Appendix.

Table 1: Case studies

Humanitarian or development sectors

Human-in-the-Loop Project Dorian	Project Jetson*	Participatory Modelling in Zimbabwe
Level of participation: Contribution	Level of participation: Consultation	Level of participation: Co-creation (frontline staff) Collaboration (local communities)
Intent behind participation: • Improving model performance	Intent behind participation: • Improving model performance	Intent behind participation: • Improving model performance • Increasing usefulness of tool for frontline users • Increasing trust and ownership of tool by all stakeholders

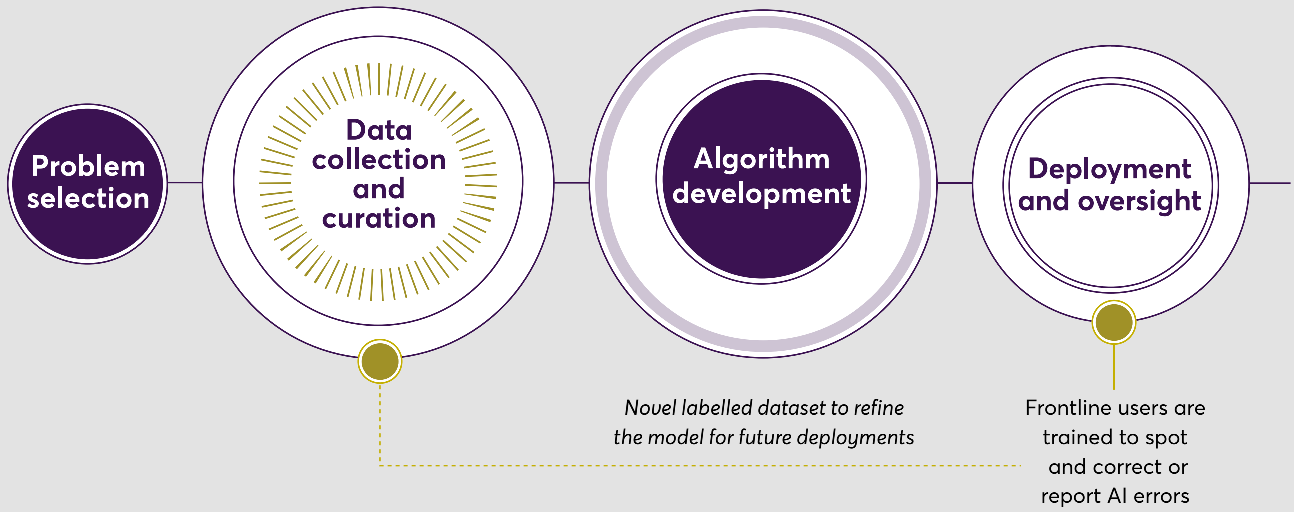
Other sectors

Sepsis Watch	WeBuildAI*	ORES*
Level of participation: Co-creation	Level of participation: Collaboration	Level of participation: Co-creation
Intent behind participation: • Improving model performance • Increasing usefulness of tool for frontline users • Increasing trust and ownership of tool for frontline users	Intent behind participation: • Optimising model performance for diverse values and preferences • Increasing trust in the tool by all stakeholders	Intent behind participation: • Optimising model performance for diverse values and preferences • Increasing usefulness of tool for frontline users • Increasing trust and ownership of tool for frontline users

*Case studies marked with an asterisk can be found in the Appendix.

CASE STUDY 1

Human-in-the-loop Project Dorian (humanitarian)



CATEGORY OF PARTICIPATION: Contribution

PAI APPROACH: Human-in-the-loop machine learning

ACTIVITIES: Labelling of datasets by frontline responders, human-in-the-loop technology development

DESIGN ARTIFACTS: Labelling platform

TECHNOLOGY STAGE: Data collection and cleaning, Deployment

PEOPLE: Montgomery County Community Emergency Response Team, Qatar Computing Research Institute (part of Hamad bin Khalifa University)

STATUS: one-off, research proof-of-concept

TIME REQUIRED: 2 weeks[d]

What's the context?

Rapid damage assessment is one of the core tasks that response organisations perform at the onset of a disaster to understand the scale of damage to infrastructure such as roads, bridges, and buildings. Traditionally experts would be sent to the disaster-affected area to conduct field assessments in order to gather relevant data, but increasingly social media data is being incorporated into crisis workflows. This helps responders assess damage without the need to gather data directly from the field, which can increase the speed of response and improve the safety of workers. The accurate interpretation of the large volumes of data that are generated on social media in the aftermath of a disaster can be challenging.

Several AI solutions have been developed to help classify social media data automatically using supervised learning, an approach that requires large amounts of labelled examples to train the algorithm. In 2019, a team of local emergency responders helped to evaluate the accuracy of an AI classifier in real time as it was used to identify social media images containing evidence of damage during Hurricane Dorian.[41]

Who is driving participation and why is it being used?

Even AI systems that have undergone extensive training on historical datasets and demonstrate high accuracy during testing can make classification errors when they encounter real-world data after they have

d. This was the time required for the participatory engagement and does not include the initial development time for the core technology, the Artificial Intelligence for Digital Response platform, which had been developed previously by the team in Qatar.

been deployed. Classification errors can have devastating consequences, particularly when they are incorporated into decision making in high-stakes contexts like criminal justice[42] or humanitarian aid. Ensuring human oversight of algorithms after deployment can help to identify and correct these errors in a timely manner. For this project, a group of volunteers from the Montgomery County Community Emergency Response Team (MCERT) helped to check and correct labels assigned by the AI classifier. This participatory approach, known as human-in-the-loop, was used to verify the classifications being made by the system and correct any mistakes in real time. The process also resulted in a novel image dataset with expert labels of damage to further train the algorithm and improve its accuracy during future deployments.

What and how of participation?

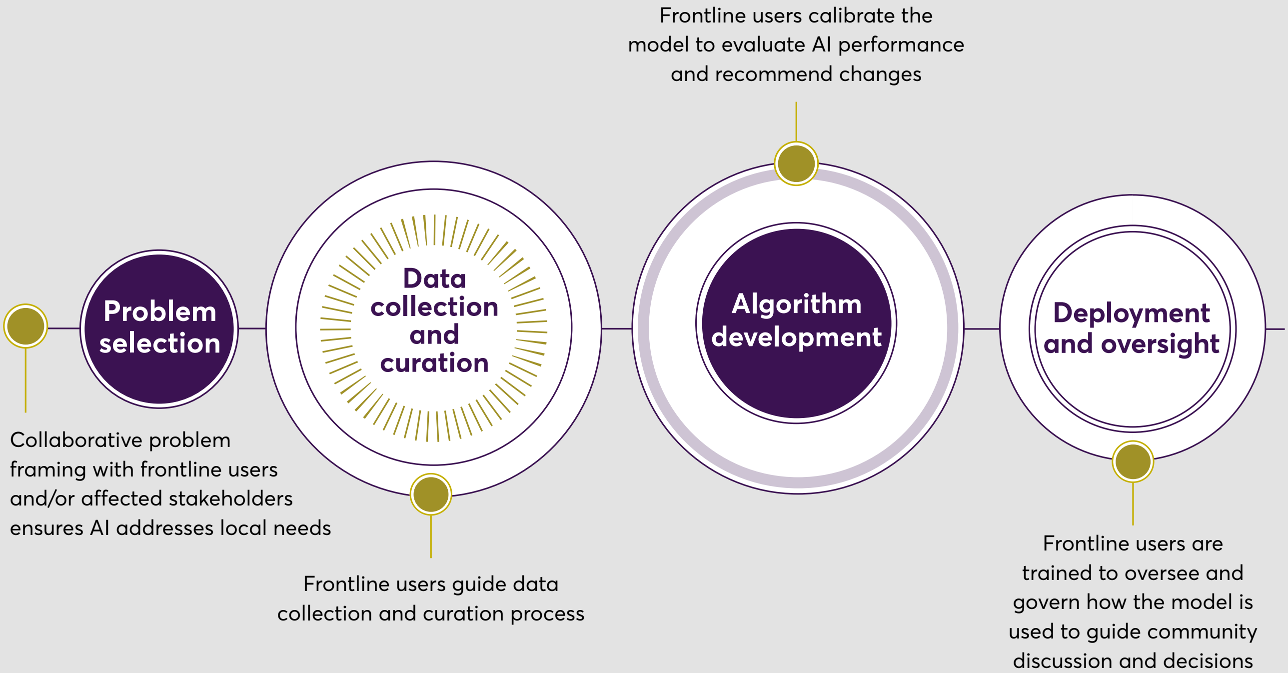
During the two-week period that followed the landfall of Hurricane Dorian, the system was deployed to collect and filter social media images. Approximately 160,000 images were tagged as relevant by the system, with ~26,000 of these containing some damage content. All images that the system ranked as severe or mild were sent for human validation. Using a dedicated web-interface for microtasking to support humanitarian efforts, local domain experts from MCERT examined this evolving sample, performing a two-stage verification. They first checked if the image contained damage content and then assigned a severity level based on a predetermined taxonomy. The research team also developed a detailed tutorial to ensure that all labellers were approaching the task in a consistent manner.

What is the impact on the AI and the problem being addressed?

The corrections made by the community responders helped to identify common types of errors made by the algorithm. For example, the AI system struggled with images taken from afar or in low-light and poor resolution settings. Isolating this weakness helped the technical team to develop targeted improvements and the response teams to better understand the potential added value of the AI tool for their operations during crisis response.

CASE STUDY 2

Participatory Modelling for Agro-Pastoral Restoration in Zimbabwe



CATEGORY OF PARTICIPATION: Co-creation with Muonde Trust (frontline staff); Contribution by local communities

PAI APPROACH: Participatory modelling

ACTIVITIES: Stakeholder workshops, Interviews, Focus groups

TECHNOLOGY STAGE: Problem selection, Data collection and curation, Algorithm development, Deployment and oversight

PEOPLE: Researchers from Santa Fe Complexity Summer School, leadership and research team from the Muonde Trust (Zimbabwean NGO dedicated to indigenous innovation), local farmers, community leaders

STATUS: One-off

TIME/RESOURCE REQUIRED: 5 years[e]

What's the context?

The knowledge and experience of local community members are vital for creating accurate models of natural resource systems,[43] particularly when competing interests and behaviours of different stakeholder groups need to be accounted for. Ecological restoration initiatives have often ignored the role of indigenous people and local groups in maintaining ecosystems and failed to account for the local history and political context. In Mazvihwa Communal Area, Zimbabwe, there are ongoing land management problems where woodland grazing areas are increasingly being converted to cropland. Despite a high level of understanding about the problem among stakeholders and the motivation of local leaders to take action, coordination between the different parties involved is difficult. To address these issues the Muonde Trust, a community-based research organisation, partnered with a group of international

e. Five years was taken for active development of the model but the researchers highlight that preliminary work that helped to trigger the development of the project had been ongoing for almost 30 years.

researchers to create an agent-based model (ABM) representing the dynamics of land use in Mazvihwa Communal Area, using a participatory modelling approach.

Who is driving participation and why is it being used?

The project was initiated by the cofounders of the Muonde Trust. One of the cofounders participated in an academic summer school where he worked with others to create the concept for the model of land-use management in Mazvihwa. Using a participatory modelling approach allowed local stakeholders to be involved in setting the parameters and data used for the model as well as exploring the impact of different interventions through simulations. Overall participation was intended to improve relevance of the model whilst also ensuring that it could be interrogated by local stakeholders to help support their discussions and decision making.

What and how of participation?

The original modelling goals were determined collaboratively by Muonde Trust and the external research team, drawing on the research questions that had emerged from the 35-year research programme of the NGO. The model used data that had been collected by Muonde researchers over three decades and they also advised the external modelling team which elements the ABM needed to include during the early stages of development. After the external researchers created the first version of the model, they held multiple workshops with the Muonde researchers to further calibrate and refine it. These workshops gave the Muonde team the opportunity to increase their understanding of how the model worked and grew their confidence in using it. Their involvement in developing the model helped ensure that enough complexity was retained to make the model useful and recognisable to the community.

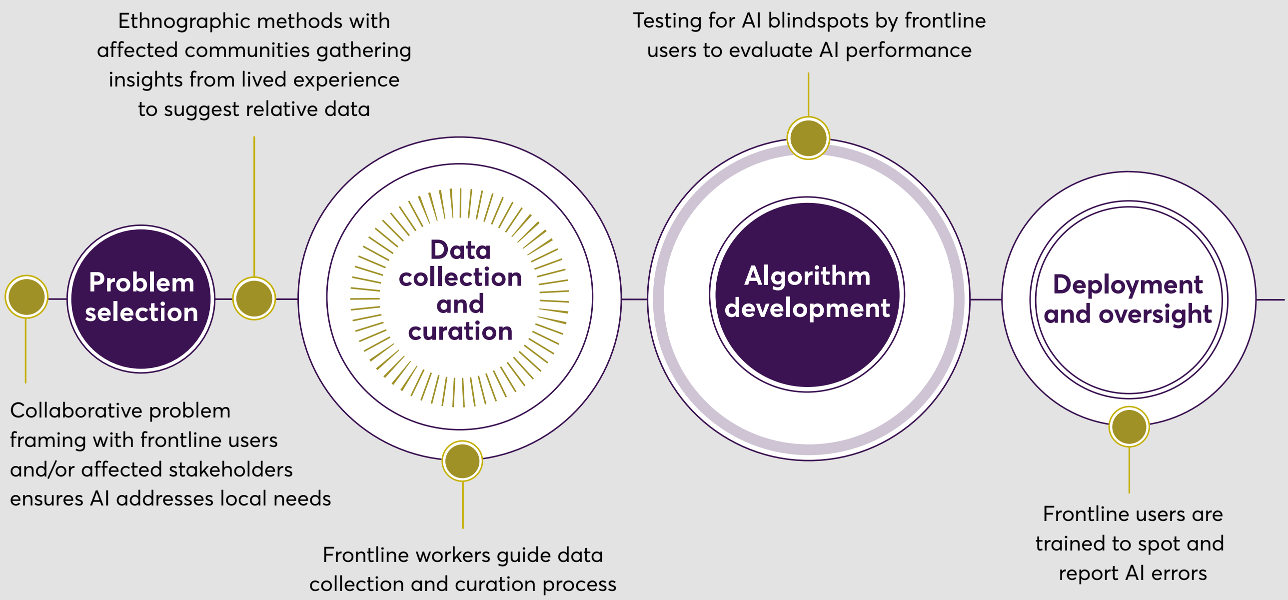
What is the impact on the AI and the problem being addressed?

After the model was finalised, the Muonde team held workshops with local traditional authorities, including the local Chief, the ward head (machinda), and the individual village heads (sabhuku). In these workshops, they used the model to visualise different actions and resulting impacts, which helped local stakeholders to discuss land management practices, past actions that had resulted in drought and the challenges associated with their agro-pastoral system. These discussions led to policy changes that allowed the re-use of fallow fields for farming. Months after these workshops, the community members reported that the model helped them realise the interaction between their different behaviours, understand the systemic drivers in the ecosystem and to take responsibility for the collective action that was necessary to improve outcomes.

Through interviews and ethnographic observation,[44] the research team established that the participatory process helped local leaders and the Muonde team to trust the model as a discussion tool, which led to more productive dialogue during workshops that eventually resulted in long-lasting pro-environmental policies. Over the two year period following the initial community workshops, 41 new homesteads had been established on previously fallow land. The Muonde team also found the model easier to understand and communicate to other local stakeholders because they had been involved in its development.[45]

CASE STUDY 3

Sepsis Watch



CATEGORY OF PARTICIPATION: Co-creation

PAI APPROACH: Human-centered design

ACTIVITIES: Stakeholder workshops, Collective governance, User training, UX feedback

DESIGN ARTIFACTS: Model fact-sheet

TECHNOLOGY STAGE: Problem selection, Data collection and curation, Algorithm development, Deployment and oversight

PEOPLE: Clinicians, Machine learning experts, nurses, Hospital management, Hospital IT staff, External innovation team

STATUS: One-off, Model is deployed and in use

TIME REQUIRED: 2.5 years

What's the context?

Machine learning technologies are increasingly being developed for healthcare settings. These models are often created by engineering teams outside of the clinical environment and implementation decisions are made at the leadership level, rather than with the involvement of clinicians or nurses who are expected to use and integrate the tool into their existing workflows. Research has also suggested that machine learning models in healthcare have poor generalisability. This means their performance suffers when they are used in real-world settings where the local data follows statistically different patterns to the datasets they were trained on. Sepsis Watch is a sepsis detection and management platform that uses deep learning to predict the likelihood of a patient developing sepsis. It was co-created by a multidisciplinary team that included hospital staff and a technical team to improve the clinical outcomes associated with sepsis, which was a major issue for the local hospital.

Who is driving participation and why is it being used?

The original proposal to develop a technology solution originated with a team of frontline doctors who secured funding to work with a local innovation team to improve detection and treatment of sepsis. The team assembled to design, develop, and integrate Sepsis Watch into routine clinical practice included a full-time innovation team, as well as implementation experts, machine learning experts, and clinical experts. Participatory design was used for multiple reasons: to improve the accuracy and appropriateness of the technology solution for the problem, overcoming the limitations of tools that had failed in the past; to cultivate institutional buy-in, trust and ownership of the tool amongst clinical teams; and importantly, to retain agency and control of decision making for clinical staff.

What and how of participation?

The Sepsis Watch model was created in close collaboration with clinicians over two and a half years, with participation integrated throughout the AI development pipeline.[46] Throughout the process the team placed a strong emphasis on developing trust between the individuals involved and also in the technology solution. A governance committee with members from the leadership across nursing, clinicians, hospital medicine, information technology, and innovation teams met monthly throughout the pilot to review progress, address concerns surfaced by frontline staff, and make decisions about changes to workflow. This helped to maintain leadership support for the project across the board.

The first twelve months of the project were used to establish the team, characterise the problem, and start designing the data pipeline and workflow for the model. First of all, clinical experts curated the local datasets and selected the parameters that the model was trained on. After this, the teams dedicated one year to developing the AI system, and integrating it into a user-facing platform which became Sepsis Watch. After a model was created, clinicians evaluated the performance of the model based on known cases of sepsis, which led to further fine-tuning. Together with nurses, the clinical experts also reviewed multiple versions of the user interface for the tool. During this process of building and tuning the model, the clinical and engineering teams met regularly.

The final six months of the process focused on training of clinical staff and integration of the model into existing workflows. The development team held regular meetings with all stakeholder groups who would interact with the model including the technology team responsible for managing data, hospital management, and the nurses who were the intended end users of the tool. Sepsis Watch was initially rolled out to a small group of nurses who provided feedback after a month of using the tool, leading to a handful of final workflow changes.

What is the impact on the AI and the problem being addressed?

The thirty month co-design process for Sepsis Watch eventually resulted in its integration into the clinical setting. This process resulted in securing institutional buy-in and retaining ownership over decision making among frontline staff. Doctors and nurses developed new processes for decision making that made use of the tool alongside their expertise. As the frontline users of Sepsis Watch, nurses also reported that the tool helped them expand their professional skills.

The researchers involved with the project credited the success to the combination of rigorously defining the problem in context, building relationships with stakeholders across multiple levels of the organisation, respecting professional discretion and creating multiple touchpoints with stakeholders to enable feedback and iteration on the design.

04

Five key design questions for participatory AI

Although the exact contours of meaningful participation in the development of AI systems are still being defined, our review of case studies revealed at least five important design considerations. These questions should be considered honestly by practitioners as they face the reality and challenges of participation in their field.

1. **Who defines the process and what counts as success?**

   Who is actively involved in defining the process and will make the decisions about what success looks like and the extent to which the views of contributors will be incorporated into the AI development? If the parameters of participation are narrowly proscribed by tech developers without acknowledgement or action to address the unequal power dynamics inherent in their role, there is a risk that participation becomes instrumental, extractive or even deliberately performative.[47]

2. **Whose participation is required?**

   How will you reach the groups whose input you're seeking? Can you adapt the process to make it easier for them to contribute? Some participatory models involve large-scale distributed crowdsourcing open to all interested members of the public or domain experts. Others focus explicitly on intended end-users from specific professional or demographic groups. The most successful collective intelligence projects typically involve some opportunity for participants to self-select. By enabling people to contribute to projects that interest them and that speak to their experience and know-how, participation becomes more robust.[48] However, this needs to be balanced with specific action to acknowledge whose participation may be missing or more difficult to achieve. Given the disproportionate impact of humanitarian crises on women,[49] and the persistent digital gender gap,[50, 51] it is especially important that measures are taken to ensure their inclusion, along with other groups whose views may be sidelined, such as individuals with disabilities, indigenous communities and ethnic minorities.

3. **What is the intent behind participation?**

   Do you want or need participation to improve your AI model's accuracy? Is it to increase the appropriateness of the AI for a particular context or user? Is it to build more trust or ownership in the AI model amongst a target community or stakeholders? Is it to spot or prevent future harms from its deployment? Or do you hope to foreground the perspectives of target communities or marginalised people throughout the AI development process? Clarifying your motivations for participation will help you determine at which stage(s) of AI development participation is needed, as well as the methods and timescale of participation desired. Referring to Table 2 which provides an overview of participatory AI in a humanitarian context may help you identify specific goals. It is important to be honest about the extent to which participants will be able to challenge unexamined assumptions and institutional agendas,[52] and therefore the degree to which participation will lead to empowerment. Funders should also consider asking these critical questions and allowing sufficient time and resources if the goal is meaningful and long-term participation.

4. **How will participants be rewarded?**

   What is the value of contributing for people taking part? Will different groups need different incentives? How will you keep track of their motivations and how they change? Many AI technologies in the commercial sector have been based on the unpaid or low paid labour of crowdworkers, particularly in the production of training data for machine learning models. Participatory AI approaches should not replicate this process of exploitation or extraction from people. Practitioners need to carefully consider the value and incentives that are offered to participants in return for their time and energy. Humanitarian organisations should be able to harness intrinsic motivations – providing people with opportunities to contribute to the collective good, but should be clear what else people will receive in return. It may be important to consider which additional incentives, from recognition, to new skills or financial remuneration, are appropriate and how the costs of participating can be offset.

5. **What is the process for closing the project?**

   How will you give feedback to participants? How will participation be evaluated and by whom? What will happen to the data that has been collected or produced? Planning for how to close the project is as important as initiating it. This includes thinking about what happens with all the data that is collected or created, as well as how to evaluate the participation and provide feedback to those who have been involved. Projects are often started without considering these questions, which can lead to unused tools,[53] poor data practices (e.g. data not archived or removed) and self-exclusion by affected communities and local stakeholders after failures of previous experiences.[54]

   A unique challenge with AI systems is that they evolve over time and may change through use. If the intention of participation is evaluation and oversight, the design process needs to plan for ongoing participation or allow participants to explore ‘possible futures’ to anticipate unwanted effects.[55]

05

Making the case for participatory AI in humanitarian settings

AI systems, in their ability to learn, predict, and in some cases make decisions, are increasingly shaping humanitarian interventions.

As we have outlined in the accompanying report on opportunities for CCI,[56] AI systems can be powerful tools for the humanitarian sector to draw on, helping to enhance the speed, scale and processing capabilities of anticipation and response to crises. But they also pose new risks to the rights and freedoms of the communities who are the recipients of humanitarian aid. Many of the concerns relate to irresponsible data practices in addition to the models themselves.

For example, the prevalence of unstructured data in CCI applications has led to the disproportionate use of deep learning algorithms which are more able to process this type of data. Deep learning approaches have been called 'black box' algorithms – meaning it is difficult for human decision makers to understand and explain how they arrive at their outputs, which is problematic in domains like crisis management where the wrong decision can affect thousands of lives.

Common critiques of AI in humanitarian settings and a potential role for participatory approaches

In Table 2 we review some common critiques of the design and deployment of AI tools for disaster and crisis management. Although many of these are not unique to the humanitarian setting, they become particularly relevant when examined alongside the humanitarian principles of neutrality, independence, impartiality and humanity. Drawing on the Core Humanitarian Standard,[57] we highlight how AI systems can jeopardise these principles and how participatory AI approaches may help to mitigate against some of the risks. The people creating new AI-enabled tools for the humanitarian sector need to spend as much time understanding humanitarian principles and the local context as developing the technology. Only then will it be possible for AI to play a part in helping humanitarians to advance the localisation agenda.

The table does not cover the full range of concerns raised by the AI ethics and critical data studies communities.[58] For example, we do not address the challenge of implementing AI in contexts where technical and network infrastructure is fragile.[59, 60] Instead we have tried to focus on the opportunities that exist for overcoming common AI and data limitations through participatory practice with local stakeholders. Of course participation alone will not be enough to address these concerns. Participatory AI will only succeed if it is developed alongside other complementary measures that strengthen the ecosystem for responsible AI. This should include new sector-wide regulation, as well as data collection and sharing practices that promote privacy-by-design and trial new models for governing data (e.g. data trusts).

Table 2: Common critiques of AI in humanitarian settings and a potential role for participatory approaches

Core humanitarian standard	Common critiques of AI in humanitarian settings	Potential role of participatory AI with example methods
Communities and people affected by crisis receive assistance appropriate and relevant to their needs. Communities and people affected by crisis have access to the humanitarian assistance they need at the right time	**The design and implementation stages of an AI project are disconnected.** AI systems may be designed without adequate contextual knowledge or input from experts in humanitarian response. This can lead to an oversimplified or over-fitted model, driven by selection of the wrong data sources and/or incorrect weighting of different variables,[61] ultimately resulting in a loss of predictive accuracy. Even when models are designed in a humanitarian context, they often cannot be easily transferred to a different location without adapting the algorithm.[62]	***AI systems are designed to meet the needs of frontline users and affected populations in specific contexts:*** • Ethnographic methods to gather insights from lived experience about the context where the tool will be used to help technical teams identify relevant data sources (**e.g Project Jetson: Case study 4**) • Co-design with frontline workers to guide the problem selection, data collection and curation process (**e.g Sepsis Watch: Case study 3**)
Communities and people affected by crisis are not negatively affected and are more prepared, resilient and less at-risk as a result of humanitarian action	**Lack of high-quality, representative, machine-readable data.**[63] Quality data is required for AI to be trustworthy and secure. Finding high-quality data has historically been more difficult in developing countries and humanitarian contexts, where technical infrastructure, resources, and expertise are often limited.[64] When the data used to train an AI model is incomplete, biased, or otherwise inadequate, the system can produce discriminatory or unfair decisions and outputs.[65] Data availability and accessibility often reflect social, economic, political, and other inequalities. **The dominance of AI methods that use supervised learning techniques may privilege certain types of data**[66] that are more readily available (e.g. social media data), limiting the types of questions that can be addressed and that receive attention for innovation.	***AI systems are able to more accurately cover a diverse range of contexts, including those typically ignored or underrepresented by:*** • Crowdsourcing of data in an ethical way that considers human factors, to ensure more diverse and representative participation (**e.g Rehumanised crowdsourcing**) • Participatory research with a community of interest to gather more complete datasets in low-resourced or under-represented topics (**e.g Masakhane** and **Common Voice**). • Crowdsourcing datasets which help to surface blindspots of an AI system (**e.g. CATS4ML**)[xx]
Communities and people affected by crisis are not negatively affected and are more prepared, resilient and less at-risk as a result of humanitarian action	**Limitations of data, get amplified throughout the model development process, further exacerbating inequalities and discrimination.**[67] Other factors, such as the selection and weighting of variables can also cause models to be discriminatory. Those in need of humanitarian assistance are already in vulnerable circumstances. Many may be suffering from discrimination and persecution, and therefore, the risk of an algorithm perpetuating past bias and discrimination is even more serious.	***Creating a collective understanding of what is meant by a fair outcome helps to avoid or identify inequality of outcomes when training data are limited:***[68] • Participatory research to crowdsource and establish new benchmarks for evaluating AI systems. (**Masakhane**) • Developers work with frontline experts to identify problems with the data or any factors that may impact data collection and document this. • Preference elicitation and prioritisation of fairness notions by affected communities (**e.g. WeBuildAI: Case study 5**)

Table 2: Common critiques of AI in humanitarian settings and a potential role for participatory approaches (continued 1)

Core humanitarian standard	Common critiques of AI in humanitarian settings	Potential role of participatory AI with example methods
Communities and people affected by crisis can expect that the organisations assisting them are managing resources effectively, efficiently and ethically.	**Lack of transparency and explainability.** The recommendations or classification of some AI systems, particularly those that use deep learning algorithms, can be opaque to human decision makers. This means that even the technology teams that develop the AI systems cannot audit or explain them to the public or those in charge of monitoring their use. Even when machine learning algorithms are interpretable, this often requires a high technical expertise that frontline staff in humanitarian contexts lack.	***User interfaces and/or other design artifacts to support end-user understanding and agency are given as much attention during development as the algorithm through:*** • Interactive tools that allow users to query the input parameters and thresholds of the algorithm and can adapt these to reflect the needs of their community (**e.g. ORES Wikipedia: Case study 6**) • Thorough documentation and/or Model Fact Sheets that explain where and how the model can and cannot be used (**e.g. Sepsis Watch: Case study 3**)
	Trust in the recommendations or predictions generated by the model is undermined by proprietary models. Developers and designers of AI systems don't always communicate clearly what is included and excluded from the algorithm or how replicable and shareable their algorithm is. The hype around the private technology sector and availability of private sector funding of AI for disaster risk management and humanitarian response can lead to overstatement of the capacities of these tools and the deployment of untested approaches in safety-critical scenarios.[69] There is a potential source of tension between corporate interests and the interests of the communities served by humanitarian actors.	***Trusted intermediaries work with end users and affected communities to interrogate and calibrate models helping to build understanding, and trust about how and when it can be used:*** • Multistakeholder community mapping efforts to create, detail and improve models representing real world environments as a tool for deliberation and decision making (**e.g. Participatory Modelling in Zimbabwe: Case study 2**) • Working with end-users and domain experts throughout the model evaluation and refinement process, to address concerns and increase confidence in the performance of the system (**e.g. SepsisWatch: Case study 3**)
Communities and people affected by crisis know their rights and entitlements, have access to information and participate in decisions that affect them.	**Loss of privacy and control over the use of data.** The ability of AI systems to make inferences from either publicly or privately accessible data can have severe impacts on privacy rights. For vulnerable persons in humanitarian contexts, participation or providing certain types of personal or personally identifiable data may be conditionally tied to aid.[70] Such populations are unlikely to know their rights regarding their data, nor be in the position to protect them.	***Communities affected by the use of an algorithm are involved in setting the constraints and benchmarks to prevent harm and preserve their rights:*** • Consultation and public deliberation to derive consensus priorities for data governance can enable a more equitable and informed approach to data use and sharing to allow individuals control over the use of their data (**e.g. Ada Lovelace deliberative dialogues on data governance and AI policy**) • Good documentation and impact assessment of data protection and management from start to finish to ensure ethical approaches to the use, storage and archiving of data (**e.g. Datasheets for Datasets**)[71] • Introducing feedback loops and consent mechanisms alongside working with communities to understand data practices and how they affect them (**e.g. Our Data Bodies**)

Table 2: Common critiques of AI in humanitarian settings and a potential role for participatory approaches (continued 2)

Core humanitarian standard	Common critiques of AI in humanitarian settings	Potential role of participatory AI with example methods
Communities and people affected by crisis have access to safe and responsive mechanisms to handle complaints and participate in decisions that affect them.	**Potential errors due to AI drift.** Complex real-world issues are constantly changing, meaning that the accuracy and appropriateness may drift as the AI system evolves through use.[72] This unpredictable behaviour after deployment of live AI systems may undermine user agency. We also know relatively little about the harms that have been caused by the use of AI in humanitarian contexts due to poor tracking and sharing of these occurrences and a general attitude not to report incidents.[73] Individuals may be unable to detect when their rights have been violated due to algorithmic errors and, as a result, may be unable to seek redress.	***Frontline users are trained to spot and report errors, and have agency to discontinue use of AI systems that do not perform as expected:*** • Human-in-the-loop feedback mechanisms are integrated into the user interface for the AI system allowing trained volunteers and frontline experts to correct classification errors in real time to train the system for future deployments (**e.g. HITL Project Dorian: Case study 1**) • Community audits of algorithms after deployment and coming to collective decisions about their continued use when outputs violate community norms (**e.g. ORES WIkipedia: Case study 6**)
Communities and people affected by crisis have access to safe and responsive mechanisms to handle complaints.	**Unclear accountability mechanisms.** Establishing accountability in data science and obtaining remedy is hard, especially when applied in a complex humanitarian situation. For example, many individuals are unaware of the scope, degree, or even existence of algorithmic decision making processes that may determine their rights. The organisational complexity in development and humanitarian projects can exacerbate this challenge. It's often not clear who is responsible when a decision or analysis is discriminatory in projects that involve a large number of actors such as funders, international organisations, contractors, private sector suppliers, local civil society partners etc.	***The governance of algorithms and their use is decided together with stakeholder groups affected by an AI system:*** • Public participation in deciding on the benchmarks the AI is tested against through modelling individual preferences and ranking (**e.g WeBuildAI: Case study 5**)
Communities and people affected by crisis can expect delivery of improved assistance as organisations learn from experience and reflection.	**Lack of evidence about impact and best practice** on AI-based systems in the humanitarian sector and absence of impact evaluation from the perspective of humanitarian principles for AI applications.	***Participatory processes and the AI system are evaluated across a broad set of qualitative and quantitative criteria, taking into account the relevance of the AI system for the specific context and problem:*** • Participatory processes are evaluated using both technical and socio-cultural dimensions, across multiple timescales (**e.g. Participatory Modelling in Zimbabwe: Case study 2**) • Community of practice develops new sector-wide standards to share best practice and impact stories (**e.g. Peer-review framework for predictive analytics by Humanitarian Data Exchange**)[74]

06

Conclusion and next steps

We hope this briefing paper is useful both for humanitarian practitioners and technology developers working across the development and humanitarian sectors. We will use the key takeaways from this rapid review to help us design the participation and evaluation approaches planned for the next phase of the CCI project.

Participatory practices for designing technology cannot be viewed in isolation. The application of participatory design to AI should be tightly coupled with responsible data practices, investment in the wider ecosystem of technology regulation and attention to the inherent power structures in society.[75, 76] To allow for meaningful engagement, technology development needs to happen alongside data literacy efforts and robust technical infrastructures that prioritise the safety and privacy of contributors rather than the interests of private companies.[77]

For anyone interested in learning more about this broader context, we recommend the summary paper from the Working Group for Responsible AI in Disaster Risk Management convened by the World Bank, Global Facility for Disaster Reduction and Recovery (GFDRR), Deltares, and University of Toronto.[78]

Appendix

CASE STUDY 4

Project Jetson

CATEGORY OF PARTICIPATION: **Consultation**

PAI APPROACH: **Human/user-centered design**

ACTIVITIES: **Ethnographic observation, Interviews with frontline workers and displaced people**

TECHNOLOGY STAGE: **Problem selection**

PEOPLE: **Focus of participation: Somali refugees, Frontline development workers at Dollo Ado**

PEOPLE: **Soliciting participation: UNHCR Innovation Service**

STATUS: **One-off, Prototype development**

TIME REQUIRED: **Months**

What's the context?[79]

The intersection of climate change, conflict and drought increasingly drives displacement worldwide. In the last decade, the combination of these factors has led to large numbers of Somalis migrating to neighbouring countries of Ethiopia, Kenya and Djibouti, with up to 2000 arriving daily, often with little prior warning. The arrival of refugees on such an unprecedented scale causes a strain on the resources available at border regions and staff in humanitarian or development organisations struggle to distribute and deliver services to those in need. In 2017 frontline workers in Dollo Ado in Ethiopia appealed to the UNHCR Innovation Service to design an algorithm that would help them to better predict the flows of displaced people in order to better plan for and manage their arrival.

Who is driving participation and why is it being used?

Forced displacement is difficult to model and predict due to the complex combination of factors that lead to it, which are often unique within any given local context. In order to accurately model displacement in Somalia, it was necessary for the technical team to identify some of the locally specific drivers of migration and other variables that could be used to train the model. Here, the purpose of participation was to help the technical team identify the locally relevant contextual factors associated with displacement in order to improve the accuracy of the model's prediction.

What and how of participation?

The technical team developing the AI solution spent many months in Dollo Ado to better understand the context for which they were developing the algorithm. They conducted detailed interviews with both frontline workers and displaced Somalis to better understand the specific factors associated with migration in the region. This allowed them to isolate ten core variables to focus on for the first iteration of the predictive model.

What is the impact on the AI and the problem being addressed?

Speaking directly to affected communities about the factors that were important for their decision to migrate and the actions they took as a consequence allowed the data scientists to uncover unique variables, such as local changes in goat prices, which were important for improving the accuracy of the model.

CASE STUDY 5

WeBuildAI – a framework for a food donation matching algorithm

CATEGORY OF PARTICIPATION: **Collaboration**

PAI APPROACH: **Value-sensitive design, Human-centered design**

ACTIVITIES: **Preference elicitation through pairwise ranking, Individual interviews, Group workshop, Think-alouds**

TECHNOLOGY STAGE: **Algorithm development**

PEOPLE: **23 affected stakeholders (donors, recipient organizations, volunteers, and staff at 412 Food Rescue)**

STATUS: **Unknown**

TIME REQUIRED: **1 year**

What's the context?

412 Food Rescue provides a 'food rescue' service that matches expiring food from donating organisations such as grocery stores with recipient organisations. Its success relies on efficient matching between donations and recipients as well as a volunteer network to transport the packages. 412 Food Rescue has rescued over three million pounds of food and expanded their model to four cities in the US including San Francisco and Philadelphia. Matching between donors and recipients has been traditionally performed by human dispatchers whose workload has increased with the success of the service. The decisions made by dispatchers also show inconsistency, which leads to certain recipients receiving a significantly larger amount of donations than others. To overcome these challenges, 412 Food Rescue worked with a group of researchers to develop an algorithmic solution for allocating donations.

Who is driving participation and why is it being used?

The participatory process was initiated by researchers in order to create value-based optimisation criteria that would be representative of the different stakeholders involved with the service to train the donation matching algorithm. They developed a participatory framework known as WeBuildAI to model the preferences of different stakeholders involved in the service and ultimately satisfy their needs and priorities. For example, if the algorithm optimised efficiency, volunteers would have shorter driving times and the areas closest to donating organisations would receive a disproportionate share of donations. Prioritising equity of distribution on the other hand could make it more difficult to recruit and retain volunteers. By involving groups who prioritised efficiency or equity to varying degrees in the algorithm design, the researchers hoped that all of the participants would continue engaging with the service and that the distribution of donations would be more equitable.

What and how of participation?

Over the course of a year, stakeholders used the participatory governance framework, WeBuildAI, to design a matching algorithm to support the services provided by 412 Food Rescue.

Problem selection, data collection and curation
First of all, the research team interviewed 31 individuals[80] connected to 412 Food Rescue's mission to identify important features that form the basis of decisions about donation allocation such as travel time, income level, poverty rate,

etc. Of these, the ones that came up most frequently and had reliable data sources were selected as features for the stakeholders to consider during the algorithm design.

Algorithm development
The research team started out by creating models that represented the 'individual beliefs' of each participant about how donations should be distributed. In the first session, participants completed pairwise comparisons to choose between different outcomes for the delivery of donations. The data from this session was used to train a model (Model 1). Fifteen participants also completed a second session where they specified explicit rules[f] to create an alternative model of their beliefs (Model 2). If their belief changed after the second session they could also provide a new set of pairwise comparisons to retrain their Model 1. During the third session, participants chose which model (Model 1 or Model 2) most closely matched their beliefs. To create the final AI system for the organisation, all of the chosen individual belief models of participants were aggregated to produce a collective preference. The choices of each of the stakeholder groups were assigned different weights according to their role. For example, representatives of 412 Food Rescue were given the highest weighting, meaning their votes carried the most influence.

Deployment and post-deployment
The resulting algorithm was integrated into 412 Food Rescue's customer management system to provide recommendations for human decision makers that assigned donations. Each recommendation was accompanied by an explanation that showed how different stakeholder's individual belief models contributed to the final decision.

What is the impact on the AI and the problem being addressed?

The findings suggest that the WeBuildAI framework successfully enabled participants to build models that they felt confident represented their own beliefs. It also helped to build trust in the algorithm and empathy towards 412 Food Rescue. The resulting algorithm improved equity in donation distribution without hurting efficiency when tested with historic data. Participants also reported that the process improved their understanding of algorithms as well as raising their awareness of biases in human decision making.

f. The study started with 23 participants but only 15 participants completed all four sessions.

CASE STUDY 6

Objective Revision Evaluation Service (ORES) by Wikipedia

CATEGORY OF PARTICIPATION: **Co-creation**

PAI APPROACH: **Value-sensitive design**[81]

ACTIVITIES: **Crowd labelling, Crowd auditing, Collective governance**

DESIGN ARTIFACTS: **Documentation, API, Labelling platform**

TECHNOLOGY STAGE: **Problem selection, Data collection and curation, Algorithm development, Deployment, Post-deployment**

PEOPLE: **Wikipedia Scoring Platform engineering team, Language communities of Wikipedia**

STATUS: **Ongoing, Multiple models deployed and in use**

TIME REQUIRED: **Months for each deployment**

What's the context?

Wikipedia is an online encyclopedia created and maintained by a global network of volunteers. This network contains multiple communities, typically grouped around the different language versions of Wikipedia. All popular content creation and sharing platforms require some form of content moderation to cover a range of functions including quality control and adherence to community values. The larger the community, the more difficult it is to monitor and regulate content manually in real time. Automated tools for content moderation help to streamline this process. These AI systems are often developed and integrated into the platform independently by internal or external technical teams. The ORES system is an alternative participatory approach designed by the technical team at Wikipedia to ensure that the content moderation systems could be adapted by each language community.

Who is driving participation and why is it being used?

The participatory model was developed by an engineering team at Wikipedia that provides multiple machine learning solutions for the Wikipedia community. They developed ORES to allow different language communities to adapt an automated content moderation system to their specific needs. Different language communities of Wikipedia have different sets of values and standards with respect to quality and content moderation. A more participatory model for designing and adapting an AI system was developed to ensure that the model served the needs of each specific language community and to encourage its integration, acceptance and usefulness.

What and how of participation?

Problem selection, data collection and curation, algorithm development

The ORES process offers multiple touchpoints for the community to provide input and feedback throughout the AI development process. Each new request for an ORES model is a collaboration between the engineering team and the community requesting the algorithm. Using several rounds of iteration, they define and scope the proposed model, collect and curate labelled training data, and suggest salient features for the model to use in its predictions. The engineering team also helps the community members think about how they might integrate ORES scores within existing or new processes they already use to update and monitor content.[82]

Deployment and post-deployment
After a model has been deployed, the community members are able to continue adapting the outputs of the AI so that it better matches their needs through two processes: they can adjust the features that the algorithm prioritises[83] and change the confidence thresholds[84] of the model to better match their own quality assurance processes or operational constraints.

Language communities with deployed ORES models also play an active role in governing the use of the algorithm. When Spanish language Wikipedians spotted that some of the results given by the ORES model did not align with their community standards, they crowdsourced an audit of the model and reached a consensus decision to block the use of the algorithm until it could be fixed.

What is the impact on the AI and the problem being addressed?

The close collaboration between the engineers and the community members throughout the design of the model helps to ensure that the content moderation models used on Wikipedia correspond to community values and priorities. The additional tight feedback loop following deployment allows any issues to be identified and addressed early, before they deter uptake. This has resulted in the continuous improvements of the model, including changing or removing features that were identified as biased against certain groups. So far, 110 classifiers have been developed for 44 of Wikipedia's language communities.

Overview of other participatory case studies referenced throughout the report

We only include here the case studies where we have not provided descriptions in the body of the briefing paper.

CATS4ML: is a crowdsourcing challenge that invited participants to submit adverse images not currently covered by benchmark training datasets. They called these adverse images AI blindspots. cats4ml.humancomputation.com/overview

Masakhane: is a grassroots community of practice made up of researchers, volunteers and students whose mission is to strengthen and spur natural language processing research in African languages. Their aim is to create new publicly available benchmark and resources to support African languages. www.masakhane.io

ORBIT: is an ongoing project to create a large image dataset for personalised AI-driven object recognition tools for visually impaired people. Researchers are working together with visually impaired users to create novel datasets that can be used to train algorithms developed directly for visually impaired users. orbit.city.ac.uk

Rehumanising crowdsourcing: a framework that allocates microtasks considering human factors of workers such as demographics and compensation. The framework routes microtasks to workers based on demographics and appropriate pay, to help mitigate biases in the contributor sample and increases the hourly pay given to contributors. dl.acm.org/doi/fullHtml/10.1145/3290605.3300773

Endnotes

Click on endnote number to go back to source page

1. Humanitarian Action. WEF Strategic Intelligence https://intelligence.weforum.org Accessed 27 March 2021
2. Metcalfe-Hough, V., Fenton, W., Willitts-King, B. and Spencer, A. (2021) The Grand Bargain at five years: an independent review. HPG commissioned report. London: ODI (https://odi.org/en/publications/the-grand-bargain-at-five-years-an-independent-review)
3. Arendt-Cassetta, L. (2021) From Digital Promise To Frontline Practice: New And Emerging Technologies In Humanitarian Action. Geneva: OCHA. (https://www.unocha.org/sites/unocha/files/OCHA%20Technology%20Report.pdf)
4. A 2020 review of predictive analytics in the humanitarian sector identified 49 initiatives. Hernandez, K. and Roberts, T. (2020). Predictive Analytics in Humanitarian Action: a preliminary mapping and analysis. K4D Emerging Issues Report 33. Brighton, UK: Institute of Development Studies.
5. Berditchevskaia, A., Peach, K., Gill, I., Whittington, O., Malliaraki, E., and Hussein, N. (2021). Collective crisis intelligence for frontline humanitarian response.
6. Deparday, V., Gevaert, C. M., Molinario, G., Soden, R., & Balog-Way, S. (2019). Machine Learning for Disaster Risk Management. http://documents1.worldbank.org/curated/en/503591547666118137/pdf/133787-WorldBank-DisasterRiskManagement-Ebook-D6.pdf
7. Crawford, K., & Finn, M. (2015). The limits of crisis data: Analytical and ethical challenges of using social and mobile data to understand disasters. GeoJournal, 80(4), 491–502. https://doi.org/10.1007/s10708-014-9597-z
8. Kulynych, B., Madras, D., Milli, S., Raji, I. D., Zhou, Angela, & Zemel, R. (2020). Participatory Approaches to Machine Learning.
9. Kate Crawford, Meredith Whittaker, Madeleine Clare Elish, Solon Barocas, Aaron Plasek, Kadija Ferryman, 2016 'The AI Now Report: The Social and Economic Implications of Artificial Intelligence.' https://artificialintelligencenow.com/media/documents/AINowSummaryReport_3.pdf
10. Sloane, M., Moss, E., Awomolo, O., & Forlano, L. (2020). Participation is not a Design Fix for Machine Learning. ArXiv:2007.02423 [Cs]. http://arxiv.org/abs/2007.02423
11. Russell, Stuart J. (Stuart Jonathan). (2010). Artificial intelligence: a modern approach. Upper Saddle River, N.J.: Prentice Hall
12. Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. Science and Engineering Ethics, 26(4), 2141–2168. https://doi.org/10.1007/s11948-019-00165-5
13. Bratteteig, T., & Verne, G. (2018). Does AI make PD obsolete?: Exploring challenges from artificial intelligence to participatory design. https://doi.org/10.1145/3210604.3210646
14. Lee, M. K., Kahng, A., Kim, J., Yuan, X., Chan, A., Lee, S., Procaccia, A., Kusbit, D., See, D., Nooth-Igattu, R., & Psomas, A. (2019). WeBuildAI: Participatory Framework for Algorithmic Governance. Proceedings of the ACM on Human-Computer Interaction, 3. https://doi.org/10.1145/3359283
15. IDEO.org. Field Guide to Human Centered Design. 1st Edition. (2015). https://d1r3w4d5z5a88i.cloudfront.net/assets/guide/Field%20Guide%20to%20Human-Centered%20Design_IDEOorg_English-0f60d33bce6b870e7d80f9cc1642c8e7.pdf Accessed 27 July 2021
16. Gasson, S., 'Human-Centered Vs. User-Centered Approaches to Information System Design', The Journal of Information Technology Theory and Application (JITTA), 5:2, 2003, 29-46.
17. Friedman, B., Kahn, P. H., Borning, A., & Huldtgren, A. (2013). Value Sensitive Design and Information Systems. In N. Doorn, D. Schuurbiers, I. van de Poel, & M. E. Gorman (Eds.), Early engagement and new technologies: Opening up the laboratory (Vol. 16, pp. 55–95). Springer Netherlands. https://doi.org/10.1007/978-94-007-7844-3_4
18. Umbrello, S., & van de Poel, I. (2021). Mapping value sensitive design onto AI for social good principles. AI and Ethics. https://doi.org/10.1007/s43681-021-00038-3
19. Interactive Machine Learning: Leveraging Human Intelligence. (n.d.). Microsoft Research. https://www.microsoft.com/en-us/research/video/interactive-machine-learning-leveraging-human-intelligence/ Accessed 27 July 2021.
20. Monarch, R. (Munro). (2021). Human-In-The-Loop Machine Learning: Active Learning and Annotation for Human-Centered AI. Manning. https://www.manning.com/books/human-in-the-loop-machine-learning
21. Voinov, A., & Bousquet, F. (2010). Modelling with stakeholders. Thematic Issue – Modelling with Stakeholders, 25(11), 1268–1281. https://doi.org/10.1016/j.envsoft.2010.03.007. See also https://participatorymodeling.org/.
22. The ACM FAccT Conference attracts a cross-disciplinary academic community and has been one of the highest-profile forums to host critical discussions about the design, development and implementation of AI systems. https://facctconference.org
23. This approach risks missing important considerations about the multiple choices that are made throughout AI development as well as the continued evolution of AI systems post-deployment. Bratteteig, T., & Verne, G. (2018). Does AI make PD obsolete? Exploring challenges from artificial intelligence to participatory design. Proceedings of the 15th Participatory Design Conference: Short Papers, Situated Actions, Workshops and Tutorial – Volume 2, 1–5. https://doi.org/10.1145/3210604.3210646
24. Several academics have proposed theoretical frameworks to guide such processes. For example the Society-in-the-loop paradigm of Rahwan, I. (2017). Society-in-the-Loop: Programming the Algorithmic Social Contract. ArXiv:1707.07232 [Cs]. https://doi.org/10.1007/s10676-017-9430-8 or community based system dynamics in Prabhakaran, V., & Martin, D. (2020). Participatory Machine Learning Using Community-Based System Dynamics. Health and Human Rights, 22(2), 71–74.
25. Gray, M. L., & Suri, S. (2019). Ghost work: How to stop Silicon Valley from building a new global underclass.

26. Irani, L.C., & Silberman, M. S. (2013). Turkopticon: interrupting worker invisibility in amazon mechanical turk. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13), 611-620. New York: ACM. doi: 10.1145/2470654.2470742
27. This has also been described as 'participation as work' in Sloane, M., Moss, E., Awomolo, O., & Forlano, L. (2020). Participation is not a Design Fix for Machine Learning. ArXiv:2007.02423 [Cs]. http://arxiv.org/abs/2007.02423
28. Crawford, K. (2021). The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. New Haven; London: Yale University Press. doi:10.2307/j.ctv1ghv45t
29. Berditchevskaia, A., Baeck, P. (2020). The future of minds and machines. London: Nesta. (https://www.nesta.org.uk/report/future-minds-and-machines/7-ci-better-ai-fifth-interaction/#content)
30. Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., & Shirk, J. (2009). Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. BioScience, 59(11), 977–984. https://doi.org/10.1525/bio.2009.59.11.9
31. Haklay, M., 2013, Citizen Science and Volunteered Geographic Information – overview and typology of participation in Sui, D.Z., Elwood, S. and M.F. Goodchild (eds.), 2013. Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice. Berlin: Springer. pp105-122 DOI:10.1007/978-94-007-4587-2_7.
32. https://www.adalovelaceinstitute.org/project/citizens-biometrics-council
33. https://www.odbproject.org
34. Lewis, T., Gangadharan, S. P., Saba, M., Petty, T. (2018). Digital defense playbook: Community power tools for reclaiming data. Detroit: Our Data Bodies.
35. UNHCR Innovation Services. A goat story. (2019) https://medium.com/unhcr-innovation-service/a-goat-story-3ed6bdd2b237 Accessed 27 July 2021
36. Common Voice in Feature on AI and Collective Intelligence Case Studies. https://www.nesta.org.uk/feature/ai-and-collective-intelligence-case-studies/common-voice Accessed 27 July 2021. See also https://commonvoice.mozilla.org/en/about
37. Albuquerque, J. P. de, & Almeida, A. A. de. (2020). Modes of engagement: Reframing 'sensing' and data generation in citizen science for empowering relationships. In Toxic truths. Manchester University Press. https://doi.org/10.7765/9781526137005.00028
38. Haklay, M. (2018). Participatory citizen science. In Haklay M., Hecker S., Bowser A., Makuch Z., Vogel J., & Bonn A. (Eds.), Citizen Science: Innovation in Open Science, Society and Policy (pp. 52-62). London: UCL Press. Retrieved August 6, 2021, from http://www.jstor.org/stable/j.ctv550cf2.11
39. ibid
40. Sloane, M., Moss, E., Awomolo, O., & Forlano, L. (2020). Participation is not a Design Fix for Machine Learning. ArXiv:2007.02423 [Cs]. http://arxiv.org/abs/2007.02423
41. Imran, M., Alam, F., Qazi, U., Peterson, S., & Ofli, F. (2020). Rapid Damage Assessment Using Social Media Images by Combining Human and Machine Intelligence. ArXiv:2004.06675 [Cs]. http://arxiv.org/abs/2004.06675
42. Ben Green. 'Good' isn't good enough,' in NeurIPS Joint Workshop on AI for Social Good. 2019.
43. Eitzel, M. V., Solera, J., Hove, E. M., Wilson, K. B., Ndlovu, A. M., Ndlovu, D., Changarara, A., Ndlovu, A., Neves, K., Chirindira, A., Omoju, O., Fisher, A. C., & Veski, A. (2021). Assessing the Potential of Participatory Modeling for Decolonial Restoration of an Agro-Pastoral System in Rural Zimbabwe. Citizen Science: Theory and Practice, 6(1), 2. https://doi.org/10.5334/cstp.339
44. Short-term impacts were characterised as those that had an immediate impact on the research and model development, medium-term impacts were the changes to land-use policy and management in the months to years that followed and long-term impacts were changes in on-the-ground land use, assessed up to four years after the participatory modelling.
45. Eitzel, M. V., Solera, J., Hove, E. M., Wilson, K. B., Ndlovu, A. M., Ndlovu, D., Changarara, A., Ndlovu, A., Neves, K., Chirindira, A., Omoju, O., Fisher, A. C., & Veski, A. (2021). Assessing the Potential of Participatory Modeling for Decolonial Restoration of an Agro-Pastoral System in Rural Zimbabwe. Citizen Science: Theory and Practice, 6(1), 2. https://doi.org/10.5334/cstp.339
46. Sendak, M., Elish, M., Gao, M., Futoma, J., Ratliff, W., Nichols, M., Bedoya, A., Balu, S., & O'Brien, C. (2019). 'The Human Body is a Black Box': Supporting Clinical Decision-Making with Deep Learning. https://arxiv.org/pdf/1911.08089
47. Sloane M, 2020, Participation Washing could be the next dangerous fad in AI https://www.technologyreview.com/2020/08/25/1007589/participation-washing-ai-trends-opinion-machine-learning Accessed July 2021
48. https://media.nesta.org.uk/documents/Collective_Intelligence_A_checklist_for_the_public_sector.pdf
49. https://www.weforum.org/agenda/2018/12/why-do-humanitarian-crises-disproportionately-affect-women
50. Facts and Figures, ICT 2017 https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2017.pdf Accessed 22 July 2021.
51. Women are 10 per cent less likely to own a mobile phone than a man in LMICs, and these numbers vary significantly by geography. Akhmatova, D.-M., & Akhmatova, M.-S. (2020). Promoting digital humanitarian action in protecting human rights: Hope or hype. Journal of International Humanitarian Action, 5(1), 6. https://doi.org/10.1186/s41018-020-00076-2
52. Albuquerque, J. P. de, & Almeida, A. A. de. (2020). Modes of engagement: Reframing 'sensing' and data generation in citizen science for empowering relationships. In Toxic truths. Manchester University Press. https://doi.org/10.7765/9781526137005.00028
53. For example the Civic Tech Graveyard documents the civic technology projects and tools that are no longer active. https://civictech.guide/graveyard
54. Cornwall, A. (2008). Unpacking 'participation': models, meanings and practices. Community Development Journal 43 (3), 269 – 283.
55. Bratteteig, T., & Verne, G. (2018). Does AI make PD obsolete?: Exploring challenges from artificial intelligence to participatory design. https://doi.org/10.1145/3210604.3210646
56. Berditchevskaia, A., Peach, K., Gill, I., Whittington, O., Malliaraki, E., and Hussein, N. (2021). Collective crisis intelligence for frontline humanitarian response. www.nesta.org.uk/report/collective-crisis-intelligence-humanitarian-frontline-response
57. Core Humanitarian Standard on Quality and Accountability. (2018). In United Nations, World Humanitarian Data and Trends 2017 (pp. 40–41). UN. https://doi.org/10.18356/4a4b59ce-en

58. Crawford, K. (2021). The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. New Haven; London: Yale University Press. doi:10.2307/j.ctv1ghv45t
59. Arendt-Cassetta, L. (2021) From Digital Promise To Frontline Practice: New And Emerging Technologies In Humanitarian Action. Geneva: OCHA. (https://www.unocha.org/sites/unocha/files/OCHA%20Technology%20Report.pdf)
60. Akhmatova, DM., Akhmatova, MS. Promoting digital humanitarian action in protecting human rights: hope or hype. Int J Humanitarian Action 5, 6 (2020). https://doi.org/10.1186/s41018-020-00076-2
61. Martin Jr., D., Prabhakaran, V., Kuhlberg, J., Smart, A., & Isaac, W. S. (2020). Participatory Problem Formulation for Fairer Machine Learning Through Community Based System Dynamics. ArXiv:2005.07572 [Cs, Stat]. http://arxiv.org/abs/2005.07572
62. Berditchevskaia, A., Peach, K., Gill, I., Whittington, O., Malliaraki, E., and Hussein, N. (2021). Collective crisis intelligence for frontline humanitarian response. www.nesta.org.uk/report/collective-crisis-intelligence-humanitarian-frontline-response
63. There are several frameworks for responsible data use in the humanitarian sector: OCHA's Data Responsibility Guidelines, the ICRC Handbook on Data Protection (chapter on Data Analytics and Big Data), and USAID's Considerations for Using Data Responsibly provide fundamental foundations on how to work responsibly with humanitarian data.
64. UN General Assembly, Roadmap for Digital Cooperation: Implementation of the Recommendations of the High-Level Panel on Digital Cooperation. Report of the Secretary-General, UN Doc. A/74/82, 29 May 2020 (Secretary-General's Roadmap), para. 6, available at: https://undocs.org/A/74/82
65. See Example 1: Assessing Earthquake Impacts in https://opendri.org/wp-content/uploads/2021/05/ResponsibleAI4DRM.pdf
66. Crawford, K., & Finn, M. (2015). The limits of crisis data: Analytical and ethical challenges of using social and mobile data to understand disasters. GeoJournal, 80(4), 491–502. https://doi.org/10.1007/s10708-014-9597-z
67. Wright, J. and Verity, A., 2020. Artificial Intelligence Principles For Vulnerable Populations in Humanitarian Contexts. Digital Humanitarian Network.
68. Given a limited number of training data, feature extraction through either automatic or expert-based approaches becomes difficult, as irrelevant features can affect the model. Recent studies show that crowdsourcing-based approaches can discover human-understandable features that significantly improve the ML process. The key insight behind this is that crowds of people are better at generating ideas and labelling subtle behavioural signals than machines.
69. Responsible Artificial Intelligence for Disaster Risk Management. (2021, April 28). OpenDRI. https://opendri.org/resource/responsible-artificial-intelligence-disaster-risk-management
70. AI for humanitarian action: Human rights and ethics. (n.d.). International Review of the Red Cross. Retrieved 6 June 2021, from http://international-review.icrc.org/articles/ai-humanitarian-action-human-rights-ethics-913
71. Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumeé III, H., & Crawford, K. (2018). Datasheets for Datasets. ArXiv:1803.09010 [Cs]. http://arxiv.org/abs/1803.09010
72. Bratteteig, T. and Verne, G., 2018, August. Does AI make PD obsolete? Exploring challenges from artificial intelligence to participatory design. In Proceedings of the 15th Participatory Design Conference: Short Papers, Situated Actions, Workshops and Tutorial-Volume 2 (pp. 1-5).
73. Humanitarian Technologies Project. (n.d.). Humanitarian Technologies Project. Retrieved 6 June 2021, from http://humanitariantechnologies.net
74. Peer Review Framework - Humanitarian Data Exchange, 2020. https://data.humdata.org/dataset/2048a947-5714-4220-905b-e662cbcd14c8/resource/76e488d9-b69d-41bd-927c-116d633bac7b/download/peer-review-framework-2020.pdf Accessed July 2021.
75. Heeks, R., & Shekhar, S. (2019). Datafication, development and marginalised urban communities: An applied data justice framework. Information, Communication & Society, 22(7), 992–1011. https://doi.org/10.1080/1369118X.2019.1599039
76. Taylor L. What is data justice? The case for connecting digital rights and freedoms globally. Big Data & Society. December 2017. doi:10.1177/2053951717736335
77. McQuillan, D. Towards Anti Fascist AI. https://www.opendemocracy.net/en/digitaliberties/towards-anti-fascist-ai Accessed 27 July 2021.
78. Soden, R., Wagenaar, D., Tijssen, A. Responsible AI For Disaster Risk Management, Working Group Summary. Geneva: World Bank. https://opendri.org/wp-content/uploads/2021/06/ResponsibleAI4DRM.pdf
79. UNHCR Innovation Services. Is it possible to predict forced displacement? (2019) https://medium.com/unhcr-innovation-service/is-it-possible-to-predict-forced-displacement-58960afe0ba1. Accessed 27 July 2021. UNHCR Innovation Services. A goat story. (2019) https://medium.com/unhcr-innovation-service/a-goat-story-3ed6bdd2b237 Accessed 27 July 2021.
80. Lee, M. K., Kim, J. T., & Lizarondo, L. (2017). A Human-Centered Approach to Algorithmic Services: Considerations for Fair and Motivating Smart Community Service Management that Allocates Donations to Non-Profit Organizations. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 3365–3376. https://doi.org/10.1145/3025453.3025884
81. The process described in this case study was analysed through the lens of value-sensitive design after it had occurred.
82. Halfaker, A., & Geiger, R. S. (2020). ORES: Lowering Barriers with Participatory Machine Learning in Wikipedia. ArXiv:1909.05189 [Cs]. http://arxiv.org/abs/1909.05189
83. This is known as the dependency injection feature.
84. This is known as the threshold optimisation feature.

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