Report: The risks and opportunities of AI on humanitarian action

Wednesday 15 – Friday 17 May 2024

Report:
The risks and opportunities of AI on humanitarian action

Wednesday 15 – Friday 17 May 2024

In partnership with
The Foreign, Commonwealth and Development Office

AI will have profound implications for humanitarians. If its potential is realised, it could help address chronic issues that have hindered effective and accountable humanitarian action. AI could transform how humanitarian actors coordinate, how decision-makers access critical information, or how accountability is provided to communities affected by humanitarian disasters.

Collaboration amongst humanitarian actors and with industry experts is key to realising a positive vision for AI. The inevitable use of AI will increasingly shape all parts of the humanitarian system as organisations rely on AI-powered systems to drive efficiency gains. Many challenges will therefore be shared and tackling them jointly will deliver better outcomes.

The recent UK AI Safety Summit provided a forum for industry and governments to come together to build consensus on how AI can safely be used for good.

While the application of AI on humanitarian action is still in a formative period, it is important to shape the use and development of AI to be consistent with humanitarian ethics, principles, and standards and determine pathways for further collaboration.

Executive summary

This Wilton Park meeting in May 2024 brought together participants from local NGOs, INGOs, industry, academic institutions, private sector, and governments to discuss the impact of AI in humanitarian contexts, how it can be harnessed, and how potential harms to vulnerable populations could be addressed. Throughout the discussions, various recurrent themes emerged which should help frame forthcoming conversations on AI and humanitarian action.

"AI will fundamentally alter every aspect of our lives. Indeed, it is already having an impact."

AI potential

AI carries huge potential for the effective delivery of humanitarian aid to greater numbers of people at a time when humanitarian needs are growing, and resources are unable to meet the current demand.

However, the risks and potential harms that AI can bring are cause for alarm including exacerbation of conflict and inequalities, erosion of trust in information, governance processes, and the humanitarian system itself, and undermining of social cohesion.

Humanitarians will need to proactively chart the right approach if they are to take advantage of the benefits, while comprehensively addressing the risk of harm.

"We are the ones figuring out what we're going to do with these incredibly powerful and transformative tools."

The importance of collaboration

Collaboration was a key theme throughout the meeting. Sharing knowledge and learning about AI applications, and ways of working with tech companies, governments and communities was seen as increasingly essential. Participants were keen to consolidate and use case studies and share experiences including failures. They worked together as partners with a common goal to classify lessons learned, avoid duplication, and foster collaboration. As well as harnessing lessons learned, collaboration would enable a more strategic approach to overcoming challenges which could shape a broader sectoral approach. Socialising definitions, building upon existing resources, and creating shared platforms for collective knowledge are relatively easy actions to implement when organisations choose to work together.

Greater transparency and collaboration can also mitigate a trend of increased polarisation on AI, sometimes characterised as a ‘silver bullet’, while others felt a need to ‘close the gate’ on the runaway advances of AI. Charting the course to a future that takes advantage of technology safely to support and empower the most vulnerable people requires careful reflection of a range of perspectives. The humanitarian sector and technology groups need to convene more productive conversations, moving from raising broad concerns to taking practical steps.

A people-centred approach

The voice, participation, and empowerment of crisis- and conflict-affected populations is paramount for effective humanitarian action. Local action and local innovation are highly valued in the humanitarian sector, while at the same time being notoriously difficult to support and implement.

“This is about people, and people must be at the centre of our approaches and conversations.”

The co-creation of AI (participatory AI) with front line responders, tech developers, local actors, and communities is important to meet real needs and to mitigate risks and harms. A people-centred approach can help identify the most appropriate uses of AI, reduce bias and harm from AI systems, remove culturally insensitive inputs, and specify guardrails for when, where, and with whom AI tools are appropriate. This approach, ideally delivered through local talent, can also enhance trust among users.

A strong business model for AI in humanitarian contexts would prioritise locally-identified needs and humanitarian principles and envision how AI could solve problems alongside identifying underlying economic incentives, and sustainability concerns.

However, implementing this approach can be challenging.

Digital public humanitarian infrastructure

A digital public infrastructure for high quality data sharing and interoperability is critically important. Systems and infrastructure for using AI in humanitarian settings are necessary and should be prioritised with long-term investment to ensure viability, sustainability, safety, and effectiveness long into the future.

> "There is vast potential for this technology to truly transform our collective humanitarian work."

High quality data is imperative for effective AI, and yet obtaining necessary local data can be challenging; it can be is expensive, can require community knowledge and data sets that do not exist.

Data selection and interoperability are crucial, and standardisation of data through repositories such as the Humanitarian Data Exchange (HDX) could drive interoperability and support better understandings of the limitations of data.

Concerns about data are multiple and interlinked, especially in the context of vulnerable communities that lack AI literacy. Consent is insufficient in the context of the risks of AI when implications of ownership and use of data by hostile actors pose serious threats to security.

Safety, ethics, and governance

When humanitarian actors experiment with AI tools and models, action should be taken responsibly and be consistent with humanitarian principles.

Humanitarians can look to other sectors to guide the assurance of safety. These include new standards that support the safe development of AI tools and a range of evaluation and assurance approaches. The humanitarian sector needs to consider how new regulations on AI fit with other existing national, regional, and global regulations and remain firmly within the context of vulnerable people who need humanitarian aid.

Managing risks within programmes and approaches, and in relationships with actors with different value systems, requires an examination of ethics and applications of existing rules and regulations. Moral underpinnings for humanitarian action have produced digital ethics in the past and knowledge from this can be harnessed and applied to AI.

“We are at a pivotal point for steering the world towards the responsible and inclusive use of AI.”

The private and humanitarian sectors are fraught with tension in relation to values, principles, and motivations around AI. Both will play an essential part in the safe use of AI in humanitarian crises. Developing relationships so that AI tools are shaped by expertise from both groups will become increasingly important.

AI Capacity

A central question for humanitarians is how to strengthen capacity on AI. A better understanding of opportunities and risks across different dimensions is needed.

Senior decision-makers, technical staff, and operational staff within organisations (local and national NGOs, governments, and international organisations) need greater understanding of both how their organisation can deliver differently and what should not change.

This should be accompanied by a requirement for technical experts to enhance their understanding of humanitarian action and principles. As well as tech organisations, humanitarians need to work with academics, regulators, and policy makers more generally to ensure that humanitarian priorities contribute to the wider conversations.

1. The state of play

“When we talk about AI, think of humanitarian principles of humanity, neutrality, impartiality, and independence.”

Participants discussed the current trends in AI technological progress, various AI models, the pace of change, regulatory spaces, and emerging risk-based approaches with implications for the humanitarian sector.

The humanitarian sector is facing some of its biggest challenges in 2024, with 293 million people estimated by the UN to need humanitarian assistance and protection. UN support is only targeted to reach 60% of them. Despite more than 12,000 organisations responding to humanitarian crises across the globe, the sector's overall capacity falls short.

With the humanitarian sector operating under huge resource constraints, the search for effective and innovative solutions inevitably turns to technology. AI offers the potential to help address this gap and unlock rapid progress. It can be used to predict when crises will happen and what the impact is likely to be on populations. It can potentially improve the breadth and depth and effectiveness of humanitarian responses.

The humanitarian sector is built on principles of humanity, neutrality, impartiality, and independence. Currently, humanitarian, development, peacebuilding and human rights organisations are staking out positions on AI, its application and governance, and their perspectives on shared global ownership.

AI is being applied in different ways in the humanitarian sector, for example, to predict natural disasters, displacement, famine, and air strikes, to identify crop pests, and to provide support to vulnerable people through chatbots.

While development in AI carries huge potential benefits to transform the humanitarian sector, participants identified major risks, challenges, and societal consequences in this uncertain space. Questions emerged around how AI might be developed and used by a range of stakeholders including malign actors; whether the humanitarian sector has enough resources to understand and harness AI; and how to build AI models that focus on putting populations who are often regarded as peripheral to humanitarian and development efforts.

Agreements with technology companies and humanitarian organisations already exist, yet there are major concerns around data storage, ownership, and use. Tension exists between collecting and owning data and protecting humanitarian principles.

“We don’t know how to be as effective as we can while also being responsible.”

How can the sector be as effective as possible to meet its humanitarian goals through AI, while also acting responsibly? The principle of ‘first do no harm’ is important. Questions emerge around the use of AI and human rights and risk, for example, the rights of populations at risk of harm, the right to security, and the right to freedom of speech. The conflict of rights is exacerbated by AI.

Governance and regulation

Governance and regulation of AI in the humanitarian sector was a major concern. It is important to get the language right around regulation, including terms like ‘interested parties’, ‘customers’, and ‘affected persons’ but the humanitarian sector uses terms like ‘beneficiaries’ and ‘recipients’. The position of people trying to survive a major disaster is not that of a ‘customer’ or ‘interested party’.

A range of global and regional regulations exist which provide standards for best practices such as EU regulations, General Data Protection Regulation (GDPR), and the new EU AI Act that sets out a range of risks from high to low. This is not therefore a legal vacuum, and the humanitarian sector needs to consider how any new regulations on AI can square with what already exists nationally and globally. However, caution is needed as GDPR, for example, has customer-focused incentives rather than considering highly vulnerable populations.

The moral underpinnings for humanitarian action have produced digital ethics in the past, and, in today’s climate, ethics and regulation for AI in the humanitarian field are urgently needed. Industry standards can be helpful and shape production and procurement.

Participants discussed governance and regulation with a view that there is no one single framework to address the problem or meet the varied outcomes the sector is seeking to achieve.

In this initial discussion, key themes of data, assurance and ethics, and governance and regulation emerged, to be explored further throughout the meeting.

2. The humanitarian world in 2030

"We have a window of opportunity to shape the future trajectory of AI in the humanitarian sector. But this window has already narrowed since 2021."

Participants worked in small groups to analyse three different scenarios that set out states for how AI might impact humanitarian action by 2030. This session was created and facilitated in collaboration with The Government Office for Science.

The scenarios informed a discussion on the potential impact that different opportunities and risks may have on humanitarian delivery and the actions that are most likely to steer humanity towards a positive future.

Wild West of AI

In the groups that discussed the scenario of a 'Wild West of AI', concerns focused inevitably on the harms caused by the proliferation of unregulated and ungoverned AI including misinformation, erosion of trust, and creation or escalation of conflict and war. Duplicate and unused solutions to problems would lead to inefficiency and wastage of precious resources. Ownership, control, and ability to validate AI and its impact could easily fall into the hands of malign actors.

"We must look at AI from a local, community-oriented perspective."

However, opportunities are huge with a catalogue of potential solutions to use, learn, and build upon with new actors and alternative power dynamics emerging. This relies upon sector collaboration characterised by transparency, knowledge sharing, and reusable applications along with strong global governance and interoperability. An AI Global Compact could provide principles and a framework to guide a network of solutions, and effective collaboration with standards for evaluation.

In this scenario, participants questioned whether humanitarian principles are as central as they should be. What do these principles mean in the modern world of AI? Humanitarian actors and technology companies and infrastructure are at the mercy of huge power imbalances, and the humanitarian sector lacks the financial resources to obtain the most impressive AI tools.

Is there an opportunity for a new type of public partnership tailored to different risk positions, allowing the sector to reframe relationships? Can the sector explore open-source data in a marketplace that allows for joint ownership, and allows all humanitarian actors to access data equally?

"Conversations about safe and responsible AI are not held strongly enough by the humanitarian sector."

'AI On a Knife Edge'

Other groups discussed 'AI on a Knife Edge', a scenario in which Artificial General Intelligence (AGI) has been achieved, and although there are early developments in safety and regulation, the risks of harms are strong. In this scenario, local NGOs are using and relying on AI for multiple operations, AI labs are developing their own humanitarian assistance, and AGI can devise its own subgoals.

The key focus of discussion was on tensions between delivering humanitarian aid as efficiently as possible, and what it means to be human. Participants identified the activities that only humans could provide including relating to others through emotional intelligence, and the ability to interact with vulnerable communities in the ways that communities prefer. Although AI systems may support negotiation and conflict resolution, there remains an element of humanity that surpasses technology in these domains.

"There is a constant mention of the need for localisation and to be embedded in communities. Why is it so difficult to do?"

The use of AI in humanitarian back-office functions and processes efficiencies would free up staff time allowing them to engage more in human activities with populations; however, donor expectations might change, and people's time may not be funded in this area.

A major question was whether the use of AI is good for localisation and the decolonisation of aid. The sector is not good at listening to local voices. Could AI provide the tools to do this better? If so, how to ensure the co-creation of AI with local actors, tech developers, and communities? It is important to build a strong business model identifying the underlying economic incentives, and sustainability concerns, along with prioritisation of locally identified needs and humanitarian principles.

Safety was a major consideration, with suggestions of a categorisation of risks that mirrors the EU AI Act, and auditing to avoid harm and AI hallucinations later down the line.

AI disappoints

Other groups discussed a scenario where 'AI disappoints', and AI capacities have developed more slowly than expected. The humanitarian sector is disappointed with the results of using AI, with bad decisions hurting groups of vulnerable people. Donors are withdrawing funding.

Groups shared reflections that this scenario is a real possibility for AI in the humanitarian sector, with concerns that expectations are too high and that the sector should not allow the hype around AI to drive engagement. There are clear issues arising around capacity and capability in the sector, with concerns over safety, ownership, access, and the constraints of working with a private sector driven by revenue and profit.

It is important to build communities of practice in the sector including learning from failures and holding continuing dialogues among trusted stakeholders to make meaningful progress. There is a need to reduce and remove interagency competition as much as possible and create a joint donor fund for the development of AI practices and learning.

3. The ‘green shoots’ of AI use in humanitarian action

AI has the potential to allow humanitarians to do more work, more effectively, with far fewer resources. However, new technologies have previously been mistaken as a panacea, often resulting in false dawns and wasted resources. Participants discussed why AI was different to other technologies that have promised to transform humanitarian delivery; how AI is already being used in humanitarian action; what are the characteristics of existing good practice; and how can this help us understand where AI tools are likely to be most effective.

“This is not the first moment we've been in a global challenge of what to do about an emerging technology - far from it.”

A key message from the discussion was that although it is positive to experiment with generative AI to potentially alleviate human suffering, it is crucial to experiment responsibly and document and share learning openly about successes, failures, and challenges. Several current use cases were discussed.

Community-led critical information and participatory AI

The use of AI is currently being explored in partnership with NGOs in three countries to provide community-led critical information through digital tools, channels, and social media. The approach provides timely, trustworthy, and accurate information to allow people to make decisions on for example, where to get identity documents, and how to access health services. A human-intensive model ensures the information that people get is trustworthy, but a team is exploring the use of generative AI to scale it up and create personalised and contextualised information for people. Work is underway to de-risk the prototype and ensure safety to test it, measure results and then take it to clients. All results will be published as part of a global public good.

“Collective Crisis Intelligence (CCI) and Participatory AI offer a pragmatic way to ensure that humanitarian AI reflects humanitarian values. Why aren’t we investing in more of this?”

AI-driven chatbot for education in crises

Another example is an AI-driven chatbot platform that delivers personalized learning and education experiences to crisis-affected children which can operate at scale within 30 days of a crisis. It offers a chatbot to reach them on platforms they already use such as WhatsApp, SMS texting, and social media. Launched in Nigeria, the approach is now integrating ChatGPT. In Syria, the use of AI with caregivers has delivered strong early childhood development outcomes.

"We can influence the trajectory AI takes. We can develop AI that gives local communities agency and increases our accountability to them."

Collective crisis intelligence and participatory AI

A further example of AI is Collective Crisis Intelligence (CCI) which combines the collective intelligence of crisis-affected populations and frontline responders. The organisation wanted to address the risks from AI by giving those local frontline responders and crisis-affected communities a major role in shaping the design, development, and evaluation of the AI tools, and called this approach Participatory AI.

Evaluations from experiences in Nepal and Cameroon showed that these CCI approaches have the potential to make local humanitarian action more timely and importantly more appropriate, and responsive to local needs.

Secondly, Participatory AI methods helped reduce some of the potential for bias and harm from AI systems from identifying model blind spots, to removing culturally sensitive inputs to the model, to specifying guardrails for when and where an AI tool could and should not be used. Using Participatory AI approaches helped build trust in the AI tool, a critical factor to the adoption and success of AI in any environment. This was all achieved through using local data, local talent, and local infrastructure.

Citizen voice and ownership for response at scale

Experience in India revealed how AI gained information from the roofs of houses in urban neighbourhoods, using the roofs as a type of QR code to identify who is vulnerable in a natural emergency. An approach to gain citizens' voices and intelligence for a better humanitarian response created a ‘disaster wallet’ at household level, with households reporting on risks and impacts of disasters. These profiles were maintained on a platform that records what kind of assistance households might need.

Some lessons learnt from this experience are that it is important to partner with the government to create a global public good with both the government and the citizens owning the data. The data is not open and there is no commercial benefit or ability for the private sector to use it for profit. Actionable insights are built according to people’s adaptive capacity, aiming to restore the agency of people. In terms of designing the approach, the team designed the system with what works at scale, rather than scaling what works.

“We use AI in the most boring ways possible but it has improved the way we work massively.”

Geospatial map data

In Kenya, an NGO is working to adapt to changes in AI and amplify local community knowledge at the same time. It is focused on closing map data gaps, ensuring that lack of geospatial data is not a barrier to humanitarian response. Integration of road tracing, mobile mapping, and community voices aims to close the gaps. For example, the NGO is supporting the government in mapping a city using AI data, drone imagery mapping, and community validation to support government planning on drainage systems and disability access. It is challenging to integrate local knowledge into AI approaches, and it is done according to context with some projects using AI for 10% of the work, and others 80%. It is also hard to use AI in high conflict situations.

Predicting the likelihood of violence

In Kenya, another AI approach is being used to predict the likelihood of violence across three areas of the country. The model looks at what factors are associated with change to a phenomenon. Over what period, and with what actors? An inclusive approach and systematic engagement of lived experience in each context is crucial to the model and its predictive ability. It is important to innovate to capture and model frequently changing phenomena. The team tracked data over time for 56 towns and cities in Africa over 5 years, developed a model of depth, and triangulated this with a household survey. This made it possible to identify what phenomena are most associated with change such as conflict events, perception of conflict or change of commodity prices. This can then inform the project, its design, and activities.

In another example from a small island state, a team trained an AI model on the three main newspapers and how they reported on crime to identify and predict volumes of crime. This gave donors the level of confidence to act.

Monitoring displacement

A team in Switzerland monitors the displacement of populations through AI, which creates massive efficiencies when identifying needs in disasters. A team of data scientists use all their capacity on data cleaning and data management and applies AI to cut down research space. When disaster strikes, the team conducts rapid needs assessments with speedy processes and simple semantic searches, which gives a probabilistic return. The tool can be used within existing processes and gives huge returns in time efficiency. The team has also reclassified and recoded two decades of data to fit the new system, so there is historical depth to the model. Access to the data is available online with a more compatible system for larger agents.

"A better distinction needs to be made between scaling up what works versus what works at scale."

Discussion

Key points included:

A common difficulty is how to move beyond the innovation and pilot projects that donors are readily willing to fund, in order to move to larger-scale efforts which meet vast humanitarian needs but require greater donor commitment. Small-scale pilots are often not scalable in wider contexts. A better distinction needs to be made between scaling up what works versus what works at scale.

Suggestions included avoiding promoting only one tool to scale up, but rather thinking about contexts, systems, and building digital public infrastructures, open structures, and open hardware and software that allows lots of different organisations to take part, with the ability to tailor appropriately to local contexts.

Another suggestion to improve the pilot-to-scale pathway was to distribute the ability to solve, by creating a platform for other innovators and agencies to enter and contribute learning and solutions.

A further suggestion was to integrate AI into existing work, rather than creating a distinct project. One team identified a use case and hired staff to integrate it into their workflow. If something is repetitive and predictable, then AI is useful and appropriate. A good level of understanding of tools and capacities of how AI can be used in general and how people can apply it to their work will allow for better integration of AI, and agencies need to promote this internally. This approach might help move away from the 'plague' of pilot projects.

"AI can only figure out from the data - but if you as humans know what works, then you need to lean into that and forget about the hype."

Guardrails, safety procedures, and rules of how AI manages itself were of concern, and questions were asked about the prevention of hallucinations and other potential harms. The tech sector might not be opposed to AI models developing in a range of ways, whereas the humanitarian sector needs clear requirements and analysis of needs and purpose, along with sturdy barriers in place to prevent AI from becoming harmful, especially to vulnerable people.

Participants also expressed concerns over the privacy of data, and the balance between AI, human resources, and the moments when human interventions are required.

4. Creating the enabling environment for safe AI uptake

Participants discussed the constraining factors and enabling environment for safe AI uptake including infrastructure requirements such as access to data, AI models, procurement skills, and technical expertise.

Exchange and learning

A review on the use of AI in the humanitarian sector revealed that increasingly AI conversations are polarised (a ‘silver bullet’ versus the need to ‘close the gate’). Charting a meaningful future requires action to relinquish overly simplistic mental models. The humanitarian sector and technology groups need more productive conversations, moving from raising broad concerns to taking practical steps.

“Where does agency stand between human and the machine?”

Coordination around initiating pilots and ensuring less duplication and competition was a common area of discussion. Participants were keen to consolidate case studies, use cases, learning, and the sharing of experiences including failures, and coming together as partners with a common interest to classify lessons, avoid duplication, and work together.

Another problem is the lack of transparency. Beyond all the hype, it is difficult to understand the extent of all the pilot projects and who is doing what within a big picture perspective. Therefore, the same mistakes may be repeated time and time again.

Evidence-based AI in humanitarian contexts is a work in progress, and the sector needs to invest human time to understand how staff are interacting with AI and using it, along with an analysis of power and politics in its application.

Locally-led action and meeting needs

Digital divides remain a concern globally, with digital gaps (hardware and software) among generations, women and girls, and other intersectional communities. Protection and freedom from violence, along with human rights in a digital age are all serious considerations.

“Are we at risk of bringing in culture debt, process debt, and reproducing our own failings as we go forward?”

An AI literacy gap exists in local communities. Are local populations sufficiently knowledgeable to deal with the complexity of risks and biases around AI? AI can deepen divides, biases, exclusions, and censorship.

AI literacy, capacity, and capability within the humanitarian sector is a big issue. Digital literacy skills are difficult to find, and when the capacity is not in-house, how far and how fast are organisations falling behind? Ensuring due diligence without AI capacity in house is problematic.

The humanitarian sector still needs to ask the basic and fundamental question: what humanitarian challenges are AI appropriate for? A plethora of AI initiatives do not seem to be guided by meaningful engagement with communities.

Some participants voiced that the use of AI use in the humanitarian sector is inherently extractive. Organisations are extracting data from vulnerable people, systems are being designed by those in power, and little genuine engagement with communities is occurring.

The issue of consent and data collection at local level is a chronic problem, and general discussions around AI have criticised business models as data theft; in this case, the humanitarian sector has lessons to learn, and it is urgent and important to ensure community engagement and participatory co-design of data collection and use.

“Don’t expect tech companies to prioritise ethics. The humanitarian sector should be doing this.”

Data concerns

The importance of high-quality data for AI cannot be underestimated. Barriers and challenges to accessing or using quality data for AI in humanitarian contexts must be solved to enable good, efficient, and responsible AI.

One major challenge to this is that sometimes data does not exist. Many AI-powered tools will only work where local data is available. This requires community knowledge, and data that has not been collected.

Here a people-centred approach, to identify what data could be invisible to technologists who may be building *for* and not *with* communities is important. Definitions also matter in data collection, for example if a category for gender is binary, then anyone who does not identify in those categories will be invisible in the data.

Where data does exist organisations may not be willing to share it with others, and sometimes organisations may not have the capability to use the data that is available. Both are barriers.

Data selection and interoperability are crucial elements for consideration, as it is important to blend local and other data. A similar key is needed to match across data sets; however, data is not often standardized to enable this.

Data lives in many places, and with varying quality, and the use of data repositories such as the Humanitarian Data Exchange (HDX) for standardisation could drive interoperability and support better understanding of the limitations of the data. Data-sharing agreements are also critical, to preserve privacy and confidentiality and to agree gradients or variations of sharing, such as sharing raw data or sharing insights, which can mitigate risk.

Data for analysis needs to be locally relevant, representative, standardised, and of high quality. In exploring how generative AI relies upon such data, there appears to be relative opacity regarding what went into the training of the massive models, and it is hard to identify biases that are created at foundational stage.

Data labelling and annotation can introduce human biases at this stage, which can have negative impacts on populations. Synthetic data which can mitigate some issues like privacy, is not a panacea.

Working with the private sector

Humanitarian actors should be careful when adopting tools and supply chains from tech companies without a deeper understanding of where the data is from, what decisions were made, and therefore what the model can and cannot do. Then humanitarians can identify the parameters of use, and likely outcomes and restrictions.

A blended, harm-mitigation approach to understand inherent biases within models is important. A common view was that it is necessary to work with tech companies to explain their process and actions. Transparency and regulation should be enforced upon companies.

> “Informing ourselves in the humanitarian sector is vital. Listen to podcasts, read tech correspondence and support investigative journalism.”

Equitable outcomes may not emerge in the AI sector. Most of the world’s population does not have access to AI in their native language, and market pressures mean a lack of commercial viability. Small language models may be critical for equitable outcomes, and tech organisations have an important part to play here.

The humanitarian sector also faces the challenge of due diligence capabilities and compliance with standards across governments and the private sector. Tech providers who work with the humanitarian sector also work with governments on surveillance and the military, and the sector must ask whether this is a conflict of interest.

There are also geopolitical dimensions to AI, with different views among China, the US and the EU which forces the sector to think about AI in the concrete, not the abstract. For example, are there risks if a humanitarian actor relies on tools that suddenly become unavailable to them due to wider geo-political shifts?

It is important to have conversations about monitoring and evaluation, and why there are positive or negative outcomes. There is a huge gap between the tech developer and the end user, and a ‘black hole’ around transparency, accountability, and advocacy.

5. Risks when humanitarians use AI

Participants discussed understanding and mitigating risks and harms that may be exacerbated as humanitarian organisations increasingly rely on AI-powered tools.

In an environment where high-risk decisions need to be made on behalf of vulnerable people, AI has the potential to return errors which may be costly for people whose survival depends on meaningful and high value information.

“It is important we are not embedding and reinforcing inequalities”

It is necessary for the sector to better understand harm in a digital world: how do we measure it, how to go beyond the equivalent of physical harm, and what are primary and secondary harms? How can the humanitarian sector quantify this and put investment into managing the narratives the sector wants in relation to its organisations, and the communities it aims to assist and protect?

Sociocultural risks are often overlooked in the AI agenda, where models are insufficiently adapted to local conditions and narrow knowledge is applied in testing. This can lead to misuse and negative impacts during crises, or to the creation of models that are not generalisable when adapted elsewhere.

Some solutions proposed were to have greater stakeholder engagement throughout the AI project lifecycle, which is complicated with multiple moving parts including project design, and model and system development.

During project design, if the project is oriented to a particular social setting, there should be a focus on understanding goals and addressing problems. The model development must define the technical output, and management in the project setting. It is important to improve the quality of data, to select and evaluate the model within community participatory processes to ensure transparency. Operationalising the model means training users and seeking continuous feedback following implantation.

Seeking the 'consent' of vulnerable people to collect data does not always ensure fairness of access or ownership of data, because of the complexities around AI and the lack of AI literacy in communities. Data justice is a vital concept, with six pillars of power, access, participation, equity, identity, and knowledge which defines how data intersects with social justice. AI must be firmly contextualised within social justice, intersectionality, and global and intercultural considerations. It is about choices, and layers of governance.

"We should not be adopting tools and supply chains without understanding the level of ethics."

Participants worked in small groups examining case studies to further the discuss the risks and harms that may be exacerbated. This session was created and facilitated in collaboration with The Alan Turing Institute.
Key summary points from these discussions included:

- Poor data quality and data governance are major risks of harms to individuals in terms of poor outcomes and risks of inequalities in access to services.
- Poor data governance could expose certain groups to stigma, discrimination, or violence.
- The failure of AI tools can lead to a lack of trust among people and communities which have implications for wider humanitarian delivery.
- The social cohesion that occurs naturally when people come together may be dissipated when there is too much emphasis on technology.
- Overreliance on tools, especially in LLMs, may create greater risks than scaling back and ensuring the technology is fit for purpose.
- Scaling up AI models can lead to a loss of nuance and specificity at the local level. AI models may not meet actual needs.
- Relying on historical data does not always provide accurate predictions for the future and can lead to bias.

- Ambitious technology is hard to get off the ground, and it is easy to go nowhere discussing the issues associated with it.
- Lack of participation by communities along the entire AI model lifecycle is a problem.
- Technical challenges may arise when infrastructure goes down.
- More pressure may be put on the frontline worker who takes decisions based on the systems, but what happens when the system is not perfect or does not work? There could be fatigue with the AI model.
- Reputational damage to organisations and a waste of resources are a risk if AI models are unsuccessful.

> "Transparency in this space is the lowest possible bar and is critical for any subsequent action."

Ways to mitigate the risks

- Put in place a risk matrix and apply a decision framework to support and formalise decision making that includes the experience of frontline workers.
- Develop operational tools that help users to practically address risks, such as risk impact assessments.
- Use risk assurance tools to ensure models work as intended.
- Apply standards and 'how-to' guides with people who are engaged across the project cycle.
- Set up a good M&E framework with feedback loops, including community, patient, and service user groups.
- Build incentives for learning and adaptation into the governance framework.
- Compare the proposal for AI with a non-tech solution.
- Deploy the model in parallel with the existing system, which helps to test the counterfactual and not penalise people who are exposed to it.
- Build expertise among communities to allow them to engage meaningfully and advise.
- Aim to build concepts of agency who defines access and who facilitates participation into practical approaches.

- Create a community of practice around AI in humanitarian contexts with a peer review system and community-based advisory committees that have AI literacy.

6. Hostile actors

Hostile actors pose risks by exploiting AI technologies in humanitarian contexts, with negative impacts on the information space including cyberattacks, disinformation, and fraud as well as growing surveillance. Participants reflected on the increasing risks to humanitarians, the eroding of trust between humanitarian actors and affected populations as well as steps to mitigate potential harms.

AI opens easy opportunities for disinformation, criminal activity, and the sowing of social discord. It is now possible to deploy deeply sophisticated scams at scale; for example, LLMs can write 20 persuasive tweets in five minutes and fake content is easy to create.

"If a hostile actor's goal is to undermine trust in a humanitarian organisation, it could change lots of things - this is not science fiction."

The information environment is polluted, overloaded, and manipulated, particularly in conflict and crises. Trends are emerging of weaponizing civilians to harm others in the real world. Both state and non-state actors can use AI to push conspiracy theories and misinformation.

Trust and safety teams are being laid off on big platforms, content moderation has worsened, and underrepresented languages are not moderated at all.

One consultation in Kenya revealed that people were very concerned about audio scams during crises and had knowledge about the risks. How is the humanitarian sector supporting communities who are dealing with misinformation and scams? An analytical framework is needed that thinks about these risks from the perspective of the community. How does the sector seek useful information and intelligence about how to do this effectively from AI and tech experts?

It is necessary to reframe the problem. Disinformation is organised crime and criminals are making inroads into places that the humanitarian sector cares about deeply. Platforms and frameworks alone will not solve this problem.

7. Ways to take action

Participants generated a list of ideas for concrete steps to support humanitarian actors to take meaningful action on AI.

> “We must harness the full potential of AI, cooperating closely to ensure it acts for the good of all, but especially for the most vulnerable who are impacted by humanitarian crises.”

Provide advice to humanitarian organisations on AI

This includes an AI advice service with the following:

- Use cases and building understanding. What can AI do and how can it be supported? What can it not do?
- Mapping AI use cases to support greater transparency
- Documentation of AI pilots and lessons learned (failures as well as successes) which would be housed and shared by a permanent, neutral and independent institution
- Useful definitions like CDACs and Nesta's definition of participatory AI
- Upskilling and guidance for individuals and institutions to help them contextualise tools and understand the limits in certain contexts
- Approaches and tools that help organisations drill down into obstacles/drivers of problems to AI uptake (are they technical or sociocultural challenges?).

Procurement

- Market assessments and the pool of vendors need to be extended, therefore increasing the number of vendors and localise supply (e.g. Nigeria, Kenya, Bangladesh, Philippines)
- Establish a framework agreement and vetted list of AI vendors (suppliers and assurance firms) agreed by multiple donors
- Provide advice on use cases

- Provide advice on what a model will cost to build, run, and maintain
- Increase understanding of intellectual property and how this should shape agreements and contracts with AI providers. Who owns the model? Who owns the data? On which tech was it built? What does the full stack look like and how does that inform risk assessments linked to procurement?
- Support transparent approaches to due diligence decisions when working with tech organisations.

Digital public humanitarian infrastructure

- Improve information and data sharing: structured and model-ready data, priority data sets, greater incentives for data sharing
- Model infrastructure: incentivise model-sharing and transparency
- Create (coordinate) sandboxes for safe practice, most likely at country level: identify data and tools and explore in a safe way.

Relationships

- Relationships with affected populations: support participatory AI (use a multidisciplinary approach, learn from non-AI participatory work); map the local tech scene (related to procurement); establish fora to enable better co-design/co-development of algorithms between Computer Science (CS) / Machine Learning (ML) engineers and humanitarian experts from local context (drawing on Indaba and similar fora)
- Build awareness of issues across humanitarian and tech actors.

Hostile actors

- An 'IPC-like' tool for the information space.

Conclusions

At the closing of the meeting, a plea was made to take not only the easiest routes of using knowledge platforms to share lessons and collaborate in new ways, but to take serious steps to embed approaches in systems and interoperable structures that will live into the future. This is challenging but necessary and requires commitment and funding from donors.

Two general calls to action to those working within the humanitarian community also emerged:

- Firstly, to use global platforms to tell powerful stories about AI in the humanitarian sector, about the harms that can be prevented, and the innovation that can be applied to create more benefits than harms.
- Secondly, to draw on the wealth of experience to date so that the AI conversation talks about the very best of humanity. It is not the first time in global history that the emergence of new technology has presented deep challenges, and the humanitarian sector must learn from the past.

Alison Dunn

Wilton Park | August 2024

Wilton Park reports are brief summaries of the main points and conclusions of a conference. The reports reflect rapporteurs' personal interpretations of the proceedings. As such they do not constitute any institutional policy of Wilton Park nor do they necessarily represent the views of the rapporteur. Wilton Park reports and any recommendations contained therein are for participants and are not a statement of policy for Wilton Park, the Foreign, Commonwealth and Development Office (FCDO) or His Majesty's Government.

Should you wish to read other Wilton Park reports, or participate in upcoming Wilton Park events, please consult our website www.wiltonpark.org.uk.

To receive our monthly bulletin and latest updates, please subscribe to www.wiltonpark.org.uk/newsletter

Wilton Park is a discreet think-space designed for experts and policy-makers to engage in genuine dialogue with a network of diverse voices, in order to address the most pressing challenges of our time.

enquiries@wiltonpark.org.uk

Switchboard: +44 (0)1903 815020

Wilton Park, Wiston House, Steyning,
West Sussex, BN44 3DZ, United Kingdom

wiltonpark.org.uk