

If AI Can Do It Better, Why Am I Still Here?



**Humanitarian expertise
in the AI era**

September 2026

If AI Can Do It Better, Why Am I Still Here? Humanitarian Expertise in the AI Era

By: Alice Zanni ([LinkedIn](#)) and Andrej Verity ([LinkedIn](#)) | [Digital Humanitarian Network](#)

Executive Summary.

Generative AI (Gen AI) has been widely available for only a few years, and already it has transformed how we approach work. Most sectors have begun asking what this means for them, but the effort is unevenly spread, with labour economics and medicine far ahead of the humanitarian sector in the depth of their literatures.

This paper asks what it means to be a humanitarian expert in the Artificial Intelligence (AI) era, drawing on current literature and interviews with practitioners in field operations, institutional governance, regulation and technology.

Since humanitarian work offers no determined way for individuals to enter the sector, recognition tends to fall back on peer nomination, credentials and institutional standing. Yet these are precisely the qualities Gen AI produces best and does so whether the content is sound or not. Defining a humanitarian expert must therefore lie outside what that system can see. This paper positions it in “**tested participation**”: the judgement accumulated by making decisions whose consequences come back to you.

The gains from integrating AI into humanitarian practice are real but conditional. They cluster in drafting and formatting, and matter most where they move hours away from a screen and towards contact with affected populations. Against them stand three costs: cognitive surrender among over-stretched staff, prioritisation judgements settled upstream in design where impartiality cannot reach them, and an adoption gap that follows existing inequalities between well-resourced and local organisations. Human sign-off, the safeguard the sector relies on to contain all three, weakens under the same pressures, as reviewers' attention erodes and as generative systems leave them less to check against.

Roles are being redefined rather than replaced, and the paper closes with recommendations separated by who can act on them and how to do so in practice. Practitioners are asked to form their own view before asking a system for one, and to defend the field time their judgement is built from. On the other hand, team leads are asked to refuse the arrangement where the tool drafts and a person relay. Lastly,

institutions are asked to make AI literacy specific to each role, and to treat entry-level posts as the place future senior judgement comes from rather than as overhead.

Overall, relevance of the expert does not reside into a single position on adoption. The question is which tasks, with what safeguards, under what conditions, and whether the people doing the work are the ones answering it.

Key Findings

1. Expertise is built by deciding and living with the result.

Practitioners learn within a mission. What matters is not only being present but being the one who decides and then facing what follows, i.e. **tested participation**. This is the foundation everything else rests on.

2. The gain from AI is time aside for tasks with affected populations.

Staff in NGOs and in the field described, more than efficiency and speed; the real gain was to spend less time on a screen and more time with people.

3. AI follows the resources rather than levelling them.

Only 13 per cent of local-agency staff know of an AI policy, against 39 percent of UN agencies. Protected time, mentorship, and the standing to reject an output describe a well-resourced organization, so the gains are achieved where capacity exists, and the risks become reality where it does not.

4. "A human will check it" is a weaker safeguard than the sector assumes.

Sign-off happens even when the reading is only a glance. And a reviewer who expects the system to be right will read to confirm. Meanwhile, AI often gives a conclusion without showing what produced it, so even an attentive reviewer has little to check against.

5. Cutting entry-level posts is how the sector loses its future experts.

Junior tasks are where judgement is built, and they are the first to be automated. Doing so saves money now against a future cost of the senior judgement needed in the next ten years

6. Versatility, not prompting fluency, is what marks the expert in the AI era.

The basics of using AI can be learned in a few hours and are easily copied, so they set no one apart. What distinguished the practitioners is the curiosity to try a tool on an unfamiliar problem and the depth to carry judgement built elsewhere into it.

Introduction

AI has shown across many fields that it can be fast, efficient, and better than humans at a wide range of tasks. This is a speed without precedent, one that redefines the possibilities of everyone's job, including those in the humanitarian sector.

A machine can sift through thousands of data points, organise them, and produce a presentation of the findings, all in one breath. Yet anyone who has asked AI to build a presentation and received something oddly hollow knows that, still, AI has a long way to go.

Fortunately, we humans appear to stay essential, at least to the quality of work.

That said, as AI absorbs more of this work, the question becomes:

***If AI can do everything better and faster than me, what then
what becomes of the relevance of a humanitarian expert?***

A first answer to this is that experts are still needed because AI makes mistakes. The harder case is, however, when an output becomes confident, deeply embedded in workflows, and indistinguishable from expert judgement in a sector that has never defined what a humanitarian expert is. In that case, what remains of the expert's intervention is far less obvious.

Scoping

This paper focuses on practitioners working in immediate field work and crisis management, and their use of generative AI. This group is an interesting case precisely because of what their work consists of. Direct interaction with affected populations is an integral part of their duties, and it is also the part of humanitarian work most often assumed to be safe from automation. Testing this assumption against the tools now arriving at practitioners' hands is what motivates this paper.

For this paper, the authors have defined the "**AI era**" as a near future in which automation spreads across tasks once carried out only by humans, and in which AI-assisted decision-making becomes common at every level of authority and for every type of task.

Lastly, a note on how the paper's question was approached. Two positions dominate the discussion, among the interviewees, but neither is adopted here. The first position holds that AI will displace humanitarian workers as it has displaced others, which overstates a

pattern that history does not support¹. The second position holds that the human character of the work protects it, which mistakes an untested assumption for a finding. This paper examines both claims against what practitioners report.

We approach it through four questions:



¹ Autor (2014)



1.

What Made Me an Expert in the First Place?

What is a humanitarian expert?

Every field argues about what counts as an expert, and the humanitarian sector is no exception. Although the definitions on offer are broad, a first approach is to consider an expert as someone highly skilled and well informed in a particular field, to which common usage adds something more tangible, since an expert is also someone whose judgement their peers accept as superior to their own².

Research on expertise has tried to be stricter and defines experts as people who consistently perform better than others on tasks that are typical of their field³. That definition works where performance can be measured, such as with a surgeon's survival rates, patient satisfaction, or a chess player's rating, all of which can be calculated. Even then the measure raises a question, since a record of good outcomes may distinguish an expert from an ordinary professional, or it may distinguish a fortunate individual from an unfortunate one.

Humanitarian work sits awkwardly against these criteria, since it produces no straightforward quantifiable outcome and offers no fixed pathway into the role. As a fallback, the sector tends to rest on recognition, which means you are an expert when enough people in the sector treat you as one.

The principles seem to offer firmer ground, since almost every humanitarian organisation is committed to humanity, impartiality, neutrality and independence⁴. Additionally, they are emphasised as to be demonstrated through action, as OCHA puts it, reciting the principles is not enough when words must be matched by practice⁵.

By inference, a humanitarian expert is therefore a practitioner who upholds them in their work with affected populations, even under sustained pressure to set them aside.

However, this leaves a gap in our definition. If experts cannot reliably identify other experts, expertise functions partly as a social label, telling us who the sector treats as an expert without telling us what made them one.

² Cambridge Dictionary (2019)

³ Anders Ericsson et al. (2006)

⁴ OCHA on Message: *Humanitarian Principles* | Ocha (2012)

⁵ OCHA on Message: *Humanitarian Principles* | Ocha (2012)

Where does expertise come from?

If recognition is not the source of expertise, the question becomes: what is? While the practitioners interviewed this paper held varied opinions, a consistent approach nonetheless emerged: expertise is a set of capacities built through years of experience, intuition, pattern recognition, mentorship and judgement.

A practitioner arrives on a first mission, works alongside someone more experienced, and eventually makes decisions small enough to survive but consequential enough to matter, discovering only afterwards which of them were wrong. As one interviewee put it, once it is "*not your first mission*", you have been "*baptised by fire*".

What accumulates through that process extends well beyond hard skills and takes the form of a repertoire made up of situations encountered and outcomes seen. Senior practitioners draw on that repertoire in ways they find it hard to explain, recognising a situation because they have "*seen this before*", and knowing what tends to follow.

How does the expert acquire this knowledge?

Humanitarian knowledge is tacit, which is to say a person holds and acts on it without being able to state the rules behind it. You know how to ride a bicycle, but could you explain in writing how you do it? However detailed the description, a first-time rider will still fall. Simply put, we know more than we can tell⁶. In humanitarian work, this covers reading a room, sensing when a rule no longer fits the situation in front of you and holding trust under pressure. Knowledge of this kind cannot be handed over, since it must be acquired by doing the work.

That much explains why practitioners learn on mission rather than in training, and it says less about machines than first appears. What cannot be written down is not, on its own, beyond them, because machine learning is precisely an attempt to infer rules from examples where nobody can state the rule. For example, a system cannot be told how to recognise a face, but it can still learn to do so if given enough labelled faces.

Rather, the obstacle lies in the examples. Learning from them requires large sets of cases with known outcomes. Humanitarian judgement produces the opposite since every context differs and outcomes are rarely recorded against the decision that produced them. There may be large inter-agency evaluation completed after a significant emergency, but nobody records the small things. Who details a checkpoint negotiation as having gone well or badly in a format that a machine could learn from?

⁶ Polanyi (1966)

Participation is necessary, but not sufficient

Knowledge can also be accumulated far from a response. Deep exposure to a field's language and debates can produce valid judgement without on-the-ground practice, and some widely respected humanitarian experts have never experienced a camp or an epidemic from the inside. How can this hold? The discipline contains different roles which carry different kinds of value, and analysis, advocacy, policy and research do not require the person doing them to have been there.

What distance cannot supply is correction. Outcomes are rarely recorded against the decisions that produced them, and nobody logs whether a checkpoint negotiation went well or badly. The practitioner who made that call learns the answer nonetheless, since the consequences come back to them whether anyone writes them down.

Tested participation at the core of the humanitarian expert

This return of consequences is what the authors call tested participation, meaning work in which the practitioner makes the decisions and the results come back to them. It rests on three things, which are arguing a call with peers who are in a position to contradict it, building knowledge of the specific context rather than of the category it belongs to, and answering for the outcome when the call turns out to be wrong.

Participation supplies the knowledge that cannot be taught, while testing converts that knowledge into judgement. In humanitarian work the testing is unusually severe, since mistakes cost lives, harm the people being assisted, or close access, and it is that severity which makes the correction stick.

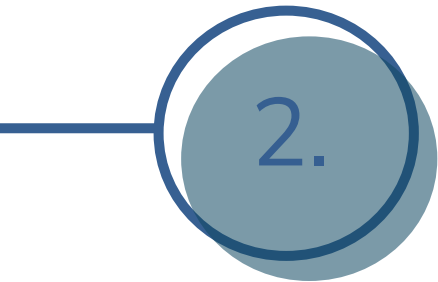
Expertise understood this way is a process rather than a possession, and it holds only for as long as the practitioner keeps making calls whose results come back to them. That is what makes the humanitarian expert, and the rest of this paper asks what happens to it once generative AI enters the work.

A working definition the Humanitarian Expert in the AI era

A practitioner who upholds the humanitarian principles under pressure, reads unfamiliar contexts well enough to act in them under uncertainty, and carries the consequences personally.

Expertise evolves and becomes the capacity for judgement rather than a stock of knowledge.

The new humanitarian experts use AI to **extend what they know, without letting it replace how and what they judge.**



2.

What is AI Changing About My Work?

AI helps under the right conditions

Integrating AI into humanitarian workflows has proven to give genuine improvements in the quality of what is produced and has changed the kind of work humanitarians spend their days on. Interestingly, these are not uniform.

What practitioners gain

Discussions with experts agree that the clearest gains cluster around, formatting and analysis, which are the parts of the job that can be written down as a procedure. Interestingly, the gains appeared to be the largest for less experienced staff, who see the improvement in speed and quality persist, so the effect works more like training than assistance.

Additionally, for teams already competent at the procedural work, the gain was notable not only in the hours saved but what those hours went to instead, which was less bureaucracy, less time at a computer, and more time with people. A clear example for this was how the head of coordination at OCHA Gaza described the shift, since teams that once spent hours cleaning data can now spend those hours collecting it and meeting the affected population, in a place where coverage is critically low.

For the experienced practitioner, the gain looks different again. It is analytical reach, and it rests on the framework the practitioner brings. A question they already knew how to ask becomes answerable at a scale hand analysis could not reach, while the judgement about which question is worth asking stays exactly where it was.

To illustrate our point, the case below shows what that looks like when an expert supplies the framework, and how much of the result depended on judgement the tool did not have.

Case study — The quad analysis at UNOCHA Somalia

From an interview with David Higgins, Head of Humanitarian Access and Civil-Military Coordination. June 2026

The context and what quad analysis is

The idea came because of team downsizing, the volume of incoming material became difficult to manage, and the office no longer held the skills to run this analysis by hand.

A quad analysis crosses two things you already measure and sorts each location into one of four resulting groups. The office held both inputs already, severity of need by location and severity of access constraints for the same locations but read them separately. Crossing them isolates the locations that are severely affected and hard to reach, which no single dataset identifies. A third dataset compared the response to what was actually being delivered, turned the picture into a finding: **aid was tracking ease of access rather than severity of need**.

Regarding methodology, he sketched the axes on a whiteboard, photographed it, and uploaded the image as his prompt, together with the datasets. He then ran what he calls loop engineering having the model draft its own prompts and iterate, while he judged each pass until the output was usable. Lastly, that draft went to country-office specialists for validation and packaged into a reusable workflow for next users.

What the AI supplied	What the expert supplied
Capacity to apply the framework to two large datasets	The framework itself and crossing over information
The processing skills	Recognising the question two separate datasets could answer together
Completion in hours, not a week	Sketching axes on a whiteboard as visual prompts
Its own prompts, rewritten across successive passes	Knowing when a pass was wrong, and contextualising the output
Packaging the workflow for reuse	Knowing which specialists could contradict him, and asking them

What a practitioner can take from this

- Design before prompting. The gain came from a technique the tool did not propose.
- Iterate against your own judgement and stop when you can defend the result.
- Choose validators who can contradict you. Review without contextual knowledge is presence rather than oversight.
- Package what worked. It makes the second run cheaper than the first, and answers the sector's absence of common analytical frameworks

What practitioners stop noticing

Nonetheless, the gains from integrating GenAI into workflow also bear its share of risks. Indeed, the same pressure that keeps experts doing more behind the screens appears to also erode the checking or validation of AI outputs.

Interviewees described colleagues accepting AI output without examining it; a behavior researchers call **cognitive surrender**. The term describes what happens when a person stops applying their own judgement to a task and instead defers to the system's answer by default, not because they have checked it and found it sound, but because checking has quietly stopped being part of the routine.

It is not a single lapse but a shift in habit, and researchers trace it to staffing shortages and the disappearance of mentorship, the everyday oversight that used to catch a wrong answer before it went out. In practice, this can look like a protection officer accepting an AI-drafted vulnerability assessment without checking whether it missed a red flag, or a coordinator using an AI-generated map of needs without asking whether the underlying data reflects the population on the ground.

This should not be read as individual failure. An under-resourced person working against a deadline finds a fluent answer easier to accept than a correct one, and the incentive to check falls away exactly as the pressure to produce rises. A meta-analysis and colleagues, covering 106 experiments, documents the behaviour as overreliance, where users treat AI suggestions as strong guidelines rather than seeking further information⁷. Furthermore, the study also found that human-AI pairings outperformed people working alone on creation tasks, while performing worse on decision tasks, the very tasks where a wrong call is hardest to catch after the fact. In that case, the losses concentrate precisely where practitioners are least likely to notice them.

Who bears the errors, and who can afford to correct them?

Inevitably, AI is bound to make errors at some point, but in that case, who bears that cost? Any system that ranks or prioritises has to be told what counts, and someone has decided which populations the training data represents, which proxies stand in for need, and where the threshold sits between a case the system flags and one it passes over. These choices are made in system design, often before the humanitarian context is known, and they arrive in a response already settled.

⁷ Vaccaro et al

For a humanitarian reader, this is a principles problem, not a quality one, since deciding whose needs come first is exactly the determination impartiality governs. A practitioner asked why one caseload was prioritised over another is expected to account for it, and that account is hard to give when the ranking rests on weightings they cannot inspect and did not set. Judgement about who is most in need has been displaced rather than removed. It now sits with people who will not answer for how it turns out, while the consequences remain with populations who have no way of contesting it.

Whether that displaced judgement gets caught depends on where the practitioner works. AI is often presented as a levelling force, extending capacity to responders who lack it, but the authors found adoption and impact to be more divided than that. Only 13% of local-agency staff report awareness of AI policy, against 39% at UN agencies, and awareness is just the visible edge of a wider gap running through training, access, and the time to check what a tool has produced.

The implication is uncomfortable. Time, mentorship, and the standing to reject an output are what make these tools safe, and all three describe a well-resourced organisation. Adoption tends to track existing inequality rather than correct it, with the gains landing where capacity exists and the errors described above landing where it does not. This is also why the remedy most often proposed that a human should check the output, deserves more scrutiny than it usually gets. It assumes a practitioner with the time to check, the training to know what to look for, and the standing to disagree with what the system produced.





3.

Doesn't Human Oversight Protects Me From This?

Adding a person is not the same as adding a safeguard

Human oversight is the sector's standard answer to most risks raised in this paper – one build on a picture of a competent person reading an output closely enough to catch what is wrong with it. Today, most AI-aware teams have built human sign-off into their workflows, and the practitioners we spoke to were right to see this as a real protection.

However, we have seen this safeguard weakened in two distinct ways. The first concerns the reviewer, whose attention erodes under conditions the sector has been creating for years. The second concerns the system, which gives the reviewer progressively less to work with even when their attention is steady.

The reviewer

Sign-off is a step in a process, so it survives on paper whatever happens to the quality of the reading behind it, and nothing in the workflow registers the difference between a reviewer who interrogated an output and a reviewer who scanned it before a deadline. Over time, the step remains, but the scrutiny may shrink, and yet safeguard looks intact in an audit.

Furthermore, oversight becomes superficial where the people doing the overseeing are themselves over-relying on AI. The safeguard degrades fastest in the environments that adopted these tools most enthusiastically. A reviewer who has come to expect the system to be broadly right reads for confirmation rather than for error, and a fluent, procedurally correct output gives them every reason to find it.

The system

The second failure does not depend on the reviewer's habits at all, and it would persist in a team of careful and well-rested people. Oversight assumes that a decision can be reconstructed afterwards, and generative AI makes reconstruction harder. A system

drawing on several sources returns a summary without showing which context, which instructions or which retrieval steps produced it. Many humanitarian decisions are poorly documented already, and when the reasoning sits inside a system that cannot expose it, a practitioner has real difficulty explaining how they arrived at the decision they took, whether the question comes from a donor, a colleague or an affected community.

In this case, the practitioner who does have the time and the expertise to check now has less to check against. The arrangement works in software teams, where the task is bounded, and the output can be read line by line, but humanitarian analysis offers neither of those conditions. The reviewer is left examining a conclusion rather than the reasoning that produced it, which is a weaker exercise than the word *review* suggests.

A note on agentic systems.

Agents and multi-agent systems are being adopted rapidly across other sectors, and humanitarian organisations tend to adopt later. That lag is an opportunity rather than a reprieve, since these systems will reach humanitarian workflows, most likely through the administrative and logistics functions where the case for them is easiest to make.

They matter here because everything above assumes a person sees the output before anything is done with it. Agentic systems remove that moment, since the point of delegation is less ongoing supervision.

The failure mode is not a wrong answer but an unintended one: agents each following their own rule correctly, yet together producing outcomes nobody intended, in domains like supply chains, payroll or food distribution. This puts into question the human in the loop approach, since a human cannot possibly be involved in each step of such systems.



4.

What Keeps Me Relevant?

The previous sections have shown that expertise is difficult to define, that it is being transformed, and that the human sign-off safeguard does not contain what is changing. So, what keeps the expert in humanitarian action relevant in the AI era?

One answer kept surfacing in our interviews. What keeps you relevant is what a system cannot hold on to your behalf, judgement built where the consequences reached you, the reading of a context nobody has written down, and the fact that you are the one who answers for the decision. In humanitarian terms, this is the difference between a system flagging a caseload as high-priority and the coordinator who has to explain to a community why one settlement was reached before another. Practitioners working in field operations, in institutional governance, in regulation and in technology arrived independently at that same formulation, which is the strongest signal in our interviews that roles are being redefined rather than replaced.

The question worth asking is which tasks, with what safeguards, under what conditions, and whether the people doing the work are the ones positioned to answer it. What follows is our attempt to answer it where the decisions are actually made, since the person choosing whether to paste a beneficiary list into a tool, the manager deciding who drafts and who reviews, and the director signing off on a hiring freeze are facing different versions of the same problem. The recommendations below are grouped accordingly.

If you are the practitioner

Acceptance is granted to people, not to systems. A tool can tell you what to say, but it cannot hold the trust that got you in the room. The moment you become a relay of AI output, you start spending the one asset you cannot delegate. So how do you use AI without eroding the relationship it depends on?

- 1. Draft first, then consult the tool.** Write down your own understandings and ideas before asking a system for one. The order in which you do these two things determines whether you are checking your reasoning or replacing it.
- 2. Treat AI as an adversarial source rather than a colleague.** Outputs may be incomplete, biased, or shaped by commercial incentives. Even internally deployed systems usually run on externally trained models.
- 3. Defend field time and the relationships in it.** *Presence* is what acceptance is granted upon and is what supplies the *tested participation* your judgement is built

from. Fight to ensure efficiency gains are reallocated to appropriate work and better judgement rather than reinvested in creating more output.

4. **Do not build your value on tool fluency.** Understanding how these systems work is necessary. But, since the tools change faster than anyone can master them, proficiency in this year's system can be a quickly depreciating asset. However, the judgement you check AI against needs no re-learning.
5. **Go narrow before you go broad.** The pattern of recognition that lets you catch a wrong answer is built inside one context rather than across many, and breadth is what you add once you have something to check against. If you start broad, it leaves you fluent everywhere but calibrated nowhere.
6. **If you would not publish it, do not paste it.** This was the simplest data rule we encountered, and it survives in most of the edge cases.
7. **Be skeptical of explainability features.** Neither explanations nor confidence displays provided by the system improved performance. Instead, the more promising direction is where tools prompt critical questions and surface hidden assumptions.

If you lead a team

Automation neither diminishes nor elevates the human tasks that remain. What matters is how the split is made. For instance, mechanisation raised the value of construction workers' judgement in safety and problem-solving, but a tech-support worker reduced to relaying scripted answers delivers a worse service than either a skilled person or a good system alone. Nobody chooses that second outcome, and yet it is where teams end up when the split is left to form on its own. That makes it a management decision rather than a technical one.

1. **Refuse the arrangement where AI drafts and a person relays.** Such a split often emerges by default rather than by decision and will mean that. Nobody on the team becomes more of an expert. Be deliberate rather than passive.
2. **Rebuild junior development around judgement rather than tool use.** Juniors gain most from these tools and are simultaneously the most exposed to over-reliance where mentorship has thinned. Preserving some tasks that juniors do AI-unaided is a future resource pipeline decision rather than a training preference.
3. **Run a regular, low-stakes session on what worked and what did not.** Teams that keep a shared log of which tools helped, for which purposes, and where they failed to build a picture faster than individuals experimenting separately. If done across a sector, the exercise can surface the failures people do not otherwise volunteer.

If you set institutional policy

Judgement is built through decisions whose consequences return to the person who made them. Unfortunately, many of the Gen AI tools are absorbing exactly those decisions at the junior level as the tasks are most easily formalised and the pressure to save time is greatest. So, what can an organization do?

1. **Make AI literacy role-specific rather than generic.** Not everyone needs to be an expert in AI. But a legal officer, an information management specialist and a WASH team member all need to recognise the risks they will likely encounter in their specific work.
 2. **Create genuine space for review.** Give reviewers the time to review and the standing to disagree. Without these, people learn nothing, people become dependent on machines, and work gets inaccurate.
 3. **Treat entry-level roles as expertise infrastructure rather than overhead.** If those positions are cut or automated now, the sector is effectively spending its future senior judgement to fund today's efficiency and, that is the big question budget-holders need to sit with.
-

Conclusion

We set out to ask what it means to be a humanitarian expert in the AI era, and the answer the interviews converged on was narrower than we expected. What keeps a practitioner relevant is not what they know, since a system will hold more of it, and not what they can produce, since a system will produce it faster.

What makes a humanitarian expert is that they were there when things went wrong, they can read a context nobody has written down, and they are the ones who answer to the decision.

A practitioner who drafts first and consults second is checking their reasoning; one who consults first is replacing it. From the outside the two look identical, and it is very likely that the sector will not be able to tell them apart. Yet the distinction matters because humanitarian decisions affect people who must be able to question them, influence them, and know who is accountable for their consequences. If human reasoning becomes indistinguishable from machine output, the sector risks weakening the trust, responsibility, and principled judgement on which its access and legitimacy depend.

And the difference will only survive if we make the right choices about time, mentorship and junior posts. Not about software.

Annex 1: Methodology & Scope

This paper is a compilation of findings from a literature review as well as key informant interviews with experts and humanitarians pushing the agenda for digital innovation in humanitarian action.

This paper is presented as a think-brief to help start a conversation or help provide a concrete stepping stone for those interested in topics of Generative AI and its use by humanitarian practitioners. The paper is not intended to be interpreted or treated as an academic or peer-reviewed paper. Instead, it is a compilation of introductory research intended for a broad audience.

Scope

The paper focuses on practitioners engaged in immediate field work and crisis management, and on their use of generative AI in that work. The AI era is taken to mean a near future in which automation spreads across tasks once performed only by humans, and in which AI-assisted decision-making becomes common at every level of authority.

Annex 2: Interviewees

The authors would like to extend our deepest gratitude to the individuals who took the time to meet with us, answer our questions, and share stories, guidance and wisdom around expertise, judgement and the use of generative AI in humanitarian work. Having conversations about the ways in which the humanitarian community is working across geographies and across sectors was both inspiring and eye opening. As this report is based predominantly on the experiences of the interviewees, it would not be possible without them.

This paper does not represent the views of the interviewees unless explicitly stated. The authors have drafted the paper based on a combination of literature review and interviews.

Interview approach

Interviews were semi-structured and conducted remotely. Participants were selected to span four vantage points on the same question: field operations, institutional governance, regulation and policy, and technology. This spread was deliberate. A finding that recurs across all four is treated as more robust than one raised from a single position.

Interviewees were asked about their own practice rather than about AI in the abstract: what they use, what changed, what they stopped doing, and where they judged a tool to be wrong. Quotations are attributed by name where the interviewee agreed to attribution and paraphrased otherwise.

Limitations

- **Sample size and selection.** The interview base is small and drawn from people already engaged in digital innovation. It over-represents practitioners who have formed a view on AI and under-represents those who have not encountered it or have quietly declined to use it.
- **Institutional weighting.** Interviewees are concentrated in UN agencies and large international organisations.
- **Self-report.** Efficiency gains and instances of over-reliance are reported by practitioners rather than measured.
- **Moving target.** The interviews were conducted over a defined window, against tools that change on a scale of months.

Interviewee List

All interviews conducted in May & June 2026.

Name	Profession
Ana Bedushi	Professor of Law, Technology and Human Rights, University of Exeter
Somya Gupta	Manager, AI Innovation and Research, Google
David Higgins	Head of Humanitarian Access and Civil-Military Coordination, UN OCHA Somalia
Heather Leson	Senior Lead, Data and AI Governance, Canadian Red Cross
Jacopo Margutti	AI Lead, 510 data and digital team, Netherlands Red Cross
Gísli Ólafsson	Founder, Gridland Consulting; author of The Crisis Leader, The Motivated Leader and The AI Fluent Leader
Cornelia C. Walter	Professor of ProSocial AI, Sunway University; Expert in Hybrid Intelligence, UNFPA
Daniela Weber	Director, NetHope Center for the Digital Nonprofit
Daniela Wuerz	Senior Programme Management Officer, UN Office at Geneva
Paola Yela	Information Management and Data Science Senior Officer, IFRC

Annex 3: Artificial Intelligence Disclosure

All ideas, concepts, and recommendations are original, but artificial intelligence supported the editing and refining process. AI tools were employed to ensure that the ideas were communicated clearly and done so in a way that would allow diverse audiences to understand the report's findings and arguments.

Annex 4: Glossary

Agentic AI - A system given an objective that then carries out a sequence of steps on its own, rather than producing one output for review. The difference that matters with generative AI the human sees the result before anything happens; with an agent, the action may already be taken.

Calibration - Knowing which tasks you judge better than the tool and which it judges better than you. Not held by default; acquired only about tasks where you have compared the two.

Cognitive surrender - Accepting AI output without checking it because checking costs time the practitioner does not have. Tracks staffing cuts and lost mentorship rather than individual failure.

Human-in-the-loop - An arrangement in which a person reviews output before it is used. This paper distinguishes the arrangement from the capacity it assumes time to look, expertise to know what is wrong, and standing to disagree.

Humanitarian principles - Humanity, impartiality, neutrality, independence. The commitments almost every humanitarian organisation holds, and the basis of this paper's working definition of an expert.

Overreliance - Treating AI suggestions as firm guidance and ceasing to seek further information. The behaviour behind cognitive surrender. Explanations and confidence displays have been tested as remedies and did not measurably help.

Tacit knowledge - Knowledge held and acted on without being able to state the rules behind it: we know more than we can tell. Reading a room, sensing when a rule no longer fits, holding trust under pressure. Acquired by doing the work, not by being told.

Tested participation - This paper's term for how tacit knowledge becomes judgement. Participation supplies what cannot be taught; testing decisions whose consequences return to you, is what converts it. In humanitarian work the testing is severe, and the severity is what makes the correction land.

Annex 5: Bibliography

Anders Ericsson, K., Charness, N., Feltovich, P. J., & Hoffman, R. R. (2006). *The cambridge handbook of expertise and expert performance*. Cambridge University Press.

Autor, D. (2014). Polanyi's paradox and the shape of employment growth. *National Bureau of Economic Research*. <https://doi.org/10.3386/w20485>

Cambridge Dictionary. (2019, October 16). *EXPERT | meaning in the cambridge English dictionary*. Cambridge.Org. <https://dictionary.cambridge.org/dictionary/english/expert>

OCHA on message: Humanitarian principles [en/ar] | ocha. (2012, June 30).

Www.Unocha.Org. <https://www.unocha.org/publications/report/world/ocha-message-humanitarian-principles-enar>

Polanyi, M. (1966). The Logic of Tacit Inference. *Philosophy*, 41(155), 1–18.
<http://www.jstor.org/stable/3749034>

Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour*, 8.
<https://doi.org/10.1038/s41562-024-02024-1>