Humanity in the Loop

Principled AI Adaptation for Humanitarian Organizations

September 2026

Humanity in the Loop: Responsible AI Adaptation for Humanitarian Organizations

By: Sophie Byrne (LinkedIn) and Andrej Verity (LinkedIn) | Digital Humanitarian Network

Key Messages

- **AI adoption is already happening in the humanitarian sector, but organizational adaptation is lagging behind.** Turning individual use into better humanitarian outcomes depends on organizational action to strengthen judgment, learning, governance, and coordination.

- **Human judgment must remain central,** supported by AI literacy, meaningful oversight, and deliberate protection of human expertise.

- **Adaptation depends on culture as much as technology.** Organizations need environments where staff can experiment safely, share lessons, and collaborate across teams.

- **Governance and responsible design should enable safe innovation** by being proportionate to risk, built in from the start, and treated as an ongoing process.

- **AI capability must be strengthened across the wider humanitarian ecosystem** to prevent duplication and ensure its benefits reach beyond the best-resourced organizations.

- **The goal is not faster or more AI use, but AI that strengthens humanitarian judgment, accountability, trust, coordination, and engagement with affected communities.**

Introduction

Across the humanitarian sector, individual adoption of Artificial Intelligence (AI) is outpacing organizational adaptation. A 2025 survey of 2,539 humanitarians across 144 countries found that around 70% already use AI in their work weekly or daily, yet only 8% report organization-wide integration and 22% say their organization has a formal AI policy.[1] A January 2026 follow-up found individual use still rising while organizational readiness barely increased.[2] This gap creates a central organizational challenge. AI is already entering humanitarian work, but often without the governance, skills, and accountability systems needed to make its use safe, principled, and effective.

AI holds real promise for humanitarian action by enabling faster analysis, broadening access to information, reducing administrative burden, and creating more time for the human engagement that matters most. However, realizing this potential is not automatic, and organizational caution is not misplaced. Alongside their benefits, AI systems can encode bias, generate misinformation, obscure the basis on which decisions are made, expose sensitive data, reinforce harmful power dynamics, and carry real environmental costs.[3] In humanitarian contexts, these are more than reputational threats, they can conflict directly with the duty to do no harm.

Organizational inaction is not a neutral choice. It allows the organization to be shaped by informal and invisible use, rather than shaping use intentionally to align with humanitarian principles and mission goals. Our interviews echo the survey results, describing AI spreading through offices, frontline responders consulting ChatGPT on their phones, staff increasingly treating these tools as necessary regardless of whether their organizations formally support them, and restrictive policies doing little but reducing the visibility of use.

---

[1] Data Friendly Space and Humanitarian Leadership Academy, "How Are Humanitarians Using AI in 2025?," Data Friendly Space, July 15, 2025, https://www.datafriendlyspace.org/resources/initial-insights-report.

[2] Data Friendly Space and Humanitarian Leadership Academy, "75% of Humanitarians Now Use AI Regularly, Yet Fewer Than One in Ten Work in an Organisation Where AI Is Widely Integrated," Humanitarian Leadership Academy, 2026, https://www.humanitarianleadershipacademy.org/news/75-percent-of-humanitarians-now-use-ai-regularly-yet-fewer-than-one-in-ten-work-in-an-organisation-where-ai-is-widely-integrated.

[3] Ana Beduschi, "Harnessing the Potential of Artificial Intelligence for Humanitarian Action: Opportunities and Risks," International Review of the Red Cross 104, no. 919 (2022): 1149–1169, https://doi.org/10.1017/S1816383122000261; Michael Pizzi, Mila Romanoff, and Tim Engelhardt, "AI for Humanitarian Action: Human Rights and Ethics," International Review of the Red Cross 102, no. 913 (2021): 145–180, https://doi.org/10.1017/S1816383121000011.

Allowing daily practice to outpace organizational understanding and governance has real costs, including "*shadow AI*" that no one can see or support, sensitive data entered into public commercial tools, uneven staff capability producing poor judgment, and duplicated effort as teams solve the same problems in isolation. The question is no longer whether AI will enter humanitarian organizations, but whether its adoption will be strategic or unmanaged.

A managed approach does not mean treating AI as a universal solution or forcing adoption. It means deciding deliberately how AI can strengthen humanitarian judgment, accountability, and trust, rather than allowing it to undermine them by default.

The Organizational Challenge: Adaptation, Not Acceleration

AI lowers the cost of producing initial drafts, summaries, code, and communications, offering significant implications for productivity. But faster output only matters if it improves the quality, timeliness, and appropriateness of humanitarian response. The more consequential shift is not how quickly work gets done, but *who can do the work.*

As information becomes increasingly accessible, more people can undertake analysis or technical projects without waiting for support from specialists. Jerri Husch describes this as a shift toward an "*initiator*" model: AI does not simply enhance existing workflows, it changes who can define a problem, perform analysis, and create a product.[4] This can bring solutions closer to operational realities, accelerate learning, and reduce bottlenecks, but it also moves the scarcity from production to judgement. Organizations must now determine how to validate outputs, align individual initiatives with organizational goals, and ensure increased production translates into better outcomes.

These are not challenges that better technology can solve on its own, they are questions of organizational design. Addressing them requires coordinated effort across four interconnected pillars: preserving **human expertise and judgment**, building systems for **organizational learning and adaptation**, ensuring **tiered governance and responsible design**, and strengthening **ecosystem coordination**. The goal is not to accelerate work for its own sake, but to adapt organizationally so that AI enhances humanitarian action and supports the uniquely human capacities at the center of the work

[4] Jerri Husch, "AI for Humanitarian Coordination: A Practical Brief for OCHA" (January 2026).

A Managed Approach: Four Pillars of Principled AI Adaptation

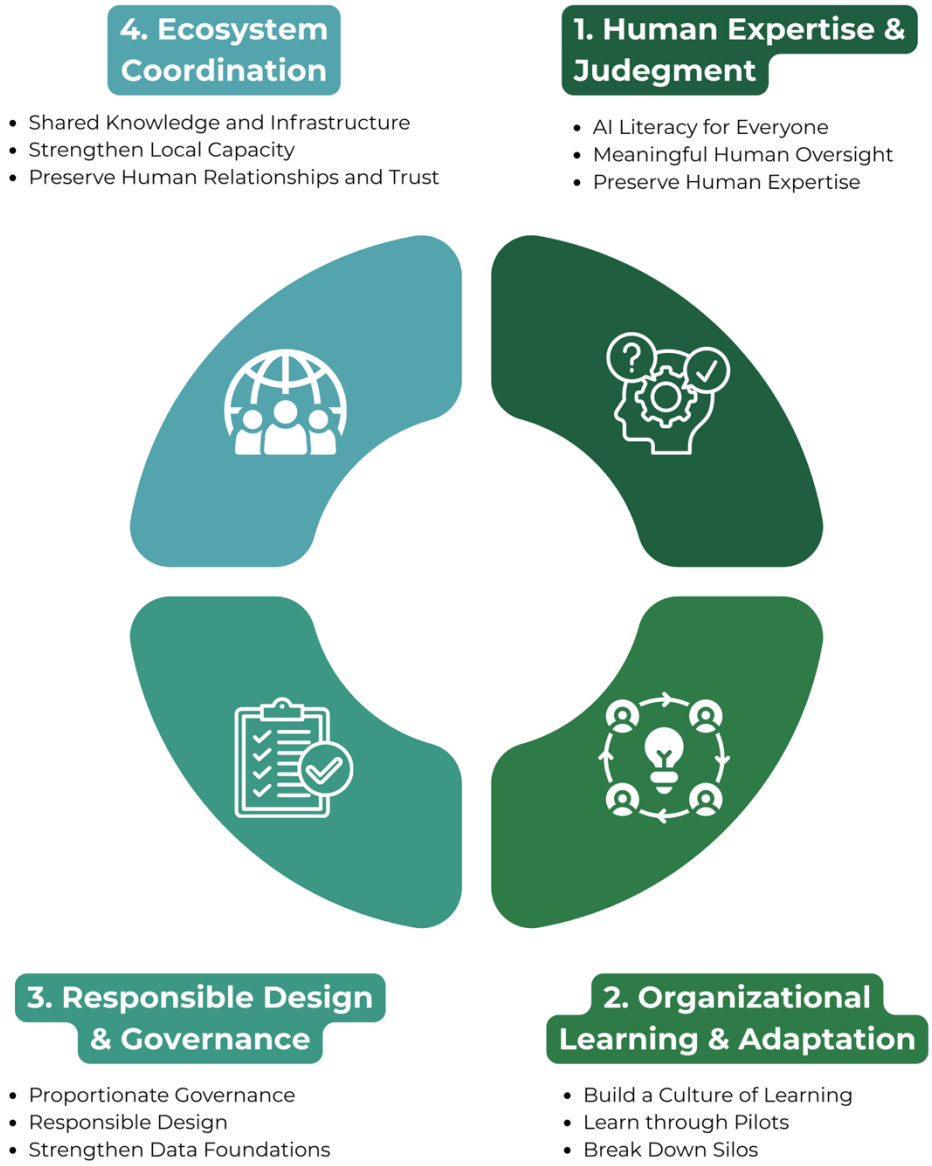

Figure 1. This framework presents four interconnected pillars for responsible AI adaptation in humanitarian organizations. Progress across human judgment, organizational learning, responsible governance, and sector-wide coordination is necessary to ensure that AI strengthens humanitarian response.

Pillar 1: Human Expertise & Judgment

Just as AI lowers barriers to producing initial analysis, it increases the value of human judgment for validating outputs, deciding when it is appropriate to use AI at all, and making the final call in AI-assisted decisions. AI systems can inherit bias from their training data, make recommendations that are difficult to trace or explain, and produce fluent, confident outputs that are nonetheless wrong. Humans, meanwhile, may trust such systems even when they are incorrect and accept their outputs without critical evaluation, a pattern that occurs even among people with deep subject-matter expertise, and intensifies with time-pressure, increased system trust, and repeated use.[5]

High workloads, staffing shortages, reporting burdens, and operational pressure may make humanitarian settings especially prone to this overreliance, and if decisions affecting people's lives are made without active human engagement, the potential for harm is significant. **Preventing AI from displacing human judgment requires organizations to protect that judgment deliberately through AI literacy, workflows built around meaningful oversight, and pathways that preserve human expertise.**

AI Literacy for Everyone

"We need to train literally every single person who works in this organization on the basics of AI literacy, ethics, and responsible use." - Ilen Madhavji

Maintaining capacity for informed human judgement depends upon people's ability to understand what they are evaluating. Whether or not staff use AI themselves, they will

[5] Hao-Ping Lee et al., "The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects from a Survey of Knowledge Workers," in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (New York: Association for Computing Machinery, 2025), https://doi.org/10.1145/3706598.3713778; Joe Pearson et al., "Examining Human Reliance on Artificial Intelligence in Decision Making," *Scientific Reports* 16 (2026): article 5345, https://doi.org/10.1038/s41598-026-34983-y; Giuseppe Romeo and Daniela Conti, "Exploring Automation Bias in Human–AI Collaboration: A Review and Implications for Explainable AI," *AI & Society* 41 (2026): 259–278, https://doi.org/10.1007/s00146-025-02422-7.

inevitably encounter AI-assisted content and knowing how to respond to it requires basic knowledge of how these systems work, where they commonly fail, and how that informs responsible use in different contexts. In a recent survey of the humanitarian sector, 73% of respondents identified training as the most important AI-related support mechanism in the coming years, while 64% reported they had received little or no organization-led training on AI.[6]

Literacy initiatives should encourage staff to treat AI outputs as a perspective that requires questioning, not a source of truth. A specific focus on the risks of AI error and overreliance may make people less likely to over trust it, and hands-on practice identifying errors in AI output can encourage continued critical evaluation.[7] But effective AI literacy depends on more than technical understanding. It also means recognizing the distinctive strengths people bring to human-AI collaboration such as contextual knowledge, empathy, ethical reasoning and creativity, and understanding how AI use may weaken or sideline these qualities.[8] Knowing where these tools break down and where human intelligence outperforms them is not abstract knowledge, but should directly shape when and how they are used.

This is not to suggest that everyone must become an expert on AI. Literacy should be oriented toward the practical realities of individual responsibilities: a universal baseline for everyone, role-specific training aligned with particular use cases, and additional pathways for staff who wish to deepen their knowledge or lead AI initiatives.

---

[6] Data Friendly Space and Humanitarian Leadership Academy, "How Are Humanitarians Using AI in 2025?".

[7] Cordula Kupfer et al., "Check the Box! How to Deal with Automation Bias in AI-Based Personnel Selection," Frontiers in Psychology 14 (2023): article 1118723, https://doi.org/10.3389/fpsyg.2023.1118723; Lucía Vicente and Helena Matute, "Warning People about the Risk of AI Error Mitigates Human Acquisition of AI Bias," Cognitive Research: Principles and Implications 11 (2026): article 36, https://doi.org/10.1186/s41235-026-00726-w; Emily Pavuluri, "Before You Trust the Output: Three Exercises for Teaching AI Evaluation," *The AI of Law*, 2026, https://vaill.substack.com/p/aievaluationexercises.

[8] Maha Bali, "What I Mean When I Say Critical AI Literacy," Reflecting Allowed, April 1, 2023, https://blog.mahabali.me/educational-technology-2/what-i-mean-when-i-say-critical-ai-literacy/; Cornelia C. Walther, "Why Hybrid Intelligence Is the Future of Human-AI Collaboration," Knowledge at Wharton, March 11, 2025, https://knowledge.wharton.upenn.edu/article/why-hybrid-intelligence-is-the-future-of-human-ai-collaboration/.

Meaningful Human Oversight

"The issue is not whether AI hallucinates. The issue is our capacity to question what it tells us."

- Jerri Husch

Many organizations assume that keeping a "human in the loop" of AI-assisted decisions is a sufficient oversight mechanism, but the reality is more complicated. Even with appropriate literacy in place, oversight can become little more than a rubber stamp if people lack the time, authority, knowledge, or confidence to challenge AI outputs.[9] A human in the loop is not the same as an *expert* in the loop.[10] It does not matter how critically someone is taught to engage with AI, if they lack the subject matter expertise to catch nuanced errors. Catching the errors will not matter if they lack the confidence to trust their own judgment over the system. And confidence will not matter if they lack the authority to override the output. Designing processes that preserve human agency and accountability therefore means more than disclosing when AI has been used or ensuring that a human is involved in AI-assisted decisions.[11] It depends on reviewers who are actually equipped to evaluate the content they are encountering and accessible escalation pathways for when outputs appear unreliable.

This should not mean documenting and scrutinizing every output equally, which would be unlikely to improve productivity or operational function. Not all AI use carries the same level of risk, summarizing internal meeting notes is very different from using AI to inform decisions about access to aid. Excessive controls create unreasonable workload and encourage staff to bypass governance mechanisms altogether, so the goal should be systems that encourage the appropriate level of evaluation for the stakes involved. People are less likely to correct AI

[9] Lee et al., "The Impact of Generative AI on Critical Thinking."
[10] Nolan Lovett, "The Tragedy of the Cognitive Commons: How AI Could Disrupt the Regeneration of Professional Expertise," Human Resource Development Review, 2026, https://doi.org/10.1177/15344843261470602; The Ai Consultancy, "Expert in the Loop: Why AI Needs Specialists, Not Just Humans," Medium, 2025, https://medium.com/@ai_93276/expert-in-the-loop-why-ai-needs-specialists-not-just-humans-5797b6500654.
[11] Audrey Tang and Caroline Emmer De Albuquerque Green, "Pack 2: Responsibility — Taking Care Of," Civic AI, Oxford Institute for Ethics in AI, Accelerator Fellowship Programme, https://civic.ai/2/.

outputs when doing so takes additional effort, making workflow design a critical determinant of whether oversight occurs in practice.[12]

Appropriate design will vary by context: peer review can help mitigate individual biases in high-risk situations, while selective auditing of uncertain cases may be more efficient elsewhere. Independent judgment can be strengthened through system design that encourages cognitive engagement, such as requiring an initial assessment before viewing AI recommendations, providing analyses without prescriptive suggestions, or incorporating understandable explanations into outputs.[13] At the same time, safeguards that are too burdensome or complex can reduce usability and even worsen performance. Meaningful oversight means ensuring that human judgment is exercised most strongly where it matters most, in cases of high uncertainty, areas where AI is prone to error or may have incomplete information, and decisions that impact operational outcomes.

Preserve Human Expertise

"If AI takes over the entry-level jobs, how do we become experts?" — Daniela Wuerz

Engaging responsibly with AI requires the contextual knowledge and self-confidence to critically evaluate its outputs, yet AI's integration into workflows may gradually erode the very expertise that effective use depends on.[14] Staff with extensive experience may become rusty at tasks they are no longer performing themselves, while junior staff may lose opportunities to develop expertise through the repetition, experimentation, and problem-solving that traditionally underpinned professional development. Our research indicated a common concern that organizations could weaken their own capacity for judgment if AI replaces the work through which expertise is built and maintained. A similar pattern has been

[12] Jacob Beck et al., "Bias in the Loop: How Humans Evaluate AI-Generated Suggestions," *Harvard Data Science Review* 8, no. 2 (2026), https://hdsr.mitpress.mit.edu/pub/nrcn4h7d/release/2.
[13] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z. Gajos, "To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-Assisted Decision-Making," Proceedings of the ACM on Human-Computer Interaction 5, no. CSCW1 (2021): article 188, https://doi.org/10.1145/3449287; Lee et al., "The Impact of Generative AI on Critical Thinking"; Romeo and Conti, "Exploring Automation Bias in Human–AI Collaboration."
[14] Isabell Steidel and Benedikt Gieger, "The Oversight Paradox: Human Control Over AI May Be Eroding," World Economic Forum, 2026, https://www.weforum.org/stories/artificial-intelligence/oversight-paradox-human-control-ai/; Lovett, "The Tragedy of the Cognitive Commons."

seen before: pilots must practice manual flying regularly, precisely because reliance on autopilot was found to degrade the skills needed when automation fails or is unavailable.

Organizations should therefore think not only about how AI changes work, but also about how to preserve the human capabilities that are most essential to their operations. Across sectors, there is growing emphasis on skill-based hiring, multidisciplinary expertise, and more flexible role structures as AI reshapes workflows.[15] Ethical reasoning, systems thinking, problem framing, and critical evaluation may become increasingly important across roles, while relationship building, collaboration, facilitation, and contextual understanding will remain at the heart of humanitarian work. The human capabilities most central to continued organizational functioning need to be identified and cultivated through hiring, mentorship, training, and deliberate practice.[16]

Pillar 2: Organizational Learning & Adaptation

For organizations, strategic AI adoption is not primarily a technical process, but an organizational learning process. It is not simply about introducing new tools, but rather creating the conditions that allow organizations to adapt as technology, work, and expertise continue to evolve. Across our research, impactful AI adoption was rarely described as a consequence of better technology alone. **Successful organizations create a culture where people are willing to innovate responsibly, learn collaboratively, and share their insights across teams.**

[15] Sue Cantrell et al., *Building Tomorrow's Skills-Based Organization: Jobs Aren't Working Anymore* (Deloitte, 2022), https://www.deloitte.com/content/dam/assets-shared/legacy/docs/Deloitte-Skills-Based-Organization.pdf; ITU AI for Good, "Human agency, AI, and the future of mission work," July 8, 2026.

[16] Antoine Buteau, "AI Can Drain the Expert Pipeline Before Anyone Notices," Antoine Buteau (blog), August 9, 2026, https://www.antoinebuteau.com/ai-can-drain-the-expert-pipeline-before-anyone-notices/; FranklinCovey, "AI in the Workplace: A Leader's Guide to Human-Centered AI Adoption," July 6, 2025, https://www.franklincovey.com/blog/embrace-ai-in-the-workplace/; International Labour Organization, Lifelong Learning and Skills for the Future, World of Work Series (Geneva: ILO, May 5, 2026), https://doi.org/10.54394/00033011.

Build a Culture of Learning

"Organizations need safe spaces to share what they do not know about AI - and learn together."

- Ka Man Parkinson

Attitudes about AI within the humanitarian sector vary widely from uncritical optimism to strong resistance. Organizations should avoid both extremes; neither mandating nor prohibiting AI use will be effective. Instead, adaptation depends upon collaboration with staff, grounded in shared learning and informed choice. This begins with open dialogue about successes, failures, ideas, concerns, and emerging practices within and across teams. Regular conversations normalize responsible experimentation, reduce duplication, surface problems earlier, and build a shared understanding of organizational needs.[17] Leaders, technical specialists, and early adopters can play an important role in creating psychological safety by openly sharing their own uncertainties and mistakes, along with their successes. Perspectives shared in these conversations should directly inform training, governance, priority use cases, and where AI is never used, ensuring decisions align with actual needs and building trust throughout the organization.

The way in which AI is introduced into an organization also shapes the culture around it. Perceived support from leaders is essential, and staff are more likely to engage responsibly when managers demonstrate curiosity, transparency, and a willingness to learn themselves.[18] Positioning "**innovation navigators**" within teams or departments ensures staff have visible points of contact for questions and concerns, helping translate organizational principles into

[17] Infosys and MIT Technology Review Insights, "Creating Psychological Safety in the AI Era," MIT Technology Review, December 16, 2025, https://www.technologyreview.com/2025/12/16/1125899/creating-psychological-safety-in-the-ai-era/; Gísli Rafn Ólafsson, "Engineer the Culture," in The AI Fluent Leader: The Art of Leading When AI Changes Everything (self-published, 2026).

[18] FranklinCovey, "AI in the Workplace."; Gallup, "Manager Support Drives Employee AI Adoption," https://www.gallup.com/workplace/694682/manager-support-drives-employee-adoption.aspx; Hong-Yu Wang, Rui-Heng Liu, and Lei Ao, "Transformational Leadership and Employee AI Usage: The Role of Perceived Organizational Support and Competitive Workplace Climate," *Frontiers in Psychology* 16 (2025): article 1581337, https://doi.org/10.3389/fpsyg.2025.1581337.

daily practice.[19] Ultimately, the strongest driver of adoption is usefulness. People will willingly engage with technology that solves real operational problems, while attempting to force adoption of tools that provide little practical value is likely to generate ongoing resistance and undermine future efforts.[20] Successful adaptation means focusing not only on technology, but on creating the trust, psychological safety, and support structures that enable people to engage with it thoughtfully.

Learn through Pilots

"Even if the pilot doesn't work and doesn't lead anywhere, it still served its purpose as a pilot... for me, it did not fail. The goal of it was to learn." - Zineb Bhaby

Organizations do not learn about AI through training and policy development alone, they learn through practice. Pilots offer an opportunity to explore use cases, surface risks, and understand operational implications before significant resources are committed. The challenge is selecting and designing pilots carefully so that they generate organizational learning rather than isolated proofs of concept.

Pilots are primarily learning mechanisms, rather than deployment mechanisms. Their product is knowledge, not a finished tool. A pilot that proves a solution is ineffective or too risky is just as valuable as one that scales, if it prevents future investment in the wrong approach. That value only materializes under three conditions: a culture where staff can admit they failed without fear, documentation of what went wrong, and safety built into the design from the beginning, so failure is not accompanied by harm.[21] These conditions allow organizations to learn from their mistakes rather than repeatedly make the same ones.

[19] James Brennan and Andrej Verity, "Getting Past the Hype: Navigating the Future of Humanitarian Innovation," Digital Humanitarian Network, August 2026, https://digitalhumanitarians.com/getting-past-the-hype-navigating-the-future-of-humanitarian-innovation-august-2026/.
[20] Gallup, "Manager Support Drives Employee AI Adoption."; Henley Business School, World of Work Institute, "AI Attitudes: A 12 Month Pulse Check" (Henley Business School, 2026), https://assets.henley.ac.uk/v3/fileUploads/HBS-survey-report_AI-pulse-check-2026_FINAL.pdf; NetHope, "Humanitarian AI: What to Expect in 2026," https://nethope.org/articles/humanitarian-ai-what-to-expect-in-2026/.
[21] Infosys and MIT Technology Review Insights, "Creating Psychological Safety in the AI Era"; Gísli Rafn Ólafsson, "Architect the Vision," in The AI Fluent Leader.

At the same time, ongoing pilots that never scale can become an exhausting and resource-intensive exercise, creating the "pilot-itis" that challenges humanitarian innovation. The underlying diagnosis is often the same: pilots fail to scale because they are selected in response to technology or donor priorities rather than organizational need, or because discussion of scaling was left until the end of the process. The solution is not necessarily fewer pilots, but more intentional design. Pilots that address a real operational problem, with context, risks, and sustainability considered from the outset, have the most potential to demonstrate immediate value and inform organizational decisions.[22] Alignment with mission goals, predefined success criteria, and explicit scale-or-stop decisions ultimately determine whether pilots contribute to shared learning and organizational adaptation.

Break Down Silos

"Getting people to collaborate across different groups is the most important thing you can do."

- Heather Leson

Adapting to AI is an organizational challenge, and it works best when expertise, experimentation, and lessons move across organizational boundaries. When they do not, organizations incur the cost of solving the same problems multiple times, repeating the same mistakes, and developing disconnected approaches to similar challenges. Breaking down silos is therefore not simply a collaboration objective, but a mechanism for accelerating organizational learning.

Well-designed AI solutions require operational, technical, and leadership perspectives from the beginning. Tools built by technical teams in isolation are often out of touch with operational realities, while solutions developed by field staff without support may be risky, unsustainable, or technically ineffective. Organizations should create structures through

[22] David McClure, Leah Bourns, and Alice Obrecht, "Humanitarian Innovation: Untangling the Many Paths to Scale," Global Alliance for Humanitarian Innovation (GAHI), 2018, https://www.alnap.org/help-library/humanitarian-innovation-untangling-the-many-paths-to-scale; Gísli Rafn Ólafsson, "Architect the Vision."

which staff across departments are brought together from the outset to determine what the operational problems are, how technology might address them, and how tools can be designed to work within the context and align with organizational priorities.[23]

Equally important is ensuring that lessons from prior experiments are visible across teams and over time. Choosing to adapt an existing solution rather than build a new one or avoiding a previously unsuccessful approach depends on knowledge being documented and accessible.[24] Creating findable archives, newsletters, discussion spaces, and visible point people ensures that staff across the organization know what has already been done and transforms lessons learned from one project into something that informs future decision

Pillar 3: Responsible Design & Governance

Innovation can appear risky in the humanitarian sector, particularly when new solutions fail to consider the implications of errors, data breaches, unforeseen consequences, or downstream impacts on affected populations. Adapting in a manner that aligns with humanitarian principles requires that ethics, risk management, and design discipline are built into initiatives from the beginning rather than added after deployment. The ability to experiment, learn, and adapt depends upon ensuring that experimentation itself does not create harm. **Responsible innovation therefore relies on governance that is practical, proportional to the stakes, grounded in strong data foundations, and carefully integrated into the design process.** The goal is not to stifle adaptation, but to ensure that it does not come at the cost of accountability, trust, equity, or the well-being of the people humanitarian organizations exist to serve.

[23] For a concrete institutional example of this kind of mechanism, see UNHCR Innovation's Incubator Programme: https://www.unhcr.org/innovation/sites/default/files/2026-07/2026%20UNHCR%20Innovation%20Incubator%20Programme%20101_0.pdf

[24] Gísli Rafn Ólafsson, "Architect the Vision."

Proportionate Governance

"There's still a lot of work to do on internal governance on AI, like internal oversight that doesn't crush innovators, but gives them clear guidance of what they can do and what they cannot do."

- Rebeca Moreno Jiménez

The purpose of governance is to enable safe innovation by creating shared understanding, meaning it must be practical, clear, and easy to apply. A long policy that nobody reads gets bypassed, and one written for technical specialists will not be consistently understood. Effective governance helps staff understand what is permitted, what needs additional review, and where to seek guidance, while leaving room for departments to determine how those principles apply within their own contexts, rather than attempting to anticipate every use case from the center.

Governance should also be proportionate to risk. Applying identical controls to every use case can be as damaging as having no controls at all, since under-governance leaves high-stakes applications unchecked, and over-governance buries low-stakes experimentation in unnecessary bureaucracy. Tiered governance allows organizations to focus limited oversight capacity where it matters most.[25] The CDAC Network's SAFE AI Framework offers a workable model, providing three tiers sorted by use-case rather than by tool: Tier 1 covers internal, low-risk uses needing only light-touch oversight; Tier 2 covers uses that shape operational decisions, calling for more formal assessment; and Tier 3 covers cases that directly affect people's access to assistance, protection, or information, where community participation, and rigorous, ongoing monitoring become essential.[26] Governance frameworks must also recognize that AI is not always the right solution. AI should never make final decisions about a person's eligibility for assistance or protection, be given sensitive data through public or unapproved tools, or inform a high-stakes judgment without meaningful

[25] DSEG, *A Framework for the Ethical Use of Advanced Data Science Methods in the Humanitarian Sector*, version 1 (April 2020), https://www.hum-dseg.org/sites/g/files/tmzbdl1476/files/2020-10/Framework%20for%20the%20ethical%20use.pdf; ICRC, *Building a Responsible Humanitarian Approach: The ICRC's Policy on Artificial Intelligence* (Geneva: ICRC, November 2024), https://www.icrc.org/en/publication/building-responsible-humanitarian-approach-icrcs-policy-artificial-intelligence.

[26] Helen McElhinney et al., *Standards and Assurance Framework for Ethical AI in Humanitarian Action (SAFE AI): A Governance Framework for Humanitarians Using AI*, version 1.2, CDAC Network, 2026, https://doi.org/10.5281/zenodo.22080937.

human accountability.[27] Refusing to use AI in inappropriate contexts is not a failure of innovation but evidence of sound decision-making.

Finally, governance must be treated as an ongoing process rather than a one-time approval exercise. As models are updated, data sources change, operational contexts evolve, and user behavior adapts, the risks and performance of an AI-enabled process may change as well. Oversight should therefore include periodic review, incident reporting, reassessment triggers, feedback mechanisms, and clearly assigned ownership for evaluating whether a use case remains within its approved risk category. This ensures that deployed systems remain safe, reliable, and aligned with organizational objectives over time.

Responsible Design

"Sometimes it's good to stop and think whether the technology is actually needed. What's the value that technology can bring and what are the potential harms that it could create? Is that worth doing?" - Ana Beduschi

The greatest failures of humanitarian innovation occur when system breakdowns, data breaches, or poorly designed interventions cause harm to the very people organizations exist to support. Ensuring that AI systems align with humanitarian principles requires building safety into their design from the outset, beginning by asking whether technology is the appropriate solution at all.[28] The question is not only whether a technology *can* solve a problem, but whether it is the safest, simplest, and most suitable way of doing so, and whether sufficient data, capacity, and oversight exist to support it. Sometimes the most responsible decision is to solve the problem in a different way.

[27] DSEG, *A Framework for the Ethical Use of Advanced Data Science Methods in the Humanitarian Sector*; ICRC, *Building a Responsible Humanitarian Approach.*

[28] Philip Brey and Brandt Dainow, "Ethics by Design for Artificial Intelligence," AI and Ethics 4 (2024): 1265–1277, https://doi.org/10.1007/s43681-023-00330-4; DSEG, *A Framework for the Ethical Use of Advanced Data Science Methods in the Humanitarian Sector*; ICRC, *Building a Responsible Humanitarian Approach.*

The interests of affected populations must remain the primary consideration throughout the design process. Humanitarian organizations should never treat vulnerable populations as subjects for experimentation or collect sensitive data simply to improve a system.[29] Tools that may impact affected populations should be developed in collaboration *with* those communities rather than *for* them.[30] Co-design contributes to both safety and performance by ensuring that systems reflect the realities, priorities, and constraints of those most likely to experience their consequences.

Responsible design also includes safe, deliberate testing. Approaches similar to regulatory sandboxes allow tools to be tested under controlled conditions before they affect real decisions.[31] Testing should involve multiple perspectives, and assess more than whether a tool is functional, but also whether it is biased, inaccessible, or a security vulnerability.[32] Technical experts can identify model failures, field practitioners can catch contextual and operational issues, and intended users can highlight barriers to usability and comprehension. Success in one context does not guarantee success in another. Humanitarian environments differ significantly in language, culture, infrastructure, and patterns of vulnerability, meaning systems must be evaluated according to the specific contexts in which they will operate.

Teams should also actively anticipate how systems might fail. Gísli Ólafsson highlights the value of the "*pre-mortem"* exercise, in which a team assumes a deployment has already failed and works backwards to identify the most plausible causes.[33] Asking a question like: "*it is two years after we deployed this system and we are in a leadership meeting discussing a serious problem it has caused — what is the most likely problem?"* helps counter optimism bias and surface risks that might otherwise be overlooked. Often risk cannot be eliminated entirely, but

[29] Kristin Bergtora Sandvik, Katja Lindskov Jacobsen, and Sean Martin McDonald, "Do No Harm: A Taxonomy of the Challenges of Humanitarian Experimentation," *International Review of the Red Cross* 99, no. 1 (2017): 319–344, https://international-review.icrc.org/sites/default/files/irrc_99_17.pdf.
[30] Helen McElhinney and Sarah W. Spencer, "The Clock Is Ticking to Build Guardrails into Humanitarian AI," The New Humanitarian, March 11, 2024, https://www.thenewhumanitarian.org/opinion/2024/03/11/build-guardrails-humanitarian-ai.
[31] ICRC,, *Building a Responsible Humanitarian Approach*; Aaron Martin and Giulia Balestra, "Using Regulatory Sandboxes to Support Responsible Innovation in the Humanitarian Sector," Global Policy 10, no. 4 (2019): 733–736, https://doi.org/10.1111/1758-5899.12729.
[32] Audrey Tang and Caroline Emmer De Albuquerque Green, "Pack 3: Competence — Care-Giving," Civic AI, Oxford Institute for Ethics in AI, Accelerator Fellowship Programme, https://civic.ai/3/.
[33] Gísli Rafn Ólafsson, "Guard the Guardrails," in The AI Fluent Leader.

it can be minimized by identifying foreseeable harms early enough for them to be addressed through design changes, additional safeguards, or, where necessary, a decision not to proceed.

Strengthen Data Foundations

"The vast chunk of all that knowledge that is generated with a lot of effort is usually just lost to the void and entropy. What I think is very interesting for organizations like ours is how to systematically institutionalize this knowledge in a way that is understandable." - Jacopo Margutti

AI systems are only as reliable as the information that supports them. Organizations that struggle to locate, understand, or trust their own information will face difficulties when attempting to build, deploy, or govern AI systems. Strengthening data foundations requires more than collecting data. It requires stewardship, documentation, knowledge management, information discoverability, quality assurance processes, and clear accountability for how information is maintained and used. These foundations enable organizations to preserve institutional memory, make knowledge more accessible, and provide AI systems with information that is accurate, current, and understandable.

Particular attention should be paid to the quality and representativeness of data.[34] Datasets frequently contain gaps, outdated information, and embedded biases that AI systems can reproduce or amplify, so organizations should understand where their data originates, who is represented within it, who may be missing, and how these limitations might influence outputs or decisions. Poor data quality can become a safety issue when decisions rest on information that is inaccurate, incomplete, or poorly documented.[35] Privacy deserves the same attention, particularly given the sensitivity of humanitarian data.[36] Organizations need to know what data an AI system touches, where it is stored, who can access it, and whether it ever leaves

[34] DSEG, *A Framework for the Ethical Use of Advanced Data Science Methods in the Humanitarian Sector*; ICRC, *Building a Responsible Humanitarian Approach.*
[35] Beduschi, "Harnessing the Potential of Artificial Intelligence for Humanitarian Action."
[36] Beduschi, "Harnessing the Potential of Artificial Intelligence for Humanitarian Action"; International Committee of the Red Cross, Building a Responsible Humanitarian Approach.

organizational control, especially through the use of commercial tools. Quality and privacy can pull in opposite directions, since improving data often means making it more detailed or identifiable, which raises the potential for harm if it is ever mishandled or exposed. Data quality cannot come at the cost of the security or consent of affected populations.[37] If high-quality data cannot be obtained without putting people at risk, AI is not the right approach for the situation.

Strong data foundations mean balancing usefulness with security to get both effectiveness and safety right. Well-managed information builds confidence in AI outputs, supports transparency and accountability, and reduces avoidable error. Without it, even a well-designed, well-governed AI system is likely to produce unreliable output and undermine trust in how the organization uses AI.

Pillar 4: Ecosystem Coordination

In the same way that AI changes coordination, information management, and expertise distribution on an organizational level, it may also reshape how these functions operate across the wider humanitarian ecosystem. Tasks that historically required large, centralized organizations may become easier to perform locally, at the same time as entirely new coordination challenges emerge. Without stronger collaboration, AI could deepen divides between well-resourced organizations and smaller or local actors, increase duplication, and erode the trust that impactful response requires. Organizations therefore need to consider their responsibilities not only internally, but across the wider humanitarian ecosystem.

[37] Mirca Madianou, "Technocolonialism: Digital Innovation and Data Practices in the Humanitarian Response to Refugee Crises," *Social Media + Society* 5, no. 3 (2019), https://doi.org/10.1177/2056305119863146; McElhinney and Spencer, "The Clock Is Ticking to Build Guardrails into Humanitarian AI."; Sandvik, Jacobsen, and McDonald, "Do No Harm."

Shared Knowledge and Infrastructure

"Who is doing what? Because we're a network, right? Let's not duplicate." - Interviewee

Just as fragmented AI adoption within an organization leads to duplication, higher costs, and repeated mistakes, fragmentation across the humanitarian sector risks producing the same outcomes at a much larger scale.[38] Many organizations face similar AI-related challenges yet lack the resources to address them independently. Collaboration offers a practical solution. Humanitarian organizations collectively possess vast amounts of contextual knowledge, data, operational learning, and technical expertise. Made accessible, these assets can support sector-wide tools, improve output quality, and spare individual organizations from repeatedly recreating the same resources. AI increases the value of shared knowledge because the performance of AI systems depends on the quality of the information and expertise available to support them.

Opportunities for collaboration extend from shared governance frameworks to open knowledge repositories, interoperable data standards, and common infrastructure. Communities of practice can enable organizations to exchange lessons and innovate together, reducing duplication while accelerating organizational learning. Open-source models can build collective capacity, give humanitarian actors greater control over how systems are built, and reduce dependence on private technology partnerships. Data-sharing arrangements, supported by appropriate privacy safeguards and informed consent, can allow organizations to pool the contextual knowledge that AI relies on, rather than collecting and maintaining similar datasets in isolation.

Collaboration also improves financial sustainability. Individual organizations may struggle to justify investment in specialized infrastructure or expertise on their own, but shared development spreads the costs and can make initiatives possible that would otherwise be unaffordable. Collective projects may also be more attractive to donors than multiple

[38] Zineb Bhaby, "AI in the Era of Humanitarian Reset: Challenges and Opportunities," Centre for Humanitarian Action, September 23, 2025, https://www.chaberlin.org/en/blog/ai-in-the-era-of-humanitarian-reset-challenges-and-opportunities/.

organizations independently developing similar tools. Building shared knowledge and infrastructure therefore serves both practical and strategic goals by reducing duplication, lowering costs, accelerating learning, and ensuring the benefits of AI are distributed across the humanitarian ecosystem.

Strengthen Local Capacity

"International organizations, including UN agencies, they have more resources... but when it comes to local actors, keeping in mind the localization agenda, they are at the forefront of any kind of crisis."

- Qasir Rafiq

AI could support local organizations by expanding access to capabilities that were previously concentrated within larger international organizations. Assistance with reporting, translation, information access, and data analysis may allow local actors to perform tasks that would otherwise require specialist support.[39] Yet, despite adopting AI most rapidly, small, local organizations often lack the resources to train staff, negotiate with technology providers, build their own systems, hire technical specialists, or participate in sector-wide governance discussions.[40] If access to the most advanced tools, technical expertise, and governance processes remains concentrated within the most well-resourced organizations, humanitarian AI may widen the existing inequalities across the sector, and it cannot be called ethical, principled, or effective on those terms.[41]

Local actors often possess the deepest understanding of social networks, cultural contexts, and community needs through lived experience and long-standing relationships. This knowledge is what keeps humanitarian action appropriate, legitimate, and responsive to local realities. Leaving it behind in AI adoption would not only reinforce existing patterns of exclusion, but undermine the quality of humanitarian response. AI often performs best when

[39] McElhinney and Spencer, "The Clock Is Ticking to Build Guardrails into Humanitarian AI."
[40] Humanitarian Leadership Academy and NetHope, "Collective Action on Localised Humanitarian AI: Unlocking Solutions Together," July 30, 2026, https://www.humanitarianleadershipacademy.org/resources/collective-action-on-localised-humanitarian-ai-unlocking-solutions-together/.
[41] Madianou, "Technocolonialism."

it is built upon the knowledge that exists locally, and local actors are usually best positioned to identify important use cases, define operational requirements, and recognize unintended consequences before they cause harm.[42] Supporting localization means treating local actors as more than recipients of technology, it means ensuring their ability to design, govern, and own these tools themselves.

Our interviews highlighted a strong desire among local organizations for training, AI literacy, and participation in sector-wide conversations that are often inaccessible to actors in the Global South. Larger organizations can contribute to a more equitable transition by openly sharing training materials, supporting sector-wide learning, providing technical assistance to organizations with fewer specialists, and investing in locally-led innovation, while remaining aware of the difference between increasing local capability and creating new forms of dependency. The ultimate objective should be to expand local actors' ability to shape and direct the use of AI within their own contexts.

Preserve Human Relationships and Trust

"Does that mean we will have less people in the field doing things or less people working on things? Not necessarily. It just means that we can provide better work or better assistance to those in need because we're not spending our time on bureaucracy." - Gísli Ólafsson

AI may automate many aspects of information management, analysis, reporting, and knowledge retrieval, reducing the time and effort organizations spend on knowledge work. This raises a strategic question for humanitarian organizations: what remains uniquely human and uniquely valuable when information itself becomes easier to generate, analyze, and distribute?

Humanitarian coordination has always relied on information, but it is not primarily an information problem. It depends on relationships, trust, negotiation, facilitation, and the ability

[42] Humanitarian Leadership Academy and NetHope, "Collective Action on Localised Humanitarian AI."

to bring actors with competing priorities around shared objectives. AI-mediated communication and self-service knowledge systems may increase efficiency, but these gains come at a cost if they reduce the interaction, collaboration, and collective sense-making through which trust is built. **AI may therefore increase the importance of relationship-centered functions**. As information becomes more abundant, shared understanding becomes more valuable. As expertise becomes more widely distributed, trust is what allows people to work together. As analytical capability becomes more accessible, the ability to convene actors around common goals becomes a distinguishing contribution.

Let this inform how AI is used within humanitarian organizations. Allow it to absorb the administrative burden and routine knowledge work, creating more capacity for engagement with affected communities, partners, and one another. **The goal is less time behind a computer and more time building the trust, understanding, and cooperation that humanitarian action depends upon**.

Recommendations for Humanitarian Organizations

- **Establish structured discussion spaces.** Create regular forums, labs, newsletters, and communities of practice where ideas, mistakes, operational needs, and concerns are shared openly.

- **Ensure AI literacy across all roles.** Provide every staff member with a practical baseline in AI limitations, risks, responsible use, and critical evaluation, role-specific guidance for particular use cases, and deeper support for those leading AI initiatives.

- **Classify risk and govern accordingly**. Identify the difference between low-stakes and high-stakes uses. Build clear, practical governance that matches oversight intensity to the stakes of the decision. Treat monitoring as an ongoing process.

- **Invest in data and knowledge foundations.** Strengthen stewardship, documentation, quality assurance, discoverability, and accountability so AI systems rely on information that is accurate, current, representative, understandable, and safe to use.

- **Design intentionally from the outset.** Start each initiative by asking whether AI is the right solution, design in response to real operational problems, test safely before deployment, and involve operational, technical, ethical, legal, and user perspectives throughout the entire process.

- **Create the conditions for organizational learning to travel.** Psychological safety to admit what failed, documentation of lessons, and cross-team circulation turn isolated experience into shared knowledge.

- **Protect human judgment where it matters most.** Implement workflows that ensure qualified reviewers have the time, confidence, authority, and escalation pathways needed to challenge AI outputs.

- **Safeguard essential human capabilities.** Name the skills and forms of expertise that AI must not erode, and develop them through hiring, mentorship, training, and practice.

- **Share what can be shared, and invest deliberately in local capacity.** Make training materials, governance approaches, infrastructure, and lessons learned available across the sector whenever possible, and support local actors to shape, govern, and own AI tools in their own contexts.

- **Keep affected populations at the centre of every decision.** AI adoption is only worth doing if it strengthens, rather than compromises, the safety, dignity, and agency of the people humanitarian organizations exist to serve.

Annex 1: Methodology & Scope

This paper synthesizes findings from a literature review and key informant interviews with experts from humanitarian, academic, technology, and diplomatic backgrounds who brought a range of perspectives on AI, digital innovation, and organizational adaptation in humanitarian action.

This paper is presented as a think-brief in order to help start a conversation or help provide a concrete stepping stone for humanitarian organizations considering how to adapt responsibly to AI. 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.

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 AI, innovation, and organizational adaptation in humanitarian action. 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 or their organizations unless explicitly stated. The authors have drafted the paper based on a combination of literature review and interviews.

Interviewees listed in alphabetical order based on their first name. Three additional interviews were conducted with individuals who requested anonymity.

Name	Organization
Ana Beduschi	University of Exeter
Brent Phillips	Humanitarian AI Today
Cornelia Walther	Sunway University
Daniela Weber	NetHope
Daniela Wuerz	UNOG
David Higgins	Personal capacity
Gautham Krishnaraj	Humanitarian Partners
Gísli Ólafsson	Gridland Consulting
Heather Leson	Canadian Red Cross
Ilen Madhavji	UNTMIS
Jacopo Margutti	The Netherlands Red Cross
Jerri Husch	2Collaborate Consulting
Ka Man Parkinson	Humanitarian Leadership Academy
Paola Yela	IFRC
Qasir Rafiq	ACF Sudan
Rebeca Moreno Jiménez	UNHCR
Sher Verick	International Labour Organization
Somya Gupta	Google
Stefano Iacus	European Commission
Zineb Bhaby	Norwegian Refugee Council

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: References

Bali, Maha. "What I Mean When I Say Critical AI Literacy." Reflecting Allowed, 2023. https://blog.mahabali.me/educational-technology-2/what-i-mean-when-i-say-critical-ai-literacy/.

Beck, Jacob, Stephanie Eckman, Christoph Kern, and Frauke Kreuter. "Bias in the Loop: How Humans Evaluate AI-Generated Suggestions." Harvard Data Science Review 8, no. 2 (2026). https://hdsr.mitpress.mit.edu/pub/nrcn4h7d/release/2.

Bhaby, Zineb. "AI in the Era of Humanitarian Reset: Challenges and Opportunities." Centre for Humanitarian Action, September 23, 2025. https://www.chaberlin.org/en/blog/ai-in-the-era-of-humanitarian-reset-challenges-and-opportunities/.

Brennan, James, and Andrej Verity. "Getting Past the Hype: Navigating the Future of Humanitarian Innovation." Digital Humanitarian Network, August 2026. https://digitalhumanitarians.com/getting-past-the-hype-navigating-the-future-of-humanitarian-innovation-august-2026/.

Brey, Philip, and Brandt Dainow. "Ethics by Design for Artificial Intelligence." AI and Ethics 4 (2024): 1265–1277. https://doi.org/10.1007/s43681-023-00330-4.

Buçinca, Zana, Maja Barbara Malaya, and Krzysztof Z. Gajos. "To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-Assisted Decision-Making." Proceedings of the ACM on Human-Computer Interaction 5, no. CSCW1 (2021): article 188. https://doi.org/10.1145/3449287.

Buteau, Antoine. "AI Can Drain the Expert Pipeline Before Anyone Notices." Antoine Buteau (blog), August 9, 2026. https://www.antoinebuteau.com/ai-can-drain-the-expert-pipeline-before-anyone-notices/.

Cantrell, Sue, Michael Griffiths, Robin Jones, and Julie Hiipakka. Building Tomorrow's Skills-Based Organization: Jobs Aren't Working Anymore. Deloitte, 2022. https://www.deloitte.com/content/dam/assets-shared/legacy/docs/Deloitte-Skills-Based-Organization.pdf.

Data Friendly Space, and Humanitarian Leadership Academy. "How Are Humanitarians Using AI in 2025?" Data Friendly Space, July 15, 2025. https://www.datafriendlyspace.org/resources/initial-insights-report.

Data Friendly Space, and Humanitarian Leadership Academy. "75% of Humanitarians Now Use AI Regularly, Yet Fewer Than One in Ten Work in an Organisation Where AI Is Widely Integrated." Humanitarian Leadership Academy, 2026. https://www.humanitarianleadershipacademy.org/news/75-percent-of-humanitarians-now-use-ai-regularly-yet-fewer-than-one-in-ten-work-in-an-organisation-where-ai-is-widely-integrated.

FranklinCovey. "AI in the Workplace: A Leader's Guide to Human-Centered AI Adoption." By Jon Lofgren. Updated July 6, 2026. https://www.franklincovey.com/blog/embrace-ai-in-the-workplace/.

Gallup. "Manager Support Drives Employee AI Adoption." 2025. https://www.gallup.com/workplace/694682/manager-support-drives-employee-adoption.aspx.

Henley Business School, World of Work Institute. "AI Attitudes: A 12 Month Pulse Check." Henley Business School, 2026. https://assets.henley.ac.uk/v3/fileUploads/HBS-survey-report_AI-pulse-check-2026_FINAL.pdf.

Husch, Jerri. "AI for Humanitarian Coordination: A Practical Brief for OCHA." January 2026.

Humanitarian Leadership Academy and NetHope. "Collective Action on Localised Humanitarian AI: Unlocking Solutions Together." July 30, 2026. https://www.humanitarianleadershipacademy.org/resources/collective-action-on-localised-humanitarian-ai-unlocking-solutions-together/.

Infosys and MIT Technology Review Insights. "Creating Psychological Safety in the AI Era." MIT Technology Review, December 16, 2025. https://www.technologyreview.com/2025/12/16/1125899/creating-psychological-safety-in-the-ai-era/.

International Labour Organization. *Lifelong Learning and Skills for the Future*. World of Work Series. Geneva: ILO, May 5, 2026. https://doi.org/10.54394/00033011.

Kupfer, Cordula, Rita Prassl, Jürgen Fleiß, Christine Malin, Stefan Thalmann, and Bettina Kubicek. "Check the Box! How to Deal with Automation Bias in AI-Based Personnel Selection." Frontiers in Psychology 14 (2023): article 1118723. https://doi.org/10.3389/fpsyg.2023.1118723.

Lee, Hao-Ping, Advait Sarkar, Lev Tankelevitch, Ian Drosos, Sean Rintel, Richard Banks, and Nicholas Wilson. "The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects from a Survey of Knowledge Workers." In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems. New York: Association for Computing Machinery, 2025. https://doi.org/10.1145/3706598.3713778.

Lovett, Nolan. "The Tragedy of the Cognitive Commons: How AI Could Disrupt the Regeneration of Professional Expertise." Human Resource Development Review, advance online publication, July 26, 2026. https://doi.org/10.1177/15344843261470602.

Martin, Aaron, and Giulia Balestra. "Using Regulatory Sandboxes to Support Responsible Innovation in the Humanitarian Sector." Global Policy 10, no. 4 (2019): 733–736. https://doi.org/10.1111/1758-5899.12729.

McClure, David, Leah Bourns, and Alice Obrecht. "Humanitarian Innovation: Untangling the Many Paths to Scale." Global Alliance for Humanitarian Innovation (GAHI), 2018. https://www.alnap.org/help-library/humanitarian-innovation-untangling-the-many-paths-to-scale.

McElhinney, Helen, and Sarah W. Spencer. "The Clock Is Ticking to Build Guardrails into Humanitarian AI." The New Humanitarian, March 11, 2024. https://www.thenewhumanitarian.org/opinion/2024/03/11/build-guardrails-humanitarian-ai.

McElhinney, Helen, Anwesha Mazumder, Marie Tjalve, Sinead Madigan, and Sarah Spencer. Standards and Assurance Framework for Ethical AI in Humanitarian Action (SAFE AI): A Governance Framework for Humanitarians Using AI. Version 1.2. CDAC Network, 2026. https://doi.org/10.5281/zenodo.22080937.

Ólafsson, Gísli Rafn**.** The AI Fluent Leader: The Art of Leading When AI Changes Everything. Self-published, 2026.

Pavuluri, Emily. "Before You Trust the Output: Three Exercises for Teaching AI Evaluation." The AI of Law, 2026. https://vaill.substack.com/p/aievaluationexercises.

Pearson, Joe, Itiel Dror, Emma Jayes, Charlotte Whordley, Sarah Mason, and Sophie J. Nightingale. "Examining Human Reliance on Artificial Intelligence in Decision Making." Scientific Reports 16 (2026): article 5345. https://doi.org/10.1038/s41598-026-34983-y**.**

Pizzi, Michael, Mila Romanoff, and Tim Engelhardt. "AI for Humanitarian Action: Human Rights and Ethics." International Review of the Red Cross 102, no. 913 (2021): 145–180. https://doi.org/10.1017/S1816383121000011.

Romeo, Giuseppe, and Daniela Conti. "Exploring Automation Bias in Human–AI Collaboration: A Review and Implications for Explainable AI." AI & Society 41 (2026): 259–278. https://doi.org/10.1007/s00146-025-02422-7.

Sandvik, Kristin Bergtora, Katja Lindskov Jacobsen, and Sean Martin McDonald. "Do No Harm: A Taxonomy of the Challenges of Humanitarian Experimentation." International Review of the Red Cross 99, no. 1 (2017): 319–344. https://international-review.icrc.org/sites/default/files/irrc_99_17.pdf.

Steidel, Isabell, and Benedikt Gieger. "The Oversight Paradox: Human Control Over AI May Be Eroding." World Economic Forum, July 2, 2026. https://www.weforum.org/stories/artificial-intelligence/oversight-paradox-human-control-ai/.

Tang, Audrey, and Caroline Emmer De Albuquerque Green. "Pack 2: Responsibility — Taking Care Of." Civic AI. Oxford Institute for Ethics in AI, Accelerator Fellowship Programme. https://civic.ai/2/.

Tang, Audrey, and Caroline Emmer De Albuquerque Green. "Pack 3: Competence — Care-Giving." Civic AI. Oxford Institute for Ethics in AI, Accelerator Fellowship Programme. https://civic.ai/3/.

The Ai Consultancy. "Expert in the Loop: Why AI Needs Specialists, Not Just Humans." Medium, April 16, 2025. https://medium.com/@ai_93276/expert-in-the-loop-why-ai-needs-specialists-not-just-humans-5797b6500654.

UNESCO. Recommendation on the Ethics of Artificial Intelligence. Paris: UNESCO, 2021.

Vicente, Lucía, and Helena Matute. "Warning People about the Risk of AI Error Mitigates Human Acquisition of AI Bias." Cognitive Research: Principles and Implications 11 (2026): article 36. https://doi.org/10.1186/s41235-026-00726-w.

Walther, Cornelia C. "Why Hybrid Intelligence Is the Future of Human-AI Collaboration." Knowledge at Wharton, 2025. https://knowledge.wharton.upenn.edu/article/why-hybrid-intelligence-is-the-future-of-human-ai-collaboration/.

Wang, Hong-Yu, Rui-Heng Liu, and Lei Ao. "Transformational Leadership and Employee AI Usage: The Role of Perceived Organizational Support and Competitive Workplace Climate." Frontiers in Psychology 16 (2025): article 1581337. https://doi.org/10.3389/fpsyg.2025.1581337.