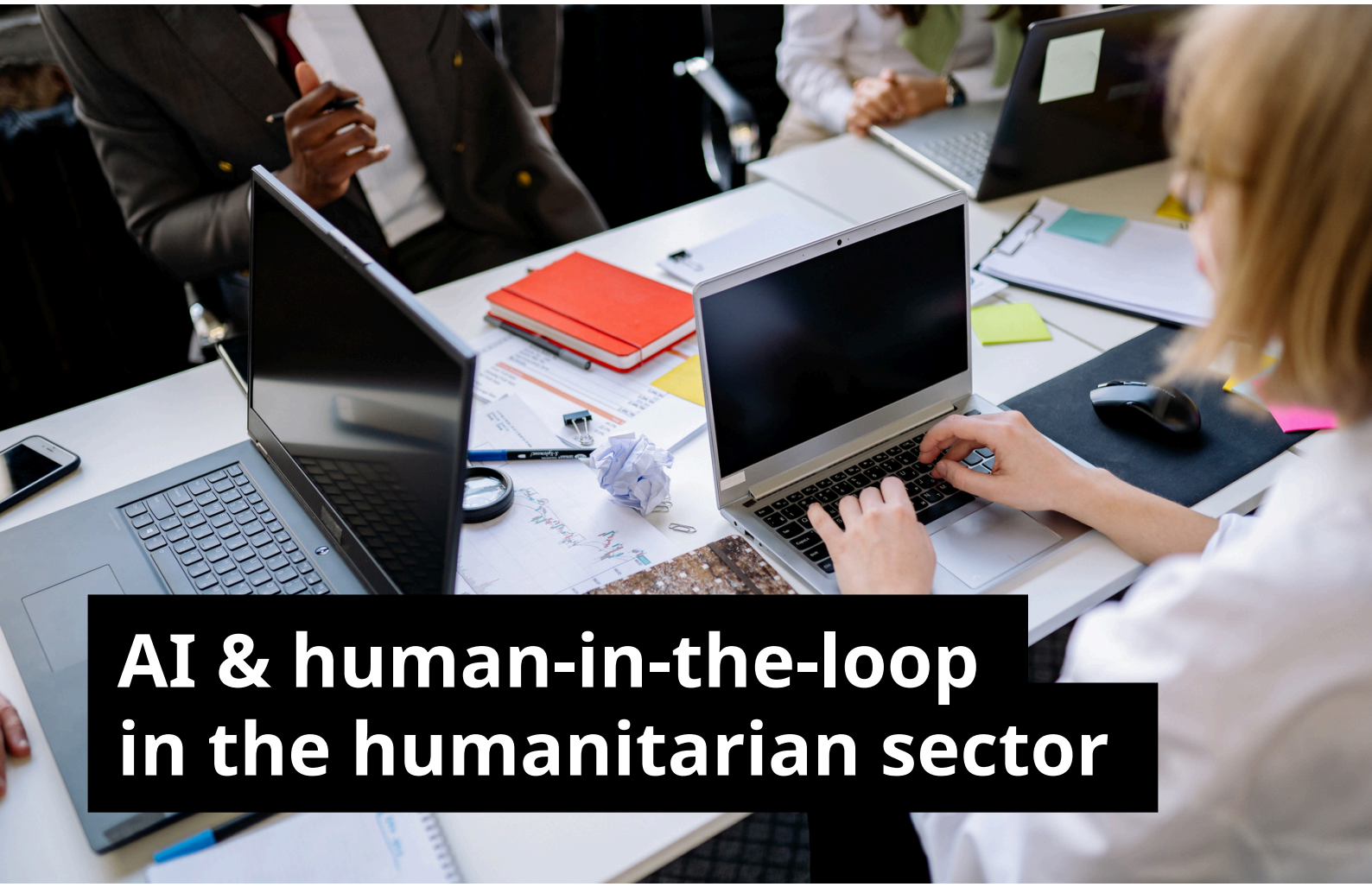
Humanitarian Leadership Academy

DATA FRIENDLY SPACE



AI & human-in-the-loop in the humanitarian sector

Confused by AI terms and not sure how they apply to humanitarian work? This practical guide with expertise from Data Friendly Space provides a beginner-friendly overview of "human-in-the-loop".

This guide has been created as follow-up to the 2025-26 joint research initiative between the Humanitarian Leadership Academy and Data Friendly Space: *Artificial intelligence in the humanitarian sector: mapping current practice and future potential.*

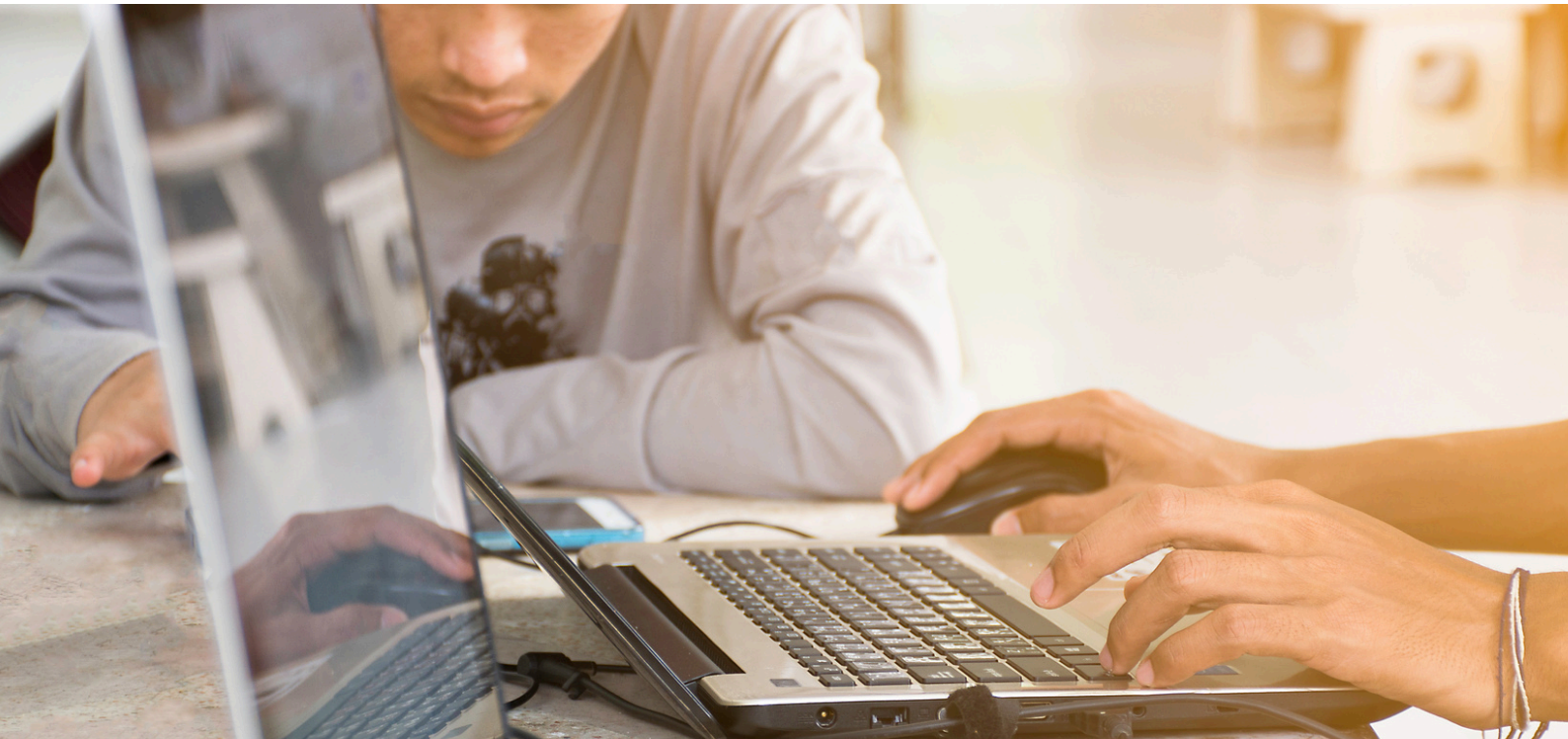
AI is being used across humanitarian work, but not always safely

In 2026, 75% of humanitarian practitioners are using AI tools regularly for work. Yet, fewer than one quarter of organisations have a formal AI policy.

When AI operates without human guidance, it can:

- Generate plausible but factually wrong information
- Miss critical on-the-ground context
- Reflect and amplify existing biases in data

In humanitarian contexts, acting on flawed AI outputs doesn't just waste resources - it can direct aid away from the people who need it most. That is why every AI output needs to be combined with human oversight.



What does ‘human-in-the-loop’ (HITL) mean?

Human-in-the-Loop (HITL) is an approach in which people actively guide, validate, improve, and/or override AI's output before action is taken.

This approach combines the strengths of both humans and machines, ensuring that complex or high-stakes tasks benefit from human judgement and oversight.

HITL is not about slowing AI down - it is about making sure AI works *for* people, not just *about* people.

HITL is not a single checkpoint. It is a continuous cycle built into how AI is used, ensuring the right people remain accountable to their organisation and to the people they serve.

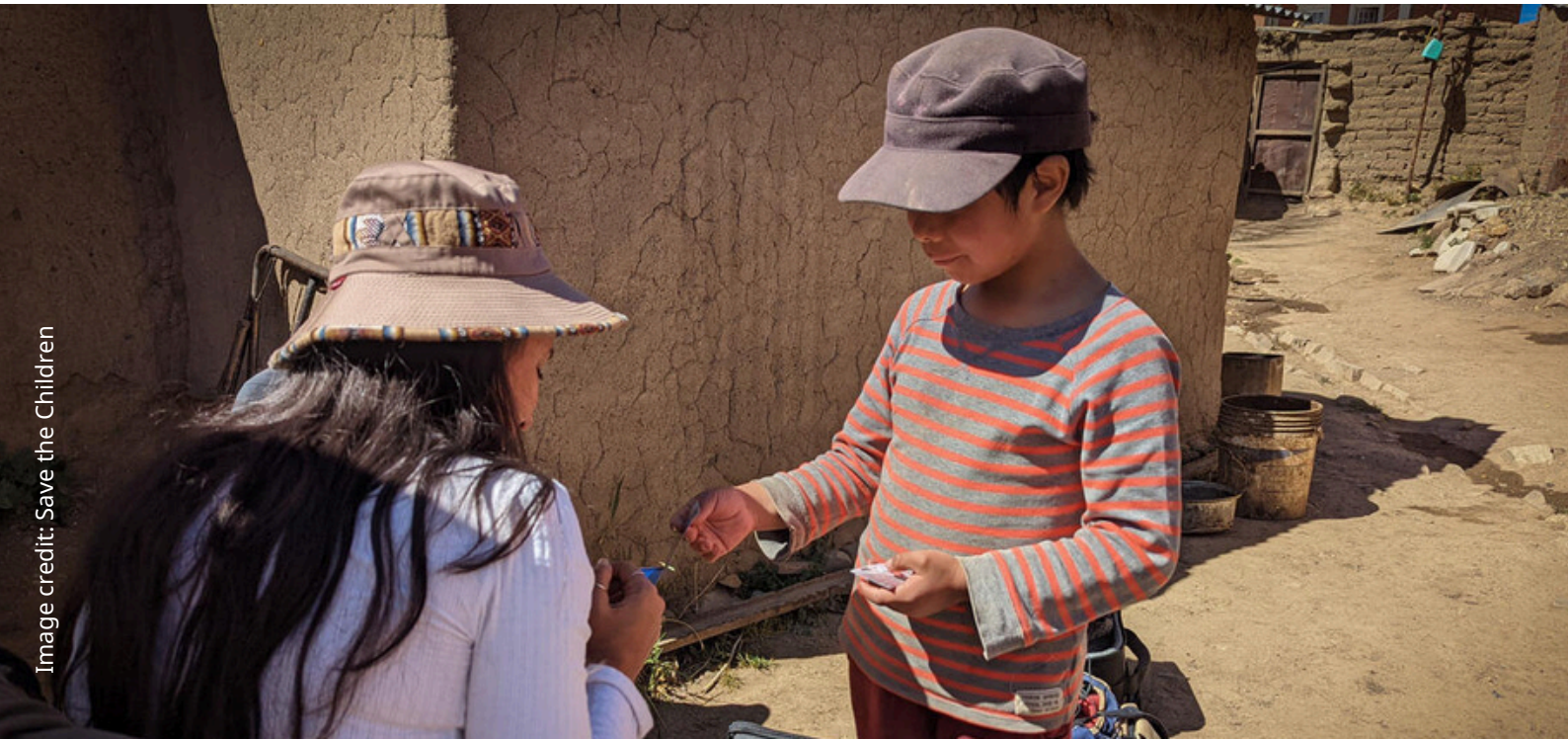


Image credit: Save the Children

Why does human-in-the-loop matter in humanitarian contexts?

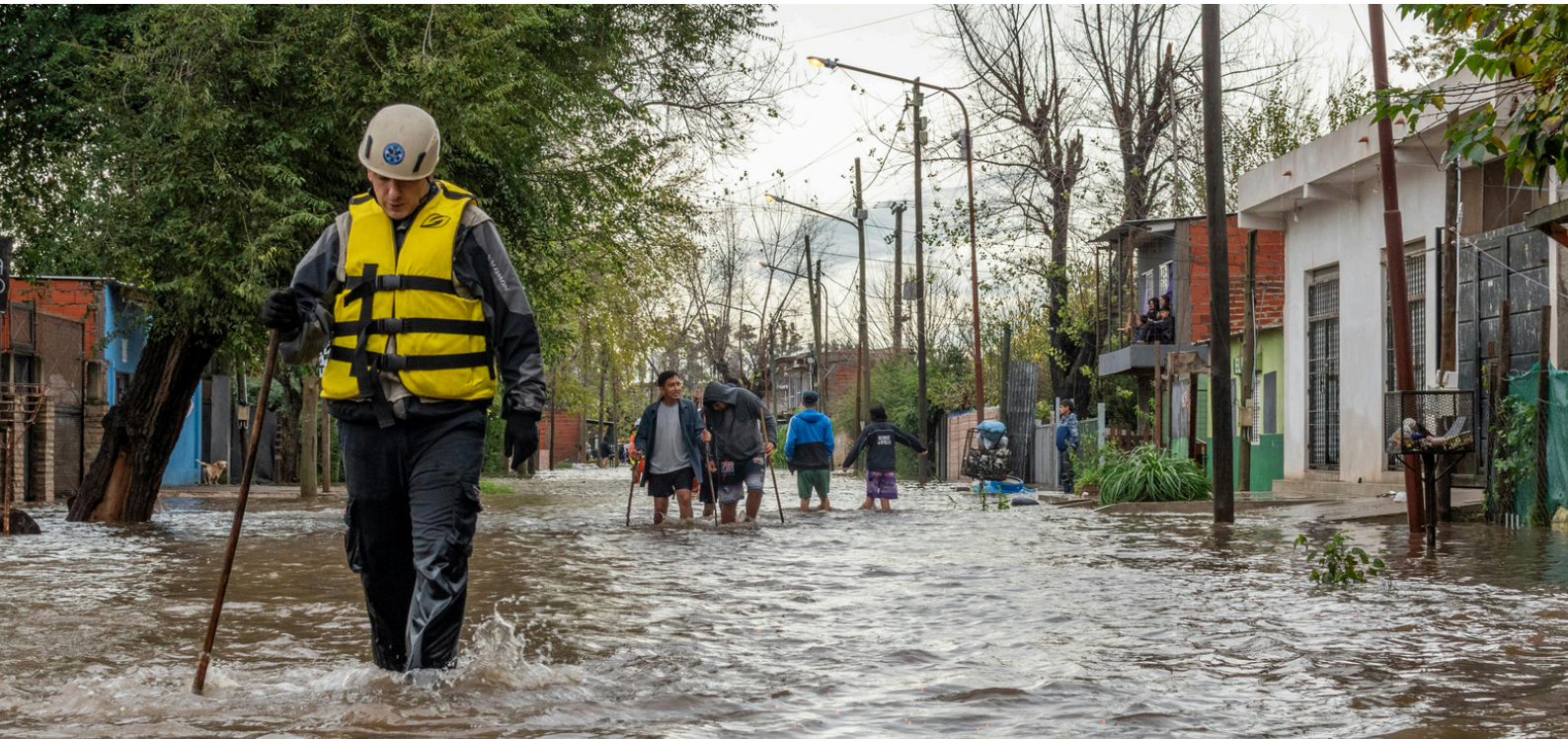
Do no harm applies to AI too.

AI can miss context. It can reflect bias in training data. AI can't fully understand the lived experience of crisis-affected people.

If an algorithm flags the wrong family for exclusion or misclassifies a protection risk, the harm is real.

Harm isn't only physical. Poorly-managed AI can expose sensitive data, re-identify individuals, or enable surveillance. Oversight protects people on every level.

Human review is how we catch that, before it reaches the people we're trying to help.



Human-in-the-loop should not be an abstract framework

Human-in-the-loop should be implemented at every level of AI usage. What does this look like?

- **Define decision points early:** Before deploying any AI tool, map out exactly where human review must happen. Don't leave it to chance or assume someone else is checking.

- **Match the reviewer to the decision:** The person reviewing an AI output should have the knowledge and authority to act on it. A WASH officer, a protection specialist, and a data analyst will each catch different things.

- **Make it easy to override:** If overriding an AI recommendation feels risky, bureaucratic, or discouraged, people won't do it. Build cultures and systems where challenging AI outputs are the norm.

- **Document every override:** When a human changes or rejects an AI output, record it. These decisions are valuable. They reveal where AI is underperforming and improve future models, and help others learn what to watch out for.

Humans bring specialised skills that AI cannot replicate

Your expertise is not supplementary to AI: it is what makes AI safe to use.

- **Local knowledge:** Data means different things in different places. A number that looks normal in one context can signal a crisis in another. Only people who know a place can make that call.

- **Ethical judgement:** Deciding what to publish, what to protect, and who might be affected is an act of accountability. That cannot be delegated to an algorithm.

- **Cultural sensitivity:** AI cannot reliably recognise inappropriate framings, harmful assumptions, or culturally specific blind spots. Human reviewers catch what models miss.

- **Reading the gaps:** The most important insight is sometimes what the data does *not* say. Who is uncounted, what is absent, what silence means. AI has no way to know what it doesn't know.

How else can we keep humans meaningfully in control?

Human-in-the-loop is only as strong as the culture and processes behind it.

- **Include affected communities:** Human-in-the-loop shouldn't only mean staff. Where possible, create channels for crisis-affected people to flag errors, ask questions, or contest decisions made about them.

- **Watch for automation bias:** People tend to trust AI outputs, especially when under pressure. Train your teams to actively question outputs, not just rubber-stamp them.

- **Revisit oversight as tools evolve:** AI systems change. The oversight process you designed six months ago may no longer be sufficient. Build in regular checkpoints to reassess whether human review is still meaningful.

- **Assign clear accountability:** Every AI-informed decision should have a named human accountable for the outcome. If no one owns it, no one will catch it when it goes wrong.

- **Act when things go wrong:** If an AI output causes harm or a serious error is made, stop, report it, and make sure the right people know. Accountability means acting on what went wrong.

Not every AI output carries the same risk: when in doubt, verify

Knowing when to look harder is a core HITL skill. The time it takes to check is always less than the cost of acting on a flawed finding. This may include situations where:

- **The output involves affected populations:** Findings about displaced communities, survivors of violence, or marginalised groups require protection-sensitive interpretation that AI cannot provide.

- **The context is fast-moving or politically sensitive:** In active crises, AI may lag behind the realities on the ground. Human verification anchors outputs in what is actually happening on the ground.

- **The finding seems surprising or suspiciously neat:** Both anomalies and overly clean results can signal errors or gaps in source coverage. Trust your instinct to dig deeper.

- **The output will inform a major decision:** Resource allocation, crisis response, and intervention strategies all carry high stakes. These outputs demand expert validation before use.

Questions every humanitarian should ask

Before acting on an AI output: a sample HITL checklist.

- Was this output produced within a human-defined framework?
- Has it been verified against trusted sources?
- Does it reflect local context, including what might be missing?
- Could this output or decision cause harm to the people it is meant to help?
- Is it safe and appropriate to share with decision-makers?
- Who trained this model, and on what data?
- Does this output reflect genuine need or a gap in our data?
- Can an affected person question or challenge this decision?
- Does this AI tool compromise our neutrality or independence?
- Is there a human who is accountable for the outcome?
- What happens when the AI gets it wrong?

Our principles don't pause when we use AI tools

The values that guide humanitarian action - do no harm, impartiality, neutrality, independence - don't stop applying when we use AI.

The human-in-the-loop framework is how we carry those principles into the age of AI.

Resources

The HLA in partnership with Data Friendly Space is leading a global study into how humanitarians are using AI. Access the research products, articles, webinars and podcasts on our Resources Hub.

kaya

Access free online learning at kayaconnect.org. Claim HPass digital badges and share them on social media to let others know about your learning in this area.

This microlearning guide was produced in May 2026 for informational purposes only. It reflects practice at the time of publication. Please consult your organisation's own AI and data protection policies.
Contact: info@humanitarian.academy or hello@datafriendlyspace.org