

How are humanitarian organisations using AI for research, information management, and knowledge processing?



Department of
**International
Development**

AUTHORS:

**ARCHIE EDWARDS, NAOMI CRYAN, ZAHRA
ALI-ASGHAR-LOKHANDWALA**

ACRONYMS

ACCORD - Austrian Centre for Country of Origin and Asylum Research and Documentation (a department of the Austrian Red Cross)

AI - Artificial Intelligence

ARC – Austrian Red Cross

CEB – Council of Europe Development Bank

COI - Country of Origin Information

EU – European Union

EDPB- European Data Protection Board

EPC- European Policy Centre

GDPR – General Data Protection Regulation

RAG- Retrieval-Augmented Generation

OCHA – United Nations Office for the Coordination of Humanitarian Affairs

UN - United Nations

UNDP - United Nations Development Programme

UNOCHA - United Nations Office for the Coordination of Humanitarian Affairs

WFP – World Food Programme

Table of Contents

Acknowledgements	1
Executive Summary	2

SECTION I: INTRODUCTION

Rationale and Project Purpose	6
Research Objectives	7
Methodology	8
Limitations of Study	10

SECTION II: CURRENT AI APPLICATION

Overview of Current AI Applications in the Sector	12
Current AI Applications Specific to Database Management	17
AI Dos and Don'ts in the Humanitarian Sector	19
The Dos: Essential Practices	20
Don'ts Based on Other Organisations' Policies	22
Common Problems Faced by AI Integration	24
Problem Area 1: Organisational Governance & AI Literacy	24
Problem Area 2: Data Bias and Transparency Deficits	25
Problem Area 3: Undocumented and Unregulated AI Use	26

SECTION III: RECOMMENDATIONS

Overview of AI Systems	29
Model Recommendations	33
Recommended AI Policy	35
Recommended Staff Training Guidelines	41
Conclusion	43

BIBLIOGRAPHY & APPENDICES

Bibliography	45
Appendix 1 – Updated Terms of Reference	49
Appendix 2 – Terms of Reference	51
Appendix 3 – Interview Questions	53
Appendix 4 – Methodology	54
Appendix 5 – Interviews	55

Acknowledgements

First, we would like to express our sincere gratitude for our coach Stuart Gordon.

We extend our sincere gratitude to all those who participated in interviews, including representatives from UNDP, CARE International UK (Mohannad Alghabra), the European Policy Centre (Samuel Goodger) and Emma Ruttkamp-Bloem.

Disclaimer

The three authors of this report are MSc International Development and Humanitarian Emergencies students attending the London School of Economics and Political Science (LSE). The ideas expressed in the following report are solely those produced by the authors and do not reflect the views of ACCORD or LSE.

Executive Summary

Background

From summarising Country of Origin Information and translating documents, to identifying patterns in humanitarian datasets, AI tools are becoming increasingly embedded in research, information management, and knowledge processing workflows. Within the humanitarian sector, where decision-making relies on a large volume of data, AI presents both opportunities for change in beneficial ways and ethical challenges.

This report examines how humanitarian organisations are currently using AI for research, information management, and knowledge processing, with the aim of informing the development of a responsible AI framework for ACCORD at the Austrian Red Cross.

The scope of this report is deliberately focused: it examines internal research functions, including document summarisation, metadata tagging, translation support, and database management, and excludes operational AI applications such as logistics or aid distribution. The ultimate aim is not to maximise efficiency, but to ensure that any AI integration strengthens research quality and safeguards the rights and protection of affected populations.

Relevance

AI adoption in the humanitarian sector is accelerating. Approximately 70% of humanitarian workers report using AI tools on a weekly basis, yet fewer than 25% of organisations have formal AI policies in place (HLA, 2025). This gap between practice and governance creates legal, ethical, and operational vulnerabilities. For ACCORD, which handles sensitive COI data used in asylum decision-making, there is a high risk. The integration of AI without a structured, evidence-based framework risks introducing inconsistency across workflows, creating ethical blind spots, and potentially undermining the rights of the populations the organisation exists to serve. This report provides the foundation for closing that gap.

Methodology

The research employs a mixed methodology combining desk-based research with semi-structured interviews. The desk-based review draws on academic literature, AI policy documents and strategy papers from humanitarian organisations. Semi-structured interviews were conducted remotely between December 2025 and March 2026 with representatives from organisations including CARE International UK, UNESCO, European Policy Centre and UNDP, as well as specialists in AI governance and humanitarian action. An initial interview with ACCORD provided internal context. Interview transcripts were analysed thematically using FATHOM. Findings were triangulated to strengthen validity and reliability. Limitations include the small, non-representative interview sample, a concentration of English-speaking organisations, and the tendency of AI policies to omit internal operational detail.

Report Outline

The report proceeds as follows. The current AI applications in the humanitarian sector are first overviewed, documenting how organisations such as CARE International UK and the Austrian Red Cross are using tools like ChatGPT and Microsoft Copilot for summarisation, translation, data verification, and database management. The AI Dos and Don'ts are summarised, emerging from the policies of seven humanitarian organisations, structured around the central principle of Do No Harm. A dedicated section examines AI applications specific to database management, directly relevant to ACCORD's work on ecoinet. The report then analyses the most common problems faced during AI integration, which include organisational governance and AI literacy deficits, data bias and transparency gaps, and the risks of undocumented and unregulated AI use. Building on these findings, the report presents technical recommendations for system design, open-source Retrieval Augmented Generation (RAG) pipeline as the most appropriate architecture for ACCORD's needs. It then sets out a recommended AI policy, followed by training guidelines for staff at all levels. The report concludes with reflections on the future of AI governance in the humanitarian sector, emphasising that responsible integration requires human oversight, data transparency, and ethical accountability as foundational commitments rather than afterthoughts.

The findings demonstrate a cautious human centred AI adoption where human oversight, data protection, transparency, and the principle of Do No Harm are key. Ultimately, the implementation of AI in humanitarian contexts should not be guided by only the improvement of efficiency, but rather to benefit society, specifically affected populations.

AI Policy Recommendations

1 - Transparency in AI usage

2 - Privacy and Confidentiality

3 - Human Oversight and Verification

4 - AI as a tool not a creator

5 - AI literacy training programme at all levels

6 - Cross-department meetings should collectively decide AI usage and a common policy on usage

7 - Establish an AI ambassador in each department

8 - Creation of an approved AI tools list

9 - An incident reporting mechanism

Section I:

Introduction



Introduction

The rise of AI has reshaped how organisations manage and interpret information. Within the humanitarian sector, where decision-making relies on a large volume of data, AI presents both opportunities for positive change and ethical challenges. From summarising Country of Origin Information and translating documents, to identifying patterns in humanitarian datasets, AI tools are becoming increasingly embedded in research, information management, and knowledge processing workflows.

1 - Rationale and Project Purpose

While several humanitarian organisations have adopted AI policies, as of December 2025, there is currently no department-wide methodological framework guiding the use of AI within ACCORD at the Austrian Red Cross. ACCORD compiles and analyses COI to support asylum and subsidiary protection procedures, operating in a particularly sensitive domain.

Given the high stakes of asylum decision-making, the integration of AI into research processes such as summarisation, tagging, and data analysis must be carefully assessed to ensure accuracy, neutrality, and adherence to the other humanitarian principles. Without a structured, evidence-based approach, AI application risks inconsistency across workflows and creating ethical blind spots.

This project examines how humanitarian organisations are currently integrating AI into research, information management, and knowledge-processing systems, with a focus on applications relevant to COI work. It excludes operational uses of AI, such as logistics or aid distribution, and concentrates instead on internal research functions.

This report presents the findings and sets out AI application recommendations for ACCORD.

2 - Research Objectives

The research is guided by the following question:

How are humanitarian organisations and humanitarian experts using AI specifically for research, information management, and knowledge processing, and what best practices, challenges, and governance mechanisms emerge from these experiences?

In response, this study analyses academic literature, organisational policies, and expert insights, triangulating secondary research with semi-structured interviews.

The report begins by outlining the study methodology, followed by an overview of the current use of AI. It then analyses ethical frameworks and practical “Dos and Don’ts” based on a synthesis of other humanitarian organisations’ policies, with an emphasis on applicability to database management and information systems. Building on this analysis, it explores common challenges associated with AI integration, including governance gaps, biases, data protection concerns and organisational capacity constraints.

The final sections present recommendations for implementation, including system design considerations, staff training guidelines and recommended AI policy, before concluding with reflections on the future of AI governance in the humanitarian sector.

Methodology

The project employed an integrated methodology of desk-based research grounded in secondary data and semi-structured interviews. This has enabled the triangulation of data sources and perspectives to create a more thorough assessment of AI practices in humanitarian organisations.

Primary Method - Desk Based Research

Given the availability of public documentation from humanitarian and related sectors on AI governance, strategy, and ethics, desk-based research based was selected as the predominant method. It involved a systemic review of academic literature produced by specialists and humanitarian organisations' AI policy.

Secondary data was sourced from various humanitarian organisation websites, AI policy papers, strategy documents, financial institutions, research institutions, and academic journals. Alongside studying the online AI policies of different organisations, searches were conducted through academic journals and libraries using targeted keywords such as “AI governance”, “AI ethics”, “AI policies”, and “AI in humanitarian action”. Each source was vetted to ensure academic rigour.

An initial literature review was conducted to evaluate the credibility, transparency, and applicability of existing AI practices to ACCORD. This allowed the identification of recurring themes and patterns in organisational approaches to AI and the gaps in current information and knowledge management practices. The findings from this review informed the development of interview questions and the selection of interview participants.

Supplementary Method: Semi-Structured Interviews

Interviews were used as a supplementary method to address limitations of secondary data and information gaps in outward-facing organisation policy. This is because - although AI policies are widely available in the public domain - they often lack in-depth information into organisational practices, internal AI systems, and the challenges organisations face. Thus, interviews allowed the research to fill these gaps and provided an insider understanding of AI policy and practice. Further, an initial interview with ACCORD helped expand understanding of the use of AI within the organisation so that this project could make the most relevant recommendations.

Limitations of Study

1) The interview sample represents a small proportion of humanitarian and related organisations and is not statistically representative.

- The report's findings are based on individuals who were willing to discuss AI policy and as such cannot be seen as characteristic of all humanitarian organisations.
- Many organisations have an internal AI policy that they are not willing to disclose

Solution: Triangulation of the primary data with secondary data and the inclusion of organisations from multiple related sectors.

2) Positionality of the authors and academic sources

- The non-random, targeted sampling approach may introduce self-selection and convenience bias

Solution: A wide variety of humanitarian organisations and specialists were contacted in the hope of widening the scope of the report. The same is true of reports and academic articles.

3) Geographic Organisation Concentration

- The majority of the organisations contacted are Western and English-speaking, this therefore only demonstrates AI application in specific regions of the world

Solution: Information was contrasted with policy from organisations that operate globally to provide a more informed response

Section II:

Current AI Application



Overview of Current AI Applications in the Sector

AI is an increasingly important tool in the humanitarian sector, particularly for information and knowledge management. While AI is often associated with futuristic applications, such as WFP use of satellite information for crop yield predictions (WFP), many humanitarian organisations are already using it for practical, low-risk tasks. These include data processing, document summarisation, and language support. As UNOCHA notes, AI “has the potential to drive operational gains for humanitarians through enhanced efficiency in routine tasks like report drafting and data formatting” (UNOCHA, 2025). This reflects a broader trend: AI is mainly supporting internal decision-making and information workflows rather than frontline operations. There is considerable debate on this topic, but it will be discussed further in later sections.

Humanitarian organisations primarily use Narrow AI, which is coded in a way to perform specific tasks such as text summarisation, language translation, or document analysis. Generative AI, which is a subtype of Narrow AI, is widely used for tasks that involve text. These include drafting reports, summarising data, and checking document consistency. Generative AI, including Large Language Models, is particularly useful for handling high-volume or repetitive tasks that would otherwise consume significant staff time. The specifics of AI models will be explored later on in the report.

Practical Applications

Document summarisation and language tasks are some of the most common applications of AI in humanitarian organisations. ACCORD, for example, uses a fine-tuned version of ChatGPT to produce brief English and German summaries for its ecoi.net database, whilst still deploying human quality assurance applied to check for style, errors or hallucinations.

Staff also use AI to improve grammar, tone, and readability, ensuring that documents are polished for internal or external use. Similarly, CARE International UK staff employ tools such as Microsoft Copilot and ChatGPT for drafting and editing emails, validating data in Excel, and reviewing project reports. In both organisations, AI outputs are treated as support tools, not replacements for human expertise: every output undergoes review to ensure accuracy and reliability.

Database Management

AI is also increasingly being applied to data management and analysis. Large datasets, such as those available through the Humanitarian Data Exchange, can be processed with AI to identify trends, filter relevant information, or flag anomalies. At CARE International UK, AI is used to cross-check project data and highlight inconsistencies, while ACCORD intends to apply AI to sift through high-volume news sources and highlight the most relevant articles. These applications allow humanitarian staff to prioritise higher-value tasks, such as interpreting findings and making decisions, rather than spending time on routine data processing.

Translation

Translation is another growing area of AI applications. In theory, AI tools can translate documents for internal comprehension, enabling teams to access information in multiple languages quickly. However, both ACCORD, CARE International UK and Ms Emma Ruttkamp-Bloem, an expert who has worked with UNESCO and other international organisations, note that AI translations require human review, particularly for languages where the system may produce Anglicised phrasing or unnatural syntax.

For high-stakes or public-facing documents, professional translation remains the standard. This concern was echoed by a non-profit organisation interviewed for this research, which emphasised that AI translation can remove cultural nuance and emotional meaning, particularly in sensitive communications. The organisation noted that direct AI translations risk flattening language and undermining trust, especially when addressing affected communities or supporters. As a result, AI is used cautiously and primarily for internal comprehension, with human translators and multiple layers of review remaining central to external communications.

This highlights a recurring theme in AI adoption: its role is assistive rather than authoritative. To quote Ms Ruttkamp-Bloem, “AI has the capability to streamline work, but only when its use is overseen.”

Cautious Adoption

Literature on the use of AI emphasises that AI adoption in all forms of humanitarian work is cautious and incremental. According to the Humanitarian Leadership Academy (HLA), while around 70% of humanitarian workers report using AI tools weekly in 2025, less than 25% of organisations have formal AI policies, and fewer than two-thirds offer any training (HLA, 2025). This demonstrates how in practice staff often drive adoption individually, identifying tasks where AI can improve efficiency. Although this may lead to increased efficiency, without ethical oversight this can have serious ethical implications. AI application in the Humanitarian sphere mirrors private-sector applications, where AI is used to automate routine processes, enhance analytical capacity, and support decision-making (Harvard Business Online, 2024).

Insights from interview with Ms Ruttkamp-Bloem reinforced this cautious approach. She emphasised that successful AI use in research and information management depends not only on technical safeguards, but on embedding ethical considerations from the earliest stages of adoption, including procurement and system design. Particular emphasis was placed on inclusion, transparency, and awareness of how data limitations, language bias, and underrepresentation can shape AI outputs, especially in contexts involving the Global South. Therefore, in both sectors, human oversight is essential to mitigate risks such as hallucinations, bias, and data errors.

Data Sensitivity

AI adoption in humanitarian organisations also considers data sensitivity and ethics, even if formal policy frameworks are still emerging. A number of interviewed organisations rely mostly on publicly available information for AI processing, avoiding sensitive beneficiary data in external systems.

This reflects broader concerns in the sector regarding privacy, security, and potential bias in AI outputs (UNOCHA, 2025). When errors occur, they are addressed through manual review rather than relying on AI to self-correct. For example, ACCORD identified instances where AI hallucinated details in COI summaries, which were subsequently corrected through human oversight. Similarly, CARE International UK uses AI primarily for verification and anomaly detection, not for generating original data, reducing the likelihood of unintentional misinformation. Ms Ruttkamp-Bloem further stressed that AI systems inevitably reflect the datasets they are trained on, meaning that data poverty, limited language coverage, or cultural bias can lead to distorted or incomplete outputs. Transparency about these limitations, including whose data is represented and whose is missing, was identified as a core principle in responsible AI governance.

Ethical Considerations

The benefits of AI in humanitarian organisations are clear; improved efficiency, enhanced language quality, and better prioritisation of tasks. Yet this adoption remains bounded by practical and technical limitations. Organisations are aware of AI's potential to introduce bias, misinformation, or errors, and current use is concentrated on tasks where human review can mitigate those risks. This aligns with broader findings in the literature which we will go on to explore in this paper: AI tools, while powerful, are not replacements for human decision-making in complex, sensitive contexts (Del Giudice et al., 2021). As Emma Ruttkamp-Bloem (2026) stated “It is not magic, it is technology”.

In summary, current AI use in humanitarian organisations largely takes on a supportive role, not an operational one. Tools like ChatGPT and Microsoft Copilot help staff manage information, summarise content, translate languages, and validate data, freeing time for higher-order analysis and decision-making. Adoption is primarily department-driven, with staff identifying tasks suitable for AI and applying human oversight to ensure accuracy. Literature and interviews consistently emphasise that AI is most effective when applied to low-risk, information-focused tasks, with human supervision mitigating errors, bias, and risks related to confidentiality.

Simultaneously, the scope of AI is gradually expanding beyond routine drafting. For example, pilot programmes have begun to use AI to generate first-cut evidence maps from hundreds of policy and evaluation reports, which researchers then refine and interrogate (Elrha, 2024). Using RAG techniques, where AI answers are based solely on documents provided, reduces the potential for hallucinations (Elhira, 2024). This reflects a shift towards greater organisational confidence, more structured internal guidance, and the systematic integration of AI tools into routine research practice. All of these themes are the topics this research project will go on to further explore, examining how organisations have incrementally adopted AI in order to inform future policy, training and governance frameworks.

Current AI Applications Specific to Database Management

The following section examines how AI is being applied to database management in the humanitarian sector. The findings are based on the trends apparent in secondary readings and interviews, in order to demonstrate the current and future possibilities of AI usage for ACCORD.

Data Ingestion and Structuring

Databases like ecoi.net rely on carefully regulated processes to ensure that documents are ingested and structured correctly. Each document is initially identified, gathered, and labelled with vital details such as country, date, topic, and source type. The system's foundation is built on these organised entries, which enable exact search and retrieval. While AI tools might help with creating descriptions or recommending categories, human reviewers are still necessary to ensure accuracy and compliance with data governance procedures. Staff at ACCORD, for example, summarise documents in both English and German, and each entry's meta data is quality-checked before being included in ecoi.net.

Metadata Assignment and Indexing

In some databases, once documents are processed, they are improved with metadata to enable easier discovery and increase analytical usefulness. Along with typical variables such as title, author, and date, metadata includes contextual tags such as ethnic groups, conflict kinds, and asylum-related topics . Metadata is vital for conducting targeted searches and producing meaningful research. AI can speed up this process by recommending tags generated from content analysis. But human validation must be employed to maintain accuracy and prevents misclassification or semantic drift. This approach ensures that search results remain relevant and trustworthy.

Summarisation and Semantic Linking

Some processes then condense the documents into a brief, searchable description. At ecoi.net, AI helps with this process by generating draft summaries

in several languages. Human reviewers check the outputs for factual accuracy, contextual sensitivity, and clarity. Although ecoi.net does not currently employ the practice, documents can also be semantically connected to similar materials, allowing users to identify patterns across countries, population groups, and conflict kinds. While AI may identify potential linkages, final verification is performed by staff to assure interpretative and factual accuracy.

Verification and Error Checking

Systematic verification and error checking are essential to database reliability. Every AI-assisted tag, summary, and translation is evaluated for inconsistencies or errors. Staff at ACCORD have documented cases in which AI-generated summaries contained plausible but incorrect information, which was recognised and corrected by humans. This stage also includes ensuring linguistic accuracy and adhering to ethical standards, especially when dealing with sensitive or confidential information.

Maintenance, Version Control and Audit Trails

Databases are dynamic systems that require constant upkeep. Tasks include updating old entries, maintaining version histories, and keeping audit trails to track who generated or edited each file. These measures promote transparency and accountability. Staff also monitor and improve AI-assisted procedures to ensure consistency and output quality over time. The database maintains its accuracy, reliability, and operational use through these continuous modifications.

Scope of Roles and Tasks

Database management entails a wide range of tasks, from data ingestion and metadata tagging to verification, summarisation, and oversight. Although AI can help with many of these activities, human skill is still required. According to interviewees, database work involves more than just technical accuracy; it also includes research support, policy relevance, and ethical responsibilities. This highlights how important database management is for operational effectiveness and informed decision-making in humanitarian and policy situations.

AI Dos and Don'ts in the Humanitarian Sector

As organisations manage increasingly large data volumes and deploy emerging technologies in the ways previously highlighted, they face more complex challenges and risks (OCHA, 2025). This part of the research paper synthesises guidance from seven major humanitarian organisations: the United Nations Chief Executives Board for Coordination (UN CEB), United Nations High Commission of Refugees (UNHCR), World Food programme (WFP), International Committee of the Red Cross (ICRC), United Nations Children's Fund (UNICEF), United Nations Office for the Coordination of Humanitarian Affairs (OCHA) and Save the Children, in order to highlight current trends in how organisations perceive AI should be used in data management and research. From this a list of common Dos and Don'ts has been created to demonstrate current ethical frameworks used by organisations in AI application.

The central Principle: DO NO HARM

Analysis of codes of conduct from these seven organisations demonstrates that all AI guidance is underpinned by one principle: "Do No Harm". The UN CEB's Principles for the Ethical Use of Artificial Intelligence explicitly grounds its framework in ethics and human rights to ensure AI serves humanity responsibly (UN CEB, 2022). UNHCR's AI approach similarly emphasises that systems must be "rights-based, inclusive, transparent and accountable" to protect displaced populations (UNHCR, 2025). This principle requires humanitarian actors to consider the future impacts of AI when it is implemented, including understanding how it will affect AI systems' lifecycles (ICRC, 2025).

The Dos - Essential Practice

DO Maintain Human Oversight and Accountability

The primary way that organisations ensure that “Do No Harm” is upheld is through human oversight. The ICRC, for example, emphasises that AI support systems “must remain tools that help and support human decision-making, rather than tools that replace human judgment” (ICRC, 2025). In the same vein, throughout UNHCR’s implementation of ARiN, a software solution that uses machine learning techniques to screen job applications, the programme was customised to comply with transparency and diversity policies. Thus, human actors retained ultimate authority over decision-making and screening (UNHCR, 2024).

Human oversight means AI tools should augment rather than substitute professional judgment. Kanoulas notes that whilst AI can process vast datasets quickly to help researchers uncover patterns in COI research, human expertise remains essential for contextualising findings (Kanoulas, 2025). Hence, in current application of AI, human oversight stays at the centre of all processes.

DO Ensure Transparency and Data Privacy

Transparency in AI application is perceived as essential to evaluate the appropriateness of AI use at each stage of a process. For example, UNICEF's guidance stipulates that AI regulations must include mechanisms for explaining how AI systems make decisions (UNICEF, 2025).

Organisations also rigorously protect data privacy through closed systems. Following a 2022 cyberattack stealing beneficiary data, the ICRC now prioritises secure AI deployments (Stoll, 2024). Save the Children employs “stringent data protection protocols that include advanced encryption and secure storage mechanisms,” collecting personally identifiable information only when essential (Save the Children, 2024). For desk-based researchers, this means public AI

systems like ChatGPT are avoided when using sensitive humanitarian data, instead closed systems are prioritised. Alternatively, datasets are anonymised.

DO Prevent Bias and Test Before Implementation

Before implementation, all AI models should be thoroughly tested. WFP's creation of an AI Sandbox exemplifies this policy, as it serves as "a testing ground for AI projects". It assesses feasibility and impact of AI systems before deployment in a closed environment (United Nations, n.d.). This is because the UN CEB framework requires organisations to conduct "oversight, impact assessments and due diligence" (WFP Innovation, 2025). This testing is essential as there is genuine concern that AI systems may "mirror racial, ethnic, gender and other inequalities present in society" (ICRC, 2021).

For desk-based research, assessments must determine whether AI tools prioritise particular information sources and adequately value different populations' opinions and statements. As bias manifests in training data, this needs to be addressed in systems implementation and training. As Stoll has emphasised, AI can be used in "parts of the world that are not necessarily accounted for in existing models, as these models are largely trained on data from western settings" (Stoll, 2024). Hence, organisations employ mitigation strategies to reduce this bias, such as using representative training data.

Don'ts Based On Other Organisations' Policies

DON'T Allow AI to Exacerbate Inequalities

As established, AI can exacerbate inequalities in desk-based research and data management. Pizzi, Romanoff and Engelhardt warn that when AI systems are used to generate analytical insight, said insights are “liable to be used as replacement for human decision making” (2021). This can be prevented by human oversight, but it emphasises how if left unchecked, AI can reproduce inequality.

If data is improperly protected it can also exacerbate vulnerabilities, for example the UNHCR notes refugees “can be identified and tracked online, jeopardising their protection” if systems lack privacy safeguards (UNHCR, 2023).

DON'T Create Dependency or Treat AI as Value Neutral

Whilst AI can save approximately 5 to 25 percent of humanitarian researchers' time (OECD, 2025), organisations must prevent hyper-dependency as human capability can be reduced as a result of hyper reliance on technologies. AI should not be seen as a cost saving measure by replacing training. This is emphasised in the ICRC's questioning of whether the sectors' metrics are shifting (or have shifted) from quality to quantity. They are concerned with trends signalling growing use of business language focused on “productivity” and “Return on Investment” rather than impact (ICRC, 2025). Over-reliance on AI leads to de-skilling and this will impact the judgment of professional researchers, reducing their ability to critically evaluate AI outputs.

Additionally, a critical error is assuming AI solutions are neutral. The ICRC emphasises that “humans develop AI and humans decide when and how to use this technology” (ICRC, 2025). As a consequence, algorithms can be seen to make choices about data prioritisation and information categorisation reflecting the values of the designers. If the designer is from outside the humanitarian

system, or a prebuilt model is used, this may not align with humanitarian principles. UNICEF warns AI systems can "uphold or undermine" children's rights depending on design and implementation (UNICEF, 2025). Researchers must interrogate AI systems in order to embed values and reduce problematic assumptions.

DON'T Deploy Without Cultural Context or Consideration of Governance

As Humanitarian contexts vary enormously an AI system appropriate in one setting may create serious risks in another. Thus, context-specific risks include security threats in conflict settings where refugees are fleeing from, connectivity limitations rendering cloud-based systems unusable, and cultural factors affecting community perceptions of AI implementation (OCHA, 2021). UNHCR highlights unique considerations around refugee data protection and cross-border information sharing risks (UNHCR, 2023). Thus, if implemented to help with research, the reaction of the group whose data the AI is categorising, must be considered.

Further, legal rules and regulations must be considered. This means that policies specifying approval processes, monitoring mechanisms, and accountability structures must be taken into account (United Nations Chief Executives Board for Coordination, 2022).

Findings:

A synthesis of different humanitarian organisations frameworks illustrates that "Dos" should focus on human oversight, which ensures transparency, data privacy protection, rigorous testing and bias mitigation. The "Don'ts" emphasise the risk of exacerbating inequalities, creating dependency, treating AI as value-neutral, deploying without governance, and ignoring context. As WFP notes, organisations are "harnessing the power of AI" whilst "upholding the highest ethical standards" (WFP Innovation, 2025). Therefore, the challenge is to use AI responsibly through what the ICRC calls a "human-centred approach" ensuring AI augments rather than undermines humanitarian effectiveness (ICRC, 2021).

Common Problems Faced by AI Integration

Problem 1 - Organisational Governance and AI Literacy

Interviews conducted with experts in AI policy and international organisations, including Emma Ruttkamp-Bloem, Samuel Goodger (EPC), Mohannad Alghabra (Care International UK), and a representative from UNDP, identified the importance of AI literacy and training to ensure responsible usage of AI.

AI systems are frequently introduced without adequate understanding at all levels of the organisation and without meaningful communication between the departments that commission them (Emma Ruttkamp-Bloem, 2026).

Boards and senior leadership hold decision-making authority over AI adoption, yet frequently they approve the implementation of AI systems without proper understanding of what these technologies are, how they function, or what risks they carry. This creates a dangerous asymmetry: those with the power to commit organisational resources to AI deployment and assume legal and ethical responsibility for it are not equipped to exercise meaningful oversight. A trend emerges where, from this knowledge gap, unrealistic expectations emerge. As Emma Ruttkamp-Bloem (2026) puts it: "Everybody expects magic to happen, but no one lifts a finger to make it".

A second governance failure is the siloed nature of AI implementation. Emma Ruttkamp-Bloem (2026) argues that projects are frequently initiated by one department without adequate consultation with the teams expected to integrate them into daily operations. This can result in resistance to new policy, workarounds, and missed opportunities to highlight practical concerns before deployment.

Therefore, boards should not delegate AI ethics oversight to technical teams. Based on this analysis questions of fairness, accountability, bias mitigation, and harm prevention require engagement at the highest level of organisational authority and boards that remain illiterate about AI are effectively abdicating that responsibility, not through intention, but through inaction.

Recommendations:

- Mandate a minimum AI literacy programme for all board members before AI initiatives are approved.
- Establish a cross-departmental AI Coordination Group that must be consulted prior to any procurement or deployment decision.
- Reframe AI as a tool requiring human effort and institutional adjustment, tied explicitly to the organisation's social mission rather than efficiency metrics alone.

Once implemented, there should be a mandatory AI literacy training programme for every individual in the company.

Problem 2 - Data Bias and Transparency Deficits

AI systems learn from data, therefore the quality of the data, how representative it is and its provenance determines the quality of AI outputs. In the humanitarian sector, data limitations can determine whether AI systems help or harm the people they are meant to serve.

There is no such thing as unbiased data. As the majority of data is collected by humans, the choices made in the process of collection embed human assumptions and biases. For example, a health dataset collected in the United States is representative of the demographics of the United States, not other countries. Applying a model trained on extracting data from a US dataset to datasets based on information on individuals in Tanzania or Austria, can risk causing harm through misrepresentation. This is not to label such systems useless, but to highlight that users must acknowledge the limitations of their AI based on the data it was trained on. Such process allows for adoption or rejection of specific models where appropriate.

Further, transparency is a legal and ethical obligation. Therefore, individuals affected by AI-driven decisions have a right to understand how those decisions were reached and the right to know whose data underpins those predictions. It

also requires clear disclosure of the populations, regions, and age groups represented in training data, and the terms under which third-party data was obtained, including how it can be deleted or accessed. This is otherwise known as “auditability”.

Recommendations:

- Require a standardised Data Transparency Statement for any AI system before deployment, covering the source, geography, demographic profile, and known limitations of training data.
- Introduce a population-specific applicability standard for systems involving vulnerable groups, requiring either adequate data representation or explicit caveats and enhanced human supervision.
- Establish clear data rights and deletion protocols and publish them in plain language.
- Where AI systems are commissioned or co-developed, require data scientists to formally declare dataset limitations, and allocate appropriate budget for data quality work rather than defaulting to whatever data is most readily available.

Problem 3 - Undocumented and Unregulated AI Use

A growing concern is the informal, untracked use of AI tools by staff to carry out tasks such as drafting emails, summarising or publishing documents and the lack of transparency of this use. This also includes the use of AI with non-approved tools and publicly available AI systems.

This can pose a significant legal and compliance risk. Under the European Data Protection Regulation (GDPR), organisations are legally obligated to maintain audibility over how personal data is stored, used and processed. This is becoming increasingly difficult, with CARE International UK noting that 90% of its staff are reportedly using AI tools, so it is important to keep each use documented (CARE International UK, 2026).

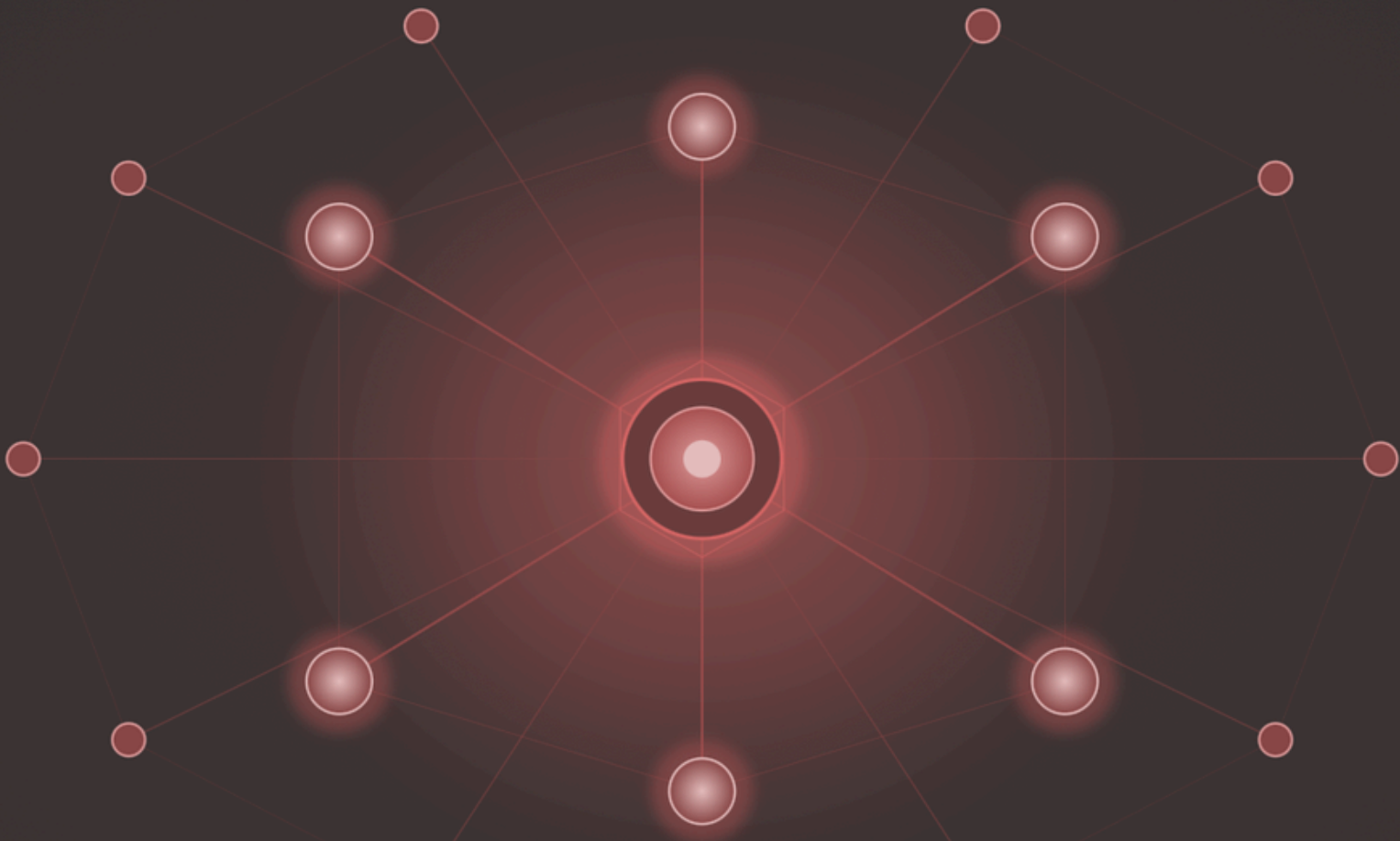
For example, if an employee uses an unapproved AI platform to summarise a document which contains personal data, the organisation could be breaching the GDPR's article 5. This article asserts that personal data should be processed lawfully, transparently and with proper security measures in place. In this case, an organisation can be exposed to potential fines of up to €20 million or 4% of global annual turnover.

Recommendations:

- Similar to academic essays, after publishing any document or completion of a task, a disclosure of AI use should be added, with staff filling in a form declaring their use of AI upon submission.
- This declaration should include details of the tools used and for what purpose.
- Staff should also actively keep a record of the prompts used throughout a project to showcase transparency of how AI contributed to the final project.

Section III:

Recommendations



Overview Of AI Systems

So far, this proposal has primarily focused on the ways in which AI is utilised by humanitarian organisations, best practice and problems with implementation. Just as crucial as this is the type of AI model that is used, with different builds being both more efficient and more ethical for different tasks. Three key binary decisions should shape ACCORD's decision on which model to implement: generative versus discriminative models, open source versus closed source deployment and pre-built versus custom models. On top of these decisions the implication of RAG and embedding models on privacy, accuracy and governance compliance must be considered.

Generative vs Discriminative

Discriminative models (such as BERT, RoBERTa, and support vector machines) learn to classify inputs by identifying boundaries between categories rather than fully modelling data distribution. This makes them more efficient than generative models and, as such, less energy intensive (Belo et al., 2025). They are also highly accurate for structured tagging tasks with well-defined labelled sets. Although it requires more oversight and data set preparation before inputting it into the AI, this process supports audibility by creating a clear processing chain. This is a key requirement under GDPR's obligation to be able to explain outcomes in automated decisions (EDPB, 2024). However, this rigidity is also the model's key weakness, as it is poorly suited to new document types outside its training data. This was emphasised through an interview with an employee at UNRWA. The said employee described how the lack of standardisation in Palestinian documentation meant that it was harder for AI models to extract key information from certain records than it was from others (UNRWA Interview, 2026).

On the other hand, Generative models, such as LLaMA, model underlying data distributions and patterns, which allows them to handle previously unseen content without prior labelling. This increased flexibility is at the cost of increased energy consumption, increased cost and increased chance of hallucinations.

As L. Huang et al. have demonstrated in their study of Large Language Models, which is a form of generative AI, such models are prone to hallucinations due to misrepresentative “data”, mistakes in “training” (or training on bias sources) and “inference” in decoding strategies (Huang et al., 2025). These hallucinations, if not caught, can lead to the spread of misinformation.

Thus, for most organisations, a hybrid approach, that employs discriminative models for high volume, routine classifications and generative models for complex or unstructured content, represents the best option (Tan et al, 2024).

Open Source versus Closed Source

For the sensitive COI information that ACCORD handles the decision between an open versus closed model has the clearest answer. Closed source AI systems, such as Gemini, require data to be transmitted to external servers possibly breaching privacy guidelines. Belo et al. highlight that such models introduce “eternal dependencies” and “privacy related concerns” when documents containing sensitive data are uploaded, instead recommending the use of an open-source system (Belo et al., 2025). The EDPB reiterates this sentiment, as personal data processed during model interaction, remain embedded in the external model and remain extractable.

Thus, open-source models, such as LLAMA and Mistral, which can be fully self-hosted, are a better option to ensure data never leaves organisational infrastructure. However, Nagle and Yue found that open models achieve an average of 90% of closed model performance, per token inference costs are on average 87% lower (running costs) (Nagle and Yue, 2025). Recent analysis has demonstrated that this gap is narrowing further with well-trained open-source models being able to surpass zero-shot GPT-4 on tasks that includes topic tagging and relevance labelling. Therefore, open-source models are more preferential with the capacity for growth.

It must be noted that this average reduction in cost is based on the price per token of each model and does not take into account running costs or the cost of the infrastructure that each model needs. Further, this analysis is assuming that sensitive data is being handled, and for more general tasks the decision between Open and Closed Source models would be more nuanced.

Pre-Built versus Custom Models

The main difference between pre-built and custom models is that the former are able to be implemented more rapidly at a lower technical overhead and price. This makes them attractive for organisations seeking quick productivity gains. However, compared to fine-tuned custom models trained on an organisation's own labelling systems and to match their wants, they are far less efficient (Carlson et al., 2025). Although tangential to ACCORD's interest in AI, Shahi and Hummel found that LLaMA 3.1 when fine-tuned, outperformed standard BERT performance in scientific data categorisation by a meaningful margin (Shahi and Hummel, 2025). ACCORD must thus decide between longevity and short-term efficiency.

Embedding models

Embedding models convert documents into numerical vectors that encode semantic meaning, enabling similarity-based search and clustering that key word approaches cannot achieve (Reimers and Gurevych, 2019). This appears to be exactly the kind of model that would be ideal for ACCORD in COI research. When deployed in a document management context this would allow a search for “at risk groups” to also retrieve semantically related documents, even if the exact phrasing is not present. For example, the search would also retrieve documents including terms such as “vulnerable groups”. Fine-tuned embedded models improve performance over general models across specific domains and would be an ideal fit for database searches for ACCORD.

Retrieval Augmented Generation

RAG combines a retrieval layer, which is powered by embedding models, with a generative model. This grounds the LLM's response in organisational documents rather than relying on a parametric memory. If ACCORD wanted to implement a generative model, this would be one of the best suited, as its reduced parameters and architecture reduces hallucination risk while keeping sensitive data within a closed, governable database (Klesel and Wittman, 2025). This allows for the

knowledge base of the system to be updated, restructured or deleted without modifying the model, making it more ideal in some sense than a fine-tuned custom model which can lack flexibility. RAG also has a low initial cost of ownership relative to fine-tuning, as it does not require the computational intensity of retraining a whole model (Klesel and Wittman, 2025). RAG thus offers a compelling balance of flexibility and auditability.

Model Recommendations

Based on the research, interviews with specialists and the specific structure and needs of ACCORD, this project recommends the following model:

An on premise, open model, RAG pipeline

On Premises, Open Model

- For a privacy sensitive organisation such as ACCORD, the best option is on premises open model.
- This means that all data will be safely contained in databases, and there is no risk of GDPR leakage.

RAG Pipeline

No single model is able to outperform the others in all categories of generation, semantic retrieval, structured classification and entity extraction. Hence multiple models are needed organised in a pipeline that would utilise the following components:

1. A fine-tuned open source generative model (e.g. LLaMA or Mistral) serving as the backbone. It would handle complex document queries without transmitting data externally.
2. An open weight embedding model converts documents into semantic vectors stored in a self-hosted (on site) database. This allows for semantic retrieval across the documents with ease.

3. RAG then grounds the generative model's response in this internal store, maintaining audibility and reducing hallucinations.

Additionally, for high volume, routine documents such as forms, a dedicated fine-tuned discriminative model can be used for labelling.

If this framework is followed, each stage of the model can be audited, updated or replaced independently without replacing the whole system. This is in line with EDPB's 2024 position that privacy in AI must be treated as a design requirement. It is also aligned with the EU AI Act around transparency.

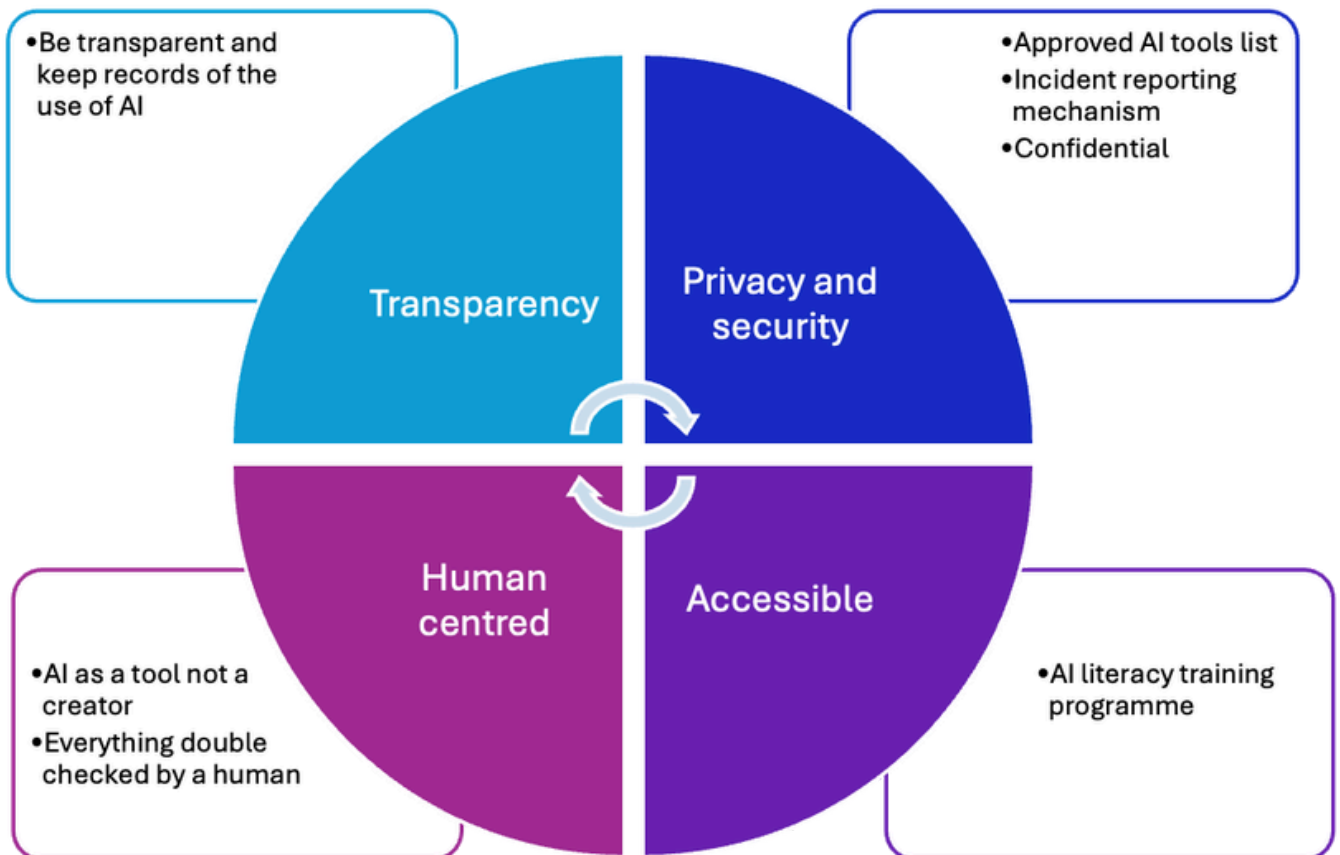
For Minimum Resources:

When handling sensitive material, a fine-tuned, open source, generative model, deployed on premises would be required to comply with EU regulations. It should have embedding layers added incrementally as the capability of available technology increases and available funds grow. However, for publicly available or non-sensitive data sets, closed-source or external models can be used.

Recommended AI Policy

Based on the report's findings, it proposes that ACCORD should consider the following policies when implementing AI for information and knowledge management.

All of the recommended policies will be underlined by the following 4 key principles:



1 - TRANSPARENCY IN AI USAGE

- All uses of AI should be documented within the workflow.
- Staff should record the tool used, the purpose of its use and the stage at which it was applied.

Maintaining a record of AI usage ensures that automated tools remain visible and traceable within research and administrative processes (European Parliament and Council of the European Union, 2024, Article 26). In humanitarian organisations where outputs may inform policy or legal decisions, it is important to be able to identify where AI contributed to the preparation of documents. This can also help build trust and credibility (European Data Protection Board, 2024). Documenting AI uses allows the organisation to monitor how these AI systems are being applied across teams and to flag any misuses or over-reliance.

2 - PRIVACY AND CONFIDENTIALITY

- Confidential, identifiable, or otherwise sensitive information must not be entered into publicly available AI systems.
- AI tools should only be used with publicly available or non-sensitive material unless the tool is self-hosted and controlled.

Sensitive and confidential information must not be entered into publicly available AI systems (European Data Protection Board, 2024). For example, AI systems intended to be used for biometric categorisation are protected under Article 9(1) of Regulation (EU) 2016/679 on the basis of biometric data (European Parliament and Council of the European Union, 2024). AI systems that are self-hosted, securely managed, and fully controlled by the organisation, with appropriate safeguards in place can however manage sensitive data. Entering internal documents, case information, or personal data into such platforms presents risks under data protection law, including the GDPR (GDPR, 2024).

3 - HUMAN OVERSIGHT AND VERIFICATION

- All AI-generated outputs should be reviewed and verified by human staff before use.

AI systems can produce text that appears credible but contains inaccuracies, omissions or fabricated information. For this reason, AI generated summaries, translations, or drafted text should be treated as preliminary material and checked against the original sources before being incorporated into reports or databases. Human oversight ensures that responsibility for accuracy remains with researchers rather than automated systems. Samuel Goodger (EPC, 2026) also notes that prompting AI tools to respond critically can reduce the likelihood of bias and hallucinations, complementing human review by encouraging more cautious and qualified outputs from the outset.

4 - AI AS A TOOL NOT A CREATOR

- AI should not be used to generate new content but rather complement existing work, such as retrieving documents, advising edits and document tagging.
- This mitigates the risks of hallucinations.

In interviews, Samuel Goodger (EPC, 2026) and Mohannad from CARE International UK (2025), emphasised that AI is to be used as a complementary tool to support human work, not as an autonomous creator of original content. In practice, this means AI use should be limited to tasks such as editing draft documents, written communications, assisting with the formatting or sending of emails and data base management under prompts given by a human.

5 - AI LITERACY TRAINING PROGRAMME AT ALL LEVELS

- Introduced to all employees to ensure Dos and Don'ts of the use of AI.
- Examples of potential hallucinations should be included.
- This needs to be regularly reviewed on a bi-annual basis, reflecting new risks given the fast pace of AI evolution.

AI literacy programme introduced for employees at all levels to build a shared understanding of how AI works, its benefits, and its limitations in the workplace (European Data Protection Board, 2026). Pilots of schemes like this have already been employed by organisations such as UNDP, which in 2024 invested in small group training sessions for staff. Similarly, Samuel Goodger of EPC (2026) cautioned that training should not become a mere box-ticking exercise. Examples of AI hallucinations, where AI systems generate incorrect, misleading, or fabricated information should be included. This is to avoid situations like that which CARE International UK highlighted, where a skill disparity in staff in AI application created uneven quality of work and increased risk in certain departments (CARE International UK, 2026).

6 - CROSS-DEPARTMENT MEETINGS SHOULD COLLECTIVELY DECIDE AI USAGE

- There should be common and consistent use of approved AI tools across all departments, with clear accountability for any AI-generated outputs.
- Equitable access to approved AI tools for all staff, regardless of role or seniority.

Regular cross-department meetings should be established to discuss AI usage and align practices across the organisation. These forums should help develop a consistent organisational approach to AI adoption and governance. Clear accountability should be defined for reviewing and validating any outputs generated by AI used in work processes.

7 - ESTABLISH AN AI AMBASSADOR IN EACH DEPARTMENT

- Within each department, there should be an AI ambassador elected.
- This individual will be the spokesperson at a bi-monthly cross department meetings.

Through the implementation of an ambassador, each department will be able to make clear how it is interacting with AI and any issues it is facing. This will allow for easy and accurate changes to AI policy and the development of AI structures as a tool that aids in the growth of the whole of the organisation (Emma Ruttkamp-Bloem, 2026).

8 - CREATION OF AN APPROVED AI TOOLS LIST

- A list of permitted tools should be created that is reviewed regularly.
- It should be ensured that any third-party AI tools used meet the organisation's data protection and security standards.

The organisation should maintain a centralised list of AI tools that are approved for use by staff. This list should be regularly reviewed and updated to reflect technological developments and organisational needs. Any third-party AI tools must meet the organisation's data protection, cybersecurity, and compliance standards before approval. Due diligence should be conducted to ensure external providers handle data securely and in accordance with relevant regulations.

9 - AN INCIDENT REPORTING MECHANISM

- A process for staff to flag AI errors or misuse without fear of blame.

Employees should be able to report AI-related issues such as errors, bias, misuse, or security concerns. Creating a culture of transparency supports responsible AI use and helps the organisation identify and address risks early.

Toon (2024) and Emma Ruttkamp-Bloem (2026) argue that datasets are inherently biased and often unrepresentative, as they are made by humans, which can introduce errors to AI outputs. Thus, reporting AI errors and creating a feedback loop would prevent hallucinations and ensure quality of AI outputs (Toon, 2024).

Recommended Staff Training Guidelines

The training guidelines were formed based on the information gathered from interviews and research based on secondary documents.

Department Based Training

AI training should be delivered at the departmental level, ensuring that instruction reflects the specific tasks carried out by staff. Different teams engage with AI in different ways, and training should therefore focus on the practical applications relevant to each role. For example, staff working on COI research and database management require guidance on verifying AI generated summaries, reviewing translations, and checking metadata suggestions. Other teams may use AI primarily for drafting, editing, or data formatting tasks and should be trained appropriately.

Interviewees emphasised that training is most effective when integrated into existing workflows. In some organisations, internal sessions involve testing AI tools on documents that staff already work with. This allows employees to observe both the capabilities and limitations of AI tools within the context of their own research processes.

Recognising hallucinations and AI Errors

Training should include practical examples illustrating hallucinations and other AI errors. LLMs generate text based on statistical probability rather than verified knowledge, which means that outputs may appear credible while containing inaccurate or fabricated information (Toon, 2024).

Interviewees noted that staff should be shown examples where AI summaries misrepresent source material or omit relevant details. Comparing AI-generated summaries with the original documents allows staff to identify inaccuracies and better understand the limitations of such systems. This is particularly important in research environments where errors could affect subsequent analysis.

Approved tools and internal practices

Interviews also highlighted the need for clear guidance on which AI tools are permitted for use. One international advocacy organisation interviewed for this research maintains an internal list of approved AI tools within the team, and staff are encouraged to use these systems rather than freely available public models

The same organisation also applies a two-person verification process when AI is used to assist with drafting or summarisation. AI assisted outputs are reviewed and fact-checked by two staff members before being finalised. This practice reinforces the principle that AI outputs should be treated as draft material rather than authoritative sources.

Data Protection and GDPR

Training should also address data protection obligations, particularly under the GDPR. Staff should be instructed not to upload identifiable information, confidential documents, or internal case material into external AI systems (GDPR, 2024).

Several interviewees indicated that AI tools are therefore primarily used with publicly available information, reducing the risk of exposing sensitive data.

Use Responsibly

Across interviews, a consistent theme was that AI should function as a support tool rather than a decision-making authority. Training should emphasise that AI outputs must be reviewed, verified, and interpreted by staff before being incorporated into research outputs or organisational documentation.

Conclusion

The integration of AI into humanitarian research, information management, and knowledge processing represents a significant opportunity to advance conventional methods. However, it is becoming increasingly apparent that its application must be done through ethical means.

Several consistent themes have emerged. First, AI adoption in the sector is cautious, incremental, and primarily assistive. Organisations including CARE International UK and UNHCR have implemented AI tools to support document summarisation, metadata tagging, translation assistance, and data verification. In each case, human oversight remains central. AI outputs are treated as drafts, not conclusions, and every automated process is subject to staff review before findings are incorporated into research or decision-making. This reflects a sector-wide consensus that AI should augment professional judgment, not substitute for it.

Second, governance frameworks are lagging behind practice. While around 70% of humanitarian workers report using AI tools on a weekly basis, fewer than 25% of organisations have formal AI policies in place (HLA, 2025). This gap creates legal, ethical, and operational vulnerabilities. Without clear guidance on approved tools, documentation requirements, and accountability structures, informal AI use increases outside institutional oversight. Under the GDPR, this exposes organisations to compliance risks that are both legally significant and reputationally damaging. Closing this gap requires not only policy development but active decisions to increase AI literacy at all levels, including senior leadership and governance boards.

Third, data bias and representational inequality within AI models have real consequences for affected populations. AI systems trained predominantly on data from Western, English-language contexts risk producing outputs that misrepresent or omit the experiences of the populations the systems are meant to

help. As Emma Ruttkamp-Bloem emphasised throughout her interviews for this research, transparency about whose data underpins a system, and whose is absent, is a foundational ethical requirement of AI systems.

For ACCORD specifically, the recommendations set out in this report offer a pathway to responsible, sustainable AI integration. An on-premises, open-source RAG pipeline represents the most appropriate technical architecture, combining privacy compliance with the flexibility required to handle varied document types. Fine-tuned discriminative models offer efficiency for high-volume, structured tasks, while embedded semantic search can meaningfully improve the quality and reach of COI database queries. Critically, these technical choices must be accompanied by robust governance and clear policies including: an approved tools list, mandatory AI usage disclosures, cross-departmental coordination, and a structured AI literacy programme reviewed on a bi-annual basis to reflect the pace of technological development.

The purpose of AI integration at ACCORD should not be efficiency alone. There is a fear in the humanitarian system that a shift in organisational metrics from quality to productivity risks shifting the sector towards a more business mindset. Whilst AI main benefit is often viewed as its time saving capabilities, its true value is realised when implemented in line with ethical guidelines, serving as a tool to achieve goals, rather than becoming the method itself.

The question guiding this research asked how humanitarian organisations are using AI for research, information management, and knowledge processing, and what best practices, challenges, and governance mechanisms emerge from those experiences. The answer is that the sector is learning in real time, building frameworks through trial, error, and reflection. The organisations doing this most responsibly are those that have placed human oversight, data transparency, and ethical accountability at the centre of their AI strategies from their inception.

Bibliography

Belo, R.C. et al. (2025) 'Benchmarking large language models for automated labeling', *Information and Software Technology*, 183.

Bryman, A. (2016) *Social Research Methods*. 5th edn. Oxford: Oxford University Press.

Calvino, F., Reijerink, J. and Samek, L. (2025) 'The effects of generative AI on productivity, innovation and entrepreneurship', *OECD Artificial Intelligence Papers*, No. 39. Paris: OECD Publishing.

Carlson, N. et al. (2025) 'The use of LLMs to annotate data in management research', *Strategic Management Journal*.

Elrha (2024). *Humanitarian Aid?: considerations for the future of AI-use in humanitarian action*. [online] Available at: <https://www.elrha.org/news-blogs/ai-use-in-humanitarian-action>

European Data Protection Board (EDPB) (2024) *Opinion 28/2024 on data protection aspects related to AI models*.

European Data Protection Board (EDPB) (2026). *EDPB and EDPS support streamlining AI Act implementation but call for stronger safeguards to protect fundamental rights* | European Data Protection Board. [online] Available at: https://www.edpb.europa.eu/news/news/2026/edpb-and-edps-support-streamlining-ai-act-implementation-call-stronger-safeguards_en

European Parliament and Council of the European Union (2024) *Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. Official Journal of the European Union. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689>

Del Giudice, M., Scuotto, V., Orlando, B. and Mustilli, M. (2021). Toward the human – Centered approach. A revised model of individual acceptance of AI. *Human Resource Management Review*, 33(1), p.100856.
doi:<https://doi.org/10.1016/j.hrmr.2021.100856>

General Data Protection Regulation (2024) GDPR fines and penalties. Available at: <https://gdpr-info.eu/issues/fines-penalties>.

Gilardi, F., Alizadeh, M. and Kubli, M. (2023) 'ChatGPT outperforms crowd-workers for text annotation tasks', *Proceedings of the National Academy of Sciences*, 120, e2305016120.

Humanitarian Leadership Academy (2025) 70% of humanitarians use AI regularly, but organisational uptake still emerging: global study of 2,500 finds. Available at: <https://www.humanitarianleadershipacademy.org/news/70-percent-humanitarians-use-ai-regularly-but-organisational-uptake-still-emerging-global-study-of-2500-finds/> (Accessed: 25 March 2026)

Huang, L. et al. (2025) 'A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions', *ACM Transactions on Information Systems*, 43(2), Article 42, pp. 1–55

International Committee of the Red Cross (2021) *Artificial Intelligence and Machine Learning in Armed Conflict: A Human-Centred Approach*. Geneva: ICRC.

International Committee of the Red Cross (2025) *Artificial Intelligence in the Military Domain*. Geneva: ICRC.

Kanoulas, E. (2025) *Unlocking Artificial Intelligence's Potential in COI Research*. Research Paper No. 44. Geneva: UNHCR.

Keraghel, I. et al. (2025) 'A review of named entity recognition: from learning methods to modelling paradigms and tasks', *Artificial Intelligence Review*.

Klesel, M. and Wittmann, H.F. (2025) 'Retrieval-Augmented Generation (RAG)', *Business and Information Systems Engineering*, 67(4), pp. 551–561. Available at:

Montebello, M. et al. (2025) 'RAG and LLMs for enterprise knowledge management: A systematic literature review', *Applied Sciences*, 16(1), p. 368.

Nagle, F. and Yue, D. (2025) The latent role of open models in the AI economy. MIT IDE.

Office for the Coordination of Humanitarian Affairs (2021) OCHA Data Responsibility Guidelines. The Hague: OCHA Centre for Humanitarian Data.

Office for the Coordination of Humanitarian Affairs (2025) Data Responsibility Guidelines: January 2025. The Hague: OCHA Centre for Humanitarian Data.

Pizzi, M., Romanoff, M. and Engelhardt, T. (2021) 'AI for humanitarian action: Human rights and ethics', *International Review of the Red Cross*, 103(913), pp. 1–31.

Reimers, N. and Gurevych, I. (2019) 'Sentence-BERT: Sentence embeddings using siamese BERT-networks', in *Proceedings of EMNLP 2019*.

Save the Children (2024) *Harnessing Artificial Intelligence for Child Protection: An Ethical Roadmap*. London: Save the Children International.

Shahi, G.K. and Hummel, O. (2025) 'On the effectiveness of large language models in automating categorization of scientific texts', in *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2025)*

Stoll, P. (2024) 'Red Cross adopts AI guidelines to contain risks', *Geneva Solutions*, 13 December.

Tan, Z. et al. (2024) 'Large language models for data annotation and synthesis: A survey', in *Proceedings of EMNLP 2024*.

The RegulatingAI Podcast (2025). UNESCO's AI Ethics Framework Ft. Prof. Emma Ruttkamp-Bloem | The RegulatingAI Podcast. [online] YouTube. Available at: https://www.youtube.com/watch?v=mMK_YeRrYrl

Toon, N. (2024). *How AI Thinks*. Random House

United Nations Chief Executives Board for Coordination (2022) *Principles for the Ethical Use of Artificial Intelligence in the United Nations System*. New York: UN CEB.

United Nations Chief Executives Board for Coordination (2024) Report on the Operational Use of AI in the UN System. New York: UN CEB.

United Nations Children's Fund (2021) Policy Guidance on AI for Children Version 2.0. New York: UNICEF.

United Nations Children's Fund (2025) Guidance on AI and Children Version 3.0. Florence: UNICEF Innocenti.

UNDP (2020) Human Development Report | UNDP HDR. (n.d.). Human Development Report 2020 | UNDP HDR. [online] Available at: <https://report.hdr.undp.org/>

UNESCO (2021). Ethics of Artificial Intelligence. [online] UNESCO. Available at: <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>

United Nations High Commissioner for Refugees (2023) UNHCR Submission for Global Digital Compact. Geneva: UNHCR.

United Nations High Commissioner for Refugees (2024) Data Innovation Programme. Geneva: UNHCR Innovation Service.

United Nations High Commissioner for Refugees (2025) UNHCR AI Approach. Geneva: UNHCR.

WFP Innovation (2025) Artificial Intelligence. Rome: World Food Programme Innovation Accelerator.

Appendix 1 - Updated Terms of Reference

NB: We have included our first and amended TORs. We had to amend this document as our first research question was too broad and the people we interviewed were also specialists in AI not just humanitarian actors.

Project Title: Artificial Intelligence in Humanitarian Organisations: Best Practices in Information and Knowledge Management

Organisations: Austrian Red Cross, ACCORD (Austrian Centre for Country of Origin and Asylum Research and Documentation)

1. Background and Context

The Austrian Red Cross (ARC) is committed to providing humanitarian aid, social support, and advocacy for people in need, both domestically and internationally. Within the ARC, ACCORD (Austrian Centre for Country of Origin and Asylum Research and Documentation) specialises in compiling and analysing country of origin information (COI) to support asylum and subsidiary protection procedures. ACCORD synthesises publicly available reports and data on human rights, security, and socio-economic conditions. It operates ecoi.net, the leading global platform for COI.

This consultancy project explores how humanitarian organisations are integrating Artificial Intelligence (AI) into their internal processes, particularly in information and knowledge management. This study aims to identify and analyse best practices, challenges, and lessons learned from organisations that have adopted AI tools to enhance reporting, data analysis, and information collection. It will exclude operational uses of AI, focusing instead on applications such as automated content tagging, summarisation, and AI-assisted desk research. Findings will help inform future implementation of AI tools into ACCORD's work and contribute to methodological guidance on AI use in humanitarian information management.

2. Purpose and Objectives

The purpose of this project is to assess and document how humanitarian organisations employ AI for research and information management, with a focus on ethical implications, accuracy, and internal governance mechanisms. The objectives include:

- To identify AI applications currently used by humanitarian organisations in desk research and information management.
- To analyse ethical, methodological, and operational challenges associated with AI adoption.
- To provide practical recommendations and DOs and DON'Ts for responsible AI integration within ACCORD and similar organisations.

3. Scope of Work

The consultancy will examine AI-related practices among humanitarian organisations. It will focus on secondary research and publicly available documentation, with expert interviews, if feasible. If insufficient information is available, we will examine the same fields in the journalistic sector, as they have similar principles in regard to information investigation

The guiding research question is as follows:

How are humanitarian organisations and experts using AI specifically for research, information management, and knowledge processing, and what best practices, challenges, and governance mechanisms emerge from these experiences?

4. Methodology

The study will primarily rely on desk-based research, including literature reviews, policy analyses, and examination of public case studies. If possible, it will include qualitative interviews with representatives from selected organisations. The methodology will be refined by the student research team in consultation with ACCORD and will adhere to ethical research standards, including GDPR data protection requirements. We will try to gather information on: UNHCR, IOM, the Norwegian or the Danish Refugee Council, WFP, Worldbank, Save the Children, Oxfam, governmental departments like SIDA, FCDO, or the German GIZ, as has been highlighted to us

5. Deliverables

- Draft Reports (provided along the way) summarizing findings, challenges, and good practices
- Final Report with recommendations and information from inception reports, DOs and DON'Ts in relation to AI, and presentation of results via Zoom

6. Timeline

The consultancy follows the LSE's internal academic schedule. The final report is due on 23 March 2025, followed by a presentation session with ACCORD representatives via Zoom. Ongoing coordination will be required.

7. Reporting and Communication

The LSE research team will report directly to ACCORD. Updates will be provided when needed or if information of importance comes to light. The Terms of Reference and subsequent deliverables will be reviewed and approved by ACCORD before finalization.

8. Roles and Responsibilities

LSE Team:

- Conduct desk research and compile findings into the final report.
- Deliver presentation of findings and recommendations to ACCORD via Zoom.

Austrian Red Cross (ACCORD):

- Review and approve the ToR.
- Provide background information and organisational context on current AI-related initiatives.
- Participate in the review of key deliverables and final presentation.

9. Ethical Standards and Data Protection

All research activities must comply with the Austrian Red Cross's ethical standards and the Red Cross Code of Conduct. Data handling will adhere to GDPR and ARC's internal data protection policies.

10. Contact Information

Project Focal Point: Boris Panhölzl

Email: boris.panhoelzl@redcross.at

Organisation: Austrian Red Cross, Headquarters – ACCORD Department

Appendix 2 - Terms of Reference

Project Title: Artificial Intelligence in Humanitarian Organisations: Best Practices in Information and Knowledge Management

Organisations: Austrian Red Cross, ACCORD (Austrian Centre for Country of Origin and Asylum Research and Documentation)

1. Background and Context

The Austrian Red Cross (ARC) is committed to providing humanitarian aid, social support, and advocacy for people in need, both domestically and internationally. Within the ARC, ACCORD (Austrian Centre for Country of Origin and Asylum Research and Documentation) specialises in compiling and analysing country of origin information (COI) to support asylum and subsidiary protection procedures. ACCORD synthesises publicly available reports and data on human rights, security, and socio-economic conditions. It operates ecoinet.net, the leading global platform for COI.

This consultancy project explores how humanitarian organisations are integrating Artificial Intelligence (AI) into their internal processes, particularly in information and knowledge management.

This study aims to identify and analyse best practices, challenges, and lessons learned from organisations that have adopted AI tools to enhance reporting, data analysis, and information collection. It will exclude operational uses of AI, focusing instead on applications such as automated content tagging, summarisation, and AI-assisted desk research. Findings will help inform future implementation of AI tools into ACCORD's work and contribute to methodological guidance on AI use in humanitarian information management.

2. Purpose and Objectives

The purpose of this project is to assess and document how humanitarian and related organisations employ AI for research and information management, with a focus on ethical implications, accuracy, and internal governance mechanisms. The objectives include:

- To identify AI applications currently used by humanitarian organisations in desk research and information management.
- To analyse ethical, methodological, and operational challenges associated with AI adoption.
- To provide practical recommendations and DOs and DON'Ts for responsible AI integration within ACCORD and similar organisations.

3. Scope of Work

The consultancy will examine AI-related practices among humanitarian organization. It will focus on secondary research and publicly available documentation, with expert interviews, if feasible. If insufficient information is available we will examine the same fields in the journalistic sector, as they have similar principles in regard to information investigation.

4. Methodology

The study will primarily rely on desk-based research, including literature reviews, policy analyses, and examination of public case studies. If possible, it will include qualitative interviews with representatives from selected organisations. The methodology will be refined by the student research team in consultation with ACCORD and will adhere to ethical research standards, including GDPR data protection requirements. We will try to gather information on: UNHCR, IOM, the Norwegian or the Danish Refugee Council, WFP, Worldbank, Save the Children, Oxfam, governmental departments like SIDA, FCDO, or the German GIZ, as has been highlighted to us.

5. Deliverables

- Draft Reports (provided along the way) summarizing findings, challenges, and good practices
- Final Report with recommendations and information from inception reports, DOs and DON'Ts in relation to AI, and presentation of results via Zoom

6. Timeline

The consultancy follows the LSE's internal academic schedule. The final report is due on 23 March 2025, followed by a presentation session with ACCORD representatives via Zoom. Ongoing coordination will be required.

7. Reporting and Communication

The LSE research team will report directly to ACCORD. Updates will be provided when needed or if information of importance comes to light. The Terms of Reference and subsequent deliverables will be reviewed and approved by ACCORD before finalization.

8. Roles and Responsibilities

LSE Team:

- Conduct desk research and compile findings into the final report.
- Deliver presentation of findings and recommendations to ACCORD via Zoom.

Austrian Red Cross (ACCORD):

- Review and approve the ToR.
- Provide background information and organisational context on current AI-related initiatives.
- Participate in the review of key deliverables and final presentation.

9. Ethical Standards and Data Protection

All research activities must comply with the Austrian Red Cross's ethical standards and the Red Cross Code of Conduct. Data handling will adhere to GDPR and ARC's internal data protection policies.

10. Contact Information

Project Focal Point: Boris Panhölzl

Email: boris.panhoelzl@redcross.at

Organisation: Austrian Red Cross, Headquarters – ACCORD Department

Appendix 3 - Interview Questions

1- Humanitarian organisations

AI systems

A- What AI tools does your organisation use and is it an in-built system?

B- Any issues encountered using these tools?

Policies, training, and Governance

C- Does your organisation have a publicly available AI policy?

D- How is hallucination risk managed?

E- What AI training exists and how is it delivered?

Risks, Bias, and translation

F- Do teams use AI for translation; how is quality ensured?

G- Have you observed bias in AI outputs?

2- Experts in AI in humanitarian sector

A- What common challenges or mistakes have you observed in organisations implementing AI in desk-based research?"

B- How would you define success when organisations or projects use AI?"

C- How does inclusion translate into practical guidance for organisations when they develop AI ethical policies?"

D- How should organisations account for the limitations such as data poverty, limited access to computing or underrepresentation in datasets

INTERVIEW DESIGN

Interviews were done remotely between December 2025 and March 2026 through TEAMS and Google Meet. Representatives from 7 different humanitarian organisations and specialists such as: Care International UK, UNESCO and UNDP were interviewed.

Said interviews were conducted using a semi-structured format, guided by a predefined interview blueprint for humanitarian organisations (see Appendix 3). This format ensured consistency across interviews while permitting flexibility for follow-up questions based on initial answers. For humanitarian and AI experts, specific ethical questions were asked (see Appendix 3).

Interviews were focused on: organisational use of AI in information and knowledge management, AI biases, ethical considerations, lessons learned from AI implementation, perceived risks, limitations, and future needs to ensure ethical AI policy. A list of all interviewees, the participant's name and role can be found in the Annex.

SAMPLING STRATEGY

A purposive sampling strategy was employed to select participants based on their relevant expertise in humanitarian, research, and policy fields, including individuals working in humanitarian organisations, research institutions, international development finance institutions, and policy-focused organisations. It also incorporated elements of convenience sampling, as participants were identified through accessible public sources like organisational websites, public staff directories, LinkedIn and direct email outreach. To complement external perspectives, an initial interview with the Austrian Red Cross (ACCORD) was conducted to understand internal AI use.

ETHICAL CONSIDERATIONS

In order to ensure ethical standards were maintained all interviewees were provided with a consent form that informed them of the purpose of the research, risks of data sensitivity and consent for recording and storage. The preference of participants who chose to remain anonymous or to not be recorded were followed.

Interview transcripts were analysed thematically using FATHOM, a qualitative analysis tool. Thematic analysis was employed to identify recurring themes and patterns, as well as divergent perspectives across organisational types (Bryman, 2016). This analysis was triangulated with findings from the literature review to strengthen the validity and reliability of results.

Appendix 4 - Interviews

Name	Organisation
Samuel Goodger	European Policy Centre
Emma Ruttkamp- Bloem	Professor of Philosophy and AI Ethics Academic Specialist in AI Ethics and African Philosophy Global AI Ethics Expert (UNESCO Contributor)
Anonymous	UNDP staff
Mohannad Alghabra	CARE International UK
anonymous	UNRWA
anonymous	anonymous
anonymous	anonymous

Theme	CARE INTERNATIONAL UK	Emma Ruttkamp-Bloem	UNDP	EPC analyst (Samuel Goodger)
AI tools used	Copilot, ChatGPT, Gemini	Not discussed	ChatGPT, internal drafting tool	Claude, Gemini, ChatGPT, Copilot
AI policy adopted internally	CARE USA policy not yet adopted	Helped developed UNESCO AI policy	Policy in progress	Internally only not public
Hallucination management	Manual cross-checks with excel for datasets	Human oversight is essential	Peer review, prompt training	Prompt AI to be critical
Translation use	Internal/low-risk use only	Dialect and cultural loss at risk	Native speaker review for publications	Verified by colleagues if sensitive data
Over-reliance risk	AI reliance could erode humanitarian judgment	<i>"It is not magic, it is technology"</i>	Poor-quality AI institutions risk degrading the outputs	Human skills more important, not less
Bias awareness	No instances observed but aware	All data inherently biased	Overcorrection risk in training	Not observed critical prompting used
Human oversight	Every step manually validated	<i>"For this to work we still need humans"</i>	Human in the loop for all outputs	Strict output verification