ETHICAL AI FOR HUMANITARIAN DATA, KOBO

INTRODUCTION

Kobo's AI-powered qualitative data tools allow humanitarian organizations to efficiently gather and analyze detailed, open-ended information from affected communities. KoboToolbox employs automatic speech recognition (ASR) to turn audio responses into text, combined with machine translation (MT) to translate text into over 100 languages. In addition, a feature leveraging large language models (LLMs) for rapid coding and analysis of qualitative data is fully operational and undergoing piloting. These tools are ethical by design, non-extractive, and integrate human verification of AI-generated outputs.

Audio recording and automated transcription and translation features are available to all 32,000 organizations and 700,000 individuals that use KoboToolbox globally (59% in developing countries). Humanitarian and other organizations are using these tools to better understand nuanced community needs and capacities to inform more inclusive aid and service delivery.

Kobo is the nonprofit behind KoboToolbox, the open-source data collection, management, and visualization platform most widely used across the humanitarian sector.

PROBLEM

In humanitarian emergencies, gathering detailed, open-ended information directly from communities in their own languages is crucial for designing programming to deliver targeted aid. However, the time and technical expertise required to process massive amounts of multilingual qualitative data frequently leaves vulnerable populations unheard and without inclusive access to aid like food, shelter, and medical care.

As an example, a refugee woman may have particular capacities and needs that she would like to express to a humanitarian worker as part of a needs assessment interview. However, if they do not share a common language and data collection involves only pre-defined, closed-ended questions, then nuances from her responses will be lost and the assistance delivered will not meet her needs. The data-collection process will also have been lacking in dignity. Up to this point, collecting and analyzing detailed, open-ended responses was generally unrealistic, however, due to the cost and time constraints of processing hundreds, or even thousands, of qualitative interviews.

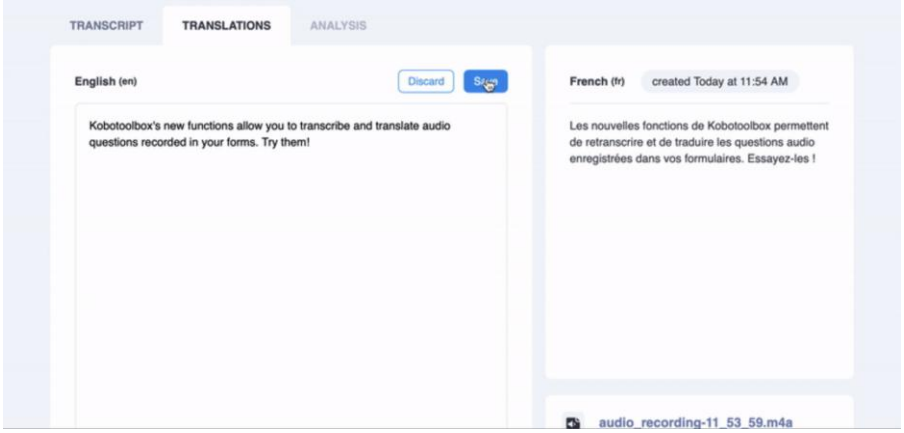
AI DEVELOPMENT

Kobo's experience working in humanitarian contexts and partnering with aid organizations quickly revealed that asking affected communities only closed-ended, quantitative questions was a recipe for unresponsive aid delivery. We first introduced our proposed innovation for using AI methods to speed up the transcription and analysis of more open-ended interview data in a collaborative paper published with colleagues from the International Rescue Committee, Translators Without Borders (now CLEAR Global), NetHope, the World Food Programme, York University, and the Harvard Humanitarian Initiative in a paper published in 2020.

To better understand the problem and prepare a response, we then interviewed more than 20 humanitarian staff with deep data collection and analysis experience, asking about operational barriers to open-ended, qualitative data collection, ethical concerns, and their recommendations for using AI to improve and expand qualitative interviews. Next, we conducted a rigorous review of potential ethical issues related to AI data processing of humanitarian data, which was published in 2025. Next, Kobo developed a prototype tool with transcription, translation, and data analysis features, which we beta tested through workshops with local

organizations. The workshops informed feature design, use cases, and ethical safeguards, ensuring the resulting tools would be user friendly, adaptable, and ethical by design. Throughout the process, Kobo's team prioritized non-extractive data practices, ensuring that sensitive information was never used for commercial model training.

Following successful refinement and community-led testing, the ASR and MT features became available on the KoboToolbox platform. The LLM-powered qualitative data analysis feature is operational and will be rolled out in 2026. These tools ensure that even with limited resources, organizations can now collect and analyze thousands of complex community testimonies–supported by responsibly designed, human-in-the-loop AI tools.



Automatic transcription and translation features to efficiently produce and translate transcripts of audio recordings on KoboToolbox.

OPERATIONALIZING AI

The automated transcription and translation tools are fully deployed and available to humanitarian organizations worldwide. Integrated directly into KoboToolbox's global data platform, the tools leverage ASR to transcribe audio responses in 80 languages and 145 dialects, and MT for translation into 129 languages. The human user reviews, edits, and validates the AI outputs within KoboToolbox. The original transcript and any translations are then added to the data spreadsheet as new columns and securely stored on KoboToolbox servers. Finally, organizations analyze and visualize the validated qualitative data to inform humanitarian decision-making, programming, and aid delivery.

A rigorous evaluation of the tools deployed by UN Women demonstrated that automated transcription and translation significantly reduced data processing time and operational costs. Field testing has also revealed high user satisfaction, with humanitarian staff highlighting the security and convenience of transcribing and translating audio directly within KoboToolbox, rather than exporting sensitive data to third-party software.

The AI-powered analysis feature is fully operational and undergoing robust piloting and extensive LLM performance and red-teaming validation. Once complete, the feature will be rolled out widely among KoboToolbox users.

LEARNINGS

Advances in AI and language technologies create an opportunity to reduce operational barriers to qualitative analysis, including through automated transcription, translation, and coding. However, a key learning is that these tools should support rather than replace human qualitative expertise and contextual interpretation. The team found that local humanitarian staff are key for supervising the transcription and translation of

interview responses from affected communities, including reviewing and editing ASR and MT outputs based on local dialects and knowledge of on-the-ground realities. Likewise, results are most robust for LLM-powered analysis of qualitative data when overseen by staff with a strong foundation in thematic coding and related qualitative methodologies.

The second key learning was to keep affected communities at the heart of the project. When deploying these tools in local communities, the team saw firsthand that affected persons often felt more comfortable speaking while their interviewer actively listened rather than watching them type responses into a data-collection app. When combined with asking respondents open-ended questions, Kobo saw a more dignified data-collection process unfold. In debriefs, interviewers reported that the flexibility allowed them to build rapport with respondents, while respondents shared that giving narrative answers to open-ended questions made them "feel heard."

PLANS FOR THE FUTURE

Future iterations of the automated transcription, translation, and analysis tools will expand language model coverage to over 200 languages and regional variants, integrating additional language model providers. A primary focus is scaling the technical framework to fine-tune custom models for more critically underserved, low-resource languages, such as Congolese Swahili and Nande. Critically, Kobo is pursuing custom language model development for Rohingya, Kurdish dialects, and other low-resource languages in partnership with UNHCR and CLEAR Global.

In addition, Kobo is establishing a reproducible, sector-wide methodology for assessing the accuracy and reliability of LLMs for coding qualitative humanitarian data. This methodology will be shared in an academic or practitioner-focused journal, as well as public-facing outlets like blog posts.

The Kobo team will also continue to monitor developments in the AI space, including assessing newer LLMs and incorporating them into this and other use cases in the future.

WHERE TO LEARN MORE

Tino Kreutzer, Chief Operating & Innovation Officer, Kobo, tino.kreutzer@kobo.ngo

Kobo.ngo