



Automating Survey Quality Control Using AI: Application to Employment and Livelihood Measurement



Nouréini Sayouti, Celine Wuyts, Steven Glazerman-Better Data for Lives

December 2025

Outline



01 Introduction

02 Multilingual Speech Transcription Pipeline

03 NLP-Based Quality Indicators

04 Pilot Implementation & Scale-Up Potential

The Research Team



Nouréini Sayouti
Research Scientist ARM
Automatic Speech Recognition



Celine Wuyts
Research Scientist, ARM
Survey Methodology



Steven Glazerman
Chief Research and Methodology Officer
Survey Methodology

01

Introduction

Motivation - Why this matters

- Employment & livelihood surveys critical for poverty measurement, program targeting, and labour market analysis in LMICs.
- But they face **distinctive measurement challenges**:
 - Informal work, household businesses, complex job histories.
 - Multilingual interviews, local languages, translation needs.
 - Cognitive demands on respondents and enumerators → risk of **systematic measurement error**.
- Accurate data where it matters most — but measurement error may be highest there.



Source: [World Bank Blogs](#)

Current Quality Control Limitations



- Many surveys now record audio — aiming for quality monitoring / auditing.
- Yet: **manual review** of recordings is cost-prohibitive → realistically check only 5–10% of interviews.
- In multilingual settings: supervisors often don't understand local languages → cannot reliably detect translation problems, interviewer deviations, comprehension issues.
- Generic ASR / transcription tools: poor support for low-resource languages or domain-specific vocabulary → transcription output unreliable.
- **Result:** large share of data remains effectively unchecked; potential biases/unobserved errors remain hidden.



Source: [SurveyCTO](#)

02

Multilingual Speech Transcription Pipeline

Proposed Solution – Overview of the Pipeline

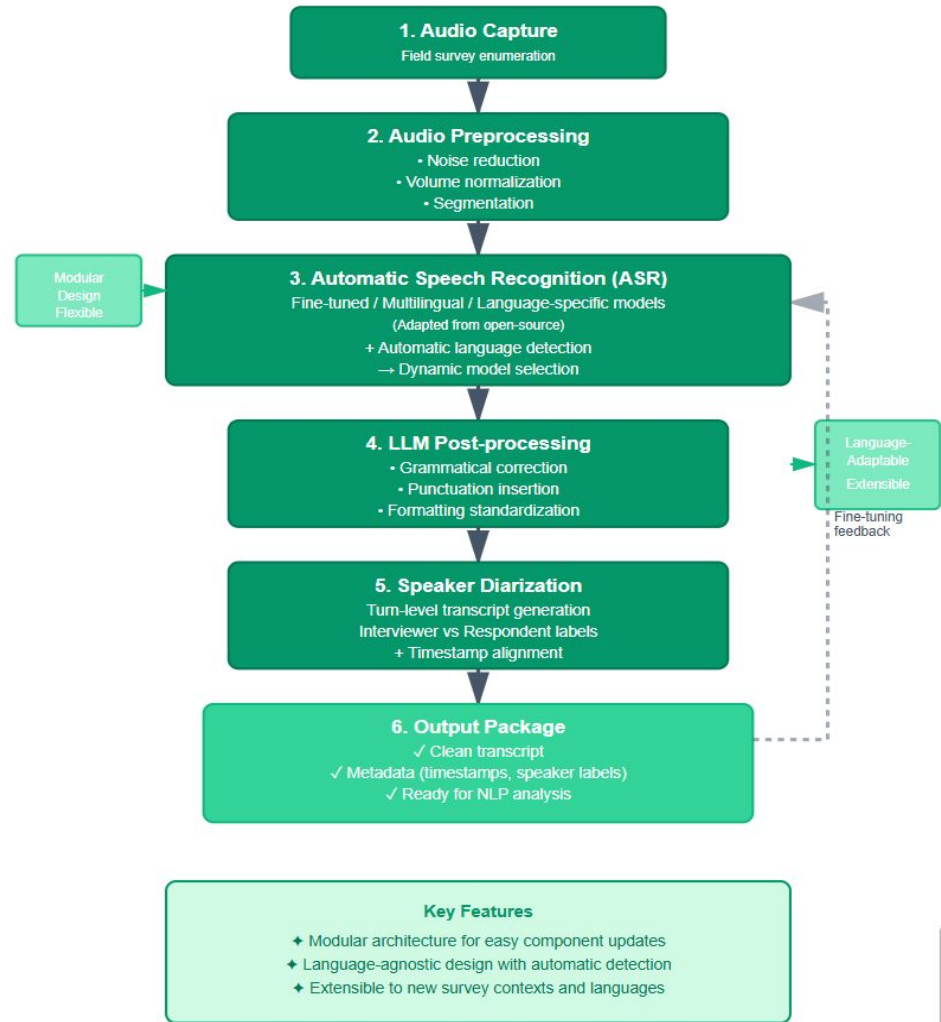


We propose an **AI-powered pipeline** that integrates:

1. Speech recognition (for multilingual, local-language interviews)
2. Post-processing via LLMs (for cleaning transcripts, formatting, correcting)
3. Speaker diarization (to separate interviewer vs respondent turns)
4. Automatic extraction of quality indicators via NLP (interaction dynamics, hesitations, translation patterns, etc.)

Goal: enable **scalable, automated quality control** covering all interviews — not just a small spot-checked fraction.

Pipeline Architecture — Process Flow



03

NLP-Based Quality Indicators

What we can measure



01 Interviewer Performance

Adherence to question script • Verbatim reading • Proper probing • Translation practices • Deviations or potential falsification

02 Respondent Cognitive Processing

Hesitations • Clarifications • Long response latencies • Signs of comprehension difficulty or recall strain

03 Emotional / Relational Dynamics

Indicators of discomfort, frustration, embarrassment, rapport — as influencing response quality

04 Interaction Dynamics

Patterns of turn-taking • Language switching • Conversational flow • Interruptions • Overlaps — reflecting interview quality and respondent engagement

Key Implications

Systematic, scalable detection of issues at interview-, question-, translation-, or interviewer-level → Enables identification of problematic items, languages, or interviewers

04

Pilot Implementation & Scale-Up Potential

Proof of Concept & Feasibility



- We ran **proof-of-concept pilots** in:
 - **Tanzania** (Swahili + English);
 - **Nigeria** (English + Nigerian Pidgin)
- Used **existing open-source ASR models** (no fine-tuning yet) to test transcription feasibility across languages and typical field acoustic conditions.
- Results: **transcription feasible** — acceptable baseline performance, demonstrating viability across languages/acoustic variability.
- System requirements modest: standard laptop (GPU beneficial), basic Python setup — accessible to many research institutions or survey teams.
- Plan: open-source release of pipeline + pre-trained models + fine-tuning protocols — enabling adaptation to new languages and contexts.

Scale-Up Potentials & Benefits



Key advantages of proposed pipeline:

- **Scalability:** ability to review 100% of interviews rather than small sample spot-checks — improving coverage and reducing risk of unobserved error.
- **Transparency & replicability:** open-source tools allow other researchers/organizations to adopt, adapt, contribute.
- **Improved data quality:** detection of measurement error sources — interviewer deviations, translation issues, recall problems — leading to more reliable labour / livelihood data.
- **Methodological insights:** structured data on interaction dynamics, respondent behaviour, translation/ language switching — useful for survey design, training, future data collection improvements.
- **Flexibility:** topic-agnostic — can be applied to other kinds of household / social surveys, beyond labour/employment.

Considerations & Challenges



- **Need for fine-tuned ASR models per language:** for high accuracy in low-resource languages or dialects — requires labelled data (audio + transcripts) for fine-tuning.
- **Ethics / privacy concerns:** storage of audio recordings, informed consent from respondents, data security and confidentiality.
- **Field logistics constraints:** audio quality (noise, environment), device management, enumerator training, consistent audio capture across contexts.
- **Technical capacity and maintenance:** need for computing resources (for processing, occasional fine-tuning), as well as human capacity to manage pipeline, interpret outputs, act on quality flags.

05

Q&A



Thank you.

nsayouti@poverty-action.org

poverty-action.org