THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

LSMS
Living Standards Measurement Study

The World Bank's Flagship Household Survey Program

Enhancing Living Standards Surveys in LMICs using Large Language Models

Philip Wollburg, Senior Economist, DECSU

SoDa Symposium, February 24, 2026

Overview of today's presentation

Part I: Overview of the Living Standards Measurement Study (LSMS) program

Part II: Survey Methods and Measurement program

Part III: Using LLMs to enhance household surveys in LMICs

THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

LSMS
Living Standards Measurement Study

The World Bank's Flagship Household Survey Program

Overview of the Living Standards Measurement Study (LSMS)

Living Standards Measurement Study (LSMS)

The World Bank's flagship household survey program

Data Production & Advisory Services

Survey Methods

Policy Research

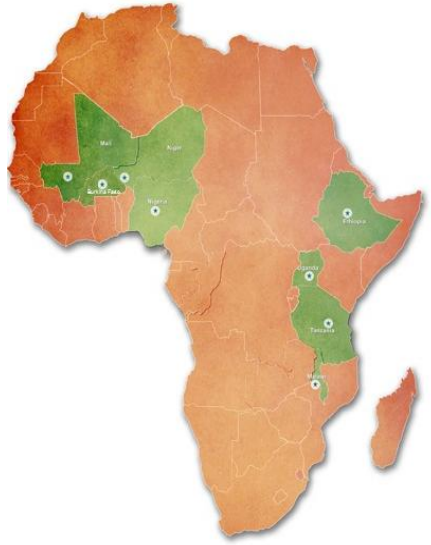
Training

OBJECTIVE

To provide a comprehensive approach to household surveys across the survey life cycle, producing timely and quality data for better-informed policy design and development investments.

LSMS Data production

8 Partner National Statistical Offices (2008-2024)

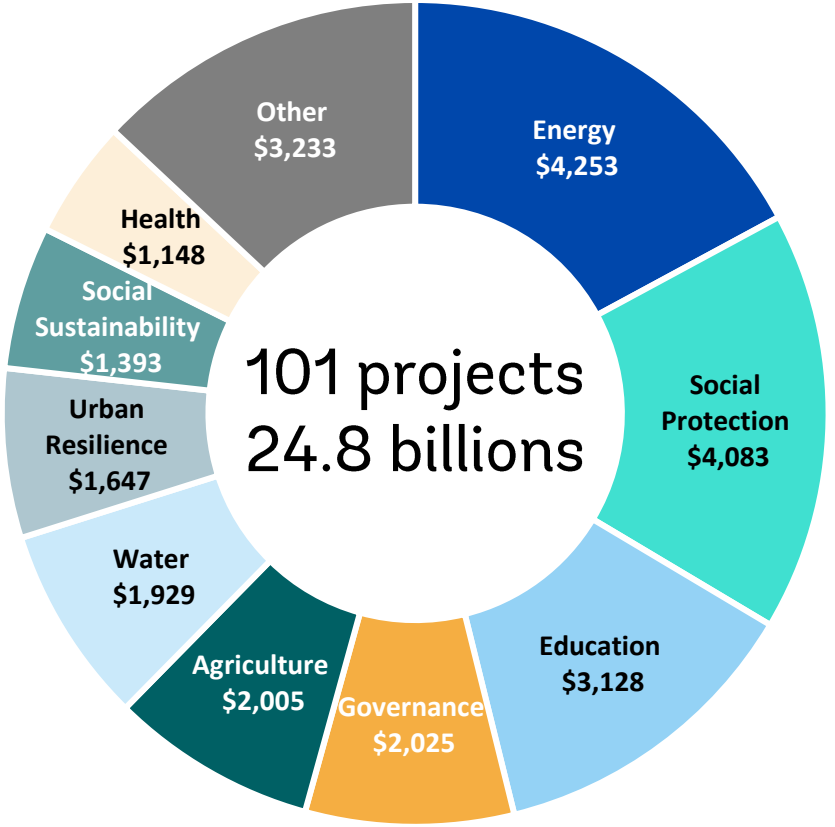


Burkina Faso, Ethiopia, Malawi, Mali, Niger, Nigeria, Tanzania, Uganda

Open-Acces Data:

- 33 face-to-face surveys (2008-2024)
- 124 phone surveys (2020-2025)
- 4 harmonized datasets
- 7,200 publications
- 117,000+ microdata library downloads
- 35% by African researchers

101 World Bank projects informed by LSMS-supported longitudinal face-to-face and phone surveys from 2008 to 2024



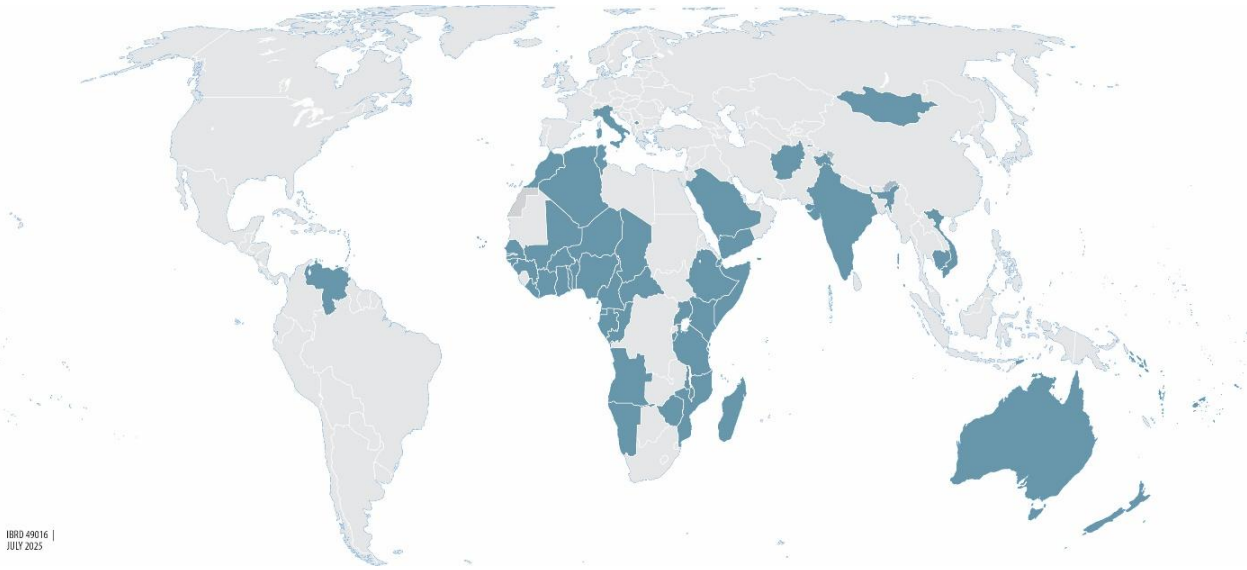
LSMS technical assistance and trainings

Global engagement:

- **Build technical capacity** in partner countries to strengthen household survey systems and improve the quality and use of microdata
- **C4D2 training initiative** to provide statisticians in LMICs with specialized skills

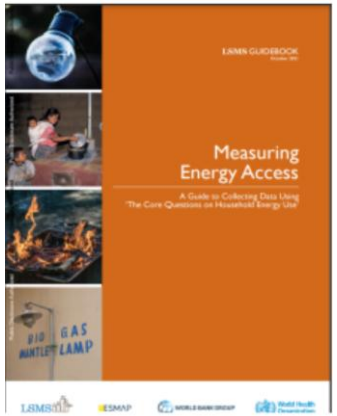
2025/26 Highlights :

- **Two training courses** under the aegis of the World Bank Data Academy
  - Mastering the Design & Implementation of Household Surveys
  - Introduction to Sampling: Theory & Practical Applications
- **60 countries received technical assistance** for face-to-face and phone surveys



LSMS Knowledge and R&D

Global outputs:



- **Guidebook series**
  - Designing household surveys
  - Employment
  - Energy Access, …
- **Publication** of peer-reviewed articles, working papers, synthesis reports, policy briefs on:
  - Policy research
  - Survey methods
- **Outreach and dissemination** of knowledge products in blogs, conferences, and seminars.

2025/26 highlights:

- **38 knowledge products:** 16 journal articles, 12 working papers, 4 reports, 3 guidebooks, 3 policy briefs
- **2 expert convenings** – on climate adaptation and on jobs, informality and unpaid work
- **2 academic conferences (**on high-frequency survey data for development research in the polycrisis area and on survey methods)

THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

LSMS
Living Standards Measurement Study

The World Bank's Flagship Household Survey Program

The LSMS Survey Methods and Measurement program

Background

- Household surveys remain the backbone of evidence generation on living standards in low- and middle-income countries (LMICs) and key tools to inform policy and investments
- But survey systems are under growing pressure to do more at lower cost as budgets contract while demands for data continue to rise: More content, more granular, faster, more frequently
- The result is growing strain on survey instruments, field teams, and respondents — with rising risks for data quality
- Innovative survey methods play key role in addressing these challenges by reducing costs and maintaining data quality and policy relevance → key principle: validation, then scale-up
- Building on previous efforts on methodological innovation in large-scale surveys, the LSMS has launched a five-year program to develop and validate a new wave of AI-enabled and AI-enabling survey methods for use in LMICs

Surveys and survey research in the LSMS

LSMS
Data
Production

- Platform for scaling up improved methods
- Learn what works in **low and middle income** and low-capacity contexts
- Help country partners **build capacity** in the latest survey methods
- Identify **new areas of study**

- Developing and validating **best practices (improved survey methods) → experimental validation**
- Developing **practical guidelines and tools** that are easily adoptable in low and middle income contexts based on the improved methods
- **Improve survey quality** through innovation and new technologies

LSMS
Research

Current Areas of Work

Methods

- Sampling for priority populations
- Mixed-mode surveys and high-frequency measurement
- Individual-level data collection
- Survey-to-survey imputation
- Surveys as ground truth for ML/AI models
- **AI and digital tools to enhance surveys in LMICs**

Topics

- Employment and small business
- Energy
- Agriculture & food security
- Resilience and adaptation
- Health

Field experiments for methods developement

- **Bangladesh**
  - Timeframe: Jan–Oct 2026
  - Method: Mixed Mode
  - Focus Areas: Adaptation and Resilience, Employment → AI
- **Nepal**
  - Timeframe: April–Dec 2026
  - Method: Sampling
  - Focus Areas: Small Business, Adaptation and Resilience, Agriculture → AI

- **Kenya**
  - Timeframe: Sep 2026–Jun 2027
  - Method: Sampling for hard-to-reach populations
  - Focus Areas: Small Business → AI
- **Malawi**
  - Timeframe: Jan-Oct 2026
  - Method: Phone
  - Focus Areas: Adaptation and Resilience, Small Business → AI

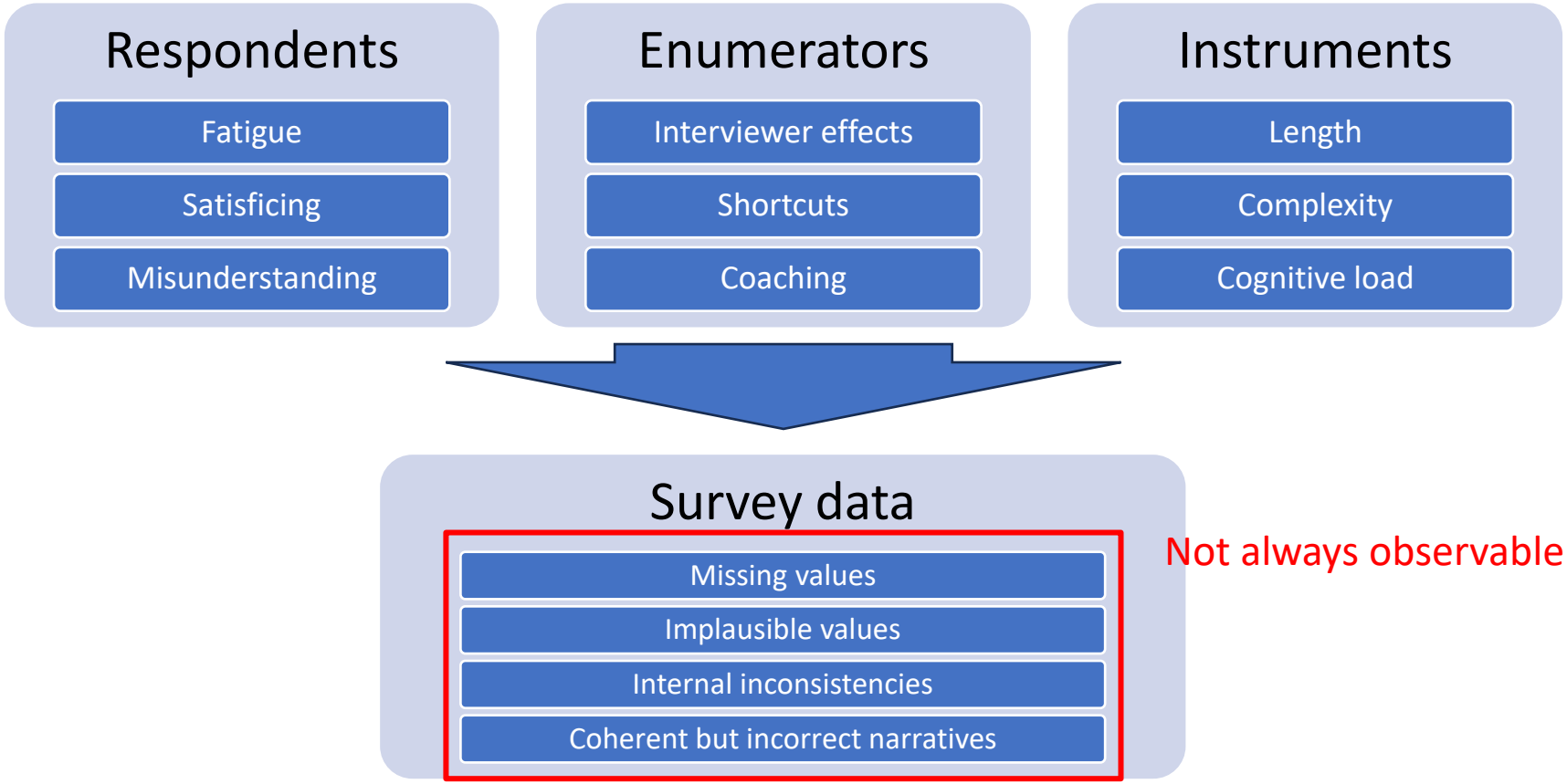
THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

LSMS
Living Standards Measurement Study

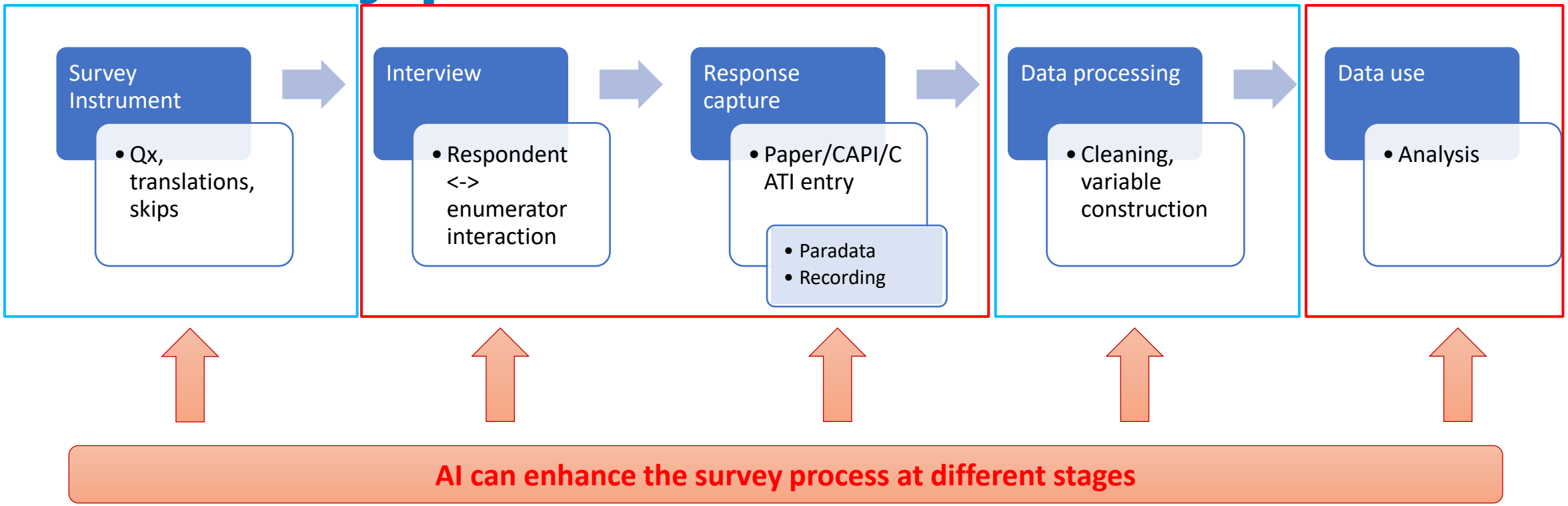
The World Bank's Flagship Household Survey Program

Using LLMs to enhance household surveys in LMICs

Data quality challenges in surveys



The survey process and AI



Principles of the LSMS AI work program

- Focused on measurement quality under constraints
- Combining engineering and survey methodology
- Relies on experimental validation
- Built for deployment, not demonstration
- Scale up in large-scale household surveys

Pillars of the LSMS AI work program

Foundational layer: Recording-Transcription-Translation pipeline

- Audio recording
- Transcription and translation using LLMs
- Linking text to micro- and paradata

Workstream 1: Detecting and reducing survey error

- Detect irregularities and faulty responses
- Patterns correlated with poor data quality

Workstream 2: Simplifying and enriching questionnaires

- Open-ended responses; automated coding
- Narrative data

Foundational layer: Recording-Transcription

Step 1: Recording of interviews

- Interviews take place in noisy, uncontrolled environments
- Devices, microphones, and field conditions vary widely → phone vs. in-person surveys
- Enumerator vs respondent audio
- Audio quality differs across enumerators, locations, and time

Challenge:

➢ How to obtain usable (and comparable) recordings across contexts?

Foundational layer: Recording-Transcription

Step 2: Transcription from speech to text

- Automatic speech recognition is imperfect in low-resource settings
- Transcription errors may be systematic, not random

Challenge:

- ➢ When is transcription good enough for downstream analysis?

Foundational layer: Recording-Transcription

Step 3: Translation

- Low-resource languages
- Translation introduces drift in meaning and emphasis
- Errors can compound with recording-transcription-translation

Challenge:

➢ How to preserve meaning and comparability across languages?

Foundational layer: Recording-Transcription

Step 4: Linking text to survey micro- and paradata

- Transcripts must align with questions, responses, and survey paradata
- Timing, skips, and corrections complicate alignment
- Misalignment undermines both quality control and analysis

Challenge:

➢ How do we reliably connect text to the underlying survey structure?

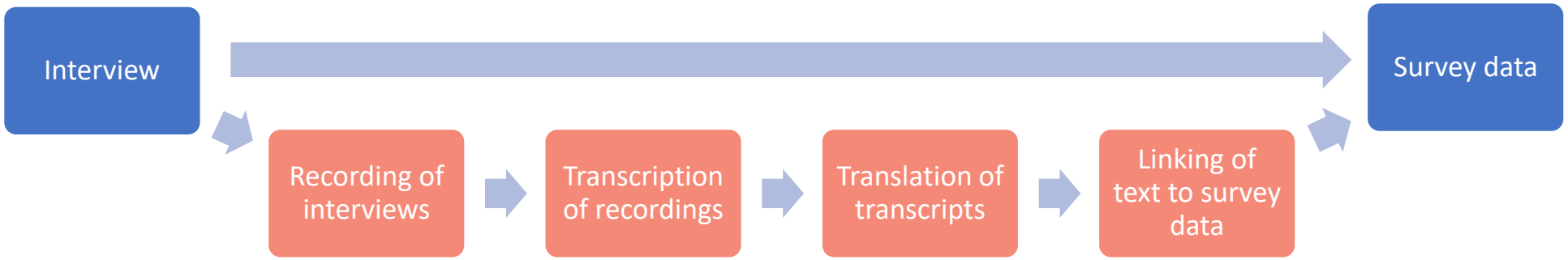
Foundational layer: Recording-Transcription

Prototype development

- We have developed a prototype for the transcription pipeline that can handle low- and high-resource languages
- Built-in layer for model training + validation

From observability to measurement

- The foundational layer makes the interview itself analyzable
- Creates new signals/data beyond the entered responses
- The new data will be used to improve measurement and analysis



Workstream 1: Detecting and reducing error in survey data

Overview

- Goal: use AI to detect survey errors that standard QC misses
- Input: structured responses + interview text + paradata
- Output: actionable quality signals during or after fieldwork

Workstream 1: Detecting and reducing error in survey data

Step 1: Obtain ground truth

- Survey data rarely reveal whether a response is truly wrong
- Large-scale surveys contain plausible but incorrect responses
- Many errors pass standard range and logic checks

Challenge:

- How do we generate reliable training and evaluation data for error detection?

Workstream 1: Detecting and reducing error in survey data

Ground truth in Bangladesh study:

- Observable ground truth: housing, WASH, infrastructure
- Consistency checks across answers
- Inverted response scales

→ Less feasible in large-scale surveys

Workstream 1: Detecting and reducing error in survey data

Step 2: Representing survey interviews for AI systems

- Survey data + transcripts combine structure, language, and interaction
- Errors often emerge across questions, not within them
- LLMs require explicit choices about alignment, granularity, and context

Challenge:

➢ How do we represent interviews so that AI can reason about quality without losing structure or interpretability?

Workstream 1: Detecting and reducing error in survey data

Step 3: Using LLMs for quality control

- LLM outputs are probabilistic and non-deterministic
- Responses depend on prompt design and context length
- Reasoning is not always transparent or stable

Challenge:

➢ How can LLMs be used as diagnostic tools for data quality — in a replicable, transparent way?

Workstream 1: Detecting and reducing error in survey data

Step 4: Benchmarking and validation

- Rule-based QC is transparent but limited
- Statistical models detect patterns but lack semantic context
- LLM-based systems may flag new errors but risk overreach → can they identify behaviors that lead to error?

Challenge:

➢ When, where, and for which types of error do LLMs add value relative to simpler methods?

Workstream 2: Simplifying and enriching questionnaires

Problem

- Household surveys rely heavily on structured questions
- Structure makes data easier to analyze, but interviews longer and performs poorly for complex concepts (e.g. coping) → surveys can lack depth
- Long questionnaires increase burden and fatigue
- Structured formats often miss nuance and context

Workstream 2: Overview

- Goal: use AI to simplify questionnaires and interpret open-ended questions at scale
- Input: open-ended questions → interview transcript
- Output: classification/structuring of survey questions

Workstream 2: Simplifying and enriching questionnaires

Step 1: Designing effective open-ended questions

- Open-ended questions vary widely in response quality
- Some respondents give rich answers, others very little
- Enumerator prompting influences what is said → enumerator effects

Challenge:

➢ How to design short, robust prompts that work across respondents and settings?

Workstream 2: Simplifying and enriching questionnaires

Step 2: Interpreting open-ended responses at scale consistently

- Respondents express similar experiences in different ways
- AI must group diverse answers into shared categories
- Categories must remain stable across surveys and countries

Challenge:

➢ How do we ensure that AI interpretations are both meaningful and comparable?

Workstream 2: Simplifying and enriching questionnaires

Step 3: Benchmarking against traditional survey questions

- Structured questions are well-understood and standardized
- Open-ended approaches promise richness but add noise
- Gains must be demonstrated empirically

Key issue:

When do AI-processed open-ended questions perform better than traditional instruments — and when do they not?

Summary and Outlook

What we are building

- A new observability layer for surveys: Audio → text → structured linkage
  - ➢ Current status: prototype built for transcription, translation; work on matching to survey data → to be deployed in Bangladesh, Nepal, Kenya, Malawi

Two measurement applications

- Workstream 1: Detecting hidden error in survey data
- Workstream 2: Simplifying questionnaires and improving analytical depth of surveys

Looking ahead

- Open technical and methodological challenges in both workstreams
- Opportunities for collaboration and student research

LSMS

Living Standards Measurement Study

The LSMS website