

Evaluation in Times of Uncertainty - AI and Professionalization in Georgia

Topic: AI AND EVALUATION: FROM PRACTICE TO PROVOCATION

 3rd June 2026 ·  11:00 AM – 1:00 PM Tbilisi  Hybrid

Presenter: Mr. Maulik Chauhan

Founder & Managing Director, Trestle Research & Consulting | India Lead, SurveyCTO

www.trestle.co.in | maulik@trestle.co.in



This Presentation Is Not About AI

It is about helping evaluators spend more time understanding people and less time fighting processes.

The Real Heroes

- Evaluators & Researchers
- Enumerators & Field Teams
- NGO & Government Professionals
- Community Workers
- Decision-Makers & Beneficiaries

AI is the supporting actor. People are the story.



Three Dangerous Myths About AI

Myth 1: AI thinks like humans

Reality: AI predicts statistical patterns
it has no understanding, intuition, or
empathy.

Myth 2: AI is unbiased

Reality: AI reflects the biases baked
into its training data, often invisibly.

Myth 3: AI always tells the truth

Reality: AI can hallucinate facts,
confidently, and convincingly.



What is the common thing in all these pics?

AI predicts patterns. Humans gives the reason, context, and the judgment



Why AI Matters for Evaluation Today

The Old Challenge

📁 Empty database

Data Scarcity, Not enough data from the field



SHIFT

The New Challenge

📊 Overflowing data

Data Overwhelm, Making sense of it quickly, accurately, and at scale

Data Scarcity

The old bottleneck

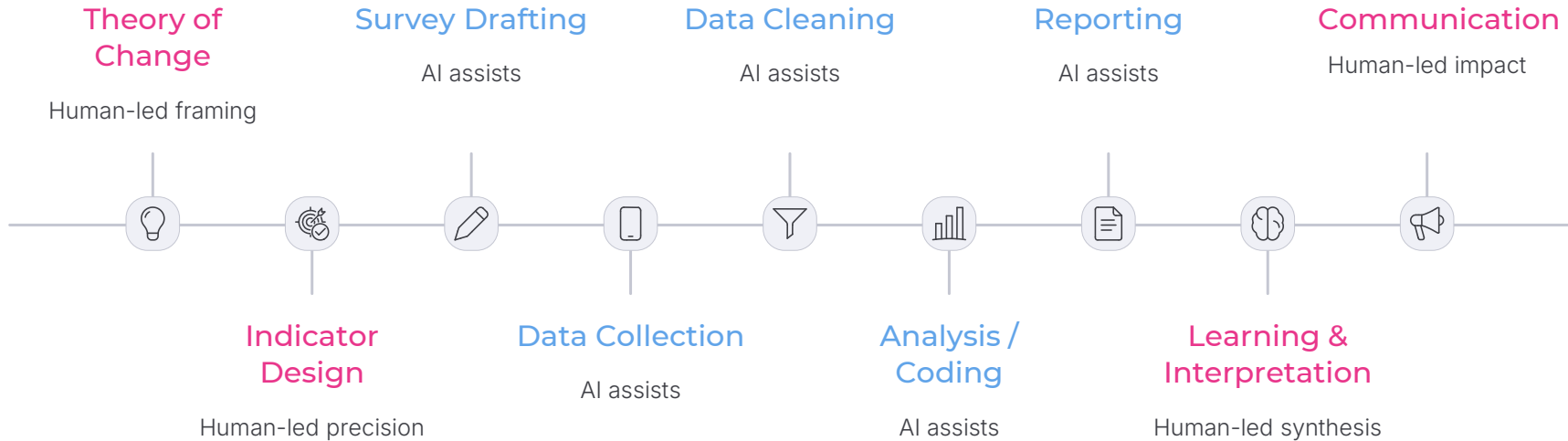
Data Sense-Making


The new bottleneck


Evaluators today face an avalanche of reports, transcripts, dashboards, and datasets. The bottleneck has shifted from data scarcity to data sense-making.



Where AI Adds Value Across the Evaluation Cycle

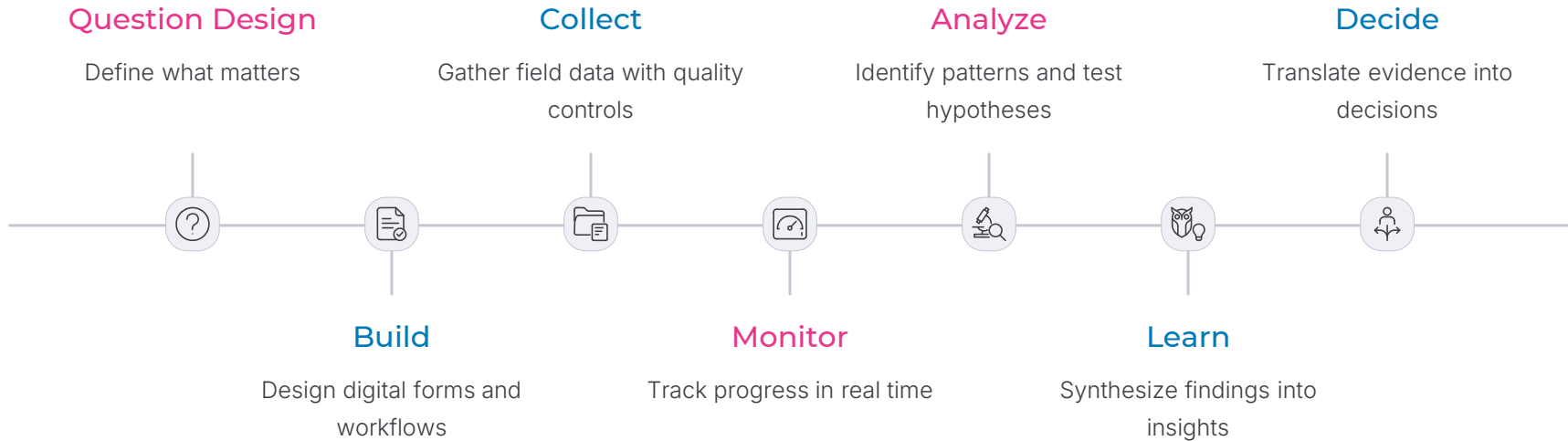


 HUMAN JUDGMENT REQUIRED

 AI ASSISTS



AI Across the Entire M&E Data Journey



AI is becoming an intelligent layer across the entire evidence ecosystem, not replacing any stage, but making each one faster, sharper, and more reliable.



Better Questionnaires. Better Data.

Before

"Is drinking water available?"

Yes / No

⚠️ Vague. Not decision-useful.

After

"How do households access drinking water?"

- Home connection
- Public tap
- Travel more than 500 meters
- Have to pay

✅ Decision-useful. Field-grounded.

Field teams were collecting data that could not answer program questions.
AI helps us move from vague questions to decision-useful questions.



From Months to Hours

Building advanced digital workflows once required external developers, long timelines, and significant costs. Field teams waited. Learning was delayed.

SurveyCTO

Digital data collection Platform

Integration with WhatsApp / SMS/Google Calendar Field Plugins

Field communication, Scheduling, and Coordination

Advanced Digital Data Collection Work Flow

Prototyped in just one afternoon

Reviewing Digitized Tools and Data Collection Work Flows

Takes minutes instead of days and months

More time for learning. Less time for technical bottlenecks.

SurveyCTO

Streamlined data flow

Collect data in
SurveyCTO



Stream data
updated in
SurveyCTO
back into the
other platform



Publish data
updated in
other platforms
to SurveyCTO

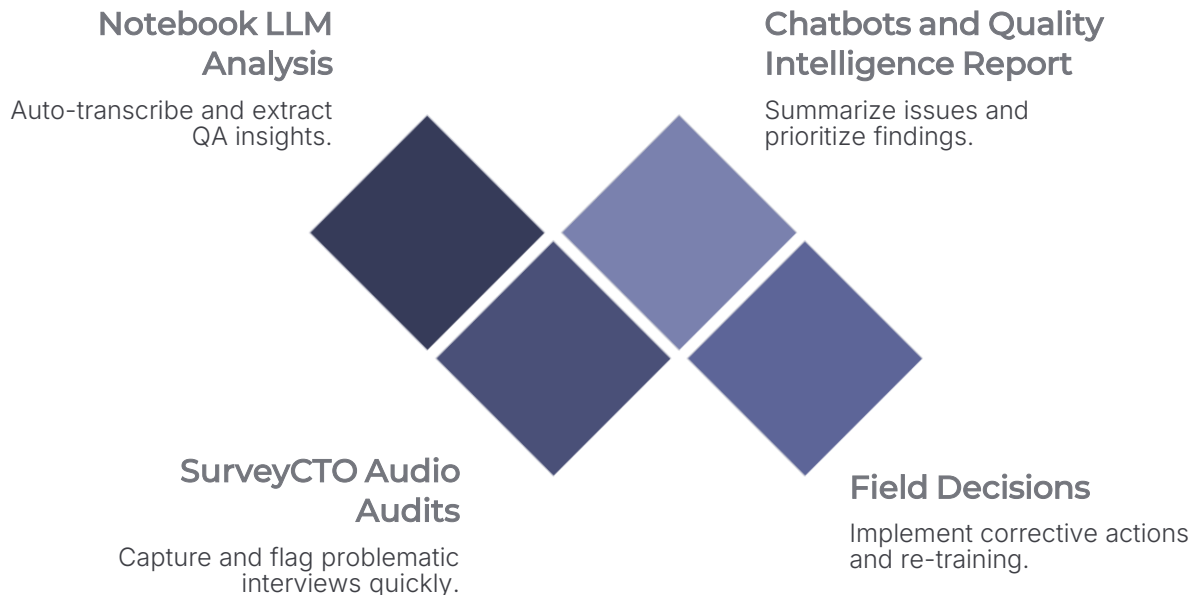


Publish data to
CRMs, HR tools,
ERPs, or other
platforms



AI-Powered Quality Assurance in Field Data Collection

Supervisors could not realistically review hundreds of interviews.
Quality issues went undetected for weeks, sometimes for entire fieldwork phases.



Before
Quality control at the end - too late to fix.

After
Quality issues identified from day-one of fieldwork.



Compliance in Minutes, Not Days

Inputs

- DPDP/GDPR Act
- Government Guidelines
- Internal Policies

Output via Notebook LLM

Actionable answers - fast, traceable, auditable.

AI is becoming a compliance co-pilot.
More confidence. Less time spent searching.

DIGITAL PERSONAL DATA PROTECTION ACT, 2023



What Didn't Work



Beautiful but Unmeasurable

Indicators that looked good on paper but could not be collected in the field.



Wrong Statistical Recommendations

AI suggested methods inappropriate for the sample design.



Missed Cultural Nuances

Context-blind outputs that misread community dynamics, local culture or community sensitivities.



Incorrect Compliance Interpretations

Regulatory guidance that required expert correction



Garbage in and Garbage out

Bad or weak data/data systems cannot be fixed by strong AI. It misguides, Garbage in and Garbage out.



The biggest AI failures happen when humans stop thinking. AI should assist judgment, not replace it.



Why Some AI Use Cases Succeed

✓ AI Works Best When...



Repetitive

Tasks done the same way, over and over



Structured

Clear inputs and predictable outputs



High Volume

Large amounts of data or content to process



Low Stakes

Errors are catchable and correctable



High Quality Data

Clean data, structured inputs produce reliable outputs

✗ AI Struggles When...



Context Matters

Local nuance and lived experience are essential



Politics Matters

Power dynamics shape what data means



Ethics Matters

Values and judgment cannot be automated



Relationships Matter

Trust is built by humans, not algorithms



Problem Statement Is Unclear

Vague instructions lead to irrelevant or misleading outputs



Five Rules for Responsible AI in Evaluation



Learn. Unlearn. Relearn.

Stay current , AI capabilities evolve rapidly.



Prompt First. Ask Later.

Invest in prompt quality before accepting outputs.
Follow RCTFE (Role, Context, Task, Format, Example)



AI is Baseline, Not Endline.

Treat AI output as a starting point, not a conclusion.



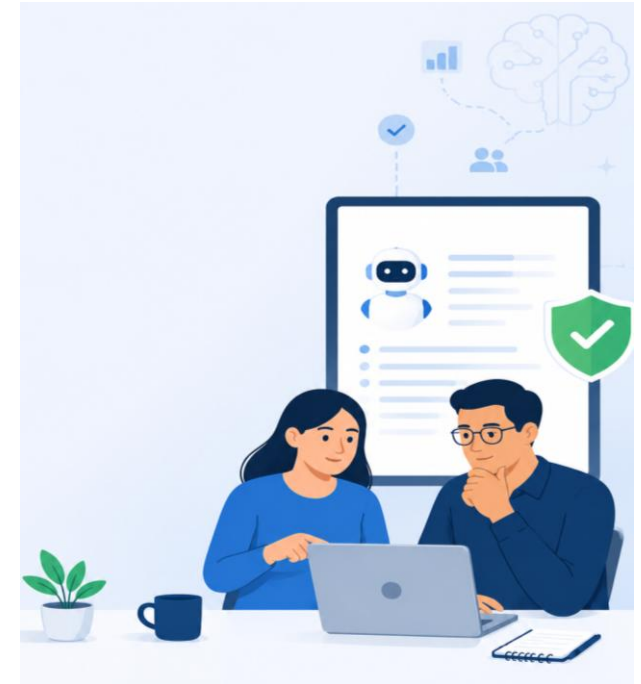
Four-Eyes Principle.

Always have a human expert review AI-generated work.



Human-in-the-Loop.

Keep human judgment at every critical decision point.



AI generates; Evaluators validate.



Future Evaluators Need



Human Judgment

Context, empathy, and ethics



Critical Thinking

Questioning outputs



AI Literacy

Knowing when to use AI



Co-Pilot, Not Autopilot

AI won't replace Evaluators

But Evaluators using AI may replace those who don't.

Technology can accelerate evidence.
Trust, empathy, judgment, and accountability remain profoundly
human.



AI: Better Evidence. Better Decisions. Better Lives.

For Evaluators

More time listening, less time on repetitive tasks.



For Communities

Evidence that truly reflects their realities and needs.



For Decision-Makers

Faster, higher-quality insights to act on with confidence.



Trestle

Bridging technology gap

Maulik Chauhan: Director, Trestle Research & Consulting | India Lead, SurveyCTO
maulik@trestle.co.in



Annexure: Some Practical AI Tools at Work



Perplexity

Live web research with citations



Claude

Writing, reasoning & document analysis



Julius

Data analysis & statistics



AI PPT

Auto-generated slide decks



Briks

Dashboards & visual reports



Google Flow & Gemini

AI-generated videos for advocacy



Plot.ly & Gemini

Interactive shareable dashboards



Grok

Real-time fact-checking & news validation



Claude for Coding

Code for pipelines & automation



ChatGPT & Gemini

Everyday drafting & research



Prezi, Gamma, & Presentation

Dynamic slide decks from outlines





global
evaluation
initiative

Thank you!