

Exhibition Introduction

EvalBlues

Everyone
wants evidence.
But not everyone
asks how the
evidence was
made.

WELCOME TO THE EVIDENCE JOURNEY

Think closely.
Question wisely.
Build trust.

Evidence

#EvalBlues

Phainom

A young evaluator
navigating the realities
of fieldwork, data
quality, AI-assisted
analysis, and
evidence use.

1



When Evidence Fails

When good intentions
meet messy data,
and the story
falls apart.

2



The Illusion of Intelligence

When AI sounds
confident, but
certainty is not
the same as truth.

3



Making Sense of Evidence

When human judgment,
context, and curiosity
turn data into
meaning.

4



Evidence that Builds Trust

When trustworthy
evidence informs
decisions and
drives real impact.

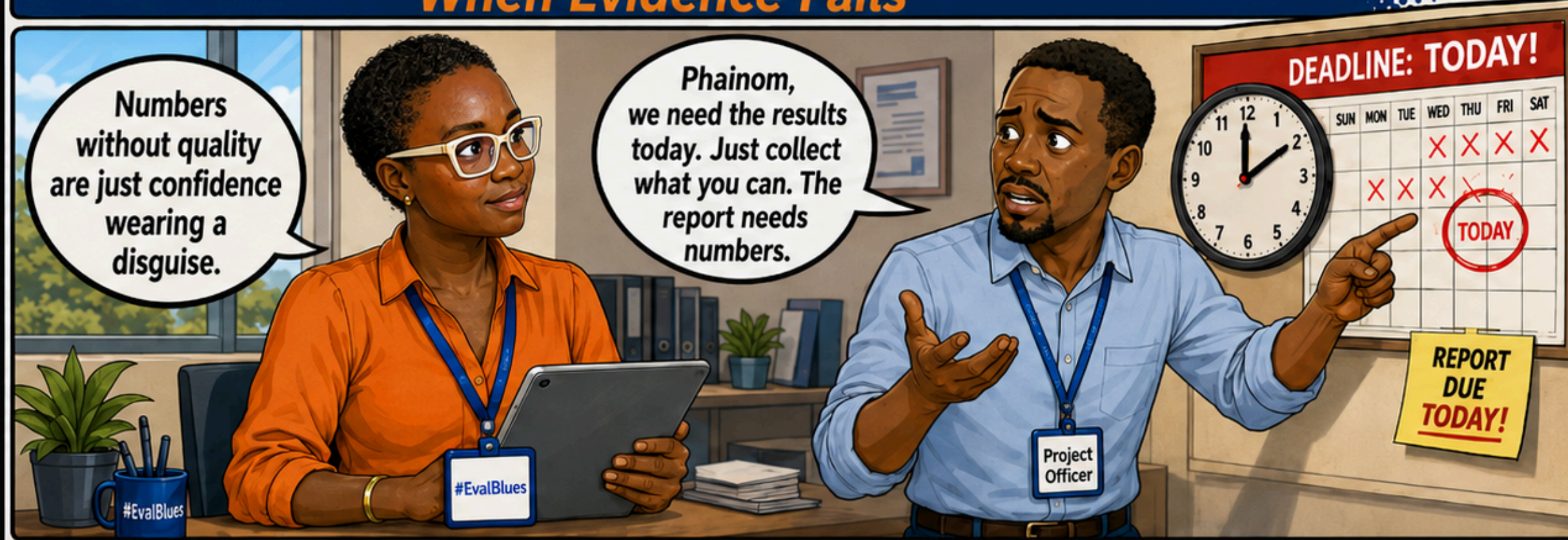
In the age of AI, evidence can move faster than ever.
Data can be cleaned, summarized, visualized, and interpreted within minutes.
But speed does not automatically create truth,
and technology does not automatically create trust.

Through **EvalBlues**, follow Phainom across four moments in the evidence journey:
when evidence fails, when AI sounds too certain, when human judgment brings
meaning, and when trustworthy evidence leads to real impact.

The Rushed Evidence

When Evidence Fails

EvalBlues



The process is rushed. Enumerators are not properly briefed. Survey questions are unclear. Some groups are underrepresented. Data is collected, but the evidence is already weak before analysis begins.



Poor evidence can come from:



- Weak sampling
- Unclear questions
- Poor training
- Missing voices
- Rushed analysis

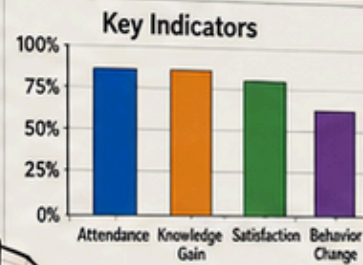
The Beautiful Mistake

When Evidence Fails

EvalBlues

The dashboard is beautiful, but the data behind it is limping.

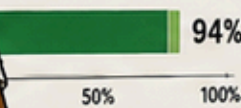
Programme Success: **87%**



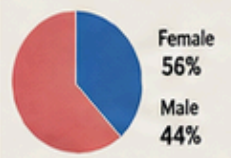
Beneficiaries Reached
12,450



Budget Utilization



Gender Breakdown



Overall Performance: **EXCELLENT** ✓

The dashboard looks excellent. Let's scale the intervention.

EVIDENCE
DRIVES
DECISIONS

DATA SOURCE: ENDLINE SURVEY

Indicator	Value	Source
Attendance	90%	Survey
Knowledge Gain	85%	Survey
Satisfaction	92%	Survey
Behavior Change	70%	Survey

⚠ small sample

⚠ missing rural data

⚠ incomplete responses

⚠ leading questions

Poor-quality data can still look convincing when presented well. But attractive visuals cannot correct weak methods. When evidence fails, resources are wasted, communities lose trust, and decisions may cause more harm than good.

THE COMMUNITY LEFT BEHIND

This is not what we said we needed.

KEY MESSAGE



Evidence is only as strong as the process that produces it.



The Magic Output

The Illusion of Intelligence

| EvalBlues

AI does not equal **Truth**

Across all groups? The dataset barely spoke to all groups.

✓ PROGRAMME SUCCESS

Programme success rate:

94%.

Beneficiaries are satisfied.

Satisfaction Rate

94%

Performance by Group



Analysis complete.
Programme performance is strong across all groups.

High Impact

Inclusive Results

Reliable Insights

rural respondents excluded

women underrepresented

language barrier

conflict-affected areas not covered

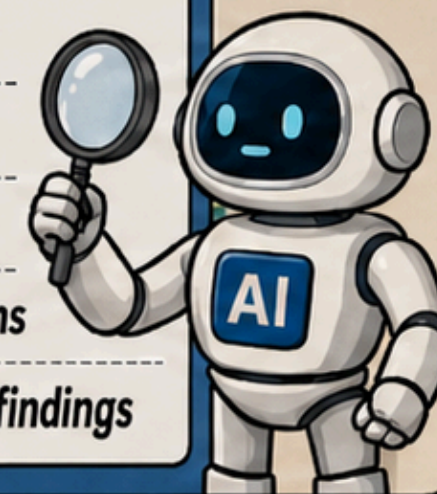
FIELD NOTES

- Listen widely.
- Ask deeper.
- Represent all.
- Decide better.

AI can process information quickly and generate polished outputs. But speed is not the same as truth. If the data is incomplete, biased, or poorly framed, the AI output may also be incomplete, biased, or misleading.

AI CAN ALSO REPRODUCE:

- ! Data bias
- ! Missing context
- ! False certainty
- ! Weak assumptions
- ! Overgeneralized findings



QUESTION THE OUTPUT. **IMPROVE THE INPUT.** STRENGTHEN THE IMPACT.



Polished does not mean truthful. Evidence matters.

Spot the Bias

The Illusion of Intelligence

EvalBlues

Now it's
your turn.
What is the
AI missing?



AI-GENERATED OUTPUT

**Beneficiaries are satisfied
because 80% of survey
respondents gave positive
feedback.**



Phainom's Field Note

70% of respondents were urban participants. Rural participants had limited access to the survey. Women's responses were underrepresented.



Identify:

- 1 What assumption is being made?
- 2 Who is missing from the data?
- 3 What bias might affect the conclusion?
- 4 What should an evaluator ask before trusting this output?

KEY MESSAGE



AI can support evaluation,
but it cannot replace
critical human judgment.



QUESTION THE OUTPUT. **IMPROVE THE INPUT.** STRENGTHEN THE IMPACT.

The Evidence Room

Making Sense of Evidence

EvalBlues



Key Message

AI can help evaluators organize large volumes of information, identify repeated patterns, summarize responses, and speed up analysis. But the evaluator must verify the output, compare it with field realities, and interpret findings with care.



AI helps with pattern recognition.



Human judgment helps with meaning, ethics, and context.

From Pattern to Meaning

Making Sense of Evidence

EvalBlues



AI may identify what is repeated. Human expertise helps explain why it is happening, what it means, and what should be done next. Together, AI and human judgment can produce more accurate, context-sensitive, and useful evidence.

Key Message

Credible evidence emerges from thoughtful interpretation, not automation alone.

Evidence Returns to the People

Evidence that Builds Trust

EvalBlues



Trustworthy evidence is not only technically sound. It is ethical, inclusive, transparent, and useful to the people it represents. When communities see how their voices shaped the findings, trust begins to grow.



Trustworthy evidence requires:

- Ethical data collection
- Strong methods
- Inclusion
- Transparency
- Clear communication

Evidence Becomes Action

We changed the approach because the evidence showed us what was not working.

PROGRAMME PLAN

BEFORE	AFTER (UPDATED)
• Meetings at 10am • No feedback channels • Limited data tools • Some groups not included	✓ Meetings at times that work for people ✓ Feedback channels introduced ✓ Improved data tools ✓ Underserved groups included

Now the programme fits our reality better.

That is when evidence becomes more than a report. It becomes a bridge.

Phainom's EvalBlues
FINDINGS BRIEF

Notes
• Listen
• Learn
• Adapt
• Improve

INCLUSIVE PRACTICE
DATA QUALITY
ETHICAL STANDARDS

COMMUNITY REPRESENTATIVE

EvalBlues

KEY MESSAGE:
Trustworthy evidence drives meaningful and sustainable change.

THE CHAIN OF TRUST

Good process → Good data → Good interpretation → Good communication → Good decisions → **TRUST**

Evidence is not just what we present. It is what we protect, question, interpret, and use responsibly.

FOUR LESSONS FROM EVALBLUES

- 1. BAD DATA LEADS TO BAD DECISIONS.**
Illustration: A trash can full of crumpled paper with a red warning triangle.
- 2. AI DOES NOT EQUAL TRUTH.**
Illustration: A robot labeled 'AI' looking at a magnifying glass over a document labeled 'NOT TRUTH'.
- 3. HUMAN JUDGMENT GIVES EVIDENCE MEANING.**
Illustration: A woman thinking, with a thought bubble showing people and a document.
- 4. TRUSTWORTHY EVIDENCE DRIVES IMPACT.**
Illustration: Three people high-fiving under a bright sun.

FINAL EXHIBITION MESSAGE:
GOOD EVIDENCE NOT ONLY INFORMS DECISIONS. IT EARNS TRUST.