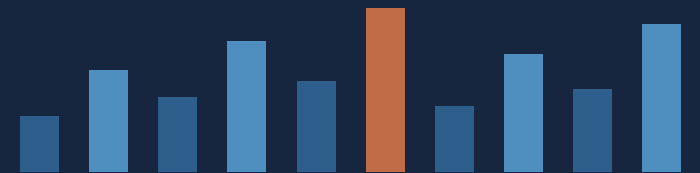


FEMINIST EVALUATION IN THE AGE OF AI

gLOCAL Evaluation Week · 2 June 2026

Infrastructures of Recognisability

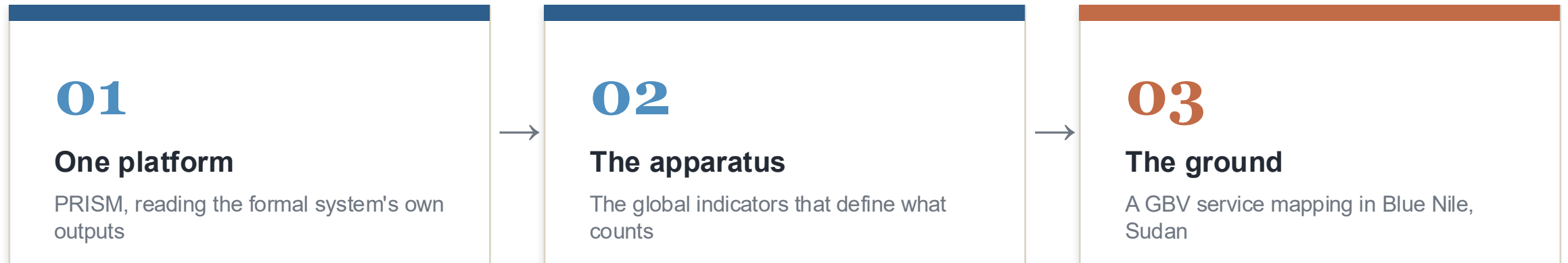
What humanitarian data sees, and what it misses: gender measurement, AI, and a GBV service mapping in Sudan.



Umutcan Yüksel Policy Leader Fellow, European University Institute

THE QUESTION

Can the datasets we use to measure gender equality recognise the women they are meant to count?



Three layers of measurement, read against each other.

PRISM: reading the system against itself

A humanitarian analytics platform that integrates the formal system's own outputs: needs overviews, cluster data, funding flows, risk and severity indices.

Open and free. Around 30 live datasets across 230+ countries, with AI-generated country situation summaries grounded in a long-term preservation archive of source documents.

IN THE AGE OF AI

Large language models read what the system publishes and summarise it. PRISM does not see past the apparatus; it reads the apparatus. By design it inherits the system's categories, and its blind spots.

~30

live datasets

230+

countries

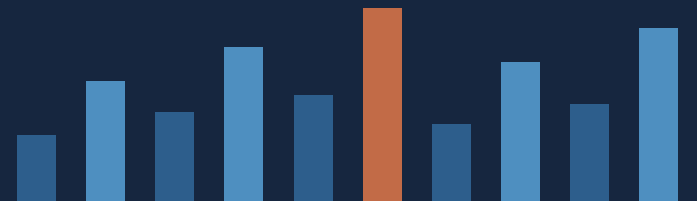
100%

sourced from the formal system

THE ARGUMENT

- The authoritative datasets we use to measure gender equality are infrastructures of recognisability.
- They make some realities legible, and systematically miss what does not fit their categories.

Legibility decides what gets counted, what gets funded, and who is seen.



How 'gender' enters the data

■ **Share of population female**

World Bank

■ **Maternal mortality ratio**

INFORM Risk

■ **Child marriage, girls before 18**

UNICEF

■ **Skilled birth attendance**

UNICEF

■ **GBV as a protection-risk exposure**

Global Protection Cluster

■ **GBV as a funding line**

OCHA FTS / HRP

■ **Funding to women-led organisations**

Localisation, GB 4.1

Seven indicators. Each one a single national number. Women enter the data as a ratio, a maternal risk, a marriage statistic, a funding code.

The instruments that define what counts

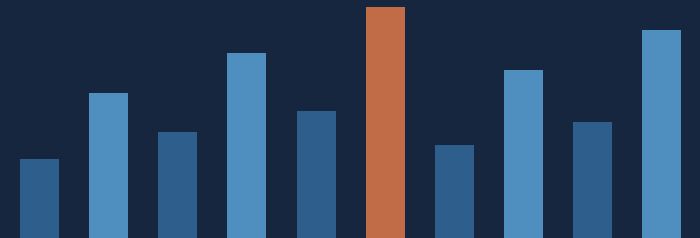
GBVIMS Six core GBV incident types	GBV AoR Five-pillar service model	4Ws Who does what, where, when	HNO / HRP People in need + severity
IASC Gender + Age Marker Projects scored 0 to 4	OECD DAC Gender Marker Aid scored 0 / 1 / 2	SDG 5 + GII Global gender targets	But a category is not a service.

Standardised, comparable, authoritative. They define what counts as GBV, and therefore what gets funded.

LAYER 3 · THE GROUND

A GBV service mapping in Blue Nile State, Sudan

Commissioned by an INGO; 19 key-informant interviews, 5 focus groups, 20 actors mapped, validated with local stakeholders. 2026.



ON PAPER, FULLY LEGIBLE

Precise in the data, absent on the ground

WHAT THE SYSTEM MEASURES

Level 4 / 8.6 of 10

GBV severity (HNO 2026 · national severity index)

23,557 · 12.07M

GBV people in need (locality · Sudan, 2026 HRP)

17.3%

of the 2026 response plan funded

WHAT EXISTS ON THE GROUND

11 (5 women)

GBV-dedicated staff — 1 per 2,142 people in need

0 / 0 / 0

safe shelters / legal aid / specialised mental health

none

formal referral pathway · GBVIMS · GBV coordination

The system measures this place to the decimal. The services its categories presuppose do not exist.

National figures via PRISM (Sudan, 2026). Locality figures: service mapping.

WHAT THE CATEGORIES CANNOT SEE

The infrastructure women actually navigate

First contact is community and women's leaders, not formal services (14/19 and 10/19).

Early marriage is counted as prevalence; the bride-price economy and marriage to settle rape are not.

Survivors rely on family, faith leaders, herbal medicine, and ad hoc cross-border journeys.

The community's own priorities, survivor suicides, HIV, childbirth deaths, are not standard GBV indicators.

Community protection networks and women's groups stay invisible to the 4Ws and GBVIMS.

The referral itself can create a new assault: rape on the route cited by 11/19.

Help-seeking runs through networks the apparatus does not count, cannot fund, and renders invisible.

Legibility is not recognition

1 Measurement routes money to categories, not to networks.

A severity score is fundable. A women's protection network the data cannot see is not.

2 PRISM reproduces this.

Reading the system's outputs means inheriting the system's blind spots.

3 AI scales the trap.

A model that summarises the apparatus launders its categories as neutral knowledge, faster and with more authority. The danger is not biased data; it is invisibility made to look objective.

What this asks of us

- 1 Evaluate the instruments, not just the programme.** Ask who each indicator makes recognisable, and who drops out.
- 2 Treat absence as data.** A zero is a question, not a finding.
- 3 Centre the informal.** Co-define indicators with the women who navigate the system.
- 4 Use AI to widen recognition, not automate the frame.** Interrogate what the model cannot see before trusting what it surfaces.
- 5 Hold the tools accountable, PRISM included.** Judge a platform by what it excludes, not only by what it shows.

“An evaluation that only sees what the system already counts will keep missing the women who never appear in it.”

Umutcan Yüksel · European University Institute · prismonitor.eu

Feminist Evaluation in the Age of AI · gLOCAL Evaluation Week · 2 June 2026

