

Data are a girls' best friend

Understanding and Addressing
Digital Inequalities

Centre for a Digital Society, EUI

Firenze, 20 November 2025

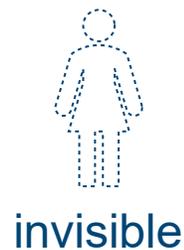


Who counts in digital finance?



Digital finance grows fast but the data behind it remain incomplete

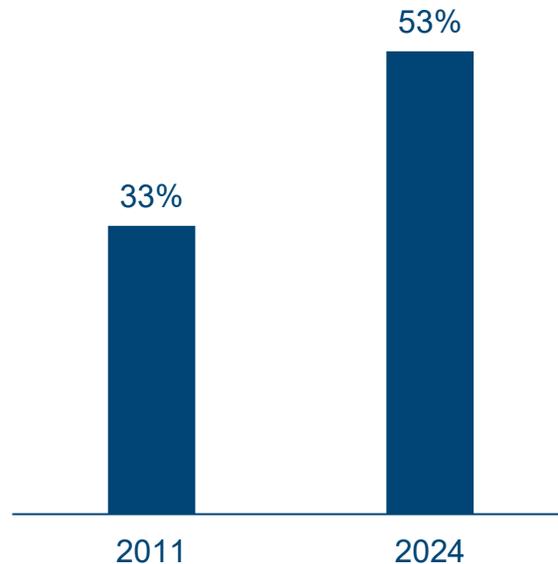
Data inequality = structural underrepresentation or misrepresentation of specific groups in the datasets that guide financial decisions



The power of being counted

Account ownership in MENA

Adults with an account (%)



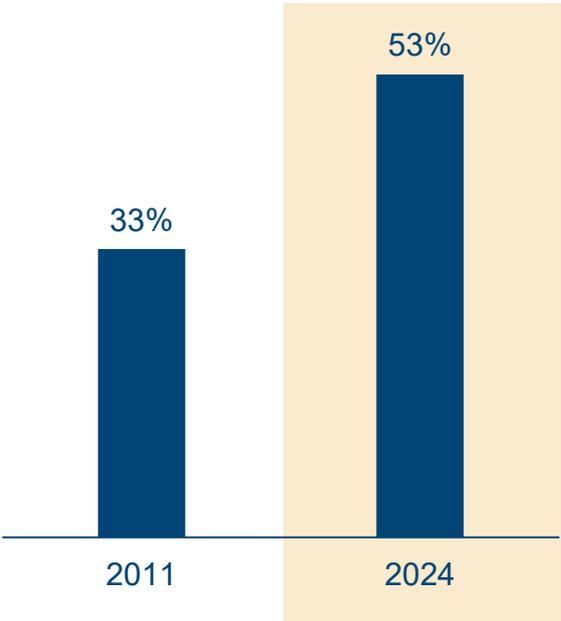
Source: Global Findex Database 2025

Main message:

Financial inclusion in MENA has grown rapidly between 2011 and 2024

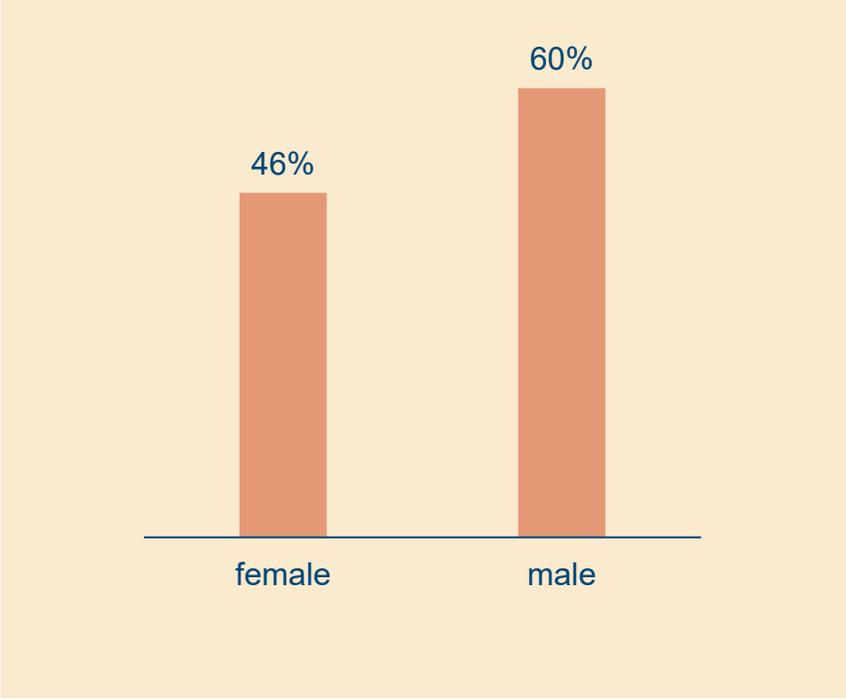
The power of being counted

Account ownership in MENA
Adults with an account (%)



Source: Global Findex Database 2025

Account ownership in MENA in 2024
Adults with an account (%)



When applying the gender lens, the message changes: Financial inclusion in MENA has grown rapidly between 2011 and 2024, yet women are still lagging behind (12pp)

“Who gets counted, counts.”

D’Ignazio & Klein (2020)

Why this matters?

Integrated data = better diagnostics,
stronger governance, fairer outcomes



Diagnose

Reveal gaps in access, usage, and outcomes



Target

Design policies and products that respond to real user needs



Monitor

Track progress and trends in inclusion across regions/institutions



Govern

Embed gender visibility into regulatory mandates and supervision



Correct

Detect and reduce systemic and algorithmic bias



What are sex-disaggregated data in banking and finance

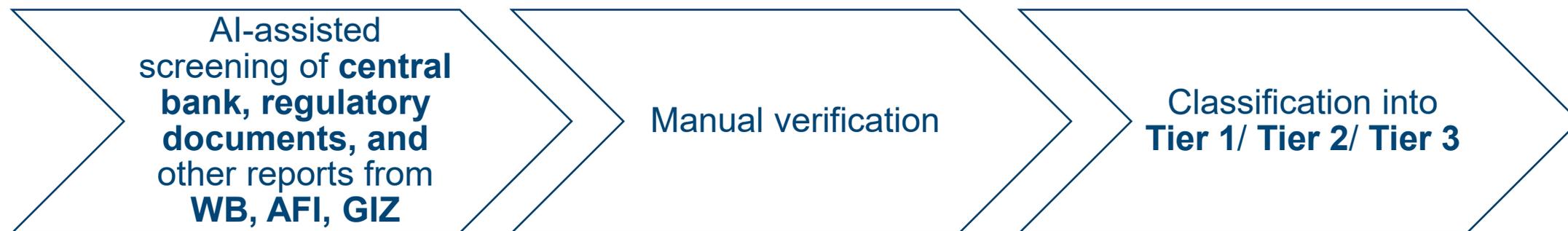
**Supply-
side SDD**

Administrative data from providers

**Demand-
side SDD**

Survey data from users

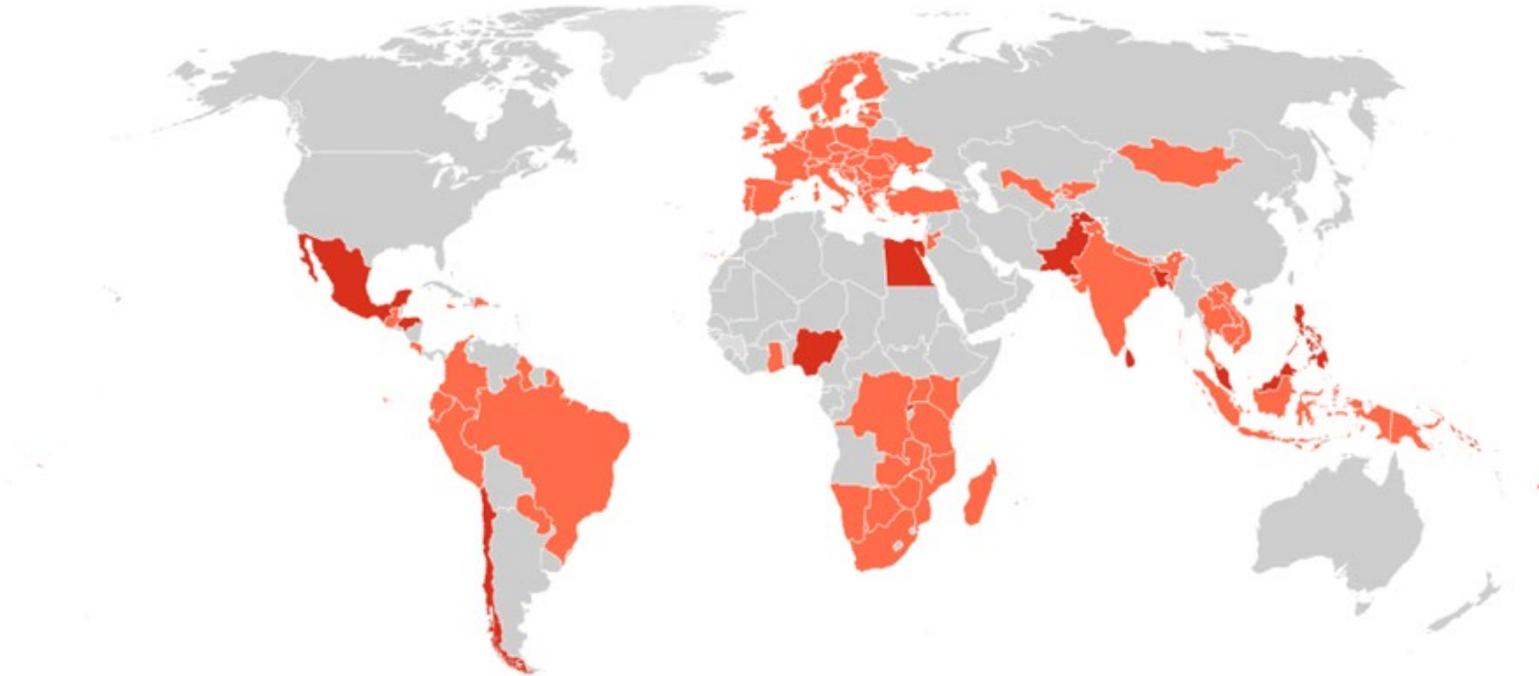
Attempting the (almost) impossible: mapping global sex-disaggregated data with a human–AI approach



Combines **scale of AI + judgment of human validation**

Global supply side data map

■ Mandated & Public ■ Emerging/Partial/Voluntary ■ None/Unknown



Tier 1 (Mandated): Regulators require financial institutions to report sex-disaggregated data

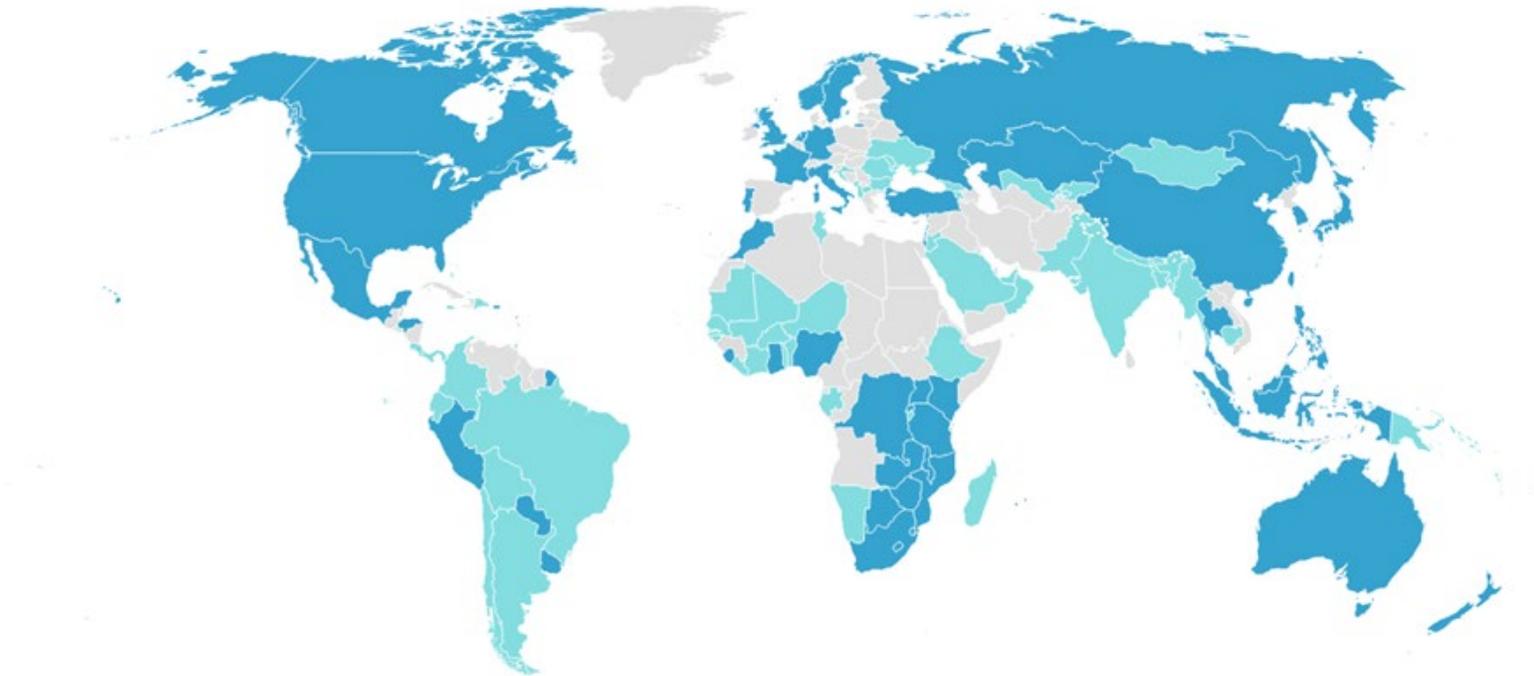
Tier 2 (Partial): Some reporting exists, but not systematic or fully mandated

Tier 3 (None): No evidence of regulatory requirements for sex-disaggregated reporting

Source: Author's elaboration based on publicly available reports • Created with Datawrapper

Global demand side data map

■ Institutionalised & recurring survey ■ One-off/recent national initiative ■ Findex-only/no national survey



Tier 1 (Recurring): National surveys consistently collect sex-disaggregated financial data

Tier 2 (Occasional): Surveys exist but irregular or donor-dependent

Tier 3 (None): No national surveys capturing sex-disaggregated financial indicators

Source: Author's elaboration based on publicly available reports • Created with Datawrapper

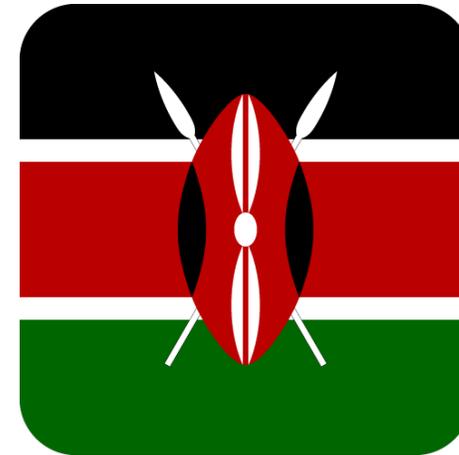
From data to inclusion: three institutional models



Chile
**Supervisory
mandates**



Mexico
**Hybrid
integration
model**



Kenya
**Emerging data
governance**

Chile: A mandated, mature system



Mandate

Long-standing mandatory sex-disaggregated reporting (SBIF/CMF) across the banking system



Coordination

Centralised under CMF; stable supervisory integration



Digital infrastructure

Granular, high-frequency supervisory dashboards; strong open-finance architecture



Transparency

Regular public reports and dashboards (*Género en el Sistema Financiero*)



Impact

Strong supply-side visibility enabled programmes like CreceMujer and sector-wide monitoring, but behavioural drivers remain less explored

Mexico: A hybrid, coordinated system



Mandate

Mandatory reporting since 2016 + national ENIF surveys + gender diagnostics



Coordination

Cross-agency alignment CNBV–INEGI–INMUJERES: gender data embedded in national policy (PROIGUALDAD, NFIS)



Digital infrastructure

Diagnostics + surveys that result into actionable insights on regional gaps, behaviours, and needs



Transparency

Regular ENIF + “Diagnóstico” that serve as evidence base for targeted reforms (e.g., Mujer PyME; provisioning adjustment)



Impact

One of the few countries where sex-disaggregated data directly shaped **prudential rules** and **credit programmes**

Kenya: An innovation-led, weakly standardised system



Mandate

No requirement for sex-disaggregated reporting from financial institutions; reliance on FinAccess surveys



Coordination

Fragmented across regulators (CBK, SACCO oversight) and private digital ecosystem



Digital infrastructure

Very rich private digital-finance data (mobile money, digital lenders) but siloed and inconsistent



Transparency

Low public reporting; FinAccess only every 3 years: **long blind spots**, especially in fast-evolving digital lending



Impact

High adoption of digital finance but **low capacity to monitor gendered risks**, reinforcing reliance on market-led data

What Chile, Mexico, and Kenya show us about sex-disaggregated data

Visibility is foundational...

Across all cases, sex-disaggregated data shape inclusion, but incentives to collect, standardise, and share them remain uneven.

..but not enough. Persistent bottlenecks:

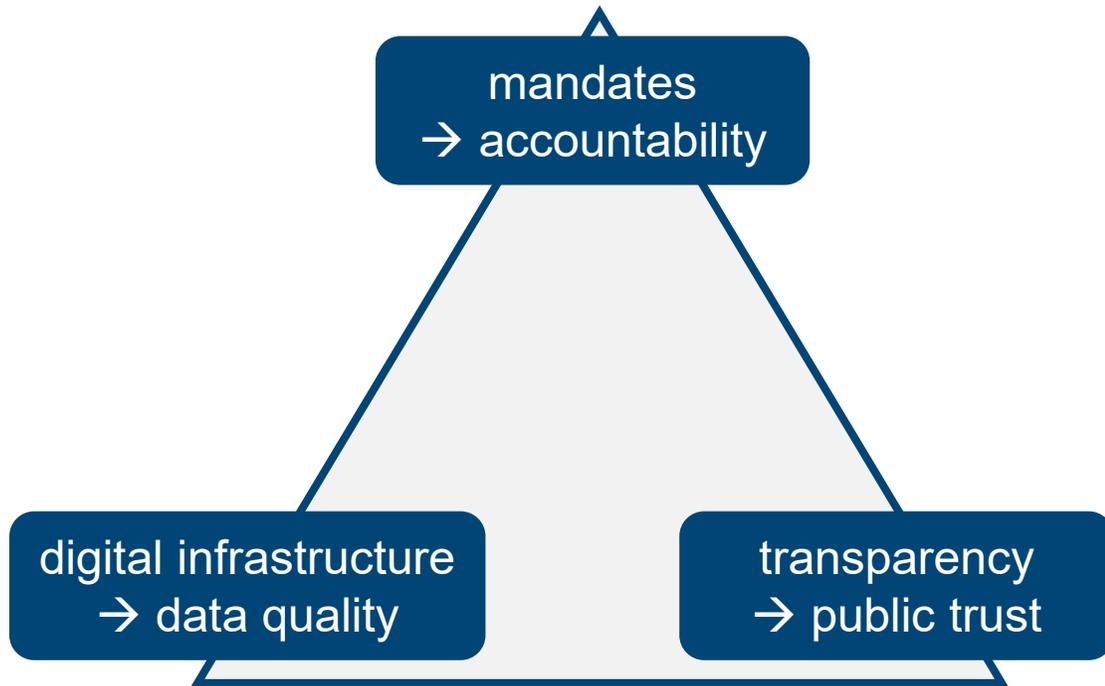
1. Fragmented institutions
2. Uneven capacity & infrastructure
3. Weak or voluntary mandates

What's needed

Sustained visibility + institutional continuity + binding mandates
→ transform data into lasting, equitable policy action



The missing architecture



Visibility becomes governance

The Self-Reinforcing Model

Mandates → generate demand for data;
Infrastructure → ensures data quality;
Transparency → converts data into action

The Goal

Make sex-disaggregated data a core component of financial governance – informing supervision, digital finance strategies, and budget decisions – so women’s experience systematically shapes financial policy.

What comes next

Algorithmic fairness: the emerging frontier

- The EU AI Act classifies credit scoring as **high-risk**, requiring bias checks, human oversight, and transparency
- Sex data can be used under safeguards to detect and prevent discrimination (art.10(5))
- Algorithmic audits must become routine to ensure digital finance advances inclusion rather than reinforcing gaps

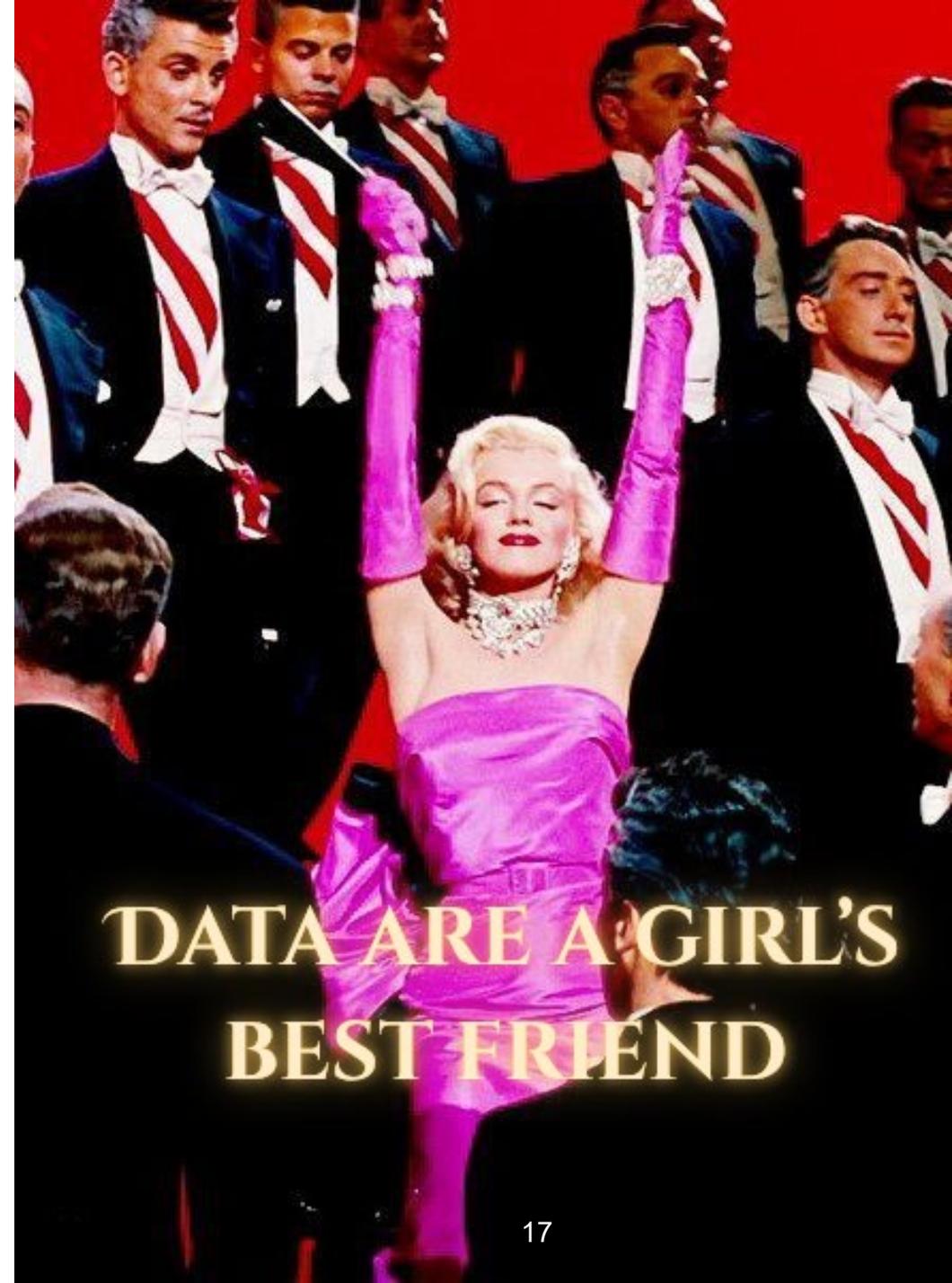
Revisiting the European paradox

- High-income economies have advanced infrastructure but **gender-blind supervision**
- AnaCredit, COREP/FINREP lack sex-disaggregated fields → supervisors can't detect gendered risk patterns
- Integrating gender metrics into **prudential reporting, stress tests, and risk analysis** would align Europe's regulatory leadership with its equality goals.



**Data are a girl's best friends –
not metaphorically but institutionally**

**When embedded in financial
governance, they turn recognition
into representation and information
into inclusion**



Thank you!



Annex



Terminology: sex and gender

Sex refers to «the biological and physiological characteristics that define humans as female, male or intersex.» (WHO) «However, many data-collection tools still ask the question ‘What is your sex?’ or ‘What is your gender?’ each of them asking seemingly for the same set of binary variables: female/male.» (PARIS21/UN Women, 2025)

Gender identity refers to «A person’s deeply felt internal and individual experience of gender, which may or may not correspond with the sex assigned at birth.» (OHCHR, 2019)

Data: details on supply-side and demand side

	SUPPLY-SIDE DATA	DEMAND-SIDE DATA
Source	Financial service providers (FSPs), central banks, and supervisory authorities	Users (individuals, households, enterprises) via surveys, focus groups, interviews
Frequency	Ongoing and frequent reporting. Data can be captured in near real-time or at regular intervals.	Periodic/episodic (one off or multi-year waves)
Insight type	Quantitative what data on transactions and accounts. Captures volumes (e.g. number/value of loans, deposits, accounts, clients, transactions, credit limits.) and usage of formal products	Mixed quantitative/qualitative why and how insights. Surveys collect usage patterns (access, account ownership, savings, credit uptake) and attitudes (perceptions of value, risk, barriers). Include socio-demographic context and informal finance.
Structure	Highly structured and standardized. Possible comparability across reports. Often requires unique IDs for matching accounts.	Variable structure. Survey data have a structured questionnaire component (quantitative indicators) and may include open-ended or qualitative sections. Less standardized and harder to harmonize across studies.
Cost/capacity	Lower cost after initial setup; integrated into IT & supervisory systems	High cost and effort. Designing and conducting representative surveys require substantial time, funding and technical expertise.
Strengths	Timely, comparable, scalable for supervision	Rich, user-level details, contextual insights, flexibility
Limitations	Coverage gaps, demographic detail, implementation hurdles	Infrequency and comparability sampling biases, resource intensive

Source: author's elaboration from Alonso et al. (2024), Segoviano et al. (2025), AFI (2017)



Methodology: Why AI supported the creation of the databases

The scope of the databases: no sex-disaggregated quantitative indicators included in the databases; AI as a support tool for **classification and documentation**.

Purpose of AI assistance: to process large volumes of regulatory and statistical documentation across 190+ countries, identify comparable fields, and generate a coherent classification framework.

How I checked AI entries in the database

- **Ex-ante check:** ensured conceptual clarity on supply-side vs demand-side data, standardising definitions and classification in Tier 1/2/3 before building the comparative framework.
- Building the database at the country level and grouping countries by region, facilitating systematic human verification
- **Ex-post check:** The variable “source” the first to be checked
 - The row was not filled in → human check
 - The source was repeated in multiple entries across the region → potential inconsistency (“propagation error”) → human check
- Introduced a reference variable (year of first implementation) to make human consistency checks faster and more reliable

How I checked AI entries in the database

- **Tier 1 category the least frequent**, making full manual verification feasible
- **Tier 3 category is the most difficult to assess**, as it corresponds to “no collection of sex-disaggregated data” - a situation that is rarely documented explicitly. It therefore functions as a **residual category**.
- **Major risk after AI-human workflow**: the possibility of **underestimating** existing sex-disaggregated data and over-classifying countries as Tier 3.
- **Additional risk**: source obsolescence in the input (i.e., not the most up-to-date document). This was addressed through systematic manual validation.
- **Additional support for manual verification** came from international organisations’ reports that assess the status of sex-disaggregated data across multiple countries.



Chile - A mature supply-side system: mandatory, granular, high-frequency data

Long-standing mandates: Two decades of compulsory sex-disaggregated reporting via unique ID → complete, granular data

What the data revealed

Women = majority of clients but <40% of microenterprise borrowers

Policy response: BancoEstado expanded *CreceMujer*: tailored credit + digital/financial support

Outcomes: +33% women-entrepreneur portfolio; women show lower delinquency and write-off rates

Remaining gaps: Savings and insurance still lag: women = 57% of depositors but ~50% of balances.

What's next: Fintech Law 21.521 embeds gender indicators into open-finance APIs → real-time monitoring and more inclusive products.

Mexico - A coordinated hybrid system: integrated diagnostics and institutional alignment

Dual system: surveys + reporting: ENIF (since 2012) + mandatory sex-disaggregated reporting (since 2016) create a combined view of access, usage, and behaviour

What the data revealed: Women's formal entry often policy-driven; persistent gaps in savings (59% vs 68%), credit (36% vs 39%), and insurance (18% vs 28%)

Supply-side insights: Women dominate development banks/microfinance but have smaller balances and loan sizes; women show lower NPLs (2.8% vs 3%)

Policy response: targeted programmes like NAFIN's *Mujer PyME* (+11,900 loans; <1% default). **CNBV reduced provisioning requirements for women's loans (2021)**

Institutional alignment: National Financial Inclusion Policy + PROIGUALDAD enable coordination across CNBV, INMUJERES, development banks; geographic dashboards guide targeting.

Takeaway: Mexico shows how combining administrative and survey data produces strong gender visibility - though sustainability depends on continued ENIF funding and regular updates.

Supervisory provisioning in Mexico

Sex-disaggregated credit data revealed consistently lower default rates among women borrowers

Building on this evidence, the CNBV adjusted prudential rules in 2021: Reduced loan-loss provisioning requirements on personal loans to women by 4 percentage points

Published in DOF, 23 July 2021 (*CNBV, 2021; DOF, 2021*)

Early evaluations show:

- Increased formal borrowing by women
- Slight improvement in portfolio quality (*Becerra-Ornelas et al., 2024*)

Kenya - A dynamic market-led system: high digital adoption, low regulatory visibility

Market-led progress: Mobile money drove near-parity in access

What enabled it: M-Pesa + products like M-Shwari and Fuliza provide savings, credit, and transaction services → visible digital histories for women

Data gap: Providers record sex, but regulators (CBK, SASRA) do not mandate sex-disaggregated reporting. Reliance on periodic surveys (FinAccess) → limited visibility

Private-sector evidence: Women borrow smaller amounts but have stronger repayment; fewer negative listings (42%) despite lower average loan size

Institutional response: CBK developing an electronic data warehouse to enable continuous supervisory data collection

Takeaway: Innovation expanded access, but without governance women's financial behaviour remains under-represented. Next step: embed sex-disaggregated data into formal supervision