

VILLA VIGONI · INTERDISCIPLINARY RESEARCH CONFERENCE

From Numbers to Narratives

A Methodological Framework for LLM-Assisted Narrative Auditing
in European Social Monitoring

HB

Hassan Bashiri

Department of Computer
Science
Hamedan University of
Technology, Iran

PA

Petra Ahrweiler

TISSS Lab, Institute of Sociology
Johannes Gutenberg University
Mainz, Germany

MR

Massimo Rusconi

InsIDE Lab, Department of
Economics
University of Insubria,
Varese, Italy

INDICATORS



CLAIMS



NARRATIVE FRAMES



Data do not speak for themselves

Decision-makers, media and the public never meet raw indicators — they meet reports, policy documents and interpretive narratives.

Statistical evidence



Eurostat

OECD

Destatis

ISTAT

Quality of life · well-being · regional inequality ·
employment · education · health · social capital ·
sustainable development

Institutional lens

- 1 Selecting
- 2 Interpreting
- 3 Highlighting
- 4 Framing

Institutional narrative

“Regional disparities are decreasing.”

Reports

Policy documents

Analyses

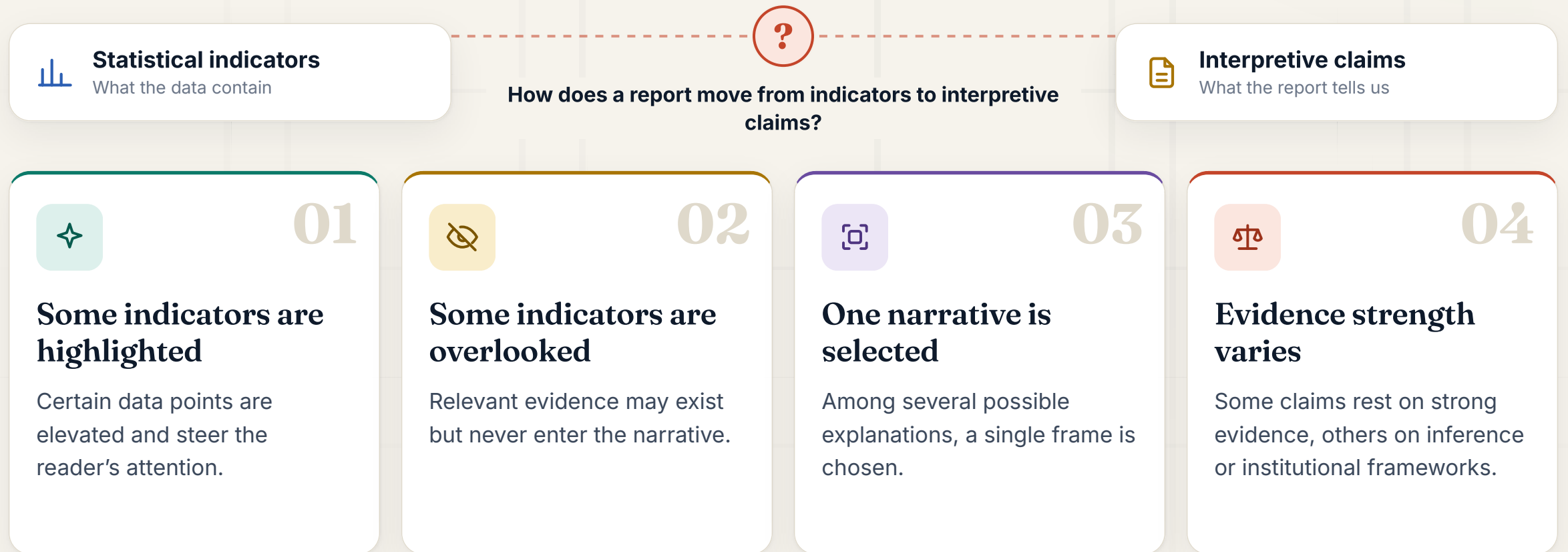
Some indicators are highlighted, others left aside; one explanation is chosen among several.



The “social condition of a region” or the “cause of an inequality” results from the **interaction between statistical evidence and institutional narratives.**

The transparency gap

The question is not whether a narrative is right or wrong — it is whether the path from indicators to claims can be seen.



👁 Readers see the narrative — **not the selection process behind it.**

From narrative analysis to **narrative auditing**

Narrative research is long-established — but rarely asks whether narratives match the evidence used in the same report.

EXISTING APPROACHES

Narrative analysis

“What narrative does this report present?”

- ✓ Describes discourse and interpretive frameworks
- ✓ Shows how social reality is represented through language
- ✓ Pays less attention to consistency with the report's own statistics

A STEP
BEYOND



PROPOSED APPROACH

Narrative auditing

“How is this narrative constructed, and to what extent is it grounded in observable evidence?”

- ✓ Examines the relation of claims, evidence and interpretive frameworks
- ✓ Makes reasoning logic and reliance on evidence explicit
- ✓ Neither political critique nor a true/false verdict

Claims

what is asserted



Evidence

what supports it



Interpretive frames

how it is organised

Four building blocks of the framework



Narrative

A coherent set of statements, interpretations and meaning frameworks that explains relations among data — a structure that gives meaning, not just a list of facts.

Data → Story



Claim

An interpretive statement about the condition, trend, cause or consequence of a social phenomenon. Not true or false in itself — it needs evidentiary support.

Regional disparities are decreasing.

Migration is the main cause of population decline.



Evidence

Empirical information, statistical indicators or documented findings used to support a claim.

Eurostat

Destatis

ISTAT

Regional indicators

Official reports



Narrative frame

The perspective through which data are interpreted. Similar data, different frame — the difference lies in the frame, not in the data.

SAME DATA

"Economic inequality"

"Demographic transformation"

What does a narrative audit do?



A **systematic process** of identifying institutional claims, tracing their supporting evidence, examining omitted evidence, and making explicit the interpretive logic through which statistical indicators become public narratives.



1

Extract the main claims

What does the report actually assert?



2

Identify the evidence cited

What data are used to support each claim?



3

Identify overlooked evidence

What relevant evidence is left out?



4

Analyse the interpretive framework

Which frame organises the story?



Purpose: transparency. The audit increases visibility of reasoning and evidence reliance — it does not judge whether narratives are correct.

A cognitive assistant — not a social analyst

THE LLM IS NOT...

- ✗ A producer of social analysis
- ✗ A content generator
- ✗ A replacement for the researcher

THE LLM HELPS THE RESEARCHER TO...



Extract claims



Classify claims



Link claims to evidence



Identify inconsistencies



Detect omitted evidence



Extract recurring patterns

THE HUMAN RESEARCHER KEEPS...



Social interpretation



Validity of narratives



Theoretical conclusions



Scientific judgment



Not a “decision-making agent” — a **tool for transparency and auditability.**

The LLM assists — the researcher decides

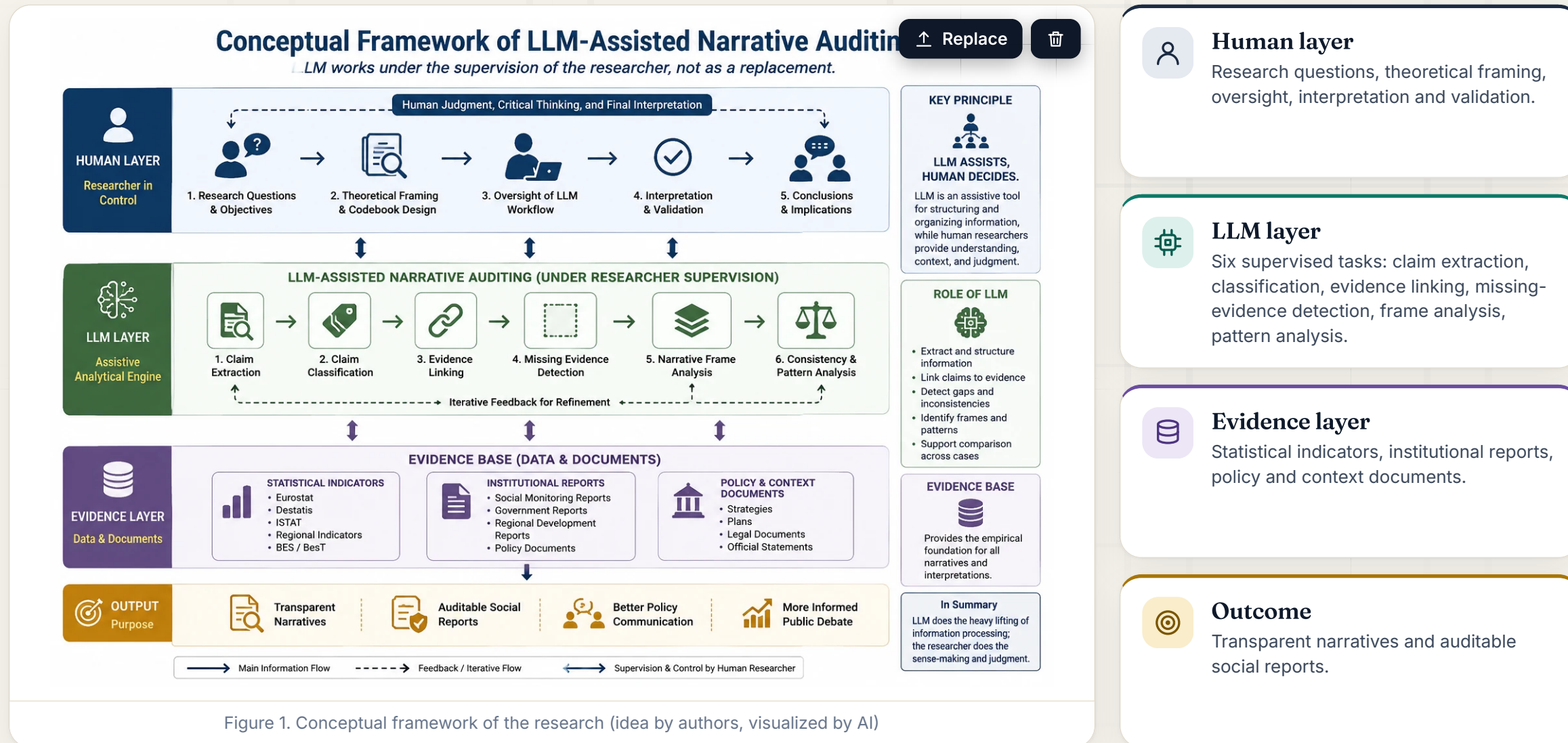


Figure 1. Conceptual framework of the research (idea by authors, visualized by AI)

From human baseline to comparative analysis

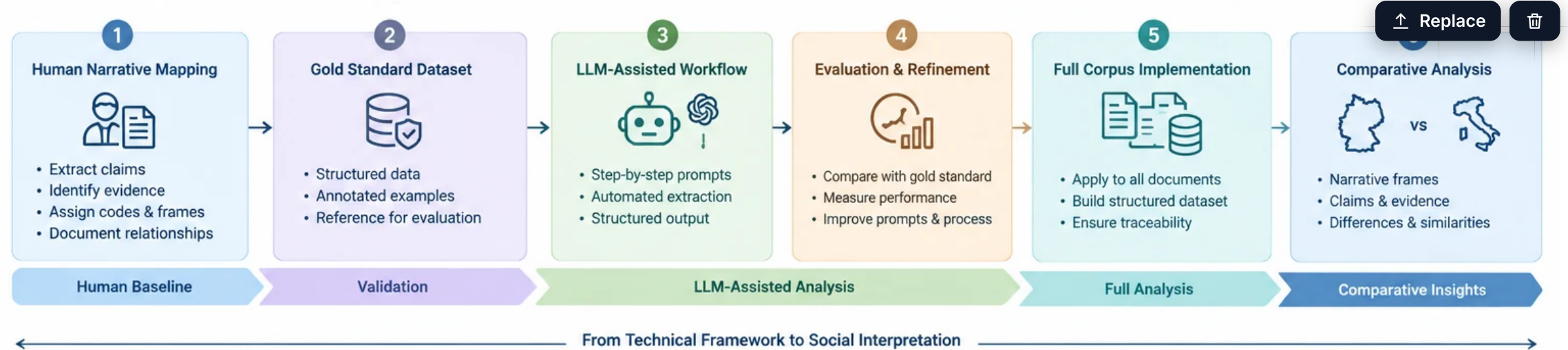


Figure 5. Research process from human baseline to comparative analysis

BUILD

Parts 1 – 3

Conceptual framework · Narrative Codebook · Evidence Corpus

TEST

Parts 4 – 6

Human baseline & Gold Standard · LLM-assisted workflow · Evaluation

INTERPRET

Parts 7 – 8

Comparative narrative analysis · Methodological reflection

A shared language for researchers and LLMs

WHY A CODEBOOK?

R1

“Economic
inequality”

R2

“Demographic
transformation”

Same report, two valid readings — without a shared vocabulary results are neither comparable nor reproducible, and LLM output becomes prompt-dependent.

THREE FUNCTIONS

- 1 Creates a shared conceptual language
- 2 Provides the basis for systematic analysis
- 3 Guides the LLM in extracting and classifying

1

Literature-derived

Concepts grounded in studies of regional inequality, quality of life, development and social monitoring — not personal interpretation.

2

Fit for both cases

Codes must allow comparison although German and Italian contexts differ.

3

Report-independent

A general framework, never designed for one particular report.

4

Human- & LLM-readable

Each code has an operational definition, scope of application and concrete examples.

Free, prompt-dependent analysis



 Codebook = controlled vocabulary



Structured · reproducible · evaluable

10 initial codes — and the anatomy of a code



DEMO Demographic Transformation

Demographic change, ageing, migration



ECON Economic Development

Economic growth, income, productivity



COH Territorial Cohesion

Regional cohesion



REC Recognition

Belonging, identity and recognition



POV Poverty & Deprivation

Poverty and deprivation



LAB Labour Market

Employment, unemployment, labour-market quality



WELL Quality of Life

Well-being, quality of life, health



INST Institutional Capacity

Institutional capacity and governance



SERV Public Services

Access to public services



EDUC Education & Skills

Education and human capital

LAB Labour Market Challenges

KEYWORDS unemployment, employment, labour force

INDICATORS Employment rate, unemployment rate

TYPICAL CLAIM “Regional unemployment remains persistently high.”

EVIDENCE Eurostat employment indicators

+ conceptual & operational definition
+ examples of claims, evidence & use

Two kinds of evidence, one traceable structure



Quantitative evidence

Regional statistical indicators

- ✓ Frequently used in social reports
- ✓ Related to regional inequality
- ✓ Comparable across both countries
- ✓ Linkable to claims in the reports

🎯 **Relevance and comparability** over data volume

METADATA RECORDED PER INDICATOR

Indicator ID Name Definition Source Country Region

Geographic level Unit Time period Data type Source URL

Narrative code



Narrative & policy documents

Where indicators become interpretation



Government reports



Social-monitoring reports



Regional-development reports



Policy documents

SEVEN DOCUMENT-SELECTION CRITERIA

Relevance

Institutional authority

Evidence content

Interpretive content

Temporal relevance

Geographic relevance

Comparability



Quality control before use: duplicates, completeness, OCR, units, definitions, missing values

2

evidence types

20–30

documents per country

40–60

documents in the initial corpus

≈10

years, common period

Two analytical testbeds — not a ranking



Germany

East–West divide

EAST

WEST

FOCUS

Reunification consequences

Demographic change

Living conditions

Employment

Regional development

SOURCES

Eurostat

Destatis

INKAR / BBSR



Similar questions,
different historical &
institutional contexts

COMPARABLE
INDICATORS



Italy

North–South divide · Mezzogiorno

NORTH

SOUTH

FOCUS

Unbalanced development

Labour market

Poverty & deprivation

Institutional capacity

Territorial cohesion

SOURCES

Eurostat

ISTAT · BES/BesT

Territorial cohesion indicators

From a claim to an actual indicator value

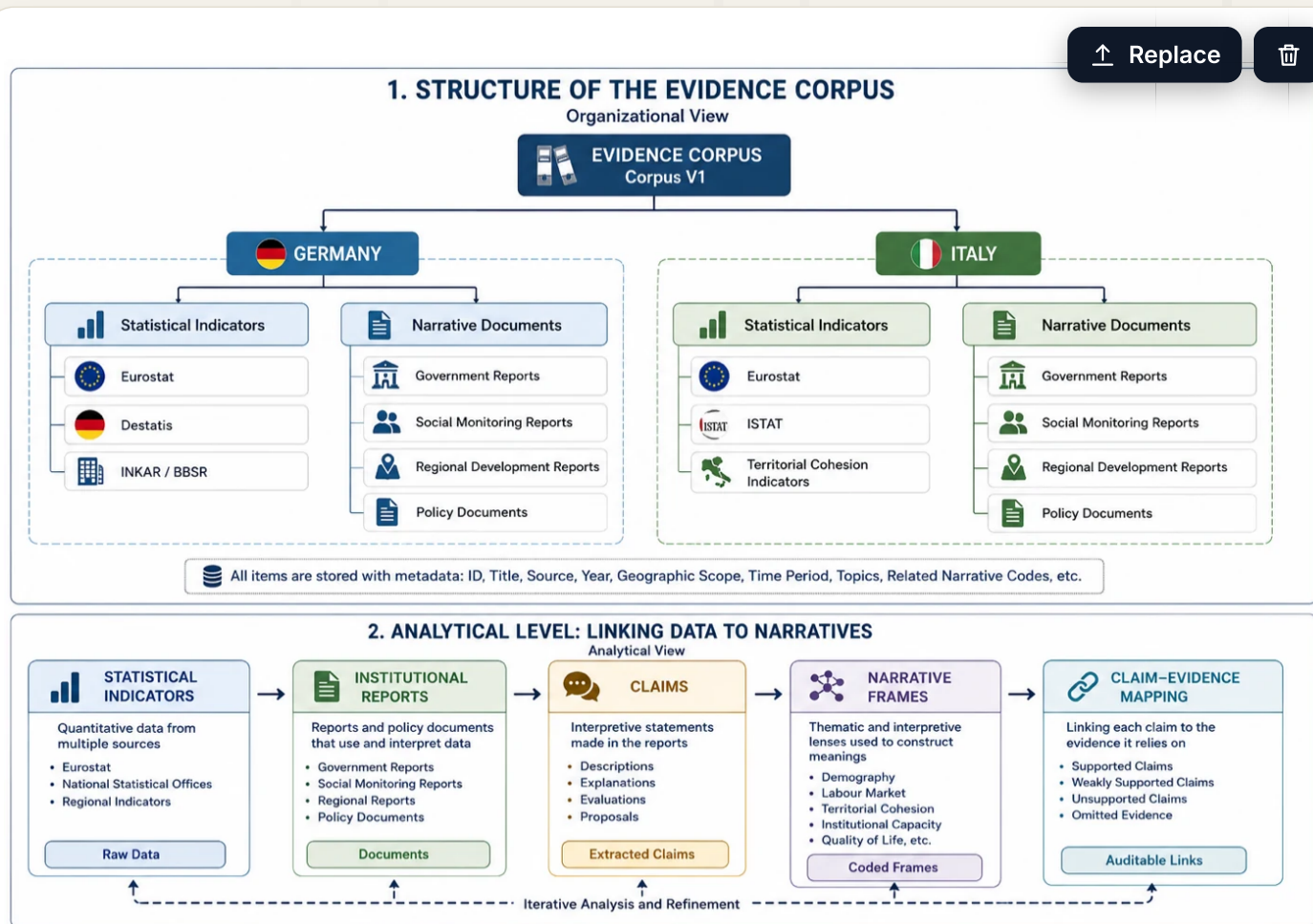


Figure 3. Corpus structure (idea by authors, visualized by AI)

“Unemployment rates in the eastern regions remain higher than in the western regions.”

CLAIM Labour-market claim

CODE LAB Labour Market

INDICATOR Unemployment rate

SOURCE Statistical source

REGION East vs West

PERIOD Time period

VALUE Actual indicator value

A human baseline becomes the Gold Standard

1

Select the sample

5 German + 5 Italian reports,
analysed in depth

2

Read the report

Purpose, institution, audience,
scope, dominant narrative

3

Extract claims

The claim is the primary unit of
analysis

FACT OR CLAIM?

FACT

"Region X had 8 % unemployment in
2022."

CLAIM

"Region X faces structural labour-market
disadvantage."

4

Classify claims

Assign a Narrative Codebook
code, e.g. DEMO, LAB, COH

5

Trace supporting evidence

Type, source, indicator, period,
region, location

6

Assess claim–evidence relation

Four levels — existence of
evidence is not enough

OUTPUT PER REPORT

★ **Narrative Map → Gold Standard
dataset**

Claim ID · text · type

Narrative code

Evidence ID · source

Evidence relation

Missing evidence

Narrative frame

Location

Annotation

7

Find potential missing evidence

Relevant, accessible, but
unused in the report

8

Identify narrative frames

Claims that share one
interpretive logic

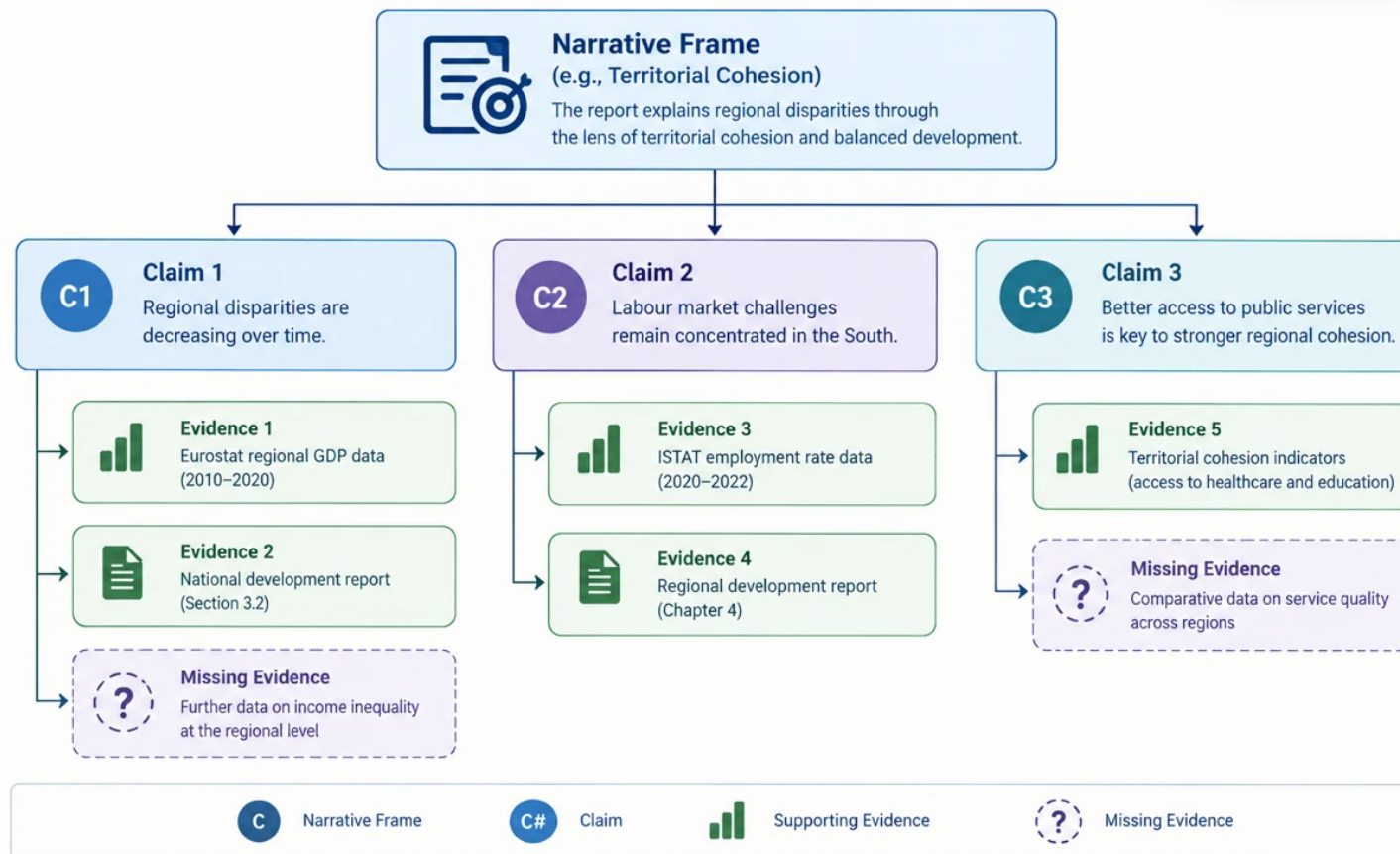
9

Analyse selection & highlighting

Pattern of emphasis — not
political "bias"

How strongly does the evidence carry the claim?

Example of a Narrative Map for One Report

[↑ Replace](#)


CLAIM-EVIDENCE RELATION

- Directly supported
- Partially supported
- Indirectly related
- Insufficient evidence

Describes the *type of relationship* — not the truth or falsity of a claim.

? POTENTIAL MISSING EVIDENCE

Recorded only if **both** hold:

1. Relevant, accessible evidence exists in the Evidence Corpus
2. It is conceptually or empirically relevant to the claim

Figure 4. General structure of the Narrative Map (idea by authors, visualized by AI)

Small, traceable steps — never one big prompt

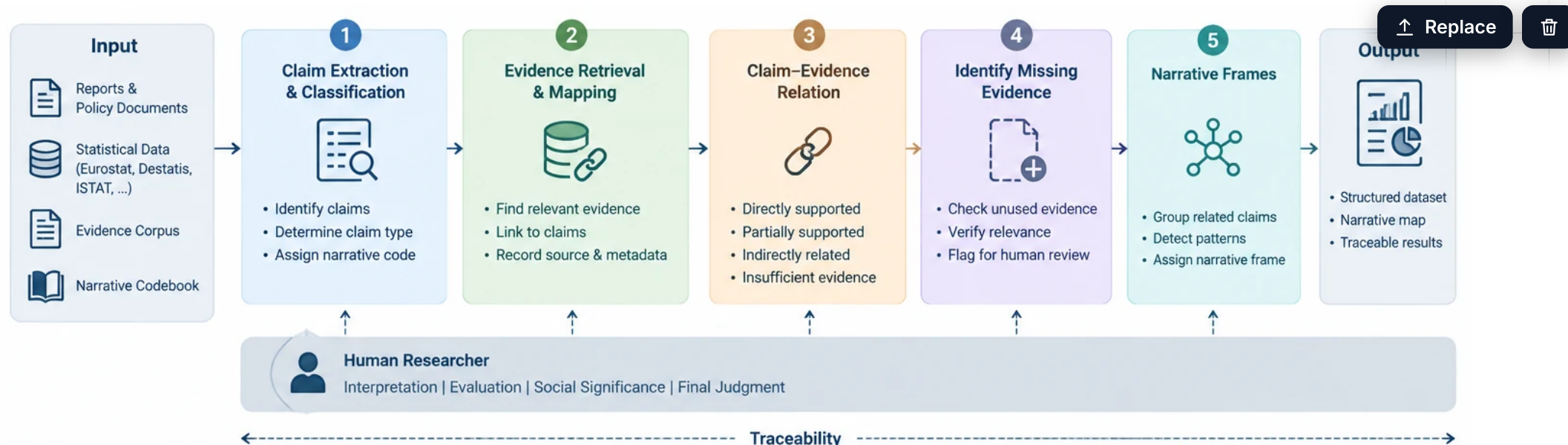


Figure 2. LLM-assisted narrative auditing workflow (idea by authors, visualized by AI)



Stepwise prompts

Extraction → classification → evidence matching → relation → unsupported claims → missing evidence → frames.



Traceable evidence

No evidence from the model's general knowledge — each item carries source, indicator, period, region, location.

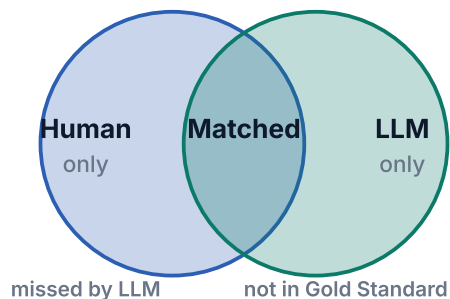


Human in the loop

Weakly supported claims are flagged for review; output schema mirrors the Gold Standard.

Testing the LLM against the human Gold Standard

EXAMPLE · CLAIM EXTRACTION



- Matched** — found by both
- Missed** — human claim the LLM did not find
- Extra** — LLM claim absent from Gold Standard

10 reports with Gold Standard
(5 Germany · 5 Italy)

SEVEN COMPARISON LEVELS

- 1 Claim extraction
- 2 Claim classification
- 3 Claim–evidence mapping
- 4 Evidence relation
- 5 Unsupported claims
- 6 Potential missing evidence
- 7 Narrative frames

Disagreements are typed: different interpretation · codebook ambiguity · LLM error.

REPORT-LEVEL MEASURES

COUNTS

- Claims
- Evidence links
- Claims without evidence
- Potential missing evidence
- Narrative frames

LLM VS HUMAN AGREEMENT

- Claim–evidence agreement
- Classification agreement
- Frame agreement



↑ Replace



Add figure or chart

Upload, drop or paste an image

Same social question, different narrative structures?

DE Germany

FRAMES TO EXAMINE

Demographic transformation

Population ageing

Labour-market change

Regional cohesion

Claims

Evidence

Potential missing evidence

CROSS-CASE COMPARISON

LEVEL 1

Common patterns

LEVEL 2

Country-specific patterns

LEVEL 3

Evidence–narrative differences

IT Italy

FRAMES TO EXAMINE

Economic development

Labour-market disadvantage

Territorial disparities

Institutional capacity

Claims

Evidence

Potential missing evidence



These themes are **hypotheses to be tested**, not predetermined findings. The comparison covers **how evidence is selected, organised and used** — not only how many claims.

Where does the LLM add value — and where not?

MORE AUTOMATION

MORE HUMAN JUDGMENT



GROUP 1

LLM-suitable

- Initial claim extraction
- Searching for evidence
- Organising information
- Initial claim–evidence links
- Initial grouping of claims

DELEGATE



GROUP 2

Human-in-the-loop

- Assessing claim–evidence relation
- Identifying potential missing evidence
- Selecting narrative frames
- Reviewing ambiguous cases
- Controlling sources and evidence

COLLABORATE



GROUP 3

Primarily human

- Social interpretation
- Historical & institutional context
- Theoretical significance
- Scientific conclusions
- Final meaning of the narrative

RETAIN

LIMITATIONS EXAMINED SYSTEMATICALLY

⚠ Hallucination

🔍 Missing evidence from general knowledge

↻ Prompt dependence

⚖ Assessing significance

📄 Social interpretation



Guiding principle: **traceability** — every finding leads back to a claim, evidence and source.

Not to automate social interpretation — but to make the path from **evidence** to **narrative** visible, traceable and open to examination.



Narrative Codebook

shared, controlled vocabulary



Evidence Corpus

indicators + documents, linked



Claim–evidence mapping

human baseline, LLM-assisted



Task allocation

LLM · human-in-the-loop · human

One workflow connecting social monitoring, narrative analysis, evidence traceability and LLM-assisted research
— tested on Germany and Italy.

Thank you

Questions & discussion

bashiri@hut.ac.ir
petra.ahrweiler@uni-mainz.de
m.rusconi8@uninsubria.it