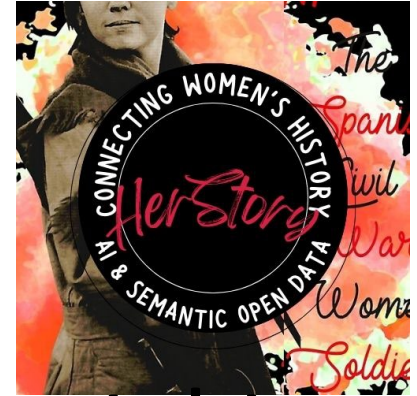




UNIVERSITAT DE
BARCELONA

HerStory (2024-27)



HerStory-NeSyAI: Designing inclusive metadata architectures with hybrid AI for epistemic justice in silenced narratives

Miquel Centelles and Núria Ferran



<http://creativecommons.org/licenses/by-nc-sa/3.0/e>



*If they erase our stories,
they will erase us twice.*

**Depurada por conducta
contraria al régimen**

- Barcelona, 1941: **Maria**, a teacher, was arrested for mixed-gender teaching, encouraging questions, and using newly forbidden books, recorded as “Depurada por conducta contraria al régimen” — *“Purged for behaviour contrary to the regime.”*
- This occurred under the Francoist dictatorship (1939–1975) after the overthrow of the Second Republic (1931–1939) and the 1936–1939 Civil War, before Spain’s 1978 constitutional transition.
- Fragments of her life persist across multiple databases, silenced archives, and records in the process of disappearing. In and out of the University of Barcelona...
- ...yet thousands share this pattern—women whose histories remain invisible, disconnected, and unrecognised.

Multiple traces of Maria

Culture and Censorship DB

 **censura**

Proyectos Miembros Publicaciones Noticias y otras informaciones Base de Datos

Autores Editores Importadores Lectores Libros presentados Libros publicados Proveedores de papel Solicitudes de publicación

Buscar


Escribe las palabras a buscar

☐ Libro

☐ Producción y distribución del libro


☐ Censura del libro

Buscar Limpiar

 **ihr.world**

DONATIONS NEWSLETTER WHO WE WORK WITH ENGLISH

IHR IN THE MEDIA
TEAM
THE RIGHT OF ACCESS TO INFORMATION
PRIVACY POLICY AND LEGAL NOTICE



1449.374 registros **BUSCAR**

THE SPANISH CIVIL WAR AND FRANCOIST SPAIN IN DATA

The **central database** is our main project. We compile and then re-publish data in order to provide access to information on the Spanish Civil War and the Franco Regime, with particular regard to victims of repression and reprisal. We also publish articles about the

Judicial Processes of Francoism DB

Democratic Memory Oral Archive

Fosses comunes a Catalunya **Legislació i Justícia**

Per motius tècnics les dades d'aquest conjunt no estan actualitzades a la última informació disponible.

Última actualització: 2 d'octubre de 2024


Dades proveïdes per: Direcció General de Memòria Democràtica

Listat de fosses comunes localitzades a Catalunya.

Continguts destacats amb aquestes dades

Drets i Memòria de la Generalitat de Catalunya


Contingut extern



Accediu a Drets i Memòria de la Generalitat de Catalunya

Fosses (mapa, plans i actuacions)

Contingut extern



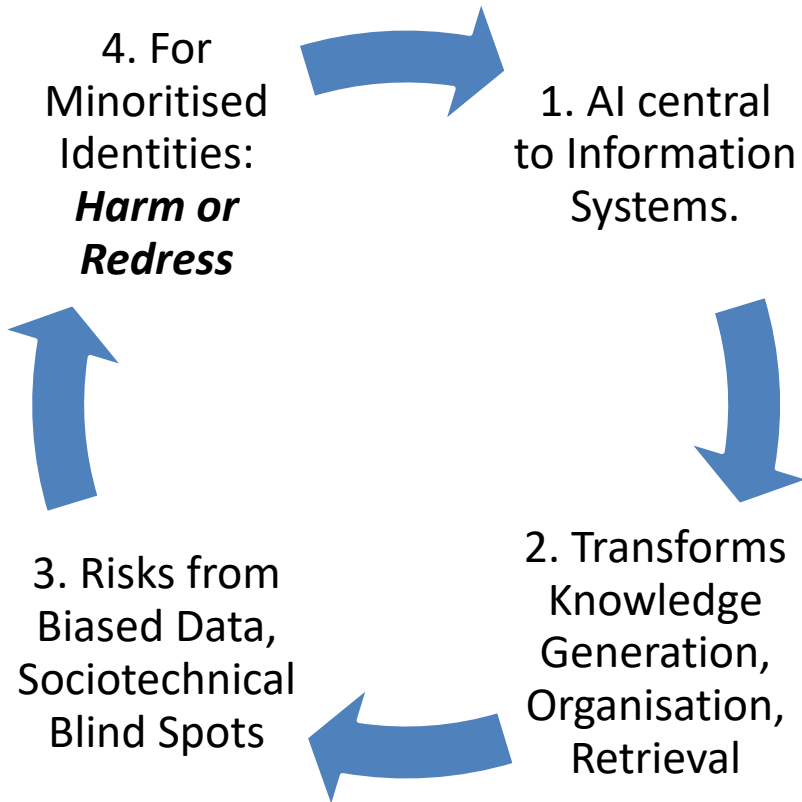
Accediu a Fosses (mapa, plans i actuacions)

In construction

Repressed Midwives and Teachers DB

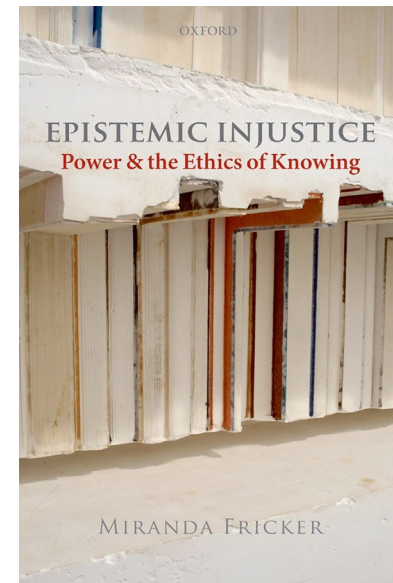
Bridging AI and Epistemic Justice

Core thesis



Main Focus

HerStory-NeSyAI project:
generative & neuro-symbolic
solution



The HerStory-NeSyAI Initiative: Context & Aims

1. The project runs from December 2024 to August 2027 at the University of Barcelona's Women and Wikipedia team.

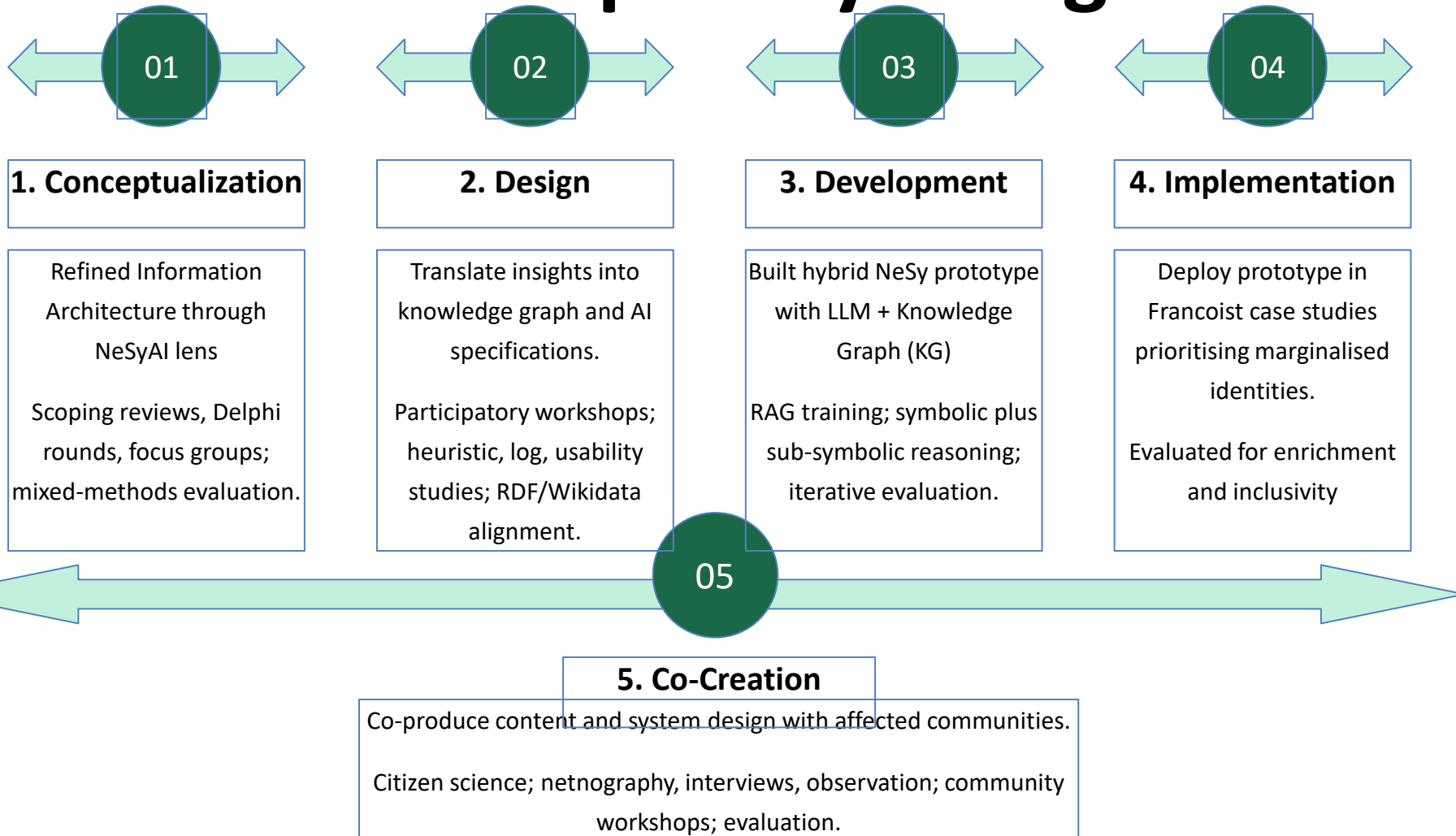
3. It focuses on gender-sensitive historical metadata from Francoist repression archives (1936–1975), integrated as linked open data.

2. It tackles database fragmentation by linking shared entities and connecting collections with Wikidata.

4. It advances content creation, curation, and access to increase visibility and equity.

5. It adopts a hybrid AI architecture—generative and neuro-symbolic—guided by human-centred research to interrogate gender narratives.

Methodology: Five-Phase, Interdisciplinary Design




Conceptual & Ethical Framework: Metadata, Paradigms, and Risks (1)

A) Metadata practices shape how AI interprets and organises knowledge.



B) Two paradigms—symbolic and sub-symbolic—are complemented by neural-symbolic AI.



C) Generative LLMs (transformers over large corpora) pose risks—imbalance, opacity, hallucination, bias—requiring fairness, transparency, and inclusion.

Conceptual & Ethical Framework: Metadata, Paradigms, and Risks (2)

B) Two paradigms—symbolic and sub-symbolic—are complemented by neural-symbolic AI.

Symbolic offers interpretable ontologies yet risks rigidity for gender diversity.



Sub-symbolic excels at pattern learning yet remains opaque and bias-prone.



Neural-symbolic pairs neural recognition with logic for flexible, explainable retrieval.

Conceptual & Ethical Framework: Metadata, Paradigms, and Risks (3)

A) Metadata practices shape how AI interprets and organises knowledge.

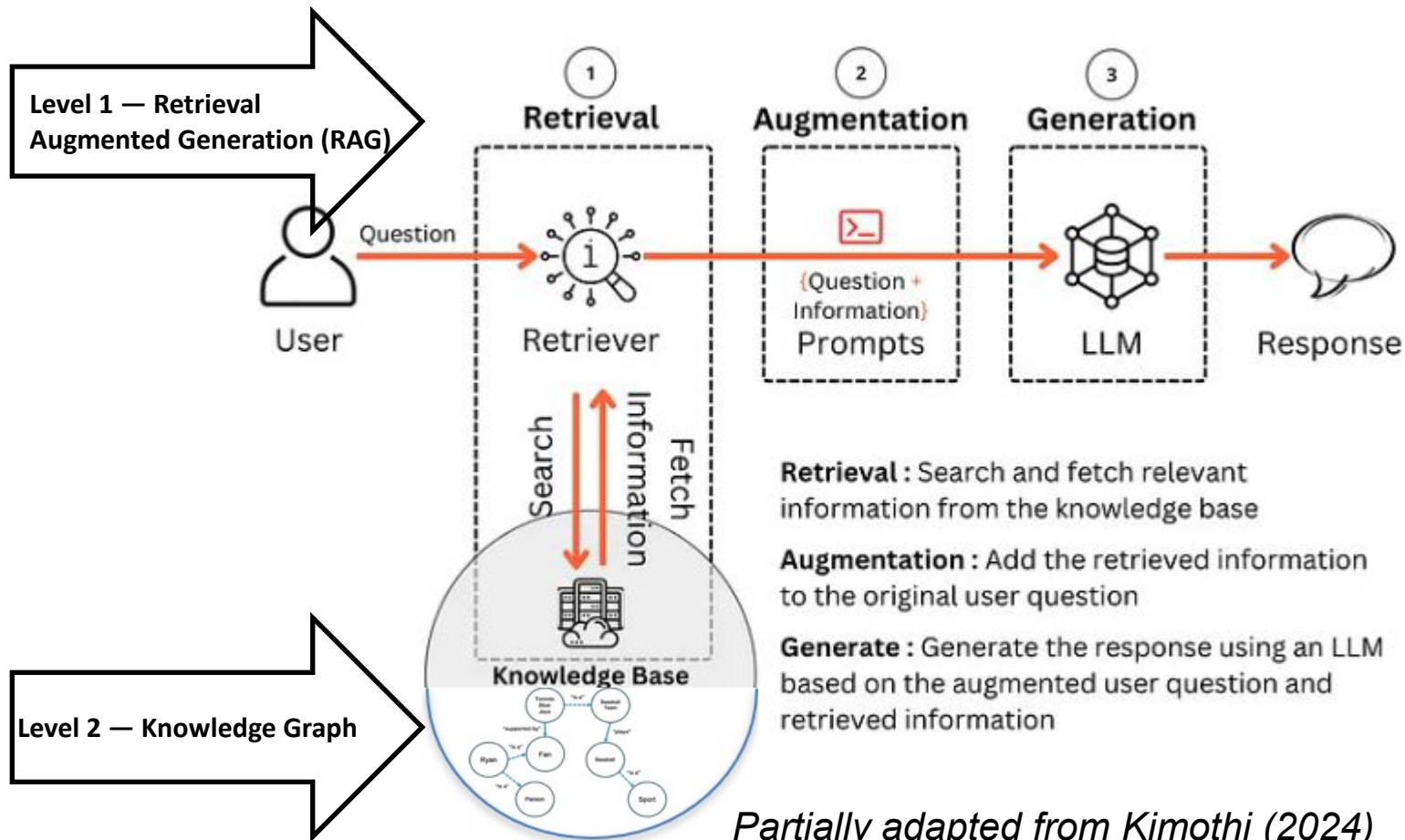
A large blue downward-pointing arrow connects the first box to the second box.

B) Two paradigms—symbolic and sub-symbolic—are complemented by neural-symbolic AI.

A large blue downward-pointing arrow connects the second box to the third box.

C) Generative LLMs (transformers over large corpora) pose risks—imbalance, opacity, hallucination, bias—requiring fairness, transparency, and inclusion.

Hybrid Metadata Architecture: Two Level Strategy



Prototype & Two-Layer Knowledge Graph



Our **prototype** operationalises the **hybrid model** by integrating an **LLM** with a custom **knowledge graph**.

It supports inclusive, transparent database management, human–bot interactions in historical data, and automated Wikipedia article generation from validated sources.

The knowledge graph includes an ontology layer enriched from multiple historical databases.



A **two-layer knowledge graph**:

It includes an ontology layer enriched from multiple historical databases.

It includes an entity layer of individuals and events supplemented by Wikidata and Wikipedia.

It captures factual content and the semantic relationships and categories that frame it, addressing bias in data and representation.



Metadata are structured using RDF and mapped to Wikidata properties via OpenRefine, with interoperability ensured through Semantic Web technologies such as SPARQL and OWL.

Human-Centred Metadata Design: Dual Perspective



HerStory-NeSyAI adopts a human-centred design that shapes both its methodology and the architecture of its AI and knowledge graph.



This approach combines advanced techniques for ontology learning with participatory, user-centred methods.

It integrates a content-centred dimension to ensure the most unbiased possible representation of historical events and figures.

It integrates a user-centred dimension tailored to the informational needs and behaviours of those who create, curate, or consult the data.

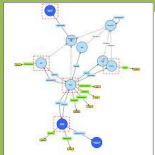


This dual perspective enables responses that are fairer, more contextually grounded, and aligned with the diversity of users and communities involved.

Human-Centred Metadata Design: Methods Across Phases



Human-Centred Design, specifically user-centred design, characterises the project and affects its methodology, the AI and knowledge graph system, and the design of the knowledge graph.



During the design phase, semi-automatic learning of ontologies and knowledge graphs from documentary corpora is combined with user experience and information architecture methods.

User participation defines goals and shapes design and implementation.

Research mixes qualitative and quantitative methods and uses scenarios and card sorting for conceptual design and prototyping.



This integration continues into development and implementation, where the knowledge graph—serving as the knowledge base of RAG—and LLM components are validated in real-world archival and research scenarios.

Reflections on Inclusive AI & Metadata for Historical Representation (1)

Pre-production status

- A conceptualised, partly developed, qualitatively tested LLM+KG prototype is not yet deployed...
- ...amid access restrictions, uneven digitisation, and ongoing interface refinement—structural challenges requiring technical innovation, institutional partnerships, and policy support.

Risks and Scenarios of Misuse

- Generative AI may hallucinate and misrepresent sensitive histories, causing epistemic harm and revisionist narratives...
- ... given Wikipedia's symbolic authority, embed explainability, traceable provenance, and participatory validation, complemented by continuous monitoring and community oversight.

Reflections on Inclusive AI & Metadata for Historical Representation (2)

Replicability and Broader Applicability

- Modular, Semantic Web–based architecture scales beyond the original case...
- ...the two-layer knowledge graph adapts to colonial archives, LGBTQ+ memory projects, and transitional justice, offering a template for critical metadata infrastructures.

Contribution to DCMI and FAIR/CARE-Aligned Practices

- Advances DCMI debates on transparent, accountable, just metadata, operationalises FAIR via open standards and Wikidata alignment...
- ... engages CARE through community participation, and integrates feminist epistemology with linked data for a replicable, ethically responsive framework.

Gràcies!

¡Gracias!

Thank you!

HerStory (2024-27)



**Your comments,
suggestions,
proposals...**

