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

Núria Ferran-Ferrer[1,†] and Miquel Centelles[1,†]

[1] *Universitat de Barcelona, Faculty of Information and Audiovisual Media, Melcior de Palau 140, 08014 Barcelona, Spain*

Abstract

A presentation of the HerStory-NeSyAI project, which designs inclusive metadata architectures using hybrid AI to address epistemic justice in silenced narratives. It combines interdisciplinary approaches from Library and Information Science, Digital Humanities, and Feminist Theory. The project aims to bridge historical and technological silences by focusing on gender-sensitive representation and ethical AI development. A hybrid neuro-symbolic AI architecture is developed to mitigate bias and enhance transparency in knowledge infrastructures. The ultimate goal is to transform AI into a vehicle for historical and social accountability.

Keywords

Inclusive Metadata, Neuro-Symbolic AI, Epistemic Justice [1]

1. Introduction

Artificial Intelligence (AI) is increasingly central to contemporary information systems, offering transformative potential in how knowledge is generated, organized, and retrieved. However, this same capacity entails critical risks when such systems are developed from biased data sets or without an awareness of their broader sociotechnical implications. In the context of women's representation and that of other minoritized identities, AI can either reinforce historical inequalities or serve as a tool for redress—if designed through ethical, inclusive, and critical feminist frameworks. This section examines how AI models, particularly generative and neuro-symbolic approaches, can contribute to more equitable knowledge infrastructures.

Drawing on the *HerStory-NeSyAI* project, it explores the intersection of technological innovation and epistemic responsibility, highlighting opportunities to reimagine historical memory through interdisciplinary and gender-aware methodologies.

* Corresponding author.
† These authors contributed equally.
aleksandr.ometov@tuni.fi (A. Ometov); t.princesales@utwente.nl (T. P. Sales); manfred.jeusfeld@acm.org (M. Jeusfeld)
0000-0003-3412-1639 (A. Ometov); 0000-0002-5385-5761 (T. P. Sales); 0000-0002-9421-8566 (M. Jeusfeld)

https://doi.org/10.23106/dcmi.952573231

2. Project context: The *HerStory-NeSyAI* initiative

The *HerStory-NeSyAI* project, funded by the Spanish Ministry of Science and Innovation (PID2023-147673OB-I00), runs from December 2024 to August 2027 at the University of Barcelona's Women and Wikipedia research team. It focuses on gender-sensitive representation of historical metadata, emphasising the digital visibility of women and minoritised identities. Methodologically, it combines an interdisciplinary approach—drawing from Library and Information Science, Digital Humanities, AI, and Feminist Theory—with the development of a hybrid AI architecture. The project is based on archives documenting repression during Spain's Francoist dictatorship (1936–1975), which are critically analysed and integrated into linked open data systems. *HerStory-NeSyAI* aims to bridge historical and technological silences, addressing the marginalisation of women and gender minorities, as well as the limitations of current digital infrastructures. While rooted in academia, its ethical, social, and epistemological implications extend beyond, concerning AI's role in historical and cultural contexts.

The project uses an interdisciplinary approach, incorporating Library and Information Science (LIS), Communication, Computer Science, the Humanities, Law, and collaboration with Wikipedia and Wikidata communities. Its focus is on content creation, curation, and access, particularly concerning digital archives of Francoist repression and censorship in Spain (1936–1975). These sources are analysed through gender and intersectional perspectives, aiming to highlight previously silenced voices.

A major challenge for *HerStory-NeSyAI* is the fragmentation of digital historical and humanities databases, often created independently within academic projects. Each initiative tends to develop isolated databases, which hampers the ability to identify common entities across collections. For instance, if a woman appears in multiple archives, there is no automatic way to link her identity. For research on women's narratives—already marked by invisibility—such connections are vital for reconstructing life stories and building a shared, inclusive memory.

Beyond this, many databases do not interface with external repositories like Wikidata, leaving knowledge confined to academic silos. This disconnect reduces visibility and limits the use of research data, particularly in gender history and digital heritage.

HerStory-NeSyAI focuses on three main areas: first, content creation involves enriching Wikidata entities, developing new Wikipedia articles, and linking data repositories to improve the visibility of women and marginalised groups in digital knowledge spaces. Second, content curation aims to improve classification, summarisation, and indexing to ensure information remains relevant, accurate, and accessible. Lastly, content access develops inclusive search and navigation tools for both public and academic use, promoting equitable knowledge dissemination. These activities are constrained by infrastructural and epistemic limitations, particularly regarding gender and minoritised identities. The invisibility of women and minorities stems not just from historical neglect but also from systemic biases in data collection, description, and categorisation. Using AI in this context risks amplifying these biases if systems are trained on flawed data and implemented without proper oversight.

https://doi.org/10.23106/dcmi.952573231

In today's data-driven ecosystems, where AI mediates access to information, there is an urgent need for inclusive, transparent, and reflexive design. Without this, digital infrastructure will perpetuate the exclusion of already marginalised groups [1, 2]. *HerStory-NeSyAI* aims not just to apply AI but to redirect its purpose towards feminist, interdisciplinary, and socially just knowledge practices.

The need for intentional, inclusive design is heightened by the particular capabilities and limitations of generative artificial intelligence (GAI). GAI is a subset of AI that creates content—text, images, audio—based on learned patterns, unlike traditional AI that recognises or predicts. It utilises large neural networks trained on vast datasets from the internet to mimic human creativity and responsiveness. Following the work of Finzel [3], two major research areas have emerged as key to improving the trustworthiness of AI: neuro-symbolic AI and human-centered research, alongside the broader field of explainable AI (XAI). There are additional precedents in the combination of these two approaches, such as Gomaa and Feld [4,5] and Mileo [6]. For the HerStory NesyAI project, these two approaches have been selected as core methodologies to support the achievement of our objectives. In the following sections, we present our own perspective on these approaches.

As these systems increasingly shape the production and sharing of historical knowledge, they present both opportunities for visibility and challenges related to bias, opacity, and epistemic harm. *HerStory-NeSyAI* examines how the architecture and training data of generative models influence what and who is represented. It investigates whether AI systems reinforce or challenge existing gender narratives, depending on how they are trained, fine-tuned, and embedded in digital heritage infrastructures.

3. Methodology

The *HerStory-NeSyAI* project adopts a multi-phase, interdisciplinary methodology that integrates qualitative, quantitative, and computational methods, with a strong emphasis on participatory design and epistemic accountability. The project unfolds through five interconnected phases: conceptualization, design, development, implementation, and co-creation. Each phase is aligned with specific methodological tools to ensure robust knowledge production, inclusive data modeling, and collaborative AI-assisted metadata generation. Moreover, each of them includes evaluation mechanisms that combine both quantitative and qualitative methods and metrics.

3.1. Conceptualization Phase

The project begins by updating the conceptual foundation of Information Architecture (IA) considering recent developments in neuro-symbolic artificial intelligence (NeSy AI). This phase includes a scoping review of literature on AI [7], metadata, and gender bias; Delphi rounds with experts to define epistemic goals [8]; and focus groups with stakeholders working on Francoist repression archives [9]. These methods serve to identify user needs, knowledge gaps, and ethical challenges in current metadata systems. Particular attention is given to how gender and intersectional perspectives can be integrated into IA frameworks.

During this phase, the project refines its theoretical and technical foundations. It applies scoping reviews, Delphi panels, and focus groups with experts and practitioners from gender

https://doi.org/10.23106/dcmi.952573231

studies, library science, and digital humanities to redefine Information Architecture in relation to neuro-symbolic AI, or NeSyAI [10]. In parallel, historical datasets on Francoist repression are assessed using qualitative analysis of metadata structures and coverage gaps, particularly regarding women and LGBTQ+ individuals.

3.2. Design Phase

In this phase, the research team translates theoretical insights and user needs into system design through a participatory, user-centered approach. Design workshops are held with key stakeholders—including editors, librarians, digital humanists, and gender-focused Wikimedians—to shape the knowledge graph architecture and guide AI system specifications. User behaviour is studied through heuristic evaluations [11], log analysis [12], and usability testing, providing empirical insights to improve navigation, labelling, and search functionalities. Simultaneously, archival datasets are semantically transformed into RDF and aligned with Wikidata properties to ensure interoperability, using tools like OpenRefine and custom entity-mapping workflows.

Methods include participatory workshops with Wikipedia and Wikidata communities, contextual interviews with librarians, historians, and archivists, and card sorting combined with scenario-based prototyping to define metadata structures and interaction models.Each author must be defined separately for accurate metadata identification. Multiple authors may share one affiliation. Authors' names should not be abbreviated; use full first names wherever possible. Include authors' e-mail addresses whenever possible.

3.3. Development Phase

In this phase, the project develops a hybrid neuro-symbolic (NeSy) AI prototype by integrating a large language model (LLM) with a two-layered knowledge graph comprising ontologies and structured entities. The LLM is trained on curated, domain-specific corpora and embedded within a Retrieval-Augmented Generation (RAG) framework to enhance context-aware responses and reasoning capabilities.

The development process includes training and evaluating the system using targeted datasets, integrating symbolic reasoning with sub-symbolic pattern recognition. Gender-aware ontologies are incorporated to ensure semantic consistency and inclusive metadata representation.

Co-design continues through iterative user testing and co-creation sessions [13], emphasizing explainability, fairness, and usability. Visualization tools are developed to render the AI's decision-making processes transparent and interpretable. Each iteration is refined through feedback loops involving both expert users and community stakeholders.

3.4. Implementation Phase

In the fourth phase, the prototype is deployed in real-world case studies focused on Francoist repression, with particular attention to women and other marginalized identities. The system is evaluated through scenario-based testing and archival use cases, assessing its capacity for metadata enrichment, content generation (such as draft Wikipedia articles), and user interaction.

https://doi.org/10.23106/dcmi.952573231

Pilot testing is carried out using data from five historical digital humanities databases. This phase involves user testing sessions with historians, librarians, and citizen archivists, followed by iterative interface refinement based on feedback. System performance is evaluated in terms of its ability to generate new content, link entities across platforms (e.g., Wikidata), and support inclusive historical representation and knowledge discovery.

3.5. Co-Creation and Citizen Science Phase

Finally, *HerStory-NeSyAI* employs a citizen science framework rooted in Digital Humanities to ensure that system design and outcomes are co-produced with affected communities. Netnography [14], contextual interviews [15], and participant observation methods [16] are applied to document the practices and perspectives of volunteer contributors. Community workshops are organized to foster collaborative article writing and metadata enrichment. These co-design processes not only improve system responsiveness but also embed community values into the architecture itself, ensuring that the resulting infrastructure serves both technical and social goals.

This phase ensures that both content and system design are co-produced with historically marginalized communities, reinforcing transparency and accountability in AI-assisted knowledge infrastructures.

4. Conceptual and Ethical Framework

Metadata practices are central to how AI systems interpret and organize knowledge. There are two main paradigms: symbolic AI and sub-symbolic AI, with a hybrid approach—neural-symbolic AI—emerging as an alternative [17, 18, 19, 20]. Symbolic AI relies on explicit rules and formal logic, valued for interpretability and predictable reasoning. For example, in gender classification within archives, it might use fixed categories like “man,” “woman,” and “other” defined through ontologies, providing clarity but limiting flexibility for non-binary identities or evolving terminology, potentially reinforcing normative frameworks. Conversely, sub-symbolic AI, such as deep learning, learns from data patterns without predefined rules, excelling at tasks like image recognition or language processing. In gender-related applications, these models infer gender from names or pronouns but often lack transparency and risk bias, reproducing stereotypes and misclassifying nonconforming identities.

To overcome the limitations of symbolic rigidity and sub-symbolic opacity, neural-symbolic AI combines neural networks’ pattern recognition with symbolic logic, enabling more flexible and explainable models. For instance, a neural-symbolic system supporting gender-inclusive retrieval might process language via deep learning while using ontologies like Wikidata to ensure diverse gender representation, promoting fairness in digital heritage platforms. This integration offers greater potential for developing inclusive, accountable AI, especially in areas related to gender equity and epistemic justice.

Building on this framework, it’s important to examine how these paradigms manifest in current AI, notably Large Language Models (LLMs) like GPT, BERT, or LLaMA. These models learn from extensive text corpora, converting language into numerical vectors to identify statistically relevant patterns. They use mechanisms called transformers, which prioritize tokens based on context, capturing meaning and relationships through attention processes.

https://doi.org/10.23106/dcmi.952573231

Generative AI systems have impressive capabilities but also pose significant challenges, especially in sensitive areas like gender representation. These issues—structural and epistemological—must be critically addressed to prevent reproducing existing exclusions.

First, data limitations are fundamental. Generative models need vast datasets, which are often unevenly distributed. In gender contexts, biographies of men—especially white, Western, and prominent figures—are overrepresented, while those of women or non-binary individuals are scarce. This imbalance causes models to favor dominant narratives and marginalize less documented voices, like overlooking women scholars in historical summaries.

Second, these systems lack transparency. Deep learning models operate as "black boxes," making their decision processes difficult to interpret —a problem in areas requiring accountability. For example, biased output might omit women's contributions or emphasize men's, with unclear reasons behind these outcomes.

Third, hallucination —the tendency to produce plausible but incorrect information—can lead to harmful representations. For instance, fabricating achievements for male figures or erasing women's contributions, thereby perpetuating epistemic erasure.

Finally, bias is pervasive. Because models learn from human language and sources filled with stereotypes, biases are inherited and often amplified. For example, leadership terms are frequently associated with men, while roles like "teacher" are linked to women, reinforcing societal inequities and influencing public discourse.

While the capabilities of generative AI are significant, their systemic limitations must be addressed as an ethical necessity—especially in areas of representation, inclusion, and justice. Technologies should be developed with critical frameworks and inclusive principles emphasizing fairness, transparency, and marginalized voices.

Empirical research shows biases in AI stem from four main sources: first, social norms and stereotypes embedded in training data reinforce dominant narratives and marginalize non-normative identities; second, skewed datasets underrepresent or misrepresent certain groups, reducing their visibility; third, design decisions often prioritize efficiency over ethical considerations, encoding structural biases; and fourth, human interventions like fine-tuning or prompt engineering can unintentionally reinforce existing biases when conducted without diversity frameworks.

These issues impact not only the accuracy of outputs but also trust, fairness, and legitimacy—crucial when AI systems influence public knowledge and institutional memory. Left unaddressed, these biases risk perpetuating historic exclusions, silencing marginalized communities in digital archives and information systems.

5. Metadata Architecture: Hybrid AI and Knowledge Graph

In response to these challenges, the *HerStory-NeSyAI* project adopts a two-tiered strategy to mitigate hallucination and bias in large language models, combining retrieval-augmented generation (RAG) with knowledge graph integration. This hybrid architecture grounds generative outputs in verifiable sources while enabling more nuanced and explainable reasoning—shifting AI from probabilistic prediction toward epistemic accountability and inclusion.

The first layer, RAG, integrates a retrieval component into the generative process. Rather than relying solely on the model's training data, it retrieves relevant documents from a curated

https://doi.org/10.23106/dcmi.952573231

corpus. This enhances output quality in four keyways: it reduces hallucination through semantic search and real-time access to trusted texts; mitigates bias by prioritizing peer-reviewed or community-validated sources; improves transparency by making each output traceable to its sources; and ensures domain specificity by grounding generation in contextually relevant information.

The second layer complements RAG by integrating knowledge graphs (KGs)—structured databases that represent entities and their relationships through nodes and edges, grounded in formal ontologies. When embedded within a RAG-based system, knowledge graphs enhance the generative process in four keyways. First, they provide semantic structure, enabling the model to retrieve information that is not only textually relevant but also logically connected through defined relationships. Second, they support contextual coherence, allowing the AI to generate outputs that reflect accurate temporal, spatial, and thematic associations—crucial when dealing with historically underrepresented groups. Third, they allow for real-time updates and interoperability, as new entities and relationships can be continuously added and linked across datasets. Finally, KGs improve explainability, offering users a transparent view of how outputs are generated by tracing connections between facts, concepts, and sources.

Together, the combination of retrieval-augmented generation and knowledge graph integration shifts the system from probabilistic language prediction to a more accountable, traceable, and semantically grounded model of knowledge generation. Rather than producing plausible-sounding but unverifiable content, the system becomes capable of generating outputs that are context-aware, transparent, and responsive to the epistemic demands of inclusion and representation.

Building on this architecture, the *HerStory-NeSyAI* project is developing a generative AI prototype that operationalizes this hybrid model. The prototype will integrate a large language model (LLM) with a custom knowledge graph to support three main functions: the creation and management of inclusive and transparent databases; the facilitation of interactions between humans and bots within historical data environments; and the automated generation of Wikipedia articles, drawing from structured, validated data. This last function is particularly significant for addressing the visibility gap of women and minoritized identities on widely consulted knowledge platforms.

At the core of the prototype lies a two-layered knowledge graph architecture. The first layer, the ontology layer, builds upon existing semantic frameworks and is enriched by transforming and aligning data from multiple historical databases. The second, the entity layer, is composed of individuals and events drawn from diverse research projects related to the Francoist dictatorship and is further supplemented with information from Wikidata and Wikipedia. This dual structure allows the system to capture not only the factual content of historical narratives but also the semantic relationships and conceptual categories that frame them—thereby addressing bias not only in data but also in representation.

Although the project is ongoing, the architecture has been tested through qualitative validation scenarios. These include simulated experiments in which structured queries are passed through the KG+LLM prototype to assess semantic alignment, bias sensitivity, and basic content traceability. These early tests have informed iterative adjustments to the knowledge graph ontology, particularly regarding gender representation and classification.

From a metadata perspective, the project explicitly structures data using RDF and maps it to Wikidata properties through OpenRefine workflows. Ontological properties are defined based

https://doi.org/10.23106/dcmi.952573231

on both standard vocabularies (e.g., Dublin Core, Wikidata schema) and feminist-informed refinements that allow non-binary and intersectional identity representation.

Metadata interoperability is ensured through adherence to Semantic Web standards (e.g., SPARQL, OWL) and alignment with Wikidata ontologies, which are updated dynamically as part of co-creation cycles with expert and community users.

This infrastructure has been explicitly designed to mitigate bias at the representational level, while ensuring that generated outputs remain transparent, traceable, and epistemically robust. By embedding ontological precision into a generative architecture, *HerStory-NeSyAI* seeks to transform AI from a neutral-seeming technical tool into a vehicle for historical and social accountability.

A key enabler of this vision is Wikidata, which plays a central role in the project's knowledge architecture. Unlike Wikipedia—where editorial dynamics often reflect sociocultural biases and negotiation—Wikidata operates through a more formalized and scalable system of knowledge organization. Its community is actively working to address issues of ontological inconsistency and representational bias, developing tools that maintain coherence across properties, classes, and labels. Of particular importance is Wikidata's growing capacity to represent gender diversity, offering properties that accommodate non-binary and non-traditional gender identities— something that remains difficult to implement in many mainstream data infrastructures. *HerStory-NeSyAI* both leverages and contributes to this ongoing evolution, aligning its design with the broader ecosystem of open, participatory, and inclusive knowledge infrastructures.

6. Human-Centered Metadata Design

In line with the principles outlined above, the *HerStory-NeSyAI* project adopts a human-centered design approach 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-centered methods [21, 22, 23, 24]. The resulting knowledge graph integrates a content-centered dimension—ensuring the most unbiased possible representation of historical events and figures—and a user-centered dimension, tailored to the informational needs and behaviors of those who create, curate, or consult the data. This dual perspective enables the generation of responses that are fairer, more contextually grounded, and aligned with the diversity of users and communities involved.

Connected to the previous points, we observe that the very conception of the Hertory project is characterized by Human-Centered Design, specifically focusing on user-centered design [25]. This emphasis profoundly impacts both the methodology applied throughout the various phases of the project and the nature of the AI and Knowledge Graph system being developed, as well as the design of its fundamental component—the knowledge graph.

Regarding methodology, during the design phase of the knowledge graph schema and AI architecture, advanced methodologies for semi-automatic learning of ontologies and knowledge graphs from documentary corpora will be combined with user experience and information architecture methods applied to information systems [26, 27]. Here, user-centered design and, consequently, user participation not only define the goals but also shape the design and implementation processes. Engaging with users through both qualitative and quantitative research, as well as conceptual design and prototyping methods such as scenarios and card

https://doi.org/10.23106/dcmi.952573231

sorting processes, will be integral. This integration of user-centered and content-focused methods will continue into the development phase (building and integrating LLM and KG components) as well as during implementation and testing (validating in real-world archival and research scenarios).

In this same vein, the knowledge graph, serving as the knowledge base of the RAG (Retrieval-Augmented Generation), incorporates two dimensions: a content-focused dimension and a user-centered dimension. The content-focused dimension ensures that entities (people, events, places, etc.) involved in the narrative are represented without bias, or at least with minimal bias reflecting the historical underrepresentation of certain groups. Meanwhile, the user-centered dimension of the knowledge graph represents the informational needs and behaviors of individuals who create, curate, search, and utilize the data integrated into the content-focused knowledge graph, again striving to be free from bias. The interplay of these two dimensions will facilitate responses to inquiries within the AI system that are (mostly) unbiased and tailored to the specific needs and preferences of its users.

7. Discussion

As *HerStory-NeSyAI* enters its second year of implementation, several critical reflections emerge regarding the challenges, risks, and broader implications of developing inclusive AI and metadata infrastructures for historical representation. While the project makes a strong conceptual and technical contribution to metadata architecture, it also confronts the inherent limitations and ethical complexities of operating at the intersection of artificial intelligence, gender justice, and cultural memory.

7.1. Limitations and Current Development Stage

The current phase of the project remains pre-production. The prototype that integrates a large language model (LLM) with a semantically enriched knowledge graph (KG) has been conceptualized, partially developed, and qualitatively tested in simulated scenarios. However, the system is not yet fully deployed in public or institutional environments. Access to certain archival datasets remains limited due to copyright restrictions, ethical concerns surrounding personal data, and uneven digitization levels across institutions. Additionally, interface design and user interaction flows are undergoing iterative refinement based on preliminary feedback from expert users and community collaborators.

These limitations reflect broader structural challenges in the field of digital heritage: the fragmentation of historical data, disparities in metadata quality, and the historical under-documentation of marginalized groups such as women, LGBTQ+ communities, and ethnic minorities. Overcoming these constraints requires not only technical innovation but institutional partnerships and policy-level support to improve data accessibility and ethical reuse.

7.2. Risks and Scenarios of Misuse

The use of generative AI in the domain of historical narrative creation poses specific risks. As highlighted in the project's conceptual framework, LLMs are prone to "hallucination," i.e., the generation of factually inaccurate or fabricated content that may sound plausible. In the context of politically sensitive or traumatic histories —such as state violence, censorship, or gender-

https://doi.org/10.23106/dcmi.952573231

based repression— such inaccuracies risk perpetuating epistemic harm, misinforming the public, or even reinforcing revisionist or exclusionary narratives.

Moreover, the symbolic authority that AI-generated content can assume on platforms like Wikipedia makes it especially urgent to embed safeguards against misrepresentation and bias. *HerStory-NeSyAI* responds to these risks by integrating explainability-by-design, traceable data provenance, and participatory validation workflows. Still, continuous monitoring and community-based oversight will be essential to prevent potential misuse or instrumentalization of the system.

7.3. Replicability and Broader Applicability

One of the project's strengths lies in its modular architecture and adherence to semantic web principles, which enable scalability beyond its original use case. The design of the knowledge graph—separated into an ontology layer and an entity layer—facilitates adaptation to other contexts involving historical silencing or exclusion. Potential domains of replication include:

- Colonial archives, where structural inequalities in metadata reflect Eurocentric biases.
- LGBTQ+ memory projects, which require flexible, non-binary representational schemas.
- Transitional justice databases, where provenance, traceability, and ethical metadata standards are crucial for legal and social accountability.

This cross-domain adaptability positions HerStory-NeSyAI not only as a tool for feminist archival intervention but as a template for critical metadata infrastructures in broader epistemically sensitive fields.

7.4. Contribution to DCMI and FAIR/CARE-Aligned Metadata Practices

HerStory-NeSyAI contributes directly to ongoing debates in the Dublin Core Metadata Initiative (DCMI) community around the role of metadata in promoting transparency, accountability, and justice in data systems. The project explicitly aligns its design with the FAIR principles (Findable, Accessible, Interoperable, Reusable) by using open standards (RDF, SPARQL), aligning properties with Wikidata ontologies, and facilitating interoperability across archives and platforms. Furthermore, it engages the CARE principles (Collective Benefit, Authority to Control, Responsibility, Ethics) by prioritizing community participation, especially from groups historically excluded from metadata systems and public knowledge production.

By integrating feminist epistemology with linked data technologies, HerStory-NeSyAI demonstrates how metadata can move beyond technical annotation to become a site of critical intervention and repair. Its methodology—combining co-creation, user-centered design, and hybrid AI—offers a replicable framework for building metadata systems that are not only technically robust but also ethically and socially responsive. In this way, the project not only advances academic discourse but also proposes practical models for inclusive, community-driven knowledge infrastructures.

https://doi.org/10.23106/dcmi.952573231

8. Conclusion and future work

This paper has presented the conceptual and technical foundations of the HerStory-NeSyAI project—a novel hybrid AI architecture that combines large language models (LLMs) with semantically enriched knowledge graphs (KGs) to support inclusive, transparent, and ethically grounded metadata systems. Rooted in a feminist and interdisciplinary perspective, the project addresses the structural underrepresentation of women and other minoritized identities within digital heritage infrastructures, with particular attention to contexts shaped by historical violence and institutional silencing, such as the Francoist dictatorship in Spain.

HerStory-NeSyAI positions metadata not simply as a technical layer, but as an epistemic practice—a way of intervening in how knowledge is structured, legitimized, and accessed. Its approach emphasizes:

- The design of ethical metadata frameworks that incorporate intersectionality and historical justice;
- The development of explainable, traceable AI systems by integrating symbolic reasoning with neural architectures;
- The centrality of participatory co-creation, embedding lived experience and community knowledge within metadata structures.

Although the system is still under active development, its architectural foundations are fully established, and preliminary validations have been conducted through simulations and expert consultations.

Next steps for the project include:

- Conducting user-centered evaluations in collaboration with archival institutions and digital humanities practitioners;
- Extending the model to other domains affected by epistemic erasure, including colonial archives, LGBTQ+ memory initiatives, and transitional justice datasets;
- Publishing open, FAIR-compliant knowledge graphs aligned with Wikidata standards to promote reuse, transparency, and interoperability.

Rather than delivering a static product, HerStory-NeSyAI seeks to offer a transferable and critically-informed framework for inclusive metadata design—one that opens space for collective reflection and shared agency in how knowledge is represented and made intelligible through AI systems.

-

References

[1] Gebru, T. Race and gender. In Markus D. Dubber, Frank Pasquale, & Sunit Das (Eds.), *The Oxford handbook of ethics of AI* (pp. 251–269). Oxford University Press, 2020. URL:

https://doi.org/10.23106/dcmi.952573231

https://books.google.com/books?hl=ca&lr=&id=8vQTEAAAQBAJ&oi=fnd&pg=PA253&dq=%22Race+and+gender%22&ots=JliKx7lww6&sig=wJPFIKhfJY05WhjmuRHBZhECT9g.

[2] Criado, J. I. Inteligencia artificial (y administración pública). *EUNOMÍA. Revista En Cultura de La Legalidad* 20 (2021), 348–372.

[3] Finzel, B. Toward trustworthy AI with integrative explainable AI frameworks. *It - Information Technology* 67, 1 (2025), 20–45. doi:10.1515/itit-2025-0007.

[4] Gomaa, A., Feld, M. Adaptive User-centered Neuro-symbolic Learning for Multimodal Interaction with Autonomous Systems (2023). doi:10.48550/arXiv.2309.05787.

[5] Gomaa, A., & Feld, M. (2023). Towards Adaptive User-centered Neuro-symbolic Learning for Multimodal Interaction with Autonomous Systems, in: ICMI'23: International Conference on Multimodal Interaction, ACM, Paris, 2023, pp. 689–694. doi:10.1145/3577190.3616121.

[6] Mileo A. Towards a neuro-symbolic cycle for human-centered explainability. Neurosymbolic Artificial Intelligence 1 (2025). doi:10.3233/nai-240740.

[7] Booth, A., Sutton, A., Clowes, M., & Martyn-St James, M. *Systematic Approaches to a Successful Literature Review* (3rd ed.). SAGE Publications Ltd, 2022. URL: https://us.sagepub.com/en-us/nam/systematic-approaches-to-a-successful-literature-review/book270933

[8] Landeta, J., Barrutia, J., & Lertxundi, A. Hybrid Delphi: A methodology to facilitate contribution from experts in professional contexts. *Technological Forecasting and Social Change* 78, 9 (2011), 1629-1641.

[9] Lazar, J., Feng, J. H., & Hochheiser, H. *Research methods in human-computer interaction*. Morgan Kaufmann, 2017. URL: https://books.google.com/books?hl=ca&lr=&id=hbkxDQAAQBAJ&oi=fnd&pg=PP1&dq=%22Research+methods+in+human-computer+interaction%22&ots=Sr149a078V&sig=JYrwQ0gx9GMgKqs0LOwTdxeFKfc

[10] Sarker, M. K., Zhou, L., Eberhart, A., & Hitzler, P. Neuro-symbolic artificial intelligence: Current trends. *AI Communications* 34, 3 (2022), 197–209. DOI:10.3233/AIC-210084

[11] Alshaheen, R., & Tang, R. User Experience and Information Architecture of Selected National Library Websites: A Comparative Content Inventory, Heuristic Evaluation, and Usability Investigation. *Journal of Web Librarianship* 16, 1 (2022), 31–67. DOI:10.1080/19322909.2022.2027318

[12] Martín, A. G., Fernández-Isabel, A., Martín De Diego, I., & Beltrán, M. A survey for user behavior analysis based on machine learning techniques: Current models and applications. *Applied Intelligence* 51, 8 (2021), 6029–6055. DOI:10.1007/s10489-020-02160-x

[13] Sanders, E. B. N., & Stappers, P. J. Co-creation and the new landscapes of design. *CoDesign* 4, 1 (2008), 5–18. DOI:10.1080/15710880701875068

[14] Kozinets, R. The method of netnography. In *SAGE internet research methods* (Vol. 3, pp. 101–119). SAGE, 2012. URL: https://www.torrossa.com/gs/resourceProxy?an=4913115&publisher=FZ7200#page=996

[15] Holtzblatt, K., & Beyer, H. *Contextual design: Defining customer-centered systems*. Elsevier, 1997. URL: https://books.google.com/books?hl=ca&lr=&id=JxQaQgOONGIC&oi=fnd&pg=PP2&dq=Contextual+design:+defining+customer-centered+systems&ots=-VU-uF4-A5&sig=0lgjTiygic-q3uMzVPsU5d2DfEw

https://doi.org/10.23106/dcmi.952573231

[16] Kawulich, B. B. Participant observation as a data collection method. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* 6, 2 (2005). URL: http://www.qualitative-research.net/index.php/fqs/article/view/466

[17] Ilkou, E., & Koutraki, M. Symbolic Vs Sub-symbolic Al Methods: Friends or Enemies?, 8, 2020.

[18] Hitzler, P., Bianchi, F., Ebrahimi, M., & Sarker, M. K. Neural-symbolic integration and the Semantic Web. *Semantic Web* 11, 1 (2020), 3—11. DOI:10.3233/SW-190368

[19] Hitzler, P., Eberhart, A., Ebrahimi, M., Sarker, M. K., & Zhou, L. Neuro-symbolic approaches in artificial intelligence. *National Science Review* 9, 6 (2022), nwac035. DOI:10.1093/nsr/nwac035

[20] Sheth, A., Roy, K., & Gaur, M. Neurosymbolic artificial intelligence (why, what, and how). *IEEE Intelligent Systems* 38, 3 (2023), 56–62.

[21] Gardasevic, S. User-driven efforts in creating knowledge graph information system. *Digital Library Perspectives* 36, 2 (2020), 97–111. DOI:10.1108/DLP-12-2019-0043

[22] Gardasevic, S., & Lamba, M. Statecarts in use: structured analysis and object-orientation, in: I. Sserwanga, A. Goulding, H. Moulaison-Sandy, J. T. Du, A. L. Soares, V. Hessami, & R. D. Frank (Eds.), *Information for a Better World: Normality, Virtuality, Physicality, Inclusivity* (Vol. 13972, pp. 473-490). Springer Nature Switzerland, 2023. DOI:10.1007/978-3-031-28032-0_36

[23] Gardasevic, S., & Lamba, M. Knowledge Graph for Academic Libraries: A User-Centered Design Approach, 2024. URL: https://www.ideals.illinois.edu/items/130413

[24] Gardasevic, S., & Lamba, M. “It answers questions that I didn't know I had": PhD students' evaluation of an information-sharing knowledge graph. *Digital Library Perspectives* 40, 4 (2024), 493–517. DOI:10.1108/DLP-02-2024-0025

[25] Ehn, P. Participation in design things. *Participatory Design Conference (PDC)*, Bloomington, Indiana, USA (2008), 92–101, 2008. URL: https://www.diva-portal.org/smash/record.jsf?pid=diva2:1408103

[26] Liu, F., Zhang, M., & Budiu, R. Al as a UX Assistant. Nielsen Norman Group, 2023. URL: https://www.nngroup.com/articles/ai-roles-ux/

[27] Moran, K., & Nielsen, J. Al for UX: Getting Started. Nielsen Norman Group, 2023. URL: https://www.nngroup.com/articles/ai-ux-getting-started/

https://doi.org/10.23106/dcmi.952573231