ADVANCES IN KNOWLEDGE ORGANIZATION
Proceedings of International ISKO Conference - ISSN: 0938-5495 - 2026 - São Paulo

Epistemological and practical implications of large-scale language models for the organization and representation of knowledge

Adriléia de Moura Lima

Federal University of Minas Gerais, Brazil

Elisângela Cristina Aganette

Federal University of Minas Gerais, Brazil

Danielle do Carmo Pimenta Rioga

University of São Paulo, Brazil

Abstract

The use of generative artificial intelligence has opened up new possibilities and challenges for the organization and representation of knowledge in libraries. Traditionally, activities such as cataloging, classification, and indexing are grounded in epistemological, ontological, and axiological principles that structure documentary representation. This paper discusses the epistemological and operational implications of using generative AI in these processes. A hybrid approach is adopted that combines theoretical reflection on the literature in Knowledge Organization with an applied study of CatalogIA BibTU, generative artificial intelligence models designed to support the generation of bibliographic records, MARC21 fields, and Dewey Decimal Classification notations. The results indicate that generative artificial intelligence represents not only a technical innovation but also an epistemological shift.

1.0 Introduction

The integration of generative artificial intelligence (GenAI), a tool capable of replicating human abilities such as creating stories, providing quick responses, offering advice, cataloging, and classifying (Khan et al. 2023; Selbach et al. 2023; Vijayakumar and Sakthivel 2024), has brought about structural changes in the practices and foundations of Knowledge Organization (KO) (Bourgeois and Ellingson 2025; García-López and Nájera-Ochoa 2025; Hussain 2022; Thellefsen et al. 2025), either by altering the way knowledge is represented or through interaction with users (García-López and Nájera-Ochoa 2025). Large Language Models (LLMs) are a type of machine learning model capable of performing a variety of tasks using natural language processing (NLP) (Khan et al. 2023), such as ChatGPT, Mistral, Lhama, Copilot (Stryker n.d.), among others, and are being used to support routine library tasks—namely, indexing, classification, and cataloging, as well as to automate them (Khan et al. 2023; Selbach et al. 2023; Vijayakumar and Sakthivel 2024), with the aim of improving information and knowledge retrieval (Vijayakumar and Sakthivel 2024).

These systems do more than just automate tasks (Khan et al. 2023). They introduce new ways of understanding and interpreting the world, based on probabilistic inference, dynamic contextuality, and statistical learning (García-López and Nájera-Ochoa 2025). Unlike the deterministic and normative models that have historically structured library science, such as the Dewey Decimal Classification (DDC) and MARC 21 formats, among others (Hjørland 2017). LLMs are trained on large volumes of textual data using deep neural networks and are able to capture complex patterns and relationships in natural language by semantically reorganizing large volumes of data (Khan et al. 2023).

In this context, this study analyzes the epistemological and practical implications of incorporating large language models (LLMs) into cataloging, based on the "CatalogIA BibTU" case study, developed at the Haydée Bittencourt Library of the University Theater at UFMG.

KO, as a theoretical and practical field within Information Science, focuses on the nature and quality of knowledge organization processes (Hjørland 2003, 2008), which underpin the principles of terminological, semantic, and normative classification control (Hjørland 2008, 2016). The emergence of Gen AI challenges these foundations by introducing, namely, i) probabilistic inferential processes; ii) dynamic contextual representations; and iii) learning based on statistical patterns extracted from massive datasets. In light of this, the following research question is formulated: How does the incorporation of generative language models into bibliographic classification (indexing) reconfigure the epistemological foundations and normative practices of KO in university libraries?

2.0 Theoretical Framework

2.1 Organization of Knowledge

Knowledge Organization (KO) encompasses activities such as the description, indexing, and classification of documents carried out by information units, such as libraries, archives, and information centers (Hjørland 2008). According to Lima and Maculan (2024), KO is a field that plays a proactive role in environments of communication and knowledge exchange. It is based on the processes of classifying and categorizing things.

Knowledge organization is inherently dependent on epistemological positions, since every documentary representation involves ontological and axiological choices; in other words, there is no neutrality in the organization of knowledge (Bowker and Star 1999; Hjørland 2008, 2016).This view reinforces the idea that organizing and making information available always involves human choices, and these choices are influenced by context.

Classification systems are not merely technical tools, but epistemological devices that stabilize meanings within discursive communities (Bowker and Star 1999; Hjørland 2008, 2016).

2.2 Generative artificial intelligence as an epistemic agent

Advances in LLMs have opened up new possibilities for the automation and optimization of knowledge management, enabling organizations to capture, distribute, and utilize information effectively to improve performance (Aggarwal et al. 2023; García-López and Nájera-Ochoa 2025).

Carvalho Junior et al. (2019, 13) draw on Floridi's (2023) concept of the infosphere to characterize the new informational environment emerging from the digitization and networking of GLAM collections. From this perspective, algorithms cease to be passive tools and begin to act as active mediators of semantic construction.

LLMs function as large statistical prediction models that repeatedly predict the next element in a sequence. They learn patterns in their texts and produce language that follows those patterns. They operate based on: probabilistic inference (text sequence prediction); distributed contextual modeling; and large-scale statistical learning (Stryker n.d.). This establishes a new computational epistemology, distinct from traditional documentary logic.

2.3 Algorithmic classification and ethical implications

AI-mediated classification incorporates structural biases present in the training data. Zhang et al. (2023) warn of the epistemological risks of LLMs, which can generate socially unacceptable content, racial discrimination, misinformation, and hallucinations. Similarly, Bender et al. (2021) and García-López and Nájera-Ochoa (2025) highlight the risk of opaque statistical models.

Thus, classification decisions are no longer exclusively human and begin to emerge from statistical patterns, which introduces ethical challenges related to transparency, accountability, and explainability (Bender et al. 2021; García-López and Nájera-Ochoa 2025; Zhang et al. 2023).

3.0 Methodology

This study is characterized as an exploratory-descriptive research project with a qualitative approach (Gil 2008), combining theoretical analysis with case study application (Yin 2015): the "CatalogIA BibTU" project, developed during a doctoral course and implemented at the Haydée Bittencourt Library (UFMG University Theater), focusing on assisted generation, namely: i) of the subject (classification, i.e., DDC notation class 700—Arts, 800—Literature), and ii) call number using the Cutter Table. The focus of this study is to investigate the epistemological and practical implications of incorporating large-scale language models into cataloging.

To this end, the case study selected 7 works physically available in the Library; these included works from the pre-test selected at random, and the other 5 were chosen

purposefully because they presented peculiarities, such as the work with an entry by entity, in the case of Grupo Galpão (which even has an incorrect entry in the system). Comic books, national and international authors. The two works related to the pre-test, which sought to adjust the initial prompt. However, several adjustments were necessary, according to the selected Gen AI models (23 models), and the others were used in the experiment and developed in 6 phases, namely: i) selection of Gen AI models; ii) development of prompts; iii) testing on the models; iv) evaluation and selection of models; v) classification and generation of call numbers; and vi) model selection.

For this study, the focus will be on analyzing phase 5, in which the three models that obtained the best results across all phases would be analyzed regarding the automated assignment of the subject and the generation of the respective call numbers, as well as the implications for the KO.

3.1 Procedures

i) Generation of the classification and request for justification by the models via LLMs; ii) comparison with the manual classification assigned by a specialist librarian; iii) qualitative analysis of discrepancies; and iv) identification of emerging epistemological patterns.

The evaluation criteria included: i) accuracy of subject assignment (DDC classification), i.e., correspondence between the notation suggested by the AI and the notation manually assigned by specialist librarians; ii) terminological consistency, coherence of subject suggestions and related terms with Table 2 and the DDC Hierarchy, 23rd edition; iii) dynamic contextuality, the meaning of information is not fixed. Rather, it changes according to the context in which it is used or interpreted; iv) descriptive completeness, the presence of all elements necessary for classifying the item and generating the call number, i.e., the relevance of the author/title combination according to the local standards of the BU/UFMG; and v) average response time, the model's efficiency in performing tasks (latency).

4.0 Data analysis and discussion

- Accuracy in subject assignment

To assign a classification number (the subject of the work), several factors are considered, such as: What is the work about? What is the profile of this library? (public, children's, university, etc.), who is the target audience? What is the system's language? Controlled vocabulary? Natural language? Among other factors. Therefore, this activity is subjective. It thus poses a challenge to the human classifier, which will also be a challenge for the machine.

According to Zhang *et al*. (2023), it is possible to train models so that they can assume roles (personas), placing them in specific scenarios, causing them to generate data that would not otherwise be produced. However, the authors warn that this increases the likelihood of

encountering problematic content in the output and that complex scenarios are obstacles for these models.

In the case study, it is evident that the prompts generated context-specific data. However, despite the same command being given to all models, they reacted differently, sometimes failing to meet the purpose due to the complexity of the activity. This corroborates the findings of the authors mentioned above.

As shown in Table 1, there is a wide range of classification scores assigned by the models, indicating significant variation in subject classification, some were more specific, while others were more general. Therefore, there is a high degree of alignment with the general theme. However, there is significant variation in the hierarchical specificity of the DDC. Consequently, accuracy will depend on the context. In this study, a higher level of topic specificity was chosen. For this reason, the expected result was partially satisfactory.

Table 1 Automatic generation of classification numbers using Gen AI models

MODEL	82	JUSTIFICATION	TITLE
ChatGPT	792.022	Number 792 refers to theater. The code .022 covers theatrical activities and institutions, including groups, centers, and cultural initiatives related to theater—which is precisely the case with the Galpão Cine Horto project, which promotes training and cultural activities in the performing arts.	
Copilot	792.09781	The work addresses cultural practices related to theater, with an emphasis on training, preservation, and promotion. According to the 23rd edition of the DDC, cultural activities related to the performing arts are classified under code 792, with the option for geographic specification. Since the focus is on Belo Horizonte (MG), Auxiliary Table 2 applies for geographic location: .09781.	Cine Horto Warehouse: A Cultural Experience
Gemini 2.5-pro	790.068	This document addresses the topic of Recreation Centers. Here is the topic and its classification: Recreation Centers - 790.068. Step-by-step classification guide. To arrive at this classification, I followed these steps from the 23rd edition of the DDC: The work “GALPÃO Cine Horto: a cultural action experience” describes the activities and experience of a specific cultural center. The main class is 700 (The Arts). Within this class, the subject belongs to 790 (Recreational Arts and Performances). Following the hierarchy, we look for facilities and organizations, which leads us to 790.06 (Organizations, facilities,	

MODEL	JUSTIFICATION	TITLE
82	recreation management). The more specific number is 790.068 (Recreation centers). This code includes "community centers as recreation centers." "Galpão Cine Horto" is a cultural center that promotes "cultural activities" (including "Cine" [cinema] and theater), fitting perfectly as a community center for recreation and the arts.	

Source: Survey data (2025).

This demonstrates that the model operates based on semantic similarity and statistical frequency, rather than through the explicit application of classification rules.

It is also worth noting that each model presents a different classification number and justification. A parallel can be drawn with human indexers, as it is commonly found in the literature that different indexers disagree on the subject assigned to the same document, and even that the same indexer may change the subject assigned to the document over time.

- Dynamic context

It was found that the models adjusted their suggestions based on additional information provided through prompts, such as the persona (a librarian with extensive experience), examples, details of the DDC, and the knowledge base the model should consult, as shown in Figure 1. The models also adjusted their suggestions when images of the book's cover, title page, table of contents, ISBN, and barcode were provided, thereby demonstrating dynamic contextual capabilities.

Figure 1. Prompt for running the models

<Missão>
Classificar os itens informados pelos usuários com base na 23ª edição da Classificação Decimal de Dewey (CDD), referente à classe 700.
</missão>

<Information>
1. Sua tarefa é gerar o número de classificação de acordo com a Tabela Cutter.
2. Você deve analisar as seguintes partes do item: página de rosto.
3. Os dados fornecidos pelo usuário serão:ISBN ou código de barras.

</Information>

<Persona>
Você atuará como o Melville Louis Kossuth Dewey, com cinco anos de idade, Melvil Dewey já organizava a dispensa da mãe de forma sistemática, com objetivo de organizar e recuperar facilmente os itens ali guardados. Durante a graduação, trabalhou como assistente na biblioteca da Faculdade e foi a partir desse momento que iniciou a sua carreira como bibliotecário. Nesse período visitou diferentes bibliotecas a fim de identificar os métodos que as bibliotecas deveriam usar na organização do seu acervo e na prestação de seus serviços. Dewey identificou algo importante: as bibliotecas normalmente armazenavam os livros por meio de localização fixa nas estantes, ou seja, cada livro tinha um lugar físico fixo em estantes e prateleiras específicas. Dewey percebeu como isso era custoso para as bibliotecas, pois à medida que seus acervos cresciam, cada livro precisava ser renumerado, o que custava tempo e dinheiro. Além disso, esse sistema gerava dispersão de livros com assuntos semelhantes pelo acervo, ao invés de mantê-los próximos uns dos outros.
</persona>

Source: Survey data (2025).

These characteristics differ radically from the static logic of traditional classification tables.

- Descriptive completeness

All the models evaluated provided the subject of the works (classification); as mentioned earlier, some were more general and others more specific. However, when asked to provide the call number in accordance with the standard format used by the BU/UFMG, the results were unsatisfactory. Even when provided with examples and a tutorial on how to form a call number, the models were unable to generate it satisfactorily.

This is one example where AI has not yet been able to handle all the nuances of bibliographic representation, and where the professional remains absolutely indispensable for verification and validation.

- Average response time

The experiment revealed a high rate of subject generation, that is, in terms of classification numbers, with a significant reduction in processing time. This may signal a shift from administrative tasks to more complex activities involving direct interaction with users, as

well as the creation of new information products and services. This finding is consistent with the studies by García-López and Nájera-Ochoa (2025).

- Epistemological implications

Given the above, it can be inferred that the classification performed by models using prompts is no longer based exclusively on explicit rules (established by manuals and classification tables), that is, on standards. On the other hand, LLM models are trained on large volumes of data and text, classifying documents by semantically interpreting the content rather than relying solely on fixed rules. Thus, the model calculates the probability that a document belongs to a certain class, and the process has become probabilistic, context-dependent, adaptive, and context-sensitive. As shown in Table 2.

Table 2 Classification Rules vs. LLMs

Criterion	Classification rules	LLMs
Foundation	Standards and tables	Machine learning
Process	Human interpretation	Computational inference
Structure	Rigid and hierarchical	Contextual and probabilistic
Consistency	High if the rules are followed	Depends on data quality, model training, and data volume.
Scale	Limited to human work	High processing capacity (can process thousands of documents)

Source: Survey data (2025).

In this sense, the algorithm acted as an interpretive agent. However, this interpretation is not clear (black-box model). There is a risk of inaccurate classifications and a lack of terminological control, necessitating human validation, which challenges principles of transparency in line with the studies by García-López and Nájera-Ochoa (2025) and documentary justifiability.

- Central ambiguity

Through the experiment, it was possible to identify some ambiguities in the indexing process; for example, the use of LLM speeds up the assignment of subject headings. However, the machine cannot perform critical reflection. This can lead to a loss of intellectual analysis of the document, a diminishing of the professional's critical role, the automatic reproduction of inappropriate classifications, and the generation of rework. That is, the risk is turning cataloging into a mechanical, automatic validation by the machine, without conceptual analysis. According to Table 3.

Table 3 Structural ambiguity

Scale	Potential	Risk
Efficiency	Operational agility	Reduction of critical reflection
Representation	Semantic expansion	Reinforcement of cultural biases
Mediation	Personalization	Opacity in decision-making

Source: Survey data (2025).

The fact that LLMs are capable of analyzing large volumes of documents and identifying complex relationships can improve subject indexing, uncover connections between topics, enrich metadata, and enhance information retrieval. However, the quality of the data can have an impact. It may be riddled with biases and stereotypes.

5.0 Concluding Remarks

The incorporation of generative AI into knowledge organization represents not only a technical innovation but an epistemological shift.

LLMs are capable of reconfiguring the concept of classificatory authority, introducing probabilistic regimes of representation, destabilizing the centrality of traditional documentary norms, and transforming the librarian's role into that of an epistemic supervisor of the algorithm.

The CatalogIA BibTU case study demonstrates that AI can enhance the efficiency and descriptive capacity of university libraries, but requires: human validation protocols; algorithmic transparency; critical training in algorithmic literacy; and ongoing ethical discussion.

It is concluded that AI should be understood as a hybrid epistemic agent, whose operation demands responsible information governance, especially in academic contexts.

ACKNOWLEDGMENTS

This study was supported by the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES), Brazil—Finance Code 001 (Author 1). The authors also acknowledge the National Council for Scientific and Technological Development (CNPq) for its research support (Grant No. 313645/2020-5, Author 2).

Artificial Intelligence Use Statement

Generative artificial intelligence tools, particularly ChatGPT (OpenAI), were used during the preparation of this manuscript to support language revision, academic writing refinement, translation, and textual organization. AI tools were not used to generate or fabricate research data, conduct the empirical analysis, or replace the authors' interpretation and

critical assessment of the results. The authors reviewed and validated the AI-assisted content and take full responsibility for the accuracy, integrity, and final content of the manuscript.

References

Aggarwal, Abhishek, Cheuk Chi Tam, Dezhi Wu, Xiaoming Li, and Shan Qiao. 2023. "Artificial Intelligence–Based Chatbots for Promoting Health Behavioral Changes: Systematic Review." Journal of Medical Internet Research 25: e40789. doi: 10.2196/40789.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21), 610–623. New York: Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922.

Bourgeois, John P., and Hannah Ellingson. 2025. "Ability of ChatGPT to Generate Systematic Review Search Strategies Compared to a Published Search Strategy." Medical Reference Services Quarterly 44, no. 3: 279–291. https://doi.org/10.1080/02763869.2025.2537075.

Bowker, Geoffrey C., and Susan Leigh Star. 1999. Sorting Things Out: Classification and Its Consequences. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/6352.001.0001.

Carvalho Junior, José Murilo Costa, Dalton Lopes Martins, and Leonardo Barbosa Germani. 2019. "GLAM e Instituições de Memória em Rede: uma 'Infosfera' de Floridi?" PragMATIZES – Revista Latino-Americana de Estudos em Cultura 9, no. 16: 11–30. doi: 10.22409/pragmatizes.v0i16.27529.

Floridi, Luciano. 2023. The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities. Oxford: Oxford University Press.

García-López, Iván Miguel, and Jessica Nájera-Ochoa. 2025. "Mapeo de Literatura sobre el Impacto de los LLM en la Gestión del Conocimiento: Tendencias y Desafíos." Vinculatégica EFAN 11, no. 4: 61–77. doi: 10.29105/vtga11.4-1149.

Gil, Antônio Carlos. 2008. Métodos e técnicas de pesquisa social. 6th ed. São Paulo: Atlas.

Hjørland, Birger. 2003. "Fundamentals of Knowledge Organization." Knowledge Organization 30: 87–111. https://doi.org/10.5771/0943-7444-2003-2-87.

Hjørland, Birger. 2008. "What Is Knowledge Organization (KO)?" Knowledge Organization 35: 86–101. https://doi.org/10.5771/0943-7444-2008-2-3-86.

Hjørland, Birger. 2016. "Knowledge Organization (KO)." Knowledge Organization 43: 475–484.

Hjørland, Birger. 2017. “Domain Analysis.” Knowledge Organization 44: 436–464. https://doi.org/10.5771/0943-7444-2017-6-436.

Hussain, Abid. 2022. “Use of WhatsApp Technology in Library Services: Case Study of National Defence University Library, Islamabad, Pakistan.” Library Philosophy and Practice, article 6899. https://digitalcommons.unl.edu/libphilprac/6899/.

Khan, Rahat, Nidhi Gupta, Atasi Sinhababu, and Rupak Chakravarty. 2023. “Impact of Conversational and Generative AI Systems on Libraries: A Use Case Large Language Model (LLM).” Science & Technology Libraries. Published online September 11. doi: 10.1080/0194262X.2023.2254814.

Lima, Gercina Ângela de, and Benildes Coura Moreira dos Santos Maculan. 2024. “Knowledge Universe: Classification and Categorization from the Perspective of Knowledge Organization.” RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação 22: e024017. doi: 10.20396/rdbci.v22i00.8675419/en.

Selbach, Clarissa Jesinska, Ana Paula Medeiros Magnus, Anamaria Ferreira, and Loiva Duarte Novak. 2023. “Inteligência artificial para catalogação na Biblioteca Central da PUCRS: reflexões introdutórias sobre o ChatGPT.” In Anais do 22º Seminário Nacional de Bibliotecas Universitárias (SNBU), 1–9. São Paulo: FEBAB. https://portal.febab.org.br/snbu2023/article/view/2841.

Stryker, Cole. n.d. “O que é LLM (grandes modelos de linguagem)?” IBM Think. https://www.ibm.com/br-pt/think/topics/large-language-models.

Thellefsen, Martin Muderspach, Bent Sørensen, and Amalia Nurma Dewi. 2025. “Generative AI and the Semiosic Reconfiguration of Knowledge Organization – A Preliminary Exploration.” Journal of Documentation 81, no. 5–6: 1181–1199. doi: 10.1108/JD-04-2025-0094.

Vijayakumar, S., and A. Sakthivel. 2024. “The Future of Libraries: Leveraging Artificial Intelligence for Knowledge and Information Management.” In Academic Publishing, Libraries and Artificial Intelligence, edited by M. Chandrashekara, B. M. Meera, H. R. Babu, and B. Nandeesha, 154–160. Mysuru: Department of Studies in Library and Information Science, University of Mysore.

Yin, Robert K. 2015. Estudo de caso: planejamento e métodos. 5th ed. Porto Alegre: Bookman.

Zhang, Jianyi, Xu Ji, Zhangchi Zhao, Xiali Hei, and Kim-Kwang Raymond Choo. 2023. “Ethical Considerations and Policy Implications for Large Language Models: Guiding Responsible Development and Deployment.” arXiv. doi: 10.48550/arXiv.2308.02678.

Works used in the pre-test

Arnheim, Rudolf. 2011. Arte & percepção visual: uma psicologia da visão criadora: nova versão. Translated by Ivonne Terezinha de Faria. São Paulo: Cengage Learning.

Maenz, Paul. 1974. Art déco: 1920–1940, formas entre dos guerras. Barcelona: Editorial Gustavo Gili.

Call number of the copy used in the pre-test: 7 M185a (TU).

Works used in the experiment

Berthold, Margot. 2008. História mundial do teatro. 4th ed. Translated by Maria Paula Vignola Zurawski et al. São Paulo: Perspectiva.

Brecht, Bertolt. 1959. Les affaires de monsieur Jules César (Die Geschäfte des Herrn Julius Caesar). Translated by Gilbert Badia. Paris: L'Arche.

Galpão Cine Horto: uma experiência de ação cultural. 2014. Belo Horizonte: Edições CPMT.

Shakespeare, William. 1977. Shakespeare em quadrinhos: Hamlet, Romeu e Julieta, A tempestade. Rio de Janeiro: Civilização Brasileira.

Suassuna, Ariano. 1985. Auto da compadecida. 21st ed. São Paulo: Agir.