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

Ontologies: the biggest challenge for large language models (LLMs)

Gabriel Machado

Federal University of Minas Gerais, Brazil

Arthur Matta

Federal University of Minas Gerais, Brazil

Jonas Reis

Federal University of Minas Gerais, Brazil

Aldo Faria Costa

Federal University of Minas Gerais, Brazil

Mauricio B. Almeida

Federal University of Minas Gerais, Brazil

Abstract

The current technological era is marked by undeniable advances in the broad field of Artificial Intelligence (AI), particularly in the field of Natural Language Processing (NLP). This work is a theoretical essay that addresses the challenge of large language models (LLMs) to follow organized structures of knowledge, even with a probabilistic nature. One of the main problems, hallucinations, is related to how LLMs function: they generate text based on a statistically constructed textual semantic relationship, and don't have a verifiable and representable model. There are various techniques that attempt to mitigate this problem (RAG, fine-tuning, etc.) by adding more context with current and relevant information. However, they do not guarantee semantic and factual consistency and do not have any auditable logical structure. This differs from an ontology, which is first and foremost a philosophical concept, which, when adopted by the field of Information Science (IS), becomes an area of knowledge concerned with conceptual models that represent a certain scope with rigor and precision.

1.0 Introduction

Human knowledge is a vast and extremely complex domain. Due to its interdisciplinary nature, Information Science (IS) offers a wide range of available tools that can be used to address this challenge. Although deeply interdisciplinary (Saracevic 1995), it has its own disciplinary core centered on the problems of organizing, retrieving and using information. Its enduring history and focus are not undermined by dialogue with other fields, but rather enriched by it.

Within the field of human-machine relationship (HMR), which is currently of great prominence, is natural language processing (NLP). It consists of a set of techniques designed to enable computers to process information by analyzing data resulting from linguistically representable human communication (Falcão et al. 2022). This prominence stems from the emergence of large language models (LLMs), whose ability to process and generate text is unprecedented in the history of NLP.

Trained on a colossal amount of data, LLMs are statistical models that capture complex linguistic relationships by generating a representation of how words (tokens) relate within a context window. The main concept behind this, known as multi-head attention (Vaswani et al. 2017), while very powerful for processing linguistic data, is also limited to this same linguistic relationship, without creating a formal and measurable representation of the knowledge and information being expressed.

Large Language Models (LLMs) have become one of the most influential technologies in contemporary Artificial Intelligence research and applications, impacting domains such as information retrieval, knowledge representation, healthcare, education, and scientific discovery. Their growing relevance is evidenced not only by their widespread adoption in academia and industry, but also by the remarkable increase in scientific publications dedicated to the topic. A search conducted in the Web of Science database covering the last five years retrieved approximately 46,600 articles related to LLMs, demonstrating the intense and rapidly expanding interest of the scientific community in this field.

Ontology is a central pillar in the process of knowledge representation. As it is the study of things that exist and the relationships among them, something fundamental to any attempt at knowledge representation. According to Guarino (1995), there are two ways of viewing ontology: the first is ontology as a discipline, which explores the philosophical concept of creating conceptual models that represent a given scope; the second is ontology as a computational artifact used as a resource to aid in the creation of information systems.

This article is a theoretical essay that investigates the integration of formal ontologies into Large Language Models (LLMs) to mitigate the inherent limitations of probabilistic technology. Although LLMs process natural language with unprecedented fluency, their statistical nature lacks a verifiable knowledge model, frequently resulting in hallucinations: information that is linguistically coherent but factually incorrect.

Throughout this work, the second section provides a preliminary review of the literature, first addressing the emergence of the study of ontologies; and, then, a review on relevant issues in the field of machine learning, ranging from simple neural networks to the transformer architecture used by LLMs and a definition of hallucination. Finally, the third section presents a conclusion on the need for integration of formal knowledge representation mechanisms within LLMs.

2.0 Theoretical framework

2.1 LLMs

What began as an experimental investigation into representing neuron behavior via mathematical models (McCulloch and Pitts 1943) laid the foundation for the deep learning architectures driving Natural Language Processing (NLP). Early architectures, such as the Multilayer Perceptron (MLP) (Rumelhart et al. 1986), proved structurally inefficient for text processing due to their inability to capture the temporal and sequential nature of language. Consequently, Recurrent Neural Networks (RNNs) (Elman 1990) emerged as a viable alternative for modeling text sequences; however, they suffered from the vanishing gradient problem, which severely hindered their capacity to retain long-term linguistic dependencies. To circumvent this limitation, Long Short-Term Memory (LSTM) networks (Hochreiter and Schmidhuber 1997) were introduced, establishing a new standard for long-sequence tasks through specialized memory gating mechanisms. Nevertheless, LSTMs inherently processed tokens sequentially, a bottleneck that precluded parallel computation and made training on expansive datasets highly resource intensive.

Concurrently with the evolution of sequential neural architectures, the NLP field underwent a paradigm shift in how words were mathematically represented, transitioning from sparse matrix models to dense, continuous word embeddings. Systems like Word2Vec (Mikolov et al. 2013) and GloVe (Pennington et al. 2014) revolutionized semantic processing by mapping discrete vocabulary tokens into low-dimensional vector spaces, where geometric proximity reflected semantic and syntactic similarity. These distributed representations allowed algorithms to capture rich linguistic regularities and overcome the computational limitations of earlier statistical models. Despite their profound utility, these static embeddings struggled to resolve polysemy, as they assigned a single fixed vector to each word regardless of its surrounding context. This limitation necessitated the development of dynamic, contextualized embedding systems, such as ELMo (Peters et al. 2018), which ultimately laid the conceptual groundwork for deeper, sequence-aware language models.

The sequential bottleneck of LSTMs (even though there were proposals for architectures to solve this problem like Schuster and Paliwal 1997) and the need for dynamic contextual representation were definitively resolved in 2017 when Vaswani et al. introduced the Transformer architecture (Vaswani et al. 2017). By completely abandoning recurrent units in favor of multi-head attention mechanisms, the Transformer enabled the simultaneous, parallel processing of entire text sequences. This extreme parallelization exponentially reduced training times and facilitated the ingestion of colossal datasets, culminating in the development of large-scale pre-trained architectures like the bidirectional BERT and the autoregressive GPT series. Achieving unprecedented performance, consistency, and stability across diverse NLP tasks, these structurally massive transformer networks are now collectively defined as Large Language Models (LLMs).

2.1.1 Hallucinations: causes and management

LLMs have proven to be a revolution in NLP problems, but they have a big issue in information retrieval. In the worst scenario, they don't have access to data that was not available during training. In this and other cases, they prove to be inaccurate and prioritize conversational fluency, without even knowing whether they are factually correct. This

phenomenon, known as hallucination, occurs in cases where LLMs produce fluent and coherent outputs that are factually incorrect or fabricated (Huang et al. 2025; Tonmoy et al. 2024; Ji et al. 2023). One of the main problems, hallucinations, relates to the actual functioning of LLMs, which generate text based on statistically constructed textual semantic relationships, not on verifiable and representable data sources (Huang et al. 2025; Farquhar et al. 2024).

They arise from probabilistic predictions based on linguistic patterns from training, prioritizing terminological correlations over factual accuracy, due to limitations in data (misinformation, biases, knowledge boundaries), training (pre-training, SFT, RLHF), and inference (imperfect decoding, overconfidence) (Huang et al. 2025; Tonmoy et al. 2024; Anh-Hoang et al. 2025).

These issues are mitigable but not eliminable through RAG for external grounding, improved RLHF, contrastive decoding, fine-tuning with reliable data, and semantic verification, reducing their incidence without total eradication (Huang et al. 2025; Tonmoy et al. 2024; Ji et al. 2023).

2.1.2 LLMs in knowledge organization

The relevance of LLMs is also reflected in leading Information Science journals, which have increasingly incorporated discussions on the topic over the last five years. In the journal Knowledge Organization, four publications addressing LLM-related issues were identified, ranging from questions of authorship and provenance (Soos and Haroutunian 2024; Cheng and Choi 2025), to the translation of SPARQL queries using LLMs (Rasheed and Aguado 2025), and bibliometric analyses supported by LLMs (Clunis and Asare 2025).

Similarly, Journal of the Association for Information Science and Technology (JASIST) presented twelve references in the same period, covering topics such as authorship and evaluation (Wagner and Jiang 2025; Lund et al. 2023; W. Wu et al. 2025; Thelwall 2025), media and social networks (Zheng et al. 2026; Yang et al. 2026), information retrieval (Hjørland 2026), the Web (Granitzer et al. 2024), classification (D. Wu et al. 2025; M. Wu et al. 2025), query decomposition (Zheng et al. 2026), and the consequences of information gatekeeping through algorithmic technologies (Potnis et al. 2025).

These studies demonstrate that LLMs are no longer restricted to Computer Science research, but have become a significant interdisciplinary topic within Information Science itself.

2.2 Ontology: a field among fields

In the last century, some scholars began to explore ontologies in greater depth. In 1993, Stanford University professor Thomas R. Gruber wrote the article “What is an Ontology?” (Gruber 1993), which became known for bringing this discussion into sharp focus. Ontology, as an academic field, is a highly contested concept across different disciplines;

however, it is essential to computing because it is how we represent knowledge that can be shared with computers. These devices can only execute logical commands and have no mechanism for handling knowledge.

2.2.1 Ontology in Philosophy

In philosophy, ontology is a term related to categories: which ones exist, how they are created, how to distinguish them, among other questions. Given their influence on contemporary theories, at least three systems of categories must be analyzed to understand the complexity of philosophical ontology: the Aristotelian, the Kantian, and the Husserlian (Thomasson 2022). Explaining these systems is beyond the scope of this project, but a summary overview can be found in Almeida (2020).

2.2.2 Ontologies in Computer Science

Although the term "ontology" has various uses, two main meanings can be identified in computer science. The first refers to the use of ontological principles to understand and model reality (Wand and Weber 1990; Wand et al. 1999). The use of the term in this case aligns with its original role in philosophy, namely, to provide a description of what exists and to characterize entities in modeling activities. The second meaning concerns the representation of a domain in a computational representation language (Staab and Studer 2013). An ontology, in this case, consists of a set of statements expressed in a representation language, which can be processed by automated inference mechanisms.

2.2.3 Ontologies in Information Science

Software engineering projects involving ontologies show parallels with theories in Information Science, such as faceted classification, controlled vocabularies, and lexicography (Vickery 1997). There are differences and similarities between ontologies in Computer Science and two other common terms in Information Science: taxonomy and thesauri. The ability to constrain natural language appears to be the point of contact between these two types of structures (Gilchrist 2003). Furthermore, even though the term ontology is sometimes used as a synonym for controlled vocabularies, the two structures have different objectives (Currás 2010).

3.0 Conclusions

Large Language Models (LLMs) represent a significant advance in Natural Language Processing, demonstrating an unprecedented capacity to generate fluent and contextually appropriate text. Nevertheless, their operation remains fundamentally based on statistical regularities extracted from large corpora rather than on explicit and verifiable representations of knowledge. As a consequence, problems such as hallucinations,

factual inconsistencies, and the absence of transparent reasoning mechanisms remain persistent limitations.

In contrast, ontologies provide formal, explicit, and logically structured representations of a domain, enabling the specification of concepts, relations, constraints, and inference mechanisms that can be independently verified and audited. From the perspective of Information Science and Knowledge Organization, ontologies offer precisely the type of semantic structure that is absent from contemporary LLM architectures.

For this reason, the integration of LLMs and ontologies should not be understood merely as a strategy for reducing hallucinations, but as an opportunity to combine complementary strengths. While LLMs excel at processing and generating natural language, ontologies contribute semantic rigor, conceptual clarity, and logical consistency. Approaches that combine probabilistic language models with ontology-based knowledge representations may enable systems capable of producing responses that are not only linguistically plausible but also semantically grounded and explainable.

Therefore, the challenge is not whether ontologies should be incorporated into AI systems, but how such integration can be achieved effectively. Future research should investigate hybrid architectures in which ontological knowledge participates directly in information retrieval, reasoning, validation, and response generation processes. Such a direction may contribute to overcoming some of the current limitations of LLMs and to the development of AI systems that are both more reliable and more consistent with the principles of knowledge organization.

AI Usage Statement: In this article, an LLM (Gemini 3.1 Pro) was used to assist the authors with language refinement, including translation and stylistic revision, in order to improve the academic English of the manuscript. All AI-generated or AI-modified content was reviewed and verified by the authors, who take full responsibility for the integrity and accuracy of the manuscript.

References

Almeida, Mauricio Barcellos. 2020. Ontologias em Ciência da Informação: teoria e método. Curitiba: CRV.

Anh-Hoang, Dang, Vu Tran, and Le-Minh Nguyen. 2025. "Survey and Analysis of Hallucinations in Large Language Models: Attribution to Prompting Strategies or Model Behavior." Frontiers in Artificial Intelligence 8: 1622292. https://doi.org/10.3389/frai.2025.1622292.

Cheng, Yi-Yun, and Inkyung Choi. 2025. "From Editorial Records to Structured Provenance Information: Documenting Warrant in Knowledge Organization Systems Using Large Language Models." Knowledge Organization 52, no. 3: 39046. https://doi.org/10.31083/KO39046.

Clunis, Julaine, and Eric Asare. 2025. "A Bibliometric Analysis of AI-Driven Healthcare Literature Containing KOS Keywords: Trends, Themes, and Gaps." Knowledge Organization 52, no. 5: 47054. https://doi.org/10.31083/KO47054.

Currás, Emilia. 2010. Ontologies, Taxonomies and Thesauri in Systems Science and Systematics. Oxford: Chandos Publishing.

Elman, Jeffrey L. 1990. "Finding Structure in Time." Cognitive Science 14: 179–211. https://doi.org/10.1207/s15516709cog1402_1.

Falcão, Luander Cipriano de Jesus, Brenner Lopes, and Renato Rocha Souza. 2022. "Absorção das tarefas de processamento de Linguagem Natural (NLP) pela Ciência da Informação (CI): uma revisão da literatura para tangibilização do uso de NLP pela CI." Em Questão 28, no. 1: 13–34. https://doi.org/10.19132/1808-5245281.13-34.

Farquhar, Sebastian, Jannik Kossen, Lorenz Kuhn, and Yarin Gal. 2024. "Detecting Hallucinations in Large Language Models Using Semantic Entropy." Nature 630: 625–630. https://doi.org/10.1038/s41586-024-07421-0.

Gilchrist, Alan. 2003. "Thesauri, Taxonomies and Ontologies—an Etymological Note." Journal of Documentation 59: 7–18. https://doi.org/10.1108/00220410310457984.

Granitzer, Michael, Stefan Voigt, Noor Afshan Fathima, Martin Golasowski, Christian Guetl, Tobias Hecking, Gijs Hendriksen, et al. 2024. "Impact and Development of an Open Web Index for Open Web Search." Journal of the Association for Information Science and Technology 75: 512–520. https://doi.org/10.1002/asi.24818.

Gruber, Thomas. 1993. Ontology. Encyclopedia of Database Systems, edited by Ling Liu and M. Tamer Özsu. Springer-Verlag, 2009.

Hjørland, Birger. 2026. "Information Retrieval or Document Retrieval? Terminological Confusions and Unrealistic Goals in Information Science, Exemplified in Relation to Generative Artificial Intelligence." Journal of the Association for Information Science and Technology 77: 714–726. https://doi.org/10.1002/asi.70057.

Hochreiter, Sepp, and Jürgen Schmidhuber. 1997. “Long Short-Term Memory.” Neural Computation 9: 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Huang, Lei, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, et al. 2025. “A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions.” ACM Transactions on Information Systems 43, no. 2: 42:1–42:55. https://doi.org/10.1145/3703155.

Ji, Ziwei, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. “Survey of Hallucination in Natural Language Generation.” ACM Computing Surveys 55, no. 12: 248:1–248:38. https://doi.org/10.1145/3571730.

Licklider, J. C. R. 1960. “Man-Computer Symbiosis.” IRE Transactions on Human Factors in Electronics HFE-1: 4–11. https://doi.org/10.1109/THFE2.1960.4503259.

Lund, Brady D., Ting Wang, Nishith Reddy Mannuru, Bing Nie, Somipam Shimray, and Ziang Wang. 2023. “ChatGPT and a New Academic Reality: Artificial Intelligence-Written Research Papers and the Ethics of the Large Language Models in Scholarly Publishing.” Journal of the Association for Information Science and Technology 74: 570–581. https://doi.org/10.1002/asi.24750.

McCulloch, Warren S., and Walter Pitts. 1943. “A Logical Calculus of the Ideas Immanent in Nervous Activity.” The Bulletin of Mathematical Biophysics 5: 115–133. https://doi.org/10.1007/BF02478259.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. “Efficient Estimation of Word Representations in Vector Space.” arXiv:1301.3781. https://doi.org/10.48550/arXiv.1301.3781.

Pennington, Jeffrey, Richard Socher, and Christopher Manning. 2014. “GloVe: Global Vectors for Word Representation.” In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), edited by Alessandro Moschitti, Bo Pang, and Walter Daelemans, 1532–1543. Doha, Qatar: Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1162.

Peters, Matthew E., Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. “Deep Contextualized Word Representations.” In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), edited by Marilyn Walker, Heng Ji, and Amanda Stent, 2227–2237. New Orleans, LA: Association for Computational Linguistics. https://doi.org/10.18653/v1/N18-1202.

Pettigrew, Karen E., and Lynne (E. F.) McKechnie. 2001. “The Use of Theory in Information Science Research.” Journal of the American Society for Information Science and Technology 52: 62–73. https://doi.org/10.1002/1532-2890(2000)52:1%3C62::AID-ASI1061%3E3.0.CO;2-J.

Potnis, Devendra, Iman Tahamtan, and Luke McDonald. 2025. “Negative Consequences of Information Gatekeeping through Algorithmic Technologies: An Annual Review of Information Science and Technology (ARIST) Paper.” Journal of the Association for Information Science and Technology 76: 262–288. https://doi.org/10.1002/asi.24955.

Rasheed, Mohammed H., and Marina Aguado. 2025. “LLM-Based Natural Language to SPARQL Translation over Domain-Specific Knowledge Graph.” Knowledge Organization 52, no. 8: 42705. https://doi.org/10.31083/KO42705.

Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. 1986. “Learning Representations by Back-Propagating Errors.” Nature 323: 533–536. https://doi.org/10.1038/323533a0.

Saracevic, Tefko. 1995. “Interdisciplinary Nature of Information Science.” Ciência da Informação 24, no. 1: 36–41.

Schuster, M., and K. K. Paliwal. 1997. “Bidirectional Recurrent Neural Networks.” IEEE Transactions on Signal Processing 45: 2673–2681. https://doi.org/10.1109/78.650093.

Soos, Carlin, and Levon Haroutunian. 2024. “On the Question of Authorship in Large Language Models.” Knowledge Organization 51: 83–95. https://doi.org/10.5771/0943-7444-2024-2-83.

Staab, Steffen, and Rudi Studer, eds. 2013. Handbook on Ontologies. Berlin: Springer.

Thelwall, Mike. 2025. “ChatGPT for Complex Text Evaluation Tasks.” Journal of the Association for Information Science and Technology 76: 645–648. https://doi.org/10.1002/asi.24966.

Thomasson, Amie. 2022. “Categories.” In The Stanford Encyclopedia of Philosophy, Winter 2022 Edition, edited by Edward N. Zalta and Uri Nodelman. Stanford, CA: Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2022/entries/categories/.

Tonmoy, S. M. Towhidul Islam, S. M. Mehedi Zaman, Vinija Jain, Anku Rani, Vipula Rawte, Aman Chadha, and Amitava Das. 2024. “A Comprehensive Survey of Hallucination Mitigation Techniques in Large Language Models.” arXiv:2401.01313. https://doi.org/10.48550/arXiv.2401.01313.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. “Attention Is All You Need.” In Advances in Neural Information Processing Systems 30, edited by Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, 5998–6008.

Vickery, B. C. 1997. “Ontologies.” Journal of Information Science 23: 277–286. https://doi.org/10.1177/016555159702300402.

Wagner, Christian, and Ling Jiang. 2025. “Death by AI: Will Large Language Models Diminish Wikipedia?” Journal of the Association for Information Science and Technology 76: 743–751. https://doi.org/10.1002/asi.24975.

Wand, Yair, Veda C. Storey, and Ron Weber. 1999. “An Ontological Analysis of the Relationship Construct in Conceptual Modeling.” ACM Transactions on Database Systems 24: 494–528. https://doi.org/10.1145/331983.331989.

Wand, Yair, and Ron Weber. 1990. “Mario Bunge’s Ontology as a Formal Foundation for Information Systems Concepts.” In Studies on Mario Bunge’s Treatise, edited by P. Weingartner and G. Dorn, 123–150. Amsterdam: Rodopi.

Wu, Dengsheng, Huidong Wu, Fan Meng, and Jianping Li. 2025. “Classifying Ultra-Short Scientific Texts Using a Hybrid Hierarchical Multi-Label Classification Framework.” Journal of the Association for Information Science and Technology 76: 1625–1646. https://doi.org/10.1002/asi.70018.

Wu, Mengjia, Gunnar Sivertsen, Lin Zhang, Fan Qi, and Yi Zhang. 2025. “Scaling Research Aim Identification: Language Models for Classifying Scientific and Societal-Oriented Studies.” Journal of the Association for Information Science and Technology 76: 1470–1487. https://doi.org/10.1002/asi.70004.

Wu, Wenqing, Chengzhi Zhang, and Yi Zhao. 2025. “Automated Novelty Evaluation of Academic Paper: A Collaborative Approach Integrating Human and Large Language Model Knowledge.” Journal of the Association for Information Science and Technology 76: 1452–1469. https://doi.org/10.1002/asi.70005.

Yang, Li, Jia Yang, Jin-Cheon Na, and Jianfei Yu. 2026. “Annotating Aspects in Text and Image: A New Task and Dataset for Multimodal Aspect-Based Sentiment Analysis.” Journal of the Association for Information Science and Technology 77: 771–794. https://doi.org/10.1002/asi.70026.

Zheng, Er-Te, Hui-Zhen Fu, Mike Thelwall, and Zhichao Fang. 2026. “Can Social Media Provide Early Warning of Retraction? Evidence from Critical Tweets Identified by Human Annotation and Large Language Models.” Journal of the Association for Information Science and Technology 77: 624–639. https://doi.org/10.1002/asi.70028