K-GRASP: Tacit Knowledge Externalisation with Language Models and Knowledge Graphs

Rafael Luna[1], Gabriel Luna[1], Carlos Eduardo Barbosa[1,2], Matheus Argôlo[1], Lucas Nóbrega[1], Luiz Felipe Martinez[1] and Jano de Souza[1]

[1]Systems Engineering and Computer Science Program of the Alberto Luiz Coimbra Institute for Graduate Studies and Research in Engineering, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

[2]Naval Systems Analysis Centre, Brazilian Navy, Rio de Janeiro, Brazil

rafaelluna@cos.ufrj.br
gabrielluna@cos.ufrj.br
eduardo@cos.ufrj.br
matheusargolo@cos.ufrj.br
lucasnobrega@cos.ufrj.br
felipem@cos.ufrj.br
jano@cos.ufrj.br

Abstract: The externalisation of tacit knowledge, defined as practical, experiential knowledge that is difficult to formalise, represents one of the greatest challenges faced by institutions that rely on the expertise of their members. Due to its subjective and contextual nature, this type of knowledge resists capture through traditional methods such as interviews and observations, which are costly, difficult to standardise, and limited in reusability. This work proposes an architecture called K-GRASP (Knowledge Graph-based Retrieval-Augmented Structured Prompting), which combines the representational power of knowledge graphs (KG) with the probabilistic reasoning capabilities of Large Language Models (LLMs) to address this challenge. The proposed solution is divided into two phases: (i) a capture phase, in which an interviewer LLM conducts structured sessions with experts and converts their responses into RDF triples stored in a semantic graph, and (ii) a retrieval phase, in which a consultative LLM uses Retrieval-Augmented Generation (RAG) to translate natural language questions into SPARQL queries (SPARQL Protocol and RDF Query Language), retrieve relevant subgraphs, and generate contextualized responses. The architecture enables the systematic, reusable, and accessible codification of tacit knowledge, allowing for its large-scale preservation and dissemination. By integrating LLMs with formal representation structures, K-GRASP offers a robust, scalable, and interpretable solution to a historically complex problem in knowledge management. In the capture phase, prompts can be shaped to elicit concrete cases, boundary conditions, and decision rationales. At the same time, responses are mapped to an agreed-upon vocabulary to reduce drift across sessions. The graph can also retain provenance and scope cues (e.g., source expert, date, and stated assumptions), which may support later review and incremental refinement. In the retrieval phase, returning both the synthesised answer and the underlying triples can make the consultation more transparent and highlight gaps or ambiguities, as well as operational concerns such as access control, privacy, and versioning.

Keywords: Tacit knowledge externalization, Knowledge graphs, Retrieval-augmented generation, SPARQL protocol, RDF query language

1. Introduction

A significant portion of valuable knowledge within organisations, academic institutions, and technical environments is tacit. This type of knowledge is grounded in practical experience, intuition, and skills developed over time. By its very nature, it is difficult to formalise, articulate, or explicitly document. As Polanyi (1966) discusses, individuals often apply knowledge they are unable to express verbally. This characteristic makes the capture, organisation, and dissemination of tacit knowledge one of the main challenges in knowledge management.

In this context, the SECI model, proposed by Nonaka and Takeuchi (1995), stands out as one of the most influential theoretical frameworks for explaining the conversion between tacit and explicit knowledge. Among its four modes, externalisation, defined as the transformation of tacit knowledge into explicit forms, is particularly difficult to operationalise. It requires translating subjective experiences into structured and reusable representations. However, traditional methods such as interviews and field observations pose limitations in terms of standardisation, operational cost, and scalability (Dalkir, 2017). However, cultural and strategic misalignments remain significant obstacles to the effectiveness of KM initiatives, even when supported by advanced infrastructures (Monteiro et al., 2025).

Recent advances in Large Language Models (LLMs) present a concrete opportunity to automate parts of this process. These models exhibit a high degree of fluency in natural language and are capable of conducting complex interactions, identifying implicit patterns, and transforming open-ended responses into structured representations. LLMs' impact lies in transforming the very nature of knowledge-intensive tasks, reshaping how expertise is articulated and reused across domains, beyond mere automation (Babashahi et al., 2024; Lima et al., 2024).

To address this challenge, this work introduces the K-GRASP (Knowledge Graph-based Retrieval-Augmented Structured Prompting) methodology. This hybrid architecture combines the probabilistic reasoning capabilities of LLMs with the semantic structure of knowledge graphs (KGs). The solution is composed of two complementary flows: a capture phase, in which an interviewer LLM conducts structured sessions with experts and converts their responses into RDF triples; and a retrieval phase, in which a consultative LLM uses Retrieval-Augmented Generation (RAG) techniques to interpret natural language queries, query the semantic graph using SPARQL (SPARQL Protocol and RDF Query Language), and generate contextualized responses based on the retrieved information.

This work is organised as follows: Section 2 reviews related work. Section 3 presents the theoretical foundations related to tacit knowledge, LLMs, and KGs. Section 4 describes the K-GRASP architecture in detail. Section 5 provides a critical discussion of its applicability and limitations. Finally, Section 6 presents the conclusions and future research directions.

2. Related Work

Several recent initiatives have addressed the challenge of externalising tacit knowledge, particularly in organisational settings that depend heavily on human expertise. The works discussed in this section converge on that objective by integrating LLMs with symbolic knowledge representations, such as ontologies and semantic graphs. These approaches aim not only to capture and formalise tacit knowledge but also to make it reusable and efficiently retrievable.

Traditional knowledge management literature emphasises qualitative methods such as structured interviews, direct observation, and ethnographic analysis (Martin, 2002; Nonaka and Takeuchi, 1995; Polanyi, 1966). Although effective, these techniques are time- and resource-intensive, with limited scalability and standardisation. In response, a growing number of studies have explored computational solutions that combine LLMs with symbolic mechanisms capable of representing and operating on tacit knowledge in a scalable and interpretable manner (Bommasani et al., 2021; OpenAI, 2023).

One such study is presented by Bendiken (2024), who introduces the KNOW ontology (Knowledge Navigator Ontology for the World), designed to structure everyday knowledge pragmatically. The ontology comprises broad conceptual categories, including social, spatiotemporal, and abstract dimensions and allows captured content to be formally represented as RDF triples for semantic retrieval. The author advocates for the convergence of symbolic representations and statistical learning, promoting a hybrid approach to knowledge generation that emphasises interpretability and reusability. This line of reasoning aligns with the architecture proposed in this work, which also integrates LLMs and KGs into RAG-based consultative flows.

Building on this hybridisation perspective, Lu et al. (2024) propose the Logic-Augmented Generation (LAG) model, which explicitly integrates Semantic Knowledge Graphs (SKGs) with LLMs to enhance the logical and factual consistency of generated content. While LLMs capture implicit patterns in language and reasoning, SKGs provide a symbolic layer of verifiable knowledge that grounds the output in structured facts. This hybrid architecture is especially valuable in domains where precision and transparency are critical, such as technical consulting and decision support. The core ideas behind LAG inspire the use of RDF triples and SPARQL queries in this work, enabling structured semantic grounding for LLM-based generation.

The convergence between automated capture and structured representation is also reflected in the work of Zhou et al. (2024), which introduces the concept of a Code Digital Twin. This system is designed to capture and preserve tacit knowledge associated with the development and maintenance of complex software systems. By combining LLMs, static analysis tools, and expert interviews, the framework aims to reconstruct the rationale behind design decisions and architectural choices. Although the application is focused on software engineering, the authors highlight the importance of transforming experiential knowledge, often passed on informally, into structured, formal representations. This objective resonates with the initial phase of this work, in which an interviewer LLM captures expert knowledge and represents it in semantic graphs.

In the context of high-stakes domains such as cybersecurity, Wang et al. (2024) explore the formation of human-machine co-teams to support adaptive and resilient operations. Their study proposes using LLMs to capture the tacit knowledge of human analysts in real time while they execute critical tasks. The effectiveness of such systems depends on the models' ability to understand, adapt to, and reuse operational expertise. This emphasis on dynamic reuse aligns with the goals of this work, which seeks to structure such knowledge as a graph that can be queried in natural language, thereby promoting precision and accessibility.

In the field of education, Sun et al. (2023) present Novobo, an AI agent designed to promote the sharing of tacit pedagogical knowledge among teachers. By positioning the agent as a learner, the system encourages educators to articulate and explain their instructional practices. This interaction functions as a mechanism for externalising practical knowledge that would otherwise remain implicit. The approach demonstrates how AI-mediated conversational strategies can support the formalisation of experiential expertise. This principle is directly aligned with this work's use of interviewer LLMs to transform expert discourse into RDF triples, which are later employed in RAG-based consultative systems.

Overall, the reviewed literature reflects a strong interest in integrating LLMs with structured knowledge mechanisms to externalise and operationalise tacit knowledge. However, many of these studies focus on narrow domains and do not provide a complete pipeline encompassing capture, formalisation, and retrieval. This work advances the state of the art by proposing a modular architecture that unifies structured elicitation, RDF triple generation, and RAG-based semantic retrieval. By combining these components into a cohesive, reusable framework, this work contributes to the development of a more general and adaptable solution for tacit knowledge externalisation across diverse domains.

3. Theoretical Background

The theoretical foundation of this work is structured around five core pillars: tacit and explicit knowledge, externalisation processes, KGs, LLMs, and the RAG technique. These components provide the conceptual basis for the K-GRASP architecture, which is designed to formalise and retrieve tacit knowledge using semantic structures and neural models. Each topic is presented below with support from the specialised literature, highlighting its relevance to the context of this work.

3.1 Tacit and Explicit Knowledge

Human knowledge is commonly categorised into two complementary types: explicit and tacit knowledge. Explicit knowledge refers to information that can be formally articulated, documented, and easily shared. It is typically codified in manuals, reports, databases, and other structured formats, enabling straightforward dissemination among individuals and organisations (Stollenwerk, 2001).

Tacit knowledge, in contrast, is deeply rooted in personal experience, skills, intuition, and contextual understanding. It is inherently subjective and difficult to formalise, as it often relies on unconscious processes and practical know-how that are not easily verbalised or codified (Stollenwerk, 2001).

The strategic significance of tacit knowledge has been extensively discussed in the literature. Nonaka and Takeuchi's SECI model highlights the dynamic conversion between tacit and explicit knowledge as a key driver of organisational knowledge creation (Nonaka and Takeuchi, 1995). Subsequent studies have confirmed the central role of tacit knowledge in domains requiring expertise and judgment, such as technical, scientific, and professional settings (Dalkir, 2017; Martin and Hecker, 2018). Despite its importance, the formalisation and systematic capture of tacit knowledge remain major challenges within knowledge management (Bratianu and Orzea, 2010).

3.2 Knowledge Externalisation

Knowledge externalisation is the process through which tacit knowledge is made explicit, allowing it to be documented, shared, and reused. This process is a key component of the knowledge creation cycle, as it enables individual insights and expertise to be communicated and institutionalised within organisations.

Traditional methods of externalisation include structured interviews, direct observation, and participatory techniques, which have been widely applied in qualitative research (Davenport and Prusak, 1998; Tiwana, 2000). These approaches focus on capturing experts' personal experiences, skills, and contextual knowledge, transforming them into formats that can be stored and disseminated.

With advances in computational techniques, artificial intelligence has increasingly been applied to support knowledge externalisation. Language models, including LLMs, can assist in guiding interviews, detecting implicit patterns in expert responses, and converting natural language expressions into structured representations (Bommasani et al., 2021). These approaches aim to enhance the efficiency, consistency, and scalability of the externalisation process.

3.3 Knowledge Graphs

Knowledge Graphs are structured semantic representations that organise information as <subject, predicate, object> triples, capturing entities and the relationships between them (Berners-Lee et al., 2001). Beyond simple data storage, KGs encode semantic meaning, enabling machines to interpret, reason over, and infer knowledge from the stored information.

KGs have been widely adopted across diverse domains, including biomedical research, Industry 4.0 applications, and recommender systems, due to their ability to represent complex relationships and support automated reasoning (Hogan et al., 2021; Ji et al., 2021). In recent years, the integration of KGs with neural models has enabled hybrid approaches that combine symbolic reasoning with statistical learning. Such integration enhances factual verification, explainability, and structured retrieval of information (Das and others, 2021). Additionally, KGs are interoperable with standardised ontologies and query languages such as SPARQL, making them suitable building blocks for architectures like K-GRASP, which require structured, machine-readable representations of knowledge.

3.4 Large Language Models (LLMs)

Large Language Models are deep learning models, typically based on transformer architectures, trained on massive volumes of textual data from diverse sources (Brown et al., 2020; Radford et al., 2019). Their primary capability is to predict token sequences with coherence and contextual relevance, making them suitable for a variety of natural language processing tasks, including text generation, translation, summarisation, and question answering.

In recent years, LLMs have been increasingly employed in applications such as semantic search systems, chatbots, and knowledge automation platforms. When combined with external knowledge sources, such as Knowledge Graphs or document corpora, LLMs can overcome limitations related to fixed parametric memory and biases inherent in their training data (Lewis et al., 2020; Xu et al., 2023). This integration of neural language models with structured knowledge representations is a key component of hybrid architectures such as K-GRASP, enabling contextualised and accurate information retrieval and generation.

3.5 Retrieval-Augmented Generation (RAG)

Retrieval-Augmented Generation is a technique that integrates external information retrieval mechanisms with generative language models. Rather than relying solely on the model's internal parameters, RAG actively searches for relevant information in structured sources, such as Knowledge Graphs, or unstructured sources, such as text corpora, before generating a response (Lewis et al., 2020). This approach enhances factual accuracy and traceability while reducing the likelihood of hallucinations in generated outputs.

RAG has been applied in increasingly complex contexts, including relational databases, formal ontologies, and Knowledge Graphs (Liu et al., 2023; Sun et al., 2023; Xu et al., 2023). Within the K-GRASP architecture, RAG is employed to interpret natural language queries, translate them into SPARQL queries, retrieve relevant subgraphs, and generate contextualised answers based on the extracted knowledge.

4. Methodology

This section presents the methodological foundations and technical components of the K-GRASP architecture, designed for the externalisation and retrieval of tacit knowledge through the integration of LLMs and KGs. The methodology is structured into four main stages: (i) capturing tacit knowledge through structured interviews with an LLM, (ii) formal representation of the responses as RDF triples, (iii) construction and storage of a semantic KG, and (iv) knowledge retrieval using the RAG mechanism. Each of these phases is described below, with emphasis on its function, purpose, and technical advantages. An overview of the architecture is presented in Figure 1.

4.1 Capture Phase

The capture of tacit knowledge begins with structured dialogues led by an LLM interviewer, which differ from free-form interactions in that they rely on prompts aligned with a domain-specific ontology. This design ensures that expert discourse remains consistent, reduces ambiguity, and facilitates later mapping into semantic structures. Beyond asking predefined questions, the LLM iteratively adapts, generating follow-ups based on prior responses to elicit justifications, contextual nuances, and decision-making rationales—elements that are typically difficult to access through traditional interviews. Methodologically, this strategy reflects the " externalisation " stage of the SECI model, as it transforms experiential knowledge into explicit, transferable representations.

By combining ontological scaffolding with the generative adaptability of LLMs, this phase achieves a balance between standardisation and expressiveness. Experts retain the freedom to articulate insights in their own words, while the conversational protocol provides comparability across sessions and domains. The resulting discourse not only captures factual content but also the contextual reasoning behind it, laying a solid foundation for subsequent formalisation into RDF triples. This stage, represented in the upper segment of Figure 1, is thus critical to the quality and reliability of the entire K-GRASP pipeline.

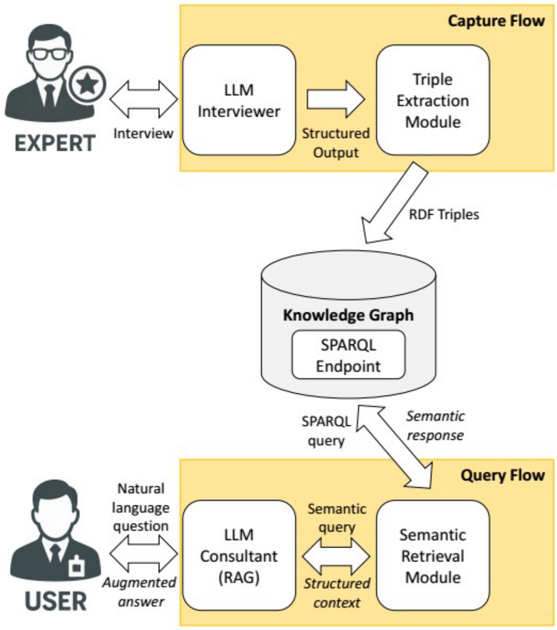

Figure 1: Proposed architecture for the externalisation and retrieval of tacit knowledge using LLMs and KGs. The upper flow represents the capture process

4.2 Semantic Formalisation

After the capture phase, the discourse produced by experts is transformed into structured representations as RDF triples of the form <subject, predicate, object>. This step converts free text into machine-readable statements that encode not only factual entities but also the semantic relations between them. The transformation may be performed using rule-based linguistic patterns, which guarantee determinism, or with the aid of LLMs fine-tuned for structured output, which provide greater flexibility in dealing with ambiguous or less standardised expressions. RDF is adopted for its broad interoperability, its compatibility with existing semantic web standards, and its capacity to support inferential reasoning beyond simple storage.

The generated triples are subsequently aligned with a reference ontology, which serves as a semantic scaffold to ensure terminological coherence and to avoid inconsistent mappings. This alignment allows the resulting knowledge base to remain interoperable with external datasets while also enabling logical inference over the represented concepts. In practice, this stage transforms the narrative flow of expert interviews into a formal semantic layer, bridging unstructured discourse and symbolic knowledge representation. As illustrated in the transition between the upper and middle layers of Figure 1, it constitutes the methodological pivot that ensures knowledge captured in natural language becomes reusable, auditable, and suitable for reasoning in downstream applications.

4.3 Knowledge Graph Construction and Storage

Once generated, the RDF triples are integrated into a semantic KG that serves as the central repository of the K-GRASP architecture. This process involves not only storing the triples but also resolving entity references, consolidating redundant nodes, and validating structural consistency to ensure semantic cohesion. The KG serves as a backbone for capturing complex relationships among entities, providing traceability and enabling automated reasoning. Its construction benefits from reusing existing ontologies, enabling alignment with standards such as FOAF, Dublin Core, and Schema.org, while also supporting domain-specific extensions.

The resulting KG is deployed in a semantic database that supports SPARQL queries, providing efficient retrieval and interoperability with external knowledge sources. Beyond functioning as a static repository, the graph serves as a dynamic semantic layer where new knowledge can be iteratively added, refined, and validated. This stage thus establishes both the persistence and scalability of the captured knowledge, ensuring that what was initially elicited through dialogue can evolve into a robust and extensible infrastructure for reasoning and retrieval. In the overall architecture, it corresponds to the middle layer in Figure 1, bridging the initial capture of tacit knowledge and the retrieval mechanisms that make it accessible to end users.

4.4 Retrieval Phase

In the final stage, tacit knowledge that has been formalised and stored in the KG becomes operationally accessible through a consultative LLM using RAG. When a user submits a natural language query, the model translates it into a SPARQL query, retrieves the relevant subgraph, and integrates the extracted content into its generative process. This hybrid workflow allows the response to remain grounded in structured, verifiable knowledge while retaining the fluency and contextual adaptation of neural generation. By combining symbolic retrieval with probabilistic reasoning, the system reduces the risk of hallucinations, increases factual accuracy, and produces outputs that can be traced back to their semantic origins.

The retrieval phase also emphasises accessibility and explainability: experts and non-specialists alike can interact with the system using natural language, while the underlying semantic infrastructure ensures transparency and auditability. Furthermore, this approach supports modularity, since different LLMs or query engines can be incorporated without altering the semantic backbone. Functionally, this phase closes the methodological cycle, completing the transition from tacit insights to structured knowledge and back to actionable answers, as represented in the lower segment of Figure 1.

5. Discussion

The K-GRASP architecture offers a systematic approach to externalising, formalising, and retrieving tacit knowledge by integrating LLMs with KGs through RAG. By combining these components, the architecture addresses longstanding challenges in knowledge management, including the capture and formal representation of experiential knowledge, and the provision of contextualised, accurate responses to queries.

Compared to traditional knowledge engineering techniques, which rely heavily on structured interviews, observations, and manual encoding, K-GRASP offers greater scalability, automation, and traceability (Davenport and Prusak, 1998; Tiwana, 2000). While conventional methods are resource-intensive and often constrained by subjectivity and limited reusability, integrating LLMs and semantic structures enables capturing tacit knowledge in a structured, machine-readable format, thereby supporting reuse across contexts and systems. This hybrid approach aligns with recent advances in AI-driven knowledge management, demonstrating the potential to enhance both efficiency and semantic richness (Bommasani et al., 2021; Xu et al., 2023).

Despite these advantages, several practical challenges remain. The quality and completeness of captured knowledge depend on active expert participation, as LLMs cannot replace human expertise in articulating tacit or context-dependent insights (Bratianu and Orzea, 2010). While studies demonstrate that LLMs can simulate expert participation in collaborative studies, producing responses that closely resemble those of real specialists (Nóbrega et al., 2025), such personas do not eliminate the need for authentic expert contributions; rather, they serve as complementary substitutes when direct access to experts is limited or unavailable. Besides, cognitive effort and individual biases may affect the diversity and generalizability of the resulting Knowledge Graph, highlighting the need for complementary mechanisms such as peer validation. Reviewing and refining the semantic representations through expert panels can ensure conceptual fidelity and ontological consistency, in line with best practices in knowledge governance and FAIR principles (Hogan et al., 2021).

Ontologies play a central role in structuring the extracted knowledge, providing formal vocabularies and relationships that facilitate interoperability, reasoning, and explainability (Gruber, 1995; Ji and others, 2021). Broad, general-purpose ontologies support integration but may lack specificity, whereas domain-specific ontologies capture detailed expertise but require maintenance and careful Evolution (Flouris et al., 2008; Noy and McGuinness, 2001). The modularity of K-GRASP allows flexible adaptation of ontologies and LLM components to new domains, making the architecture extensible and robust.

Compared with prior work, K-GRASP stands out for combining structured knowledge representation with neural generative models within a cohesive framework. While previous studies have explored LLMs for knowledge extraction or KGs for formal representation separately, the RAG-based integration provides traceable, contextually accurate outputs and enhances the interpretability of results (Das and others, 2021; Lewis et al., 2020). This positions K-GRASP as a versatile tool for knowledge-intensive domains, including technical, scientific, and professional environments, where tacit knowledge is critical yet traditionally difficult to formalise.

Overall, the proposed architecture demonstrates that hybrid approaches leveraging both symbolic and neural methods can significantly advance the externalisation and reuse of tacit knowledge. Future work may explore enhancements such as automated verification mechanisms, adaptive ontologies, and expanded evaluation in real-world organisational settings to further validate and extend its applicability.

6. Conclusion

Tacit knowledge is a critical yet difficult-to-formalise asset within organisations, as it is often embedded in practical experience. This work addresses this challenge by proposing an architecture that combines LLMs and KGs to externalise and systematise such knowledge, offering a structured and scalable alternative to traditional elicitation methods.

Throughout this work, the K-GRASP architecture was developed and described, consisting of two main flows: a capture phase, where an interviewer LLM conducts sessions with human experts and converts their responses into RDF triples; and a retrieval phase, in which a consultative LLM uses the RAG technique to query the semantic graph and generate contextualised responses. Each module was designed to ensure semantic consistency, reusability, and transparency in knowledge processing.

The main contribution of this work lies in the theoretical and architectural articulation of a pipeline that transforms tacit knowledge into machine-readable structures. By adopting ontologies and SPARQL-compatible queries, the system enables accurate, auditable, and interpretable information retrieval, overcoming limitations commonly associated with opaque neural models.

Key contributions include integrating symbolic and neural approaches within a unified architecture, using structured prompts for guided elicitation, and formally representing experiential knowledge via RDF triples aligned with ontologies. This combination supports the generation of verifiable, structured, knowledge-based responses, thereby enhancing the explainability and reliability of AI-driven decision-support systems.

Despite the advances presented, the architecture still depends on the availability and clarity of domain experts during the elicitation phase, and the semantic quality of the generated triples may be affected by ambiguities or biases in the responses. Additionally, the effort required to define or adapt ontologies for specific domains remains a practical challenge, particularly in rapidly evolving fields.

Future work includes empirical validation of the architecture in domains such as engineering, healthcare, and education, which may reveal specific requirements for prompt design, ontology selection, and evaluation protocols. We also propose developing peer validation mechanisms and automated metrics to assess the consistency and usefulness of extracted knowledge, thereby enhancing the trustworthiness of LLM-based systems for tacit knowledge management.

Ethics Declaration: Ethical clearance was not required for the development of this research.

AI Declaration: AI tools such as Grammarly and ChatGPT were used solely for language revision. The authors' analysis and interpretations are their own.

References

Babashahi, L., Barbosa, C.E., Lima, Y., Lyra, A., Salazar, H., Argôlo, M., Almeida, M.A. de, Souza, J.M. de, 2024. AI in the workplace: A systematic review of skill transformation in the industry. Administrative Sciences 14, 127.

Bendiken, A., 2024. KNOW: A Real-World Ontology for Knowledge Capture with Large Language Models. arXiv preprint arXiv:2405.19877.

Berners-Lee, T., Hendler, J., Lassila, O., 2001. The Semantic Web. Scientific American 284, 28–37.

Bommasani, R., et al., 2021. On the Opportunities and Risks of Foundation Models. arXiv preprint arXiv:2108.07258.

Bratianu, C., Orzea, I., 2010. Tacit Knowledge Sharing in Organizational Knowledge Dynamics, in: Proceedings of the 2nd European Conference on Intellectual Capital. Academic Publishing International, pp. 107–114.

Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al., 2020. Language models are few-shot learners, in Advances in Neural Information Processing Systems.

Dalkir, K., 2017. Knowledge Management in Theory and Practice. MIT Press.

Das, A.S., others, 2021. Knowledge graphs and explainable AI. Communications of the ACM 64, 64–73.

Davenport, T.H., Prusak, L., 1998. Working Knowledge: How Organizations Manage What They Know. Harvard Business Press.

Flouris, G., Manakanatas, D., Kondylakis, H., Plexousakis, D., Antoniou, G., 2008. Ontology change: Classification and survey. The Knowledge Engineering Review 23, 117–152.

Gruber, T.R., 1995. Toward principles for the design of ontologies used for knowledge sharing. International Journal of Human-Computer Studies 43, 907–928.

Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., de Melo, G., Gutiérrez, C., Kirrane, S., Gayo, J.E.L., Navigli, R., Neumaier, S., others, 2021. Knowledge graphs. ACM Computing Surveys (CSUR) 54, 1–37.

Ji, S., others, 2021. A Survey on Knowledge Graphs: Representation, Acquisition and Applications. IEEE Transactions on Neural Networks and Learning Systems 33, 494–514.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Kuttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., Kiela, D., 2020. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks, in: Advances in Neural Information Processing Systems (NeurIPS).

Lima, C.D.C., Salazar, H., Lima, Y., Barbosa, C.E., Argôlo, M., Lyra, A., de Souza, J.M., 2024. Generative AI impact on the future of work: Insights from software development, in: 2024 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, pp. 3887–3892.

Liu, W., et al., 2023. RAG-Duo: Effective and Efficient Retrieval-Augmented Generation over Tabular and Textual Data. arXiv preprint arXiv:2310.01863.

Lu, Z., others, 2024. Logic Augmented Generation: Integrating Semantic Knowledge Graphs with Large Language Models. arXiv preprint arXiv:2411.14012.

Martin, J., 2002. Organizational Culture: Mapping the Terrain. Sage Publications, Thousand Oaks, CA.

Martin, R., Hecker, A., 2018. Tacit Knowledge in Organizations: An Integrative Review. Academy of Management Annals 12, 612–640.

Monteiro, A. da F., Argôlo, M., Barbosa, C.E., Nóbrega, L., Martinez, L.F., Lima, Y., Souza, J. de, 2025. Knowledge Management in the Tech Sector: A Rapid Review of Benefits, Challenges, and Metrics, in: Proceedings of the 26th European Conference on Knowledge Management (ECKM). Academic Conferences International Limited, Lisbon, Portugal, pp. 659–667.

Nóbrega, L., Martinez, L.F., Lima, Y., Barbosa, C.E., Argôlo, M., Salazar, H., Xexéo, G., De Souza, J.M., 2025. Towards a Knowledge Management Framework for LLM-Generated Personas in Collaborative Systems, in: 2025 28th International Conference on Computer Supported Cooperative Work in Design (CSCWD). IEEE, pp. 539–545.

Nonaka, I., Takeuchi, H., 1995. The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation. Oxford University Press.

Noy, N.F., McGuinness, D.L., 2001. Ontology Development 101: A Guide to Creating Your First Ontology (No. KSL-01-05). Stanford Knowledge Systems Laboratory.

OpenAI, 2023. GPT-4 Technical Report.

Polanyi, M., 1966. The Tacit Dimension. University of Chicago Press.

Radford, A., Wu, J., et al., 2019. Language Models are Unsupervised Multitask Learners.

Stollenwerk, M.F.L., 2001. Gestão do conhecimento: conceitos e modelos, in: Tarapanoff, K. (Ed.), Inteligência Organizacional e Competitiva. UNB, Brasília, pp. 143–163.

Sun, Z., Ren, X., Wang, W.Y., 2023. Knowledge-intensive natural language processing with large language models: A survey. arXiv preprint arXiv:2303.08896.

Tiwana, A., 2000. The Knowledge Management Toolkit: Practical Techniques for Building a Knowledge Management System. Prentice Hall PTR.

Wang, Y., others, 2024. Towards AI-Driven Human-Machine Co-Teaming for Adaptive and Resilient Cybersecurity Operations. arXiv preprint arXiv:2505.06394.

Xu, C., Liu, Z., Sun, M., 2023. Retrieval-augmented generation for knowledge-intensive NLP tasks: A survey. arXiv preprint arXiv:2302.05110.

Zhou, X., others, 2024. Code Digital Twin: Empowering LLMs with Tacit Knowledge for Complex Software Maintenance. arXiv preprint arXiv:2503.07967.