Revista Eletrônica de Iniciação Científica em Computação, 2026, 24:1
ISSN: 3085-8461 • doi: 10.5753/reic.2026.8503

Research Paper

Multimodal RAG with Knowledge Graphs for Portuguese Maintenance Manuals in the Context of Industry 4.0

Christian Freitas [Universidade Federal de São Paulo | *christian.freitas@unifesp.br*]
Lilian Berton [Universidade Federal de São Paulo |*lberton@unifesp.br*]

Institute of Science and Technology, Universidade Federal de São Paulo (UNIFESP), São José dos Campos, SP, Brazil.

Navigating dense, multimodal technical documentation remains a critical bottleneck for industrial maintenance in non-English contexts. This paper presents a multimodal Retrieval-Augmented Generation (RAG) framework for Portuguese maintenance manuals, combining dense vector search over text and tables with knowledge graph traversal. We evaluate a local multilingual transformer (`paraphrase-multilingual-MiniLM-L12-v2`) and OpenAI's `text-embedding-3-small` across 8 configurations (2 models × 4 retrieval modes) using a ground-truth set of 50 domain-specific queries. The best configurations (`openai/text_only` and `openai/text_table`) achieve a BERTScore-F1 of 0.70, ROUGE-L of 0.36, and an MRR of 0.65, answering 45 of 50 queries. Analysis shows that while high-dimensional proprietary embeddings excel at capturing specialized technical jargon, standard Reciprocal Rank Fusion (RRF) introduces rank contamination when fusing sparse graph signals with dense text vectors. Source code is publicly available.

Keywords: Retrieval-Augmented Generation, Knowledge Graphs, Multimodal RAG, Industry 4.0, Portuguese NLP

Received: 17 June 2026 • **Accepted:** 17 June 2026 • **Published:** 10 July 2026

1 Introduction

Industry 4.0 represents the convergence of advanced digital technologies, such as the Internet of Things (IoT), artificial intelligence, big data analytics, and cyber-physical systems, into manufacturing environments, creating smart and interconnected production ecosystems [Ghobakhloo, 2020]. Within this context, industrial maintenance operations become critical, as the complexity of equipment and the volume of data generated demand rapid and precise access to technical information [Bousdekis *et al.*, 2019]. This directly connects to the reliance on technical documentation, including equipment manuals, maintenance procedures, safety guidelines, and spare parts catalogs. These documents, often dense and multi-layered, contain narrative text, structured tables, technical diagrams, and component images. The ability to efficiently locate and interpret accurate information within this corpus is essential for minimizing equipment downtime and ensuring safe, reliable operations in the era of Industry 4.0.

This perspective directly resonates with the Sustainable Development Goals related to industry, innovation, and infrastructure (SDG 9) [United Nations, 2015], by fostering more efficient and safer maintenance practices. The ability to navigate and interpret complex technical documentation helps reduce equipment downtime, enhance operational reliability, and drive industrial modernization within the framework of Industry 4.0, thereby strengthening competitiveness and sustainability in the Brazilian industrial sector.

Retrieval-Augmented Generation (RAG) [Lewis *et al.*, 2020] is a paradigm for question answering over document collections that combines dense retrieval with generative language models. Most RAG implementations focus on English-language, text-only corpora. Industrial settings in Brazil introduce two additional challenges: (i) documents are primarily in Portuguese, where general-purpose English embedding models produce lower-quality representations, and (ii) relevant information is often stored in non-textual modalities such as tables and technical diagrams.

This work is part of a research line on intelligent access to technical documents in Brazilian Portuguese. Prior work in this line addressed RAG applied to epidemiological reports in Portuguese [Freitas *et al.*, 2025a], demonstrating that retrieval-augmented generation is effective for technical Portuguese text. A parallel line investigated structured data access through Text-to-SQL with Portuguese LLMs and LangGraph for industrial databases [Freitas *et al.*, 2025b], and an extended version integrating RAG with Text-to-SQL for industrial data retrieval is currently under review [Freitas *et al.*, 2025c]. Linguistic resources supporting Portuguese NLP were also developed as part of this effort [Freitas *et al.*, 2026]. The present paper extends this line by introducing multimodal retrieval over unstructured industrial maintenance documents, combining dense vector search over text and tables with knowledge graph traversal.

The contributions of this work are:

- A four-phase pipeline covering PDF extraction with semantic chunking, knowledge graph construction, multimodal embedding generation, and hybrid retrieval with Reciprocal Rank Fusion (RRF).
- A comparative ablation study evaluating two embedding models across four retrieval modes (8 configurations) on a domain-specific query set in Portuguese.
- Empirical results showing that OpenAI embeddings produce higher answer quality across all categories, while multilingual sentence transformers remain competitive in safety-related retrieval ranking.
- An open-source implementation available at `https://github.com/ChristianSF/multimodal-rag-industrial`.

2 Related Work

Retrieval-Augmented Generation: Lewis *et al*. [2020] introduced RAG as a framework that combines generative models with non-parametric retrieval. Subsequent work explored dense passage retrieval [Karpukhin *et al*., 2020], hybrid retrieval combining sparse and dense signals [Arivazhagan *et al*., 2023], and reranking strategies including Reciprocal Rank Fusion [Cormack *et al*., 2009].

Multimodal RAG: Extending RAG to incorporate images, tables, and structured data has been explored in different directions. *RAG Beyond Text* [Mei *et al*., 2025] surveys approaches that process heterogeneous document modalities and reports improvements in answer quality in technical domains when tables and figures are included. Query-Driven Multimodal GraphRAG [Bu *et al*., 2025] further combines dynamic knowledge graph construction with vector retrieval over documents, images, and tables, reporting improvements over standard RAG and GraphRAG on visual question answering benchmarks. The present work applies a similar multimodal retrieval strategy to the specific domain of industrial maintenance documentation in Portuguese.

Knowledge graph-augmented retrieval: The RAP framework [Kagaya *et al*., 2024] incorporates structured knowledge graphs into the retrieval pipeline and reports improvements on multi-hop queries. SiQA [Liu *et al*., 2025] shows that situated question answering benefits from explicit relational context. Most closely related to the present work, a recent study from Bosch [Zhang *et al*., 2025] proposes a semi-automated knowledge graph construction pipeline combining rule-based methods, Small Language Models, and LLMs for maintenance question answering in manufacturing environments. While that work focuses on graph construction quality, the present work evaluates retrieval strategies empirically across multiple modes and embedding models on Portuguese documents.

Multilingual embeddings: Challenges of retrieval in non-English languages, particularly Portuguese technical text, have been reported in recent benchmarks. Sentence transformers trained on multilingual corpora [Reimers and Gurevych, 2019] produce better representations for Portuguese retrieval tasks compared to English-only models, which motivates the comparative evaluation conducted in this work.

This work: The primary novelty lies in implementing a multimodal RAG framework specifically optimized for Portuguese within the industrial maintenance domain, addressing a gap where most state-of-the-art solutions focus on English. Unlike traditional approaches limited to dense text, this proposal integrates vector search across heterogeneous modalities—text and structured tables—with knowledge graph traversal to resolve complex, multi-hop queries. This research conducts a comparative empirical evaluation between multilingual and OpenAI embedding models to identify which strategy best captures the technical nuances of Brazilian industrial terminology.

3 Methodology

The proposed framework consists of four sequential phases, illustrated in Figure 1.

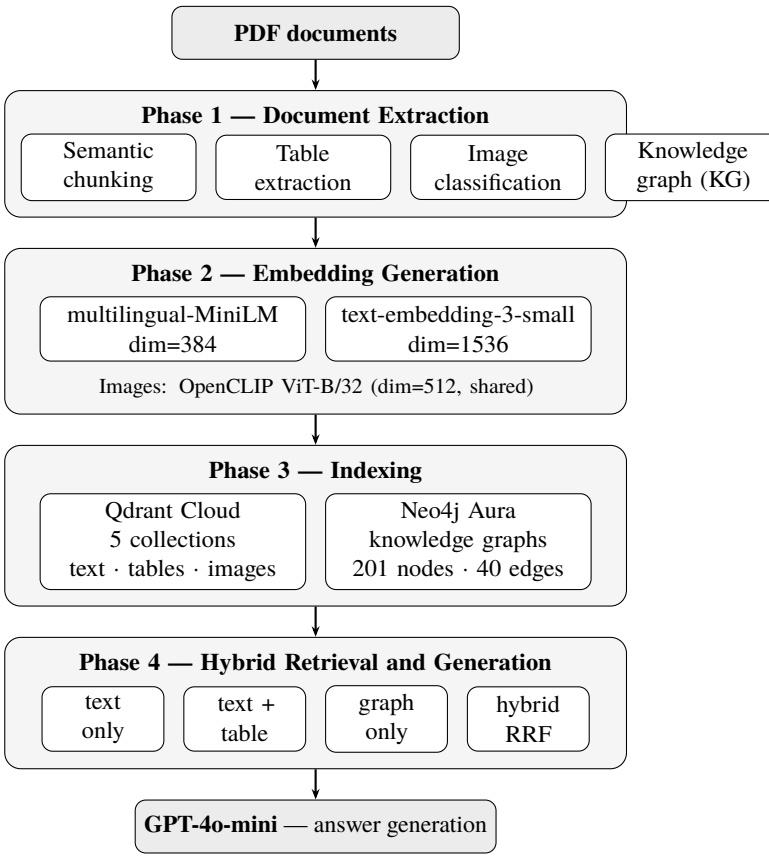

Figure 1. Overview of the four-phase multimodal RAG pipeline.

3.1 Phase 1: Document Extraction

Raw PDF documents are processed using PyMuPDF. Each document undergoes the following steps: (i) **semantic chunking** by section, preserving structural boundaries rather than applying fixed-size windows; (ii) **table extraction**, where tabular regions are detected with PyMuPDF's `find_tables()` and each table is serialized as a Markdown-formatted text block that preserves its row and column structure—rather than as a JSON object of cells—so that it can be embedded as a dense vector alongside the text chunks (Phase 3) and queried in the `text_table` mode; (iii) **image extraction and classification** into three categories (diagram, photograph, table-image) using heuristics based on aspect ratio and pixel density; and (iv) **knowledge graph construction** per document, where entities (equipment, components, spare parts) and their relationships are extracted using GPT-4o-mini with structured output prompts. The node types used are `Equipment`, `Component`, `SparePart`, and `Document`, connected by semantic relations (`SAME_STRUCTURAL_ASSEMBLY`, `SAME_ELECTRICAL_ASSEMBLY`, `SAME_MECHANICAL_ASSEMBLY`, `SAME_HYDRAULIC_ASSEMBLY`).

Processing four test documents produced 238 text chunks, 2,763 extracted images, 108 tables, and 18 knowledge graphs containing 201 nodes and 40 semantic edges.

3.2 Phase 2: Multimodal Embedding Generation

Embeddings are generated using two text models in parallel:

- **Model A (multilingual):**

`paraphrase-multilingual-MiniLM-L12-v2` [Reimers and Gurevych, 2019], a sentence transformer trained on over 50 languages, producing 384-dimensional vectors. The model runs locally without API costs.

- **Model B (OpenAI):** `text-embedding-3-small` [OpenAI, 2024], producing 1,536-dimensional vectors via API.

Images are embedded using OpenCLIP ViT-B/32 [Cherti *et al.*, 2023], producing 512-dimensional vectors shared across both model configurations.

3.3 Phase 3: Indexing

Vector embeddings are indexed in Qdrant Cloud across five collections separated by modality and model:

Collection	Dim	Points
`maintenance_text_multilingual`	384	238
`maintenance_text_openai`	1536	238
`maintenance_tables_multilingual`	384	108
`maintenance_tables_openai`	1536	108
`maintenance_images`	512	2763

Knowledge graphs are stored in Neo4j Aura with node types `Equipment`, `Component`, `SparePart`, and `Document`, connected by the semantic relations described in Phase 1.

3.4 Phase 4: Hybrid Retrieval and Generation

At query time, the system operates in one of four retrieval modes:

- **text_only:** Dense vector search over text chunks.
- **text_table:** Dense vector search over text and table chunks.
- **graph_only:** Keyword-based node-first traversal over Neo4j knowledge graphs.
- **hybrid:** The retrieval results from the three modalities are combined using a weighted variant of Reciprocal Rank Fusion (RRF) [Cormack *et al.*, 2009]. The combined score for each document d within the document corpus D is calculated as follows:

$$RRF\_Score(d \in D) = \sum_{m \in M} w_m \cdot \frac{1}{k + r_m(d)} \quad (1)$$

where $M = \{\text{text}, \text{table}, \text{graph}\}$ represents the set of retrieval modalities, w_m is the weight assigned to modality m ($w_{\text{text}} = 1.0$, $w_{\text{table}} = 0.8$, and $w_{\text{graph}} = 0.9$), $k = 60$ is a constant regularization parameter, and $r_m(d)$ is the rank of document d in the result list of modality m. If a document does not appear in the top retrieved results for a given modality, its reciprocal rank contribution for that modality is set to zero.

The top-k retrieved context ($k = 5$) is passed to GPT-4o-mini for answer generation, with a system prompt instructing the model to answer exclusively from the provided context and cite sources.

4 Experiments and Results

4.1 Experimental Setup

The pipeline was evaluated on four industrial maintenance documents in Portuguese: two welding equipment manuals (`manual-migpulse-2001-dp.pdf` and `Certo-09-07-2019-08-07-27_2_doc_25.pdf`), one industrial machine manual (`Manual_Maquina.pdf`), and one cooling system cleaning procedure (`Certo-Limpeza-Geral-Sistemas-de-Resfriamento.pdf`). The evaluation set contains 50 domain-specific queries covering five categories: specification, safety, procedure, maintenance, and troubleshooting. Queries and reference answers were manually formulated by the authors based on direct reading of the source documents, targeting information that requires locating specific sections, tables, or procedural steps. We evaluated 8 configurations resulting from the combination of 2 embedding models—multilingual MiniLM and OpenAI `text-embedding-3-small`—with 4 retrieval modes: `text_only`, `text_table`, `graph_only`, and `hybrid`. For each query, the system retrieved the top-k contexts ($k = 5$) and generated an answer with GPT-4o-mini.

4.2 Evaluation Metrics

Three complementary metrics are reported:

- **BERTScore-F1**: semantic similarity between generated and reference answers using `bert-base-multilingual-cased`.
- **ROUGE-L**: lexical overlap via longest common subsequence.
- **MRR** (Mean Reciprocal Rank): retrieval ranking quality, checking whether the correct source document appears in the top-k retrieved results.

A response is classified as answered when it contains domain-specific content rather than a refusal.

4.3 Overall Results

Table 1 summarizes the quantitative results across all 8 configurations.

Table 1. Overall results for the 8 evaluated configurations (50 queries).

Model	Mode	Ans.	BS-F1	RL	MRR
multilingual	text_only	38/50	0.582	0.295	0.591
multilingual	text_table	38/50	0.583	0.296	0.591
multilingual	hybrid	31/50	0.466	0.220	0.462
multilingual	graph_only	3/50	0.042	0.014	0.308
OpenAI	text_only	**45/50**	**0.697**	0.361	**0.652**
OpenAI	text_table	**45/50**	0.696	**0.361**	**0.652**
OpenAI	hybrid	42/50	0.639	0.316	0.458
OpenAI	graph_only	3/50	0.040	0.011	0.308

The best overall configurations were `openai/text_only` and `openai/text_table`, both answering 45 of 50 queries and achieving nearly identical scores: BERTScore-F1 of 0.697 and 0.696, ROUGE-L of 0.361, and MRR of 0.652. Both `graph_only` configurations answered only 3 of 50 queries and obtained the lowest scores,

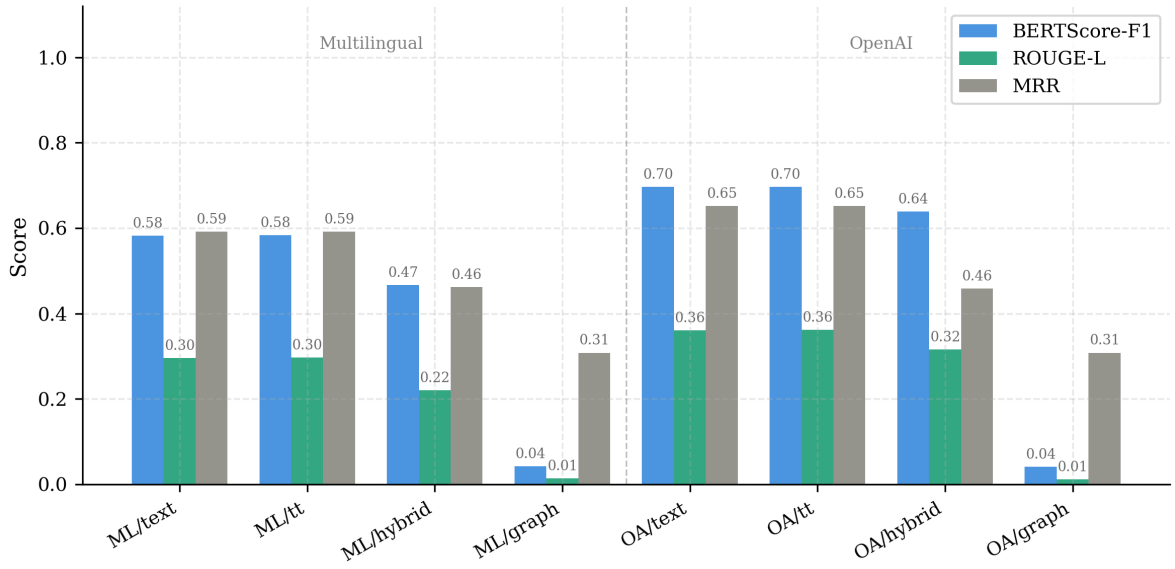

Figure 2. Comparison of the 8 evaluated configurations using BERTScore-F1, ROUGE-L, and MRR.

confirming that graph-only retrieval is insufficient as a standalone strategy.

Figure 2 illustrates the three metrics across all configurations.

4.4 Category-Level Analysis

Figure 3 presents BERTScore-F1 by query category for the `text_only` mode. OpenAI achieved higher scores in all five categories: specification (0.73 vs. 0.60), safety (0.67 vs. 0.59), procedure (0.68 vs. 0.51), maintenance (0.66 vs. 0.53), and troubleshooting (0.77 vs. 0.76).

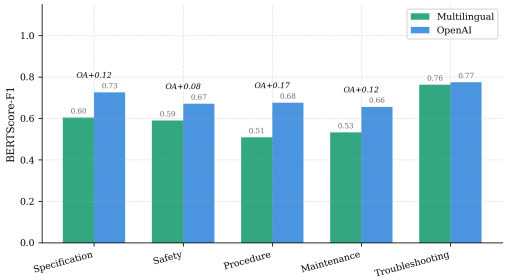

Figure 3. BERTScore-F1 by query category (`text_only` mode). OpenAI achieves higher scores across all categories.

Figure 4 presents MRR by query category. The multilingual model achieved higher retrieval ranking in safety (0.72 vs. 0.70) and comparable performance in maintenance (0.67 vs. 0.71). OpenAI led in troubleshooting (0.83 vs. 0.60),

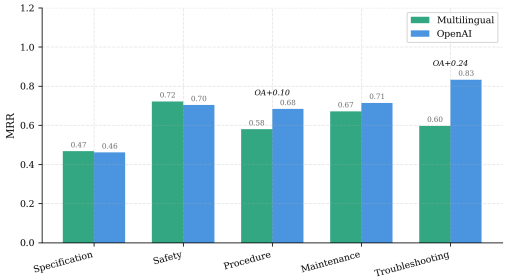

Figure 4. MRR by query category (`text_only` mode). The multilingual model leads only in safety retrieval ranking.

procedure (0.68 vs. 0.58), and specification (0.46 vs. 0.47). These results show that OpenAI embeddings provide stronger overall performance, while multilingual embeddings remain competitive in safety-related queries.

4.5 Qualitative Example

Figure 5 shows a representative query and the answer generated by the `openai/text_table` configuration, alongside the retrieved source diagram. The system correctly identifies the three-step procedure and retrieves the associated technical illustration.

5 Discussion

Overall model comparison. OpenAI embeddings produced the strongest overall results. Both `openai/text_only` and `openai/text_table` achieved BERTScore-F1 of 0.697, ROUGE-L of 0.361, and MRR of 0.652, answering 45 of 50 queries. The strong performance of `openai/text_only` shows that dense text retrieval alone is already effective in this dataset.

Effect of table retrieval. Adding tables produced marginal differences for both models. For OpenAI, BERTScore-F1 changed from 0.697 to 0.696 with answer coverage unchanged at 45/50. For the multilingual model, results were similarly unchanged (0.582 vs. 0.583). This negligible effect is explained by the composition of the benchmark relative to the four manuals: of the five query categories, only *specification* questions explicitly target tabular numerical data, and even those values typically appear *verbatim* in the surrounding narrative text (e.g., within procedure descriptions or section headings). The remaining categories—safety, procedure, maintenance, and troubleshooting—are answered from running text. Consequently, although the four documents yielded 108 tables, few of them held an answer span that was not already retrievable from the text chunks, leaving `text_table` virtually identical to `text_only`. A tighter coupling between table content and natural-language queries therefore warrants further investigation.

Hybrid retrieval. The empirical results reveal that the

Query: Quais são os passos básicos para realizar a soldagem MIG sinérgica?

Source: manual-migpulse-2001-dp.pdf, p. 8

Answer: Os passos básicos são: (1) escolher o número de PRG de acordo com o diâmetro do fio, o tipo de material e o gás; (2) regular as funções no submenu de "Funções de serviço"; (3) regular a velocidade do fio e a tensão de soldagem utilizando o manípulo AI.

16- Fac. (fatory).
A finalidade é a de fazer regressar a soldadora à configuração de fábrica. Depois de seleccionada a função, o visor **AM** mostra **noP** = faz regressar a soldadora à configuração de fábrica ignorando os programas memorizados; **Prg** = apaga todos os programas memorizados e **ALL** = faz regressar a soldadora à configuração de fábrica. Para confirmar a função desejada basta premir por 3 segundos o botão **AV**, a sigla mostrada no visor **AM** começa a piscar o, após alguns segundos, um som confirmará a memorização.

6 INSTALAÇÃO

A instalação da soldadora deve ser executada por pessoal especializado. As ligações devem ser todas executadas respeitando as leis de prevenção de acidentes em vigor (CEI 26-23 - IEC/TS 62081).

6.1 COLOCAÇÃO

A soldadora pesa cerca de **22 kg**.
Colocar o aparelho numa zona que assegure uma boa estabilidade, um arejamento suficiente e de modo a evitar que possa entrar pó metálico (por ex. limalhas).

6.2 PREPARAÇÃO PARA O FUNCIONAMENTO

Montar a ficha no cabo de alimentação prestando especial atenção na ligação do condutor amarelo-verde ao pólo de terra. Verificar se a tensão de alimentação corresponde à nominal da soldadora.
Dimensionar os fusíveis de proteção em função dos dados indicados na chapa dos dados técnicos.
Ligar o tubo do gás da máquina de soldar ao redutor de pressão da botija. Montar o maçarico.
Para verificar se o bordo dos rolos corresponde ao diâmetro do fio utilizado, abrir a parte lateral móvel, tirar a cobertura **CA**, soltar os rolos de pressão do fio no manípulo de regulagem da pressão**BN**, substituir os rolos e montar a cobertura **CA** (Ver a fig. 4).

Fig. 4

Montar a bobina do fio e enfiar o fio na tração e na baínha do maçarico.

Fixar os rolos de pressão do fio no manípulo **BN** e regular a pressão.
Ligar a máquina.
Regular o gás utilizando o botão **AV** , depois avançar o fio com o botão **AW**.

7 SOLDAGEM

Soldagem **Mig pulsada sinérgica**, LED **AP** aceso.
Escolher o número de **PRG** em função do diâmetro do fio a utilizar, o tipo, a qualidade do material e o tipo de gás, utilizando a instrução colocada no interior da parte lateral móvel.
Regular as funções presentes no submenu como indicado no parágrafo "**Funções de serviço**".
A regulagem dos parâmetros de soldagem é efetuada no manípulo **AI**.
Soldagem **Mig sinérgica**, LED **AQ** aceso.
Escolher o número de **PRG** em função do diâmetro do fio a utilizar, o tipo, a qualidade do material e o tipo de gás, utilizando a instrução colocada no interior da parte lateral móvel.
Regular as funções presentes no submenu como indicado no parágrafo "**Funções de serviço**".
Regular a velocidade do fio e a tensão di soldagem tramite o manípulo **AI**.
Soldagem **Mig** convencional, LED **AR** aceso.
Escolher o número de **PRG** em função do diâmetro do fio a utilizar, o tipo, a qualidade do material e o tipo de gás, utilizando a instrução colocada no interior da parte lateral móvel.
Regular as funções presentes no submenu como indicado no parágrafo "**Funções de serviço**".
Regular a velocidade do fio e a tensão de soldagem respectivamente nos manípulos **AI** e **AN**.

8 ACESSÓRIOS

8.1 MAÇARICO MIG ART. 1242
Maçarico MIG 280 A arrefecido a ar 3,5 m.

8.2 MAÇARICO PUSH-PULL UP/DOWN arrefecido a ar ART. 2003.

8.3 Carro para transporte do gerador, art. 1656.

8.4 Kit para bobinas de 300 mm de diâmetro, 15 kg, art. 114.

9 MANUTENÇÃO

Cada trabalho de manutenção deve ser executado por pessoal qualificado respeitando a norma CEI 26-29 (IEC 60974-4).

Figure 5. Query, generated answer, and retrieved diagram (Fig. 4 of the MigPulse manual, p. 8).

hybrid retrieval mode systematically underperformed compared to the text-based baselines, a behavior that requires a more nuanced mathematical explanation than a simple mismatch in score scales. Because the standard Reciprocal Rank Fusion (RRF) algorithm scores documents strictly based on their relative positions within individual retrieval lists rather than their raw similarity scores, it is highly sensitive to the quality of the incoming rankings. In our framework, the *graph_only* mode exhibited severe signal sparsity, successfully answering only 3 out of 50 queries. Consequently, for the vast majority of queries, the graph traversal returned either empty sets or highly noisy, low-confidence node lists. When these sparse and irrelevant graph results were fused with the high-quality textual rankings, the aggressive weight assigned to the graph component ($w_{\text{graph}} = 0.9$) forced unrelated documents to artificial prominence at the top of the combined hybrid list. This rank contamination effectively penalized the robust textual signals, introducing significant noise into the top-k context passed to the generative stage and directly causing the observed drop in BERTScore-F1 and answer coverage.

Graph-only retrieval. The `graph_only` setting was not competitive as a standalone mode. Both models answered only 3 of 50 queries, with BERTScore-F1 near 0.04, confirming that the knowledge graph alone captures only a limited portion of the information required by the benchmark. Graph retrieval must therefore be complementary to vector search.

Dataset size. The current corpus of four documents represents an acknowledged limitation of this work. While the 50-query benchmark with five balanced categories enables a controlled ablation study, the small document set limits the generalizability of the findings. Expanding the corpus is a direct target for future work in order to validate the observed results across a broader range of Portuguese industrial documentation.

Linguistic and Architectural Comparison of Embedding Models. The substantial performance gap between the embedding models warrants a deeper analysis of their architectural and linguistic characteristics. OpenAI's *text-embedding-3-small* significantly outperformed the *paraphrase-multilingual-MiniLM-L12-v2* across almost all evaluation categories. This behavior can be attributed to two main factors: dimensionality and vocabulary capacity. The multilingual MiniLM operates on a compressed 384-dimensional space and relies on a highly shared multilingual vocabulary. Consequently, its tokenization process often over-segments domain-specific Brazilian Portuguese industrial terms into generic sub-tokens, failing to capture the tight semantic boundaries of technical jargon such as *"soldagem MIG sinérgica"*. Conversely, OpenAI's model benefits from a significantly larger embedding space (1536 dimensions) and a richer vocabulary representation. This combination allows it to preserve the semantic nuances and localized syntactic structures of specialized manufacturing documentation, proving more robust against the domain-shift typical of industrial maintenance corpora.

6 Conclusion and Future Work

This paper presented a multimodal RAG framework for Portuguese industrial maintenance manuals, combining dense vector search over text and tables with knowledge graph traversal and LLM-based generation. Evaluation across 8 configurations on 50 domain-specific queries showed that `openai/text_only` and `openai/text_table` achieved the best overall performance, with BERTScore-F1 of 0.70, ROUGE-L of 0.36, and MRR of 0.65, answering 45 of 50 queries. Category-level analysis showed that OpenAI embeddings produce higher answer quality across all categories, while the multilingual model remains competitive in safety-related retrieval ranking. Graph-only retrieval proved insufficient as a standalone strategy, and hybrid retrieval via RRF reduced answer quality due to score scale mismatches during fusion.

Future work includes: (i) expanding the document corpus to cover more manufacturers and equipment categories, improving generalizability of the findings; (ii) correcting the RRF score normalization to prevent hybrid retrieval degradation; (iii) applying vision-language models (e.g., Gemini or GPT-4o vision) to caption technical diagrams and enable images to participate in dense retrieval; (iv) enriching the knowledge graph schema with additional entity types such

as `Procedure`, `Parameter`, and `FailureMode` to improve coverage of procedural and parametric content; (v) evaluating Portuguese-specific embedding models such as BERTimbau alongside the multilingual and OpenAI models; (vi) introducing tree-based indexing as an alternative to flat vector search for hierarchically structured manuals; (vii) conducting human evaluation of answer quality with domain experts; and (viii) integrating the pipeline with time-series anomaly detection and structured maintenance history databases [Freitas *et al*., 2025c], and addressing security and LGPD compliance requirements for industrial deployment.

Declaration of Generative AI in the Writing Process

During the preparation of this work, the authors used Claude and Gemini in order to improve the language, grammar, and flow of the manuscript. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declarations

Acknowledgements

The authors would like to thank FAPESP (Number 2020/09850-0) for partially funding this research.

Funding

This research received no specific external funding.

Authors' Contributions

C.F. contributed to the conception of the study, designed and implemented the pipeline, conducted the experiments, and wrote the manuscript. L.B. provided supervision, contributed to the methodology design, and reviewed the manuscript.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The source code and evaluation scripts generated during the current study are available at `https://github.com/ChristianSF/multimodal-rag-industrial`. The ground-truth query set will be made available upon request.

Further relevant information

This work does not involve human subjects research, and therefore did not require ethics committee approval.

References

Arivazhagan, M. G., Liu, L., Qi, P., Chen, X., Wang, W. Y., and Huang, Z. (2023). Hybrid hierarchical retrieval for open-domain question answering. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 10680–10689. Association for Computational Linguistics. DOI: 10.18653/v1/2023.findings-acl.679.

Bousdekis, A., Apostolou, D., and Mentzas, G. (2019). Predictive maintenance in the 4th industrial revolution: Benefits, business opportunities, and managerial implications. *IEEE Engineering Management Review*, 48(1):57–62. DOI: 10.1109/EMR.2019.2958037.

Bu, C., Chang, G., Chen, Z., Dang, C., Wu, Z., He, Y., and Wu, X. (2025). Query-driven multimodal GraphRAG: Dynamic local knowledge graph construction for online reasoning. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 21360–21380. Association for Computational Linguistics. DOI: 10.18653/v1/2025.findings-acl.1100.

Cherti, M., Beaumont, R., Wightman, R., Wortsman, M., Ilharco, G., Gordon, C., Schuhmann, C., Schmidt, L., and Jitsev, J. (2023). Reproducible scaling laws for contrastive language-image learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2818–2829. IEEE. DOI: 10.1109/CVPR52729.2023.00276.

Cormack, G. V., Clarke, C. L. A., and Buettcher, S. (2009). Reciprocal rank fusion outperforms condorcet and individual rank learning methods. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 758–759. ACM. DOI: 10.1145/1571941.1572114.

Freitas, C., Rabonato, R. T., and Berton, L. (2025a). Enhancing epidemiological insights with RAG for SIREVA-SUS reports. In *Anais do Encontro Nacional de Inteligência Artificial e Computacional (ENIAC)*, pages 1364–1375. SBC. DOI: 10.5753/eniac.2025.11796.

Freitas, C., Real, L., Berton, L., and de Paiva, V. (2026). Towards a universal dependencies corpus for Portuguese epidemiological reports. In *Proceedings of the 17th International Conference on Computational Processing of Portuguese (PROPOR 2026) – Vol. 2*, pages 228–237. Association for Computational Linguistics. DOI: 10.18653/v1/2026.propor-2.31.

Freitas, C., Vega-Oliveros, D. A., and Berton, L. (2025b). Enhancing industrial data access with text-to-SQL using Portuguese LLMs and LangGraph. In *2025 12th International Conference on Soft Computing & Machine Intelligence (ISCMI)*, pages 278–282. IEEE. DOI: 10.1109/ISCMI67495.2025.11358511.

Freitas, C., Vega-Oliveros, D. A., and Berton, L. (2025c). Orchestrating industrial data retrieval: A RAG-enhanced text-to-SQL pipeline for Brazilian Portuguese. *Enterprise Information Systems*. Under review.

Ghobakhloo, M. (2020). Industry 4.0, digitization, and opportunities for sustainability. *Journal of Cleaner Production*, 252:119869. DOI: 10.1016/j.jclepro.2019.119869.

Kagaya, T., Yuan, T. J., Lou, Y., Karlekar, J., Pranata, S., Kinose, A., Oguri, K., Wick, F., and You, Y. (2024). RAP: Retrieval-augmented planning with contextual memory for multimodal LLM agents. *arXiv preprint arXiv:2402.03610*. DOI: 10.48550/arXiv.2402.03610.

Karpukhin, V., Oguz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., and Yih, W.-t. (2020). Dense passage retrieval for open-domain question answering. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6769–6781. Association for Computational Linguistics. DOI: 10.18653/v1/2020.emnlp-main.550.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., and Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Ad-*

vances in Neural Information Processing Systems, volume 33, pages 9459–9474. Curran Associates, Inc.. DOI: 10.5555/3495724.3496517.

Liu, J., Tao, Y., Wang, F., Li, H., and Qin, X. (2025). SiQA: A large multi-modal question answering model for structured images based on RAG. In *2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE. DOI: 10.1109/ICASSP49660.2025.10888359.

Mei, L., Mo, S., Yang, Z., and Chen, C. (2025). A survey of multimodal retrieval-augmented generation. *arXiv preprint arXiv:2504.08748*. DOI: 10.48550/arXiv.2504.08748.

OpenAI (2024). Embeddings — OpenAI API documentation. `https://platform.openai.com/docs/guides/embeddings`. Accessed: 2026-03-29.

Reimers, N. and Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992. Association for Computational Linguistics. DOI: 10.18653/v1/D19-1410.

United Nations (2015). Transforming our world: The 2030 agenda for sustainable development. General Assembly 70th Session. Accessed: 2026-03-29.

Zhang, H., Schmidt, W. J., Shen, X., Cao, Q., Monka, S., and Paschke, A. (2025). Knowledge graph construction towards a graph RAG-enhanced intelligent maintenance chatbot. In *International Workshop on Scaling Knowledge Graphs for Industry 2025*.