

Optimize RAG Results: Best Embeddings, LLMs, Prompts, & Retrievals

One of the most powerful aspects of Nuclia is its ability to **find the perfect combination of embeddings, LLMs, prompts, and retrieval strategies tailored to any specific use case**. Whether you are dealing with highly specialized industry data or general-purpose information, Nuclia's flexible framework allows you to configure and optimize every aspect of your RAG process. This means that no matter your specific requirements, you can rely on Nuclia to deliver the most effective and efficient results.

To deliver the best possible results, we need to consider numerous factors, including table/image captioning, embeddings, LLM usage, prompt design, and the use of visual LLMs. Additionally, it's crucial to determine the optimal RAG strategy in real-time while balancing the associated costs. Our mission at Nuclia is to provide a tailored RAG solution that meets your specific data needs. Here's how we do it:

1. Visual LLM RAG Support

Nuclia offers support for visual LLMs at both query and extraction times. This capability is essential for processing data that involves images or layouts, ensuring that your RAG solution can handle all types of information effectively.

2. Multiple Embedding Options

Choosing the right embedding model for your data is critical. Nuclia allows for real-time evaluation of multiple models, giving you the flexibility to select the best option for your specific use case. This approach ensures that you get the most accurate and relevant results.

3. Best-in-Class Nuclia Models

Out of the box, Nuclia provides top-tier reranking, embedding, segmentation, and extraction models. These models are designed to offer superior performance, allowing you to extract maximum value from your data.

4. Access to ALL public LLMs and embeddings

Nuclia gives you access to all publicly available Large Language Models (LLMs) and embeddings in the market. This extensive selection enables you to choose the tools that best fit your needs and to experiment with different configurations to find the optimal setup.

5. Enrich your data with entity relations (NERs) and Classification

During data ingestion, Nuclia can perform entity relation extraction, classification, and custom summarization to enrich your data. These features ensure that your data is not only well-organized but also highly informative and actionable.

6. REMi for Configuration Effectiveness

To evaluate the effectiveness of each configuration, Nuclia utilizes [REMi](#), an open-source RAG evaluation model. This tool helps you understand which strategies are working best and where there might be room for improvement.

7. Flexible RAG Strategies

Nuclia supports multiple text block combinations and [RAG strategies](#), allowing you to customize your approach to data retrieval and generation. Whether you need full-text search, Knowledge Graph integration, or vector search, Nuclia has you covered.

8. Domain-Specific Fine-Tuning

Fine-tuning your LLM for specific domains is crucial for achieving high accuracy and relevance. Nuclia provides the tools you need to fine-tune your models, ensuring that they are perfectly aligned with your industry or field.

9. Hybrid Search Options

Nuclia offers hybrid search options that combine full-text, Knowledge Graph, and vector search capabilities. This multi-faceted approach ensures that you can retrieve the most relevant information quickly and efficiently.

Achieve the Best Results with Nuclia

With Nuclia, you can upload your data, define your ingestion and RAG configurations, and achieve the best possible results. Whether you prefer to manually configure your setup or let our system try multiple approaches to deliver the highest REMi Score based on your specific queries, Nuclia provides the flexibility and tools you need.

Nuclia empowers you to configure your RAG solution with precision, delivering the tools and strategies necessary to define the best possible approach for your data. No matter the complexity or specificity of your needs, Nuclia ensures that your RAG process is as effective and efficient as possible.