

Bia

Sistema Inteligente de Recuperação de Informações sobre o Bacharelado em Inteligência Artificial

Alunos: André Cerqueira, Dayane
Rodrigues, Hugo Pessoni e Pedro
Rabelo

Índice

- 1. Introdução**
- 2. Problema**
- 3. Objetivo**
- 4. Metodologia**
- 5. Resultados**
- 6. Conclusões**

Introdução

O Bacharelado em Inteligência Artificial (BIA) da UFG, lançado em 2020, é pioneiro e desperta grande interesse. Contudo, devido à sua recente criação, muitos alunos e vestibulandos não têm acesso fácil a informações sobre o curso, o que é IA e as oportunidades profissionais na área, tornando necessária uma solução para facilitar esse acesso.



**Curso Recente
e Inovador**



**Empregabilidade
e CEIA?**



**Desinformação
sobre o Curso**

Resultados do Formulário de Pesquisa

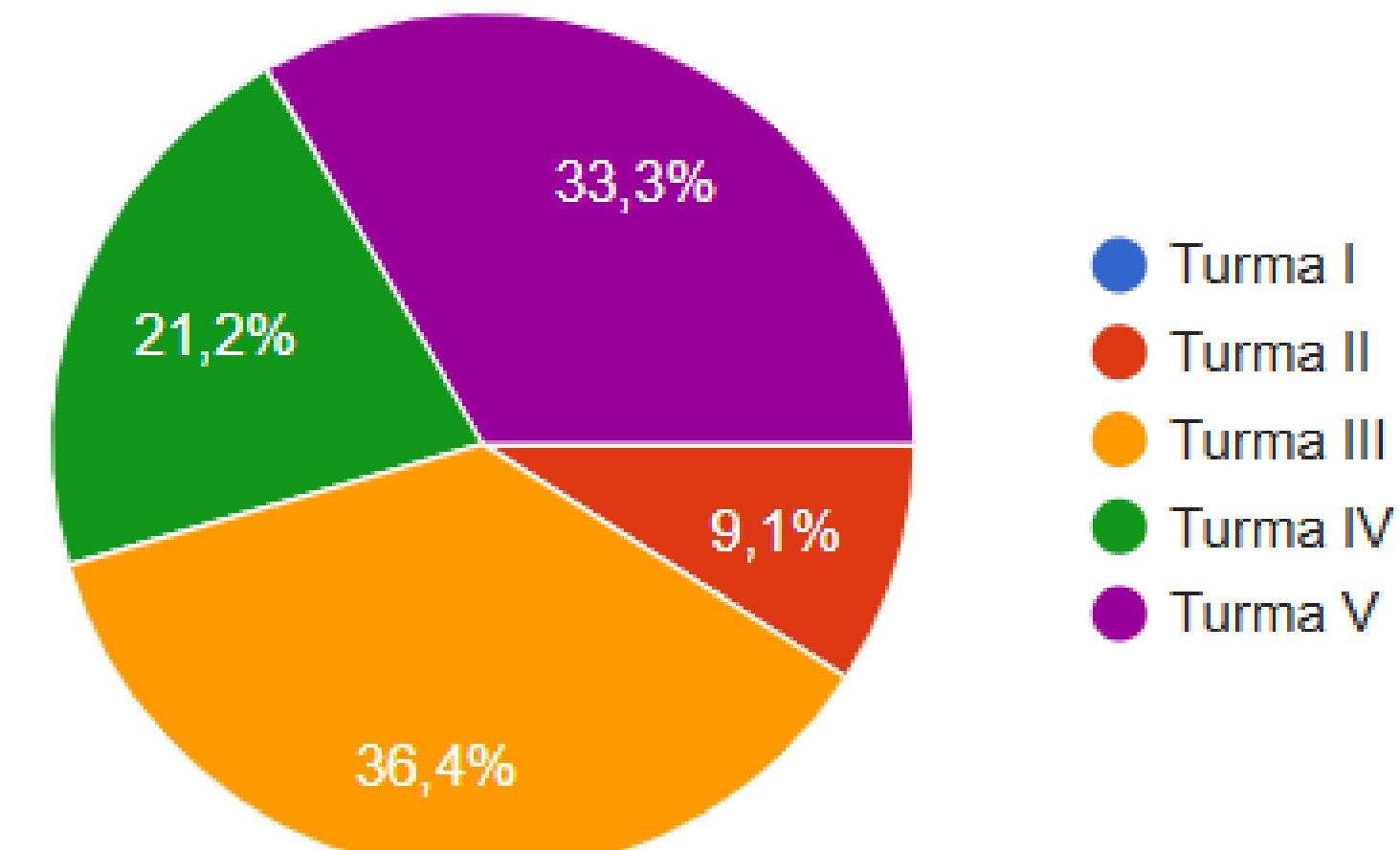
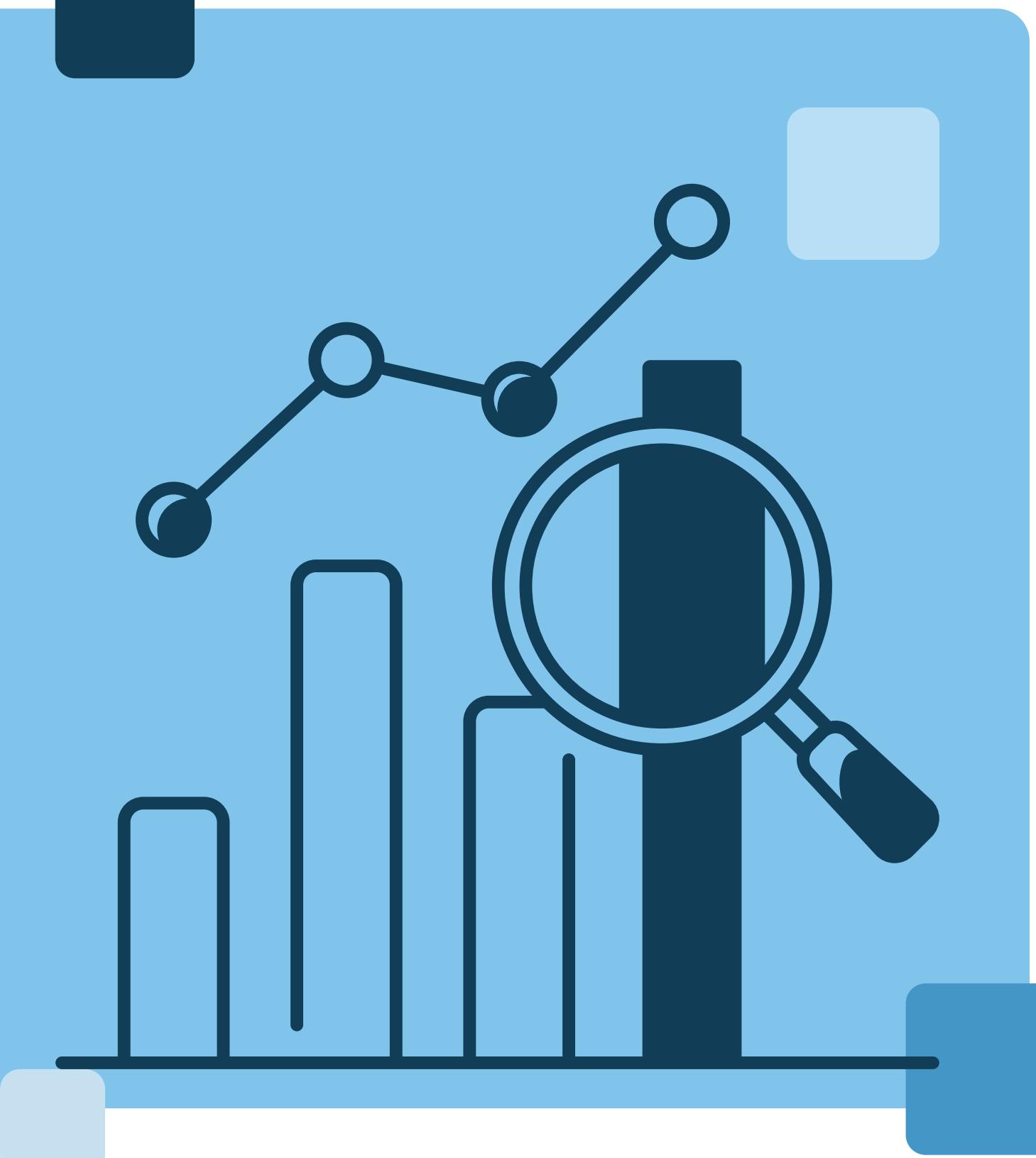
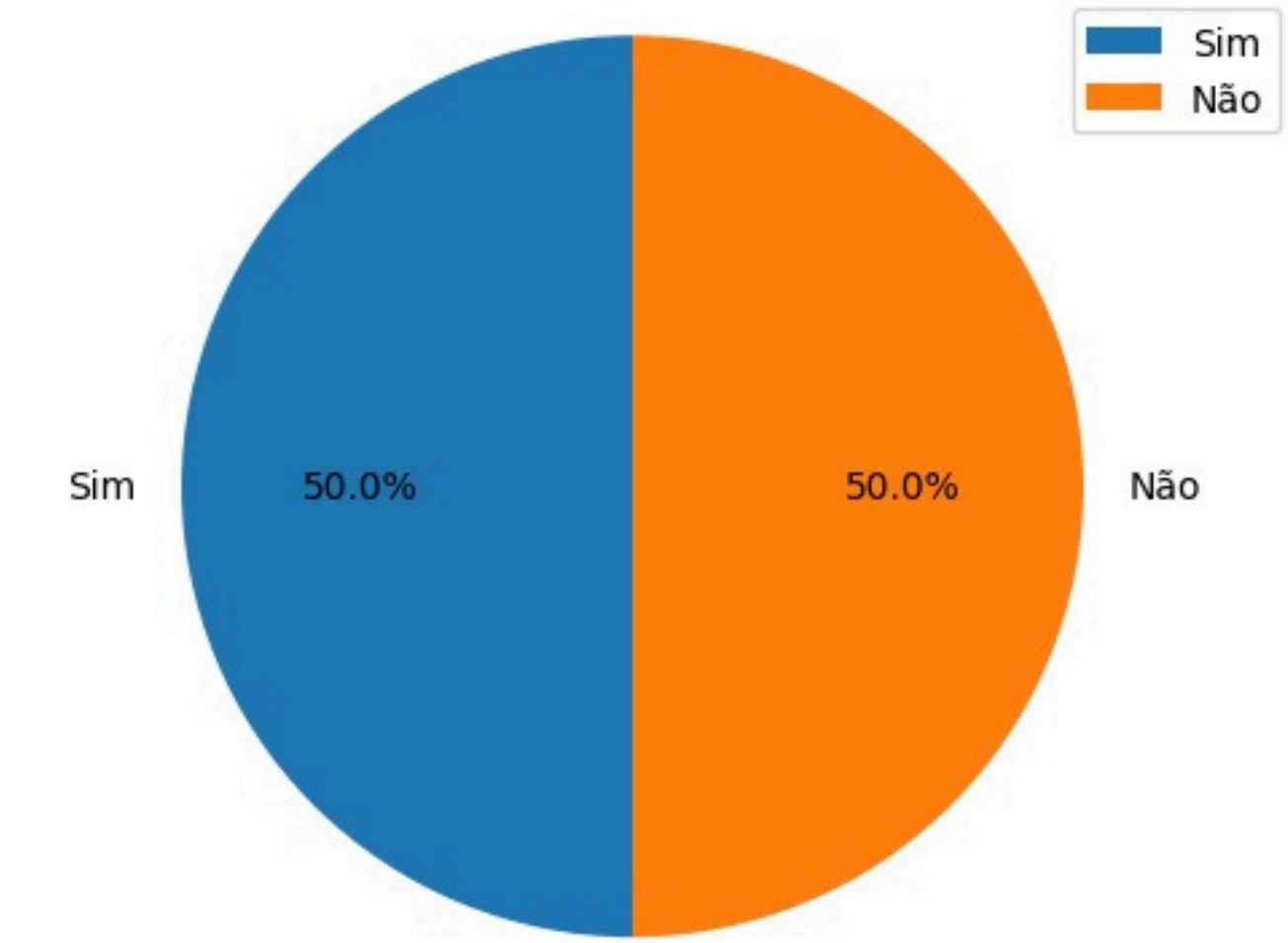
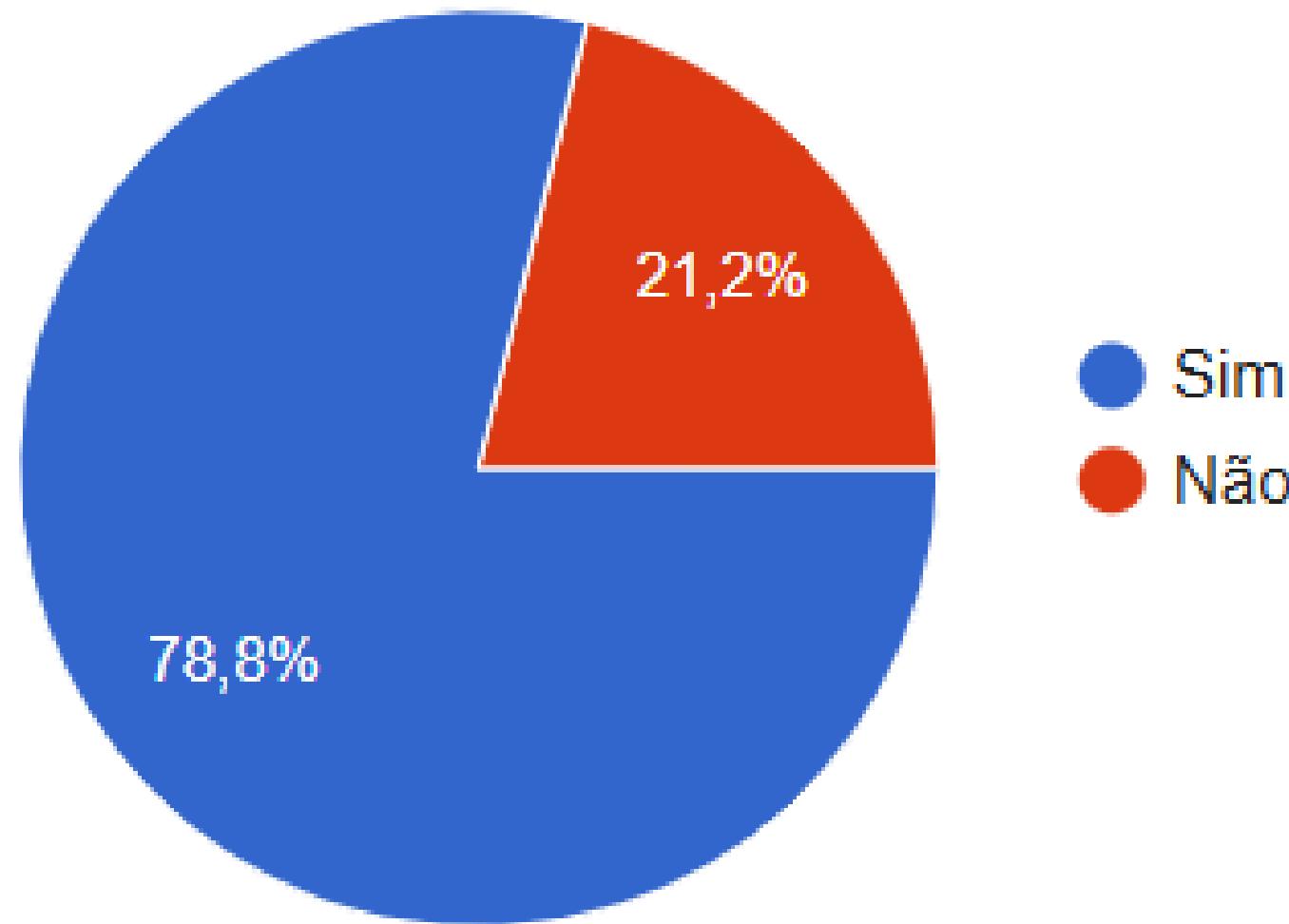


Gráfico 1: Presença das Turmas

Alunos no fluxo curricular das matérias



Insights dos Estudantes

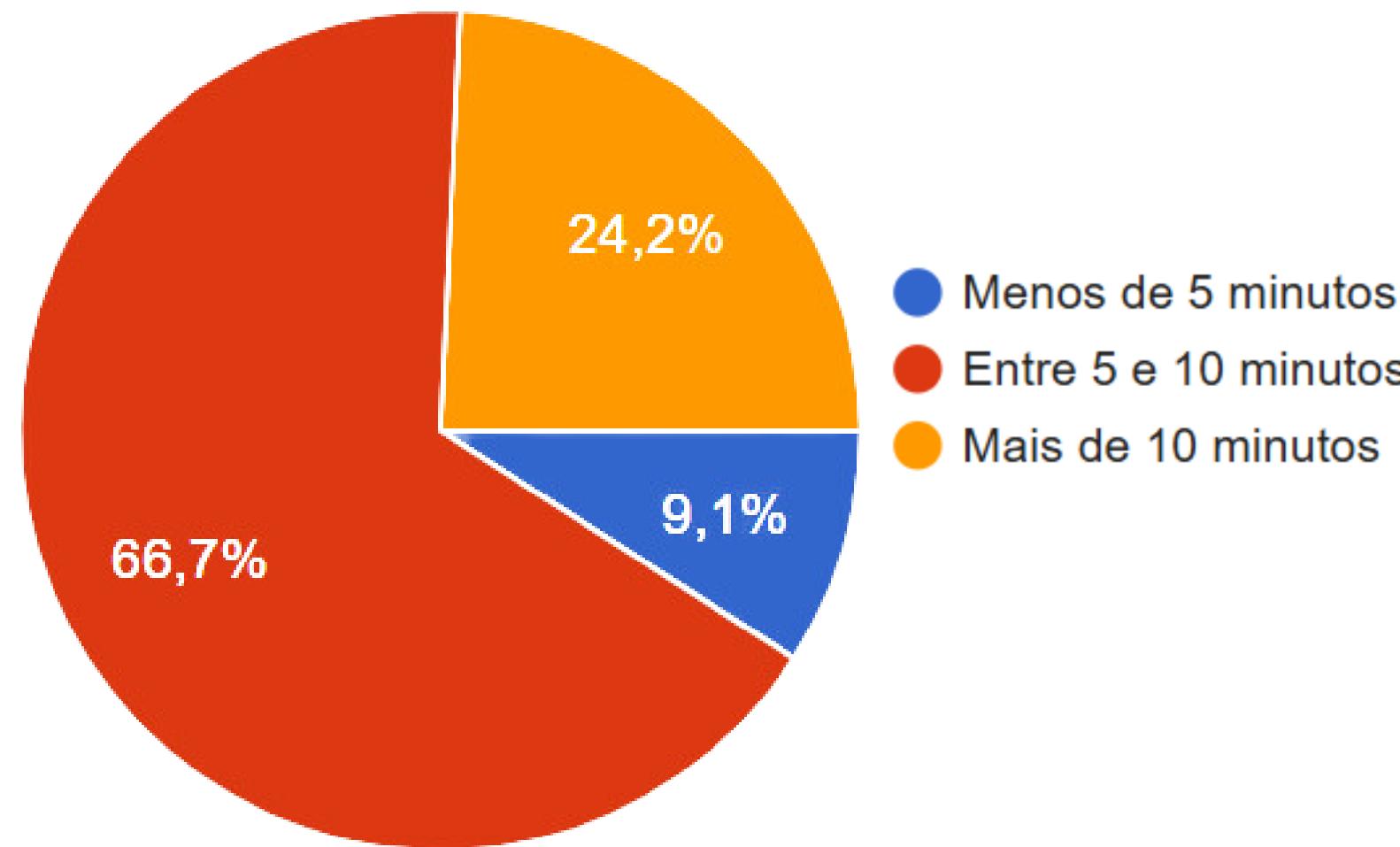


Gráfico 4: Tempo dos Alunos
para encontrarem informações



Gráfico 3: Leitura dos Alunos
do PPC do curso

Feedback dos Estudantes

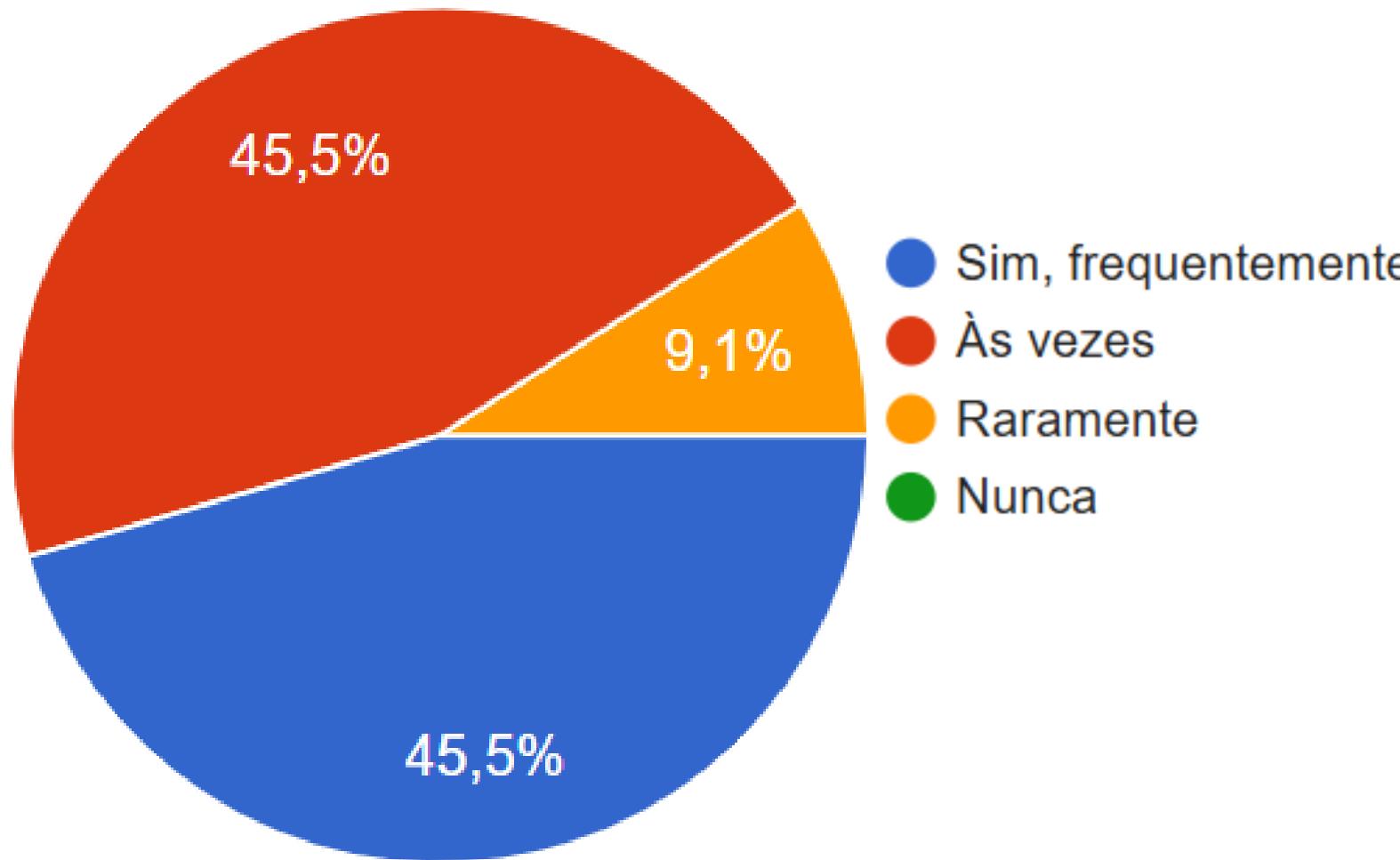


Gráfico 5: Dúvidas dos Alunos do que será abordado na matéria



Gráfico 6: Opnião sobre a criação de uma ferramenta para dúvidas do curso

Problema

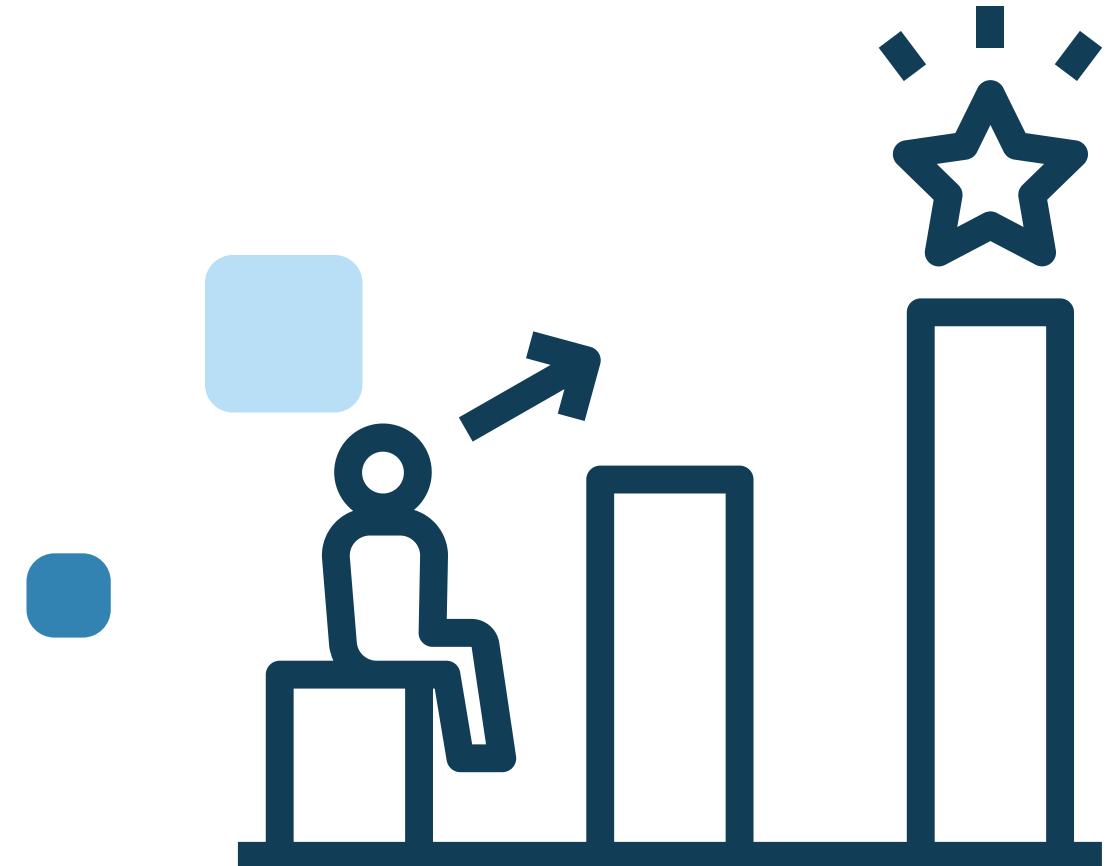
Abaixo estão os principais problemas enfrentados pelos estudantes ao buscar informações acadêmicas:

**Falta de Centralização
das Informações
Acadêmicas em uma
Única Plataforma de
Fácil Acesso**

**Respostas Genéricas
e Inadequadas para
Questões Específicas
dos Estudantes**

**Dificuldade em Acessar
Informações Atualizadas
e Alinhadas com as
Mudanças Recorrentes
nos Cursos**

Objetivos



Objetivo principal

- Desenvolver uma aplicação eficiente para recuperação de informações sobre o BIA.

Objetivos específicos

- Implementar e validar os métodos de RAG Tradicional e Graph RAG para recuperar informações sobre o curso de BIA.
- Avaliar eficiência, precisão e relevância das respostas.

O que é RAG (Retrieval-Augmented Generation)?

- **Definição:** Combinação de recuperação de dados e geração de texto.
- **Funcionamento:** Busca informações relevantes antes da geração.
- **Uso de Dados de Domínio:** Utiliza dados específicos do domínio para **aumentar a precisão nas respostas.**

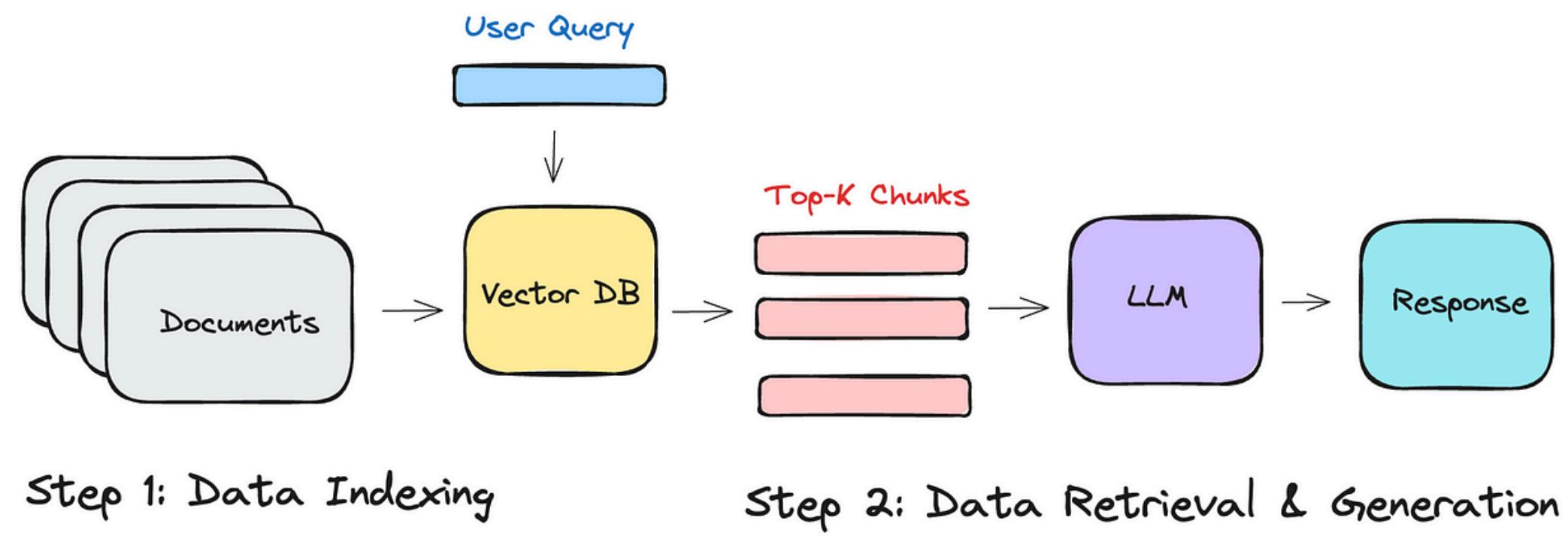
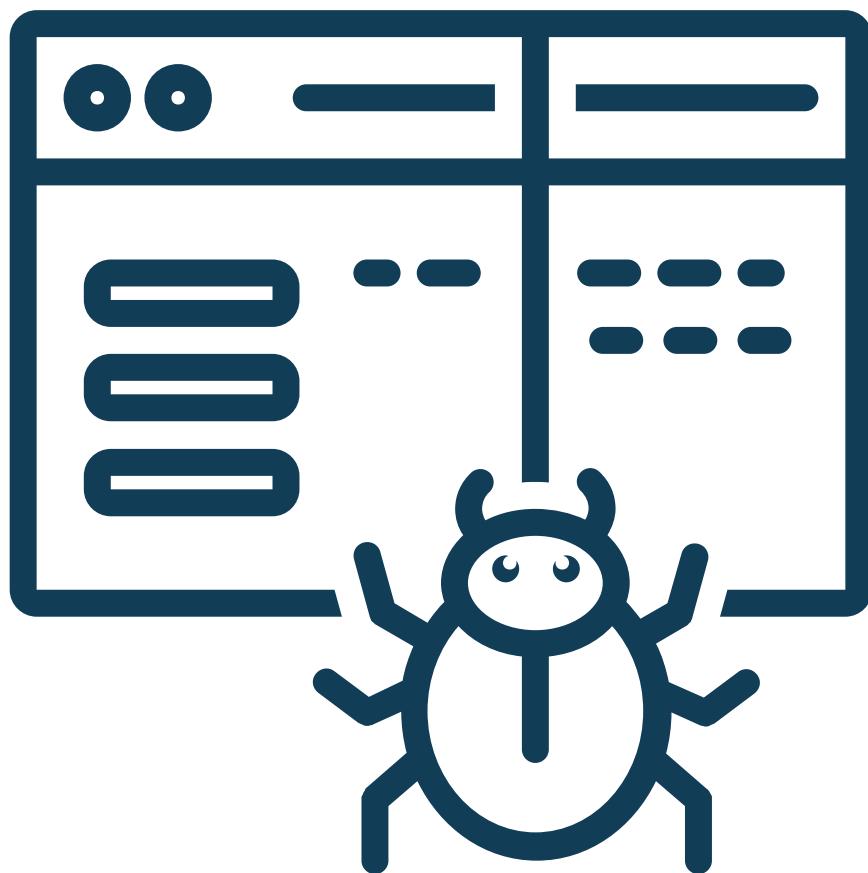


Figura 1: Arquitetura básica do RAG.

Metodologia



Metodologia



Crawler



Coleta de Dados

Fontes de Dados do Projeto (txt) - 224 páginas

- **Projeto Pedagógico do Curso (PPC):** Documento oficial com informações sobre a estrutura e a grade curricular do BIA.
- **Entrevista com o Coordenador Anderson Soares:** Transcrição do quadro Guia das Profissões do Brasil Escola.
- **Fontes Web:**
 - Site oficial do BIA.
 - Notícia sobre a criação do curso.
 - Publicação da UFG relacionada ao BIA.

Metodologia



RAG TRADICIONAL

- **Indexação**
 - Dados são indexados em *chunks*, sendo cada *chunk* um nó (texto).
- **Biblioteca de Implementação:**
 - **LlamalIndex**
 - Facilita a construção de pipelines RAG;
 - Integra recuperação e geração de forma eficiente.





Banco de Dados Vetorial:

- **Qdrant**

- Versão Gratuita atende nossos requisitos;
- Excelente performance na indexação e recuperação;
- Suporte a busca híbrida (esparsa + densa) na versão pro (próximos passos);

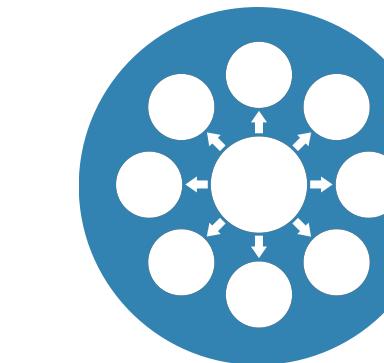


- **Tipo de Busca:**

- **Densa:**

- Tipo: Similaridade de cosseno:
 - Modelo de *Embedding*: text-embedding-3-small (OpenAI).
 - Custo-benefício e desempenho.

Metodologia



Graph Rag

- Alucinações;
- Falta de explicações;
- Falta de conhecimento de domínio;
- Responder “não”;
- Versão avançada;
- Ao invés de vetores, grafos;
- Nós e Arestas;
- Relações hierárquicas e não hierárquicas;
- Maior contextualização.

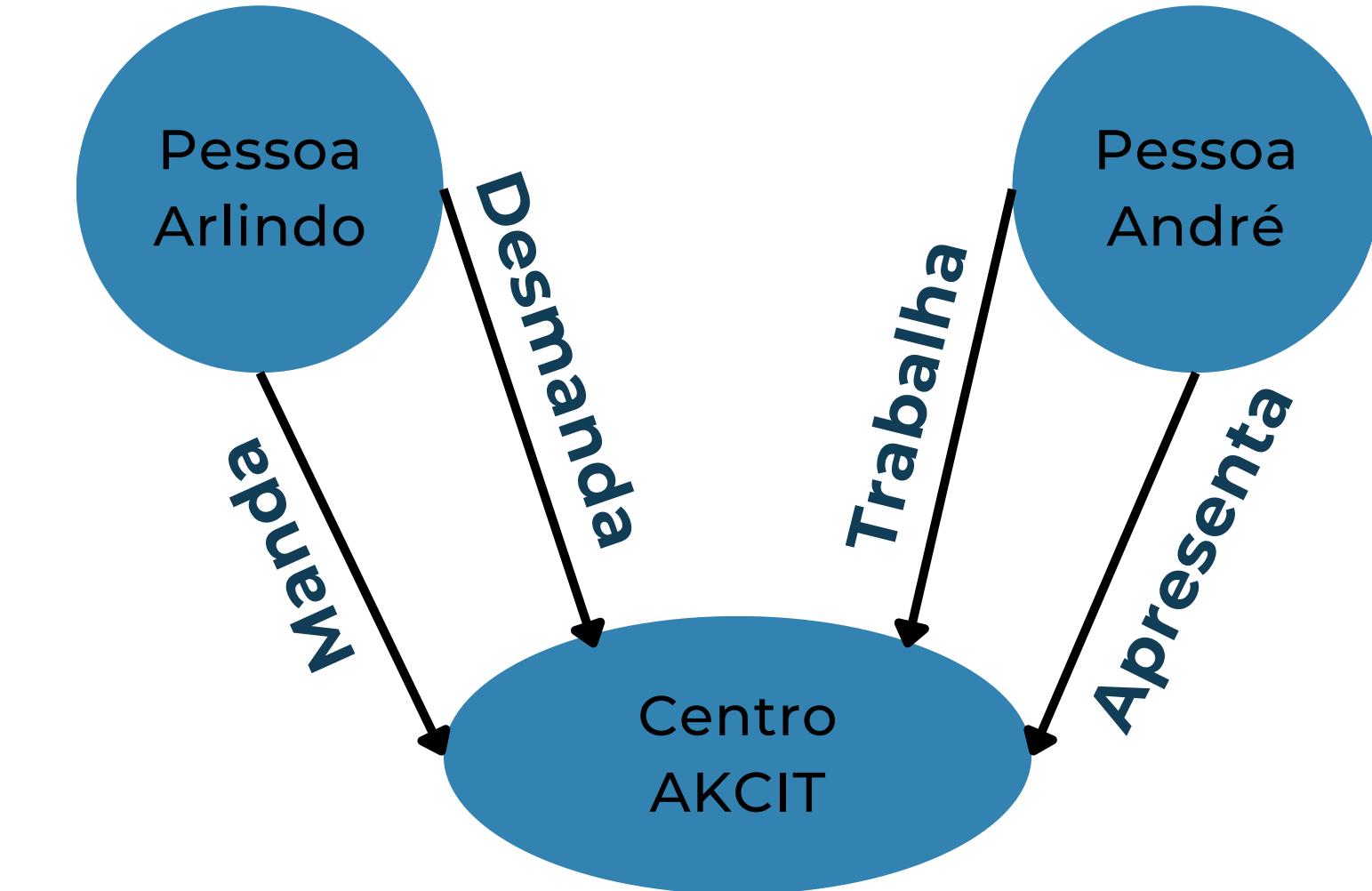
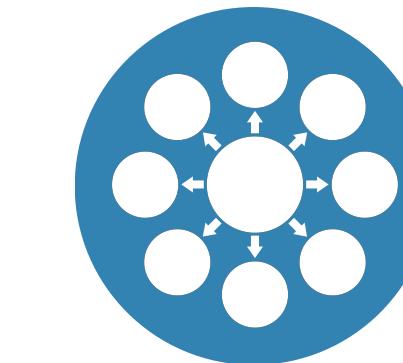


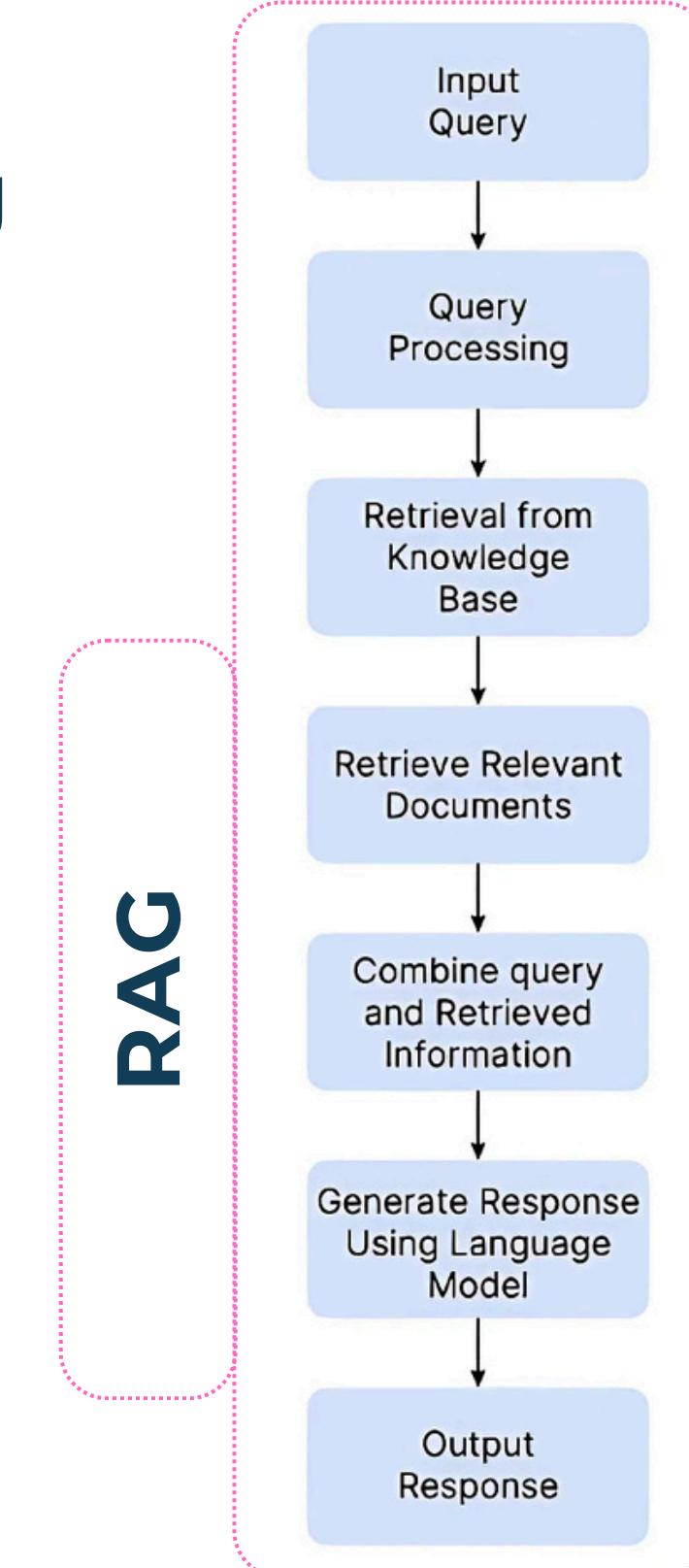
Figura 2: Exemplo de Grafo.

Metodologia



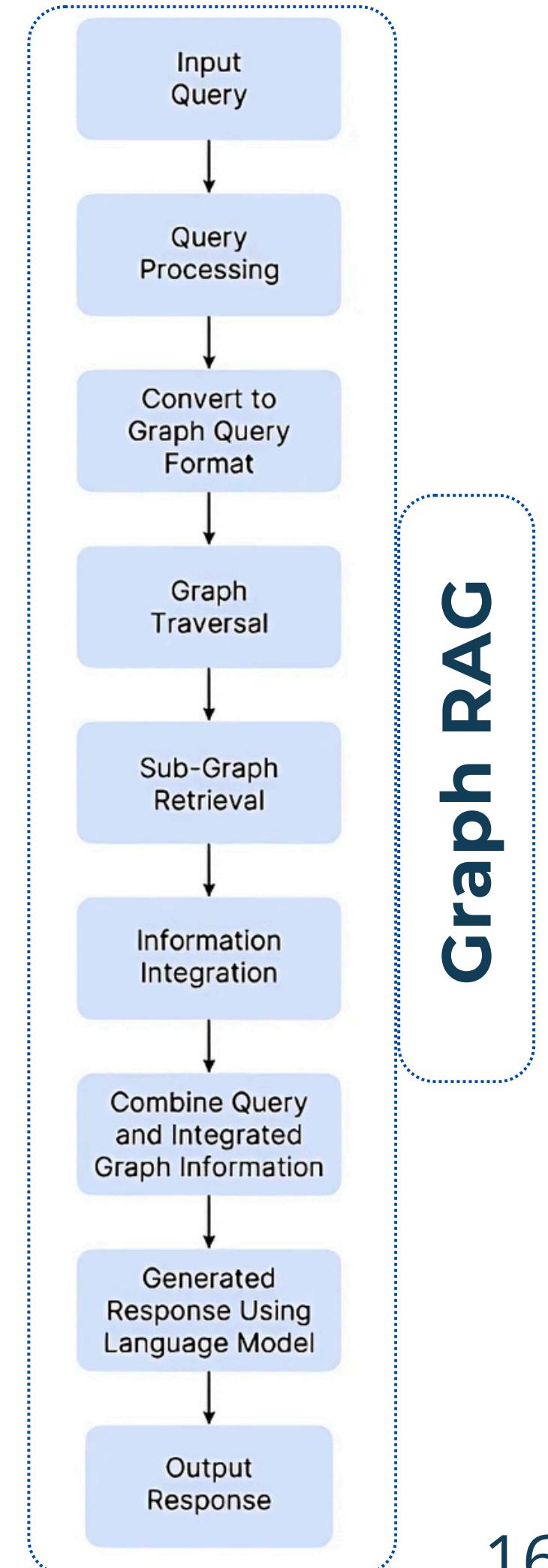
- E como funciona na prática?

Graph Rag



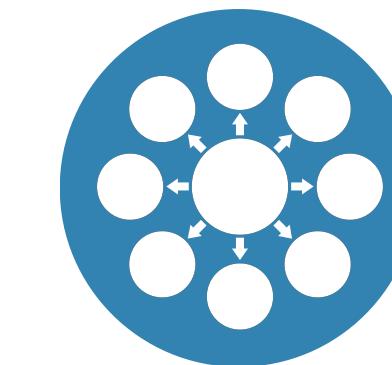
RAG

Figura 3: Pipilene Tradicional RAG vs GraphRAG



Graph RAG

Metodologia



Graph Rag

- **E como funciona na prática?**
- Query Processing: converte uma query para o formato de consulta em grafos.
- Graph Traversal: explora a estrutura do grafo seguindo as relações mais relevantes para encontrar infos conectadas.
- Subgraph Retrieval: em vez de buscar infos isoladas, busca por sub-grafos.

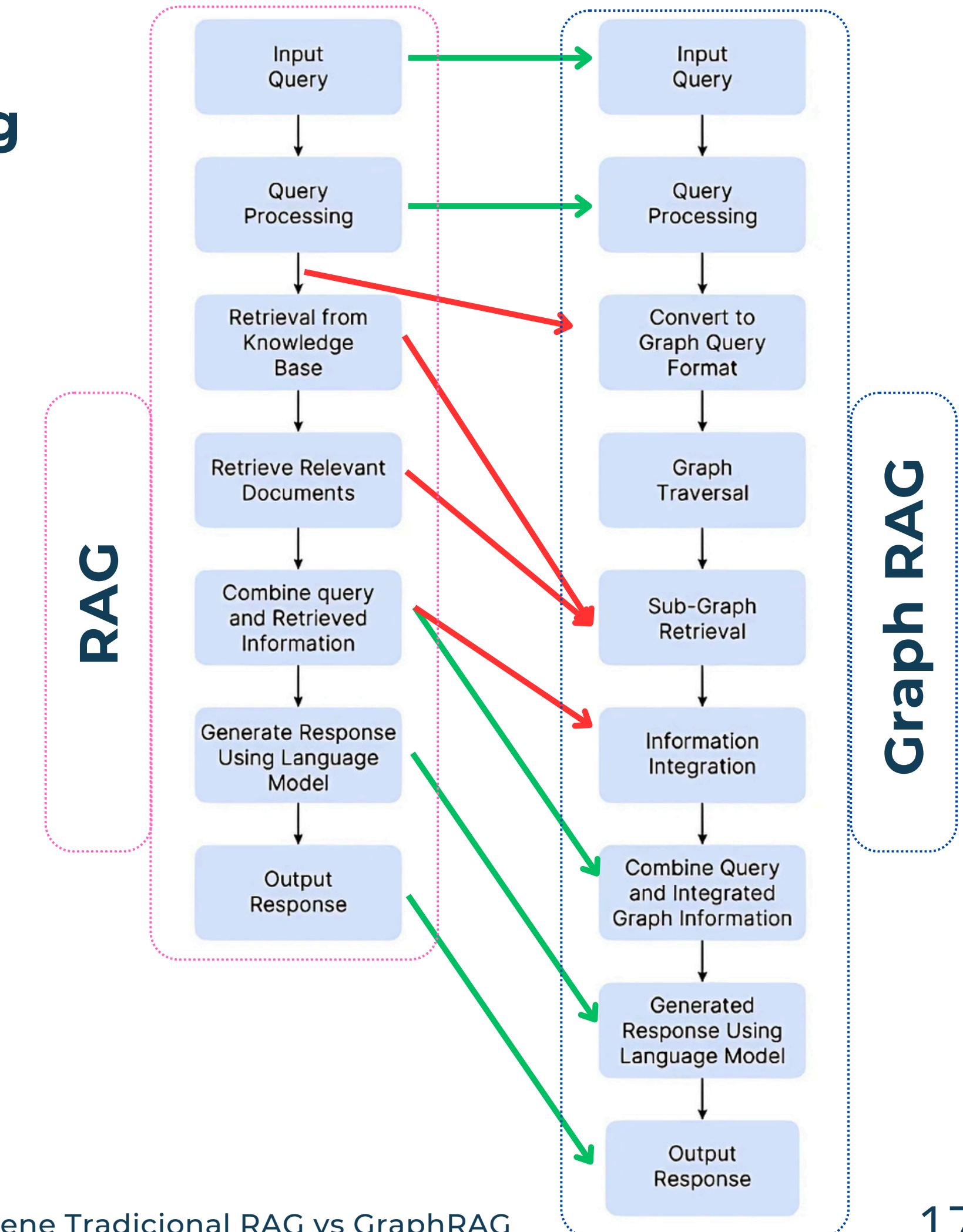
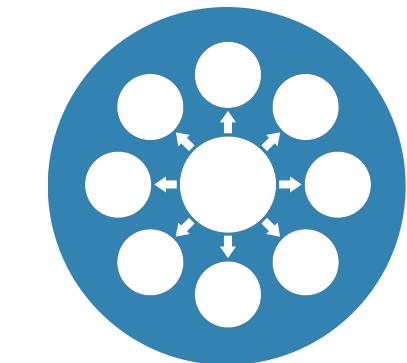


Figura 3: Pipilene Tradicional RAG vs GraphRAG

Metodologia



Graph Rag

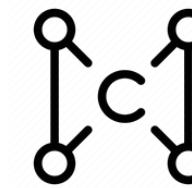
5 documentos interligados.

A screenshot of the Neo4j Graph Store interface. The main area displays a message: "No instance connected" with the sub-instruction "Start exploring the graph by connecting to a instance." Below this is a "Connect to instance" button. To the left, there is a small illustration of a flask with a dropper. A green cursor arrow is visible on the left side of the screen.

- **Graph Store:**



- Visualização;
- Consultas;
- Manipulação;
- Rastreamento;
- Cypher.



- Custo de geração do grafo;
- Sensível a dados ruidosos ;
- Tempo de resposta maior;

Vídeo 1: Graph Store no Neo4j.



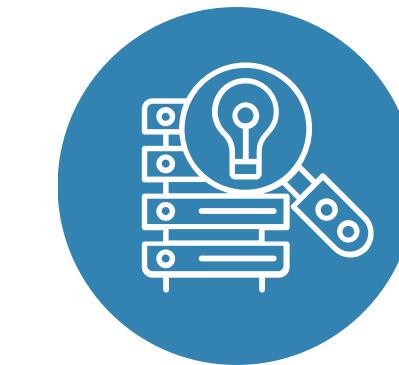
- **Recuperação esparsa**

- Relevância determinada por palavras chaves.
- Utiliza representações esparsas de texto.
- Ex: TF-IDF e BM25.

- **Recuperação densa**

- Busca semântica;
- Baseado em redes neurais e modelos pré-treinados (como BERT).
- Usa representações densas geradas por embeddings.

A recuperação híbrida combina os dois métodos acima!



Reranking reorganiza os nós recuperados para garantir que os resultados mais relevantes apareçam no topo.

- **Funcionamento**

- **1º) Dada a recuperação inicial** pelo recuperador (similaridade);
- **2º) Reranking:** modelo mais sofisticado é utilizado para reavaliar a relevância dos resultados
- **3º) Resultado final:** documentos mais importantes são levados ao topo da lista (rerankeamento)

- **Modelo escolhido: Cohere Rerank**

- **Motivos:** desempenho e facilidade de uso.



Metodologia



RAG Tradicional

- **Vector Store:**
 - Qdrant
- **Implementação:**
 - LlamalIndex
- **Busca Densa:**
 - Similaridade de Cosseno
 - text-embedding-3-small
- **Reranking:**
 - Cohere Reranker.
- **LLM:**
 - gpt-4o-mini (**Custo Benefício**)
- **Chunk_size:** 256
- **Top k:** 5

RESUMO DAS PRINCIPAIS DIFERENÇAS

GraphRAG

- **Graph Store:**
 - Neo4j
- **Implementação:**
 - LlamalIndex
- **Busca Híbrida:**
 - Similaridade de Cosseno
 - text-embedding-3-small + BM25
- **Reranking:**
 - Cohere Reranker.
- **LLM:**
 - gpt-4o-mini (**Custo Benefício**)
- **Chunk_size:** 256
- **Top k:** 5



Personalização do Agente e Template

Você é um assistente especializado no curso de Bacharelado em Inteligência Artificial (BIA). Seu objetivo é fornecer respostas objetivas, gentis e formais às dúvidas apresentadas, com base no contexto fornecido.

Contexto: {context}

Pergunta: {question}

- Seja objetivo, claro e preciso.
- Use uma linguagem formal e educada.
- Baseie-se no contexto fornecido para criar respostas completas e informativas.
- Se a resposta não estiver claramente presente no contexto, diga que apenas "Não posso responder à sua pergunta: " (cite a pergunta)". Estou à disposição para esclarecer outras dúvidas que você possa ter."

Avaliação: LLM como Juiz

Notas:

- 0.0 = documento não relevante;
- 0.5 = documento razoavelmente relevante;
- 1.0 = documento relevantes

Pergunta	Nós Recuperados (n)	Answer Relevance (n=5)
Qual é o perfil do egresso do curso?	node1, node2, ..., noden	[0.0, 1.0, 1.0, 0.5, 0.5]
Há disciplinas optativas no curso?	node1, node2, ..., noden	[1.0, 0.5, 1.0, 0.0, 0.0]
Quais são os temas abordados na disciplina de ?	node1, node2, ..., noden	[0.5, 0.0, 0.0, 1.0, 1.0]
O curso tem alguma ligação com empreendedorismo?	node1, node2, ..., noden	[1.0, 0.5, 0.5, 1.0, 0.0]
Em qual período os estudantes aprendem NLP?	node1, node2, ..., noden	[0.5, 0.5, 0.0, 0.0, 1.0]

Tabela 1: Avaliação do LLM como juiz.

Resultados (métricas)

- Queries = 50

PERGUNTA	SAÍDAS IRREGULARES	PRECISION @K (k=5)	MEAN RECIPROCAL RANK	AVERAGE SIMILARITY	LATÊNCIA MÉDIA
RAG TRADICONAL	0	0.53	0.76	0.53	1,8s
GRAPH RAG	0	0.52	0.75	0.54	2,5s

Tabela 2: Resultados das Métricas.

● Queries e Respostas

RAG Tradicional

- @andrecqcastro → /workspaces/SIRI_BIA/Tradisional_RAG (main) \$ python retriever.py
Pergunta: Quais são os pré-requisitos da matéria de NLP?

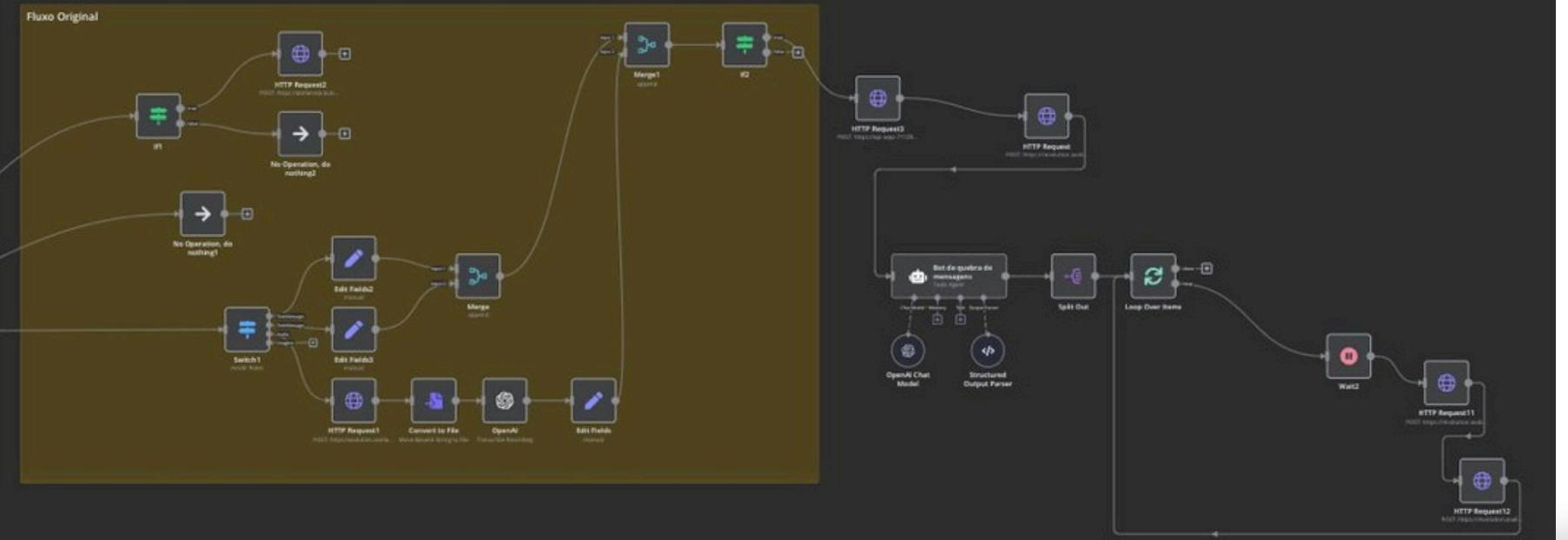
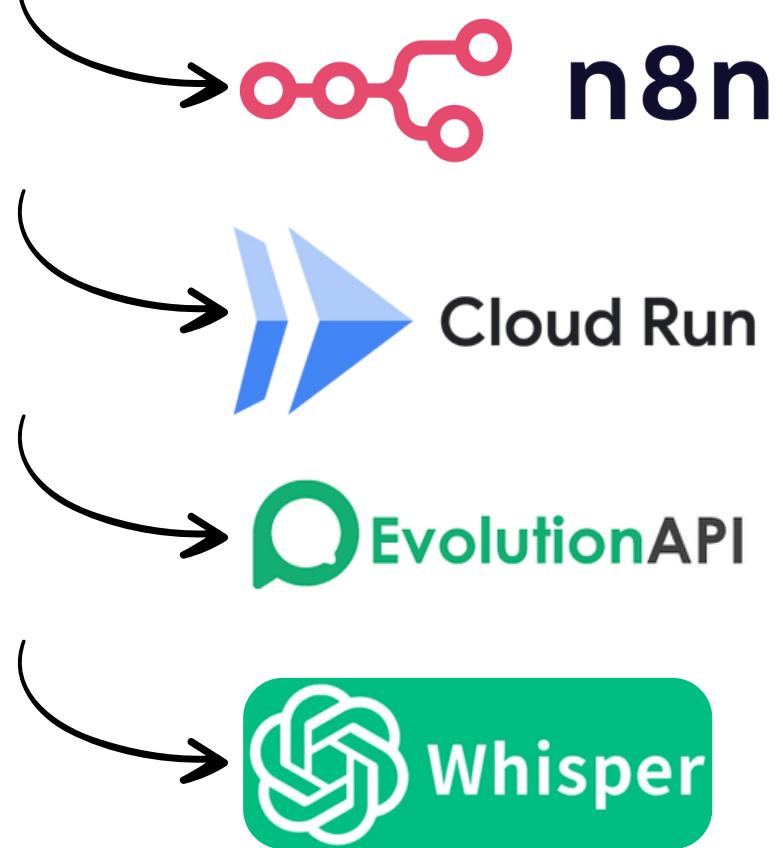
Resposta: Os pré-requisitos da matéria de Processamento de Linguagem Natural (NLP) são: Redes Neurais Profundas. Estou à disposição para esclarecer outras dúvidas que você possa ter.

GRAPH RAG

- @andrecqcastro → /workspaces/SIRI_BIA/Graph_RAG (main) \$ python retriever.py
Pergunta: Quais são os pré-requisitos da matéria de NLP?

Resposta: Os pré-requisitos da disciplina de Processamento de Linguagem Natural (NLP) são Redes Neurais Profundas.

ChatBIA



Limitações



Limitações do Sistema RAG

- Atualização dos Dados;
- Informações Extraoficiais;
- Senso Comum Universitário;
- Acesso a Recursos Computacionais.

Conclusões

Os resultados mostram que o **RAG Tradicional** teve melhor Precision@K (0.53) e Mean Reciprocal Rank (0.76), sendo mais **eficiente em consultas diretas**. Já o **Graph RAG** apresentou maior Average Similarity (0.54), capturando **melhor as relações contextuais**. O sistema provou ser eficaz, equilibrando precisão e relevância na recuperação de informações.

OBRIGADO!