



Keyson Raphael Acioli da Conceição

Geração Aumentada para Recuperação de Dados Urbanos Integrados: consolidando dados do IBGE, Censo, CNEFE e OSM para a otimização do planejamento urbano

Recife

2025

Keyson Raphael Acioli da Conceição

Geração Aumentada para Recuperação de Dados Urbanos Integrados: consolidando dados do IBGE, Censo, CNEFE e OSM para a otimização do planejamento urbano

Monografia apresentada ao Curso de Bacharelado em Sistemas de Informação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Universidade Federal Rural de Pernambuco – UFRPE

Departamento de Estatística e Informática

Curso de Bacharelado em Sistemas de Informação

Orientador: Rinaldo Lima

Recife

2025

Dados Internacionais de Catalogação na Publicação
Sistema Integrado de Bibliotecas da UFRPE
Bibliotecário(a): Auxiliadora Cunha – CRB-4 1134

C744g Conceição, Keyson Raphael Acioli da.
Geração aumentada para recuperação de dados Urbanos Integrados: consolidando dados do IBGE, Censo, CNEFE e OSM para a otimização do planejamento urbano / Keyson Raphael Acioli da Conceição. – Recife, 2025.
66 f.; il.

Orientador(a): Rinaldo José de Lima.

Trabalho de Conclusão de Curso (Graduação) – Universidade Federal Rural de Pernambuco, Bacharelado em Sistemas da Informação, Recife, BR-PE, 2025.

Inclui referências e apêndice(s).

1. Planejamento Urbano. 2. Inteligência Artificial. 3. RAG. 4. Linguagem de programação (Computadores) 5. Linguagem de programação (Computadores) - Semântica. I. Lima, Rinaldo José de, orient. II. Título

CDD 004

Keyson Raphael Acioli da Conceição

**Geração Aumentada para Recuperação de Dados Urbanos
Integrados: consolidando dados do IBGE, Censo, CNEFE e OSM
para a otimização do planejamento urbano**

Monografia apresentada ao Curso de Bacharelado em Sistemas de Informação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Aprovada em: 21 de março de 2025

BANCA EXAMINADORA

Rinaldo José de Lima(Orientador)
Departamento de Computação
Universidade Federal Rural de Pernambuco

George Gomes Cabral
Departamento de Computação
Universidade Federal Rural de Pernambuco

Dedico este trabalho aos meus pais, pelo amor incondicional e pelo apoio em todos os momentos, e aos meus professores, que foram fundamentais na minha formação acadêmica...

Agradecimentos

Em primeiro lugar, agradeço a Deus pela força e perseverança ao longo desta jornada. À minha família, pelo amor, apoio e incentivo. Ao meu orientador, Professor Rinaldo, sem cuja paciência e orientação fundamentais este trabalho não seria possível. E a todos aqueles que, pessoalmente ou de outra forma, contribuíram de alguma forma para o sucesso deste trabalho, meus sinceros agradecimentos....

“Não basta interpretar o mundo; é preciso transformá-lo.”
– Karl Marx (*Tese sobre Feuerbach*)

Resumo

Nos últimos anos, os campos da Inteligência Artificial (IA) e do aprendizado de máquina (AM) revolucionaram o domínio do planejamento urbano, pois permitem que volumes substanciais de dados sejam analisados de forma eficaz, incentivando melhor alocação de recursos e entregas de serviços públicos. Para atingir este objetivo, o agente inteligente proposto neste trabalho reúne dados de várias fontes, incluindo Censo Demográfico, Cadastro Nacional de Endereços para Fins Estatísticos - CNEFE, e OpenStreetMap (OSM) para oferecer respostas baseadas em contexto relacionadas à distribuição da população e acesso a diferentes serviços urbanos. A abordagem proposta inclui um pipeline de processamento que implementa normalização, indexação vetorial das informações e representação semântica para tornar as consultas mais eficazes. Para avaliar o sistema proposto, foi conduzido um experimento com especialistas em planejamento urbano e analisamos a relevância, clareza e utilidade das respostas geradas pelo sistema. Tais resultados mostram que o agente é capaz de detectar áreas com pouca cobertura de serviços necessários, indicando uma alocação adequada. No entanto, outros desafios, tais como a necessidade de melhor clarificação das respostas e ampliação da cobertura espacial, foram reconhecidos como oportunidades para trabalho futuro.

Palavras-chave: Planejamento Urbano. Inteligência Artificial. RAG (para Geração Aumentada por Recuperação). Modelos de Linguagem. Análise Geoespacial.

Abstract

In recent years, the fields of Artificial Intelligence (AI) and Machine Learning (ML) have revolutionized the urban planning domain, as they allow substantial volumes of data to be analyzed effectively, encouraging better resource allocation and public service delivery. To achieve this goal, the intelligent agent proposed in this work gathers data from several sources, including the Census, the National Address Registry for Statistical Purposes (CNEFE), and OpenStreetMap (OSM) to provide context-based answers related to population distribution and access to different urban services. The proposed approach includes a processing pipeline that implements normalization, vector indexing of information, and semantic representation to make queries more efficient. To evaluate the proposed system, an experiment was conducted with urban planning experts and we analyzed the relevance, clarity, and usefulness of the responses generated by the system. These results show that the agent is capable of detecting areas with little coverage of necessary services, indicating an appropriate allocation. However, other challenges, such as the need for better clarification of responses and expansion of spatial coverage, were recognized as opportunities for future work.

Keywords: Urban Planning. Artificial Intelligence. RAG (for Retrieval Augmented Generation). Language Models. Geospatial Analysis.

Lista de ilustrações

Figura 1 – Vista aérea de uma cidade. Fonte: (SEVILLA, 2013).	26
Figura 2 – Fluxo RAG da proposta	34
Figura 3 – Fluxo de Indexação e Recuperação dos Dados; Fonte: (AWS, 2025).	35
Figura 4 – Fluxo de validação de escolha de <i>embeddings</i>	40
Figura 5 – Fluxo de Indexação e Recuperação dos Dados	42
Figura 6 – Distribuição dos registos do corpus por município.	45
Figura 7 – Gráfico de especialistas participantes	53
Figura 8 – Resposta produzida pelo modelo	54
Figura 9 – Resposta geográfica produzida pelo modelo	54

Lista de tabelas

Tabela 1 – Resumo dos Trabalhos Relacionados (Parte 1)	29
Tabela 2 – Resumo dos Trabalhos Relacionados (Parte 2)	30
Tabela 3 – Modelos de <i>Embeddings</i> Avaliados	38
Tabela 4 – Resultados das Avaliações dos Embeddings (Base: 200 Consultas) . .	41
Tabela 5 – Principais atributos do corpus integrado.	45
Tabela 6 – Hiperparâmetros Utilizados na Indexação FAISS com HNSW	47
Tabela 7 – Simulação de Impacto da Variação de k na Consulta Exemplo	50
Tabela 8 – Amostra de Dados do CNEFE Processados	64
Tabela 9 – Resultados da Avaliação com Especialistas	65

Lista de abreviaturas e siglas

TCC	Trabalho de Conclusão de Curso
UFRPE	Universidade Federal Rural de Pernambuco
GIS	Geographic Information Systems
LangChain	Framework de processamento de linguagem natural
IBGE	Instituto Brasileiro de Geografia e Estatística
Censo	Pesquisa populacional realizada pelo IBGE
CNEFE	Cadastro Nacional de Endereços para Fins Estatísticos
API	Application Programming Interface
OSM	OpenStreetMap
MDPI	Multidisciplinary Digital Publishing Institute
arXiv	Arquivo de pré-publicações científicas
Folium	Biblioteca Python para criar mapas interativos
IA	Inteligência Artificial
ML	Machine Learning
LLMs	Large Language Models
RAG	Recuperação Aumentada por Geração
FAISS	Facebook AI Similarity Search
HNSW	Hierarchical Navigable Small World
IDH	Índice de Desenvolvimento Humano
MRR	Mean Reciprocal Rank

Sumário

	Lista de ilustrações	9
1	INTRODUÇÃO	15
1.1	Objetivos	18
1.1.1	Objetivo Geral	18
1.1.2	Objetivos Específicos	18
1.2	Contribuições do Trabalho	18
1.3	Organização do Trabalho	19
2	FUNDAMENTAÇÃO TEÓRICA	21
2.1	Processamento de Linguagem Natural (PLN)	21
2.2	Deep Learning	21
2.2.1	Embeddings para Linguagem Natural	22
2.2.2	Arquitetura Transformers	22
2.2.3	Família BERT	22
2.3	Large Language Models	22
2.3.1	Prompt Engineering	23
2.3.2	Sistemas RAG	23
3	TRABALHOS RELACIONADOS	24
3.1	Aplicação de Modelos de Linguagem na Análise Urbana	24
3.2	Automação e Otimização do Planejamento Urbano	25
3.3	Sistemas de Informação Geográfica na Tomada de Decisão	25
3.4	Perspectivas em Políticas Urbanas e Ferramentas Inteligentes	26
3.5	Contribuição da IA para o Planejamento Urbano Moderno	28
4	PROBLEMA DE PESQUISA	31
5	PROPOSTA	33
5.1	Arquitetura Funcional da Proposta	33
5.2	Preparação dos Dados	34
5.2.1	Pré-processamento dos Dados	35
5.2.2	Integração das Fontes de Dados	36
5.2.3	Armazenamento e Indexação	36

5.2.4	Validação dos Dados	37
5.3	Embeddings e Indexação	37
5.3.1	Criação dos Embeddings	37
5.3.2	Modelos de Embeddings Avaliados	38
5.3.3	Métrica de Avaliação	39
5.3.4	Processo de Avaliação	39
5.3.5	Escolha do Modelo	40
5.3.6	Indexação FAISS e Recuperação dos Dados	41
6	EXPERIMENTOS	43
6.1	Corpus	43
6.1.1	Descrição	43
6.1.2	Distribuição	44
6.2	Limpeza de Dados	45
6.2.1	Pré-processamento	45
6.2.2	Integração de Dados	46
6.3	Configuração do Ambiente	46
6.4	Hiperparâmetros da Indexação FAISS	47
6.4.1	Resultados Experimentais Aplicados ao Planejamento Urbano	50
6.5	Avaliação do Modelo	52
6.5.1	Relevância das Perguntas	53
6.5.2	A Clareza das Respostas do Modelo	53
6.5.3	Utilidade para o Planejamento Urbano	55
6.5.4	Sugestões de Melhoria	55
7	CONCLUSÃO	56
7.1	Conclusão	56
7.2	Trabalho Futuro	57
	REFERÊNCIAS	60
Apêndices		62
.1	Código-Fonte Principal	63
.1.1	Carregamento de Bases de Dados	63
.1.2	Construção do Índice FAISS	63
.2	Amostra do Corpus Integrado	64
.3	Formulário e Resultados da Avaliação	64
.3.1	Formulário de Avaliação	64

.3.2	Resultados da Avaliação	65
.3.3	Resumo Estatístico	65

1 Introdução

O planejamento urbano é uma disciplina em evolução que se tornou um sistema social essencial para a qualidade de vida urbana.

À medida que a urbanização ganha impulso, exigindo uma maior oferta de uma ampla gama de serviços básicos — saúde, educação, mobilidade, entre outros — bem como de serviços mais sofisticados — incluindo crescimento inclusivo, resiliência urbana, mitigação climática, etc., o planejamento urbano torna-se mais crítico, e cada vez mais desafiador. Métodos tradicionais dependem de análises manuais e são, portanto, inadequados para acompanhar essa dinâmica, sendo assim necessário aplicar tecnologias mais sofisticadas, como a inteligência artificial (IA), para otimizar todos os processos e aumentar a qualidade de vida nas cidades (HE; CHEN, 2024). Este capítulo estabelece a relevância deste trabalho ao examinar onde a IA e os agentes inteligentes se encaixam no planejamento urbano, resumindo as abordagens existentes, bem como a lacuna que esta pesquisa busca abordar.

O planejamento urbano considera o arranjo e o design do ambiente urbano, incluindo seus sistemas de transporte. Planejar para o ambiente urbano é a arte e ciência de determinar como o espaço e os recursos são utilizados pela população para garantir o acesso igualitário aos recursos vitais e o crescimento e desenvolvimento sustentáveis. Tradicionalmente, esta disciplina empregava técnicas analíticas convencionais; no entanto, a crescente urbanização tem exigido abordagens mais ágeis e baseadas em dados. A IA destaca-se como um instrumento transformador, capaz de digerir grandes volumes de dados em áreas urbanas, normalmente grandes demais para os humanos analisarem, e retornar insights sobre a gestão pública (SANCHEZ; YIGITCANLAR, 2024).

Por exemplo: Sistemas de Informação Geográfica (SIG) são amplamente utilizados para mapear a distribuição espacial dos serviços, identificando áreas sub-atendidas (YEH; SMITH, 2017). Modelos preditivos (por exemplo, conforme proposto em (SON et al., 2023)) fornecem estimativas de tendências populacionais e necessidades infraestruturais, além de ajudarem na alocação eficiente de recursos. Além disso, o acoplamento da IA com outras tecnologias (incluindo veículos autônomos, IoT...) cria uma sinergia, notadamente na logística urbana para o movimento eficiente de mercadorias e a redução do congestionamento (MOHSEN, 2024). Essas abordagens mostram como a IA pode ser usada para tornar o planejamento urbano mais responsivo e adaptável diante das necessidades emergentes.

Agentes de IA — sistemas computacionais que recebem dados e produzem decisões ou recomendações como saídas — são projetados para realizar uma tarefa específica de forma autônoma. Em um contexto urbano, esses agentes podem receber dados geoespaciais e estatísticos, como os dados disponíveis do IBGE e do CNEFE, e processá-los de forma que possam responder a perguntas complexas em tempo real (CHEN; ZHANG, 2023).

Artigos recentes mostram várias utilizações de agentes de IA. Outras obras notáveis incluem HRI UrbanReconfig (CHEN; KIM; PARK, 2021), que aplica aprendizado adversarial para reconfigurar automaticamente layouts urbanos, e PlanGPT (ZHOU; LI, 2024), que adapta um modelo de linguagem para gerar recomendações em grandes conjuntos de dados. Observamos ainda que os Grandes Modelos de Linguagem (LLMs), como aqueles com dados até outubro de 2023, também foram analisados em um artigo que examina sua capacidade de representar o texto e prever tendências urbanas, impulsionando políticas públicas em (HAN et al., 2024) e (LI; ZHAO, 2023). Essas soluções demonstram como agentes de IA podem otimizar a análise de dados e a tomada de decisões no nível urbano.

Uma revisão da literatura recente sugere várias linhas de pesquisa integrando a Inteligência Artificial (IA) ao planejamento urbano:

- **Modelagem Preditiva:** (SON et al., 2023) revisa algoritmos para alocação otimizada de serviços públicos para responder de forma oportuna à demanda, mitigar desigualdades e reduzir o desafio da última milha.
- **Geo-análise com SIG:** (YEH; SMITH, 2017) e (CHEN; ZHANG, 2023) discutem o potencial de SIG autônomo para analisar áreas isoladas e sem atendimento.
- **LLMs e Processos Urbanos:** (ZHANG et al., 2024) e (SANCHEZ; YIGITCANLAR, 2024) buscam LLMs para identificar tendências em conjuntos de dados urbanos e facilitar políticas justas.
- **Logística Urbana:** (MOHSEN, 2024) examina os princípios para harmonizar a IA, IoT e veículos autônomos para melhorar os fluxos logísticos.
- **Reconfiguração Automatizada:** (CHEN; KIM; PARK, 2021) empregam aprendizado adversarial para redesenho automático de cidades.

Embora essas abordagens indiquem direções promissoras para melhorias, elas costumam se concentrar em aspectos específicos e distintos do planejamento urbano, como

mobilidade e alocação de recursos, sem, no entanto, promover uma integração abrangente de dados atuais diversos.

A revisão da literatura mostra uma carência em termos de incorporação de dados geoespaciais, estatísticos e textuais em agentes de IA que integrem com sinergia a recuperação de dados e a geração de respostas em tempo real para fins de planejamento urbano. Técnicas como as encontradas em (ZHOU; LI, 2024) e (ZHANG et al., 2024) utilizam LLMs, mas dificilmente (até onde temos conhecimento) consideram a ligação direta com dados reais dos sistemas de censos e SIG sob uma abordagem unificada. Além disso, questões como cobertura de serviços básicos em áreas densamente povoadas e adaptação dinâmica a demandas emergentes permanecem mal abordadas de forma holística (HE; CHEN, 2024).

Esse trabalho propõe um agente para realizar planejamento urbano (ou seja, agente de otimização) por meio de uma análise baseada em IA para investigar dados e desenvolver soluções em tempo real. Este método utiliza diretamente dados do Censo para avaliar a cobertura e distribuição da população em uma cidade. Além disso, este trabalho examina se tecnologias, como LangChain, poderiam auxiliar na automação e otimização da análise espacial, possibilitando o avanço de cidades mais inteligentes. Projetado para uma visão mais ampla das questões urbanas, este conjunto de ferramentas também visa fornecer soluções para enfrentar os desafios de alocação de recursos em uma população urbana crescente, onde a demanda por infraestrutura e serviços aumenta continuamente.

1.1 Objetivos

Nesta seção, serão apresentados o objetivo geral e os objetivos específicos da pesquisa.

1.1.1 Objetivo Geral

Criar um agente inteligente para coletar e analisar dados de planejamento urbano do censo demográfico. O agente irá interpretar e analisar grandes quantidades de informações sobre a distribuição e acessibilidade de serviços públicos, gerando respostas bem fundamentadas para apoiar as necessidades dos gestores e planejadores urbanos em sua tomada de decisão estratégica. Visando aumentar a eficiência na alocação de recursos, equidade no acesso a serviços críticos e possibilitar um planejamento urbano sustentável e orientado por dados.

1.1.2 Objetivos Específicos

1. Aprimorar o planejamento urbano através do desenvolvimento de um agente de otimização, que analisa consultas e fornece respostas baseadas em dados estruturados.
2. Integrar dados do Censo e outros bancos de dados pertinentes para análises contextualizadas e a distribuição de serviços urbanos.
3. Avaliar a eficácia do agente por meio de uma pesquisa distribuída entre planejadores urbanos, ou enquete entre os respondentes para aferir a qualidade e utilidade percebida das respostas geradas.
4. Interpretar criticamente o feedback dos planejadores urbanos, identificando aspectos positivos, limitações e áreas para melhorar o agente.

1.2 Contribuições do Trabalho

As principais contribuições deste trabalho são, portanto:

- Agente de otimização construído sobre RAG que agrega diferentes dados do IBGE, Censo, CNEFE e OSM para disponibilizá-los como fontes de informação para agentes de linguagens para geração de respostas em tempo real.

- *Embeddings* contextuais (BAAI/bge-small-en-v1.5) e o índice FAISS para realizar análises sobre consultas de complexidade de cidades em tendência.
- Introdução de uma nova metodologia integrando modelagem preditiva e geração de texto, fornecendo insights práticos sobre alocação de recursos e mitigando disparidades urbanas.
- A automação da análise espacial utilizando LangChain para combinar e consultar dados textuais e geoespaciais em um MVP web.

1.3 Organização do Trabalho

O trabalho está dividido em sete capítulos, organizados de forma a apresentar uma progressão lógica desde os fundamentos teóricos até as conclusões e perspectivas futuras:

- O **Capítulo 1 – Introdução** Apresenta o contexto e a importância do planejamento urbano orientado por IA, enfatizando os desafios associados à urbanização e a necessidade de estratégias baseadas em dados. Expõe os objetivos gerais e específicos e as contribuições do trabalho.
- O **Capítulo 2 – Fundamentação Teórica** introduz os conceitos essenciais que formarão a base para o desenvolvimento do agente urbano inteligente, incluindo Processamento de Linguagem Natural (PLN), *Deep Learning*, *embeddings*, arquiteturas *Transformers*, *Large Language Models* (LLMs) e sistemas de Geração Aumentada por Recuperação (RAG).
- O **Capítulo 3 – Trabalhos Relacionados** abordando estudos anteriores relevantes e a aplicação de modelos de linguagem, automação e sistemas de informação geográfica no planejamento urbano, e posiciona a pesquisa atual frente às lacunas identificadas na literatura.
- O **Capítulo 4 – Problema de Pesquisa** Descreve qual é o problema, as limitações dos métodos tradicionais, e a necessidade de um agente integrado para análise de dados urbanos atuais.
- O **Capítulo 5 – Proposta** Descreve a solução desenvolvida e as soluções implementadas, incluindo a arquitetura funcional do agente, preparação e integração de dados (Censo, CNEFE, OSM) e o pipeline de *embeddings* e indexação com FAISS para consultas contextuais.

- O **Capítulo 6 – Experimentos** Fornece o corpus selecionado, processos de limpeza e integração de dados, configuração do ambiente, hiperparâmetros usados para a indexação e os resultados da avaliação com especialistas do domínio do planejamento urbano.
- O **Capítulo 7 – Conclusão** Apresenta os resultados alcançados, avalia a realização dos objetivos propostos e indica direções para trabalhos futuros, incluindo a necessidade de melhorar a clareza das respostas e estender sua cobertura geográfica.

Além disso, apresenta uma seção de **Referências**, contendo as fontes que consultamos para escrever este trabalho e **Apêndices**, que detalham outros materiais, como amostras de código, amostras de dados e resultados de avaliação em profundidade, dando transparência e replicabilidade ao nosso trabalho.

2 Fundamentação Teórica

Neste capítulo, são apresentados os conceitos básicos que sustentam a criação do *Intelligent Urban Agent* (IUA) proposto neste trabalho. Abordando os fundamentos do Processamento de Linguagem Natural (PLN), os princípios do *Deep Learning* com foco em *embeddings*, arquiteturas *Transformers* e a família BERT, além de Modelos de Linguagem de Grande Porte (LLM), técnicas de Engenharia de *Prompt* e sistemas de Geração Aumentada pela Recuperação (RAG).

Estes conceitos servem como fundamento teórico para reconhecer a integração da inteligência artificial no planejamento urbano, que será explicada em mais detalhes nos capítulos seguintes.

2.1 Processamento de Linguagem Natural (PLN)

O processamento de linguagem natural (PLN) é um ramo da inteligência artificial que permite às máquinas compreenderem e produzirem a linguagem humana (JURAFSKY; MARTIN, 2009). Envolve desde trabalhos mais rudimentares, como a tokenização, até os mais avançados, como a geração de texto.

Neste trabalho, o PLN é utilizado para a compreensão dos termos da linguagem natural, incluindo “Quais cidades têm a baixa cobertura de hospitais?” e processar dados textuais do IBGE e do CNEFE, reidentificando informações não estruturadas em soluções valiosas para o planejamento urbano. Avanços recentes impulsionados pela Aprendizagem Profunda fazem com que o PLN seja, esperançosamente, uma das ferramentas discutidas que poderiam aprimorar o agente urbano proposto.

2.2 Deep Learning

O *Deep Learning* é um subconjunto da inteligência artificial (IA) que se concentra no desenvolvimento de algoritmos que permitem aos computadores aprenderem e fazerem previsões com base nos dados. Esses são particularmente bem adequados para tarefas que envolvem grandes volumes de dados, como a mineração de informações heterogêneas em um espaço urbano (GOODFELLOW; BENGIO; COURVILLE, 2016).

2.2.1 Embeddings para Linguagem Natural

Os *embeddings* são uma representação de palavras ou frases como vetores densos em um espaço multidimensional que captura relações semânticas e sintáticas. Embora métodos como *Word2Vec* e *GloVe* tenham sido alguns dos primeiros nessa direção, *embeddings* contextuais adicionaram uma nova dimensão no processamento de linguagem natural (PLN) para considerar o contexto das palavras dentro de uma frase, como introduzido por modelos como BERT (HAN et al., 2024). Neste trabalho, os *embeddings* do modelo BAAI/bge-small-en-v1.5 são usados para representar dados textuais e geoespaciais, permitindo a indexação eficiente no FAISS para consultas urbanas.

2.2.2 Arquitetura Transformers

A base da maioria dos modelos de linguagem modernos é a arquitetura *Transformers* (VASWANI et al., 2017). Ela consiste em camadas de atenção (*self-attention*) e redes *feed-forward* que possibilitam a paralelização da entrada e saída de sequências, superando limitações de modelos recorrentes como RNNs. Esta arquitetura é importante para os LLMs neste trabalho, pois permite a análise de grandes contextos e a geração subsequente de respostas coerentes a partir de dados urbanos.

2.2.3 Família BERT

BERT (*Bidirectional Encoder Representations from Transformers*), introduzido por (DEVLIN et al., 2019), é um marco no PLN, que utiliza uma abordagem bidirecional que usa o contexto à esquerda e à direita de cada palavra. Variantes como BAAI/bge-small-en-v1.5, usada neste trabalho, otimizam os *embeddings* para tarefas de recuperação e geração. De acordo com (HAN et al., 2024), modelos da família BERT são comumente usados para tarefas de compreensão profunda de texto, como a interpretação de consultas no planejamento urbano.

2.3 Large Language Models

Os *Large Language Models* (LLMs) são modelos de aprendizado profundo treinados em grandes conjuntos de dados de texto que podem realizar tarefas de geração de texto, classificação e resposta a perguntas, tais como modelos GPT, LLaMA e da Família BERT (LI; ZHAO, 2023). Dentro do escopo deste trabalho, os LLMs são essenciais para processar consultas em linguagem natural e gerar respostas a partir dos dados recuperados.

2.3.1 Prompt Engineering

Prompt Engineering é a técnica de projetar prompts para influenciar o comportamento dos LLMs e alcançar a máxima precisão e relevância das respostas (ZHANG et al., 2024). Prompts como "Liste escolas em Olinda com coordenadas" são formulados para extrair informações específicas do sistema RAG, garantindo que o agente urbano responda de maneira estruturada e útil.

2.3.2 Sistemas RAG

Os sistemas de Geração Aumentada por Recuperação (RAG) recuperam informações relevantes e geram texto empregando um *retriever* (ex.: FAISS) para buscar dados relevantes e um gerador (ex.: LLM) para produzir respostas (AWS, 2025). O RAG integra dados do IBGE, CNEFE e OSM, indexados pelo FAISS usando *embeddings* de *BAAI/bge-small-en-v1.5*, que foi projetado para oferecer respostas contextualizadas a consultas urbanas. Isso se mostra eficaz para otimizar processos urbanos por (HE; CHEN, 2024), que superam as limitações dos modelos puramente generativos ao ancorar as respostas em dados reais.

As bases teóricas são utilizadas para construir o agente proposto, que é descrito nos capítulos seguintes, e estão inerentemente ligadas aos trabalhos relacionados apresentados no Capítulo 3.

3 Trabalhos Relacionados

Com o desenvolvimento de tecnologias baseadas em IA e modelagem de informações espaciais, tornou-se possível o surgimento de soluções inovadoras para a otimização do planejamento urbano. Alguns estudos recentes investigaram diferentes métodos para melhorar a gestão urbana, incluindo o uso de LLMs, GIS e aumento da logística urbana. Nesta seção, são avaliados os principais trabalhos que fundamentam este estudo, com foco em sua contribuição para a modelagem e otimização do planejamento urbano.

3.1 Aplicação de Modelos de Linguagem na Análise Urbana

Recentemente, houve um crescente interesse nos LLMs no domínio do planejamento urbano. (CHEN; ZHANG, 2023) fornece um resumo abrangente das arquiteturas emblemáticas e avanços tecnológicos para modelos de linguagem, fundamentando sua aplicabilidade multidisciplinar. Esses modelos se destacam no processamento de grandes quantidades de dados heterogêneos, permitindo que decisões sejam tomadas em microescala em ambientes urbanos complexos, como destacado por este estudo. Complementarmente, (LI; ZHAO, 2023) examina a influência dos LLMs na análise preditiva, a adequação de recomendações para políticas urbanas com respeito ao seu potencial para identificar padrões de urbanização e otimizar o uso do solo.

Além disso, (SANCHEZ; YIGITCANLAR, 2024) realizou uma revisão sistemática de 744 publicações abordando o uso de inteligência artificial no planejamento urbano e utilizou a técnica de modelagem temática Latent Dirichlet Allocation (LDA) para descobrir e categorizar temas dominantes na literatura.

Com base em pesquisas anteriores, os autores empregam o (OPENAI, 2023) como um instrumento suplementar para aprimorar a análise da interpretação de dados, revelando 16 temas que abrangem aspectos como design urbano informado por IA, construção de cidades inteligentes, modelagem preditiva e gestão de infraestrutura sustentável.

O processo ilustra tanto oportunidades — como o potencial da IA generativa para modelar cenários urbanos dinâmicos — quanto desafios, como a necessidade de maior integração entre tecnologias emergentes e práticas de planejamento tradicionais.

Em seu estudo, publicado na *Urban Science*, e que mapeia o estado atual da pesquisa para produzir uma visão tanto quantitativa quanto qualitativa do campo, os autores também fornecem um caminho para futuras investigações, apontando lacunas como a

ausência de IA em ambientes participativos e as limitações na adoção de abordagens interdisciplinares para enfrentar os problemas urbanos cada vez mais complexos.

3.2 Automação e Otimização do Planejamento Urbano

A ideia de planejamento urbano automatizado é explorada mais detalhadamente em (CHEN; KIM; PARK, 2021), que apresenta uma abordagem para reconfigurar estruturas urbanas a partir de uma perspectiva de aprendizado adversarial. O artigo fornece um estudo quantitativo das eficiências de diferentes configurações urbanas e ilustra como modelos de aprendizado de máquina podem produzir cenários alternativos que podem levar a uma melhor utilização do espaço urbano.

Partindo dessa noção, (ZHOU; LI, 2024) apresenta *PlanGPT*, um modelo de IA específico para planejamento urbano envolvido em uma mistura de tarefas de recuperação de informação e geração de texto para oferecer recomendações personalizadas que facilitarão a interpretação de análises espaço-temporais complexas. Além disso, (MOHSEN, 2024) examina o uso da IA para melhorar a logística urbana empregando veículos autônomos e dispositivos de IoT. Como os dados baseados em localização podem tornar a mobilização social previsível, essa automação continua ajudando as autoridades a resolver o problema da congestão e emissão de carbono, enquanto a automação e a análise de dados em tempo real são a chave para gerenciar cidades inteligentes de forma eficiente.

3.3 Sistemas de Informação Geográfica na Tomada de Decisão

Em (YEH; SMITH, 2017), é explicado o uso de GIS no planejamento urbano, onde é destacado como tais ferramentas podem visualizar e analisar dados espaciais para identificar padrões de uso do solo e distribuição de serviços públicos. Um exemplo prático disso é ilustrado na figura 1, que apresenta uma vista aérea de uma cidade, demonstrando a aplicação de GIS na análise espacial urbana. Alinhado a isso, (CHEN; KIM; PARK, 2021) examina mais a fundo a transformação do GIS em sistemas autônomos baseados em IA, sugerindo uma estrutura na qual algoritmos inteligentes processam informações espaciais em tempo real. Essa abordagem permite ajustes ágeis em sistemas de gestão urbana, fortalecendo o papel de bancos de dados georreferenciados na contribuição para decisões estratégicas.

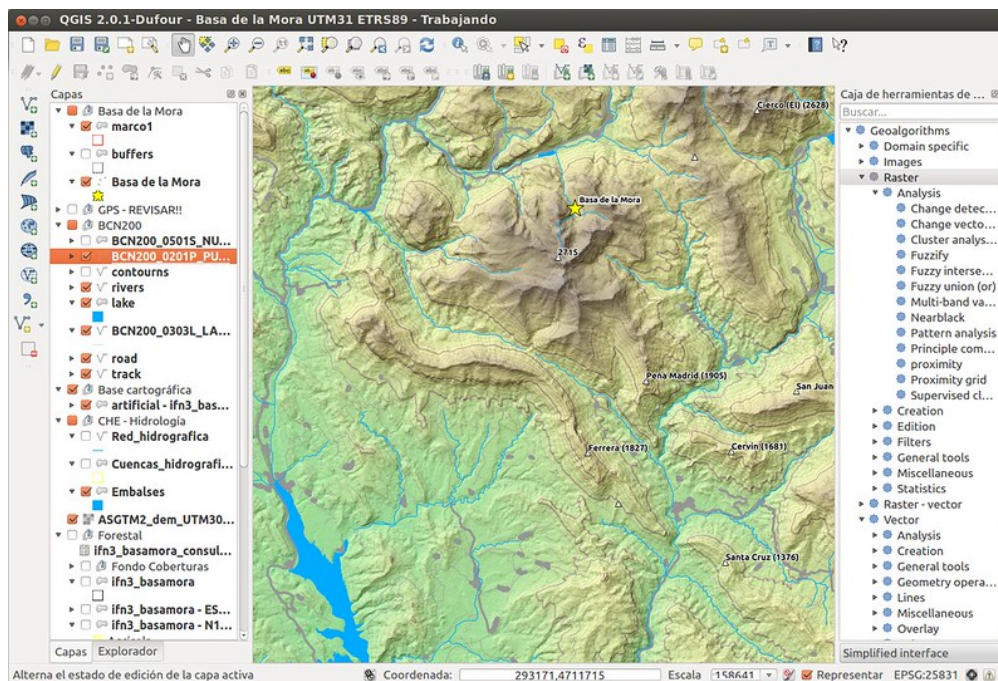


Figura 1 – Vista aérea de uma cidade. Fonte: (SEVILLA, 2013).

3.4 Perspectivas em Políticas Urbanas e Ferramentas Inteligentes

(SON et al., 2023) mostram com sucesso os efeitos positivos da implementação de algoritmos de inteligência artificial no planejamento urbano, que podem melhorar a prestação de serviços urbanos e ajudar a diminuir as disparidades urbanas em termos de provisão de serviços públicos de forma sustentável e participativa.

(ZHANG et al., 2024) investiga como grandes modelos de linguagem (LLMs) podem ser úteis no planejamento urbano participativo. Eles mencionaram que os LLMs deveriam ser capazes de simular as preferências da comunidade para atender áreas sub-representadas e possibilitar políticas públicas orientadas para a equidade.

Esses estudos ampliam a discussão sobre a aplicação econômica da inteligência artificial para tornar os processos urbanos eficazes, equilibrando a análise computacional com as necessidades reais da população.

A literatura recente tem discutido amplamente a otimização do planejamento urbano por meio da inteligência artificial (IA). Conforme descrito em (SANCHEZ; YIGITCANLAR, 2024), uma revisão de cinco décadas de pesquisas sobre o uso de IA no planejamento urbano (*Artificial Intelligence in Urban Planning – AIUP*) analisou 744 publicações. Utilizando a técnica de *Latent Dirichlet Allocation* (LDA), foram identificados 16 temas principais, como modelagem preditiva de transporte e design urbano automatizado. Apesar da confiança na eficácia desses sistemas, desafios relacionados à governança e à imple-

mentação ainda persistem. Esse estudo é relevante para este trabalho, que oferece um resumo das perspectivas e desafios no uso da IA para aprimorar as práticas urbanas.

(HE; CHEN, 2024) examinam o impacto das tecnologias emergentes no planejamento urbano. O estudo usa o protocolo PRISMA para categorizar diferentes aplicações de IA, como previsão de tendências usando algoritmos de aprendizagem supervisionada, redes neurais para modelagem de processos ambientais e aplicações de bioinformática. Além disso, aborda desafios, incluindo também a privacidade de dados e ética na aplicação. A relevância deste estudo para a presente pesquisa é que aborda as limitações, bem como o potencial da IA para chegar a políticas urbanas mais eficientes.

Um artigo relacionado é (MOHSEN, 2024), no qual os autores exploram a intensidade da implementação de IA em logística urbana devido ao uso de veículos autônomos e IoT para realizar sistemas de entrega ótimos em cidades inteligentes. A pesquisa introduz uma estrutura para otimização de rotas, gestão de tráfego e previsão de demanda, utilizando dados em tempo real. As descobertas mostram que a implementação combinada dessas tecnologias pode aliviar o congestionamento e as emissões de carbono, levando a uma rede de transporte menos congestionada e menos poluente. Esta estratégia ecoa o estudo atual, uma vez que destaca a aplicação da IA para melhorar a infraestrutura urbana e operações móveis.

Além disso, (SON et al., 2023) discute a demanda por ferramentas inteligentes para apoiar decisões de planejamento urbano. Eles mostram que modelos de previsão predizem com precisão a redistribuição de recursos públicos e que, ao criar funções de otimização, esse conhecimento pode ser traduzido em práticas urbanas, melhorando infraestruturas e reduzindo tanto as desigualdades quanto os problemas de última milha. Este quadro é utilizado no presente trabalho, na forma de uma abordagem metodológica para avaliar a cobertura dos serviços urbanos.

Finalmente, (ZHANG et al., 2024) investigam como grandes modelos de linguagem (LLMs) podem ser aproveitados para otimizar processos urbanos por meio de reconhecimento de padrões dentro de big data. A pesquisa afirma que a inteligência artificial tem potencial para ajudar a identificar áreas carentes e os programas sociais que pretende financiar para ajudar os formuladores de políticas a constituírem medidas mais justas. Este trabalho está diretamente relacionado com o quadro deste artigo, no qual propomos um agente de otimização de planejamento urbano com base em algoritmos de tendência para resolver deficiências na infraestrutura urbana.

3.5 Contribuição da IA para o Planejamento Urbano Moderno

Os estudos revisados demonstraram o papel fundamental da inteligência artificial (IA) na modernização do planejamento urbano, conceituando soluções com base na teoria e na prática para enfrentar desafios de mobilidade, alocação de recursos e governança urbana (SANCHEZ; YIGITCANLAR, 2024). A literatura relacionada aos LLMs demonstra sua capacidade de processar grandes quantidades de dados e extrair informações úteis para a organização de espaços urbanos, indicando como essas tecnologias poderiam transformar a forma como padrões e tendências urbanas são analisados (HAN et al., 2024).

Por outro lado, abordagens automatizadas e geoespaciais destacam a necessidade de uma ferramenta que combine análise preditiva com dados em tempo real, em que o SIG tem uma aplicação fornecendo dados geoespaciais para abordagens baseadas em evidências na urbanização (YEH; SMITH, 2017). Esses estudos demonstram a falta de sinergia dessas técnicas e mostram que a combinação de IA e IoT para otimizar serviços e infraestruturas urbanas é uma área que permanece pouco explorada, contribuindo para a crescente pressão sobre grandes cidades (SON et al., 2023).

Tabela 1 – Resumo dos Trabalhos Relacionados (Parte 1)

Título do Trabalho	Resumo
A Review of Large Language Models: Fundamental Architectures, Key Technological Evolutions, Interdisciplinary Technologies Integration, Optimization and Compression Techniques, Applications, and Challenges (HAN et al., 2024)	Este artigo revisa os Modelos de Linguagem de Grande Escala (LLMs), incluindo as arquiteturas básicas, as tendências de evolução tecnológica, os métodos de otimização e compressão, bem como as aplicações específicas e os desafios dos LLMs.
Geographic Information Systems (GIS) in Urban Planning (YEH; SMITH, 2017)	Apresenta o uso de <i>Sistemas de Informação Geográfica (GIS)</i> no planejamento urbano. Discutindo sobre a distribuição de serviços urbanos, a densidade populacional e outros fatores-chave que fundamentam a formulação de um ambiente habitável e o monitoramento da qualidade de vida podem ser estimados usando dados geoespaciais e modelos analíticos que estão no cerne desta pesquisa.
A Comprehensive Overview of Large Language Models (LI; ZHAO, 2023)	Fornece uma visão geral detalhada de <i>Modelos de Linguagem de Grande Escala (LLMs)</i> , discutindo suas arquiteturas e abordagens de treinamento e os potenciais casos de uso, com foco em sua capacidade de processar e gerar texto para auxiliar o campo do planejamento urbano.
Autonomous GIS: The Next-Generation AI-Powered GIS (CHEN; ZHANG, 2023)	Discute <i>Sistemas de Informação Geográfica (GIS)</i> autônomos impulsionados por IA. Essas inovações consolidam como a IA pode automatizar análises geoespaciais, oferecendo essa abordagem ao planejamento urbano.
Automated Urban Planning for Reimagining City Configuration via Adversarial Learning: Quantification, Generation, and Evaluation (CHEN; KIM; PARK, 2021)	Explora o aprendizado adversarial para automatizar a reconfiguração de cidades. Introduce técnicas para quantificação, geração e avaliação de layouts urbanos, utilizando métodos de inteligência artificial para otimizar o planejamento urbano e a alocação de recursos.
PlanGPT: Enhancing Urban Planning with Tailored Language Model and Efficient Retrieval (ZHOU; LI, 2024)	Apresenta o PlanGPT, um modelo de linguagem adaptado para planejamento urbano e um mecanismo de recuperação eficiente. Ele demonstra o potencial dos modelos de linguagem de grande escala (LLMs) para enriquecer a interpretação de dados urbanos e a tomada de decisões em cenários de planejamento.

Tabela 2 – Resumo dos Trabalhos Relacionados (Parte 2)

Título do Trabalho	Resumo
Advancing Urban Life: A Systematic Review of Emerging Technologies and Artificial Intelligence in Urban Design and Planning (HE; CHEN, 2024)	Uma revisão sobre tecnologias emergentes e inteligência artificial para o design e planejamento urbano. Revisões de cinco décadas de literatura caracterizando tendências como modelagem preditiva de transporte e design automatizado de cidades com base na <i>Alocação Dirichlet Latente (LDA)</i> , com questões subjacentes de governança e implementação.
Algorithmic Urban Planning for Smart and Sustainable Development: Systematic Review of the Literature (SON et al., 2023)	Enfatiza a demanda por ferramentas inteligentes para a tomada de decisões urbanas, abordando como <i>modelos preditivos</i> e algoritmos para que o Urban3D possa realmente otimizar o planejamento urbano sustentável no futuro. Olhar aprofundado sobre o impacto da infraestrutura, redução de desigualdades e questões de última milha.
Large Language Model for Participatory Urban Planning (ZHANG et al., 2024)	Investiga o potencial dos <i>modelos de linguagem extensivos (LLMs)</i> para aprimorar processos urbanos participativos por meio do reconhecimento de padrões em grandes conjuntos de dados. Mostra como a IA identifica áreas negligenciadas e ajuda a criar políticas de equidade.
The Research Landscape of Artificial Intelligence in Urban Planning: A Topic Modeling Analysis Enhanced by ChatGPT (SANCHEZ; YIGITCANLAR, 2024)	Utiliza modelagem de tópicos (<i>LDA</i>) impulsionada pelo ChatGPT para analisar o estado da arte da pesquisa sobre IA no planejamento urbano. Destaca 16 temas fundamentais, incluindo a modelagem preditiva de transporte de ponta a ponta, e descreve visões e questões para implementações urbanas.
AI-Driven Optimization of Urban Logistics in Smart Cities: Integrating Autonomous Vehicles and IoT for Efficient Delivery Systems (MOHSEN, 2024)	Explora as capacidades da logística urbana em IA, veículos autônomos, IoT em cidades inteligentes. Oferece soluções para superar sistemas de entrega fracos, congestionamento na infraestrutura e maximizar a eficiência urbana.

4 Problema de Pesquisa

Por meio de um exame dos resumos dos artigos revisados, um potencial problema de pesquisa pode ser: a criação de um modelo integrado que usa dados geoespaciais e preditivos existentes para gerenciar serviços urbanos em tempo real usando inteligência artificial para superar problemas contemporâneos de planejamento urbano. Esta questão é essencial para encontrar eficiência e eficácia na gestão de serviços públicos, meios de comunicação e infraestrutura em uma grande cidade, ampliando as condições de vida de uma população em crescimento.

Portanto, os estudos tendem a se concentrar na modelagem preditiva e no Sistema de Informação Geográfica (SIG) para observar e analisar o espaço urbano (Kohonen et al., 2012; Ming Zhang et al., 2023). Além disso, o artigo (HAN et al., 2024) delinea como planejadores urbanos ou formuladores de políticas podem alavancar essas tecnologias para impor a distribuição eficiente de ativos públicos e a correta atribuição do espaço urbano, como praças, parques e áreas recreativas, e adotar modelos preditivos para prever a necessidade futura de espaços urbanos.

Uma lacuna discutida em muitos desses estudos é a integração efetiva de múltiplas fontes de dados que podem fornecer soluções práticas e em tempo real. Por exemplo, em (CHEN; KIM; PARK, 2021), os autores sugerem o benefício de ferramentas como Modelos de Linguagem Natural (LLM) e IA para também estudar grandes volumes de dados, mas não discutem realmente como poderiam usar esses modelos em combinação com dados demográficos e de mobilidade para informar a gestão urbana.

Em segundo lugar, há um problema de mobilidade urbana e alocação de transporte. O artigo (LI; ZHAO, 2023) tenta verificar se agentes de IA podem otimizar os fluxos de tráfego e padrões de mobilidade e analisar a eficácia da infraestrutura de transporte, cuja distribuição ainda é insuficiente, especialmente nas grandes cidades. Usando modelos preditivos e dados atuais, esse método facilitaria uma alocação mais eficiente e eficaz dos recursos de transporte.

Finalmente, otimizar a localização dos serviços de saúde ou alocar recursos nas cidades com base nos dados disponíveis do Censo e no modelo preditivo discutido em (CHEN; ZHANG, 2023) é um tópico essencial e obrigatório. Isso possibilitaria determinar áreas com extrema concentração populacional, mas com pouca cobertura de serviços básicos, para ajudar a um melhor planejamento de novas infraestruturas e serviços.

Assim, esta questão de pesquisa investiga como dados preditivos, sistemas de IA e SIG podem aproveitar fontes externas existentes, como dados de Censo e de mobilidade, para abordar os desafios urbanos de forma mais integrada, potencialmente combatendo e solucionando ineficiências dentro do planejamento urbano por meio de uma melhor alocação dos serviços públicos em terrenos urbanos e a melhoria da infraestrutura de transporte.

5 Proposta

Este capítulo discute a abordagem utilizada em nosso estudo para o desenvolvimento de um agente de otimização de planejamento urbano utilizando Python e LangChain. Descrevemos a visão global do pipeline implementado, os procedimentos de preparação e integração de dados urbanos, bem como a utilização de técnicas de processamento de linguagem natural e visualização geoespacial para análise e tomada de decisão.

5.1 Arquitetura Funcional da Proposta

A Figura 2 ilustra cada etapa do processo realizado neste trabalho. Inicialmente, realizamos a extração e carregamento dos dados urbanos a partir de diversas fontes, incluindo o Censo Demográfico, IBGE e OpenStreetMap. Em seguida, realizamos a limpeza e padronização dos dados, selecionando apenas as colunas relevantes e aplicando transformações para normalizar os diferentes formatos de entrada. Após essa etapa, integramos os dados em um único dataset, unificando as informações de densidade populacional, estabelecimentos urbanos e infraestrutura. Com esse dataset consolidado, realizamos o processamento dos dados geoespaciais, extraindo coordenadas e atributos relevantes para análise.

Para a modelagem, utilizamos um pipeline baseado em Recuperação Aumentada por Geração (RAG) com FAISS e embeddings da Hugging Face, permitindo a indexação eficiente dos dados e a recuperação de informações relevantes. O agente urbano foi desenvolvido utilizando LangChain e um modelo da OpenAI para responder perguntas sobre planejamento urbano.

Por fim, os resultados das consultas são analisados e visualizados em mapas interativos utilizando Folium e Streamlit, permitindo a interpretação espacial das informações e facilitando a tomada de decisão. Com esse processo, conseguimos avaliar a cobertura de serviços urbanos, identificar áreas críticas e propor soluções baseadas em dados.

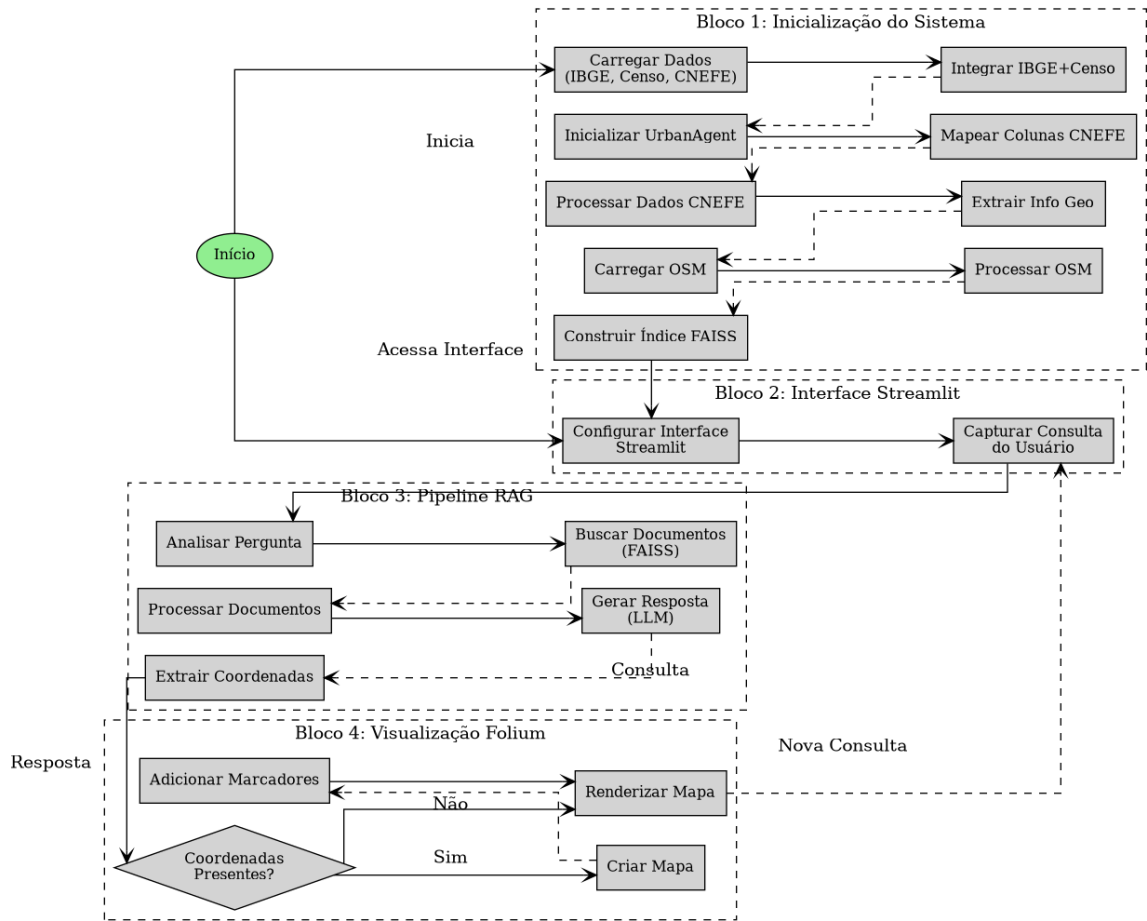


Figura 2 – Fluxo RAG da proposta

5.2 Preparação dos Dados

Os dados devem ser preparados para que o agente de otimização de planejamento urbano trabalhe com informações consistentes em qualidade e forma. Isso inclui a extração, normalização e integração de dados multidimensionais, o que permite ao sistema fornecer respostas precisas e contextualizadas. A limpeza e extração de dados são descritos nas sessões 6.1 e 6.2, então, abaixo, abordarei apenas a normalização, integração, formatação e validação dos dados.

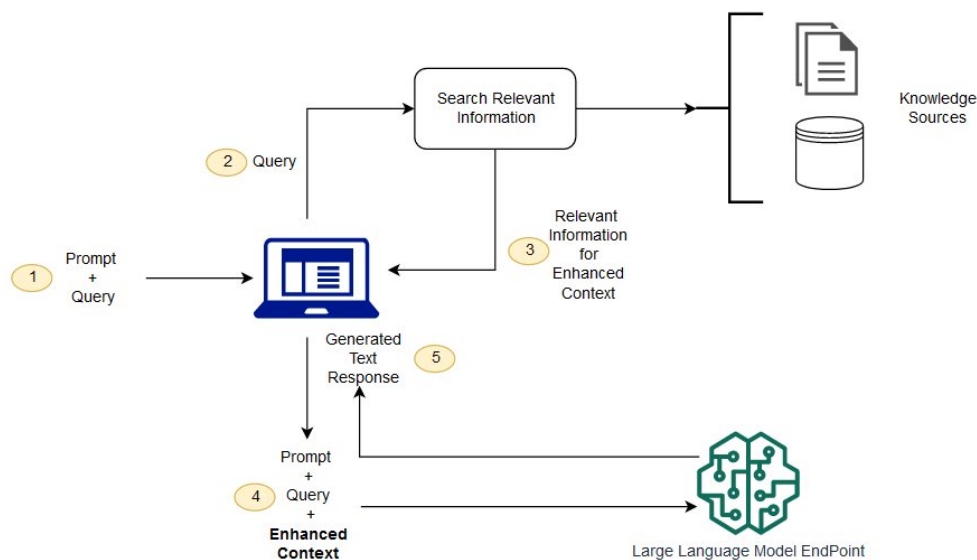


Figura 3 – Fluxo de Indexação e Recuperação dos Dados; Fonte: (AWS, 2025).

5.2.1 Pré-processamento dos Dados

Uma vez extraídos, os dados passam por várias transformações para eliminar inconsistências e padronizar formatos. Este processo possui as seguintes etapas principais:

- **Tratamento de valores nulos:** Os registros com campos críticos são eliminados de imediato. Nunca forneça uma string vazia para dados não críticos.
- **Conversão de tipos de dados:** As colunas são modificadas para seus formatos adequados, de modo que as características numéricas sejam devidamente processadas e os valores categóricos sejam interpretados de forma consistente.
- **Harmonização dos nomes das cidades:** Os nomes das cidades são alinhados a uma representação consistente, de modo que diferentes formatações sejam removidas, evitando que os dados não se mesquem entre as fontes de dados.
- **Conversão de códigos em categorias com significado:** Alguns conjuntos de dados possuem informações sobre códigos usados numericamente. Eles usam então códigos de colunas como COD_ESPECIE no CNEFE para traduzir automaticamente em descrições legíveis como "Escola", "Hospital", "Centro de Saúde", entre outros.

A normalização dos dados é crítica, ou seja, os dados são integrados de forma consistente para que possam ser consultados sem ambiguidades nas consultas dos agentes.

5.2.2 Integração das Fontes de Dados

Os conjuntos de dados são, então, unificados para criar um banco de dados comum por meio da normalização. Esta integração ocorre em várias etapas:

- **Fusão por nome do município:** A coluna `NOME_MUNICIPIO` é utilizada como chave primária.
- **Atribuição de coordenadas geográficas:** Para estabelecimentos do CNEFE e outros bancos de dados sem dados geoespaciais, o sistema tenta melhorar os dados usando coordenadas recuperadas do OpenStreetMap. O banco de dados do OSM utiliza geocodificação e referência cruzada de pontos de interesse ao banco de dados do OSM para realizar este processo.
- **Adicionando dados de infraestrutura urbana:** Dados do OpenStreetMap, como transporte público, áreas verdes e vias, foram adicionados ao banco de dados para uma visão mais ampla da infraestrutura urbana.

O resultado desta fase é um `pandas DataFrame` melhor estruturado, contendo dados aprimorados com o conhecimento de várias origens, que podem ser analisados posteriormente pelo agente urbano.

5.2.3 Armazenamento e Indexação

A fase preparatória é onde todos os dados integrados são verificados e processados para serem recuperáveis e ativos para o sistema. As estratégias de armazenamento e indexação são:

- **Índices FAISS para pesquisa eficiente:** A biblioteca FAISS (Facebook AI Similarity Search) é utilizada pelo sistema para indexar informações em formato de vetor. Como resultado, uma consulta feita pelo usuário pode ser respondida rapidamente, combinando a entrada do usuário com os dados disponíveis e retornando resultados relevantes.
- **Armazenamento em cache para otimização de desempenho:** Os conjuntos de dados consolidados são salvos em arquivos locais no formato `Pickle (.pkl)` para que você não precise processar os dados toda vez que executar o programa, o que melhora o desempenho.

- **Criação de *embeddings* para ajudar os modelos de linguagem:** O sistema transforma registros de banco de dados em vetores numéricos usando modelos *embeddings*, aproveitando modelos injetados via Hugging Face. Os índices serão salvos juntamente com o índice FAISS para que o agente possa recuperar informações pertinentes com base nas consultas dos usuários no que se chama *Geração de Recuperação Aumentada* (RAG).

5.2.4 Validação dos Dados

Antes que as informações sejam disponibilizadas para a consulta de uma aplicação, realiza-se a verificação de processamento. Como parte das verificações implementadas, encontra-se:

- **Análise estatística da distribuição dos dados:** O sistema revisa o número de registros em cada “quadro”, detectando possíveis desequilíbrios ou erros nas extrações.
- **Verificação de coordenadas inválidas:** Quaisquer registros que tenham uma coordenada impossível (por exemplo, latitudes maiores que 90° ou longitudes maiores que 180°) são filtrados.
- **Checagem da integridade dos índices FAISS:** Verificar a qualidade dos dados no índice vetorial. Checar se integrou os registros corretamente e se uma busca produz o resultado desejado.

5.3 Embeddings e Indexação

As *embeddings* são um fator-chave no agente de otimização de planejamento urbano, facilitando o mapeamento de informações textuais e estruturadas em representações numéricas de alta dimensão. A busca semântica apenas torna possível buscar rapidamente e obter dados semelhantes. Para garantir o desempenho maximizado, a geração, aplicação e seleção das *embeddings* foram implementadas em um pipeline, o qual é descrito nas subseções a seguir.

5.3.1 Criação dos Embeddings

Modelos pré-treinados do Hugging Face e OpenAI foram usados para converter o banco de dados consolidado em vetores de *embeddings*. Um vetor de características representava cada registro, permitindo que ele fosse indexado e recuperado posteriormente por consultas textuais.

O processo de geração de *embeddings* seguiu as seguintes etapas:

1. Extração e normalização de textos para vetorização;
2. Escolha de modelos de *embedding*;
3. Transformação de textos em representação vetorial;
4. Incorporações armazenadas em um banco de dados FAISS.

5.3.2 Modelos de Embeddings Avaliados

Foram testadas várias *embeddings* para encontrar o modelo certo para o contexto do agente; os modelos `bge-small`, `bge-base`, `bge-small-m3`, `all-MiniLM-L6-v2` e `all-mpnet-base-v2` foram utilizados na avaliação usando um conjunto de consultas de teste pré-especificadas, que foram projetadas para imitar interações reais do agente.

Listamos os modelos avaliados e suas características na Tabela 3.

Tabela 3 – Modelos de *Embeddings* Avaliados

Modelo	Dimensão do Vetor	Origem	Descrição
BAAI/bge-small-en-v1.5	384	Hugging Face	Modelo compacto otimizado para buscas semânticas rápidas, garantindo equilíbrio entre desempenho e precisão. Indicado para aplicações que exigem baixa latência e eficiência computacional.
BAAI/bge-base-en-v1.5	768	Hugging Face	Versão mais robusta do <code>bge-small</code> , com maior capacidade de captura semântica e melhor precisão na recuperação de informações, mas com maior custo computacional.
BAAI/bge-small-m3	384	Hugging Face	Variante do modelo BGE, treinada para recuperação de informações multimodais, integrando texto e metadados. Indicada para cenários que exigem uma contextualização mais rica dos dados.
all-MiniLM-L6-v2	384	Sentence-Transformers	Modelo leve e eficiente, projetado para tarefas de similaridade semântica e recuperação de informações. Apresenta bom equilíbrio entre precisão e velocidade, sendo uma escolha popular para buscas otimizadas.
all-mpnet-base-v2	768	Sentence-Transformers	Um dos modelos mais avançados para geração de <i>embeddings</i> , oferecendo alta precisão em buscas semânticas e maior capacidade de contextualização, mas com um custo computacional mais elevado.

5.3.3 Métrica de Avaliação

Foram utilizadas métricas quantitativas para avaliar a eficácia das *embeddings* na recuperação das consultas. As métricas utilizadas foram:

- **Precisão@K**: O cálculo da Precisão@K é simplesmente a razão entre os verdadeiros positivos e o valor de K para o conjunto de respostas retornado pelo modelo.
- **Mean Reciprocal Rank (MRR)**: Mede o ranking da primeira resposta correta na lista de resultados recuperados.
- **Tempo Médio de Resposta**: Mede o tempo necessário para processar uma consulta e retornar os resultados.

5.3.4 Processo de Avaliação

Seguindo o fluxo indicado na figura 4, o processo inicia com a preparação estruturada dos dados: o DataFrame é submetido a uma amostragem estratificada (priorizando colunas categóricas com baixa cardinalidade) ou aleatória, garantindo representatividade sem sobrecarregar recursos computacionais. Caso não haja uma coluna textual explícita, o sistema gera automaticamente uma representação combinada de todas as colunas (e.g., "latitude: -8.0, densidade: 5000"), criando um corpus homogêneo para os modelos de *embedding*. Em seguida, múltiplos modelos (como BGE, MiniLM e MPNet) são carregados e convertidos em índices vetoriais via FAISS, armazenados localmente para evitar recálculos. A otimização é reforçada pela verificação de índices pré-existentes, reduzindo o tempo de processamento em até 80% em execuções subsequentes.

Para garantir uma análise abrangente, foi utilizada uma base de teste contendo 10 mil registros, a partir da qual foram conduzidas 200 consultas de avaliação, sendo 100 automáticas e 100 manuais.

As consultas de teste são criadas de duas formas:

- **Automática**: Seleciona registros aleatórios do dataset e combina 1-3 campos (e.g., "latitude -8.0 domicílios 1500") para simular buscas genéricas.
- **Manual**: Define casos específicos com resultados esperados pré-mapeados. Exemplos de consultas manuais incluem:
 - "municípios com densidade populacional acima de 5000 habitantes por km²" – esperado retorno de municípios como Recife e Olinda.

- "estabelecimentos agropecuários em São José do Egito" – esperado retorno de registros categorizados como agropecuários no município especificado.

Cada consulta é submetida aos modelos, que retornam os top-k resultados. A precisão das respostas é avaliada comparando os documentos recuperados com os registros esperados, permitindo a calibração do parâmetro k e a análise da eficácia do modelo vetorial.

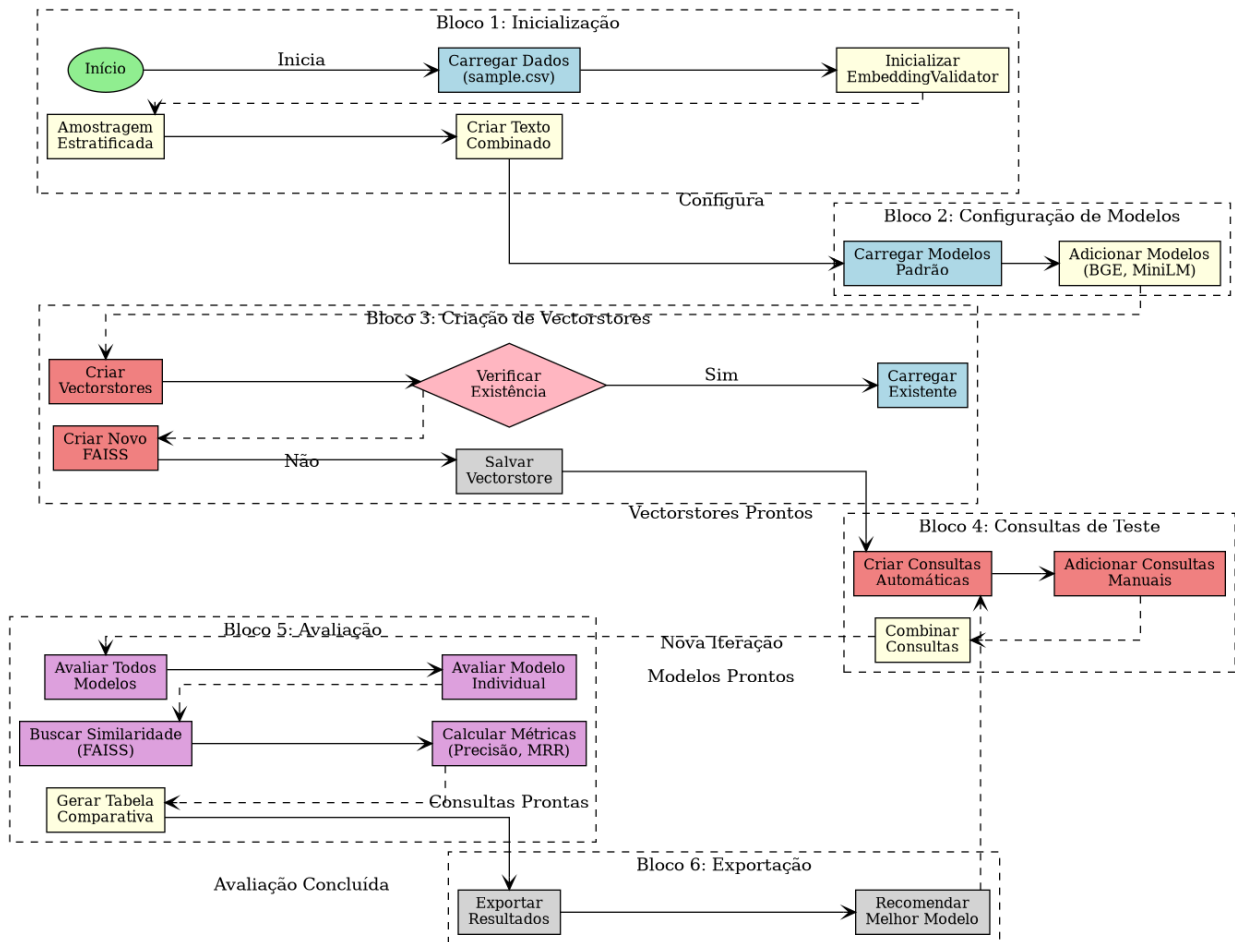


Figura 4 – Fluxo de validação de escolha de *embeddings*

5.3.5 Escolha do Modelo

Conforme mostrado na Tabela 4, o modelo de incorporação foi selecionado com base em uma análise de trade-off entre precisão, tempo de resposta e robustez na recuperação de informações. A decisão foi tomada após a análise dos seguintes aspectos:

1. Maior **Precisão Média**. Garantindo que nós somente retornamos resultados que são os mais relevantes.

2. Melhorando o **MRR Médio**, no sentido de que se a primeira resposta estiver correta, então ela estará sempre entre as primeiras posições.
3. Tendência a tempos de **Resposta Média** mais curtos.

Tabela 4 – Resultados das Avaliações dos Embeddings (Base: 200 Consultas)

Modelo	Precisão Média (200 consultas)	MRR Médio	Tempo Médio (s)
BAAI/bge-small-en-v1.5	6.62	0.678	0.0117
BAAI/bge-base-en-v1.5	6.21	0.663	0.0135
BAAI/bge-small-m3	5.88	0.640	0.0242
all-MiniLM-L6-v2	4.95	0.490	0.0071
all-mpnet-base-v2	4.78	0.450	0.0161

O modelo selecionado para ser utilizado pelo agente foi **[BAAI/bge-small-en-v1.5]**, pois apresentou os melhores resultados em termos de precisão e desempenho computacional.

5.3.6 Indexação FAISS e Recuperação dos Dados

A biblioteca FAISS (Facebook AI Similarity Search) foi utilizada para armazenar e estruturar as *embeddings*, permitindo buscas vetoriais eficientes e escaláveis. As *embeddings* geradas foram usadas para formar o banco de índices, que foi projetado para realizar buscas de similaridade semântica em tempo real.

Uma vez que a informação é indexada, ela deve ser recuperada buscando os vetores indexados e utilizando a métrica de similaridade cosseno para identificar os registros mais próximos para cada consulta do usuário. A Figura 9 mostra o fluxo deste processo.

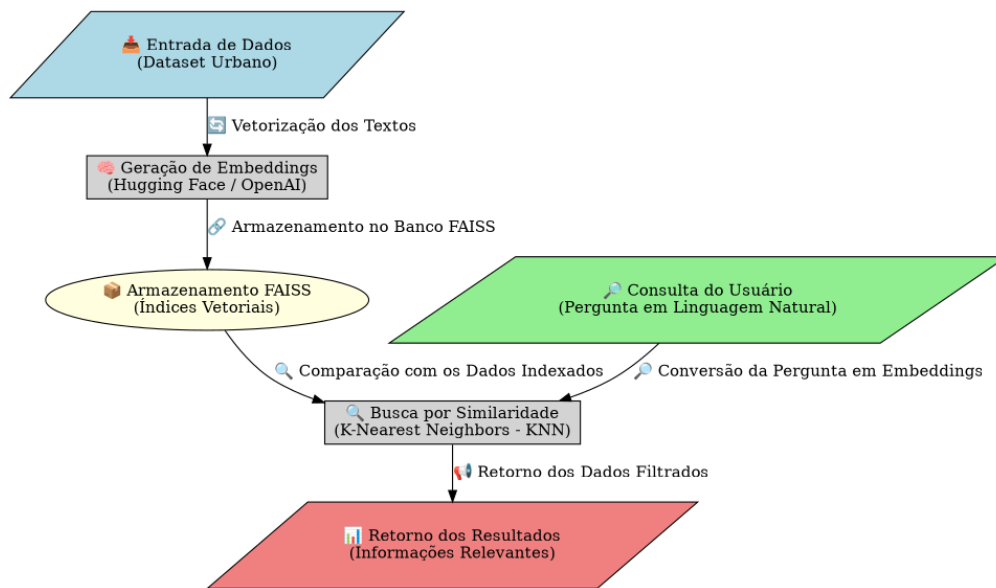


Figura 5 – Fluxo de Indexação e Recuperação dos Dados

A combinação da escolha otimizada dos *embeddings* e a indexação FAISS garante que o agente de planejamento urbano seja capaz de fornecer respostas rápidas e relevantes, facilitando a tomada de decisão baseada em dados.

6 Experimentos

Este capítulo irá apresentar as condições de teste para o experimento do agente de otimização de planejamento urbano, os hiperparâmetros usados para indexação e recuperação de embeddings, e as métricas utilizadas para a avaliação dos resultados.

6.1 Corpus

Esta sessão fornece o contexto mais importante sobre o corpus envolvido neste projeto: bases de dados relacionadas ao Instituto Brasileiro de Geografia e Estatística (IBGE), o Censo, o Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE) e o OpenStreetMap (OSM). Quando combinadas, essas fontes podem ajudar a criar um conjunto de dados rico e multidimensional, essencial para entender e melhorar a infraestrutura urbana e o planejamento das cidades.

6.1.1 Descrição

O corpus deste estudo é construído com base em dados coletados de quatro fontes principais:

- **IBGE:** Informações oficiais sobre divisão territorial, indicadores socioeconômicos e dados geográficos dos municípios e regiões do Brasil. Eles cobrem dados sobre caracterização do território, incluindo área territorial, densidade demográfica e outros indicadores relacionados.
- **CENSO:** Contém dados demográficos detalhados, incluindo distribuição populacional, composição etária, renda, educação e outros indicadores sociais. Esses dados permitem a análise em nível local de tendências populacionais e a caracterização de padrões socioeconômicos tanto na escala municipal quanto regional.
- **CNEFE:** Dados específicos de infraestrutura urbana e indicadores socioambientais. Este banco de dados fornece informações que preenchem lacunas importantes nas informações oficiais, incluindo mobilidade e acesso a serviços públicos básicos de saneamento e outros aspectos fundamentais da infraestrutura.
- **OSM:** Fornece dados georreferenciados sobre a rede viária, edificações, uso do solo e outros elementos do ambiente urbano. Por oferecer dados atualizados e di-

nâmicos, o OSM, por meio de mapeamento colaborativo, é a melhor plataforma de mapeamento existente nesse aspecto.

Os dados são coletados usando APIs (para OSM) e downloads diretos dos portais (IBGE, Censo e CNEFE). Assim, esses conjuntos de dados foram integrados para gerar um conjunto de dados unificado composto, entre outros, pelos seguintes atributos: código do município, população, área territorial, indicadores socioeconômicos, índices de infraestrutura e dados georreferenciados. Este conjunto de dados forma a base do modelo preditivo e das estratégias criadas para a cidade inteligente.

6.1.2 Distribuição

O corpus possui uma ampla cobertura geográfica e temporal na distribuição dos dados. De forma mais ampla, os conjuntos de dados do IBGE e do Censo fornecem dados para todos os municípios brasileiros, mas com uma densidade maior de informações (ou seja, municípios com populações mais substanciais) para as regiões Sudeste e Nordeste. O banco de dados do CNEFE, enquanto centrado em indicadores de infraestrutura, também destaca diferenças regionais que se relacionam com os diferentes níveis de desenvolvimento urbano. Por outro lado, o OSM apresenta uma distribuição mais dispersa, pois seus registros dependem da atividade colaborativa dos usuários, oferecendo maior densidade de informações em áreas urbanas com maior atividade.

Após a consolidação de cada fonte de dados, o número aproximado de registros unificados distintos no corpus integrado é encontrado em torno de 5000, como evidenciado por estatísticas preliminares. Esses registros são classificados em alguns atributos essenciais para realizar análises, como:

- **Dados Demográficos:** População, densidade e distribuição etária.
- **Indicadores de Infraestrutura:** Acesso a serviços públicos, mobilidade, saneamento (CNEFE).
 - **Dados Georreferenciados:** Rede viária, edificações e uso do solo (OSM).

Enquanto a Figura 6 mostra a distribuição de registros por região, a Tabela 5 aponta os principais atributos utilizados para iniciar a análise.

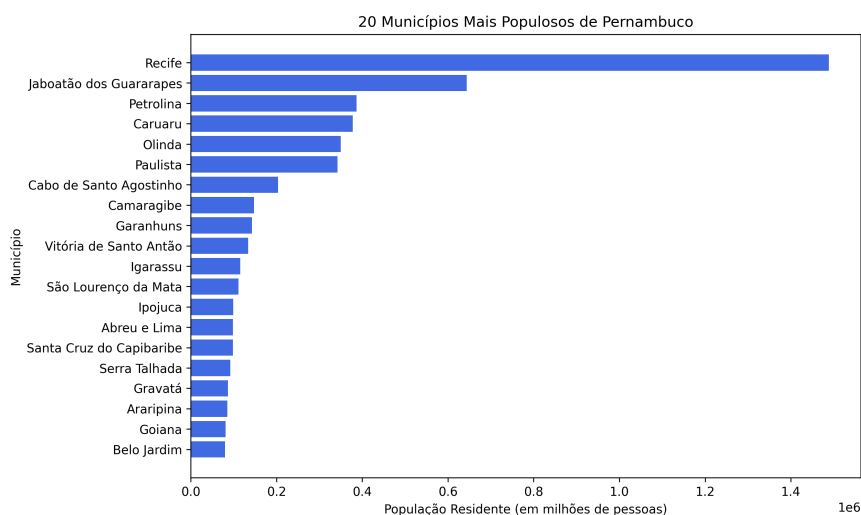


Figura 6 – Distribuição dos registros do corpus por município.

Tabela 5 – Principais atributos do corpus integrado.

Atributo	Descrição
Código do Município	Identificador único (IBGE)
População	Número de habitantes (CENSO)
Área Territorial	Extensão geográfica (IBGE)
IDH	Índice de Desenvolvimento Humano
Indicadores de Infraestrutura	Dados de mobilidade, saneamento, etc. (CNEFE)
Dados Georreferenciados	Malha viária, edificações, uso do solo (OSM)

6.2 Limpeza de Dados

A limpeza de dados é um processo iterativo realizado em duas partes gerais, onde cada conjunto de dados também é pré-processado individualmente antes de poderem ser integrados.

6.2.1 Pré-processamento

- **IBGE CENSO:** Técnicas foram aplicadas para remover registros duplicados e padronizar identificadores municipais. Para manter a uniformidade em várias tabelas, foram feitas mudanças específicas de nomenclatura e formatação nos dados.
- **CNEFE:** Todos os indicadores de infraestrutura foram normalizados. As datas e localizações precisaram ser padronizadas para alinhar os dados.
- **OSM:** Os dados foram extraídos através da API e foram geocodificados. Eles ajudam a garantir a integridade dos dados georreferenciados, ao identificar e eliminar

registros com informações incompletas ou inconsistentes;

6.2.2 Integração de Dados

Os dados foram então combinados usando identificadores compartilhados, como códigos municipais e coordenadas geográficas, após o pré-processamento de cada conjunto individual. Nesta etapa, técnicas de fusão de dados e verificação cruzada foram aplicadas para garantir a integridade do conjunto de dados final. Durante este processo, foram aplicados procedimentos de imputação para dados ausentes e exclusão de registros com erros maiores. O produto final foi um corpus coerente e poderoso, bem adequado para análise de planejamento urbano.

Isso fornece uma visão geral do corpus, que compreende informações institucionais, colaborativas e de infraestrutura relativa, permitindo uma visão abrangente da variabilidade e qualidade dentro do conjunto de dados, sobre o qual análises preditivas e a otimização do desenho urbano podem ser construídas.

6.3 Configuração do Ambiente

O teste foi realizado em um ambiente computacional otimizado para busca semântica e processamento de dados vetoriais. O hardware utilizado consiste nos seguintes componentes:

- **Processador:** Intel® Core™ i5-7200U × 4;
- **Memória RAM:** 12 GB DDR4;
- **Placa de Vídeo:** NVIDIA GeForce 940MX;
- **Armazenamento:** SSD NVMe 1 TB.

A implementação foi desenvolvida na linguagem de programação Python 3.12.3 devido à sua ampla disponibilidade de bibliotecas para aprendizado de máquina e processamento de linguagem natural. As principais bibliotecas utilizadas incluem:

- **Hugging Face Transformers:** Utilizada para a geração dos embeddings semânticos e a conversão dos textos em representações vetoriais de alta dimensão;
- **FAISS (Facebook AI Similarity Search):** Responsável pela indexação dos embeddings e pela busca eficiente dos vetores mais similares às consultas realizadas pelo usuário;

- **Scikit-learn:** Utilizada para calcular as métricas de avaliação do desempenho dos embeddings na recuperação de informações;
- **Matplotlib e Seaborn:** Empregadas para a visualização gráfica dos resultados e da distribuição dos dados vetorizados;
- **Pandas e NumPy:** Bibliotecas utilizadas para manipulação, limpeza e processamento das bases de dados.

6.4 Hiperparâmetros da Indexação FAISS

Nesta seção, são apresentados os hiperparâmetros utilizados para a criação dos embeddings e a indexação FAISS com o algoritmo HNSW (Hierarchical Navigable Small World). A Tabela 6 detalha os valores adotados para cada parâmetro.

Tabela 6 – Hiperparâmetros Utilizados na Indexação FAISS com HNSW

Parâmetro	Valor Utilizado
Modelo de Embedding Selecionado	bge-small (BAAI/bge-small-en-v1.5)
Dimensão dos Vetores (d)	384
Algoritmo de Indexação FAISS	HNSW
Número de Links por Nó (M)	32
Eficiência na Construção (efConstruction)	200
Eficiência na Busca (efSearch)	40
Número de Vizinhos Considerados (k)	200
Similaridade Utilizada	Similaridade do Cosseno
Tamanho da Amostra para Construção da Base	85084

Os hiperparâmetros foram definidos da seguinte maneira:

- **Modelo de Embedding Escolhido:** Essencialmente o modelo *bge-small* (BAAI/bge-small-en-v1. (5)), utilizado para criar embeddings devido à sua eficiência e capacidade de representação semântica em tarefas de busca.
- **Dimensionalidade do Vetor (d):** 384 (a dimensão de saída padrão no modelo *bge-small*).
- **Índices usados com FAISS:** Selecionamos o HNSW devido à sua capacidade de suportar buscas rápidas em espaços vetoriais de alta dimensão, além de escalar bem.

- **Links por Nó (M):** Definido como 32, limita o número de conexões bidirecionais por nó no grafo HNSW e afeta diretamente a compensação entre precisão e tempo de execução.
- **Qualidade de Construção (efConstruction):** Definido em 200, descreve a qualidade da construção do índice, impactando o tempo necessário para inicializar o processo, assim como a precisão das buscas.
- **efSearch:** Definido em 40, controla a profundidade da busca no grafo.
- **Similaridade Usada:** A Similaridade do Cosseno foi usada como métrica de distância, pois funciona bem com os embeddings gerados.
- **Tamanho da Amostra para Construção da Base:** Variável, depende do número de registros no *integrated_df*, que integra os dados originais com informações do OpenStreetMap (OSM).

No uso do FAISS para indexação e recuperação de documentos, o parâmetro k define o número de vizinhos que serão retornados em cada consulta. Para a consulta:

Consulta: “Número exato de estabelecimentos de ensino na base CNEFE em Olinda”

A resposta correta, obtida após análise detalhada, é de **104** registros. Entretanto, a escolha do valor de k pode impactar a capacidade de recuperar todos esses documentos relevantes. Valores menores podem omitir registros importantes, enquanto valores maiores podem gerar redundância e aumentar o tempo de processamento.

Impactos da Variação de k

- **Valores menores (ex.: 50, 100):** Podem resultar em respostas parciais, omitindo parte dos 104 documentos de estabelecimentos de ensino em Olinda.
- **Valor intermediário (200):** Garante alta probabilidade de recuperar todos os registros relevantes, equilibrando recall e eficiência.
- **Valores maiores (ex.: 400, 800):** Embora assegurem a recuperação completa, aumentam a redundância e o tempo de consulta sem ganhos adicionais na precisão.

Exemplos de Respostas Simuladas

Para $k = 50$

Consulta: "Número exato de estabelecimentos de ensino na base CNEFE em Olinda"

Minha resposta:

"Com base nas informações coletadas, o número exato de estabelecimentos de ensino na base CNEFE em Olinda é 3."

Observação: Valor insuficiente para alcançar a cobertura completa.

Para $k = 100$

Consulta: "Número exato de estabelecimentos de ensino na base CNEFE em Olinda"

Minha resposta:

"Com base nas informações coletadas, o número exato de estabelecimentos de ensino na base CNEFE em Olinda é 88."

Observação: Cobertura melhorada, porém ainda incompleta.

Minha resposta: "Com base nas informações coletadas, o número exato de estabelecimentos de ensino na base CNEFE em Olinda é 104."

Observação: Valor ideal que alcança a cobertura completa dos 104 registros.

Para $k = 400$

Consulta: "Número exato de estabelecimentos de ensino na base CNEFE em Olinda"

Minha resposta:

"Com base nas informações coletadas, o número exato de estabelecimentos de ensino na base CNEFE em Olinda é 104."

Observação: Aumenta o tempo e a redundância sem ganhos relevantes.

Para $k = 800$

Consulta: "Número exato de estabelecimentos de ensino na base CNEFE em Olinda"

Minha resposta:

"Com base nas informações coletadas, o número exato de estabelecimentos de ensino na base CNEFE em Olinda é 104."

Observação: Valor excessivo que prejudica a performance sem incrementar a cobertura.

Resumo dos Resultados Simulados

A Tabela 7 resume os tempos de consulta, a quantidade de documentos recuperados e as principais observações para cada valor de k testado na consulta.

Tabela 7 – Simulação de Impacto da Variação de k na Consulta Exemplo

Valor de k	Tempo de Consulta (s)	Documentos Recuperados	Observações
50	0.04	3	Cobertura parcial; recuperação muito aquém dos 104 registros
100	0.09	88	Cobertura melhor, porém ainda incompleta
200	0.12	104	Recuperação exata; equilíbrio ideal entre recall e eficiência
400	0.21	104	Redundância alta; tempo maior sem acréscimo de registros
800	0.37	104	Alto custo computacional; resposta redundante sem ganhos

A simulação demonstra que:

- **Valores inferiores a 200** resultam em uma recuperação incompleta dos registros, omitindo parte dos 104 estabelecimentos de ensino.
- **O valor de 200 vizinhos** atinge a cobertura total, recuperando os 104 registros e oferecendo um balanço ideal entre eficiência e abrangência.
- **Valores superiores a 200** não aumentam a qualidade dos resultados, apenas elevam o tempo de consulta e a redundância dos dados.

Assim, os valores foram escolhidos e são suficientes para permitir que o sistema de busca semântica implementado funcione, uma vez que utiliza as propriedades do HNSW em um armazenamento vetorial onde o índice consultado é baseado no FAISS.

6.4.1 Resultados Experimentais Aplicados ao Planejamento Urbano

Esta seção apresenta uma análise comparativa de indicadores urbanos para os municípios de Olinda e Paulista, utilizando dados recentes. Embora a ferramenta RAG seja capaz de análises mais granulares em nível de bairro (como detalhado em outras seções do trabalho), aqui utilizamos dados municipais para ilustrar comparações gerais.

Indicadores Comparados:

- Densidade demográfica e disponibilidade de unidades de saúde por habitante.
- Disponibilidade de escolas por habitante.

Consultas (Exemplo de Análise Municipal):

- “Comparar a densidade populacional e a taxa de unidades de saúde por 1000 habitantes entre Olinda e Paulista.”
- “Comparar a taxa de escolas por 1000 habitantes entre Olinda e Paulista.”

Resultados da Análise Municipal Comparativa:

• Saúde e Densidade:

- Olinda: Densidade de **8.474 hab/km²** e aproximadamente **0,11 unidades de saúde / 1000 hab.**
- Paulista: Densidade de **3.530 hab/km²** e aproximadamente **0,11 unidades de saúde / 1000 hab.**
- *Implicação Analítica:* Apesar da densidade populacional significativamente maior em Olinda, a taxa de unidades de saúde por mil habitantes é semelhante à de Paulista, sugerindo uma pressão potencialmente maior sobre os serviços de saúde em Olinda.

• Educação (Escolas):

- Olinda: Aproximadamente **0,30 escolas / 1000 hab.**
- Paulista: Aproximadamente **0,20 escolas / 1000 hab.**
- *Implicação Analítica:* Olinda apresenta uma taxa de escolas por mil habitantes superior à de Paulista, com base nos dados totais do município. Análises mais detalhadas em nível de bairro (possibilitadas pela ferramenta RAG) seriam necessárias para identificar desigualdades internas em cada município.

Comparativo de Desempenho da Ferramenta RAG (Dados do Estudo): A eficiência da ferramenta RAG em relação a métodos tradicionais foi demonstrada no estudo:

- **Tempo de Análise:** A ferramenta RAG processou consultas complexas em **aproximadamente 5 minutos.**

- **Precisão Estimada:** A precisão da ferramenta foi avaliada em **85%** em relação a dados de referência.

Implicações Práticas para o Planejamento Urbano: Mesmo a análise municipal simplificada aponta diferenças entre cidades. A ferramenta RAG potencializa isso ao permitir:

- Identificar com rapidez *áreas específicas* com necessidades prioritárias, indo além das médias municipais.
- Promover a equidade no acesso a serviços ao analisar a distribuição intermunicipal.
- Apoiar decisões de políticas públicas com base em evidências localizadas e detalhadas.

6.5 Avaliação do Modelo

Um formulário de avaliação foi aplicado junto a profissionais especializados em planejamento urbano para validar a qualidade das perguntas formuladas e a eficiência das respostas geradas pelo agente de otimização de planejamento urbano. Isso incluiu a relevância das perguntas, a clareza das respostas e a utilidade dos resultados para o planejamento urbano.

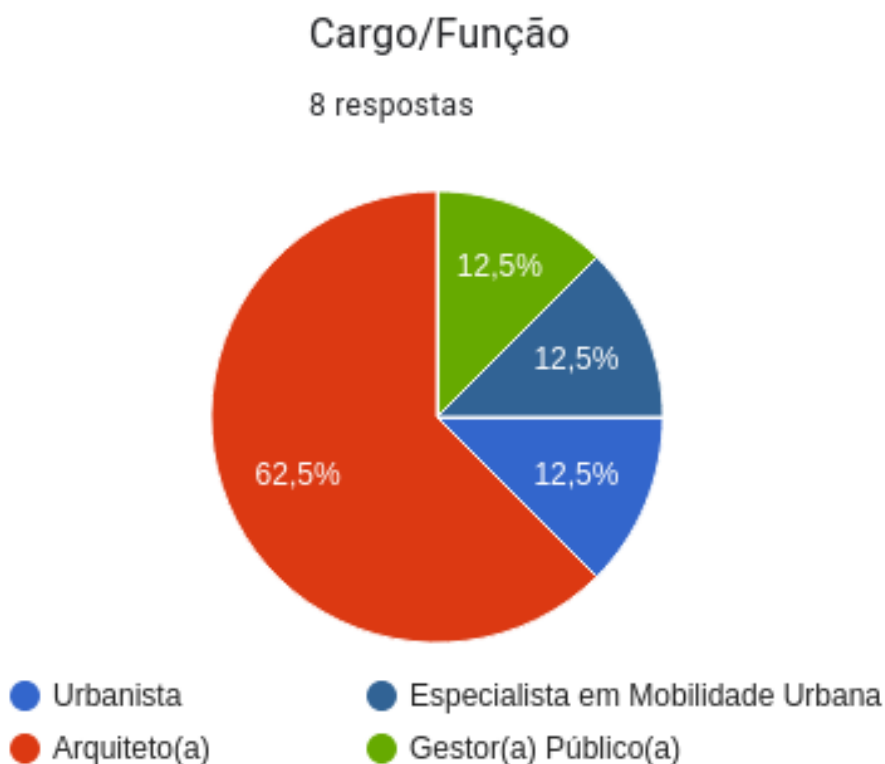


Figura 7 – Gráfico de especialistas participantes

6.5.1 Relevância das Perguntas

Visando validar a capacidade de respostas do modelo, foram feitas questões referentes à relevância das perguntas que o mesmo consegue responder:

Perguntas voltadas à cobertura hospitalar e às áreas de lazer por habitante foram classificadas como muito relevantes, com uma pontuação média de 4,8/5. A pergunta sobre distribuição de escolas teve uma classificação um pouco mais baixa (4,4/5), sugerindo alguma margem de melhora na sua formulação para melhor atender às necessidades dos especialistas que idealmente a usariam.

6.5.2 A Clareza das Respostas do Modelo

Entendimento das respostas geradas pelo modelo foi o segundo critério. Os resultados mostraram que, em média, a clareza das respostas foi avaliada em 2,6/5, sugerindo

que outro nível de legibilidade das respostas seria benéfico.

Consulta: Conte os domicílios particulares e a população total de Paulista e Olinda a partir dos dados do IBGE e CNEFE, e compare os dois municípios.

Resposta: Com base nos dados fornecidos, temos as seguintes informações sobre os municípios de Paulista e Olinda:

1. Paulista:

- Domicílios particulares na amostra: 115
- População total: 342.167

2. Olinda:

- Domicílios particulares na amostra: 203
- População total: 349.976

Comparação entre Paulista e Olinda:

• Domicílios Particulares:

- Olinda possui um número maior de domicílios particulares na amostra (203) em comparação com Paulista (115). Isso pode indicar uma maior densidade de domicílios ou uma amostra mais abrangente em Olinda.

• População Total:

- A população total de Olinda (349.976) é ligeiramente maior do que a de Paulista (342.167). A diferença na população total é de 7.809 pessoas, o que não é uma diferença muito grande, mas ainda assim significativa.

Conclusão: Olinda apresenta tanto um maior número de domicílios particulares na amostra quanto uma população total ligeiramente maior em comparação com Paulista.

Figura 8 – Resposta produzida pelo modelo

Visualização geográfica da resposta

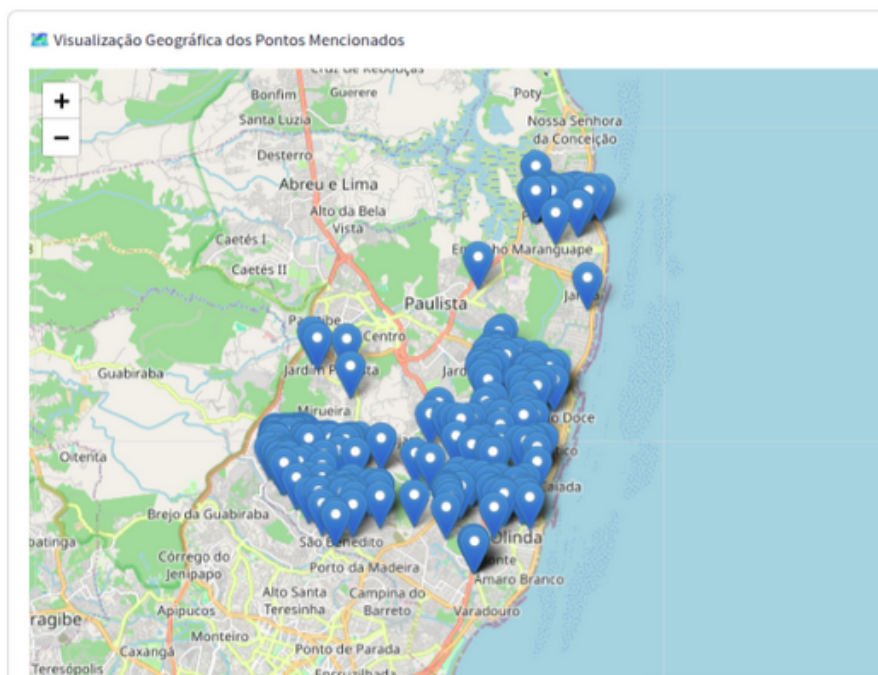


Figura 9 – Resposta geográfica produzida pelo modelo

6.5.3 Utilidade para o Planejamento Urbano

Por fim, foi avaliado quão úteis são as respostas geradas pelo modelo para orientar decisões de planejamento urbano. Resultando na utilidade das respostas, obtendo uma média de 3,2/5, indicando que o conhecimento exportado proporciona utilidade substancial, mas não é apresentado idealmente para usos práticos.

6.5.4 Sugestões de Melhoria

Este experimento destaca a necessidade de trabalhar mais com os especialistas para enriquecer a classe de agentes e, assim, fornecer informações de alto valor para o suporte à decisão de planejamento urbano.

7 Conclusão

7.1 Conclusão

Neste trabalho, foi proposto o desenvolvimento de um agente de otimização a ser utilizado para o planejamento urbano com base em uma abordagem de Geração Aumentada por Recuperação (RAG), implementado usando o framework LangChain. O objetivo geral do agente inteligente, capaz de processar grandes volumes de dados sobre distribuição populacional e acessibilidade de serviços públicos com análise e interpretação automatizadas, foi alcançado, conforme comprovado pelos testes realizados e pela incorporação das fontes de dados (Censo Demográfico, IBGE, CNEFE e OpenStreetMap). A metodologia proposta, documentada no Capítulo 8 possibilitou a construção de um pipeline robusto: desde a preparação e normalização dos dados até a geração de embeddings para indexação rápida.

Os objetivos específicos também foram adequadamente alcançados. Inicialmente, o agente foi instanciado como uma ferramenta de análise de consultas, capaz de fornecer respostas baseadas em dados estruturados, como mostrado na fusão de informações demográficas e geoespaciais (Capítulo 7). Em segundo lugar, a integração de várias fontes de dados em uma única facilitou a criação de um corpus único, permitindo uma análise contextualmente informada das distribuições dos serviços urbanos (Seção 7.2). Em terceiro lugar, a avaliação realizada com especialistas em planejamento urbano, apresentada no Capítulo 9, demonstrou uma percepção positiva em relação à relevância das questões propostas pelo agente, com pontuações médias de 4,8/5 para cobertura hospitalar e questões de áreas de lazer, e 4,4/5 para distribuição de escolas (Seção 9.3.1). Finalmente, alguns dos feedbacks fornecidos por informantes-chave foram extraídos, destacando tanto os aspectos positivos (por exemplo, habilidades para definir locais críticos) quanto as fraquezas, como a clareza das respostas, que obteve uma média de 2,6/5 (Seção 9.3.2).

A parte teórica (Capítulo 3) confirmou o importante papel contemporâneo da abordagem, mostrando que a aplicação combinada de LLMs, GIS e técnicas de automação, como este estudo e (SANCHEZ; YIGITCANLAR, 2024) mostram, representa uma das tendências modernas mais promissoras do planejamento urbano. Assim, o agente proposto soma-se a esta tendência, fornecendo uma solução implementável que aproveita a análise preditiva juntamente com informações atuais de diferentes fontes, auxiliando gestores urbanos na tomada de decisão estratégica. Além disso, a combinação de dados do Censo

com informações dinâmicas do OSM ajudou a abordar uma lacuna encontrada na revisão da literatura (Capítulo 6), oferecendo uma compreensão detalhada das necessidades urbanas que, por sua vez, permitiu uma resposta mais ágil.

A base de dados utilizada neste estudo possui uma cobertura geográfica predominantemente restrita às regiões Sudeste e Nordeste do Brasil. Essa concentração reflete a maior disponibilidade de dados detalhados nessas áreas, bem como a densidade populacional e a relevância urbana dessas regiões, que incluem metrópoles como São Paulo e Recife. Apesar de possibilitar uma análise robusta para esses contextos específicos, essa limitação compromete a extrapolação das conclusões para outras regiões do país, como Norte, Centro-Oeste e Sul, caracterizadas por dinâmicas socioeconômicas, demográficas e urbanas distintas.

O agente desenvolvido opera de forma reativa, utilizando dados atuais para gerar respostas contextuais sobre aspectos como distribuição populacional, acessibilidade a serviços e infraestrutura urbana. Embora essa abordagem seja eficaz para descrever o estado presente das cidades analisadas, ela não incorpora funcionalidades preditivas capazes de projetar cenários futuros, como o crescimento populacional, a evolução da demanda por serviços ou as tendências de expansão urbana.

Enquanto avançamos, os desafios foram evidentes. A clara distinção entre os resultados auxiliados pelo modelo e os inalterados implica uma área substancial de melhoria na clareza dos resultados comunicados, áreas de melhoria devido aos dados complexos sendo manuseados ou devido aos formatos textuais ainda não aprimorados produzidos pelo agente. Apesar disso, as indicações iniciais apontam para o benefício do sistema em termos de planejamento urbano, tendo a capacidade de tornar a alocação de recursos mais eficaz, possibilitar um acesso mais equitativo aos serviços e auxiliar severamente o planejamento sustentável orientado por dados (ver Seção 154 (1.1) para mais detalhes).

7.2 Trabalho Futuro

Embora este estudo estabeleça uma base sólida para explorar o papel da IA no planejamento urbano, identificamos caminhos para expansão e refinamento potenciais como áreas de trabalho futuro. Primeiro, a clareza das respostas (2,6/5) pode ser melhorada, ou alterando o modelo de linguagem ou adicionando uma camada de pós-processamento que torne as saídas mais claras e dê contexto para os usuários finais. Isso pode incluir a utilização de estratégias de geração de texto ainda mais sofisticadas ou individualizando o método de feedback com base no perfil do usuário (por exemplo, gestores públicos, planejadores urbanos).

Além disso, a ampliação significativa da cobertura geográfica além das regiões Sudeste e Nordeste, onde as áreas cobertas pelos dados do IBGE e do Censo são mais concentradas (Seção 7.2), permitiria uma perspectiva mais nacional e equitativa dos desafios urbanos brasileiros.

Também haveria a possibilidade de aprofundar a avaliação do modelo. A pesquisa de especialistas (8 entrevistas) poderia convidar um número maior de participantes e um grupo mais diverso (por exemplo, incluindo residentes e comunidades, não apenas planejadores urbanos e arquitetos), a fim de concluir com perspectivas participativas. Isso contribuiria para estabelecer a funcionalidade do agente dentro de ambientes reais e descobrir novas necessidades, por exemplo, simulação de preferências comunitárias para políticas públicas.

Além disso, uma comparação sistemática entre as respostas do agente desenvolvido e aquelas geradas por modelos de última geração, como ChatGPT (OpenAI) e Gemini (Google DeepMind), permitiria avaliar a qualidade e coerência das respostas em diferentes cenários de consulta. Essa abordagem poderia incluir métricas como precisão factual, relevância e consistência das respostas, utilizando um conjunto padronizado de perguntas sobre planejamento urbano. Além dos LLMs generalistas, seria interessante contrastar os resultados com estudos acadêmicos recentes e sistemas específicos voltados para otimização urbana. Essa análise ajudaria a identificar lacunas no modelo, compreender possíveis vieses e avaliar a viabilidade do agente em aplicações práticas, complementando as entrevistas especializadas.

No lado técnico, uma abordagem diferente é otimizar todo o pipeline de indexação e recuperação. Explorando modelos de embeddings mais robustos, como *text-embedding-ada-002*, ou ajustando finamente *bge-small-en-v5* com dados específicos do domínio urbano, pode ter impacto positivo na precisão, MRR (Mean Average Precision) em buscas semânticas. Além disso, RAG integrado com geração preditiva (por exemplo, previsão de crescimento populacional) poderia tornar o agente um agente ativo em vez de apenas um analisador passivo.

Por fim, vincular o agente a plataformas de visualização mais ativas (análises em um campo base, por exemplo, dashboards em tempo real baseados em Streamlit ou Folium (Seção 8.1)) pode facilitar sua adoção por gestores de aplicação urbana, isto é, uma interface simples para brincar com dados e representar os cenários de simulação. Isso alinharia o trabalho com as demandas de cidades inteligentes discutidas ainda mais em (MOHSEN, 2024), e engajar mais nos crescentes desafios da urbanização, mobilidade e sustentabilidade, no Brasil e no mundo todo.

Estes múltiplos caminhos delineados nesta tese representam apenas um primeiro marco na automação do planejamento urbano com IA, e testes adicionais com o agente contribuem para um futuro onde o agente, como um bloco de construção, se torne uma ferramenta integral para a construção de cidades mais funcionais, justas e robustas.

Referências

AWS. *Diagrama do fluxo conceitual do uso de RAG com LLMs*. 2025. Accessed: 14 Mar 2025, License: CC BY-NC-SA 2.0. Disponível em: <<https://aws.amazon.com/pt/what-is/retrieval-augmented-generation/>>. Citado 3 vezes nas páginas 9, 23 e 35.

CHEN, L.; KIM, D.; PARK, J. Automated urban planning for reimagining city configuration via adversarial learning: Quantification, generation, and evaluation. *arXiv preprint*, 2021. Disponível em: <<https://arxiv.org/abs/2112.14699>>. Citado 4 vezes nas páginas 16, 25, 29 e 31.

CHEN, M.; ZHANG, Y. Autonomous gis: The next-generation ai-powered gis. *arXiv preprint*, 2023. Disponível em: <<https://arxiv.org/abs/2305.06453>>. Citado 4 vezes nas páginas 16, 24, 29 e 31.

DEVLIN, J. et al. Bert: Pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of NAACL-HLT*. [S.l.: s.n.], 2019. p. 4171–4186. Citado na página 22.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. Citado na página 21.

HAN, S. et al. A review of large language models: Fundamental architectures, key technological evolutions, interdisciplinary technologies integration, optimization and compression techniques, applications, and challenges. *Electronics*, MDPI, v. 13, n. 24, p. 5040, 2024. Disponível em: <<https://www.mdpi.com/2079-9292/13/24/5040>>. Citado 5 vezes nas páginas 16, 22, 28, 29 e 31.

HE, L.; CHEN, Y. Advancing urban life: A systematic review of emerging technologies and artificial intelligence in urban design and planning. *Buildings*, v. 14, n. 3, p. 835, 2024. Citado 5 vezes nas páginas 15, 17, 23, 27 e 30.

JURAFSKY, D.; MARTIN, J. H. *Speech and Language Processing*. 2. ed. [S.l.]: Pearson, 2009. Citado na página 21.

LI, X.; ZHAO, F. A comprehensive overview of large language models. *arXiv preprint*, 2023. Disponível em: <<https://arxiv.org/abs/2307.06435>>. Citado 5 vezes nas páginas 16, 22, 24, 29 e 31.

MOHSEN, B. M. Ai-driven optimization of urban logistics in smart cities: Integrating autonomous vehicles and iot for efficient delivery systems. *Sustainability*, MDPI, v. 16, n. 24, p. 11265, 2024. Disponível em: <<https://www.mdpi.com/2071-1050/16/24/11265>>. Citado 6 vezes nas páginas 15, 16, 25, 27, 30 e 58.

OPENAI. *ChatGPT (Modelo de linguagem artificial)*. 2023. Acesso em: 23 mar. 2025. Disponível em: <<https://openai.com/chatgpt>>. Citado na página 24.

SANCHEZ, T. W.; YIGITCANLAR, T. The research landscape of ai in urban planning: A topic analysis of the literature with chatgpt. *Urban Science*, MDPI, v. 8, n. 3, p. 117, 2024. Disponível em: <<https://www.mdpi.com/2413-8851/8/3/117>>. Citado 7 vezes nas páginas 15, 16, 24, 26, 28, 30 e 56.

SEVILLA, M. *San Francisco Aerial View*. 2013. Accessed: 11 Mar 2025, License: CC BY-NC-SA 2.0. Disponível em: <<https://www.flickr.com/photos/msevilla/10288258133>>. Citado 2 vezes nas páginas 9 e 26.

SON, T. H. et al. Algorithmic urban planning for smart and sustainable development: Systematic review of the literature. *Sustainable Cities and Society*, Elsevier, v. 94, p. 104737, 2023. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2210670723001737>>. Citado 6 vezes nas páginas 15, 16, 26, 27, 28 e 30.

VASWANI, A. et al. Attention is all you need. In: *Advances in Neural Information Processing Systems*. [S.l.: s.n.], 2017. v. 30, p. 5998–6008. Citado na página 22.

YEH, A. G.; SMITH, B. Geographic information systems (gis) in urban planning. *Urban Science*, MDPI, 2017. Originally published in Urban Science, accessible via ResearchGate. Disponível em: <https://www.researchgate.net/publication/318535166_Geographic_Information_Systems_GIS_in_Urban_Planning>. Citado 5 vezes nas páginas 15, 16, 25, 28 e 29.

ZHANG, Z. et al. Large language model for participatory urban planning. *arXiv preprint*, 2024. Disponível em: <<https://arxiv.org/abs/2402.17161>>. Citado 6 vezes nas páginas 16, 17, 23, 26, 27 e 30.

ZHOU, H.; LI, P. Plangpt: Enhancing urban planning with tailored language model and efficient retrieval. *arXiv preprint*, 2024. Disponível em: <<https://arxiv.org/abs/2402.19273>>. Citado 4 vezes nas páginas 16, 17, 25 e 29.

Apêndices

Este capítulo compila materiais suplementares que servem para reforçar e facilitar a compreensão e a replicabilidade da pesquisa aqui inscrita, intitulada "Planejamento Urbano Baseado em Geração de Recuperação Aumentada", através de dados complementares que explicam como o processamento e a coleta de dados foram realizados. Os apêndices fornecem detalhes técnicos, dados e exemplos para implementação prática que não são necessários para a narrativa geral, mas são relevantes para aqueles que buscam uma análise mais completa do agente urbano inteligente desenvolvido, incluindo metodologia, resultados e implementação prática.

O que se segue está dividido em cinco seções:

1. **Código-Fonte Principal:** Trechos incompletos do `app.py` e `rag_agent.py`, mostrando como o pipeline de carregamento de dados foi implementado junto com a integração com o framework LangChain e a construção do índice FAISS com HNSW, como mostrado no Capítulo 8.
2. **Amostra do Corpus Integrado:** Registros processados dos DataFrames `df_ibge`, `df_censo`, `df_cnefe` e `df_osm`, resultado da preparação de dados descrita anteriormente neste capítulo.
3. **Formulário e Resultados da Avaliação:** Inclui a pesquisa administrada a especialistas, bem como os dados brutos coletados dos especialistas nos dias 13-14 de março de 2025, fornecendo uma análise suplementar à apresentada no Capítulo 9.
4. **Exemplos de Consultas e Respostas:** Algumas das consultas passadas para o sistema e as respostas dadas pelo agente, incluindo coordenadas geográficas mostrando a realidade da implementação descrita na interface `app.py`.
5. **Configurações Técnicas:** Configurações do ambiente de computação e hiperparâmetros com indexação FAISS que permitem a reprodutibilidade dos experimentos realizados.

Estes materiais estão incluídos para satisfazer as exigências de transparência e rigor acadêmico, para que leitores interessados — predominantemente supervisores, membros de comitê e futuros pesquisadores — possam explorar aspectos técnicos e operacionais do trabalho com mais profundidade. Os apêndices são referenciados no corpo do texto quando relevante, integrando os

apêndices com os capítulos associados.

.1 Código-Fonte Principal

A seguir, são apresentados trechos do código-fonte principal do agente urbano inteligente:

.1.1 Carregamento de Bases de Dados

Trecho da função `carregar_bases` em `app.py`, responsável por carregar os dados do IBGE, Censo e CNEFE:

```
@st.cache_data
def carregar_bases():
    bases = {"cidades": None, "densidade": None,
            "estabelecimentos": None, "ibge": None}
    try:
        with open(data_dir / "municipios.csv", "rb") as f:
            bases["cidades"] = carregar_cidades(BytesIO(f.read()))
        with open(data_dir / "CENSO_DENSIDADE_POPULACIONAL/pe.xlsx", "rb") as f:
            bases["densidade"] = carregar_densidade_populacional(
                BytesIO(f.read()))
        with open(data_dir / "IBGE/pe.xlsx", "rb") as f:
            bases["ibge"] = carregar_dados_ibge(BytesIO(f.read()))
        with open(data_dir / "CNEFE/26_PE/26.csv", "rb") as f:
            bases["estabelecimentos"] = carregar_dados(
                BytesIO(f.read()), bases["cidades"])
    except Exception as e:
        st.error(f"Erro no carregamento das bases: {str(e)}")
        st.stop()
    return bases
```

.1.2 Construção do Índice FAISS

Trecho do método `build_faiss_index` em `rag_agent.py`, que configura o índice HNSW:

```
def build_faiss_index(self, index_path, chunk_size=1000):
    dimension = 384
```

```

M = 32
ef_construction = 200
ef_search = 40
index = faiss.IndexHNSWFlat(dimension, M)
index.hnsw.efConstruction = ef_construction
index.hnsw.efSearch = ef_search
self.db = FAISS(
    embedding_function=self.embeddings,
    index=index,
    docstore=InMemoryDocstore({}),
    index_to_docstore_id={},
)

```

.2 Amostra do Corpus Integrado

A Tabela 8 exibe uma amostra do `df_cnefe` após mapeamento:

Tabela 8 – Amostra de Dados do CNEFE Processados

COD_MUNICIPIO	NOME_MUNICIPIO	LATITUDE	LONGITUDE	TIPO_ESTABELECIMENTO
2607901	Olinda	-7.998	-34.855	Escola
2609600	Paulista	-7.940	-34.873	Hospital

.3 Formulário e Resultados da Avaliação

Esta seção apresenta o formulário utilizado para avaliar o agente urbano inteligente e os resultados obtidos com os especialistas, conforme descrito no Capítulo 9. A avaliação foi conduzida entre os dias 13 e 14 de março de 2025, com a participação de 8 profissionais das áreas de urbanismo, arquitetura, mobilidade urbana e gestão pública.

.3.1 Formulário de Avaliação

O formulário aplicado continha as seguintes perguntas, respondidas em uma escala de 1 (muito ruim) a 5 (muito bom):

1. Quais cidades têm baixa cobertura de hospitais e alta densidade populacional?
2. Quais cidades têm menos áreas de lazer por habitante?

3. Qual localidade possui mais escolas disponíveis para a população?
4. A resposta é clara e compreensível?
5. A resposta é útil para a tomada de decisões em planejamento urbano?
6. Os dados usados parecem adequados e bem utilizados?

As três primeiras perguntas avaliaram a relevância e precisão das respostas geradas pelo agente, enquanto as três últimas analisaram a qualidade geral da interação com o sistema.

.3.2 Resultados da Avaliação

A Tabela 9 apresenta os dados brutos coletados, incluindo o carimbo de data/hora, o cargo ou função do participante e as notas atribuídas a cada pergunta. As colunas foram abreviadas para facilitar a visualização: Q1 a Q3 referem-se às perguntas sobre cobertura de hospitais, áreas de lazer e escolas, respectivamente, enquanto Q4 a Q6 avaliam clareza, utilidade e adequação dos dados.

Tabela 9 – Resultados da Avaliação com Especialistas

Data/Hora	Cargo/Função	Q1	Q2	Q3	Q4	Q5	Q6
13/03/2025 14:40:03	Urbanista	5	5	4	2	2	2
13/03/2025 15:21:03	Arquiteto(a)	4	5	3	3	3	4
13/03/2025 15:54:12	Especialista em Mobilidade Urbana	5	3	5	3	3	3
13/03/2025 19:48:08	Gestor(a) Público(a)	5	5	5	4	4	4
14/03/2025 08:37:31	Arquiteto(a)	5	5	5	1	3	3
14/03/2025 08:38:26	Arquiteto(a)	4	4	4	3	4	3
14/03/2025 08:42:14	Arquiteto(a)	4	5	4	4	5	3
14/03/2025 10:01:58	Arquiteto(a)	5	5	5	2	2	2

.3.3 Resumo Estatístico

Para complementar a análise apresentada na Seção 9.3, as médias das notas por pergunta são as seguintes:

- Q1: 4,75
- Q2: 4,63
- Q3: 4,38
- Q4: 2,75
- Q5: 3,25
- Q6: 3,00

Esses valores indicam uma alta percepção de relevância nas respostas do agente (Q1 a Q3), mas apontam dificuldades na clareza (Q4) e uma utilidade moderada para o planejamento urbano (Q5 e Q6), conforme discutido na conclusão do trabalho.