

UNIVERSIDADE FEDERAL DE VIÇOSA
CAMPUS VIÇOSA
CIÊNCIA DA COMPUTAÇÃO

ANA CLARA GUERRA TORRES

**APLICAÇÃO DE GRANDES MODELOS DE
LINGUAGEM NA CLIMATOLOGIA E
AGROMETEOROLOGIA: ANÁLISE DAS MUDANÇAS
CLIMÁTICAS E SEUS IMPACTOS NA ESTAÇÃO
CHUVOSA NA AMAZÔNIA E NO CERRADO**

VIÇOSA
2025

ANA CLARA GUERRA TORRES

APLICAÇÃO DE GRANDES MODELOS DE LINGUAGEM NA
CLIMATOLOGIA E AGROMETEOROLOGIA: ANÁLISE DAS
MUDANÇAS CLIMÁTICAS E SEUS IMPACTOS NA ESTAÇÃO
CHUVOSA NA AMAZÔNIA E NO CERRADO

Monografia apresentada à Universidade
Federal de Viçosa como parte das exigên-
cias para a aprovação na disciplina Tra-
balho de Conclusão de Curso I.

Orientador: DANIEL LOUZADA FER-
NANDES

Coorientador: JULIO CESAR SOARES
DOS REIS

VIÇOSA

2025

Resumo

As mudanças climáticas representam um dos maiores desafios científicos e sociais da atualidade, impactando diretamente os ecossistemas e a produção agrícola. A Amazônia e o Cerrado, regiões fundamentais para a regulação climática na América do Sul, vêm sofrendo alterações expressivas em função do desmatamento e da mudança no uso do solo, fatores que afetam a dinâmica da estação chuvosa e a sustentabilidade da produção agrícola. Nesse contexto, os Grandes Modelos de Linguagem (LLMs) surgem como ferramentas promissoras para análise e síntese de informações complexas, integrando dados climáticos, geoespaciais e relatórios ambientais. Este trabalho tem como objetivo investigar a aplicação de LLMs na climatologia e agrometeorologia, com foco na análise dos impactos das mudanças climáticas na estação chuvosa da Amazônia e do Cerrado. A pesquisa será conduzida a partir de uma revisão bibliográfica sistemática, seguida da proposição de um *framework* conceitual baseado em técnicas como *Retrieval-Augmented Generation* (RAG) e *instruction tuning*. Espera-se, ao final, fornecer bases metodológicas para o uso de LLMs em análises climáticas regionais, contribuindo para a mitigação de riscos e para a tomada de decisão em políticas ambientais e agrícolas.

Palavras-chaves: Mudanças Climáticas; Amazônia; Cerrado; Modelos de Linguagem; Agrometeorologia.

Abstract

Climate change is one of the greatest scientific and social challenges of our time, directly affecting ecosystems and agricultural production. The Amazon and the Cerrado, key regions for climate regulation in South America, have been experiencing significant changes due to deforestation and land-use change, which alter the rainy season dynamics and compromise agricultural sustainability. In this context, Large Language Models (LLMs) emerge as promising tools for analyzing and synthesizing complex information, integrating climatic, geospatial, and environmental data. This work aims to investigate the application of LLMs in climatology and agrometeorology, focusing on the analysis of climate change impacts on the rainy season in the Amazon and Cerrado. The research will be carried out through a systematic literature review, followed by the proposal of a conceptual framework based on techniques such as Retrieval-Augmented Generation (RAG) and instruction tuning. The expected outcome is to provide methodological foundations for the use of LLMs in regional climate analysis, supporting risk mitigation and decision-making in environmental and agricultural policies.

Key-words: Climate Change; Amazon; Cerrado; Large Language Models; Agrometeorology.

Sumário

1	Introdução	5
1.1	Objetivo Geral	5
1.2	Objetivos Específicos	6
2	Referencial Teórico	7
2.1	Mudanças Climáticas e Estação Chuvosa	7
2.2	Grandes Modelos de Linguagem (LLMs)	7
2.2.1	Adaptação de LLMs a Domínios Específicos	7
2.3	Integração com Dados Externos: RAG	8
3	Trabalhos Relacionados	9
3.1	LLMs de Uso Geral	9
3.1.1	Survey of LLMs	9
3.1.2	GPT-3	9
3.2	LLMs de Domínio Específico	9
3.2.1	AgriGPT	9
3.2.2	ClimateChat	10
3.3	LLMs e Geoprocessamento	11
3.3.1	GeoGPT	11
3.3.2	GeoLLM	11
3.4	Quadro Comparativo	12
3.5	Tabela Comparativa de Propriedades	13
4	Metodologia	14
4.1	Avaliação e Métricas	14
5	Cronograma	16
	Referências	17

1 Introdução

As mudanças climáticas estão entre os maiores desafios do século XXI, com impactos profundos nos ecossistemas e na produção agrícola. A Amazônia e o Cerrado, em especial, ocupam papel central na regulação climática regional e global.

Contudo, o desmatamento e a mudança no uso do solo têm modificado o ciclo hidrológico, afetando diretamente a estação chuvosa. Isso gera atrasos no início das chuvas, irregularidade na distribuição pluviométrica e aumento de extremos climáticos, com sérias consequências para a agricultura e a segurança hídrica.

Os métodos tradicionais da climatologia e agrometeorologia enfrentam limitações diante da complexidade e do volume dos dados. Nesse contexto, os Grandes Modelos de Linguagem (LLMs) apresentam-se como alternativa promissora para apoiar análises e interpretações, dada sua capacidade de sintetizar informações complexas em linguagem natural. Trabalhos como o de Chen et al. (2025) demonstram que a adaptação de LLMs a domínios específicos, como o ClimateChat, pode oferecer análises mais detalhadas e factualidade aprimorada ao integrar a técnica de Retrieval-Augmented Generation (RAG) para contextualizar as respostas a partir de um corpus especializado (CHEN et al., 2025).

A hipótese deste trabalho é que LLMs, quando especializados para climatologia e agrometeorologia, podem contribuir na análise dos impactos do desmatamento e das mudanças no uso do solo sobre a estação chuvosa na Amazônia e Cerrado.

1.1 Objetivo Geral

Investigar a aplicação de Grandes Modelos de Linguagem (LLMs) na climatologia e agrometeorologia, com foco na análise das mudanças climáticas e seus impactos na estação chuvosa da Amazônia e Cerrado.

1.2 Objetivos Específicos

- a) Mapear o estado da arte sobre o uso de LLMs aplicados à climatologia, agrometeorologia e geociências;
- b) Analisar metodologias existentes para adaptação de LLMs em domínios especializados;
- c) Identificar lacunas em trabalhos relacionados à aplicação de LLMs em mudanças climáticas;
- d) Propor um framework conceitual de uso de LLMs para análise da estação chuvosa na Amazônia e Cerrado;
- e) Estruturar um cronograma de pesquisa de setembro/2025 a maio/2026.

2 Referencial Teórico

2.1 Mudanças Climáticas e Estação Chuvosa

A Amazônia exerce papel essencial na reciclagem de umidade e regulação climática. O desmatamento tem atrasado o início da estação chuvosa e aumentado a irregularidade pluviométrica, com impactos diretos para a agricultura no Cerrado.

2.2 Grandes Modelos de Linguagem (LLMs)

Os Grandes Modelos de Linguagem (Large Language Models – LLMs) surgiram a partir da evolução dos modelos estatísticos de linguagem para arquiteturas baseadas em *deep learning*, tendo os Transformers como marco fundamental (VASWANI et al., 2017). LLMs são redes neurais com bilhões de parâmetros, treinadas em grandes volumes de dados textuais, capazes de capturar padrões linguísticos e produzir textos coerentes, traduzir idiomas, responder perguntas e realizar inferências complexas.

Esses modelos apresentam capacidades emergentes como *few-shot learning* e *zero-shot learning* (BROWN et al., 2020), permitindo executar tarefas não vistas no treinamento apenas com instruções em linguagem natural.

2.2.1 Adaptação de LLMs a Domínios Específicos

Embora os LLMs gerais apresentem grande poder, sua aplicação em domínios especializados (como climatologia e agrometeorologia) requer adaptação. As principais estratégias incluem:

- **Fine-Tuning:** treinamento adicional de todo o modelo ou de camadas específicas usando um corpus especializado;

- **Parameter-Efficient Fine-Tuning (PEFT)**: métodos como LoRA (*Low-Rank Adaptation*), que reduzem custo computacional ao ajustar apenas parte dos parâmetros;
- **Instruction Tuning**: adaptação baseada em exemplos de instruções e respostas, tornando o modelo mais alinhado a tarefas específicas;
- **RLHF (Reinforcement Learning with Human Feedback)**: técnica que utiliza feedback humano para ajustar respostas, aumentando alinhamento e confiabilidade.

2.3 Integração com Dados Externos: RAG

Um dos maiores desafios na aplicação de LLMs é lidar com a obsolescência e a falta de dados especializados no treinamento original. Nesse contexto, o paradigma de *Retrieval-Augmented Generation* (RAG) torna-se relevante.

O RAG combina um modelo de busca em bases externas (repositórios científicos, dados geoespaciais, relatórios climáticos) com a geração de linguagem natural. Assim, o LLM não precisa memorizar todo o conhecimento, mas pode recuperar informações relevantes no momento da consulta e usá-las para gerar respostas contextualizadas e atualizadas.

Essa técnica é especialmente útil na climatologia, onde dados regionais, séries temporais e relatórios técnicos mudam frequentemente e precisam ser integrados às análises.

3 Trabalhos Relacionados

Esta seção apresenta os principais trabalhos que dialogam com a proposta deste estudo. Cada subseção detalha o foco, a metodologia, as vantagens, as limitações e as diferenças em relação ao presente trabalho.

3.1 LLMs de Uso Geral

3.1.1 Survey of LLMs

O trabalho de [Zhao et al. \(2025\)](#) apresenta uma revisão abrangente sobre os avanços de LLMs, incluindo arquiteturas, técnicas de adaptação (fine-tuning, LoRA, instruction tuning) e desafios éticos e computacionais. **Vantagens:** fornece uma visão panorâmica do campo, sistematizando conceitos. **Limitações:** não possui aplicação prática em climatologia ou geociências. **Diferença:** serve como base conceitual para este trabalho, mas não aborda o problema climático.

3.1.2 GPT-3

O modelo GPT-3 ([BROWN et al., 2020](#)) mostrou a eficácia do *few-shot learning*, sendo capaz de executar tarefas sem treinamento específico. **Vantagens:** demonstra o poder de generalização de LLMs. **Limitações:** não é especializado em domínios científicos, podendo gerar respostas imprecisas em temas técnicos. **Diferença:** enquanto GPT-3 é generalista, este trabalho busca adaptação a climatologia e agrometeorologia.

3.2 LLMs de Domínio Específico

3.2.1 AgriGPT

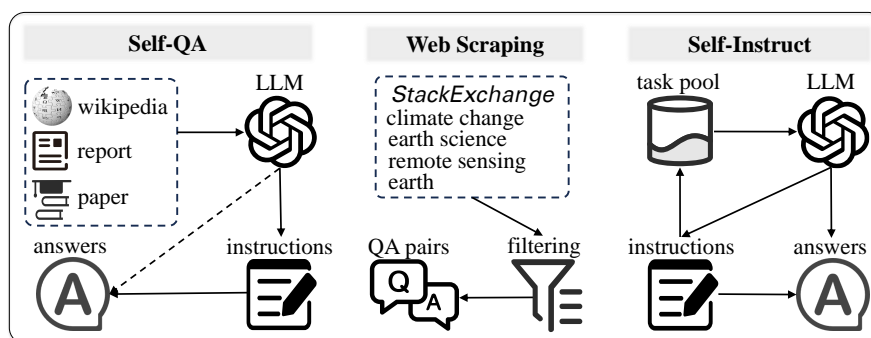
O AgriGPT ([YANG et al., 2025](#)) foi proposto como um ecossistema especializado em agricultura, construído a partir de um corpus de 342 mil artigos

científicos e adaptado via Tri-RAG. **Vantagens:** fornece respostas contextualizadas e relevantes para agricultura, com benchmarks específicos. **Limitações:** não aborda mudanças climáticas ou a dinâmica da estação chuvosa. **Diferença:** inspira este trabalho na ideia de construir corpus especializado, mas foca em agricultura e não em clima.

3.2.2 ClimateChat

O ClimateChat (CHEN et al., 2025) adaptou LLMs para mudanças climáticas utilizando um corpus climático e o método de *Retrieval-Augmented Generation* (RAG). Esse trabalho serve como uma inspiração metodológica crucial, demonstrando como a adaptação a um domínio específico pode aprimorar significativamente a factualidade e a capacidade de análise dos modelos, em contraste com o desempenho de LLMs generalistas. O estudo de Chen et al. (2025) estabelece um fluxo de trabalho detalhado para a criação de um corpus especializado através de técnicas como *Self-QA* e *Web Scraping*, além de utilizar o RAG para contextualizar as respostas. Conforme ilustrado na Figura 1, o processo de construção do corpus é uma etapa fundamental que direciona a especialização do modelo.

Figura 1 – Fluxograma de construção do ClimateChat-Corpus.



Fonte: Adaptado de Chen et al. (2025).

Vantagens: Oferece análise aprofundada em relatórios climáticos, com respostas contextualizadas e baseadas em evidências. **Limitações:** Não possui um recorte regional específico, como a Amazônia e o Cerrado, o que é uma lacuna

a ser preenchida. **Diferença:** O presente trabalho adapta e aplica a metodologia do ClimateChat, propondo uma especialização regional para a análise dos impactos do desmatamento e das mudanças no uso do solo sobre a estação chuvosa na Amazônia e no Cerrado. O foco é na coleta e organização de um corpus de dados brasileiros sobre agrometeorologia e clima regional, e na aplicação do RAG como a principal estratégia de integração para o LLM.

3.3 LLMs e Geoprocessamento

3.3.1 GeoGPT

O GeoGPT (ZHANG et al., 2024) integrou LLMs a sistemas de informações geográficas (GIS), permitindo consultas em linguagem natural a mapas e bases espaciais. **Vantagens:** automatiza tarefas de geoprocessamento com grande precisão. **Limitações:** foca em GIS genérico, não em análises climáticas. **Diferença:** inspira integração GIS + clima, mas não trata da estação chuvosa.

3.3.2 GeoLLM

O GeoLLM (MANVI et al., 2024) utilizou LLMs para extração de conhecimento de dados geoespaciais e socioeconômicos (como OpenStreetMap). **Vantagens:** melhora a análise de padrões espaciais complexos. **Limitações:** foco em aplicações urbanas e socioeconômicas, não ambientais. **Diferença:** este trabalho aplica a lógica a dados climáticos, com foco em impactos regionais.

3.4 Quadro Comparativo

Trabalho	Foco Principal	Método / Dataset	Vantagens	Limitações
Survey of LLMs (ZHAO et al., 2025)	Revisão ampla sobre LLMs	Revisão sistemática; análise de arquiteturas, tuning e avaliação	Síntese abrangente das técnicas e desafios; base conceitual sólida	Não aplica em climatologia
GPT-3 (BROWN et al., 2020)	Modelo generativo e few-shot	GPT-3 (175B) pré-treinado em grande corpus da web	Capacidade de few/zero-shot; alta generalização	Não especializado; possíveis alucinações e vieses
AgriGPT (YANG et al., 2025)	Agricultura e agrometeorologia	Corpus Agri-342K; Tri-RAG (recuperação densa + esparsa + grafo)	Corpus curado; benchmarks específicos	Foco agrícola amplo; não analisa estação chuvosa regional
ClimateChat (CHEN et al., 2025)	Mudanças climáticas	ClimateChat-Corpus (Self-QA, Web Scraping, Self-Instruct) + RAG	Especialização temática; melhor factualidade via RAG	Não tem recorte Amazônia/Cerrado
GeoGPT (ZHANG et al., 2024)	Integração LLM + GIS	LLM orquestrador + ferramentas GIS (buffer, intersect, clip...)	Automatiza workflows GIS via NL; bom para operações espaciais	Depende de ferramentas externas; não foca análise climática temporal
GeoLLM (MANVI et al., 2024)	Extração de conhecimento geoespacial	LLM + dados auxiliares (ex.: OpenStreetMap)	Melhora previsões espaciais; extrai padrões latentes de LLMs	Aplicações voltadas a ambientes urbanos/socioeconômicos
Nosso Trabalho	Mudanças climáticas na Amazônia e Cerrado	Corpus especializado + RAG	Recorte regional inédito; integração de dados climatológicos; suporte à pesquisa aplicada	Corpus em construção; avaliação inicial restrita a especialistas

Quadro 1: Comparação dos principais trabalhos relacionados.

3.5 Tabela Comparativa de Propriedades

Tabela 1 – Comparação de propriedades dos principais trabalhos relacionados.

Trabalho	Corpus Especializado	RAG	Foco em Climatologia	Few/Zero-Shot	Integração Ferramentas	Cobertura Regional
Survey of LLMs (ZHAO et al., 2025)	X	X	X	✓	X	X
GPT-3 (BROWN et al., 2020)	X	X	X	✓	X	X
AgriGPT (YANG et al., 2025)	✓	✓	X	✓	X	X
ClimateChat (CHEN et al., 2025)	✓	✓	✓	✓	X	X
GeoGPT (ZHANG et al., 2024)	X	X	X	✓	✓	X
GeoLLM (MANVI et al., 2024)	X	X	X	✓	✓	X
Nosso Trabalho	✓	✓	✓	✓	X	✓

4 Metodologia

A pesquisa será conduzida em três etapas principais, seguindo uma abordagem inspirada em trabalhos como o de [Chen et al. \(2025\)](#) sobre a adaptação de LLMs para domínios específicos:

1. **Revisão Bibliográfica Sistemática:** Mapeamento do estado da arte sobre o uso de LLMs aplicados à climatologia, agrometeorologia e geociências. Esta etapa visa identificar as principais arquiteturas, técnicas de treinamento (como *fine-tuning* e RAG) e lacunas existentes na literatura.
2. **Proposição de *Framework* Conceitual:** A partir da revisão, será proposto um *framework* para a construção de um LLM especializado em climatologia e agrometeorologia regional. Esta etapa incluirá:
 - Escolha de um LLM base (ex.: LLaMA 3, Mistral), justificando a opção por modelos de código aberto ou via API.
 - Construção de um corpus especializado com foco na Amazônia e no Cerrado, utilizando metodologias de coleta de dados como as empregadas no estudo do ClimateChat (*Self-QA*, *Web Scraping* e *Self-Instruct*).
 - Definição da estratégia de integração, com foco na implementação do *Retrieval-Augmented Generation* (RAG) como método principal para garantir a factualidade e o acesso a dados atualizados.

4.1 Avaliação e Métricas

A validação da proposta será realizada a partir de um processo de avaliação cega conduzido por especialistas da área de climatologia. Um conjunto de pesquisadores selecionará previamente perguntas relacionadas a mudanças climáticas, estação chuvosa e impactos do desmatamento na Amazônia e no Cerrado.

Essas perguntas serão submetidas tanto ao modelo proposto (LLM especializado em climatologia/agrometeorologia na região citada) quanto a modelos

generalistas de referência (ainda a serem definidos). As respostas geradas serão apresentadas de forma anônima aos avaliadores, de modo que não seja possível identificar a origem de cada resposta.

Cada avaliador atribuirá notas de acordo com critérios como:

1. Acurácia científica: correção dos conceitos apresentados;
2. Clareza e coerência: estrutura e compreensão da resposta;
3. Relevância: grau em que a resposta aborda efetivamente a pergunta;
4. Utilidade prática: aplicabilidade no contexto da climatologia/agrometeorologia.

Os resultados dessa avaliação permitirão comparar a performance do LLM especializado em relação aos generalistas, fornecendo uma métrica qualitativa da sua eficácia.

5 Cronograma

Tabela 2 – Cronograma previsto (setembro/2025 – maio/2026).

Mês/Ano	Revisão Bibliográfica	Estudo de Técnicas	Proposição do Framework	Texto Parcial	Defesa
Set/2025	X				
Out/2025	X				
Nov/2025		X			
Dez/2025		X			
Jan/2026		X			
Fev/2026			X		
Mar/2026			X		
Abr/2026			X	X	
Mai/2026					X

Referências

BROWN, T. B. et al. Language models are few-shot learners. **Advances in Neural Information Processing Systems**, v. 33, p. 1877–1901, 2020. Acesso em: 01 out. 2025. Disponível em: <<https://arxiv.org/abs/2005.14165>>.

CHEN, L. et al. Climatechat: Domain-adapted large language models for climate change research. **Environmental Modelling & Software**, 2025. Preprint arXiv:2406.09876. Acesso em: 21 set. 2025. Disponível em: <<https://arxiv.org/abs/2506.13796v1>>.

MANVI, R. et al. Geollm: Harnessing large language models for geospatial knowledge extraction. **Transactions in GIS**, 2024. Preprint arXiv:2401.04567. Acesso em: 26 set. 2025. Disponível em: <<https://arxiv.org/abs/2310.06213>>.

VASWANI, A. et al. Attention is all you need. **Advances in Neural Information Processing Systems**, v. 30, 2017. Acesso em: 02 out. 2025. Disponível em: <<https://arxiv.org/abs/1706.03762>>.

YANG, X. et al. Agrigpt: A domain-specific large language model for agriculture. **Computers and Electronics in Agriculture**, 2025. Preprint arXiv:2404.01234. Acesso em: 01 out. 2025. Disponível em: <<https://arxiv.org/abs/2508.08632v1>>.

ZHANG, H. et al. Geogpt: Integrating large language models with geographic information systems. **International Journal of Geographical Information Science**, 2024. Preprint arXiv:2312.06789. Acesso em: 20 set. 2025. Disponível em: <<https://arxiv.org/abs/2307.07930>>.

ZHAO, W. X. et al. A survey of large language models. **Artificial Intelligence Review**, 2025. Preprint arXiv:2303.18223. Acesso em: 28 set. 2025. Disponível em: <<https://arxiv.org/abs/2303.18223>>.