BRACIS
2025

Semantic Clustering in the Context of Legislative Amendments

Pedro Lucas C. de Andrade[1]
André de Carvalho[1]
Ellen Souza[2]
Nádia Silva[3]

[1] *ICMC(USP)*
[2] *UFRPE*
[3] *UFG*

October 1, 2025

Overview

Ulysses Project: AI Solutions

- The Ulysses project was initiated in 2019 in partnership with USP and the Chamber of Deputies.
- Focus on developing AI solutions to assist the plenary in its demands.
- This Work: Exploration of the use of GPT-4o and GPT-4o-mini models to cluster legislative documents.

The Problem

Introduction to the Brazilian Legislative Process

In Brazil, the legislative process includes drafting, analyzing, and voting on Bills (PL) and Constitutional Amendments (PEC). Amendments are proposed to modify original texts. Given the large volume of amendments, automation offers an efficient solution for organizing them.

Methodology

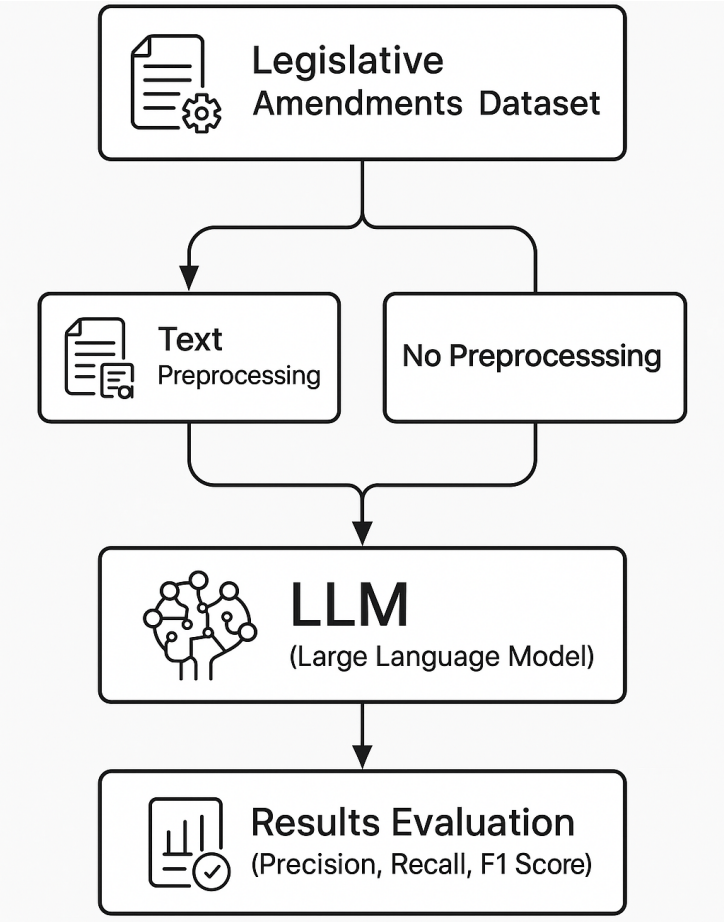

Figure: Project's Pipeline

- Corpus: Amendments from Article 21 of PL 280/2020 and Article 4 of PL 8167/2022.
- Preprocessing: stopwords removal.
- LLMs: GPT-4o and GPT-4o-mini.

Table

Treatments	Response 1	Response 2
Treatment 1	0.0003262	0.562
Treatment 2	0.0015681	0.910
Treatment 3	0.0009271	0.296

Table: Table caption

Results

Temp.	GPT-4o	GPT-4o-mini	GPT-4o [Prep.]	GPT-4o-mini [Prep.]
0.0	**0.7789**	0.6960	0.7085	0.6846
0.25	**0.8000**	0.7031	0.7178	0.6883
0.5	**0.7854**	0.7043	0.7201	0.6787
0.75	**0.7671**	0.7105	0.7109	0.6865
1.0	**0.7664**	0.7235	0.6937	0.6727

Table: Clustering Precision for PL 280-2020

Temp.	GPT-4o	GPT-4o-mini	GPT-4o [Prep.]	GPT-4o-mini [Prep.]
0.0	**0.8007**	0.7660	0.7223	0.6960
0.25	**0.7857**	0.7518	0.7179	0.7032
0.5	**0.7756**	0.7308	0.7067	0.7043
0.75	**0.7517**	0.7494	0.6858	0.7105
1.0	**0.7614**	0.7190	0.6984	0.7235

Table: Clustering Precision for PL 8167-2022

Conclusion

Figure: AI Hallucination Illustration

1. Temperature Impact: Lower values produced more accurate and consistent clusters.
2. Preprocessing: Amendments without textual treatment produced better results than the preprocessed ones.
3. Limitations:
   - Hardware and API Cost.
   - Risk of Hallucination.

References

Barros Melo

ChatGPT: Brasil é o 4º país que mais utiliza a inteligência artificial. Disponível em: https://www.barrosmelo.edu.br/noticia/chatgpt-brasil-e-o-4o-pais-que-mais-utiliza-a-inteligencia-artificial. Acesso em: 03 nov. 2024.

OpenAI

API Pricing. Disponível em: https://openai.com/api/pricing/. Acesso em: 04 nov. 2024.

Vayadande, K., Bhat, A., Bachhav, P., Bhoyar, A., Charoliya, Z., and Chavan, A. (2024). Ai-powered legal documentation assistant. *In Proceedings of the 4th International Conference on Pervasive Computing and Social Networking (ICPCSN). IEEE.*

Contact

- Email: pedroandrade@usp.br
- Repository: https://github.com/Pedrest15/TCC