

# Universidade de Pernambuco Programa de Pós-Graduação em Engenharia da Computação (PPGEC)

## Proposta de Dissertação de Mestrado

**Área: Computação Inteligente**

**Título: Integração de *Large Language Models* e Métodos Tradicionais de Processamento de Linguagem Natural para Análise Qualitativa em Pesquisas Sociais**

**Orientador – Fernando Buarque de Lima Neto ([fbln@ecomp.poli.br](mailto:fbln@ecomp.poli.br))**

**Co-orientador – Luís Filipe de Araújo Pessoa ([lfap@ecomp.poli](mailto:lfap@ecomp.poli))**

### Contexto

Análises temáticas e de conteúdo [1] em pesquisas sociais são duas das principais metodologias utilizadas para análise de dados qualitativos. Contudo, a análise manual é intensiva em tempo e recursos, limitando a escala e velocidade de pesquisas sociais importantes. Assim, diversos métodos estatísticos e de Inteligência Computacional (IC) para o Processamento de Linguagem Natural (NLP) têm sido tradicionalmente aplicados para a identificação de padrões (e.g., sentimentos, opiniões, temas) e geração de insights relevantes, possibilitando a análise eficaz mesmo em contextos com grande volume de dados [2]. O surgimento de *Large Language Models* (LLMs) [3] para conversação e *embeddings* com capacidades avançadas de processamento textual e compreensão semântica [4], além de suporte a contextos cada vez maiores, oferece novas possibilidades para análise qualitativa automatizada.

### Problema

Embora LLMs apresentem potencial significativo, sua aplicação em pesquisas sociais introduz limitações metodológicas críticas: inconsistências e alucinações que geram interpretações não fundamentadas nos dados, limitações de explicabilidade e interpretabilidade, problemas de replicabilidade, e eventuais dificuldades no tratamento de características regionais e culturais específicas. Por outro lado, é possível explicar as premissas metodológicas e o funcionamento dos métodos tradicionais de NLP, assim como verificar e interpretar seus resultados.

### Hipótese

A integração estruturada de LLMs com métodos tradicionais de NLP, através de um framework que incorpore supervisão metodológica e feedback humano especializado, pode combinar a eficiência dos LLMs com o rigor metodológico dos métodos tradicionais, resultando em um sistema híbrido adequado para análise qualitativa que atenda aos padrões científicos necessários para pesquisas sociais rigorosas.

### Pergunta Principal

Como desenvolver um framework híbrido que integre eficazmente LLMs com métodos tradicionais de NLP para análise temática e de conteúdo em pesquisas sociais, incorporando supervisão metodológica e feedback humano para garantir rigor científico?

### Perguntas Secundárias

- Quais métodos tradicionais de NLP têm sido aplicados de forma efetiva e responsável para as diferentes etapas e tarefas da análise temática e de conteúdo em pesquisas sociais?
- Como combinar as capacidades de LLMs com métodos tradicionais de NLP (e.g., modelagem de tópicos, análise de sentimentos, classificação textual) para diferentes etapas do processamento dos dados e da análise temática e de conteúdo?
- Como implementar mecanismos de supervisão metodológica e *feedback* humano para permitir que pesquisadores sociais validem, corrijam e refinem resultados de análises híbridas?

### Objetivos

Desenvolver uma ferramenta híbrida que integre LLMs com métodos tradicionais de NLP para análise temática e de conteúdo em pesquisas sociais, incorporando supervisão metodológica e *feedback* humano para garantir rigor científico e aplicabilidade prática.

### Produtos esperados

- Revisão estruturada das metodologias e tecnologias de NLP aplicadas às diferentes tarefas de análise temática e de conteúdo em pesquisas sociais.
- Modelo para integração de LLMs com métodos tradicionais de NLP para a análise qualitativa.
- Modelo para integração de *feedback* humano para permitir a validação, correção e refinamento dos resultados.
- Teste e avaliação de aplicação da contribuição em estudos de caso em pesquisa social aplicada.

### Referências

- [1] Souza, L. K. Pesquisa com análise qualitativa de dados: conhecendo a Análise Temática. Arquivos Brasileiros de Psicologia, v. 71, n. 2, Rio de Janeiro, 2019, pp. 51-67.
- [2] Macanovic, A. (2022). Text mining for social science–The state and the future of computational text analysis in sociology. Social Science Research, 108, 102784.
- [3] Naveed, H. et al. (2023). A comprehensive overview of large language models. ACM Transactions on Intelligent Systems and Technology.
- [4] Gardazi, N. M. et al. (2025). BERT applications in natural language processing: a review. Artificial Intelligence Review, 58(6), 1-49.