

# Universidade de Pernambuco

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

### Proposta de Dissertação de Mestrado

**Área:** Computação Inteligente ou Modelagem Computacional

**Título:** Explorando Interações Sociais com Múltiplas Pessoas Usando Large Visual-Language Models

**Orientador(a):** Pablo Barros (pvab@ecom.poli.br)

**Co-Orientador(a):** Bruno Fernandes (bjtf@ecom.poli.br)

#### Descrição:

As interações sociais humanas são fenômenos complexos que envolvem múltiplos conceitos interdependentes, como teoria da mente, normas sociais, papéis sociais e dinâmicas de relações interpessoais [1]. Com o avanço da inteligência artificial, há uma oportunidade de utilizar modelos de linguagem visual (Visual Large Language Models, VLLMs) para entender e modelar essas interações, especialmente em cenários envolvendo múltiplas pessoas [2]. No entanto, desafios permanecem: como esses modelos podem integrar informações visuais e linguísticas para capturar nuances como intenções ocultas, expectativas sociais e hierarquias de papéis? Além disso, a avaliação desses modelos em contextos de interações reais é limitada por datasets que frequentemente não abrangem a complexidade das relações sociais em ambientes diversos.

Este projeto visa desenvolver uma abordagem baseada em Visual Large Language Models para modelar os conceitos fundamentais de interações sociais. Utilizando a base de dados *KokoMind* [3], rica em interações sociais complexas geradas e verificadas por especialistas, o objetivo será treinar e avaliar VLLMs na capacidade de interpretar cenas visuais, entender diálogos e responder a perguntas relacionadas à dinâmica social [4]. A metodologia inclui a adaptação de VLLMs para incorporar representações visuais detalhadas e linguísticas, permitindo uma análise mais profunda de aspectos como empatia, regras implícitas e decisões morais em contextos sociais. Além disso, será proposta uma métrica personalizada para medir o desempenho dos modelos na compreensão de interações multi-party, com foco em diversificação e escalabilidade.

O candidato deve possuir uma base sólida em aprendizado de máquina, com ênfase em modelos de linguagem e visão computacional. Experiência em frameworks como PyTorch ou TensorFlow será essencial, além de conhecimento prático em processamento de linguagem natural (NLP) e visão computacional (CV). Habilidades em engenharia de dados, incluindo a manipulação de datasets complexos como o *KokoMind*, serão críticas para garantir o sucesso do projeto. Por fim, será necessário um bom entendimento de inglês, já que o projeto será incluído em colaborações internacionais.

#### Referências Bibliográficas:

- [1] Leng, Y., & Yuan, Y. (2023). Do LLM Agents Exhibit Social Behavior?. *arXiv preprint arXiv:2312.15198*.
- [2] Karanjai, R., & Shi, W. (2024). LookALike: Human Mimicry based collaborative decision making. *arXiv preprint arXiv:2403.10824*.
- [3] Shi, W., Qiu, L., Xu, D., Sui, P., Lu, P., & Kokomind, Z. Y. Can large language models understand social interactions?, July 2023. URL <https://chats-lab.github.io/KokoMind>.
- [4] Menon, S., & Vondrick, C. (2022). Visual classification via description from large language models. *arXiv preprint arXiv:2210.07183*.