

# Universidade de Pernambuco

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

### Proposta de Tese de Doutorado

**Área: Computação Inteligente**

**Título: Uso de Modelos de Larga Escala para Análise de Dados em Saúde**

**Orientador – Carmelo José Albanez Bastos Filho ([carmelo.filho@upe.br](mailto:carmelo.filho@upe.br))**

**Coorientador – Marcia R. O. B. C. Macedo ([marcia.macedo@upe.br](mailto:marcia.macedo@upe.br))**

**Descrição** – A área da saúde tem sido beneficiada há alguns anos pelo avanço das técnicas de inteligência computacional, principalmente a partir do uso de técnicas de aprendizado de máquina, mais especificamente a utilização das abordagens de Recurrent Neural Networks (RNNs) e Convolutional Neural Networks (CNNs). Dois exemplos clássicos de aplicação destas que podem ser citados respectivamente são as análises de prontuários eletrônicos via RNN para descoberta de padrões e predição de desfechos clínicos futuros. E ainda a aplicação de CNNs para investigação de padrões em exames de imagens ou histopatológicos. Porém, o aprimoramento das técnicas de processamento de linguagem natural precisa ser fortalecido para a interação com o paciente. Aliado a isso, os modelos de larga escala (LLMs) e as IAs generativas podem contribuir para isso (Ayers, Poliak & Dredze, 2023). Os LLMs são algoritmos de inteligência artificial que reconhecem, resumem e geram linguagem humana a partir de grandes conjuntos de dados baseados em texto. Os LLMs são capazes de extrair informações de big data. Esses modelos são aprendendo padrões sobre como as palavras e frases são comumente usadas juntas. Quando alimentado com uma nova entrada de texto, um LLM tentará prever ou gerar a continuação mais provável desse texto com base no que aprendeu durante o treinamento (Miner, Milstein & Hancock, 2017). Embora os LLMs já existam há algum tempo, ganharam a mídia através do ChatGPT, interface de chat para modelos LLM GPT-3 e GPT-4. Além disso, outra questão a ser estudada refere-se à pandemia de COVID-19, que acelerou a adoção de cuidados de saúde virtuais, concomitante com um aumento de 1,6 vezes nas mensagens eletrônicas dos pacientes, com cada mensagem acrescentando 2,3 minutos de trabalho no registro eletrônico de saúde e mais trabalho fora do horário de expediente (Holmgren et al., 2022). Volume adicional de mensagens prevê aumento do esgotamento para os médicos e aumentam a probabilidade de que as mensagens dos pacientes não sejam respondidas ou recebam respostas inúteis (Duziak, 2023). O objetivo deste trabalho é investigar e desenvolver soluções para responder questões como: Qual melhor equilíbrio em termos de requisitos de interpretabilidade para um modelo de LLM para diagnóstico de doenças em seres humanos? Como buscar um *trade-off* entre precisão de previsões e confiança no modelo? Quanto a redução de dimensionalidade de dados pode contribuir para o equilíbrio satisfatório entre precisão e transparência do processo de previsão do modelo? Como gerar explicações em âmbito local (indivíduo) e global (população) sobre diagnóstico e progressão de doenças progressão, tendo-se um *trade-off* razoável em termos de interpretação dos resultados e tempo de execução de máquina. Pretende-se que estas perguntas sejam respondidas a partir de experimentações usando bases dados geradas junto à plataforma de cidades inteligentes da UPE, “Na Palma da Mão”.

#### Referências Bibliográficas

1. Ayers J. A, Poliak, A., Dredze, M.. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum JAMA Intern Med. 2023;183(6):589-596. doi:10.1001/jamainternmed.2023.1838
2. Dudziak, E. A. (2023). O efeito ChatGPT baseado em Inteligência Artificial e o ensino nas Universidades. Portal ABCD. Disponível em: <https://www.abcd.usp.br/noticias/ia-e-o-ensino-nas-universidades>
3. Miner, A. S, Milstein, A, Hancock, J. T. Talking to machines about personal mental health problems. JAMA. 2017;318(13):1217-1218. doi:10.1001/jama.2017.14151
4. Holmgren, A. J., Downing, N. L., Tang, M., Sharp, C., Longhurst, C., Huckman, R. S. Assessing the impact of the COVID-19 pandemic on clinician ambulatory electronic health record use. J Am Med Inform Assoc. 2022;29(3):453-460.