Sebastian Joseph

952-649-8738 | sebyjoseph2015@gmail.com | Austin, TX
github.com/SebaJoe | https://www.linkedin.com/in/sebajoe | sebajoe.github.io

EDUCATION

The University of Texas at Austin, Austin, TX May 2024
Master of Science, Computer Science | *Bachelor of Science, Computer Science* | *Business Minor*
GPA *3.9*, Five-year Integrated Masters Program
Relevant Coursework: Spoken Language Technologies (ASR), Grounded Natural Language Processing, Natural Language Generation (OpenAI, LLM), Natural Language Processing (Pytorch, TensorFlow, HuggingFace, Data mining), Computer Vision (Matlab, Python), Computer Graphics (C++, OpenGL), Advanced Operating Systems (C, Bash), Concurrency (Go, C, Rust, CUDA), Human-Computer Interaction (HTML/CSS, JS, JQuery, Django), Symbolic Programming (Clojure)

SKILLS

Technical Skills:

- **Highly Proficient:** C, C++, Java, Python
- **Web Development:** HTML, CSS, Javascript, JQuery, Django
- **Data mining:** SQL, Pandas, Matplotlib, and SciPy
- **Machine Learning**: PyTorch, TensorFlow, HuggingFace, OpenAI API, LLM Prompting
- **Other Languages**: Git, Clojure, Go, Rust, CUDA, Bash, MATLAB
- OpenGL graphics pipeline
- Natural Language Processing, Computer Vision

Languages: Malayalam (Native), Japanese (Intermediate), Spanish (Intermediate)
Certifications: Cisco Certified Network Associate in Routing and Switching **(CCNA),** Microsoft Technology Certificate in Networking **(MTA)**

EXPERIENCE

The University of Texas at Austin, Austin, TX January 2022 - Present
Research Assistant

- Contributing to an effort to improve access to factual medical information through automatic simplification methods.
- Developed a novel annotation interface for annotating simplification datasets.
- Led a research initiative in developing a multilingual medical simplification dataset.
- Published papers for premier computational linguistics conferences (ACL, EMNLP).

The University of Texas at Dallas, Richardson, TX November 2018 - June 2019
Research Assistant

- Created large datasets for entity classification to help classify the usefulness of product reviews.
- Created a rule-based program to analyze sentiment in product reviews using semantic and syntactic relations.

RELEVANT PROJECTS

- **Multilingual Medical Simplification (Skills Used: Machine Learning, Python, LLM Prompting)**
  - Developed a multilingual medical simplification dataset containing sentence aligned medical abstracts and plain-language summaries across various languages.
  - Performed supervised finetuning on several language models (LM) to create systems that can perform multilingual medical simplification.
- **QUD-based Elaborative Simplification Tool (Skills Used: LLM Prompting, Flask, HTML/CSS)**
  - Built an interactive web tool where users can highlight parts of complex or simplified text and provide queries for additional clarifications.
  - Answers to user queries are generated through a large language model and are seamlessly rewritten into the original text.
- **Visual Storytelling with CLIP (Skills Used: Machine Learning, Reinforcement Learning, Python)**
  - Trained a multimodal model for the task of generating a story about a sequence of input images.
  - Used a reward model that combined a CNN-based architecture with CLIP to score text-image pairs.
- **Sentence Alignment & Labeling Annotation Tool (Skills Used: Python, HTML/CSS)**
  - Built a web tool to help users align sentences between two similar texts and label them if necessary.
  - Users had the option to allow the tool to rank alignments based on similarity measures to aid alignment.