Sebastian Joseph

952-649-8738 | sebyjoseph2015@gmail.com | Austin, TX
github.com/SebaJoe | linkedin.com/in/sebastian-joseph-a30228171 | sebajoe.github.io

EDUCATION

The University of Texas at Austin, Austin, TX May 2024
Integrated B.S/M.S., Computer Science
GPA *3.9*
Minor in Business
Relevant Coursework: Natural Language Generation, Natural Language Processing, Computer Vision, Computer Graphics, Advanced Operating Systems, Human-Computer Interaction, Symbolic Programming

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: English (Native), 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 Natural Language Processing related research
- Created a rule-based program to analyze sentiment using semantic relations

RELEVANT PROJECTS

- **Multilingual Medical Simplification**
  - 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**
  - 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.
- **Sentence Alignment & Labeling Annotation Tool**
  - 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.