Sebastian Joseph

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

EDUCATION

The University of Texas at Austin, Austin, TX   May 2024
Integrated M.S./B.S., Computer Science

   GPA *3.91*

   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
- **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
- Leading a group of annotators in creating a factuality-driven medical simplification dataset

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

- **Synthetic Non-Factual Text Generation:** This is an ongoing project to develop novel methods to generate diverse and difficult factual errors from a provided text. This is a part of the current work in improving the factuality of text simplification models. Several methods, such as backtranslation, paraphrasing, and named entity substitutions have been attempted. Current development is done using PyTorch, HuggingFace, and other Python libraries.
- **Text Alignment Annotation Interface:** I developed a novel annotation interface for annotating sentence alignments and factuality. The tool allows users to align similar sentences, with assistance, between texts. This annotation interface was used for the greater ongoing project of creating a factuality-driven medical simplification dataset in our research group. The interface was created using HTML, CSS, and Javascript.
- **Named Entity Recognition Model:** I developed and trained a named entity recognition model to identify named entities in tweets. The model architecture consisted of a base BERT transformer, whose hidden states were passed to a linear layer and then a CRF layer. I trained this model on a modestly-sized annotated dataset from Twitter. The model was developed using PyTorch and HuggingFace. The resulting model placed within the top five of the class leaderboard with an F1 score of 0.625.