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 2027
Ph.D. in Computer Science

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

SKILLS

Technical Skills:

- **Highly Proficient:** Python, C, C++, Java
- **Machine Learning**: PyTorch, TensorFlow, HuggingFace, OpenAI API, LLM Prompting
- **Data mining:** SQL, Pandas, Matplotlib, SciPy
- **Web Development:** HTML, CSS, Javascript, JQuery, Django, React, Vue
- **Other Languages & Skills**: Git, Clojure, Go, Rust, CUDA, Bash, MATLAB

PUBLICATIONS

Joseph, S. A., Chen, L., Trienes, J., Göke, H. L., Coers, M., Xu, W., Wallace, B. C., and Li, J. J. (2024). "FactPICO: Factuality Evaluation for Plain Language Summarization of Medical Evidence". In *arXiv preprint arXiv:2402.11456.*

Jan Trienes, **Sebastian Joseph**, Jörg Schlötterer, Christin Seifert, Kyle Lo, Wei Xu, Byron C. Wallace, and Junyi Jessy Li. (2024). "InfoLossQA: Characterizing and Recovering Information Loss in Text Simplification". In *arXiv preprint arXiv:2401.16475***.**

Sebastian Antony Joseph, Kathryn Kazanas, Keziah Reina, Vishnesh J Ramanathan, Wei Xu, Byron C Wallace, & Junyi Jessy Li (2023). Multilingual Simplification of Medical Texts. In *The 2023 Conference on Empirical Methods in Natural Language Processing*.

Shaib, C., Li, M., **Joseph, S.**, Marshall, I., Li, J., & Wallace, B. (2023). Summarizing, Simplifying, and Synthesizing Medical Evidence using GPT-3 (with Varying Success). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 1387–1407). Association for Computational Linguistics.

HONORS & AWARDS

- **2021 ConocoPhillips Innovation Challenge:** 4th place
- **UT Research Assistantship (2022)**
- **UT Professional Development Award (2023)**

EXPERIENCE

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

- Advised by **Dr. Jessy Li** (University of Texas at Austin), **Dr. Byron Wallace** (Northeastern University), and **Dr. Wei Xu** (Georgia Institute of Technology).
- Contributing to an effort to improve access to factual medical information through automatic simplification methods.
- Developed a novel annotation interface as a web application for annotating simplification datasets.
- Led a research initiative in developing a multilingual medical simplification dataset.
- Published research papers in premier conferences for computational linguistics **(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.