

Derek Andersen

📍 United States ✉ andersen.derek@proton.me 📞 812-648-3494 🌐 in/derekcanderson ☁ dechrissen.github.io/

SUMMARY

Experienced NLU Developer with 3 years in enhancing AI-driven voice assistant technologies. Improved model accuracy through precise error analysis and data optimization efforts. Seeking a Computational Linguist role to leverage skills in Python, linguistics, data annotation, and natural language understanding to advance innovative language solutions.

EXPERIENCE

NLU Developer

Cerence

September 2021 – August 2024, Remote

- Developed context-free grammars in the Java Speech Grammar Format (JSGF) for artificial input data synthesis to be used in machine learning (seq2seq) model training.
- Improved the accuracy of AI-based language models used in automobile voice assistant technology by 9%. Performed regular error analysis and bug fixing to correct patterned problems in source data.
- Navigated a Linux environment daily utilizing Bash, Git, regular expressions, and several other tools and techniques on the command line.
- Developed Python scripts for data processing and task automation.
- Led data-cleaning effort which resulted in a 9% increase in accuracy on customer test sets.
- Facilitated the weekly publication of customer deliverables in the form of voice assistant models which are deployed in automobiles from manufacturers such as Audi, BMW, Mercedes, and Toyota. Ensured on-time delivery as the leader of this project, addressing any road-blocks which prevented publication.

Linguist

Defined.ai

September 2020 – March 2021, Remote

- Annotated prosody data in English utterance audio data using the ToBI annotation guidelines to create a dataset for human speech model training.
- Performed spectrogram analysis using Praat and annotated consonants and vowels using the International Phonetic Alphabet.

SKILLS

Python, Linux, Git, Bash, Java Speech Grammar Format, HTML, CSS, JavaScript, Node, SQL, Agile development, Data annotation

PROJECTS

News Headline Toolkit

Personal • github.com/Dechrissen/hstk • March 2024 – June 2024

- Developed toolkit for creating and interfacing with a database of news headlines.
- Utilized Tesseract OCR backend for image-to-text conversion.
- Includes language modeling, text synthesis, data analysis, and data visualization tools.

Blog Templating Engine

Personal • derekandersen.net/blog/new-blog-engine • December 2021 – January 2022

- Developed Node-powered backend engine for my personal blog.
- Utilized Showdown.js to auto-generate HTML blog posts via Markdown conversion.
- Utilized Mustache.js as a templating engine.

Recommendation Engine

Stony Brook University • github.com/Dechrissen/buzzrec • September 2019 – January 2021

- Developed content-based filtering recommendation engine for academic papers in linguistics from the LingBuzz database.
- Utilized similarity matrix based on TF-IDF vectorization values to score and measure similarity between papers.
- Learned about UX best practices when developing user interfaces.

EDUCATION

M.A. Computational Linguistics

Stony Brook University • Stony Brook, NY • 2021 • 3.93 GPA

- Graduated Summa Cum Laude
- Additional coursework in Computer Science & Japanese

B.A. Linguistics

Stony Brook University • Stony Brook, NY • 2018 • 3.98 GPA

- Graduated Summa Cum Laude