Nathan James Anderson

anderson.nathanjames@gmail.com | (704) 604-4165 | www.linkedin.com/in/njanders/

EDUCATION

Carnegie Mellon University, School of Computer Science — Pittsburgh, PA
MS, Language Technologies — August 2022
GPA: 3.90
Coursework: Neural Networks for NLP, Multilingual NLP, Data Science, Discourse Modeling, Machine Learning in Practice, Grammar Formalisms

Brigham Young University — Provo, UT
BA, Linguistics, cum laude — April 2020
Minors: Computer Science, Linguistic Computing

TECHNICAL SKILLS

Programming Languages: Python, Java, Javascript, C++, SQL
Deep Learning Frameworks: PyTorch, TensorFlow, Keras, Dynet
Libraries: NumPy, Pandas, scikit-learn, Matplotlib, Seaborn, Plotly, Hugging Face, spaCy, NLTK, Gensim
Cloud Platforms: Google Cloud/BigQuery, Microsoft Azure, AWS

WORK EXPERIENCE

Brigham Young University — Provo, UT (Remote)
NLP Researcher (Independent Contractor) — October 2022 – *current*

- Trained readability classifier for low-resource languages using PyTorch, achieving 0.85 F1 score across 3 classes
- Leveraged pretrained language models (e.g. RoBERTa) to generate language proficiency exams
- Ran a user-study to correlate student performance on these exams with established proficiency metrics (ACTFL)

Lowe's — Mooresville, NC (Remote)
Data Science Graduate Intern — June 2021 – August 2021

- Built Siamese Network using TensorFlow for a related items recommendation system
- Developed pipeline to rank most relevant product features in over 3,000 catalogs to improve model robustness
- Pitched model which was adopted internally for testing

The Church of Jesus Christ of Latter-day Saints — Riverton, UT
Software Development Intern — July 2019 – December 2019

- Created Speech-to-Speech Interpreter to interpret live speech from English to 38 other languages
- Scraped and aligned >50 hours of audio/video/text data to fine-tune custom Automatic Speech Recognition models, reducing word error rate from 7.2% to 3.3%

PROJECTS

- Hiligaynon-English Neural Machine Translation: Incorporated neural model of Spanish loanwords to improve BLEU score by 5.4 points over NMT baseline.
- Neural OCR Post-Correction for Low-resource Languages: Leveraged multi-input attention and data augmentation to reduce character error rates by 8.5% for Ainu and 16.9% for Griko (low resource languages).

PUBLICATIONS

Anderson, Nathan, Caleb Wilson and Steve Richardson, "Lingua: Addressing Scenarios for Both Real-Time Interpretation and Automatic Dubbing." In *Conference of the Association for Machine Translation in the Americas*. 2022.

Pimentel, Tiago, Nathan Anderson, et al. "A Pipeline for Generating Plausible Wug Words." In *Transactions of the Association for Computational Linguistics*. 2022. (submitted)

Crandall, Jacob W., Nathan Anderson, et al. "Human-swarm Interaction as Shared Control: Achieving Flexible Fault-tolerant Systems." In *International Conference on Engineering Psychology and Cognitive Ergonomics*. 2017.