Alejandro Ciuba

Department of Computer Science
University of Pittsburgh
135 N Bellefield Ave
Pittsburgh, PA 15213

Email: alejandrociuba@pitt.edu
Phone: (412) 874-1838
Homepage: alejandrociuba.github.io
GitHub: github.com/AlejandroCiuba

Education

University of Pittsburgh August 2023 – Present
Ph.D. in Computer Science Pittsburgh, PA
- **Research:** Community-focused natural-language processing
- **Advisors:** Aakash Gautam and Lorraine (Xiang) Li

University of Pittsburgh August 2019 – May 2023
B.S. in Computer Science Pittsburgh, PA
Minors in Linguistics and Spanish GPA: 3.97

Research Experience

Human-Computer Interaction Research: Low-Resource Language Preservation January 2024 – Present
Graduate Student Researcher under Dr. Aakash Gautam Pittsburgh, PA
- Developing a literature analysis of NLP community engagement practices using the PRISMA framework
- Conducting interviews and thematic analysis to better understand how technology can benefit communities

Natural Language Processing Research: Automatic Essay Evaluation May 2024 – Present
Graduate Student Researcher under Dr. Lorraine (Xiang) Li Pittsburgh, PA
- Performing baseline training/fine-tuning on a custom BERT-based model via PyTorch and HuggingFace
- Conducting exploratory data analysis on essay corpora via feature correlation and hypothesis testing

ASR Transcription Research & Technical Support August 2022 – August 2024
Research Assistant under Dr. Dan Villarreal Pittsburgh, PA
- Developed custom ASR pipeline with PyTorch Lightning and `pyannote` to automatically segment audio
- Developed a Python API to work with ELAN files with `pytest` integration

Classroom Discussion Analysis Research January 2023 – April 2023
Research Assistant under Dr. Diane Litman Pittsburgh, PA
- Managed dataset cleaning pipeline for Word Documents, focusing on anonymization and XML tag extraction
- Examined teacher-student classroom discussion differences through data analysis and hypothesis testing

Frederick Jelinek Memorial Summer Workshop June 2022 – August 2022
Undergraduate Researcher under Drs. Anthony Larcher and Santosh Kesiraju Baltimore, MD
- Created a trilingual corpus (English, French and Tamasheq/Tamahaq) for low-resource machine translation
- Ran low-resource model training simulations (e.g. for T5-Small and mBART)

Linguistics Research January 2022 – April 2022
Undergraduate Research Assistant under Drs. Alan Juffs and Na-Rae Han Pittsburgh, PA
- Parsed video game scripts for pragmatics analysis using `spaCy`, `nltk` and `beautifulsoup`
- Aided in the cleaning of the unpublished PELIC Speech Dataset

Work Experience

Teaching Assistant Positions   August 2021 – April 2022
Computational Linguistics and Data Structures & Algorithms   Pittsburgh, PA
- Ran recitations/office hours for students, covering fundamental subject concepts
- Ran a class lecture on machine translation with Jupyter Notebook demo

Carnegie Library ESL Class   August 2019 – March 2020
Volunteer ESL Instructor for the Carnegie Public Library   Pittsburgh, PA
- Guided adult learners through vocabulary and grammar exercises
- Maintained a safe environment for English learners

Publications

INTERSPEECH
Santosh Kesiraju, Marek Sarvaš, Tomáš Pavlíček, Cécile Macaire, and **Alejandro Ciuba**. "Strategies for Improving Low Resource Speech to Text Translation Relying on Pre-trained ASR Models". In: *INTERSPEECH 2023*. ISCA, Aug. 2023. DOI: 10.21437/interspeech.2023-2506

Horizons 2020 Project
Anthony Larcher, Yannick Estève, Mickael Rouvier, … Cecile Macaire, **Alejandro Ciuba**, Peter Vickers, Luis Vicente, Victoria Mingote, and Pablo Gimeno. *Multi-Lingual Speech to Speech Translation for Under-Resourced Languages*. Tech. rep. Le Mans Université, Aug. 2022

Awards

Southern Regional Educational Board Member   August 2024 – Present

K. Leroy Irvis Fellow   August 2023 – Present

CSC Hacks Winner + **Best Beginner Project**   November 2020
Submission: *Flatland: The Game*

Provost Academy Member   August 2019 – May 2023

Skills

Programming & Computation
- C, Python, R, Go, Java
- Git, GitHub and GitLab
- Compute cluster usage via Slurm
- Linux, Docker and bash script creation

User Research Skills
- Quantitative analysis via inferential statistics
- Qualitative analysis via coding and thematic analysis
- User interview and survey creation

Foreign Languages & Linguistics
- Conversational Spanish
- Basic Portuguese Comprehension
- Manual document translation experience