

EDUCATION

University at Buffalo, SUNY	present-06/2025
Master of Sciences in Computational Linguistics GPA: 3.33/4	Buffalo, NY
Coursework: NLP, Information Retrieval, Machine Learning, Computational Linguistics, Phonetics, Syntax I, Semantics II	
Nanjing Normal University	09/2020 - 06/2023
Master of Arts in English Language and Literature GPA: 3.86/4	Nanjing, Jiangsu
China Three Gorges University	09/2016 - 06/2020
Bachelor of Arts in English Language and Literature GPA: 3.59/4	Yichang, Hubei

SKILLS & INTERESTS

- Technical Skills & Tools: Python; Huggingface; JavaScript; Apache Solr; Doccano; MS Office Suite; Praat; WordPress
- Language Skills: Chinese (native); English (fluent); French (basic)

PROFESSIONAL EXPERIENCES

Web Content Assistant	02/2024 - 06/2024
Nursing Clio, University at Buffalo	Buffalo, NY
<ul style="list-style-type: none">• Assist in editing articles, reports, and other written materials for clarity, style, and consistency; maintaining weblinks, updating relevant images with WordPress and ensuring the overall presentation aligns with editorial standards	
Language Engineer Intern	09/2022 - 12/2022
iFLYTEK Shanghai	Shanghai, China
<ul style="list-style-type: none">• Contributed to the construction and development of a TTS corpus for an in-car AI voice assistant; managed the OOV database for lexicon refinement in LLM training; assisted in evaluating and selecting voice assistant speakers.• Collected, categorized and checked speech data for specific foreign languages and dialects; analyzed speech phenomena and set pronunciation rules based on data in different types of corpora• Created and optimized grammar-clear conversation prompts for human-interactive systems in educational and industrial scenes; managed and trained remote interns in human labeling projects	
Teaching Assistant, Course: Selected Readings of English Poetry	03/2021
Department of English, Nanjing Normal University	Nanjing, Jiangsu
<ul style="list-style-type: none">• Instructed over 95 undergraduate students weekly in the course section "From Wordsworth to Browning"; developed and assigned homework aligned with course materials; designed and proctored exams ensuring academic integrity	

PROJECTS

Wikipedia Information Retrieval Chatbot (GCP, Python, Flask, Apache Solar)	08/2024- 12/2024
<ul style="list-style-type: none">• Designed and developed an end-to-end chatbot with document retrieval and chitchat functions; scrapped and indexed 50,000 unique Wikipedia documents on 10 topics with Solar; integrated document retrieval system and chit-chat LLM• Ensured accurate query classification and effective error handling of the chatbot; developed a Flask-based web application with a user-friendly interface for chatbot interaction and chat data visualization	
Annotation on Political Dogwhistles (Doccano)	02/2024- 05/2024
<ul style="list-style-type: none">• Developed, annotated, and analyzed a stereotype-based dogwhistle corpus comprising 60 keywords sourced from news and social media platforms (e.g., Reddit, X); established comprehensive annotation guidelines and labels• Visualized the annotation results, revealing that intentional overt dogwhistles are the most prevalent type, characterized by their deliberate purpose	
Uli Language Phonetic Report with K-means (Pratt, Python)	08/2024- 12/2024
<ul style="list-style-type: none">• Manually transcribed endangered language Uli with IPA based on audio documents, created Uli IPA phonetic chart and lexicon; classified and analyzed vowels based on sound formant features with K-means model	
English TTS Text Normalization Annotation Project	11/2022 - 12/2022
<ul style="list-style-type: none">• Spearheaded the development of text normalization rules for English TTS corpus annotation; led test set data collection and creating comprehensive annotation guidelines for texts with special characters, optimizing data quality for ML applications• Selected and trained a team of 17 annotators to ensure compliance with established data annotation standards, providing daily feedback to supervisors and effectively resolving ambiguities and disagreements during the annotation process	