

Shan Chen

Greater Boston, MA, USA | +1 (507) 210-1913 | shaaan@brandeis.edu | Web: shanchen.dev | [Linkedin](#)

EDUCATION

Brandeis University , Waltham, Massachusetts	Expected: May 2022
Master of Science, Computational Linguistics (STEM Designated, NLP/Data Science/Linguistics)	3.84/4.0
St. Olaf College , Northfield, Minnesota	Sep 2016 - May 2020
Bachelor of Arts, Cum Laude with distinctions, Mathematics and Japanese Majors.	3.7/4.0
<ul style="list-style-type: none">Honors/Awards: Japanese National Honor Society, Pi Mu Epsilon, JASSO Scholarship, National Japanese Exam Silver Prize, Henry Luce Research Grant RecipientSemester abroad: Nagoya University School of Engineering, Nagoya, Japan	

RELEVANT COURSES

Algorithms & Data Structures, Applied Algorithms, Graph Theory, Graph Mining, Neural Computation, Natural Language Annotation for Machine Learning, Object-Oriented Software Design, Information Retrieval, Statistical Machine Learning & NLP, Natural Language Processing, NLP System, Syntactic Theory, Semantics

LANGUAGES AND FRAMEWORKS

Python, PyTorch, Pandas, C++, Linux, R Studio, Java, Flask, Git, Docker, E2, SQL, variety of NLP packages
Chinese (Native), English (Native), Japanese (Professional working proficiency)

WORK EXPERIENCE

NLP Intern , Boston Children's Hospital & Harvard Medical School , Longwood, MA	Jun 2021 - Present
<ul style="list-style-type: none">Clinical NLP adaptation and pipeline developments, training over 500 millions EHR data with HPCBuild system predicting ICU length of stay based on hourly updated EHR and medical device data, with the goal to alleviate future ICU operation pressurePublished with a postdoc on a BioCreative VII track3 Workshop, developed a drug name NER system for automatic extraction of medication names in tweets challenge	
Graduate Research Assistant , Brandeis Xue Lab , Waltham, MA	May 2021 - Present
<ul style="list-style-type: none">Developed systematic way to built millions level Chinese COVID news corpusAdapt Chinese news documents temporal & event relations extraction models within the BBN system	
Data Engineering Intern , Elphi , Cambridge, MA	Dec 2020 - Feb 2021
<ul style="list-style-type: none">Integrated XGBoost model using GCP Pipeline for predicting mortgage closures into Elphi's loan origination platform, to help users make more informed decisions on loan approvals.	
Software Developer , Unismart ・ユニスマ, Tokyo, Japan	Jun 2020 - Sep 2020
<ul style="list-style-type: none">Developed a cross-platform app using Flutter and Firebase framework to connect 10k+ Waseda - University students during the COVID-19 pandemic with an international teamSupported real-time push notification, and course calendar system with Firebase API	
Linguistics Research Assistant , St. Olaf College , Northfield, MN	Jun 2018 - Jul 2019
<ul style="list-style-type: none">Conducted data analysis and synthesized data into a published paper (publication listed below) The Paper was published as a linguistics book chapter in 2019 by Lexington Press as: "The Development of Willingness to Communicate in L2 Chinese Writing"Generated and analyzed Japanese web-sourced linguistics data for Ito, Ph.D.'s linguistics research	

SELECTED PROJECTS

Washington Post Information Retrieval System	Apr 2021 - Jun 2021
<ul style="list-style-type: none">Web app retrieving information from Washington post corpus with boolean and vector space searchFine-tuned FastText and BERT embeddings, adding spell checks and query expansion to improve TREC retrieving results by 20% using NDCG metricsLeveraged frameworks: Flask + Python + Huggingface for front & backend implantations	