

# HAOLIANG CHEN

1700 Summit Avenue, Apt 201, Seattle, WA 98122  
+86-18867532207 ◊ haelc1023@gmail.com ◊ https://haelchan.me

## EDUCATION

---

### Northeastern University

*Expected May 2021*

Master of Science in Computer Science

### Zhejiang University

*September 2015 - July 2019*

Bachelor of Engineering in Measurement Control Technology and Instruments

Overall GPA: 3.56/4.0

The last two years GPA: 3.90/4.0

## EXPERIENCE

---

### Laboratory of Translation and Localization Technologies, School of International Studies, Zhejiang University

*September 2018 - February 2019*

- Wrote an essay *A Corpus-based Study of Differences in Use of Structural Auxiliary Word 'de' in Native Chinese and Translated Chinese*
- Wrote a post *Processing Corpus with Python*

### Nai Ding's Lab, College of Biomedical Engineering and Instrument Science, Zhejiang University

*April 2018 - August 2018*

- Read papers on word embeddings and language models
- Learned linguistics and natural language processing online

## PROJECTS

---

### Language Model Based Punctuation Prediction (Summer, 2019)

- Designed punctuation prediction algorithms based on two pre-trained language models: OpenAI GPT and BERT
- Tested on Europarl v7 and achieved  $F_1$  scores of 0.602 (OpenAI GPT) and 0.732 (BERT)

### Intelligent Dictionary (Summer, 2018)

- Used a 1-million-word corpus and wrote a spelling corrector referring to Peter Norvig's blog
- Implemented simple 1-letter and 2-letter word completion
- Called API provided by Microsoft Azure and Baidu Translate to recognize handwritten letters and translate text into Chinese

### Image Processing Tool (Fall, 2017)

- Implemented some image processing algorithms (grayscale, Gaussian blur, etc.) both from scratch and using MATLAB's Image Processing Toolbox
- Used K-means algorithm learned from Machine Learning to compress images

### Hypertension Management System (Spring, 2017)

- Built GUI with Java Swing
- Designed database using SQL Server 2012

## SKILLS

---

C++, Python, C, Java, MATLAB

PyTorch, spaCy, NLTK

Markdown, L<sup>A</sup>T<sub>E</sub>X, Jupyter Notebook, Git