Personal Statement

For a long time, I have been driven by a belief: how to use the world's information network to improve our lives. People have been mining information in various dimensions and applying them in different fields of artificial intelligence(AI). My research interests lie in natural language processing (NLP), knowledge acquisition, and representation. My experience has given me a great background and determination to pursue further research based on my strong self-motivation. I am applying to USC's Ph.D. in CS to conduct research with the top experts in the field. I entered my postgraduate studies in 2016. During this time, I developed an interest in artificial intelligence as a career, where the beauty of algorithms attracts me. After graduation, I joined an AI company (Mobvoi) in 2019 as a speech algorithm engineer. So far, I have mainly been involved in NLP and text-to-speech(TTS). I will describe some of the related research I have done in the following passages.

Research on Semantic NLP: In Mobvoi, based on cross-language(Mandarin, English, and other dialects), I led the construction and optimization of several modules, including Tokenization, text normalization(TN), polyphonic disambiguation(Polyphone), Prosody, etc. These joint modules convert raw text into features containing pronunciation, pause, stress, and emotion that directly affects TTS intelligibility and naturalness. First, we developed each module with simple models, such as conditional random fields(CRF), finite state transducers(FST), and LSTM. Then, I found that Tokenization, Polyphone, and Prosody are all based on the semantics of the text, and our system is limited. Therefore, I proposed a model architecture called Unified Frontend, which aims to transform complex joint models into end-to-end(E2E) models. We use a pre-trained backbone model to obtain text representation and combine this information into my unified structure. Finally, we constructed the final model by adopting a multi-task scheme. In this process, I verified the maximum capability of each model and designed a reasonable multi-task training scheme. My solution solves the problems of out-of-vocabulary(OOV), ambiguity, and model complexity. During developing this model, I have covered data mining, semantic analysis, pre-training, sequence model, knowledge distillation, multi-task, etc. This work led to a co-author conference submission to *PRML 2022* and five patents.

Research on Emotional NLP: In Mobvoi, I am researching textual factors affecting the naturalness and emotion of TTS, focusing on AudioBook. I divide this work into several stages: First, we introduced textual semantic information based on the pre-trained model, which can effectively improve the naturalness of TTS, such as exclamations and questions. Second, we found that introducing the context of the dialogue is not helpful because irrelevant or non-obvious information is not contributing to the emotion of the sentence. Then, we adopted the concept of style and established a corpus criterion by referring to the predecessor's experience and communicating with professional broadcasters. With a small amount of data and a semi-supervised way, a good style extraction model is obtained and used as a guide for TTS. Finally, we propose a progressive granularity stress model. Specifically, using a two-stage pipeline to construct the coarse-to-fine stress model with auxiliary granularity supervision, which decreases the model deviation. The result is remarkable, and TTS gets better expressiveness. Our research has attracted the attention of users and other industries. In this work, I have covered linguistics learning, contrastive learning, pre-training, knowledge distillation, NLU(dialog analysis), etc.

The work led to a first-author conference submission to *ICASSP 2023(reviewing)* and two patents.

Research on Knowledge Tracking: In TAL AI Lab, I built a deep knowledge tracing(DKT) model to monitor student knowledge states or skill acquisition levels based on historical practice records. We considered the inherent side relations of the questions more than the sequential dependencies. Firstly, we build a weight graph of questions based on semantic and statistical knowledge. Then, we use the graph embedding (LINE/node2vec) method to represent the question. In addition, we create a Laplacian matrix of the question to capture the correlation. Our optimal model has a 4%-5% improvement over the base model. This method demonstrates the effectiveness of side information. It proposes an idea to extract sequence memory based on graph embeddings, which can be easily extended to similar issues. We validate it on the MovieLens corpus. We also applied this method to predicting the answering situation and planning an adaptive learning path, which is helpful in intelligent tutoring systems. The above research was completed by a Ph.D. and me and led to a co-author conference submission to *AIED 2019*.

About My General Life: During my undergraduate years, I learned the basics of computer science and mathematics, which gave me a strong research background. In the last semester of graduate school, I actively participated in academic activities organized by enterprises and universities. I was engaged in other research areas, including computer vision, recommendation systems, data mining, and knowledge graph. All of these experiences strengthened my understanding of machine learning. I also met many great people, enriching my life experience. During the past few years of working, I have worked with several partners and enhanced my problem-solving ability. I have mentored five interns, where I developed strong self-motivation and some management skills. Over the years, I developed a good coding style and documentation skills. Most importantly, my original belief remains unchanged. I aspire to make breakthrough contributions to my interested field.

Future work: I believe USC can help me with my aspirations because of the strength of its faculty and students. Specifically, I would be excited to work with professor Muhao Chen, and I have learned about his recent work, including the publications in *EMNLP* and *NAACL*. I am interested in knowledge acquisition, robust representation, and utilization of information. I think processing the world's information is a mysterious and fascinating subject that I never limit myself to exploring. As I have explored relevant fields, I have the confidence to work with professors and partners to solve practical problems and make further contributions in the future.