Personal Statement | **Xinru Wang**

For as long as I remember, I was the only girl in my Go class. Slightly introverted, I only had as my best friends my two boxes of elegant, black and white Go pieces made of agate and purple jade. Six years of training and tournaments made me a 4-dan at 10 and No.1 in the women's Go game in Hangzhou city, one step away from entering the 5-dan rank and becoming a professional chess player. Representing the Youth Group of Jiangxi Province Go Team because of my aggressive style and formidable calculation skills, I had been enjoying all kinds of games and contests. But everything changed when AlphaGo easily defeated Lee Sedol and Ke Jie, both former world No. 1s, in 2016 and 2017 respectively. AlphaGo Zero, the updated version of AlphaGo, defeated its predecessor after training by playing against itself for ONLY three days, with no prior knowledge whatsoever. That was why I chose to study computer science in college, which seemed a far more exciting game. Since graduation, I have been working in Meituan, one of the leading e-commerce platforms in China, for over a year as a software developer. However, the enormous amount of data and incredible operational and managerial complexity I encountered daily at work convince me of the necessity of more rigorous graduate-level training to better harvest AI power.

I am confident in my preparedness to embark on rigorous graduate training given my extensive research and professional experience, such as the Summer School for AI at Cambridge and the Data Science Camp of Alibaba Cloud. More importantly, I secured a research opportunity at Columbia University's Cathaypath Institute of Science in May 2022. Our goal was to conduct a sentiment analysis of Amazon Kindle Store's classified product reviews. It was a typical natural language processing (NLP) task, involving tokenization, lemmatization, features vectorization with TF-IDF, and emotion scoring based on fuzzy string matching and NRCLexicon dictionary. To unlock NLP's full potential, we incorporated a time series analysis to understand how long it took reviews' sentiment to affect sales results. Then we employed grid search for optimal lag-length while training a Decision Tree, a Support Vector Machine, and a Naïve Bayes classifier to predict the category of a sale change (i.e., great increase, medium increase, no increase, medium decrease, great decrease). With our models achieving prediction accuracies of 72%, 68%, and 69%, respectively, Dr. Patrick Houlihan, our supervisor, encouraged us to submit our work, with myself as the first author. I am pleased to mention that the paper has been published[1].

Despite enjoying my work with machine learning, I felt the need for industrial experience. That was why I took the job offer from Meituan's Daojia Business Group (DBG), which operates the core business of Meituan, including online ordering and delivering food, pharmaceuticals, and other retail goods. Summarizing and analyzing massive data from more than 50 million online orders and 10 million delivery riders on a daily basis, DBG heavily depends on back-office applications, such as data management, business process management, operational analysis, etc. Thousands of such applications supported the daily operation of tens of thousands of teams, groups, and divisions from all levels of geolocations. Noticing the workflow similarity of these applications, we created a low-code, cloud-based development platform named WOLF for back-office application development and deployment to lower the workload for both front-end and back-end developers. My contribution was to transform the DSL-based configuration method into a graphic one, with which any business idea could materialize into workflows well-defined by a UI element tree. To further refine its composition, I created the event configurator for rapid selection of proper event attributes from a multitude of options and the event link for upstream and downstream

[1] **Wang, Xinru** & Jin, Xinran & Wu, Yongqiang & Liao, Yajing & Lei, Dongchen. (2023). Predicting product sales based on sentiment analysis using reviews: The case of Amazon's kindle. *Applied and Computational Engineering*. 6. 1641-1651. 10.54254/2755-2721/6/20230784.

events to easily compose or branch events. Other than that, I devised an expression configurator, a style configurator, and a manager for adapters and interceptors; I invented the overall interaction logic of the continuous integration and delivery pipeline. By the end of 2022, WOLF had helped 1319 developers in 200+ teams and hosted 1064 projects. Because of my performance, I was promoted to L6 within one year, compared to an average of three years for others.

While working on WOLF, I found that all levels of business operators and managers basically fed on data. To monitor the execution and results of just one specific operation, business analysts would have to import raw data from 10 to 20 data sources, spend 0.5 to 4 hours processing them with formulas and expressions in Excel, and waste another 30 minutes manually updating the table daily. To address these concerns, DBG launched a front-end development platform, Diting (諦聽), named after Kshitigarbha's divine beast capable of listening to a person's heart. Based on WOLF, the first version of Diting enabled operators and managers to create flexible, customized data report applications. However, with the expansion of the business, more data sources, more complicated business logic, and more refined operational granularities necessitated a major update. Accordingly, I employed a new multi-tenant architecture and enforced the "Leave No Trace Principle" for incremental refactorization, progressively decoupling business logic from data and hiding the complex data structures. Other improvements included route separation for different business lines, singleton patterns and adaptors for business logic, a state manager to centralize data source storage, and encapsulation of data access into functions for data independence of different business lines. Immediately, Diting decreased the average cost of developing a front-end application for new business lines from 7 persons per day (P/D) to 0.5 P/D and reduced the cost of a data-report application from 5 P/D to 0 P/D. As of July 2023, we saw a remarkable 943% growth in the number of regional business collaborators and an 186% increase in the number of business developers actively using Diting.

With the advent of ChatGPT and Copilot, I am impressed by their performance in responding to specific commands but disappointed by their stark insufficiency in abstraction, logical reasoning, and project architecting. However, their ambitions are way bigger than my WOLF low-code platform. While I aimed at low-code, they aspire for "no-code." As a seasoned developer and a mature architect, I share the same vision. Therefore, I must receive graduate-level training to dive deeper. To that end, the Master of Science in Engineering in Computer and Information Science (CIS/MSE) program at UPenn's Department of CIS attracts me for many reasons. I am first attracted by the rigorous curriculum systematically covering programming, logic, and AI. In particular, Advanced Programming (CIS5520), Programming Paradigms (CIS5540), and Advanced Topics in Programming Languages will guide me to the very foundation of modern programming. Topic in Logic (CIS5180) and Friendly Logics (CIS6820) will bestow me with an in-depth understanding of descriptive and computational complexity. And Artificial Intelligence (CIS5210) and Deep Learning for Data Science (CIS5220) will take me to the forefront of industry trends. Moreover, I admire CIS' excellent research on machine learning and AI and the abundant research opportunities from its numerous centers and institutes. I especially look forward to joining the ASSET Center for AI-Enabled Systems; its pioneering research on Algorithmic Reasoning in Large Language Models and Faithful Chain-of-Thought Reasoning excites me profoundly.

Upon completing my training at UPenn, I will seek full-time employment at a leading-edge AI-focused company like OpenAI or Google DeepMind. Committed to the vision of a world where programmers don't need to code, I am open to any challenges.