In retrospect, I was directed to dive into computer science from my data-centered professional experiences. Although majoring in computer science from the early start of my academic career, my vision was to become a data analyst who utilizes data to enhance sales for a company. Deeply involved in data-oriented academic research and internships, I got a great sense of achievement, which caught me in the illusion that I was born for data. However, after witnessing a web platform collapse due to the data burst, I perceived the indispensableness of robust distributed systems for companies. During that brief period of passive pause, I grasped a clear picture of the role I'm dedicated to, a frontier computer system engineer and scientist.

My interest in computer systems originated from a massive distributed system crash during my second internship. When Baidu acquired YY, the network platform I worked on collapsed due to the internal segmented cloud services and unprocessed big data. At that moment, I realized how data analysts' work was significantly correlated with and influenced by the distributed systems. Regardless of individual capability, all the data analysts can do during a system pause is wait for system engineers to optimize and fix the system at the lower level. Following the passive technical stuck, my professional pursuit was directed to a computer systems engineer and scientist, where I could understand and transform the basis of all data in greater depth.

Guided by my passion and ambition towards the computer system field, I took the initiative to practice the theories in academic studies. Getting the privilege of working with Prof. Yang Wang and Prof. Mike Bond, my operating systems course instructors at the Ohio State University, I tested the synchronization correctness of the open-source distributed system Hadoop. The project aimed to investigate, in real-world applications, which configuration parameters cannot be set on run time in a heterogeneous manner. Over the six months, I developed an automatic online testing pipeline that reconfigured the Hadoop cluster without manual interference and found more than 30 parameters related to synchronization correctness within the datanode. While modifying the distributed system, I got exposed to the automated testing procedure constructions and potential system performance improvement approaches, such as network, storage, etc.

Counting the research journey, I was introduced to Prof. Haryadi S. Gunawi at the University of Chicago for exploring tail latency reduction in storage devices, which complies with my data analysis skills and my interest in computer systems. Throughout the study, I recognized this unconventional imbalance learning problem and pointed out the importance of the false negative rate metric. By building an automated training pipeline based on the previous research insights, I introduced a new model with a continual learning concept that improved the efficacy of the false negative rate by more than 20% and accelerated the response time by three times.

With all these formative work and research experiences, I found my genuine interest lay in operating systems related, covering distributed systems, high-performance computing, and storage device research. Motivated and dedicated, I look forward to exploring the breadth and width of the computer systems fields through systematic postgraduate training. After completing the master's degree, I plan to become a computer scientist at a research institution or university, innovating high-performance system platforms on enormous data scales.

Prospectively, I wish to materialize my vision of accelerating data responsiveness and computational speed at the system level, benefiting all the data-oriented work and the information industries.