

Statement of Purpose

Kayo Yin (kayoyin.github.io)

When my mother first started using a smartphone, she only used it to call. She wouldn't use other functionalities, such as looking up directions, until I taught her to orally ask the virtual assistant what she needs. This led me to believe Natural Language Processing will bridge the gap between humans and modern technology. My goal is to make Artificial Intelligence benefit as many people as possible. Language plays a central role in our lives, developing models to process all these textual data is a challenge I am determined to tackle. I believe advancements in NLP will be a game changer in how data can be exploited, how communication can be enhanced, and if computers understand our language, using technology would be as simple as maintaining a conversation.

To achieve this ambition, I have worked hard to gain experience undertaking research in AI. During my summer internship at Illuin Technology, a French company that develops solutions using AI with a focus on industrializing NLP, I researched unsupervised question answering. After becoming familiar with language modeling and self-attention, I built on recent research where I used cloze translation to synthesize question answering data. I also learned to code BiDAF and BERT models in PyTorch and perform cloud computing. I created my own question answering pipeline for XLNet and my unsupervised approach even obtained performance surpassing the scores reported in the paper "Unsupervised Question Answering by Cloze Translation".

My internship ended with many promising leads in my research, so during my exchange in Taiwan, I continued working with the company to extend my work to French. I added new methods for question generation based on French syntax and developed novel unsupervised approaches using neural machine translation that, unlike existing works, can be applied to languages without available language analysis tools. I also implemented CamemBERT, a French pre-trained RoBERTa model, for question answering, contributed the French Reading Comprehension (FReaC) dataset, and reported my research in a paper to publish. I hope to extend my work to low-resource languages or specific domains such as medical question answering.

I believe my remarkable performance is founded on my academic abilities and determination. By pursuing a demanding double major in mathematics and computer science, I completed advanced coursework in programming and theory of computation, as well as in all major areas of mathematics. Prior to discovering computer science, I frequently competed in math competitions with exceptional results. My rich background in mathematics gives me a deep understanding of machine learning theory and integrate them in my research: I use my knowledge in linear algebra to manipulate matrices and kernels, calculus and optimization to derive gradient descent and object function convexity, probability and statistics to build algorithms based on information theory and Bayesian statistics.

My determination makes me highly productive. I learned to manage my time by maintaining a 4.0 GPA every year while participating in over a dozen student associations, I even ranked 1st in machine learning and all four of my math courses last semester. With minimal experience in AI, I secured my NLP internship after attending 4 networking events, distributing 50 resumes and having 10 interviews. As soon as my internship started, I efficiently applied course knowledge to a computer vision project where I used an ensemble of image processing and CNN for document cleaning. I also experimented with semi-supervised CycleGAN to overcome the challenge of limited data, and I finished the task within two weeks instead of the estimated two months. During my exchange semester, I ambitiously chose graduate level courses in machine learning where foundations acquired in my undergraduate studies allowed me to win the first places in class challenges, such as image classification on limited data where I used transfer learning, snapshot ensembles and anti-aliasing.

During my coursework, I also practiced leadership and project management. In my first programming project, I lead a team of 8 to build a LSTM neural network in C++ for music composition. I had no machine learning experience at the start of the project, and in 4 months, I assimilated the technical skills and machine learning notions needed to build and train the neural network. I also learned to design a machine learning project, from data retrieval and embedding to user interface for the model demo, and distribute these tasks within a team efficiently. The results of the project exceeded all expectations and each member received A.

In addition, I believe the transmission of knowledge is as important as the research behind it, which is why I practice public speaking at debate competitions and my ability to break down complex topics by tutoring mathematics and teaching Japanese, French and English. I first started posting articles about AI during my internship for the company blog, which received great success. This motivated me to continue publishing articles as a writer for Towards Data Science. I even earned the title “Top writer in Artificial Intelligence” on Medium and recently, NVIDIA researchers invited me to deploy my code from my article on RAPIDS AI. During my graduate studies, I want to gain more teaching experience through teaching assistantships.

My goal in language technologies is to create solutions that are generalizable, usable and beneficial. I aim to employ multilingual models and unsupervised learning to develop approaches generalizable to low-resource languages and specific domains. I want to improve how machines understand our language to make technology usable by anyone. I am determined to leverage NLP to identify injustice in social contexts like employment and law enforcement, and address bias towards marginalized populations. With this goal in mind, my next large-scale project is my Bachelor’s thesis on neural machine translation of sign language glosses to natural language.

I have demonstrated my ability to apply course knowledge in various topics of machine learning, such as data processing and computer vision, to academic and industrial projects, as well as the capacity to quickly comprehend and incorporate new notions of NLP in my research. To undertake longer-term research with contributions on par with my ambition, I need more research experience and a wider and deeper understanding of language technologies.

I strongly believe that [dream university] is the best suited to help me realize my dream. The CS department has the most comprehensive course offering I find essential for my research goals such as spoken language processing and social issues in NLP, and for my related interests like AI in healthcare and symbolic musical information. The faculty is also composed of scientists whose interests greatly align with my own, whose works I look up to and inspire my research.

I am particularly determined to join the [university NLP lab] and research NLP for social good under the guidance of [really cool Prof #1], multilingual language analysis tools and reading comprehension with [really cool Prof #2]. I also want to incorporate [really cool Prof #3]’s work in [really cool paper] to my research on unsupervised question answering for adversarial training of robust unsupervised QA systems.

I am confident in my academic abilities to follow your demanding courses and in my skillful creativity to contribute innovative insights to your research. I look forward to getting involved in your vibrant student community with my unique culture and passion for social causes, orchestra and soccer. I am well prepared and determined to complete my [dream Master program] at [dream university] to strengthen my competence for a PhD and a research career.
