Summary of my projects

My first research experience was an exploratory study on the text-2-code project. I conducted an in-depth evaluation of the Alpaca model's performance on Python programming tasks to uncover the model's limitations and potential paths for improvement in code generation capabilities of LLMs scenarios. Due to the lack of a suitable existing dataset, I systematically gathered a diverse set of programming problems and their corresponding solutions from Python textbook exercises and various beginner-friendly programming websites by employing advanced web scraping techniques. By automating the data collection process, I was able to efficiently aggregate a large volume of high-quality examples, ensuring diversity in problem types, difficulty levels, and solution approaches. After systematically analyzing and categorizing errors generated by the model's code, I fine-tuned the model to improve its accuracy by 17% on Python programming problems. From delving into transformer architectures and practical applications of PyTorch to independently designing and implementing research ideas, this project expanded my technical vision in NLP and machine learning and cultivated my capacity for autonomous scientific inquiry. Also, it taught me that the true essence of research is not about pursuing perfection, but about remaining open and resilient in the process of constant exploration.

This NLP project aroused my keen interest in the vast prospects of artificial intelligence, and the research I conducted on the maximum clique-based alignment algorithm design gave me the opportunity to explore another important area of AI research: computer vision (CV). In this project, we investigated the innovative potential of maximum clique algorithms in optimizing point cloud registration (PCR) by studying a non-deep learning method from the best paper candidate at CVPR 2023. I was responsible for two key stages: maximum clique reconstruction and graph structure optimization. I applied the Otsu thresholding method to determine an optimal cutoff value that distinguishes between significant high-frequency features and background noise. This adaptive thresholding ensures that only the most relevant feature points are retained. Points with wavelet coefficients below the identified threshold were systematically removed, effectively isolating the high-frequency feature points that are crucial for accurate point cloud matching and alignment. Our experiments showed a 40% reduction in computation time without losing alignment accuracy and robustness. This project not only gave me insights into algorithm optimization, but also allowed me to experience the collaborative research for the first time. By leading the technical solution design and advancement, I learned how to clearly articulate innovative ideas, effectively guide discussions, and transform technical ideas into specific research paths in a team.

As my interest in multimodal fields intensified, I applied to join the Soft Computing Lab at Yonsei University and engaged in a project on enhancing neural symbolic visual question answering through logical reasoning. As the second author of the paper, I led the experimental design and implementation, validating that logic-based model interactions effectively address modality reasoning challenges. We developed a novel framework that constructs reasoning premises from the image captions generated by vision-language models (VLMs) and derives conclusions utilizing the internal commonsense knowledge of LLMs. Logical rules are established through connectives, and the scores obtained via fuzzy logic operations are subsequently provided to the LLM for answering questions. After using a penalty mechanism to reduce dependence on linguistic information to address the hallucination issue of VLM when answering absurd questions—where the questions are entirely unrelated to the image, yet the model still generates premises relevant to the question—and to resolve the threshold problem in fuzzy logic by my proposed optimizations solution, the framework achieved state-of-the-art performance in common VQA datasets. This research marked my first deep involvement in the complete laboratory paper publication process, teaching me publication standards. It also introduced me to neuro-symbolic AI, a frontier area in multimodal learning. These experiences inspired my current NLP research direction, where I have already validated a novel research concept through preliminary experiments and plan to pursue further investigations.