Hyunwoo Yoo

CONTACT INFORMATION	Email: hty23@drexel.edu Phone: (+01) 215 252 2342

EDUCATION

Drexel University
Graduate Student in Electrical and Computer Engineering

Philadelphia, United States
September 2023 - Current

Sogang University
M.S., Computer Science and Engineering,

Seoul, South Korea
September 2017 - August 2019

Sogang University
B.S., Computer Science and Engineering

Seoul, South Korea
March 2013 - August 2017

Monash University
Exchange Student, Information Technology

Victoria, Australia
February 2016 - June 2016

RESEARCH INTEREST

Natural Language Processing, Bioinformatics, Natural Language Understanding, Large Language Models, Machine Learning

EXPERIENCE

Naver Boostcamp AI Tech
Natural Language Processing Track

Remote
March 2022 – August 2023

- Developed movie recommendation and movie review analysis system based on ColBERT and elastic search

Hyperlounge
Business Platform team

Seoul, South Korea
October 2022 - February 2023

- Developed automatic business data crawling and summarization system.

Tmax Soft Group
ERP/CRM Research team

Gyeonggi, South Korea
July 2019 - October 2022

- Developed algorithms and tools for embeddings and linking on advertising documents.
- Improved a chatbot and an OCR system by exploiting embeddings and named entity tags to augment training data.
- Designed and optimized tools and APIs for processing and correcting income statements and expense claims.

PUBLICATIONS

H. Yoo. Adversarial Examples for DNA Classification, arXiv preprint, 2024.

H. Yoo. Fine Tuning Text-to-Image Diffusion Models for Correcting Anomalous Images, arXiv preprint, 2024.

H. Yoo, M. S. Refahi, R. Polikar, B. Sokhansanj, J. R. Brown, and G. Rosen. iSeqsSearch: Incremental Protein Search for iBlast/iMMSeqs2/iDiamond, bioRxiv preprint, 2024.

M Refahi, BA Sokhansanj, JC Mell, J Brown, **H Yoo**, G Hearne, G Rosen. Scorpio: Enhancing Embeddings to Improve Downstream Analysis of DNA sequences, bioRxiv preprint, 2024

H. Yoo, B. Sokhansanj, J. R. Brown, and G. Rosen. Predicting Anti-microbial Resistance using Large Language Models, arXiv preprint, 2024.

H. Yoo, M. Kang, and K. Oh. A Semantic Search Model Using Word Embedding, POS Tagging, and Named Entity Recognition, In proc. of the International Conference on Computational Science and Computational Intelligence, 2018.

H. Yoo, S. Kim, and K. Oh. Ontology-Based Semantic Search System for Movie Object, In proc. of the 20th International Conference on Artificial Intelligence, 2018.

H. Yoo, D. Kim, and T. Kim. Measurement Value Model and System based on Kinect Sensor for Sitting Position Calibration, In proc. of the Korean Information Processing Society Conference, 2017.

TEACHING EXPERIENCE	Teaching Assistant, Drexel University -ECE303 Lab(ECE303) -Senior Design(ECE491)	Fall 2023 - Fall 2024
	Teaching Assistant, Sogang University -Introduction to Artificial Intelligence(CSE4185) -Computational Thinking(COR1011) -Capstone Design II(CSE4187)	September 2017 - August 2019
AWARDS	Patrick Dewar PhD Fellowship Award Drexel University	February 2024
	Best Wikipedia Award Hackathon, Sogang University Prototyped Healthy Computer - The Spine Health Care System	January 2017
TECHNICAL SKILLS	Programming Languages: Python, C, C++, Java, SQL, and UNIX shell scripting Toolkits: PyTorch, TensorFlow, Django, Flask, FastAPI, and Spring	

Hyunwoo Yoo

I have been deeply fascinated by the process of quantifying human language and have explored various applications of natural language processing (NLP). In particular, my academic and professional journey has been driven by research on computational models that help understand and generate complex human language. This journey has led me to deeply dive into large language models (LLMs), genomic models, and various fine-tuning techniques.

After earning a bachelor's degree in computer science, I pursued a master's degree focused on embedding-based semantic search in NLP applications. Subsequently, I gained practical experience at TmaxSoft Group, applying NLP technologies in the accounting domain, and at Hyperlounge, where I conducted research integrating NLP with management information systems (MIS). These experiences not only deepened my understanding of practical NLP applications but also ignited my passion for exploring the theoretical and technical foundations of language models.

During the Naver Boostcamp program, I explored state of the art LLMs such as ChatGPT and Naver Clova, solving real world challenges like enhancing search systems and data augmentation. This was a turning point in my NLP research, revealing the transformative potential of LLMs in addressing diverse computational problems.

In my current graduate program in the United States, I am focusing on cutting-edge research in LLMs and diffusion models, tackling technical challenges such as memory constraints during large-scale training. Specifically, I have worked on integrating LoRA-based fine-tuning techniques with DNABERT2 and Nucleotide Transformer to achieve significant performance improvements. I have also conducted instruction tuning for domain-specific models like BioGPT, pioneering an innovative approach that merges text-based models like BioBERT with DNA-based models like Nucleotide Transformer. In this process, I implemented genomic sequence analysis by leveraging sequence algorithms and databases such as BLAST and MMseqs2, enabling the utilization of diverse nucleotide sequence information.

In addition, I have researched text generation and transformation using diffusion models. Recently, I addressed the issue of generating bizarre images with specific prompts by fine-tuning diffusion

Hyunwoo Yoo

models using DreamBooth. This exploration of text generation and style transformation highlighted the synergies between NLP and diffusion models, further solidifying my research direction toward extending LLM and generative AI technologies into broader applications.

My extensive experience in applying NLP technologies to various domains includes developing prediction and classification models based on nucleotide and amino acid sequences in textual formats, designing embedding-based search systems for mapping accounting and management terminology, and fine-tuning diffusion models with DreamBooth to prevent abnormal image generation from specific prompts. These experiences have allowed me to practically explore the possibilities of NLP applications and generative AI technologies.

The work conducted by Professor McKeown's research team to integrate diffusion models with NLP has left a profound impression on me. The potential of diffusion models in text generation and style transfer in NLP resonates with the multimodal data applications and generative AI opportunities I have explored in my diffusion model research. This combination of fundamental NLP research and novel techniques aligns closely with my research goals.

Professor McKeown's research on text generation, summarization, and sentiment analysis provides a strong foundation for tackling fundamental problems in NLP and expanding into diverse application areas. Particularly inspiring is the team's work leveraging large language models (LLMs) to deeply analyze the context and meaning of text data. I also believe that my domain-specific expertise and technical skills can contribute to the fusion of generative AI technologies, including diffusion models, with NLP.

At Columbia University, I aspire to deep dive into fundamental NLP research while proposing new research directions based on my extensive domain experience and technical skills. I look forward to joining your team to address core NLP challenges and bridge the gap between fundamental research and practical applications.