Sehyun CHOI

Address: Modni At Home 101-401, Baekokdaero 2332-31, Cheoin-gu, Yongin-si, Gyeonggi-do, South Korea
Email: choisehyun98@gmail.com | **Telephone**: +82 01088194520 | **Github**: https://github.com/syncdoth

EDUCATION

The Hong Kong University of Science and Technology (HKUST)	**Hong Kong**
Master of Philosophy (MPhil) in Computer Science	09/2022 – 06/2024 *(Exp.)*
Bachelor of Engineering (BEng) in Computer Science, Minor in Bioengineering	09/2017 – 06/2022

- **GGA**: 4.01/4.30 (First Class Honours; Dean's List & Academic Achievement Medal Awardee)
- **Relevant Coursework**: Machine Learning for NLP, Deep Learning in Computer Vision, Big Data Management, Honours Software Engineering, Honours Design and Analysis of Algorithms, Honours Object Oriented Programming & Data Structures, Linear Algebra

Handong International School	**South Korea**
Advanced Placement (AP) by College Board & ACT	02/2011 – 02/2016

- **Subject**: Physics I (*5/5*), Physics II (*5/5*), Calculus BC (*5/5*), Statistics (*5/5*) | ACT (*34/36*)

PROFESSIONAL & RESEARCH EXPERIENCE

Naver Corporation	**South Korea**
Machine Learning Research Intern	12/2021 – 02/2022

- Experimented with various Deep Active Learning approaches of querying the most important data points from the unlabelled pool of datasets in Computer Vision and Natural Language Understanding fields to Machine Translation (MT) tasks
- Improved the efficiency of fine-tuning process by selecting and training on the most informative data examples, utilizing corpus statistics, pre-trained MT Transformer model's output uncertainty, and the model's encoder representation of sentences
- Identified important aspects of data selection processes in Active Learning such as uncertainty, representativeness & diversity and devised metrics to quantify them in effective manner

KAIST Artificial Intelligence (AI) Lab	**South Korea**
Summer Research Intern	07/2021 – 08/2021

- Supported the language model detoxification project by researching the literatures of eXplainable AI to detect potentially offensive concepts captured in the language models and ablate them
- Proposed a novel framework for detecting the activation of offensive concepts in language models from non-offensive sentences

HKUST Undergraduate Research Opportunity Program (UROP)	**Hong Kong**
Student Researcher (https://github.com/HKUST-KnowComp/CSKB-Population)	02/2021 – 12/2021

- Developed novel methods using Graph Neural Networks (GNN) and pre-trained language models (PTLMs) applied on Commonsense Knowledge Graphs (CKG) and achieved a significant improvement of performance over the previous baselines in CKG Population task
- Performed hyperparameter optimization and ablation study over network components such as different PTLMs or GNN design to find the best performing model and attribute the improvement to each component quantitatively

Skelter Labs	**South Korea**
Machine Learning Engineer Intern	12/2019 – 07/2020

- Conducted experiments to improve the quality of deep learning based speech synthesis models, such as introducing a new module into network structures, employing data augmentations, and fine-tuning with different datasets
- Utilized data pipelines, microservices and cloud services such as Spark, Kubernetes, and GCP to perform large scale training jobs
- Participated in collaborative software development processes using Git and code review systems

PUBLICATION

Benchmarking Commonsense Knowledge Base Population with an Effective Evaluation Dataset
EMNLP 2021 Main Conference (3rd Author) | **Doi**: *http://dx.doi.org/10.18653/v1/2021.emnlp-main.705*

AWARD & CERTIFICATE

Naver Clova AI Rush 2021	**South Korea**
First Runner-Up in Smart Grammar Editor Task	05/2021 – 07/2021

- Improved the benchmark GLEU score by 33% over the vanilla Transformer baseline in Korean Grammatical Error Correction (GEC) tasks by implementing pseudo-data creation with back-translation, two-stage pre-training, and data-mixing techniques in fine-tuning

EY Next Wave Data Science Competition	**Hong Kong**
Country Finalist - Hong Kong (https://github.com/syncdoth/EY_DataWave_Challenge)	04/2019 – 05/2019

- Performed feature engineering and cleaning of raw data using data handling packages and developed a Recurrent Neural Network Model (LSTM) using eras to predict the behavior of city travelers at a specific time window

EXTRACURRICULAR ACTIVITY

HKUST KSA Machine Learning Study Club	**Hong Kong**
Founding Member & President (https://github.com/syncdoth/ML_STUDY_2020)	10/2020 – 12/2021

- Organized a machine learning study club of more than 20 students and currently teaching with a self-implemented curriculum, covering Python language, linear algebra, deep learning in tensorflow API, and real-world application projects using state-of-the-art models

J.P. Morgan – Code for Good 2020	**Hong Kong**
Participant	10/2020 – 10/2020

- Designed a NoSQL schema for Firestore database and developed React front-end components that visualizes the submissions of each student within the administratie website for centralized communication and progress reporting

SKILL & INTEREST

Language: English (Native) | English (Fluent)
Technology: Python | PyTorch | Tensorflow | Transformers | git | shell scripting & Linux | OOP
Interest: NLP | XAI | Knowledge Graph | Reasoning Ability of AI | AI Privacy | Music | Photography | Neuroscience