YUNSOO KIM

(+44)75-8731-6332 | Email : yunsoo.kim.23@ucl.ac.uk | Website | LinkedIn

EDUCATION

University College London, *London, UK*
PhD in Health Informatics — Sep. 2023 - Present
Thesis: "Transformer in Health Informatics: Multimodal Model"
Supervisors: Dr Honghan Wu, Dr Adam Levine, Professor Robert Stewart, Dr Nikolas Pontikos

Imperial College London, *London, UK*
M.Sc. in Bioinformatics and Theoretical Systems Biology — Oct. 2015 - Sep. 2016
Thesis: "A Clustering-based Short Time Series Analysis Methodology in Omics"

Case Western Reserve University, *Cleveland, OH*
B.S. with Honors in Systems Biology (Concentration in Bioinformatics, Pre-Health Plan) — Sep. 2012 - May. 2015
Honors Thesis: "Assembly of Marama chloroplast from Next-gen sequencing reads"

RESEARCH INTEREST

Deep Learning in computational biology and health informatics, specifically in language model and multimodal model using transformer models.

EXPERIENCE

UCL, London, United Kingdom

KnowLab, PhD Student — Sep. 2023 - Present

LG CHEM, Seoul, Korea

CTO Platform Technology Artificial Intelligence PJT — Apr. 2020 – Sep. 2023
Professional. NLP Group Leader. Work Description: Leading natural language processing (NLP) group at LG Chem — Mar. 2022 – Sep. 2023
Specialist. NLP Group Leader. Work Description: Leading natural language processing (NLP) group at LG Chem — Apr. 2020 – Feb. 2022

- Working on language model in chemistry, materials science, and synthetic biology patents and literature to accelerate research
  - Pre-training language models (BERT, RoBERTa, deBERTa) on patents and literature (using custom vocabulary)
  - Fine-tuning a multitask model for tasks such as domain-specific entity recognition, summarization, and document classification.
- Working on chemical knowledge graphs construction from the extracted entities (nodes) and relations (edges)
  - Link prediction for unknown edges in knowledge graphs using RGCN.
- Developing an artificial reading system to recommend research candidates for researchers
  - Mongo DB of documents collected from multiple sources in chemistry, materials science, and synthetic biology fields.
  - Internal web search system empowered by the multitask model
  - Successful cases on different research topics (battery, petrochemical, biopolymer, etc).
- Leading and supervising NLP group members
  - Organizing weekly meetings for checking group members' progress and giving them feedback regarding their action steps.
  - Merging pull requests in internal Gitlab, and Improving skills of group members through group seminars.

LG AI Master Program Pilot at LG AI Research — Sep. 2019 – Jan. 2020
LG Chem representative researcher. Program Description: Solving the challenges using AI and becoming a leader in AI project

- Improved language model in biotechnology, especially in Green Biology (agricultural)
  - Used BERT for pre-training on Green Biology literature and fine-tuning of 9 bionlp benchmark tasks
  - Reached SOTA in 7 bionlp benchmark tasks (Task types: Named Entity Recognition, Relation Extraction, Question Answering)
- Developed a web service for the NER analysis of biology text with user input.

CTO Green Bio GBT1 PJT at LG Chem — Mar. 2019 – Mar. 2020
Specialist. Work Description: Developing deep learning models to accelerate biotechnology research

- Developed plant disease recognition and plant disease severity diagnosis deep learning models (ResNext, EfficientNet ensemble)
  - Plant disease recognition model for seed sales department as a customer service tool
  - Plant disease severity diagnosis for seed R&D department to accelerate breeding disease-resistant seed

CTO Corporate R&D Future Technology GBT PJT at LG Chem — Jul. 2017 – Feb. 2019
Associate. Work Description: Analyzing -omics data to accelerate biotechnology research

- Developed an early allergen and toxin assessment system for the candidate gene
  - Built in-house allergen and toxin database for the sequence alignment
  - Developed a neural network model for the prediction of protein allergenicity and toxicity with improved performance
- Developed a DNA marker for beta-Carotene Chinese cabbage to be used in marker-assisted selection
- Developed a web service for the analysis of genome editing results using targeted sequencing data

Alternative military service, Korea

Alternative military service at LG Chem, Military Manpower Administration Aug. 2017 – Aug. 2020

Technical Research Personnel. Description: Alternative military service for the Korean compulsory military service

- A company designated by the commissioner of the Military Manpower Administration can utilize a research personnel (a master's degree holder or higher) for the advancement of science and technology in South Korea
- Was hired as both a full-time employee and technical research personnel for LG Chem

Torrey Pines Co., Ltd., Cheonan, Korea

Co-Founder. Work Description: Giving advice regarding microbiome science in cosmetics May. 2015 – Present

CEO. Work Description: Planning a New Brand and Leading Export Oct. 2016 – Jul. 2017

- Planned *Bota nouveau* Microbiome science hair care cosmetics brand
  - One of the pioneers in hair care cosmetics that introduced microbiome science in the field (main product is hair loss care shampoo)
- Exported to Vietnam and Cambodia
  - Led meetings with buyers in many countries such as Hong Kong, China, Taiwan, Vietnam, and Cambodia, and led export process

Imperial College London, London, UK

Computational Bioinformatics Laboratory Jun. 2016 – Oct. 2016

Research Description: Developing methodology to analyze omics time-series data

- Adopted existing R methodology in python to analyze differential metabolomics time series using smooth splines
- Tested the methodology with varying parameters and scenarios regarding the limitations of short time series and on *Schistosoma mansoni* metabolomics cohort data
- Extended the methodology to include network inference features

Theoretical Systems Biology Laboratory Jan. 2016 – Jun. 2016

Research Description: Developing a novel methodology to analyze metagenomic abundance data

- Developed a novel methodology to infer OTU (Operational Taxonomic Unit) co-occurrence and differential networks
- Benchmarked 8 other inference methods for co-occurrence networks by our novel method (the best precision)
- Tested on multiple EBI datasets including diabetes and chron's disease
- Developed a web service for the analysis to be used in a wider community

Dankook University, Cheonan, Korea

Bioinformatics Laboratory Jul. 2015 – Oct. 2015

Visiting Researcher Research Description: Analyzing –omics data using machine learning

- Identified genetic markers of MERS-COV using Marmoset data by detecting alternative splicing events on experimental data
- Integrated semi-supervised learning of RNA-seq and ChiP-seq data to cluster regions of a chromosome such as exons

Case Western Reserve University, Cleveland, OH, US

Undergraduate Bioinformatics Research Jan. 2014 – May. 2015

Research Assistant Research Description: Identifying the Marama chloroplast genome

- Designed a workflow for the assembly of Marama chloroplast genome
- Conducted *de novo* assembly of contigs, alignment of contigs to different chloroplast genomes, and reassembly of the mapped contigs
- Presented at SOURCE Summer 2014 Intersections, Case Western Reserve University
- Presented at Michelson-Morley Undergraduate Research Competition

New York University Medical Center, New York, NY, US

WISE Program at Rusk Institute Feb. 2012 – Jun. 2012

Research Assistant Research Description: Improving rehabilitation for stroke patients

- Worked on eye-hand coordination improving methods for stroke patients rehabilitation process
- Analyzed statistical data from an experiment on the effect of music during the rehabilitation process for stroke patients
- Presented on experience and research project in a public forum

PUBLICATION (INCLUDING PATENTS)

Chemical Language Understanding Benchmark

First Author Conference: ACL2023 Industry Track

Accepted in 2023

Chemical entity name recognition system and method using artificial intelligence

Main Inventor Patents Priority : KR 10-2023-007009

Application filed in 2023; Publication in progress

Method and system for recognizing chemical entities through sentence length adjustment of the training data

Main Inventor Patents Priority : KR 10-2023-0070096

Application filed in 2023; Publication in progress

Document classification method and system using named entity recognition score
Main Inventor — Patents Priority : KR 10-2023-006571
Application filed in 2022; Publication in progress

Method to build chemical language model and nature language processing document analysis system and method applying it
Main Inventor — Patents: KR20230037335A
Application in 2021; Publication in 2023

Chemical Named Entity Recognition Method and System through Data Preprocessing
Main Inventor — Patents: KR20230030337A
Application in 2021; Published in 2023

Crop protection products virtual library construction and browsing system
Co-Inventor — Patents: KR20230029084A
Application in 2021; Published in 2023

Protein Toxicity Prediction System and Method Using Artificial Neural Network
Main Inventor — Patents: KR20200126715A
Application in 2019; Published in 2020

A novel inversion in the chloroplast genome of marama (Tylosema *esculentum*)
First Author — Journal: Journal of Experimental Botany
Accepted in 2016

HONORS AND AWARDS

LG Chem On-Spot Incentive	2022
LG Chem DX Good Practice	2021
LG Chem AIT Organization Manager Discretionary Award	2021
LG AI IDEAthon Digital Transformation Frontier (Applications 1st place; Overall 2nd place)	2021
LG Chem CTO Digital Transformation Best Practice	2021
LG Chem Seoul National University AI Hackathon 1st place	2021
LG Chem DX Best Practice	2020
LG AI Advanced problem solving Best Practice mentor	2020
LG Chem Platform Technology On-Spot Incentive	2020
LG Chem CEO Excellent R&D Project Award	2020
LG Chem AIT Organization Manager Discretionary Award	2020
LG Chem Platform Technology Best Memorandum (Research Report)	2020
LG AWARDS (Best Award given by LG Group Chairman)	2020
LG AI Master Best Practice	2019
LG Chem Green Biology Best Memorandum (Research Report)	2019
LG Chem Green Biology On-Spot Incentive	2019
Case Western Reserve University Scholarship	2012 – 2015
Ogelbay Summer Research Grant	2014
Case Western Reserve University Dean's Honor List, 3 semesters	2012 – 2014
Case Western Reserve University Dean's High Honor List, 1 semester	2013
Victor Ridder Scholarship	2012

OTHER EXPERIENCE PRESENTATIONS

Invited Talks

University of Edinburgh Clinical NLP Group Talk on Large Language Model Research in Medicine	2023
ACL2023 Industry Track Poster Presentation	2023
LG Chem CTO Digital Transformer talks	2022
LG Chem Platform Technology Workshop talks	2022
LG Chem Platform Technology Artificial Intelligence annual meeting speaker	2021
LG Chem Tech Fair talks	2021
LG DX Fair 2021 speaker	2021
LG Chem Digital Transformation talks	2021
LG Chem CTO Digital Transformer talks speaker	2021
LG Chem Platform Technology Workshop talks speaker	2021
LG Chem Digital Transformation Best Practice talks	2020
LG Academy Digital Transformation talks panelist	2020
LG Housys AI Master talks speaker	2019
LG AI Master talks speaker	2019
LG Chem Green Biology quarterly meeting speaker	2019

Teaching Experience, Workshops, Seminars

LG Chem Community of Practice for ChatGPT 2023
LG Chem Community of Practice for Systems Metabolic Engineering AI 2022
LG Chem transformer model study group lead organizer 2021 – 2023
LG AI Advanced problem solving grader, mentor 2020
LG Chem Platform Technology Python Instructor 2020
LG AI weekly study group member, organizer 2019 – 2022
LG Chem Bio Big Data and AI Community of Technology member, organizer 2017 – 2019
Case Western Reserve University Volunteer Peer Tutoring in Biology 2013 – 2014

- Taught two students for biology classes at university-level