(Danny) To Eun Kim

Information Retrieval | NLP | Dialogue Systems | Knowledge Graphs | Search and QA

✉ to.kim.17@ucl.ac.uk 🎓 Google Scholar 🌐 https://kimdanny.github.io/

Education	**University College London (UCL)** Sep 2018 - Aug 2022

University College London (UCL)   Sep 2018 - Aug 2022
M.Eng. Computer Science (4-year integrated master's degree)   **First-Class Honours**
Minor: Modern Applications of Engineering Mathematics
Thesis: Multi-Task Neural User Simulator for Task Oriented Dialogue System

Ongoing Work

Response ranking with conversational look-ahead (Targeting SIGIR '23 - 1st author)
User satisfaction estimation with Hawkes Transformer (Targeting ACL '23 - 2nd author)
Interactive Grounded Language Understanding (IGLU) - NeurIPS 2022 challenge

Publications

A Multi-Task Based Neural Model to Simulate Users in Goal-Oriented Dialogue Systems
T.E. Kim, A. Lipani [paper] [poster] [code]
45th ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 2022

Condita: A State Machine Like Architecture for Multi-Modal Task Bots
J. Ramos*, T.E. Kim*, Z. Shi, X. Fu, F. Ye, Y. Feng, A. Lipani [paper]
Alexa Prize TaskBot Challenge Proceedings, 2022

Attention-based Ingredient Phrase Parser
Z. Shi, P. Ni, M. Wang, T.E. Kim, A. Lipani [paper] [code]
European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN), 2022

Under Review

Exploring the potential of automating the process of clustering smell stories
T.E. Kim, S. Mitroi, R. Kumar, N. Ranade, K. Selvaratnam, L. Xu, E. Maggioni, M. Obrist
International Journal of Human-Computer Studies (IJHCS), 2023

Workshops & Presentations

Recent Advances in Applications of Knowledge Graphs. To be presented at the 5th Meet-up of Knowledge Graphs Interest Group at Alan Turing Institute, 27th Oct., 2022.

A Multi-Task Based Neural Model to Simulate Users in Goal-Oriented Dialogue Systems. Presented at Search-Oriented Conversational AI Workshop, 2022 (SCAI '22). [recording] [slides]

Research Experience

UCL Web Intelligence Group - Graduate Research Assistant   May 2021 - Present
Conducting and assisting research with Prof. Emine Yilmaz and Dr. Aldo Lipani .

- Highly contributed (first authored) to the *Alexa Prize TaskBot Challenge*, taking charge of engineering/DevOps challenges, such as AWS services integration, deployment of ranking/search/QA models, and making effective user feedback loops.
- Published three papers in top academic venues, and extending the research using novel ideas.
- Leading the *Viral Content Ranking and Generation* project. Aiming to construct a neural learning-to-rank model that ranks marketing articles in terms of future virality in social media.
- Participating in a NeurIPS 2022 challenge - Interactive Grounded Language Understanding in a Collaborative Environment (IGLU) with PhD students.
- Join in regular reading group.

Amazon Alexa Prize TaskBot Challenge - Team Condita   May 2021 - Mar 2022
Co-led the team to the Quarter-final stage.

- Engineering a Task-oriented Dialogue System by using Alexa Skills Kit, Evi, DynamoDB, Lambda, CodePipeline, and SageMaker. Constructed and maintained CI/CD pipelines, deployed multiple micro-services with Flask and Docker, dealing with thousands of user conversations (details are addressed in the Master's thesis).
- Designed and implemented a novel state machine-like architecture for dialogue policy decision.
- Research were focused on Information Retrieval, Knowledge Graph, Question-Answering, Ranking, and Dialogue State Tracking.

UCL (Human-Computer) Interaction Centre - bachelor's group thesis   Oct 2020 - Apr 2021
Project *Olfactory-NLP* with group memebers and Prof. Marianna Obrist .
- Extracted smell-related phrases and clustered multiple passages into different olfactory groups.
- Literature Review: *An NLP Approach to Classify Smell Experiences.* [paper]
- *Exploring the potential of automating the process of clustering smell stories.* [paper] [code]
- Edited version is recently submitted and under review in the International Journal of Human-Computer Studies.

Industry Experience	**Vector AI** Jun 2022 - Jul 2022

Vector AI   Jun 2022 - Jul 2022
Machine Learning Engineer Intern   London, UK
- Cross-functional intern in both Core-ML and Forward-deployed engineering teams.
- Tabular information extraction from bill of ladings and business registration documents based on OCR results for freight forwarding automation.
- Use of modern dev tools, e.g., Gitlab, Jira, Sentry, Docker, Kubernetes, GCP.

KB Securities   Jun 2020 - Aug 2020
Data Analyst Intern   Seoul, South Korea
- Rule-based Investment Funds Recommender System based on Network Analysis, Hierarchical Clustering and Association Rule Mining.
- Stock Price Prediction System using LSTM Network and Natural Language News Data. [code]
- Piloted a Social Media Monitoring System using KB's NLP Engine and Automation (RPA) software. Real-time sentiment monitoring in the comment sections.
- Heavy use of SQL in the Hadoop ecosystem (Map-reduce, Hive, and Hue).

Royal College of General Practitioners (RCGP)   Oct 2019 - Mar 2020
Software Engineer - UCL Industrial Exchange Network   London, UK
- Built a tree-based task-oriented chatbot that assists in checking software compliance regimes of British healthcare organizations.
- Built a RESTful API for user intents and entities detection using Dialogflow and React, Express Node, and MongoDB.
- Python scripts for chatbot integration testing. [code]

Teaching Experience

UCL COMP0189 Applied Artificial Intelligence, *Teaching Assistant*   (Spring 2023)
UCL COMP0071 Software Engineering, *Teaching Assistant*   (Spring 2023)
UCL CEGE0004 Machine Learning for Data Science, *Teaching Assistant*   (Spring 2023)
UCL CEGE0096 Geospatial Programming, *Teaching Assistant*   Fall 2022
UCL Artificial Intelligence Society, *Lecturer*   2020 - 2022
- Founder, maintainer, and lecturer of Machine Learning Tutorial Series
- Season 20/21: [GitHub] [Lecture Recordings] || Season 21/22: [GitHub]

Awards

2022 Automated Knowledge Graph Construction (AKBC) Conference – Registration fee waiver
2022 SIGIR Student Travel Grant – Registration fee waiver
2021 UCL Student Union best societies award – *Centenary Award (UCL AI Society)*
2021 UCL CS Undergraduate Summer Research Internship Fund – £3,200

Community Involvement

Automated Knowledge Graph Construction (AKBC) Conference, *volunteer*   Nov 2022
Alan Turing Institute Knowledge Graphs Interest Group, *member*   Sep 2022 - Present
CIKM 2022, *Sub-reviewer*   July 2022
UCL Artificial Intelligence Society, *Head of Tutorials Team*   Sep 2019 - Jun 2022
NLP paper reading group (Jiphyeonjeon), *Season 2,3 Organizer*   May 2021 - Nov 2022
- Season 2: Survey on Task-Oriented Dialog System and Knowledge Graph construction. [slides] [recording] (spoken in Korean)
- Season 3: Recent Advances and Challenges in Task-oriented Dialogue Systems. [slides] [recording] (spoken in Korean)

UCL CS Transition Mentor, *Mentor - helping 1st year CS students*   Sep 2020 - Jun 2022
UCL Korean Society, *IT Director - webmail automation*   Jun 2020 - May 2021

Selected Coursework & Projects	**COMP0087 Statistical Natural Language Processing** Taught by Prof. Sebastian Riedel, received high mark (76.62%). • *Fake reviews generations via large language models.* [paper]	Spring 2022

COMP0087 Statistical Natural Language Processing — Spring 2022

Taught by Prof. Sebastian Riedel, received high mark (76.62%).

- *Fake reviews generations via large language models.* [paper]

Unified API of Apache Kafka & Google Pub/Sub — Nov 2021

- Simple ML model is served and does inferences in an event-driven way. [code]

COMP0022 Database and Information Management Systems — Spring 2021

Best group coursework nominated - Movie Analytics Web App.

- Microservice architecture consisting of a NestJS REST API, PostgreSQL DB, and Redis cache. Designed a normalized ERD for DB and optimized queries, seeing more than 10% speed increase. APIs can pre-cache anticipated requests, enabling extremely low latency infinite scrolling. [code]

BENG0095 Data Mining and Analysis — Winter 2020

Best group coursework nominated, received high mark (79.88%).

- Data visualization, model training and selection, hyperparameter tuning, and evaluation. [paper] [code]

Building Conversational AI for Messenger Hackathon — Dec 2019

Hackathon at Facebook London

- Built a Facebook's messenger-bot that checks if user's reply to one's lover is okay to send or not by fine-tuning a BERT model for sentiment analysis. [code]

Propositional Semantic Tableau Validator — Dec 2019

- Parse propositional logic, build a semantic tableau, and check its logical satisfiability (implemented in C) [code]

Professional Training

Google Developers Korea Machine Learning Bootcamp	Oct 2020 - Jan 2021
Deep Learning Specialization (Prof. Andrew Ng on Coursera)	Dec 2020
UCLQ Summer School in Quantum Technology	Sep 2020
Introduction to AI in the Data Center (NVIDIA)	Mar 2020
KAIST Artificial Intelligence Bootcamp [Project 1] [Project 2]	Jun 2019 - Aug 2019

Major Skills

Programming:	Python, Java, C, LaTeX
ML/DL:	PyTorch, TensorFlow/Keras, Scikit-Learn, OpenCV
Data skill:	SQL, Hadoop, Numpy/Pandas/Matplotlib
SW Dev:	Apache Kafka, Node, Express, React, Flask, MongoDB, PostgreSQL
Platforms:	Linux, Git, Docker, AWS, GCP

References

Prof. Emine Yilmaz
Professor and Turing Fellow at UCL, Dept. of Computer Science, **Email:** emine.yilmaz@ucl.ac.uk

Dr. Aldo Lipani
Assistant Professor in Machine Learning at UCL, **Email:** aldo.lipani@ucl.ac.uk

Prof. Marianna Obrist
Professor of Multisensory Interfaces at UCL, Dept. of Computer Science, **Email:** m.obrist@ucl.ac.uk