

(Danny) To Eun Kim

Information Retrieval | NLP | Dialogue Systems | Retrieval-Enhanced ML | Search and QA

✉ toeunkim@cmu.edu

🎓 [Google Scholar](#)

🐙 <https://kimdanny.github.io/>

Education	Carnegie Mellon University (CMU) Aug 2023 - (2028) Ph.D. Language and Information Technology at Language Technologies Institute (LTI).
	University College London (UCL) Sep 2018 - Sep 2022 M.Eng. Computer Science (4-year master's integrated undergrad. degree) First-Class Honours Minor: Modern Applications of Engineering Mathematics Thesis: Multi-Task Neural User Simulator for Task Oriented Dialogue System
Conference Publications	A Multi-Task Based Neural Model to Simulate Users in Goal-Oriented Dialogue Systems T.E. Kim , A. Lipani [paper] [poster] [code] <i>45th ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)</i> , 2022 Conditia: 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] <i>Alexa Prize TaskBot Challenge Proceedings</i> , 2022 Attention-based Ingredient Phrase Parser Z. Shi, P. Ni, M. Wang, T.E. Kim , A. Lipani [paper] [code] <i>European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)</i> , 2022
Preprints	When and What to Ask Through World States and Text Instructions: IGLU NLP Challenge Solution Z. Shi*, J. Ramos*, T.E. Kim , X. Wang, H. Rahmani, A. Lipani [paper] <i>arXiv</i> , 2023
Research & Industry Experience	UCL Web Intelligence Group - Research & Teaching Assistant May 2021 - Present Research with Prof. Emine Yilmaz and Dr. Aldo Lipani at UCL Centre for AI <ul style="list-style-type: none">• Highly contributed (first authored) to the <i>Alexa Prize TaskBot Challenge</i>.• Published papers in top academic venues, and extending the research using novel ideas.• Leading the <i>Viral Content Ranking and Generation</i> project. Aiming to construct a neural learning-to-rank model that ranks marketing articles in terms of future virality in social media.• NeurIPS 2022 IGLU challenge – Research Prize \$650.• Unified model for different schema in task-oriented dialogue systems (Targeting ACL '23) Raft AI Jun 2022 - Jul 2022 Machine Learning Engineer Intern London, UK <ul style="list-style-type: none">• Cross-functional intern in both Core-ML and Forward-deployed engineering teams.• Tabular information extraction from bill of lading 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. Amazon Alexa Prize TaskBot Challenge - Team Conditia Co-lead May 2021 - Mar 2022 Co-led the team to the Quarter-Final stage. <ul style="list-style-type: none">• 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 with feedback loop. (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.

	<p>UCL (Human-Computer) Interaction Centre - bachelor's group thesis Oct 2020 - Apr 2021</p> <p>Project <i>Olfactory-NLP</i> with group members and Prof. Marianna Obrist .</p> <ul style="list-style-type: none"> • Extracted smell-related phrases and clustered multiple passages into different olfactory groups. • Literature Review: <i>An NLP Approach to Classify Smell Experiences</i>. [paper] • <i>Exploring the potential of automating the process of clustering smell stories</i>. [paper] [code]
	<p>KB Securities Jun 2020 - Aug 2020</p> <p>Data Analyst Intern Seoul, South Korea</p> <ul style="list-style-type: none"> • 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).
	<p>Royal College of General Practitioners (RCGP) Oct 2019 - Mar 2020</p> <p>Software Engineer - UCL Industrial Exchange Network London, UK</p> <ul style="list-style-type: none"> • 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	<p>UCL Intense Python Programming course, Course Co-Organiser and Lecturer (March 2023)</p> <ul style="list-style-type: none"> • Designing and will lead a new intensive programming course for MSc dissertations. <p>UCL COMP0189 Applied Artificial Intelligence, <i>Teaching Assistant</i> (Term2 22/23)</p> <ul style="list-style-type: none"> • Making weekly practical contents and coursework [GitHub] <p>UCL COMP0071 Software Engineering, <i>Teaching Assistant</i> (Term2 22/23)</p> <p>UCL CEGE0004 Machine Learning for Data Science, <i>Teaching Assistant</i> (Term2 22/23)</p> <p>UCL CEGE0096 Geospatial Programming, <i>Teaching Assistant</i> (Term1 22/23)</p> <p>UCL Artificial Intelligence Society, <i>Lecturer</i> 2020 - 2022</p> <ul style="list-style-type: none"> • Founder, maintainer, and lecturer of Machine Learning Tutorial Series • Season 20/21: [GitHub] [Lecture Recordings] Season 21/22: [GitHub]
Workshops & Presentations	<p>NeurIPS '22 competition: <i>IGLU: Interactive Grounded Language Understanding in a Collaborative Environment</i>, Research Prize solution presentation, 6th Dec., 2022</p> <p>NLP Interest Group at The Alan Turing Institute, <i>Can Dialogue Systems Have Theory-of-Mind Ability?</i>, 24th Nov., 2022. [recording][slides]</p> <p>Knowledge Graphs Interest Group at The Alan Turing Institute (5th meet-up), <i>Personal Knowledge Graphs: Towards Better Conversational Assistants</i>, 27th Oct., 2022. [recording] [slides]</p> <p>Search-Oriented Conversational AI Workshop (SCAI '22), <i>A Multi-Task Based Neural Model to Simulate Users in Goal-Oriented Dialogue Systems</i>, 15th Jul., 2022. [recording] [slides]</p>
Awards	<p>2022 NeurIPS Interactive Grounded Language Understanding (IGLU) – Research Prize – \$650</p> <p>2022 Automated Knowledge Graph Construction (AKBC) Conference – Registration fee waiver</p> <p>2022 SIGIR Student Travel Grant – Registration fee waiver</p> <p>2021 UCL Student Union best societies award – <i>Centenary Award (UCL AI Society)</i></p> <p>2021 UCL CS Undergraduate Summer Research Internship Fund – £3,200</p>
Community Involvement	<p>The Alan Turing Institute NLP Interest Group, <i>member</i> Nov 2022 - Present</p> <p>The Alan Turing Institute Knowledge Graphs Interest Group, <i>member</i> Sep 2022 - Present</p> <p>Automated Knowledge Graph Construction (AKBC) Conference, <i>volunteer</i> Nov 2022</p> <p>ICLR 2023 (<i>Reviewer</i>), CIKM 2022 (<i>Sub-reviewer</i>) July, Oct 2022</p> <p>UCL Artificial Intelligence Society, <i>Head of Tutorials Team</i> Sep 2019 - Jun 2022</p> <ul style="list-style-type: none"> • Co-organised an international competition, <i>Climate Hack.AI</i> [GitHub] [media coverage]

	<p>NLP paper reading group (Jiphyeonjeon), <i>Season 2,3 Organizer</i> May 2021 - Nov 2022</p> <ul style="list-style-type: none"> 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) <p>UCL CS Transition Mentor, <i>Mentor - helping 1st year CS students</i> Sep 2020 - Jun 2022</p> <p>Google Developer Student Club, <i>UCL core team - event with Marc Cohen</i> Sep 2021 - Jun 2022</p> <p>UCL Korean Society, <i>IT Director - webmail automation</i> Jun 2020 - May 2021</p>
Selected Coursework & Projects	<p>COMP0087 Statistical Natural Language Processing Term2 21/22</p> <p>Taught by Prof. Sebastian Riedel, received high mark (76.62%).</p> <ul style="list-style-type: none"> <i>Fake reviews generations via large language models.</i> [paper] [Presentation Recording] <p>Unified API of Apache Kafka & Google Pub/Sub Nov 2021</p> <ul style="list-style-type: none"> Simple ML model is served and does inferences in an event-driven way. [code] <p>COMP0022 Database and Information Management Systems Term2 20/21</p> <p>Best group coursework nominated - Movie Analytics Web App.</p> <ul style="list-style-type: none"> 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] <p>BENG0095 Data Mining and Analysis Term1 20/21</p> <p>Best group coursework nominated, received high mark (79.88%).</p> <ul style="list-style-type: none"> Data visualization, model training and selection, hyperparameter tuning, and evaluation. [paper] [code] <p>Building Conversational AI for Messenger Hackathon Dec 2019</p> <p>Hackathon at Facebook London</p> <ul style="list-style-type: none"> 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] <p>Propositional Semantic Tableau Validator Dec 2019</p> <ul style="list-style-type: none"> Parse propositional logic, build a semantic tableau, and check its logical satisfiability (implemented in C) [code]
Professional Training	<p>Turing Commons: AI Ethics and Governance (at The Alan Turing Institute) [website] Nov 2022</p> <p>Google Developers Korea Machine Learning Bootcamp Oct 2020 - Jan 2021</p> <p>Deep Learning Specialization (Prof. Andrew Ng on Coursera) Dec 2020</p> <p>UCLQ Summer School in Quantum Technology Sep 2020</p> <p>Introduction to AI in the Data Center (NVIDIA) Mar 2020</p> <p>KAIST Artificial Intelligence Bootcamp [Project 1] [Project 2] Jun 2019 - Aug 2019</p>
Major Skills	<p>Programming: Python, Java, C, L^AT_EX</p> <p>ML/DL: PyTorch, TensorFlow/Keras, Scikit-Learn, OpenCV</p> <p>Data skill: SQL, Hadoop, Numpy/Pandas/Matplotlib</p> <p>SW Dev: Apache Kafka, Node, Express, React, Flask, MongoDB, PostgreSQL</p> <p>Platforms: Linux, Git, Docker, AWS, Alexa Skills Kit, GCP</p> <p>Languages: English (professional; IELTS overall 8), Korean (native)</p>
References	<p>Prof. Emine Yilmaz Professor and Turing Fellow of Computer Science in UCL, Email: emine.yilmaz@ucl.ac.uk</p> <p>Dr. Aldo Lipani Assistant Professor of Machine Learning in UCL, Email: aldo.lipani@ucl.ac.uk</p> <p>Prof. Marianna Obrist Professor of Multisensory Interfaces and Computer Science in UCL, Email: m.obrist@ucl.ac.uk</p>