

# Linqing Liu

90 High Holborn, London WC1V 6LJ, UK

+44-07878510487 | likicode@gmail.com



## EDUCATION

---

### University College London

Ph.D. in *Computer Science*

London, UK

Sept 2020 - Present

- Supervisors: Pontus Stenetorp, Sebastian Riedel
- Area: Natural Language Processing

### University of Waterloo

MMath in *Computer Science (thesis-based)*

Waterloo, Canada

Sept 2018 - Jul 2020

- Supervisor: Jimmy Lin
- Related Courses: *Theory of Deep Learning, Data-Intensive Distributed Computing, Optimization for Data Science, IoT & Intelligent Connectivity*

### Tongji University

B.Eng. in *Software Engineering*

Shanghai, China

Sept 2013 - Jun 2017

- Related Courses: *Data Mining & Analysis, Data Warehouse Technology, Object-Oriented Programming, Data Structures, Operating System, Compiler Principle, Computer Architecture, Software Engineering*

## PUBLICATIONS

---

(\* Equal Contribution)

1. **Query Expansion Using Contextual Clue Sampling with Language Models**  
Linqing Liu, Minghan Li, Jimmy Lin, Sebastian Riedel, Pontus Stenetorp. Arxiv Preprint 2022
2. **When Do Flat Minima Optimizers Work?**  
Jean Kaddour\*, Linqing Liu\*, Ricardo Silva, Matt J. Kusner. **NeurIPS 2022**
3. **Challenges in Generalization in Open Domain Question Answering**  
Linqing Liu, Patrick Lewis, Sebastian Riedel, Pontus Stenetorp. **NAACL Findings 2022**
4. **PAQ: 65 Million Probably-Asked Questions and What You Can Do With Them**  
Patrick Lewis, Yuxiang Wu, Linqing Liu, Pasquale Minervini, Heinrich Kuttler, Aleksandra Piktus, Pontus Stenetorp, Sebastian Riedel. **TACL 2021**
5. **Controllable Abstractive Dialogue Summarization with Sketch Supervision**  
Chien-Sheng Wu\*, Linqing Liu\*, Wenhao Liu, Pontus Stenetorp, Caiming Xiong. **ACL-IJCNLP Findings 2021**
6. **MKD: a Multi-Task Knowledge Distillation Approach for Pretrained Language Models** Linqing Liu, Huan Wang, Jimmy Lin, Richard Socher and Caiming Xiong. Arxiv Preprint 2020
7. **Incorporating Contextual and Syntactic Structures Improves Semantic Similarity Modeling**  
Linqing Liu, Wei Yang, Jinfeng Rao, Raphael Tang, Jimmy Lin. **EMNLP 2019**
8. **Bridging the Gap between Relevance Matching and Semantic Matching with Hierarchical Co-Attention Network**  
Jinfeng Rao, Linqing Liu, Yi Tay, Wei Yang, Peng Shi, Jimmy Lin. **EMNLP 2019**
9. **Distilling Task-Specific Knowledge from BERT into Simple Neural Networks**  
Raphael Tang\*, Yao Lu\*, Linqing Liu\*, Lili Mou, Olga Vechtomova, Jimmy Lin. Arxiv preprint 2019
10. **Generative Adversarial Network for Abstractive Text Summarization**  
Linqing Liu, Yao Lu, Min Yang, Qiang Qu, Jia Zhu. **AAAI abstract 2018**

## HONOURS & AWARDS

---

- 20' **EfficientQA NeurIPS competition**, won 2 tracks with the UCL-FAIR team
- 18' **David R. Cheriton Graduate Scholarship**, University of Waterloo, based on academic excellence
- 18' **AAAI Student Scholarship**, AAAI
- 16' **Google Anita Borg Memorial Scholarship: Asia-Pacific**, Google Inc (only 9 recipients in China)

## WORK EXPERIENCE

---

### DeepMind

Research Scientist Intern

**London, UK**

*Jun 2022 - Oct 2022*

- Supervisors: Arthur Mensch, Igor Babuschkin, Laurent Sifre
- *Retrieval augmented language model*: Implementing the language model architecture to scale with trillions of tokens retrieval database, and improving the model on the downstream knowledge-intensive tasks such as question answering.

### Salesforce Research

Research Intern

**Palo Alto, USA**

*Sept 2019 - Aug 2020*

- Supervisor: Caiming Xiong
- *Controllable Abstractive Dialogue Summarization*: a two-stage generation strategy that first generates a preliminary summary sketch identifying the interaction between speakers and salient information in each turn, then controls the granularity of the final summary by predicting amount of information contained in the source dialogue.
- *Multi-Task Knowledge Distillation Approach for Pretrained Language Models*: a general knowledge distillation framework under multi-task learning setting that transfers knowledge from large PLMs to small student model. The approach is evaluated on two different student model architectures and shows advantage in both performance and inference speed.

## PROJECTS

---

### Efficient Open-Domain Question Answering

*Sept 2020 - Dec 2020*

### EfficientQA Competition

- ODQA is primarily tackled through retrieve-and-read systems, which is computationally expensive and slow. We instead generate and index a large corpus of question-answer pairs in advance. To operate on our KBs of QA-pairs, we further develop a retriever model that can be optimised for memory, speed or accuracy. With additional system compression techniques, we won 2 tracks: *best accuracy while system size under 500MiB* and *smallest system to get 25% accuracy*.

### Text-to-Image Information Retrieval from Large-Scale Web Archives

*Nov 2018 - Jan 2019*

### JCDL demo

- The first integration of deep learning models with a toolkit for exploring web archives (Archives Unleashed Toolkit) to support content-based image analysis at scale. By broadcasting the pretrained model to all Spark executors in parallel, we are able to extract images of common objects from a 4TB web archive of GeoCities, which we then compile into browsable collages.

## TECHNICAL STRENGTH

---

**Proficiency:** Python (PyTorch, NumPy, spaCy, HuggingFace Transformers), L<sup>A</sup>T<sub>E</sub>X, Shell, SLURM.

**Prior Experience:** Jax, Tensorflow, Javascript, Scala, MapReduce, Spark, HTML/CSS, R, Cocos2d.