

# Yuqing Xie

PH.D. STUDENT IN COMPUTER SCIENCE (NATURAL LANGUAGE PROCESSING)

✉ [yuqing.xie@uwaterloo.ca](mailto:yuqing.xie@uwaterloo.ca) | 🏠 [amyxie361.github.io](https://github.com/amyxie361) | 📧 [amyxie361](https://twitter.com/amyxie361) | 🌐 [yuqing-xie-8bb213170](https://www.linkedin.com/in/yuqing-xie-8bb213170) | 🎓 [Google Scholar](https://scholar.google.com/citations?user=...)

## Education

### Cheriton School of Computer Science, University of Waterloo

Waterloo, ON, Canada

Ph.D. Student in Computer Science

Sept 2018 - Present

- GPA:91.62/100
- Supervisor: [Prof. Ming Li](#) and [Prof. Jimmy Lin](#).
- Research interest: NLP applications, including **Question Answering**, **Information Retrieval**, and **Model Robustness**.

### School of Mathematical Sciences, Fudan University

Shanghai, China

B.S. in Mathematics and Applied Mathematics

Sept 2014 - June 2018

- GPA:3.6/4.0 Major 3.5/4.0(Overall)
- In the Honor Class of the National Basic Subject Top-notch Talent.
- Thesis: Biological Question Answering System based on BiDAF (Supervisor: [Yiming Wei](#) and [Shanfeng Zhu](#))

## Publication

### Applying Structural and Dense Semantic Matching for the ARQMath Lab

ARQMath Lab 2021, CLEF

Wei Zhong\*, **Yuqing Xie\***, Jimmy Lin (\*Both authors contributed equally)

Sept 2022

### An Embedding-Based Grocery Search Model at Instacart

SIGIR eCom 2022

**Yuqing Xie**, Taesik Na, Xiao Xiao, Saurav Manchanda, Young Rao, Zhihong Xu, Guanghua Shu, Esther Vasiete, Tejaswi Tenneti, Haixun Wang

July 2022

### Regression Bugs Are In Your Model! Measuring, Reducing and Analyzing Regressions In NLP Model Updates

[ACL 2021](#)

**Yuqing Xie**, Yi-an Lai, Yuanjun Xiong, Yi Zhang, Stefano Soatto

May 2021

### Segatron: Segment-Aware Transformer for Language Modeling and Understanding

[AAAI 2021](#)

He Bai, Peng Shi, Jimmy Lin, **Yuqing Xie**, Luchen Tan, Kun Xiong, Wen Gao, Ming Li

Feb 2021

### Approximate Nearest Neighbor Search and Lightweight Dense Vector Reranking in Multi-Stage Retrieval Architectures

[SIGIR 2020](#)

Zhengkai Tu\*, Wei Yang\*, Zihang Fu\*, **Yuqing Xie**, Luchen Tan, Kun Xiong, Ming Li, Jimmy Lin (\*All authors contributed equally)

Sept 2020

### Distant Supervision for Multi-Stage Fine-Tuning in Retrieval-Based Question Answering

[WWW 2020](#)

**Yuqing Xie\***, Wei Yang\*, Luchen Tan, Kun Xiong, Nicholas Jing Yuan, Baoxing Huai, Ming Li and Jimmy Lin (\*Both authors contributed equally)

Apr 2020

### Rapid Adaptation of BERT for Information Extraction on Domain-Specific Business Documents.

[arXiv:2002.01861](https://arxiv.org/abs/2002.01861)

Ruixue Zhang, Wei Yang, Luyun Lin, Zhengkai Tu, **Yuqing Xie**, Zihang Fu, Yuhao Xie, Luchen Tan, Kun Xiong, and Jimmy Lin

Feb 2020

### End-to-End Open-Domain Question Answering with BERTserini

[NAACL 2019](#)

Wei Yang\*, **Yuqing Xie\***, Aileen Lin, Xingyu Li, Luchen Tan, Kun Xiong, Ming Li, Jimmy Lin (\*Both authors contributed equally)

June 2019

### Data Augmentation for BERT Fine-Tuning in Open-Domain Question Answering.

[arXiv:1904.06652](https://arxiv.org/abs/1904.06652)

Wei Yang\*, **Yuqing Xie\***, Luchen Tan, Kun Xiong, Ming Li, Jimmy Lin

Apr 2019

## Industry Experience

---

### Instacart

Machine Learning Engineer - Ph.D. Internship

Remote - US

Aug 2021 - Dec 2021

- Developed a self-adversarial **embedding based grocery search model**
- Achieves 2.06% relative improvement in recall@20 for offline human evaluation dataset comparing with previous embedding based systems.
- Paper accepted by SIGIR eCom 22'.

### Amazon AWS AI

Applied Scientist Intern

Remote - US

Aug 2020 - Apr 2021

- Explored **regression-free model update problem**, which aims to reduce errors introduced by new neural network models.
- Proposed methods including constraint optimization and meta adaptor to effectively reduce regression in NLU tasks.
- Published a paper in ACL 21' and will submit another to EMNLP 21' that relates to the project.

### RSVP.ai

Machine Learning Engineer - Part-time

Waterloo, Canada

Dec 2018 - June 2020

- Implemented **paraphrase generation** models and conducted human labelling experiment comparing different models to verify the rationality of current evaluation methods.
- Improved the performance of **open domain question answering** system with paraphrase-augmented queries, retriever augmented context, and named-entity-filtered examples.
- Constructed an end-to-end question answering system that integrates BERT (Tensorflow) with the open-source Anserini (a Lucene IR toolkit) information retrieval toolkit both in English and Chinese and create new state of the art.
- Deployed the QA system's Elastic Search API and provided real-time online service.

### Yitu-Tech

Machine Learning Algorithm Intern

Shanghai, China

Feb 2018 - Aug 2018

- Improved the accuracy of the **Single Shot MultiBox Detector** model for car detection.
- Implemented the HOG-SVM for digital recognition in car license identification.

## Projects

---

### Tree-structure Information Integration for Question Generation

University of Waterloo

Advisor: Ming Li

Jan 2020 - Apr 2020

- Explored integrate parsing tree knowledge for more logical question generation.
- The proposed method qualitatively introduces structure information compared with universal sentence encoder.

### Paper Recommendation via GraphX

University of Waterloo

Advisor: Adam Roegiest

Jan 2019 - Apr 2019

- Applied GraphX to build an academic paper recommendation system.
- Implemented PageRank, keyword filtering, and pattern finding algorithms in GraphX and compared the framework against MapReduce on Hadoop.
- Applied the algorithm on a citation network to recommend according to users' interest.

### Contextual Decomposition for Rationalizing LSTM Predictions

University of Waterloo

Advisor: Yaoliang Yu

Sept 2018 - Nov 2018

- Decomposed and analyzed LSTM models in token level to understand the behaviour on named entity detection task.
- Implemented and modified multi-view learner to analyze the difference of the two-directions of LSTM model.

## Skills

---

**DevOps** AWS(EC2, EFS, S3, Sagemaker), Docker, Git, MapReduce

**Frameworks and Tools** PyTorch, Tensorflow, Keras

**Programming Languages** Python, Bash, SQL, LaTeX, MATLAB, C, Java, Scala

**Languages** Chinese, English

## Serving

---

Dec, 2022 **Reviewer**, EMNLP 2022

Abu Dhabi

Oct, 2022 **External Reviewer**, COLING 2022

Gyeongju, Korea

Sept 2022 **Committee member**, CLEF

Bologna, Italy

June, 2022 **Reviewer**, Springer Nature: Artificial Intelligence and Law