Kun Qian

United States | +1-831-239-8201

kunqian.usa@gmail.com

🏠 personal website | 🐙 linkedin |

PERSONAL STATEMENT

I am a computer scientist and researcher who has been working on a variety of data-centric AI problems including Data Integration, Data Exchange, Active Learning, NLP, Explainabile AI, and Information Retrieval. I have more than 30 peer reviewed publications and 1 granted patent (4 more are under review). My work at IBM and Amazon has influenced hundreds of millions of customers around the world. Besides industrial work, I served as PC members and referees for many top computer science conferences and journals, and I also mentored several junior researchers.

HIGHEST EDUCATION

UNIVERSITY OF CALIFORNIA, SANTA CRUZ
2012-2017 / USA

PH.D. IN COMPUTER SCIENCE

•Advisers: Balder ten Cate, Phokion Kolaitis, Wang-Chiew Tan

BEIHANG UNIVERSITY
2007-2010 / China

MASTER OF SOFTWARE ENGINEERING

CHONGQING UNIVERSITY
2003-2007/ China

BACHELOR OF SOFTWARE ENGINEERING

EXPERIENCE

APPLE
2021.06 - current

Sr. Machine Learning Engineer and Researcher

•I am a member of the Open Domain Knowledge Extraction team at Apple Knowledge Platform team, where I work on features that improve the richness and correctness of Apple KG that powers many Apple products.

AMAZON.COM, INC.
2021.03 - 2022.06

Applied Scientist

•I was part of the Global Search Quality team at Amazon Search. I worked on improving Amazon customers' search experiences. I led a project that brings in high-quality product substitutes to search matchset to fulfill customers' shopping intents worldwide. Techniques used to solve this problem include large-scale graph algorithm, cross-encoder based defect classifier, and graph neural networks. My work has been delivered to Amazon product that has impacted close to 1 billion Amazon customers, increased annual revenue at 100+ million scale.

IBM RESEARCH AI
2017.02 - 2021.03

Research Staff Member

•I was the technical lead for a couple of research projects on low-resource machine learning frameworks for entity matching and entity normalization. By leveraging active learning, transfer learning, and weak supervision, our frameworks can produce models that are comparable to the state-of-the-art approaches while using only 1% to 10% of the training data.
- Demos of systems developed by me:
  - **PARTNER** (https://www.youtube.com/watch?v=6DDXARJezz4).
  - **SystemER** (https://www.youtube.com/watch?v=5ENye9hg-UA)

•My research has been transferred to IBM's products including IBM Watson Health and IBM Watson NLP.

•I led a research project on explainable AI for NLP, which systematically reviewed and summarized the state-of-the-art work on this topic, and have given two well-received technical tutorials at two top AI conferences: KDD'2021 and AACL'2020. Project website: https://xainlp.github.io/

•I had more than 20 publications/tutorials/demos at prestigious research conferences such as AAAI, AACL, ACL, CIKM, COLING, DIS, EMNLP, ICDE, ISWC, IUI, KDD, PODS and VLDB including a best demo award at ISWC'2020. Also filed 5 patents (1 granted).

RECENT PUBLICATIONS

Check my full publication records at my **DBLP** and **GoogleScholar** profiles.

1. Shipi Dhanorkar, Marina Danilevsky, Yunyao Li, Lucian Popa, **Kun Qian**, Anbang Xu
   Explainability for Natural Language Processing. (tutorial).
   (SIGKDD 2021) The 27th ACM SIGKDD Conference on Knowledge Discovery Data Mining.
   Equal contribution, authors are listed in alphabetical order)).

2. **Kun Qian**, Marina Danilevsky, Yannis Katsis, Ban Kawas, Erick Oduor, Lucian Popa, Yunyao Li
   XNLP: A Living Survey for XAI Research in Natural Language Processing.
   (IUI 2021) Annual Conference on Intelligent User Interfaces (demo track).

3. **Kun Qian**, Poornima Chozhiyath Raman, Lucian Popa, and Yunyao Li
   Learning Structured Representations of Entity Names using Active Learning and Weak Supervision.
   (EMNLP 2020, Acceptance rate: 16.7%) The 2020 Conference on Empirical Methods in Natural Language Processing.

4. Marina Danilevsky*, **Kun Qian***, Ranit Aharonov, Yannis Katasis, Ban Kawas, Prithviraj Sen
   A Survey of the State of Explainable AI for Natural Language Processing.
   (AACL-IJCNLP 2020) The 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics.
   (* Equal contribution)

5. **Kun Qian**, Poornima Chozhiyath Raman, Yunyao Li, and Lucian Popa.
   PARTNER: Human-in-the-loop Entity Name Understanding with Deep Learning.
   (AAAI-2020) The 34th AAAI Conference on Artificial Intelligence (demo).

6. **Kun Qian**, Lucian Popa, and Prithviraj Sen.
   SystemER: A Human-in-the-loop System for Explainable Entity Resolution.
   (VLDB-2019) The 45th International Conference on Very Large Data Bases.

7. Jungo Kasai, **Kun Qian**, Sairam Gurajada, Yunyao Li, Lucian Popa.
   Low-resource Deep Entity Resolution with Transfer and Active Learning.
   (ACL-2019) The 57th Annual Meeting of The Association for Computational Linguitics.

8. Phokion G. Kolaitis, Lucian Popa, and **Kun Qian***.
   Knowledge Refinement via Rule Selection.
   (AAAI-2019) The 33rd AAAI Conference on Artificial Intelligence .
   (Authors are listed in alphabetical order)

9. **Kun Qian**, Nikita Bhutani, Yunyao Li, H.V. Jagadish, Mauricio Hernandez.
   LUSTRE: An Interactive System for Entity Structured Representation and Variant Generation.
   (ICDE 2018) 34th IEEE International Conference on Data Engineering.

10. **Kun Qian**, Lucian Popa, and Prithviraj Sen.
    Active Learning for Large-Scale Entity Resolution.
    (CIKM 2017) The 2017 ACM on Conference on Information and Knowledge Management.

PATENTS

1. Nikita Bhutani, Mauricio Hernandez-Sherrington, Yunyao Li, Min Li, and **Kun Qian**.
   Entity Structured Representation and Variant Generation.
   U.S. Patent 10,585,986
   Status: **Granted**

PROFESSIONAL AFFILIATIONS AND SERVICES

Membership
- ACM, AAAI

Conference Program Committee Member
- PVLDB 2022, ACL-NAACL 2022
- EMNLP 2021, DaSH@NAACL 2021, IUI 2021 (demo), NAACL 2021, AAAI 2021
- ACL 2020, IJCAI 2020, ICDE 2020 (industry), AAAI 2020
- IEEE BigData 2019
- WebDB 2018, CIKM 2018
- CIKM 2017, KDD 2017, AAAI 2017, ADAMA 2017

Journal Referee
- ACM TODS (2018, 2019), IEEE TKDE (2019)

RECENT AWARDS

BEST DEMO PAPER AWARD - ISWC 2020	*2020*
IBM CLASS-A RESEARCH ACCOMPLISHMENT	*2017*

Skills

Programming Languages	Python | Java | Javascript
Deep Learning & Datea Science	PyTorch | Pytorch-Transformers | NLTK | Jupyter | Matplotlib | Numpy | Pandas
Web Programming	Angular | Angular Material | W3.css | Typescript | Django
Cloud & Distributed Computing	AWS | PySpark
Others	IBM SystemT | LaTeX