

# Ke Yang

Ph.D. Candidate, Computer Science, New York University

ky630@nyu.edu  
+1 2159005854  
[Google Scholar page](#)  
github: [KeYang0923](#)  
[keyang0923.github.io](#)

---

## Education

**Ph.D. candidate in Computer Science**, New York University expected 05.2021

- Advisor: Julia Stoyanovich, GPA 3.89
- Dissertation: Fairness, diversity, and interpretability of rankings
- Committee: Julia Stoyanovich (chair), Sebastian Schelter, Daniel Neill, and Oded Nov

**Ph.D. candidate in Computer Science**, Drexel University (transferred) 2015-2018

- Advisor: Julia Stoyanovich, GPA 3.98

**M.S.E in Computer Science**, Beijing Technology and Business University, 2015

- Advisors: Qian Mo and Zhongming Han, GPA 3.51
- Thesis: Detecting web spammers in online shopping websites
- Graduated summa cum laude

**B.Eng. in Software Engineering**, Beijing Technology and Business University 2012

- Advisor: Qian Mo, GPA 3.45
  - Graduated summa cum laude
- 

## Professional Experience

**Research Assistant**, Tandon School of Engineering, New York University 2019-2021

Supervisor: Julia Stoyanovich

- Current project: the impact of technical bias on model serving in data science pipelines.
  - Selected past project: [causal intersectionality for fair rankings](#).
- More projects at [keyang0923.github.io/projects](#).

**Research Intern**, AT&T Research, New York Summer 2019

Supervisors: Emily Dodwell, Ritwik Mitra, and Balachander Krishnamurthy

- Worked on a project to ensure fairness and transparency for internal AT&T machine learning projects. Developed a diagnostic tool for bias-related issues that can be used by project managers and data scientists at AT&T.

**Research Assistant**, College of Computing & Informatics, Drexel University, 2015-2018

Supervisor: Julia Stoyanovich

- Worked on a project to quantify fairness in rankings through equalized representation across groups, and proposed a mitigation framework to ensure a fair ranking outcome.

**Research engineer**, Elite & Resource (start-up company) 2014-2015

Supervisor: Peng Sun

- Worked on a project to model historical data of flood disaster and develop techniques to help with flood prevention. Proposed a model to estimate the probability of future floods using water flow rate of small watershed torrents, which is recognized as a significant signal of potential flood. Our model is integrated as a core component of a national watershed data management system.

## Research Interest

I am interested in ethical concerns, such as fairness, accountability, transparency, interpretability, and in the social impacts of data science.

---

## Peer-Reviewed Publications

1. **Ke Yang**, Biao Huang, Julia Stoyanovich, and Sebastian Schelter, “Fairness-Aware Instrumentation of Preprocessing Pipelines for Machine Learning”, in *Proceedings of the 4th Workshop on Human-In-the-Loop Data Analytics, in conjunction with the 2020 ACM SIGMOD/PODS Conference (HILDA@SIGMOD 2020)*, Portland, OR, USA.
2. **Ke Yang**, Vasilis Gkatzelis, and Julia Stoyanovich, “Balanced ranking with diversity constraints”, in *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI 2019)*, Macao, China.
3. **Ke Yang**, Julia Stoyanovich, Abolfazl Asudeh, Bill Howe, H.V. Jagadish, and Gerome Miklau, “A Nutritional Label for Rankings” demo in *Proceedings of the 2018 ACM International Conference on Management of Data (SIGMOD 2018)*, Houston, USA.
4. Julia Stoyanovich, **Ke Yang**, H.V. Jagadish, “Online Set Selection with Fairness and Diversity Constraints” in *Proceedings of the 21th International Conference on Extending Database Technology (EDBT 2018)*, Vienna, Austria.
5. **Ke Yang**, Julia Stoyanovich, “Measuring Fairness in Ranked Outputs” in *Proceedings of the 29th International Conference on Scientific and Statistical Database Management (SSDBM 2017)*, Chicago, USA.
6. Zhongming Han, **Ke Yang**, Xusheng Tan, “Analyzing Spectrum Features of Weight User Relation Graph to Identify Large Spammer Groups in Online Shopping Websites” in *Chinese Journal of Computers*, 40(4): 939-954 (2017) (in Chinese).
7. Zhongming Han, **Ke Yang**, Fengmin Xu, Dagao Duan, “Probabilistic Graphical Model for Detecting Spammers in Microblog Websites” in *Journal of Embedded System*, 8(1): 12-23 (2016).
8. Kai Wang, **Ke Yang**, “A method and application system of detecting web spammers” in *China Invention Patent*, (2015) (in Chinese), No. CN201510012860.
9. Qian Mo, **Ke Yang**, “Overview of Web spammer detection” in *Journal of Software*, 25(7): 1505-1526 (2014) (in Chinese).

## Working Papers

10. Meike Zehlike, **Ke Yang**, Julia Stoyanovich, “Mapping Normative Frameworks to Fair Ranking”, in progress.
11. **Ke Yang**, Joshua R. Loftus, and Julia Stoyanovich, “Causal intersectionality for fair ranking”, [arxiv.org/abs/2006.08688](https://arxiv.org/abs/2006.08688).
12. **Ke Yang**, Sebastian Schelter, Julia Stoyanovich, “FairDeploy: Disparate Impact of Technical Bias in the Deployment of Machine Learning Models”, in progress.
13. Meike Zehlike, **Ke Yang**, Julia Stoyanovich, “Fairness in Rankings: A Survey”, in progress.

---

## Open Source Tools

- **Mirror Data Generator**: a python library that generates synthetic data to mirror data distribution problems due to sampling and societal bias.  
[github.com/DataResponsibly/MirrorDataGenerator](https://github.com/DataResponsibly/MirrorDataGenerator)
- **Ranking Facts**: a web-based tool that generates a “nutritional label” for rankings.  
[github.com/DataResponsibly/RankingFacts](https://github.com/DataResponsibly/RankingFacts)
- **FairDAGs**: a web-based tool that extracts a directed acyclic graph (DAG) representation of data science pipelines, and tracks data distribution changes after each operation.  
[github.com/DataResponsibly/fairDAGs](https://github.com/DataResponsibly/fairDAGs)

## Teaching & Advising Experience

**Teaching assistant** at Drexel University

- Fundamentals of Databases (graduate level), Winter 2017
- Database Systems (undergraduate level), Spring 2017

**Research advisor** at New York University

- Jensine Raihan (BS in Computer Science), Fall 2020, “Ranking Facts in real datasets”
- Biao Huang (MS in Data Science), Spring 2020, “Fairness-aware instrumentation of preprocessing pipelines for machine learning”
- Samasth Norway Ananda (MS in Computer Science), Spring and Summer 2020, “FairPrep: promoting data to a first-class citizen in studies on fairness enhancing interventions”

---

## Academic Service

**Conference Organization**

- Technical chair: 4th Workshop on Data Management for End-to-End Machine Learning (DEEM) at SIGMOD 2020

**Journal Reviewer**

- Information Systems 2019

**Program Committee Member**

- ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT) 2020
- Workshop on Data Management for End-to-End Machine Learning (DEEM) at SIGMOD 2020

**External Reviewer**

- ACM Conference on Human Factors in Computing Systems (CHI) 2020

**Student volunteer**

- ACM Conference on Fairness, Accountability, and Transparency (ACM FAT\*) 2019

---

## Fellowships, Awards & Honors

- Selected participant, ACM Conference on Fairness, Accountability, and Transparency (ACM FAT\*) Ph.D. consortium, 2019
- Selected participant, AT&T graduate student symposium, 2018
- Invited participant, Fairness in ML Workshop at Google, 2018
- Student scholarship, ACM Conference on Fairness, Accountability, and Transparency (ACM FAT\*) 2018
- Fellowship for Doctoral Study, Graduate School, Beijing Technology and Business University, 2015
- Dissertation Award, Graduate School, Beijing Technology and Business University, 2015