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   2015-2018
- Advisor: Julia Stoyanovich, GPA 3.98

M.S.E in Computer Science, Beijing Technology and Business University,   06.2015
- 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   06.2012
- 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 the 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 Lab Research, New York   2019 Summer
Supervisors: Emily Dodwell, Ritwik Mitra, and Balachander Krishnamurthy
- Worked on a project to insure fairness and transparency for internal AT&T machine learning projects. This project aims to ensure that the projects' outcome meets reasonable guidelines for fairness without illegal bias and discrimination. We developed a diagnostic tool for bias-related issues that can be used by project managers and data scientists in 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 we 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. In this project, we proposed a model to predict the probability of future flood 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 national watershed data management system.

Research Interest	I'm interested in ethical concerns, such as fairness, accountability, transparency, interpretability, and the social impact of the algorithms in data science pipelines.

Peer-Reviewed Publications

1. FairDeploy: Disparate Impact of Technical Bias in the Deployment of Machine Learning Models, working in progress.

2. **Ke Yang**, Joshua R. Loftus, and Julia Stoyanovich, "Causal intersectionality for fair ranking", arxiv.org/abs/2006.08688.

3. **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.

4. **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.

5. **Ke Yang**, Julia Stoyanovich, Abolfazl Asudeh, Bill Howe, HV 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.

6. Julia Stoyanovich, **Ke Yang**, HV Jagadish, "Online Set Selection with Fairness and Diversity Constraints" in *Proceedings of the 21th International Conference on Extending Database Technology* (EDBT 2018), Vienna, Austria.

7. **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.

8. 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).

9. 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).

10. Kai Wang, **Ke Yang**, "A method and application system of detecting web spammers" in *China Invention Patent*, (2015) (in Chinese), No. CN201510012860.

11. Qian Mo, **Ke Yang**, "Overview of Web spammer detection" in *Journal of Software*, 25(7): 1505-1526 (2014) (in Chinese).

Open Source Tools

- **Mirror Data Generator**: a python script that generates synthetic data to mirror issues, such as sampling and societal bias. The issues are described by the correlation between features. github.com/DataResponsibly/CIFRank

- **Ranking Facts**: a web-based tool that generates a "nutritional label" for rankings. Each label shows a fact about the ranking. For example, a fact about fairness explains whether the ranking shows statistical parity between groups that are defined by a user-specified feature. github.com/DataResponsibly/RankingFacts

- **FairDAGs**: a web-based tool that extracts directed acyclic graph (DAG) representation of data science pipelines and tracks the changes of the distributions of targets and groups due to each operation. The groups are often defined by a user-specified feature in the dataset. github.com/DataResponsibly/fairDAGs

Teaching & Advising Experience	**Lab section leader** at New York University - Special Topics in Data Science: Responsible Data Science, 2019 Spring **Teaching assistant** at Drexel University - Fundamentals of Databases (graduate level), 2017 Winter - Database Systems (undergraduate level), 2017 Spring **Research advisor** at New York University Independent Studies: - Jensine Raihan, 2020 Fall, "Ranking Facts in real datasets" - Biao Huang, 2020 Spring, "Fairness-aware instrumentation of preprocessing pipelines for machine learning" Research assistant: - Samasth Norway Ananda, 2020 Spring and Summer, "FairPrep: promoting data to a first-class citizen in studies on fairness enhancing interventions"
Academic Service	**Conference Organizing Committee:** - Technical chair: Workshop on Data Management for End-to-End Machine Learning (DEEM) at SIGMOD 2020 **Journal Review:** - Information Systems 2019 **Conference Proceedings Review:** - ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT) 2020 - Workshop on Data Management for End-to-End Machine Learning (DEEM) at SIGMOD 2020 - 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	- A selected participant at ACM FAT* Ph.D. consortium, 2019 - A selected participant at AT&T graduate student symposium, 2018 - Invited participant at Fairness in ML Workshop at Google, 2018 - Student Scholarship at 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