

# Huiji Gao

---

## CONTACT INFORMATION

*E-mail:* ghjmon@gmail.com

*Cell:* 480-334-5423

*Homepage:* <http://www.nini2yoyo.com>

*LinkedIn:* <https://www.linkedin.com/in/huiji-gao>

## AREA OF EXPERTISE

Data Mining, Machine Learning, Artificial Intelligence

- Large-scale Recommender Systems, Information Retrieval, Computational Advertising
- Deep Natural Language Processing, Conversational AI
- Location-Based Mobile Social Networking, Social Computing

Citations (Google Scholar): 4798; h-index:25; i10-index: 31

## EXPERIENCE

Jul 2020 - Present: **Senior Engineering Manager - ML and AI**, LinkedIn Corporation, CA.

- Team Focus: Search and Recommender Systems
- Lead AI Algorithms Foundation team (about 20 talents) to develop efficient and effective horizontal AI technologies on Natural Language Processing, Multimedia Mining, Ranking, and Personalization to power LinkedIn search and recommender systems.
- Drive the development of unified modeling framework GDMix-DeText that enables end-to-end solution on deep natural language understanding and personalization modeling.
- The framework *GDMix* is open sourced (<https://github.com/linkedin/gdmix>) together with *DeText* (<https://github.com/linkedin/detext>) and being adopted by many LinkedIn recommender systems (Ads, Feed, Learning, etc.) and search systems (Job Search, People Search, etc.) with significant product impact.
- Improve pre-training efficiency of *LiBERT* (foundation language model at LinkedIn based on BERT) with advanced optimizers.

Aug 2018 - June 2020: **Engineering Manager - ML and AI**, LinkedIn Corporation, CA.

- Team Focus: Search and Recommender Systems
- Lead AI Algorithms Foundation team (about 15 talents) to develop efficient and effective horizontal AI technologies on Natural Language Processing, Ranking, and Personalization to power LinkedIn search and recommender systems.
- Drive the development and open source of *DeText* (<https://github.com/linkedin/detext>), a deep NLP framework for ranking, classification, and language generation. Deployed end-to-end *DeText* solutions to LinkedIn products on search ranking, intent modeling, query auto complete, query suggestion, spell check, and conversational AI related applications, with significant product impact. *DeText* has received CIKM Best Paper Award and multiple media reports with around 1000 stars on GitHub.
- Drive the development of *LiBERT*, BERT model trained and calibrated on LinkedIn data, as the foundation language model for company use. The work has been reported by Fortune magazine.
- Drive the development *GDMix*, a deep ranking personalization framework, that enables efficient modeling on user/item level features such as id feature modeling, incremental learning for large-scale recommender systems.

Mar 2018 - Jul 2018: **Staff Machine Learning Engineer**, LinkedIn Corporation, CA.

- Project Focus: Information Retrieval
- Establish LinkedIn's international search relevance foundation w.r.t. language detection, query tagging, and query suggestion powered by Machine Learning and Artificial Intelligence.

Jun 2016 - Feb 2018: **Senior Machine Learning Engineer**, LinkedIn Corporation, CA.

- Project Focus: Computational Advertising, Information Retrieval
- Worked on Ads relevance products including audience expansion behavior modeling, campaign performance optimization, and CTR prediction. Developed and deployed several machine learning models that have generated double-digit annualized lift of Ads revenue. Led people search

ranking model and data quality improvement, achieved double-digit key search metrics lift.

Feb 2015 - May 2016: **Applied Research Engineer**, LinkedIn Corporation, CA.

- Project Focus: Computational Advertising, CTR Prediction
- Predict click-through rate (CTR) of LinkedIn's sponsored updates w.r.t. cold-start and warm-start models. Developed features including LDA and GBDT under Spark and Hadoop MR. Implemented offline/online evaluation metrics for model decision making.

May 2014 - Aug 2014: **Software Engineering Intern**, LinkedIn Corporation, CA.

- Research Focus: Data Mining, Machine Learning, and Recommender Systems
- Develop company interest models for job recommendation and recruiter search on Hadoop Distributed File System (HDFS) using pig and java.

May 2013 - Aug 2013: **Research Intern**, IBM Almaden Research Center, CA.

- Research Focus: Social Media Mining
- Developed matrix factorization based algorithm to model online users' attitude towards controversial topics in online social media in terms of sentiment polarity, opinion preferences, and online actions with mathematical models.

Sep 2007 - Jun 2009: **Research Assistant**, PRIS Lab, BUPT, China

- Research Focus: Information Retrieval
- Developed document retrieval component for project COSE (Campus Object Search Engine), an campus oriented information retrieval system of BUPT

Dec 2005 - Jun 2007: **Research Assistant**, Innovation Lab, BUPT, China

- Research Focus: Embedded Systems
- Developed heartbeat pattern detection module for project FAES (First-Aid Expert System), an embedded system designed to monitor the health status of patients
- Led the project See-Saw, an intelligent system for improving study efficiency of primary students

## EDUCATION

**Arizona State University**, Tempe, AZ, U.S

Ph.D., Computer Science, GPA 4.00/4.00, 2009 – 2015

Thesis: Personalized POI Recommendation on Location-Based Social Networks

Advisor: Huan Liu

**Beijing University of Posts and Telecommunications**, Beijing, China

M.S, Signal and Information Processing, GPA: 90.05/100, Rank: 2/500, 2007 – 2009

Advisor: Jun Guo

- *Admitted without the Graduate Entrance Examination*

**Beijing University of Posts and Telecommunications**, Beijing, China

B.S, Information Automation, GPA: 84.73/100, Rank: 2/59, 2003 – 2007

## SELECTED HONORS AND AWARDS

- Best Paper Award (DeText: A Deep Text Ranking Framework with BERT) in CIKM Applied Research Track, 2020
- LinkedIn Engineering Showcase on “Natural Language Intelligence at Scale” (Top 1 in AI Org), 2019
- Best Model Award for Job Recommendation, Relevance Hackathon, 2016
- ASU Graduate Education Dissertation Fellowship (Awarded to 1 Student per School), 2014
- ASU President's Award for Innovation, 2014
- The 3rd Place in Task 2: Next Location Prediction of Nokia Mobile Data Challenge, 2012
- Best Performer in Baseline Ranking of TREC Blog Faceted Blog-Distillation Task, 2009

- “Samsung” Scholarship (Awarded to Top 3 of 500 Students), 2008
- World Top 200 in “Microsoft Imagine Cup Embedded System Competition”, 2007
- The Third Prize in “Technology Paper” Award of “Innovation Competition”, BUPT, 2007.
- The Third Prize in “Artistic Design” Award of “Innovation Competition”, BUPT, 2007.
- Outstanding Undergraduate Student of Beijing, China, 2007
- First Prize Scholarship (Awarded to Top 5% Students), BUPT, 2006
- First Prize in “‘INTEL Cup’ - National College Students Embedded System Competition”, 2006
- The Second Prize University Scholarship, BUPT, 2003-2005.

## PUBLICATIONS

### Tutorials

1. Zhoutong Fu, Huiji Gao, Weiwei Guo, Sandeep Jha, Jun Jia, Xiaowei Liu, Jun Shi, Sida Wang, Mingzhou Zhou, Bo Long, “Deep Learning for Search and Recommender Systems in Practice”, the 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (SIGKDD 2020).
2. Weiwei Guo, Huiji Gao, Jun Shi, Bo Long, Liang Zhang, Bee-Chung Chen, Deepak Agarwal, “Deep Natural Language Processing in Search and Recommender Systems”, the 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (SIGKDD 2019).
3. Weiwei Guo, Huiji Gao, Jun Shi, Bo Long, “Deep Natural Language Processing in Search Systems”, the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2019).
4. Huiji Gao, Jiliang Tang, and Huan Liu, “Personalized Location Recommendation on Location-based Social Networks”, the ACM Recommender System Conference (RecSys 2014).

### Book

1. Huiji Gao and Huan Liu. “Mining Human Mobility in Location-Based Social Networks”, Morgan & Claypool Publisher, April, 2015.

### Book Chapters

1. Huiji Gao and Huan Liu. “Data Analysis on Location-Based Social Networks” in Mobile Social Networking: An Innovative Approach. Editors: Alvin Chin and Daqing Zhang, Springer, 2014.
2. Nitin Agarwal, Shamanth Kumar, Huiji Gao, Reza Zafarani and Huan Liu. “Analyzing Behavior of Influentials Across Social Media” in Behavior Computing: Modeling, Analysis, Mining and Decision, Springer, 2012.

### Conferences

1. Weiwei Guo, Xiaowei Liu, Sida Wang, Huiji Gao, Ananth Sankar, Zimeng Yang, Qi Guo, Liang Zhang, Bo Long, Bee-Chung Chen and Deepak Agarwal. “DeText: A Deep Text Ranking Framework with BERT”, the 29nd ACM International Conference on Information and Knowledge Management (CIKM), 2020. (**Best Paper Award**)
2. Michael Kazi, Weiwei Guo, Huiji Gao, and Bo Long. “Improving Search Query Suggestion With User Feedback”, the 29nd ACM International Conference on Information and Knowledge Management (CIKM), 2020.
3. Sida Wang, Weiwei Guo, Huiji Gao, and Bo Long. “Efficient Neural Query Auto Completion”, the 29nd ACM International Conference on Information and Knowledge Management (CIKM), 2020.
4. Jianling Zhong, Weiwei Guo, Huiji Gao, and Bo Long. “Personalized Query Suggestions”, The 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 2020.

5. Huiji Gao, Jiliang Tang, Xia Hu, and Huan Liu. “Content-aware point of interest recommendation on location-based social networks”, the 29th AAAI Conference on Artificial Intelligence (AAAI), 2015.
6. Fred Morstatter, Huiji Gao, and Huan Liu. “Discovering Location Information in Social Media”, IEEE Data Engineering Bulletin Special Issue on Location-based Social Media Analysis (IEEE-DEB), 2015.
7. Xinxin Zhao, Huiji Gao, Lingjun Li, Huan Liu, and Guoliang Xue. “An efficient privacy preserving location based service system”, IEEE Global Communications Conference (GLOBECOM), 2014.
8. Xia Hu, Jiliang Tang, Huiji Gao, and Huan Liu. “Social spammer detection with sentiment information”, 2014 IEEE International Conference on Data Mining (ICDM), 2014.
9. Huiji Gao, Jalal Mahmud, Jilin Chen, Jeffrey Nichols, and Michelle Zhou. “Modeling User Attitude toward Controversial Topics in Online Social Media”, the Eighth International AAAI Conference on Weblogs and Social Media (ICWSM), 2014.
10. Jalal Mahmud and Huiji Gao. “Why Do You Spread This Message? Understanding Users Sentiment in Social Media Campaigns”, the Eighth International AAAI Conference on Weblogs and Social Media (ICWSM), 2014.
11. Huiji Gao, Jiliang Tang, Xia Hu, and Huan Liu. “Modeling Temporal Effects of Human Mobile Behavior on Location-Based Social Networks”, the 22nd ACM International Conference on Information and Knowledge Management (CIKM), 2013.
12. Huiji Gao, Jiliang Tang, Xia Hu, and Huan Liu. “Exploring Temporal Effects for Location Recommendation on Location-Based Social Networks”, the 7th ACM Recommender Systems Conference (RecSys), 2013.
13. Jiliang Tang, Huiji Gao, Xia Hu, and Huan Liu. “Context-Aware Review Helpfulness Rating Prediction”, the 7th ACM Recommender Systems Conference (RecSys), 2013.
14. Huiji Gao, Xufei Wang, Jiliang Tang, and Huan Liu. “Network Denoising in Social Media”, the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2013.
15. Jiliang Tang, Xia Hu, Huiji Gao, and Huan Liu. “Exploiting Local and Global Social Context for Recommendation”, the 23rd International Joint Conference on Artificial Intelligence (IJCAI), 2013.
16. Xia Hu, Jiliang Tang, Huiji Gao, and Huan Liu. “Unsupervised Sentiment Analysis with Emotional Signals”, the 22nd International World Wide Web Conference (WWW), 2013.
17. Xia Hu, Jiliang Tang, Huiji Gao, and Huan Liu. “ActNeT: Active Learning for Networked Texts in Microblogging”, the 13th SIAM International Conference on Data Mining, 2013.
18. Jiliang Tang, Huiji Gao, Xia Hu, and Huan Liu. “Exploiting Homophily Effect for Trust Prediction”(hTrust), the 6th ACM International Conference on Web Search and Data Mining (WSDM), 2013.
19. Huiji Gao, Jiliang Tang, and Huan Liu. “gSCorr: Modeling Geo-Social Correlations for New Check-ins on Location-Based Social Networks”, the 21st ACM International Conference on Information and Knowledge Management (CIKM), 2012.
20. Jiliang Tang, Huiji Gao, Huan Liu, and Atish Das Sarma. “eTrust: Understanding Trust Evolution in an Online World”, the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2012.
21. Huiji Gao, Jiliang Tang, and Huan Liu. “Mobile Location Prediction in Spatio-Temporal Context”, Nokia Mobile Data Challenge, 2012.

22. Huiji Gao, Jiliang Tang, and Huan Liu. “Exploring Social-Historical Ties on Location-Based Social Networks”, the Sixth International AAAI Conference on Weblogs and Social Media (ICWSM), 2012.
23. Jiliang Tang, Huiji Gao, and Huan Liu. “mTrust: Discerning Multi-Faceted Trust in a Connected World”, the Fifth ACM International Conference on Web Search and Data Mining (WSDM), 2012.
24. Huiji Gao, Xufei Wang, Georey Barbier, and Huan Liu. “Promoting Coordination for Disaster Relief - From Crowdsourcing to Coordination”, Social Computing, Behavioral-Cultural Modeling, and Prediction (SBP), 2011.
25. Xufei Wang, Lei Tang, Huiji Gao, and Huan Liu. “Discovering Overlapping Groups in Social Media”, the Tenth IEEE International Conference on Data Mining series (ICDM), 2010.
26. Lei Tang, Huiji Gao, and Huan Liu. “Network Quantification Despite Biased Labels”. Workshop on Mining and Learning with Graphs, KDD, 2010.

### Journal Papers

1. Huiji Gao, Jiliang Tang, and Huan Liu. “Addressing the Cold-Start Problem in Location Recommendation using Geo-Social Correlations”, Data Mining and Knowledge Discovery (DMKD), 29(2), 2014.
2. Jiliang Tang, Xufei Wang, Huiji Gao, Xia Hu and Huan Liu. “Enriching Short Texts Representation in MicroBlog for Clustering”. Frontiers of Computer Science, 6(1), 2011.
3. Geoffrey Barbier, Reza Zafarani, Huiji Gao, Gabriel Fung and Huan Liu. “Maximizing Benefits from Crowdsourced Data”. Computational and Mathematical Organization Theory, 2011.
4. Huiji Gao, Geoffrey Barbier, and Rebecca Goolsby. “Harnessing the Crowdsourcing Power of Social Media for Disaster Relief”, *IEEE Intelligent Systems*, 26(3), 2010.

### WORKSHOPS

- Workshop Co-chair of Deep Natural Language Processing for Search and Recommendation  
Chairs: Bo Long, Jieping Ye, Zang Li, Huiji Gao, and Sandeep Jha. Held in conjunction with SIGIR 2020, July 25–30, 2020, Xian, China (Virtual).

### PATENTS

20+ patents applied and granted: <https://nini2yoyo.github.io/huiji-gao/patents/>

### MEDIA REPORT

- DeText: A deep NLP framework for intelligent text understanding, is reported by Fortune, VB (Venturebeat.com), and TheNewStack (“DeText: LinkedIn Open Source Deep Learning Framework for Natural Language Processing”), 2020
- GDMix: A Deep Ranking Personalization Framework, is reported by VB (Venturebeat.com) (“LinkedIn open-sources GDMix, a framework for training AI personalization models”), 2020
- LiBERT, BERT model trained and calibrated on LinkedIn data, is reported by Fortune Magazine: A.I. breakthroughs in natural-language processing are big for business, 2020

### INVITED TALKS

- “DeText: A Framework for Deep Natural Language Understanding at LinkedIn”, AI Camp and NYNLP Meetup, 2020.
- “Deep Natural Language Processing in Search Systems”, AI Tech Meetup – by LinkedIn, Uber and Facebook, LinkedIn, 2019.
- “Personalized POI Recommendation on Location-Based Social Networks”, Beijing University of Aeronautics and Astronautics, 2014.

### TECHNOLOGY SKILLS

- Programming Language: Pig, Spark, Python, TensorFlow, Java, C#, Matlab
- Web Development: ASP, JSP, PHP, Mysql, Pinot, Dreamwaver
- Video Production and Animation: Premiere, After Effects