BHAVYA
e-mail: bhavya@ibm.com Google Scholar: https://tinyurl.com/3c98tv22 phone: +12179040345

RESEARCH SUMMARY

I am broadly interested in developing natural language processing and text mining algorithms and interactive systems to assist humans in various domains including education, healthcare and IT. I'm experienced in various text mining and NLP techniques (e.g., large language models), developing applications (e.g., web-based platforms, mobile apps) based on those algorithms, and conducting user studies for need-finding and evaluation.

PEER-REVIEWED PUBLICATIONS

- Saurabh Jha, Rohan Arora, Yuji Watanabe, Takumi Yanagawa, Yinfang Chen, Jackson Clark, **Bhavya** et al. "ITBench: Evaluating AI Agents across Diverse Real-World IT Automation Tasks" In *Proceedings of the Forty-Second International Conference on Machine Learning (ICML).* 2025.
- **Bhavya**, Chris Palaguachi, Yang Zhou, Suma Bhat, ChengXiang Zhai. "Long-Form Analogy Evaluation Challenge" In *Proceedings of the 17th International Natural Language Generation Conference: Generation Challenges.* 2024.
- **Bhavya**, Shradha Sehgal, Jinjun Xiong, Chengxiang Zhai. "AnaDE1.0: A Novel Data Set for Benchmarking Analogy Detection and Extraction" In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*. 2024.
- **Bhavya**, Yang Zhou, Shradha Sehgal, Suma Bhat, Chengxiang Zhai. "Analego: Let's build analogies together!" Demo presentation at the *AAAI 2024 Workshop on AI for Education (AI4Ed)*. 2024.
- **Bhavya,** Paulina Toro Isaza, Yu Deng, Michael Nidd, Amar Prakash Azad, Larisa Shwartz, ChengXiang Zhai. "Exploring Large Language Models for Low-Resource IT Information Extraction" In *IEEE International Conference on Data Mining Workshops (ICDMW)*. 2023.
- **Bhavya**, Jinjun Xiong, Chengxiang Zhai. "CAM: A Large Language Model-based Creative Analogy Mining Framework" In *Proceedings of the ACM Web Conference (WWW).* 2023.
- **Bhavya**, Jinjun Xiong, Chengxiang Zhai. "Analogy Generation by Prompting Large Language Models: A Case Study of InstructGPT." In *Proceedings of the 15th International Conference on Natural Language Generation (INLG).* 2022.
- **Bhavya**, Si Chen, Zhilin Zhang, Wenting Li, Chengxiang Zhai, Lawrence Angrave, and Yun Huang. "Exploring collaborative caption editing to augment video-based learning." *Educational technology research and development.* 2022.
- Sandeep Puthanveetil Satheesan, **Bhavya,** Adam Davies, Alan B. Craig, Yu Zhang, and ChengXiang Zhai. "Toward a big data analysis system for historical newspaper collections research." In *Proceedings of the Platform for Advanced Scientific Computing Conference (ACM PASC)*. 2022.
- Zhilin Zhang*, **Bhavya*,** Lawrence Angrave, Ruihua Sui, Rob Kooper, Chirantan Mahipal, Yun Huang. "How Students Search Video Captions to Learn: An Analysis of Search Terms and Behavioral Timing Data", *ASEE Annual Conference & Exposition.* 2021.
- **Bhavya**, Jinfeng Xiao, Chengxiang Zhai. "Scaling Up Data Science Course Projects: A Case Study", In *Proceedings of the Eighth ACM Conference on Learning @ Scale (L@S)*. 2021.
- **Bhavya**, Jessie Chin, Chengxiang Zhai, Chung-Yi Chiu. "A critical review of consumer health search engine functionalities for patients with chronic health conditions", Poster Presented at *International Symposium on Human Factors and Ergonomics in Health Care.* 2021.
- **Bhavya**, Assma Boughoula, Aaron Green, and Chengxiang Zhai. "Collective Development of Large Scale Data Science Products via Modularized Assignments: An Experience Report", In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education (SIGCSE).* 2020.

- **Bhavya** and Chengxiang Zhai. "Explanation Mining", In *Proceedings of the Seventh ACM Conference on Learning @ Scale (L@S)*. 2020.

- Sahiti Labhishetty*, **Bhavya***, Kevin Pei*, Assma Boughoula, and Chengxiang Zhai. "Web of Slides: Automatic Linking of Lecture Slides to Facilitate Navigation.", In *Proceedings of the Sixth ACM Conference on Learning@ Scale (L@S)*. 2019.

- Sahiti Labhishetty*, **Bhavya***, Kevin Pei*, Assma Boughoula, and Chengxiang Zhai. "WOSView Demo: A Tool to Explore the Web of Slides." In *Proceedings of the Sixth ACM Conference on Learning@ Scale (L@S)*. 2019.

* *denotes equal contribution*

PATENT

Bhavya, Yu Deng, Md Faisal Mahbub Chowdhury, Paulina Toro Isaza, Michael Elton Nidd, Amar Prakash Azad, Harshit Kumar, and Larisa Shwartz. "*Automated Generation of Mitigation Information."* U.S. Patent Application 18/129,368. 2024.

BOOK

Chengxiang Zhai, Sean Massung, and **Bhavya**. "*Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining. Second Edition."* ACM Books. 2025.

WORK EXPERIENCE

Research Scientist AI| IBM @ Yorktown Heights *October 2024-Present*

Research Scientist Intern, AI| IBM @ Yorktown Heights| Mentor: Dr. Yu Deng *May 2023-Aug 2023*
- Developed LLM retrieval-augmented generation (RAG) pipeline for answering IT support questions

Research Intern, Hybrid Cloud| IBM @ Yorktown Heights| Mentor: Dr. Yu Deng *May 2022-Aug 2022*
- Explored semi-supervised learning for sequence tagging of IT tickets for mitigation guidance generation

Technical Advisor for Search| UI Health, Chicago *May 2021-Dec 2021*
- Mentored six BS and MS students with Dr. Karl Kochendorfer on developing a federated search engine based on Elasticsearch for physicians

Data Scientist| Gartner Inc. *Jul 2016-Jul 2018*
- **Summary:** Led the design and development of multiple Text Mining and Data Analytics platforms. Responsible for understanding business requirements, devising, and executing data science solutions, communicating complex technical models with the business leaders, and mentoring ten interns and junior employees.
- **Highlights:**
  - Deployed an in-house unified text mining platform thereby saving $200k/year on vendor expenses
  - Deployed a recommender system with Alternating Least Squares, Collaborative Filtering ensemble models. This resulted in an increase in all KPIs like click-through-rates, user retention by double digits

Part-time Data Analyst| Gartner Inc. *Oct 2015-Jun 2016*
- Lead intern for analyzing text data using techniques like Dependency Parsing, Hierarchical Topic Mining

EDUCATION

University of Illinois at Urbana-Champaign
- Ph.D. in Computer Science | Advisor: Dr. ChengXiang Zhai| GPA: 4.0/4.0 *Aug 2020-2024*
- MS (with Thesis) in Computer Science | Advisor: Dr. ChengXiang Zhai| GPA: 4.0/4.0 *Aug 2018-May 2020*
- BS in Computer Science with Honors | GPA: 3.63/4.0 *Aug 2016*

RESEARCH EXPERIENCE

SavvyHealth App| Drs. Jessie Chin, ChengXiang Zhai, Chung-Yi Chiu *Jun 2020-2024*

- Led the development of an intelligent mobile application to assist people with multiple sclerosis with self-management. App includes features such as web search and organization of search results, health tracking. Link to the app: https://testflight.apple.com/join/3sC7uUPF
- Conducted user studies on the app's feasibility and usability and found patients perceived the app as useful

EducationalWeb| Dr. ChengXiang Zhai *Aug 2018-2024*

- Led three students to develop a web-based educational platform that is being used by students in one course
- Received funding from the Office of the Vice Chancellor for Academic Affairs and Provost at UIUC

Query-driven Relation Extraction| Dr. Kevin Chang *Aug 2019-May 2020*

- Explored knowledge graph embeddings for online web-scale extraction and ranking of open-domain relations between entities of users' interests

Family Behavioral Support App| Drs. Karrie Karahalios, Hedda Meadan *Aug 2019-May 2020*

- Maintained and enhanced an iOS app that helps families implement Functional Assessment (FA)-based strategies to prevent challenging behaviors in young children
- Collaborated with the Special Education Dept. at UIUC, University of Washington, University of Vanderbilt
- Pushed multiple versions and analyzed log data for beta testing and Randomized-Controlled Trials (RCTs)

Neuroscience Image Processing| Dr. Justin Rhodes *Summer 2015, Jan-Mar 2017*

- Collaborated with neuroscience researchers for quantitative analysis of the impact of exercise on aging in mice
- Built desktop applications to analyze terabytes of brain slice images, such as curve fitting
- Expanded the team's analytical capabilities by 2x and reduced the analysis time by 30-100x

TEACHING EXPERIENCE

Teaching Assistant| Instructor: Dr. ChengXiang Zhai| Course: Text Mining *Spring 2019, Fall 2018, 2020*

- Head TA during Fall 2020 leading a team of 8 TAs and 400 undergraduate and graduate students
- Deployed course programming assignments on an in-house cloud-based virtual lab platform, integrated the platform with Coursera using Learning Tools Interoperability (LTI), designed a large-scale course project with peer-assisted grading
- Graded assignments, conducted office hours, answered questions on discussion forums for both online (on Coursera) and on-campus offerings

ACHIEVEMENTS AND OUTREACH ACTIVITIES

- Research Advisor, Data Inspired Young Analysts (DIYA) Research 2024, 2025
- Invited Panelist, The AI Effect: Evolving Roles and Skills in Technology Careers, ACM's Committee on Women in Computing (ACM-W) 2024
- Outstanding TA Award, UIUC *2021*
- Outstanding Employee, Gartner Inc. *2017-2018*
- Selected Technical Recruiter, Gartner Inc. *2017-2018*

PROFESSIONAL SERVICE

- Program Chair, Workshop on Secure and Safe AI Agents for Big Data Infrastructures, IEEE Big Data Conference 2025
- Organizer, Tutorial on An open platform for AI-driven systems research on cloud system management, Symposium on Operating Systems Principles (SOSP) 2025
- Reviewer, Association for Computational Linguistics (ACL) 2025
- Organizer, Shared Task on Long-Form Analogy Evaluation Challenge, GenChal, International 2024

Natural Language Generation Conference (INLG)

- Web Chair, Workshop on Knowledge Discovery and Data Mining in IT Operations, IEEE Big Data Conference 2022
- Reviewer, ACM SIGCSE 2021

SELECTED RESEARCH MENTORSHIP

Shradha Sehgal Jun 2023-May 2024

- Supervised MS thesis on "Educational analogy generation and extraction"

Victor Szabo Jan 2022-May 2022

- Development of a search engine for semantic retrieval of IT forum data. Student presented this work at the UIUC Illinois Scholars Undergraduate Research (ISUR) Poster Expo.