BHAVYA

e-mail: bhavya2@illinois.edu website: https://bhaavya.github.io/ 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 and healthcare. I'm experienced in various text mining and NLP techniques (e.g. large language models like GPT), building systems (e.g., web-based platforms, mobile apps) based on those algorithms, and conducting user studies for need-finding and evaluation. I also enjoy collaborating with stakeholders from multiple disciplines (e.g., health sciences, social science, IT) to carry out impactful research.

EDUCATION

University of Illinois at Urbana-Champaign

- Ph.D. in Computer Science | Advisor: Dr. ChengXiang Zhai| GPA: 4.0/4.0 *Aug 2020-2024 (expected)*
- 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*
- Relevant Coursework: Advanced NLP, Information Retrieval, Machine Learning, Data Mining, HCI for ML

PEER-REVIEWED PUBLICATIONS

- **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. (To Appear)
- **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): 1-25.
- 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)*, pp. 1-11. 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, pp. 1-25. 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), pp. 311-314. 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), pp. 1200-1206.* 2020.
- **Bhavya** and Chengxiang Zhai. "Explanation Mining", In *Proceedings of the Seventh ACM Conference on Learning @ Scale (L@S), pp. 321-324. 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)*, *pp. 1-4. 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)*, *pp. 1-3. 2019.*

* *denotes equal contribution*

ACHIEVEMENTS AND LEADERSHIP ACTIVITIES

- Outstanding TA Award, UIUC *2021*
- Outstanding Employee, Gartner Inc. *2017-2018*
- Selected technical recruiter for intern hiring, Gartner Inc. *2017-2018*

RESEARCH EXPERIENCE

Concept Extraction for IT Mitigation Guidance| IBM AIOps, Dr. ChengXiang Zhai *Sep 2021-Ongoing*

- Proposed prompting large pre-trained language models for extracting sequences of semantic concepts from IT support data in low resource settings

Analogy Generation| Drs. Jinjun Xiong (IBM), **ChengXiang Zhai** *Jun 2021-Ongoing*

- Proposed prompting large pre-trained language models for generating analogies to explain a concept. Conducted a study on Amazon Mechanical Turk, where participants found GPT-3 generated analogies to be as meaningful as human-written analogies
- Published a first-author paper at INLG, 2022

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

- Leading 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
- Presented a poster at HFES Health Care Symposium, 2021, where I reviewed and identified the limitations of existing consumer health search engines in fulfilling the informational needs of patients with chronic conditions and suggested potential AI techniques to fill the gaps

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

- Leading the development of a web-based educational platform called EducationalWeb, which is currently being used by students in one course
- Developed and deployed algorithms to link lecture slides for flexible navigation through slides of multiple courses and retrieve explanations of concepts on slides from textbooks. Students reported in surveys that they preferred this way of lecture slide navigation compared to other tools. Published findings in two short papers and one demo paper at the ACM Learning@Scale conference, 2019 and 2020
- Received funding from the Office of the Vice Chancellor for Academic Affairs and Provost at UIUC

Collaborative Caption Editing| Drs. ChengXiang Zhai, Lawrence Angrave, Yun Huang *Oct 2020-May 2022*

- Designed a collaborative editing activity where students corrected errors in automatically generated lecture video transcripts. Conducted student interviews and analyzed log data to understand their editing behavior
- Proposed leveraging ML to assist editors and evaluated the feasibility of ML models trained on student log data
- Led to a first-author publication at the Educational Technology Research and Development journal

Text Mining for Social Science| Drs. Alan Craig, Yu Zhang (CSU, Fresno), **ChengXiang Zhai** *Oct 2020-Aug 2021*

- Built an interactive web-based text mining system to analyze millions of historical newspaper articles including post-OCR correction, topic modeling, text annotation and classification, and embedding-based lexical analysis. System is being used by a few researchers studying how the concept of "juvenile delinquency" has been discussed in historical newspapers. Link to the system: https://newsanalysis.web.illinois.edu/
- Led to a second-author paper at ACM PACS, 2022

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

*all research advisors are at UIUC, unless mentioned otherwise

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. Published a paper in ACM L@S, 2021 as a case study describing the project design
- Designed a novel modular assignment and published the experience report in SIGCSE, 2020. The students created a useful search engine from scratch via the assignment
- Graded assignments, conducted office hours, answered questions on discussion forums for both online (on Coursera) and on-campus offerings with >300 students

WORK EXPERIENCE

Research Intern, Hybrid Cloud| IBM @ Yorktown Heights *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

Junior 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

SERVICE

- Peer-reviewer, ACM SIGCSE 2021
- Web Chair, Workshop on Knowledge Discovery and Data Mining in IT Operations, virtually co-located with IEEE Big Data Conference 2022

RESEARCH MENTORSHIP

Alexander Wang, Sarnabh Mukhopadhyay, Richa Meherwal Jun 2022-Ongoing

- Expansion and enhancement of the design and performance of the EducationalWeb system

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.

Zhaoyu Cheng Nov 2020-May 2021

- Development of the web-based text mining system for social science research

SKILLS

- **Programming Languages*:*** Python; Swift; JavaScript; SQL; Java; C++; C
- **Frameworks/Libraries*:*** ElasticSearch; Kibana; Apache Spark, Solr; Python Pytorch, scikit-learn, spaCy, NLTK, Gensim