

Rajarshi Haldar

PhD Candidate · Computer Science · University of Illinois Urbana-Champaign

Champaign, IL • rhaldar2@illinois.edu • [Google Scholar](https://scholar.google.com/citations?user=8YUW380AAAAJ) • github.com/rajarshihaldar • [linkedin.com/in/rhaldar](https://www.linkedin.com/in/rhaldar) • rajarshihaldar.com

RESEARCH INTERESTS

Evaluating Long-Form Text · Code Intelligence · Structure & Knowledge in NLP

EDUCATION

University of Illinois Urbana-Champaign, Urbana, IL Aug 2018 — Present

Ph.D. in Computer Science · Advisor: Julia Hockenmaier · Natural Language Processing

Indian Institute of Technology Kharagpur, Kharagpur, India Jul 2014 — May 2018

B.Tech. in Computer Science and Engineering

- Undergraduate research on knowledge graphs over the ACL Anthology (NAACL-HLT 2018 demo) and on venue recommendation for event-based social networks (IEEE TKDE).

RESEARCH EXPERIENCE

University of Illinois Urbana-Champaign, Urbana, IL Aug 2018 — Present

PhD Candidate & Graduate Research Assistant · Advised by Prof. Julia Hockenmaier

- Quantified how much of an LLM-as-a-judge benchmark ranking is evaluator noise rather than model quality (EMNLP 2025).
- Characterized where large language models succeed and fail at code summarization (LREC-COLING 2024).
- Introduced a multi-perspective neural architecture for semantic code search (ACL 2020; granted US patent).
- Built the dataset, training, and evaluation pipelines in PyTorch and Hugging Face Transformers.
- Designed a system to auto-grade code-reading assignments, collaborating with four researchers to devise a validated scoring rubric and to produce a dataset of code snippets, student responses, and quality scores (SIGCSE 2020).

Schlumberger, Menlo Park, CA Summer 2022

Digital Intern

- Built a knowledge-graph embedding model that learns from limited labeled data via labeling functions and Snorkel AI.
- Shipped a query-answering model on those embeddings that replaced the team's system at lower latency and overhead.

IBM Research, Yorktown Heights, NY Summers 2019 & 2020

Graduate Research Intern

- Led an automated tech-support agent that resolves IT queries by extracting events from unstructured incident text.
- Designed an unsupervised event-representation model and graph-based schema induction — both granted as US patents.
- Derived a novel event schema for tech support by analyzing an IT corpus with statistical modeling, then extracted entities, their states, and their relationships via event extraction.

TEACHING

Teaching Assistant, CS 447 — Natural Language Processing

- Designed a Google Colab coding assignment on machine translation and wrote its auto-grader on Gradescope.
- Held office hours twice a week and answered student questions on Campuswire.

University of Illinois Urbana-Champaign, Urbana, IL

Summers 2024 & 2025

Teaching Assistant, CS 416 — Data Visualization

- Graded a midterm and a final project in which students built dashboards in Tableau and D3.js.
- Held weekly office hours and answered student questions on Campuswire.

PUBLICATIONS

[7] **Rating Roulette: Self-Inconsistency in LLM-As-A-Judge Frameworks** Findings of EMNLP 2025

Rajarshi Haldar, Julia Hockenmaier

[6] **Analyzing the Performance of Large Language Models on Code Summarization** LREC-COLING 2024

Rajarshi Haldar, Julia Hockenmaier

[5] **A Multi-Perspective Architecture for Semantic Code Search** ACL 2020

Rajarshi Haldar, Lingfei Wu, JinJun Xiong, Julia Hockenmaier

[4] **Deep Learning Driven Venue Recommender for Event-Based Social Networks** IEEE TKDE 2020

Soumajit Pramanik, Rajarshi Haldar, Anand Kumar, Sayan Pathak, Bivas Mitra

[3] **A Validated Scoring Rubric for Explain-in-Plain-English Questions** SIGCSE 2020

Binglin Chen, Sushmita Azad, Rajarshi Haldar, Matthew West, Craig Zilles

[2] **Cluster Aware Mobility Encounter Dataset Enlargement** IWCMC 2019

Rajarshi Haldar, Salih Safa Bacanlı, Moayad Aloqaily, Adel Ben Mnaouer, Damla Turgut

[1] **CL Scholar: The ACL Anthology Knowledge Graph Miner** NAACL-HLT Demo 2018

Mayank Singh, Pradeep Dogga, Sohan Patro, Dhiraj Barnwal, Ritam Dutt, Rajarshi Haldar, Pawan Goyal, Animesh Mukherjee

PATENTS

Unsupervised Event Extraction US 12,001,896 B2

Rajarshi Haldar, Yu Deng, Lingfei Wu, Ruchi Mahindru, Shu Tao · IBM · Granted Jun 2024

Graph-Based Event Schema Induction for Information Retrieval US 11,615,152 B2

Rajarshi Haldar, Yu Deng, Lingfei Wu, Ruchi Mahindru, Julia Hockenmaier, Sinem Guven Kaya · IBM · Granted Mar 2023

Multi-Perspective, Multi-Task Neural Network Model for Matching Text to Program Code US 11,132,512 B2

Lingfei Wu, JinJun Xiong, Julia Hockenmaier, Rajarshi Haldar · IBM · Granted Sep 2021

SKILLS

Languages & Tools: Python, SQL, C, Git, Linux, LaTeX, Typst**ML & NLP:** PyTorch, Hugging Face Transformers, TensorFlow, scikit-learn**Research:** Large Language Models, LLM Evaluation, Code Intelligence, Information Retrieval, Knowledge Graphs

SERVICE

- Reviewer, Annual Meeting of the Association for Computational Linguistics (ACL)
- Reviewer, Conference on Empirical Methods in Natural Language Processing (EMNLP)

- Reviewer, North American Chapter of the ACL (NAACL)