P | K

PRIYANKA **KARGUPTA**

pk36@illinois.edu pkargupta.github.io pkargupta pkargupta

EDUCATION

University of Illinois, Urbana-Champaign
Ph.D. Student in Computer Science (2022-)
Overall GPA: 4.0/4.0
Advisor: Professor Jiawei Han

University of California, Berkeley
B.A. in Computer Science, Minor in Data Science:
Overall GPA: 3.773/4.0; Upper Division: 3.89/4.0
Advisor: Professor Ren Ng

EXPERIENCE

Data Mining Group: PhD Candidate

University of Illinois, Urbana-Champaign / 08/2022 – Current

• Funded by NSF Graduate Research Fellowship

• Designing models that exploit structured knowledge to enhance both their own critical thinking and human users', in both the education & scientific domains.

Berkeley AI Research: Research Assistant

University of California, Berkeley / 08/2020 - 05/2022

• Funded by the Intel SRC Research Scholar Program

• Developed multi-object tracking and segmentation methods for self-driving applications.

• Explored dependencies in coordinate-based neural networks, applied to optimized reconstruction of neural radiance fields.

IBM T.J. Watson Research Center: Research Intern

Efficient & Resilient Systems @ Yorktown Heights / 05/2021 - 09/2021

• Worked on collaborative detection and tracking for autonomous vehicles

• Led demo creation & production for the 2021 DARPA Electronics Resurgence Initiative Summit

NASA Ames Research Center: Research Intern

Data Sciences Group, Intelligent Systems Division / 05/2017 - 02/2018

• Designed and implemented deep image classification algorithms for swarms of drones in loosely coupled environments.
• *Poster presentation at the Ames Center symposium.*

Intel: Algorithms Research Intern & Product Manager

3DAT - Olympics Technology Group / 01/2022 - 07/2022

• Designed algorithms and optimize ML models for inferred 3D pose estimation for athlete tracking.

• Developed product roadmaps and manage the group's 2D skeletal product.

Principles and Techniques of Data Science: Student Instructor

University of California, Berkeley / 01/2021 - 05/2022

• Taught discussion and lab sections of 30 students each, hold office hours, and assist with various aspects of running the course of over 900 students.

Vyncs (under Agnik): Project Manager

05/2018 - 12/2021 (Full-time summers, part-time school year)

• Led development of sensor-based on-demand location detection and sharing mobile application w/ 70,000+ users.

• Led design & development of an Android augmented reality app which renders visual location data (e.g. nearby restaurants, traffic accidents) via user's phone camera.

AWARDS

• 2024 NSF Graduate Research Fellowship
• 2020-2021 Intel Semiconductor Research Corp. Research Scholar
• 2021 Computer Graphics (CS184) Showcase Winner
• 2020 Global Top 15 Winner for Health & Wellness: The Global Hack (12,000+ participants from 100+ countries)
  ◦ Development Team Lead
• 2018 National Center for Women & IT (NCWIT) Maryland Winner
• 2018 NCWIT National Honorable Mention (Top 10% out of 3,600 applications)
• 2017 National Top 15 Company & JA Maryland Company of the Year
• 2016 Rep. Elijah Cummings's Congressional App Challenge Winner
• 2016 MIT Media Lab/QuHacks Hackathon Winner: 1st Place

SELECTED PUBLICATIONS

Instruct, Not Assist: Improving Socratic Questioning for Active Multi-Turn Code Debugging
Priyanka Kargupta, *Ishika Agarwal, Dilek Hakkani-Tur, Jiawei Han*
• *EMNLP'24 Findings; Poster @ NeurIPS'24 Workshop on Educational Assessment*

• Designing a framework to re-orient a language model towards instructor-like responses and questions for helping users debug code in an educational domain. Guides conversations using a state-representation-based dynamic tree.

MEGClass: Text Classification with Extremely Weak Supervision via Mutually-Enhancing Text Granularities
Priyanka Kargupta, *Tanay Komarlu, Susik Yoon, Xuan Wang, Jiawei Han*
• *EMNLP'23 Findings (Poster)*

• Designed an extremely weakly-supervised text classification method that leverages multiple, mutually-enhancing text granularities. Learns a contextualized document representation that captures the most discriminative class indicators.

Extensive experiments on seven benchmark datasets demonstrate that MEGClass outperforms other weakly and extremely weakly supervised methods by even 7% and 15x faster than the SOTA.

REACTION MINER: An Integrated System for Chemical Reaction Extraction from Textual Data
Ming Zhong, Siru Ouyang, Yizhu Jiao, ***Priyanka Kargupta****, Leo Luo, Yanzhen Shen, Bobby Zhou, Xianrui Zhong, Xuan Liu, Hongxiang Li, Jinfeng Xiao, Minhao Jiang, Vivian Hu, Xuan Wang, Heng Ji, Martin Burke, Huimin Zhao, Jiawei Han*
• ***EMNLP'23 Demo Track***

• Designed a system specifically designed to interact with raw scientific literature and automatically extract precise and more informative chemical reactions.

Our segmentation module beat GPT-4 by improvements of 16.2% and 9.1% for two different metrics.

Analyzing Driving Data using the ADAPT Distributed Analytics Platform for Connected Vehicles
Hillol Kargupta, ***Priyanka Kargupta***
• ***2019 IEEE International Conference on Data Science and Advanced Analytics (Oral Presentation)***

• Designed a randomized contraction mapping algorithm that exploits self-similarity, in order to generate a lower-dimensional, covariance-preserving representation.

ONGOING PUBLICATIONS

Unsupervised Episode Detection for Large-Scale News Events (*Lead) - Submitted to KDD'25*

• Proposing a new task, episode detection, which seeks to detect episodes from a news corpus containing key event articles.
• Designed unsupervised framework which automatically partitions a key event corpus into cohesive episode sequences.

Tree-of-Debate: Multi-Persona Debates Elicit Critical Thinking for Scientific Comparative Analysis (*Lead*)

• Determine the novelties, similarities, and equivalent ideas within papers through the reasoning induced by author personas debating. Structures the debates as a tree of rounds, augmented with fine-grained retrieved paper segments.

Corpus-Explorative Taxonomy Expansion with Large Language Models (*Lead*)

• A self-supervised framework which expands a taxonomy based on whether a corpus indicates that a node needs further *exploration*. Proposes a joint enrichment-classification-expansion method.

SKILLS

• ***Programming:*** Python, Java, HTML5/CSS, C, SQL, JavaScript, Android Development, C++, C#, R
• ***Development Tools:*** PyTorch, Transformers, HuggingFace, scikit-learn, scikit-image, OpenCV, AWS (EC2, S3, Lambda), Alexa Skill Development, Spark, GeoSpark, Pandas, Plotly, Caffe, .NET, Linux, Firebase, Git, Docker
• ***Graphic & Video Design:*** Blender***,*** Adobe Photoshop, Adobe Illustrator, Adobe Premiere Pro, Camtasia