

Upasana Dutta

Profile

Ph.D. student in Computer and Information Science, with expertise in data science, applied machine learning, natural language processing, computational social science, and network analysis. Experienced in conducting interdisciplinary research with large-scale longitudinal data using large language models, statistical modeling, and social network analysis. Seeking to apply advanced analytical skills to high-impact real-world challenges.

Education

- 08/2022 - 05/2027 **PhD in Computer and Information Science, University of Pennsylvania** GPA: 3.92/4.0
Advisors: Duncan Watts, Aaron Clauset
- 2019-2022 **Master of Science in Computer Science, University of Colorado Boulder** GPA: 4.0/4.0
- 2015-2019 **B.Tech in Computer Science and Engg., Heritage Institute of Technology, India** GPA: 9.36/10

Internship

- May-August 2025 **Machine Learning SWE Intern**
Meta, Menlo Park, CA
Team: Profile Relevance (Facebook)
Developed end-to-end ranking models for Facebook users' friend-lists across four view types (self, friend, friend-of-friend, non-friend-of-friend), beating production model AUC by 18.3%; ran A/B tests showing significant engagement gains; launched model to production for 3 billion Facebook users.

Publications and Research

- 2025 **Asymmetric Framing Strategies Drive Partisanship in Television News** (under review)
Upasana Dutta, Homa Hosseinmardi, Amir Ghasemian, Aaron Clauset, Duncan Watts
Leveraged student-teacher model setup for training neural networks on LLM labels for detecting partisanship and framing strategies, showing media bias in 400,000+ TV news episodes (2013-2025).
- 2025 **LLM-Simulated Experimental Analysis of Ideological Bias in News** [Extended Abstract]
Upasana Dutta, Josh Nguyen, Samar Haider, Bryan Li, Neil Fasching, Duncan Watts
Ran experiments using LLM simulations with GPT-4o and LLaMA-3.1 across multiple treatment arms, showing differences in narrative bias across four U.S. national news outlets.
- 2025 **Dynamics of Ingroup and Outgroup Narratives on Telegram** (under review)
Upasana Dutta, Yasmine Hourri
Conducted multi-channel analysis of 300,000 Telegram messages to quantify the prevalence of ingroup and outgroup narratives and statistically comparing its propagation in direct vs. forwarded messages.
- 2025 **News-like Information Ecosystem on YouTube** [Extended Abstract]
A. Ghasemian, H. Hosseinmardi, U. Dutta, J. Allen, D. Rothschild, D. Watts
Analyzed 385,903 YouTube videos using LLM classifications and unsupervised clustering, showing higher user engagement with non-traditional content creators than traditional media outlets for news.
- 2025 **LLM-Assisted Recommendations: Impact on Human-AI Ecosystems** [Extended Abstract]
A. Ghasemian, H. Hosseinmardi, U. Dutta, J. Allen, D. Rothschild, D. Watts
Audited an LLM-assisted recommender system for YouTube, finding that LLM-driven personalization may promote harmful videos; and proposed a prompt-engineered solution to mitigate this risk.
- 2025 **Political Slant and Causal Overreach in the Behavioral Sciences** (under review)
Calvin Isch, Upasana Dutta, Cory Clark, Phil Tetlock, Eric Uhlmann
Analyzed 194,606 behavioral science research articles using LLM classifications for political slant and causal claims, finding that politically slanted papers are more likely to make overreaching causal claims.
- 2024 **Comparison of Real-World Network Scaling with Random Graph Models** (in preparation)
Upasana Dutta, Alexander Ray, Aaron Clauset
Analyzed the scaling trends of 254 real-world networks and assessed the extent to which null models, such as the configuration model and the stochastic block model, explain their scaling behavior.

- 2024 **Algorithmic Systems, Human Agency, and the Future of Platform Research** (in preparation)
H. Hosseinmardi, [Upasana Dutta](#), David Rothschild, Duncan Watts
A perspective piece challenging the narrative of algorithmic curation by recommendation systems, highlighting that user's self-selection instead plays a critical role in their online content consumption.
- 2023 **Sampling random graphs with specified degree sequences** [[Publication](#)] [[Code](#)] [[Python Package](#)]
[Upasana Dutta](#), Bailey K. Fosdick, Aaron Clauset
Published in the **Journal of Computational and Graphical Statistics 2025**
Developed a method for sampling networks from the degree-preserving configuration model by detecting convergence in a double-edge swap Markov chain sampler.
- 2021 **Analyzing Twitter Users' Behaviour Before and After Contact by Russia's Internet Research Agency** [[Publication](#)] [[Code](#)] [[CU Boulder Today](#)]
[Upasana Dutta](#), R. Hanscom, J. Zhang, R. Han, T. Lehman, Q. Lv, S. Mishra
Published in the **Proceedings of the ACM on Human-Computer Interaction, CSCW 2021**
Employed statistical methods to analyze changes in user behavior on Twitter after they engaged with Russian bots backed by the Internet Research Agency (IRA) during the US 2016 presidential election.

Technical Projects

- Spring 2024 **Causal Effects of Exposure to Ideologically-biased TV news** [[Slides](#)]
Developed causal model to detect if partisan TV exposure affects future Fox News/MSNBC viewing.
- Fall 2023 **Evaluating Performance of LLMs on Gun Violence Database** [[Report](#)] [[Presentation](#)]
Evaluated LLMs (zero/few-shot T5, BERT, GPT-3, GPT-4) for gun violence reports data extraction.
- Fall 2022 **Restuarant Recommender System** [[Report](#)]
Analyzed ML models (KNN, matrix factorization, neural networks) for restaurant recommendations.
- Spring 2021 **Statistical Analysis on Rural Underrepresentation in CU Boulder** [[Slides](#)] [[Report](#)] [[Video](#)]
Leveraged Data Analytics to evaluate underrepresentation of small/rural communities in CU Boulder.
- Fall 2019 **Study of user activity on Question-Answering Platform : Stack Exchange** [[Slides](#)] [[Report](#)]
Analyzed user activity on Q/A platform for latent community structure with Stochastic Block Modeling.
- Fall 2019 **Movie Recommendation System : Kaggle Dataset** [[Report](#)]
Utilised user demographics and film-viewing preferences to recommend new movies to users.

Awards

- Nov' 2024 **GAPSA Research Grant**, University of Pennsylvania
- Apr' 2022 **Abel Lukens Stout Fellowship**, University of Pennsylvania
- Apr' 2022 **Bell Foundation Outstanding Research Award**, Department of Computer Science, CU Boulder
- Mar' 2022 **CS Annual Research Expo Award**, Department of Computer Science, CU Boulder [[Poster](#)]
- Dec' 2021 **NCWIT Collegiate Award 2022 Finalist** [[Video](#)]
- Feb' 2021 **CS Publication Recognition Award**, Department of Computer Science, CU Boulder [[Publication](#)]

Skills

Data Science: Hypothesis testing, Multivariate Analysis, Statistical Inference, Sampling Methods, Anomaly Detection, Time-Series Analysis

Machine Learning: Regression Analysis, Supervised/Unsupervised Learning, Ensemble Methods, Model Training and Fine-Tuning, Model Evaluation, Active Learning, Deep Learning

NLP: Text Preprocessing, Topic Modeling, Transformers, Embeddings, Large Language Models (LLMs), Retrieval-Augmented Generation (RAGs), LLM Agents, Prompt Engineering

Network Analysis: Centrality Measures, Null models, Community detection, Cascade models

Causal Inference: Experimental Designs, A/B testing, RCTs, Diff-in-Diff, Instrumental Variables

Toolkits: PyTorch/Tensorflow, Huggingface, pandas, scikit-learn, matplotlib, seaborn

Computer Languages: Python, SQL, R