

| | | |
|--------------------|---|---|
| CONTACT | ✉ dr.debjyotipaul@gmail.com in /in/debjyotipaul385/ | 🏠 www.debjyotipaul.in 🌐 www.github.com/debjyoti385 |
| RESEARCH INTERESTS | Automatic Speech Recognition System, Natural Language Processing and Sentiment Analysis, Graph Representation Learning, Spatio-Temporal Public Health Analysis. | |
| WORK EXPERIENCE | <p>[March 2020 - Current] Meta, Inc., Research Scientist, Menlo Park, United States Working in the AI Speech team responsible for creating, improving, and deploying state-of-the-art Automated Speech Recognition, Inverse Text Normalization and Punctuation-Capitalization models.</p> <p>[Summer 2019] Alibaba Group., AI Research Intern, Sunnyvale, United States Workload-aware database tuning with Machine Learning.</p> <p>[Summer 2018] Facebook, Inc., Research Scientist Intern, Seattle, United States Improving metrics for Facebook's Search Ranking via rated label inferences with Machine Learning.</p> <p>[Summer 2017] Amazon AI Research, Research Intern, New York, United States Hyperparameter optimization techniques on MXNET.</p> <p>[2013-2015] Flipkart, Data Engineer, Bangalore, India Served as data engineer at data platform team <i>flipkart.com</i>, India's biggest e-commerce company. Facilitated scalable environment for Big data analytics with processing pipelines for batch and stream.</p> | |
| EDUCATION | <p>[2015-2020] Ph.D., University of Utah, School of Computing, Salt Lake City, Utah, USA Major : Computer Science & Engineering Advisor : <i>Prof. Feifei Li</i> Thesis : <i>Data-Driven Spatio-Temporal Analysis for Multimode Data</i> GPA : 3.98/4.0</p> <p>[2011-2013] Masters of Technology, Indian Institute of Technology Kanpur, Kanpur, India Major : <i>Computer Science & Engineering</i> Advisor : <i>Late Professor Sanjeev K Aggarwal</i> Thesis : <i>Multi-constraint Job scheduling in Grid Computing</i> GPA : 8.67/10.0 (Rank: 3)</p> <p>[2007-2011] Bachelors of Technology, West Bengal University of Technology, Kolkata, India Major : <i>Computer Science & Engineering</i> College : <i>Institute of Engineering and Management</i> CGPA : 8.93/10 (Rank: < 10)</p> | |
| THESES | <p>[2020] Ph.D. Thesis, Advisor: Dr. Feifei Li <i>Data-Driven Spatio-Temporal Analysis for Multimode Data.</i> The objective of this thesis is to build a framework that enables a data-driven approach for spatio-temporal analysis for multimode data with AI models and facilitate workload-aware support for effective large-scale processing.</p> <p>[2013] M.Tech Thesis, Advisor: Late Dr. Sanjeev Kumar Aggarwal <i>Multi-constraint Job scheduling in Grid Computing</i> The objective is to efficiently schedule Jobs on Grid Computing to achieve maximum utilization of resources with energy efficient approach. Modeling real world computing & storage grid on Multi-objective Evolutionary Algorithm satisfying hard constraints on jobs constraints, resources constraints, and soft constraints on cost, and energy consumption.</p> | |
| PUBLICATIONS | <ul style="list-style-type: none"> + Improving Data Driven Inverse Text Normalization using Data Augmentation, Laxmi Pandey, Debjyoti Paul, Pooja Chitkara, Yutong Pang, <i>ACL ARR 2022</i> (In Review). + Database Workload Characterization with Query Plan Encoders, Debjyoti Paul, Jie Cao, Feifei Li, Vivek Srikumar, To appear in <i>48th International Conference on Very Large Data Bases (VLDB 2022)</i>. + Semantic Embedding for Regions of Interest, Debjyoti Paul, Feifei Li, Jeff M. Phillips, <i>The International Journal on Very Large Data Bases (VLDBJ 2021)</i>. DOI: 10.1007/s00778-020-00647-0. | |

- + Bursty Event Detection Throughout Histories, **Debjyoti Paul**, Yanqing Peng, Feifei Li, *35th IEEE International Conference on Data Engineering (ICDE 2019)*, Macau, China, 2019. DOI: [10.1109/ICDE.2019.00124](https://doi.org/10.1109/ICDE.2019.00124)
- + Compass: Spatio Temporal Sentiment Analysis of US Election, **Debjyoti Paul**, Feifei Li, Murali Krishna Teja, Xin Yu, Richie Frost, *What twitter says!, 23rd SIGKDD Conference on Knowledge Discovery and Data Mining (SIGKDD)*, Aug 13-17, 2017, Halifax, Canada. DOI: [10.1145/3097983.3098053](https://doi.org/10.1145/3097983.3098053).
- + AI Pro: Data Processing Framework for AI Models, **Debjyoti Paul**, Richie Frost, Feifei Li, *35th IEEE International Conference on Data Engineering (ICDE)*, Macau, China, 2019.
- + Geotagged US Tweets as Predictors of County-Level Health Outcomes, 2015–2016, Quynh C. Nguyen, Matt McCullough, Hsien-wen Meng, **Debjyoti Paul**, Dapeng Li, *American Journal of Public Health (AJPH)*, September, 2017, DOI: [10.2105/AJPH.2017.303993](https://doi.org/10.2105/AJPH.2017.303993).
- + Social media indicators of the food environment and state health outcomes, Quynh C. Nguyen, Hsien-wen Meng, Dapeng. Li, Matt McCullough, **Debjyoti Paul**, Kanokvimankul. P, Nguyen. T, Li. Feifei, *American Public Health Association (APHA)*, 148, 120-128., 2017, DOI: [10.1016/j.puhe.2017.03.013](https://doi.org/10.1016/j.puhe.2017.03.013).
- + Multi-objective Evolution based Dynamic Job Scheduler in Grid, **Debjyoti Paul**, Sanjeev K. Aggarwal, *The 8th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS)*, IEEE, July 2nd - 4th, 2014, Birmingham, UK. DOI: [10.1109/CISIS.2014.50](https://doi.org/10.1109/CISIS.2014.50).
- + RCached-tree: An Index Structure for Efficiently Answering Popular Queries, Manash Pal, Arnab Bhattacharya, **Debjyoti Paul**, *ACM International Conference on Information and Knowledge Management (CIKM 2013)*, Oct. 27 - Nov. 1, 2013, San Francisco, CA, USA. DOI: [10.1145/2505515.2507817](https://doi.org/10.1145/2505515.2507817).
- + Lightweight Security Enhancement Protocol for Radio Frequency Identification (RFID), **Debjyoti Paul**, Sumana Basu, Punit Beriwal, *IEEE SPSITM International Conference 2011*, Kolkata, India.

POSITION OF RESPONSIBILITIES

- [2021] Program Committee Member for 47th International Conference on Very Large Data Bases (VLDB) 2021, Industrial Track.
- [2021] Program Committee Member for 37th IEEE International Conference on Data Engineering (ICDE) 2021, Industrial Track.
- [2018-2021] Reviewer: IEEE TKDE 2018, IEEE TKDE 2019, IEEE ICDE 2021, PVLDB 2021.
- [2016-2019] External Reviewer: EDBT 2016, PVLDB 2019.
- [2015-2018] Teaching Mentor: Natural Language Processing (CS6340), Data Mining (CS 6140) at the University of Utah. USA.
- [2012-2013] Member and organizer of SIGDATA, a Special Interests Group on data management and mining at the IIT Kanpur, India.
- [2011-2013] Teaching assistant: Software Engineering (CS455), Fundamentals of Computer Science (ESC101) at IIT Kanpur.

HACKATHON TROPHIES

- | | |
|----------------------|--|
| [2017] HackTheU | HackTheU MLH Hackathon Winner (around 40 teams) |
| [2016] EMC2 Code | Mars Challenge Hackathon Winner (around 45 teams) |
| [2015] Goldman Sachs | Air Quality Hackathon 1st Runners-up (around 30 teams) |
| [2014] InMobi | Freedom Hack Worldwide Hackathon 1st Runners-up (around 160 teams) |
| [2013] Yahoo | Yahoo HackU 2013 Hackathon Winner (around 40 teams) |

ACHIEVEMENTS

- [2015-2020] **Ph.D. Fellowship** at University of Utah.
- [2012] **All India Rank 7** in Indian Space Research Organization (ISRO) Junior Research Fellow Exam.
- [2012] Secured **All India Rank 228** out of 160,000+ participants in Graduate Aptitude Test for Engineering (GATE) 2012, Computer Science.
- [2011] Secured **All India Rank of 7** in Indian Space Research Organization (ISRO)'s Junior Research Fellow exam out of 200,000+ candidates.
- [2011] Secured **All India Rank 223** out of 150,000+ participants in Graduate Aptitude Test for Engineering (GATE) 2011, Computer Science.