







CONTACT  dr.debjyotipaul@gmail.com  [debjyotipaul385](#)  [xSy7SxUAAAAJ](#)
 [debjyotipaul.in](#)  [debjyoti385](#)  [debjyoti-paul](#)

RESEARCH INTERESTS Speech understanding models, large language models, representation learning, spatial-temporal data sets, database optimization with AI

WORK EXPERIENCE **[2020 - Present] Meta, Inc.**, Menlo Park, United States
[2025 - Present] Staff Research Scientist
[2022 - 2025] Senior Research Scientist
[2020 - 2022] Research Scientist
 Working towards advancements in speech technologies and multimodal large language models, includes hands-on participation in training and refinement of foundational models such as herd of Llama models and sophisticated Automatic Speech Recognition (ASR) systems. The work requires expertise in architecture exploration and performance optimization.

[Summer 2019] Alibaba Group., AI Research Intern, Sunnyvale, United States
 Workload-aware database optimization with AI.

[Summer 2018] Facebook, Inc., Research Scientist Intern, Seattle, United States
 Improving metrics for Facebook's Search Ranking via rated label inferences with Machine Learning.

[Summer 2017] Amazon AI Research, Research Intern, New York, United States
 Hyperparameter optimization techniques on MXNET.

[2013-2015] Flipkart, Data Engineer, Bangalore, India
 Served as data engineer at data platform team *flipkart.com*, India's biggest e-commerce company.
 Facilitated scalable environment for Big data analytics with processing pipelines for batch and stream.

EDUCATION **[2015-2020] Ph.D., University of Utah, School of Computing**, Salt Lake City, Utah, USA
 Major : Computer Science & Engineering
 Advisor : *Prof. Feifei Li*
 Thesis : *Data-Driven Spatio-Temporal Analysis for Multimode Data*
 GPA : *3.96/4.0*

[2011-2013] Masters of Technology, Indian Institute of Technology Kanpur, Kanpur, India
 Major : *Computer Science & Engineering*
 Advisor : *Late Prof. Sanjeev K Aggarwal*
 Thesis : *Multi-constraint Job scheduling in Grid Computing*
 GPA : *8.67/10.0 (Rank: 3)*

[2007-2011] Bachelors of Technology, West Bengal University of Technology, Kolkata, India
 Major : *Computer Science & Engineering*
 College : *Institute of Engineering and Management*
 CGPA : *8.93/10 (Rank: < 10)*

THESES **[2020] Ph.D. Thesis, Advisor: Dr. Feifei Li**
Data-Driven Spatio-Temporal Analysis for Multimode Data.
 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.

[2013] M.Tech Thesis, Advisor: Late Dr. Sanjeev Kumar Aggarwal
Multi-constraint Job scheduling in Grid Computing
 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.

PUBLICATIONS

- + VowelPrompt: Hearing Speech Emotions from Text via Vowel-level Prosodic Augmentation, Yancheng Wang, Osama Hanna, Ruiming Xie, Xianfeng Rui, **Debjyoti Paul**, Maohao Shen, Xuedong Zhang, et al. *ICLR 2026, April 23–27, 2026*, Rio de Janeiro, Brazil.
- + R2S: Representation Learning for Speech Recognition with Guidance from Synthetic Data, Minh Tran, **Debjyoti Paul**, Yutong Pang, Laxmi Pandey, Kevin Jiang, Jinxi Guo, Ke Li, Shun Zhang, Xuedong Zhang, Xin Lei. *Interspeech 2025*, Rotterdam, The Netherlands.
- + A Domain Adaptation Framework for Speech Recognition Systems with Only Synthetic data, Minh Tran, Yutong Pang, **Debjyoti Paul**, Laxmi Pandey, Kevin Jiang, Jinxi Guo, et al. *IEEE ICASSP*, 6-11 April 2025, Hyderabad, India.
- + LLaMA based Punctuation Restoration With Forward Pass Only Decoding, Yutong Pang, **Debjyoti Paul**, Kevin Jiang, Xin Lei, Xuedong Zhang CS arXiv:2408.11845 (2024).
- + Towards Scalable Efficient On-device ASR with Transfer Learning, Laxmi Pandey, Ke Li, Jinxi Guo, **Debjyoti Paul**, Arthur Guo, Jay Mahadeokar, Xuedong Zhang. CS arXiv:2407.16664 (2024).
- + Language Agnostic Data-Driven Inverse Text Normalization, Szu-Jui Chen, **Debjyoti Paul**, Yutong Pang, Peng Su, Xuedong Zhang. *Interspeech 2023*, Dublin, Ireland, August 2023.
- + Recovering from Privacy-Preserving Masking with Large Language Models, Arpita Vats, Zhe Liu, Peng Su, **Debjyoti Paul**, Yingyi Ma, Yutong Pang, Zeeshan Ahmed, Ozlem Kalinli. (*In Review.*)
- + Language Agnostic Data-Driven Inverse Text Normalization, Szu-Jui Chen, **Debjyoti Paul**, Yutong Pang, Peng Su, Xuedong Zhang. *Interspeech 2023*, Dublin, Ireland, August 2023.
- + Improving Data Driven Inverse Text Normalization using Data Augmentation and Machine Translation, **Debjyoti Paul**, Yutong Pang, Sray Chen, Xuedong Zhang, *Interspeech 2022 Show and Tell*, Incheon, Korea, September 2022.
- + Improving Data Driven Inverse Text Normalization using Data Augmentation, Laxmi Pandey, **Debjyoti Paul**, Pooja Chitkara, Yutong Pang, Xuedong Zhang et al. CS arXiv:2207.09674 (2022).
- + Database Workload Characterization with Query Plan Encoders, **Debjyoti Paul**, Jie Cao, Feifei Li, Vivek Srikumar, *48th International Conference on Very Large Data Bases (VLDB 2022)*, Sydney, Australia, 2022. DOI: 10.14778/3503585.3503600
- + Semantic Embedding for Regions of Interest, **Debjyoti Paul**, Feifei Li, Jeff M. Phillips, *The International Journal on Very Large Data Bases (VLDBJ 2021)*. 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
- + 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.
- + 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. DOI: 10.1109/ICDE.2019.00219.
- + 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.
- + 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.
- + 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.

POSITION OF
RESPONSIBILITIES

- [2016-2025] Reviewer: ISCA Interspeech (2023, 2024, 2025), ICASSP (2023, 2024, 2025), IEEE ASRU (2024 2025), IEEE ICDE 2021, PVLDB (2019, 2021), IEEE TKDE (2018, 2019, 2020), EDBT 2016,
- [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.
- [2015-2019] Teaching Mentor and Lecture: Natural Language Processing (CS6340), Data Mining (CS 6140), Database Systems (CS 6530) 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

- [2012] Secured **All India Rank 228** out of 160,000+ (99.86 percentile) participants in Graduate Aptitude Test for Engineering (GATE) 2012.
- [2011] Computer Science. Secured **All India Rank 223** out of 150,000+ (99.85 percentile) participants in Graduate Aptitude Test for Engineering (GATE) 2011, Computer Science.
- [2011] **All India Rank 7** (99.99 percentile) in Indian Space Research Organization (ISRO) Junior Research Fellow Exam.