Gauranga Kumar Baishya

Chennai Mathematical Institute (CMI)

Whitefield, Bangalore, Karnataka, India

✉ gaurangakumarbaishya@gmail.com • https://gaurangakrb.github.io/

https://www.linkedin.com/in/gauranga-kumar-baishya-4a12731ba/

Education

Program	Institution/Board	%/CGPA	Year
M.Sc. (Data Science)	Chennai Mathematical Institute (CMI) *Chennai, Tamil Nadu*	**8.50/10**	2023-25
B.Sc.(Mathematics)	Tezpur Central University (TU) *Napaam, Tezpur, Assam*	**80.20%**	2019-22
Higher Secondary (Science)	Shrimanta Shankar Academy (SSA) *Dispur, Guwahati, Assam*	**87.60%**	2016-18
High School	Shrimanta Shankar Academy (SSA) *Dispur, Guwahati, Assam*	**10/10**	2015-16

Key Projects & Experience

1. *Project 1* **May 2024 - present**

Guide: Prof. Yu-Hua Tseng *Harvard Medical School, Brookline, MA, USA*

- ML-based deconvolution pipeline for bulk RNA-seq of deep & superficial cervical adipose tissue, modeling cell-type composition and expression signatures to resolve depot-specific transcriptional variance & differential pathways. †

2. *Project 2* **May 2025 - present**

Guide: Prof. Amitava Das *Birla Institute of Technology (BITS) Pilani, Goa*

- Developed differential geometry-based framework for transformers that, across layers, measures (i) *curvature and path length* of the model's predictive distribution via Fisher–Rao on logits, (ii) *gauge-invariant parallel transport of hidden states*, and (iii) attention as a *depth-indexed transport operator*, producing task-aligned gradient fields that localize where and how the network reorganizes *its beliefs*. Project page

3. *Project 3* **Aug 2022 - Apr 2023**

Guide: Prof Sandeep Choubey & Prof. Dhiraj Hazra *Institute of Mathematical Sciences (IMSc.), Chennai*

- Modeled cross-omic dependencies in single-cell hematopoietic differentiation by applying *Gaussian Process Regression* and *hierarchical Bayesian inference* with *MCMC sampling* to learn nonlinear covariation structures among DNA, RNA, and protein modalities across temporal developmental trajectories. Github Repo

Industrial Training

1. Lucid Systematic Literature Review (SLR) engine **Aug 2025 - present**

Dr. Amrita Ostawal, CEO *Arete Access*

- Developing an AI-driven SLR screening model that uses domain-tuned transformers for inclusion/exclusion, RAG over guidelines, & structured information extractors to generate audit-ready, regulatory-grade evidence. †

2. Modeling of Cadmium Eco-toxicity **Sep 2025 - present**

Dr. Sunil Nagpal & Mrs. Mitali Merchant *Tata Research, Development & Design Centre (TRDDC)*

- Modeling of cadmium ecotoxicity using probabilistic graphical models to identify and predict toxicological pathways and environmental risk patterns. †

Key Course Projects

1. *Course: Applied Machine Learning* **Jan-Apr 2025**

Instructor: Dr. Raghav Kulkarni, Principal Data Scientist, Success in Cloud Inc., USA *CMI*

- Empirical benchmark of *classical ML* vs *modern DL* (including custom ANN and BERT/DistilBERT) for Goodreads review-popularity prediction, showing boosting can rival transformer recall under severe class imbalance. Github Repo

2. *Course: Guided project/Causal Inference & Modelling* **Jan-Apr 2025**
Prof. M. R. Srinivasan, Visiting faculty, CMI *CMI*

- Causal inference framework integrating *potential outcomes framework*, *DAG-based structural equations modeling*, and *DoubleML orthogonalization* to estimate race-conditioned effects in COVID-19 telemedicine adoption using inverse propensity weighting and counterfactual simulation. Github Repo

3. *Course: Natural Language Processing* **Aug-Nov 2024**
Instructor: Prof. Ramaseshan Ramachandran, Visiting faculty, CMI *CMI*

- Integrated computational linguistics & representation learning framework combining statistical text analysis, probabilistic semantics, and deep sequence modeling - from *Zipf–Heaps* corpus characterization to *COALS* and skip-gram embeddings through RNN/LSTM-driven abstract generation on ≈ 50,000 COVID-19 scientific articles. Github Repo

4. *Linear Algebra & its applications* **Jan-Apr 2024**
Instructor: Prof. Priyavrat Deshpande, faculty, CMI *CMI*

- Implemented graph-theoretic image segmentation using spectral clustering; formulating *normalized cut* via Laplacian eigen decomposition & integrating SLIC superpixels with RAG-based recursive partitioning to achieve perceptual grouping; based on a breakthrough paper by *Shi & Malik*. Github Repo

Online Courses

- Deeplearning.AI: *Tensorflow Developer Specialization*, *Deeplearning Specialization*, *Machine Learning Specialization*
- Coursera: *Probabilistic Graphical Models Specialization*, *Introduction to Genomic Data Science*
- Others: *Deepmind x UCL Introduction to RL by David Silver*, *Introduction to Large Language Models (NPTEL)*, *Introduction to Synthetic Biology (IIT Madras)*, *Reinforcement Learning (NPTEL)*

Course Work

Key Courses **August 2019-April 2022**
(Core) *Tezpur University*

- Real Analysis - I&II, Calculus - I&II, Introductory ODE&PDE, Linear Algebra - I, Physics - I,II,III, Introductory Statistics and Probability, Introduction to Optimisation, Numerical Methods and Boolean Algebra, Scientific Computing

Key Courses **August 2023-April 2025**
(Core and electives) *CMI*

- Programming & Data Structures with Python, RDBMS & SQL, Linear Algebra & its Applications, Regression & Classification, Algorithm Design Techniques, Advanced Machine Learning, NLP, Multivariate Statistics, Applied Machine Learning

Technical Skills

- *OS*: Windows, MAC, Linux
- *Programming Language*: C, R, Matlab, Python
- *Tools*: LaTeX, Anaconda, Adobe, Git, MS Office
- *Database System*: MySQL

Positions of Responsibility, Volunteering & Social Work

- *Course Instructor:* Mathematics for Machine Learning at *Upmenta* (2025 - present).
- *Managing Director* of Non-Profit Organisation *Sankhya* (2024 - 2025).
- *Lead AI/ML Strategy* at *Arete Access*. (2024 - present).
- *Academic mentor* at *Topmate.io top 1%* (2024 - present).
- *Placement Co-ordinator* for *M.Sc. Data Science Course, Chennai Mathematical Institute (CMI)* (2023-24).
- *Course instructor:* Mathematics, Statistics & CS for *Cheenta School of Statistics & Data Science, India* (2023-25).
- *Event Co-ordinator* at *Polymath Jr. REU, US* (Jun - Aug 2021).

Workshops and Conferences Attended

- Attended *Data Science, Probabilistic & Optimisation Methods - II* conference (2025) funded by *Google Deepmind, ICTS-Bangalore & Ashoka University.*
- Attended *Geometry of Machine Learning Conference* (2025) conference conducted by *CMSA, Harvard University.*
- Attended the Symposium on *Artificial Intelligence and Pharmaceutical Medicine (2025)*, jointly organized by *Pfizer and Indian Institute of Technology (IIT) Madras.*
- Attended the workshop on *Language & Vision with AI/ML* (2025), conducted by *Multimedia Lab, Indian Institute of Technology (IIT) Guwahati.*

- Attended *Accelerating statistical inference and experimental design with ML* (2025) conference, conducted by *Isaac Newton Institute for Mathematical Sciences, University of Cambridge.*
- Attended the *CCAIM AI and ML Summer School in medicine* (2024) & *in healthcare* (2023), (with scholarship), conducted by the *University of Cambridge, UK.*
- Attended the workshop on *Synthetic Biology* (2024), conducted by *Indian Institute of Technology (IIT) Madras.*
- Participated in the *5 day Generative AI - intensive course*, conducted by *Google.*
- Attended *International Conference on Security & Privacy* (2020), conducted by *National Institute of Technology Jamshedpur.*
- Attended *International Conference on Number Theory and Discrete Mathematics* (2020), conducted by *Ramanujan Mathematical Society.*

Invited Talks & Workshops

- Conducted a 2-day workshop on the *Fundamentals of Machine Learning* for B.Tech - CSE students, conducted by the *IEEE students council, GLA University, Mathura* (2025).
- Appeared for *a podcast*, a conversation with Mohammed Yousef Shaik (CEO of Approtors Technology Pvt. Limited) about my experiences in Data Science (2025).
- Gave *a guest talk*, on *Causal Inference*, organized by the *Google Developer Group, IIT Madras.*
- Gave a *guest public talk*, on the Life and Work of *John Von Neumann* for high school kids on the occasion of National Mathematics Day (2023), conducted by the *District Commissioner (DC)*, West Karbi Anglong, Assam.
- Conducted a *workshop* for state math olympiad winners conducted by the *Assam Academy of Mathematics (AAM).*
- Gave a *talk* as a part of the *Polymath REU Jr.* program 2021 listed by the *American Mathematical Society (AMS).*

Achievements/Awards

- Awarded the prestigious *Khorana Scholars Program Fellowship* 2024 (acceptance rate: 0.2%).
- Secured *North-East India Rank 1* (2020) in Madhava Mathematical Competition (equivalent to the William Lowell Putnam Mathematical Competition, USA).
- Qualified *National Talent Search Examination (NTSE)* Level-1, 2017 (top 0.5%).
- Secured *All India Rank (AIR) 89* in the Indian Institute of Technology (IIT) - JAM Mathematics (top 0.6%).
- Awarded *Dr. Subratananda Dowerah Memorial Gold Medal - 2016* by Assam Academy of Mathematics (AAM).
- Awarded *Dr. Durgeswar Das Memorial Award - 2016* by Assam Academy of Mathematics (AAM).
- Qualified for *Indian National Mathematical Olympiad (INMO)* four times in 2014, 15, 16 & 17.
- Selected for the *JNCASR Summer Research Fellowship* (2020).
- Selected for the *Mathematics Training and Talent Search (MTTS)* Programme (2021).
- Secured *Rank 1* in State Chemistry Olympiad (2016).
- Awarded *The Best Academian Award* by my school, in 2016 for outstanding academic excellence.
- Awarded *School Math Topper Award* (2013) in International Mathematics Olympiad (IMO) conducted by Science Olympiad Foundation (SOF).

Others

- Hobbies: Solo-Travelling, Book Reading, Teaching
- Languages: English, French (A1), Hindi, Assamese

† Due to confidentiality agreements, I am unable to share the code repositories.