Soumadeep Saha, PhD

Indian Statistical Institute, Kolkata

soumadeep.saha_r@isical.ac.in • soumadeep.saha97@gmail.com • (+91) 869 737 3806 • • • •

Summary

My PhD research centered on enhancing deep learning systems with the strengths of symbolic AI techniques for greater logical coherence and domain constraint adherence. While my primary focus is *natural language processing*, I actively collaborate across domains including vision, medicine, biology, astrophysics, and business applications. My *current research* explores techniques to better understand reasoning with large language models. I am also interested in alignment, and my recent work has received support from *OpenAI*.

Education

Indian Statistical Institute, Kolkata — Kolkata, India
PhD (Computer Science) *2020–2025*
- ❏ Supervisor: **Prof. Utpal Garain**
- ❏ Dissertation: **"Domain Obedient Deep Learning"**.

Indian Institute of Science Education and Research, Kolkata — Kolkata, India
Integrated BS & MS in Physics (Minor in Mathematics) *2015–2020*
- ❏ Awarded the Kishore Vigyan Pratoshan Yojna (**KVPY**) fellowship from the Department of Science and Technology, Govt. of India (GoI).
- ❏ Master's dissertation: **"Towards Robust Deep Learning Systems"**.

Bhavan's G.K. Vidyamandir — Kolkata, India
10+2 (Pre-University Secondary Education) *2002–2015*
- ❏ Recipient of the prestigious National Talent Search Examination (**NTSE**) Scholarship from NCERT, GoI.
- ❏ CBSE (10th standard): 91.2%; AISSCE (12th standard): 92%.

Publications

❏ sudoLLM: On Multi-role Alignment of Language Models — **EMNLP 2025** (findings)
S. Saha, A. Chaturvedi, J. Mahapatra, U. Garain

❏ Language Models are Crossword Solvers — **NAACL 2025** (main)
S. Saha, S. Chakraborty, S. Saha, U. Garain

❏ Analyzing Semantic Faithfulness of Language Models via Input Intervention on Conversational Question Answering — **Computational Linguistics** (2024), 50(1): 119-155
A. Chaturvedi, S. Bhar, **S. Saha**, U. Garain, N. Asher

❏ On Measuring Intrinsic Causal Attributions in Deep Neural Networks — **CLeaR 2025**
S. Saha, D. V. Rathore, **S. Saha**, D. Doermann, U. Garain.

❏ VALUED - Vision and Logical Understanding Evaluation Dataset — **DMLR** (2024), (13):1-18.
S. Saha, S. Saha, U. Garain

❏ MedTric : A clinically applicable metric for evaluation of multi-label computational diagnostic systems — **PLOS One** (2023), 18(8): e0283895
S. Saha, U. Garain, A. Ukil, A. Pal, S. Khandelwal.

❏ DOST – Domain Obedient Self-supervised Training for Multi Label Classification with Noisy Labels — **AAAI 2024 Workshop (W3PHIAI)** (2024), 1164: 117-127
S. Saha, U. Garain, A. Ukil, A. Pal, S. Khandelwal.

❏ LADDER: Revisiting the Cosmic Distance Ladder with Deep Learning Approaches and Exploring its Applications — **ApJS** (2024), 273(2): 27
R. Shah, **S. Saha**, P. Mukherjee, U. Garain, S. Pal

❏ KisMATH: Do LLMs Have Knowledge of Implicit Structures in Mathematical Reasoning? — Preprint. 10.48550/arXiv.2507.11408
S. Saha, A. Chaturvedi, S. Saha, U. Garain, N. Asher — Under review.

❏ Deep Learning Based Recalibration of SDSS and DESI BAO Alleviates Hubble and Clustering Tensions — Preprint. 10.48550/arXiv.2412.14750
R. Shah, P. Mukherjee, **S. Saha**, U. Garain, S. Pal — Under review at ApJ.

Patents

❏ Method and System for Contradiction Avoided Learning for Multi-Class Multi-Label Classification — **US Patent - US12038949B2**
S. Saha, U. Garain, A. Ukil, A.Pal — Granted Jul. 2024

❏ Method and System for Evaluating Clinical Efficiency of Multi-Label Multi-Class Computational Diagnostic Models — *Pending*
U. Garain, **S. Saha**, A. Ukil, T. Deb, S. Richa A. Pal, S. Khandelwal — Filed Sep. 2022.

Experience

Helmholtz Visiting Researcher — Jul '24–Sep '24
Awarded the Helmholtz Information and Data Science Academy (HIDA) visiting researcher grant to work at the Institute of Aerospace Medicine, DLR (German Aerospace Center). Conducted research on improving AI-based retinal diagnostics for spaceflight applications.

Alleima — Mar '24–May '24
Developed a custom end-to-end computer vision-based solution for assembly-line object detection/tracking, enabling autonomous logging of crucial data with existing infrastructure.

TCS Research — Nov '21 – Jul '22
Conducted research as part of a team on ECG-based diagnosis of cardiovascular diseases. Independently identified key gaps in the literature and developed innovative, state-of-the-art solutions to address them, resulting in two patents, and publications.

Talks and Presentations

- ❏ Organized a 4-day workshop for DataLab, **Capital One**, Bangalore (2023).
- ❏ Subject matter expert for Be10x—designed MOOCs, and delivered live seminars to 500+ participants (2024).
- ❏ TA for Natural Language Processing course at ISI Kolkata (2022–2025).
- ❏ Instructor at the Winter School of Deep Learning (WSDL), ISI Kolkata (2021–2024).
- ❏ Instructor for the Comprehensive Course on Business Analytics, ISI Kolkata (2022).
- ❏ Presented my work on Logically Coherent Deep Learning at Amazon Research Days (ARD '22).
- ❏ Presented my work on domain-obedient self-supervision at W3PHIAI, AAAI 2024, Vancouver, Canada.
- ❏ Presented my work on constrained LLM inference at NAACL 2025, Albuquerque, USA.

Skills

❏ **Deep Learning**: Expert in deep learning theory and practice, from low-level optimization to advanced architectures and training paradigms.
❏ **Python**: 10 years of experience. Proficient in `pytorch`, `Hugging Face`, `nympy`, etc., with extensive experience developing custom modules, training loops, etc., and scaling for high-performance GPU clusters.
❏ **NLP**: Significant LLM expertise (in-context learning, RAG, PEFT, Q-LoRA, instruction tuning, RLHF, DPO, etc.)
❏ **Deep RL**: DQN, PPO, MCTS, GRPO, etc.
❏ **Programming**: In addition to python, experienced in C++, FORTRAN, bash, etc.
❏ **Vision**: Vision transformers, CNNs, VLMs, object detection, segmentation, etc.

Interests

❏ **Robotics/DIY**–Active interest in 3D printing, robotics, electronics, etc. I have conducted introductory workshops on robotics and was the Secretary of the Robotics and Astronomy Club at IISER Kolkata.
❏ **Sports**–Represented my college in national-level sports meets in basketball and volleyball. Played in my state's Senior Division Men's Basketball League.
❏ **Music**–Classically trained pianist.
