Bhati, Sunny

sunnybhati123@gmail.com
https://sunnybhati.github.io/

EDUCATION

Indiana University, Bloomington — Indiana, USA
• M.S in Data Science with concentration in AI — Sep. 2019 - May. 2021

Indian Institute of Technology, Madras — Chennai, India
• Dual Degree *(B.Tech + M.Tech)* in Engineering Design — July. 2010 - May 2015
Advisors: Dr. Sandipan Bandyopadhyay
Thesis: Dynamic Modelling of Macpherson and double wishbone suspension systems

PUBLICATIONS

1. **Rethinking Dataset Distillation: Hard Truth about Soft Labels**
   Priyam Dey, **Sunny Bhati**, Aditya Sahdev, R.Venkatesh Babu, Konda Reddy Mopuri
   Conference on Computer Vision and Pattern Recognition (CVPR), 2026[**Oral Presentation and Best Paper Finalist (Top 15 out of 16000 submission)**]

PROFESSIONAL ACTIVITIES

Reviewer (Journals)
T-PAMI – Transactions on Pattern Analysis and Machine Intelligence — since 2025
TMLR – Transactions on Machine Learning Research — since 2025
JMLR – Journal of Machine Learning Research — since 2025

RESEARCH EXPERIENCE

Indian Institute of Science — Bangalore, India
• Pre-Doctoral Researcher — Sep 2024 - December 2025
Advisor: Dr. R Venkatesh Babu
Project: **Rethinking Dataset Distillation: Hard Truths about Soft Labels**

- Conducted research on dataset distillation and coreset learning, revealing fundamental limitations of soft-label evaluation protocols for measuring distilled dataset quality
- Developed CAD-Prune and CA2D, compute-aware methods for efficient data selection and distillation, achieving improved performance on ImageNet-1K under hard-label evaluation.
- Established comprehensive benchmarks comparing state-of-the-art distillation and coreset methods across multiple compression regimes.
- Accepted as an **Oral Presentation** and **Best Paper Finalist** at **C**VPR 2026.

*Project:***Positional Bias in Self Supervised Models**

- Currently conducting research on self-supervised learning, with a focus on mitigating positional bias in vision foundation models.

TEACHING ASSISTANT

Indiana University Bloomington
ENGR 511 – Machine Learning for Signal Processing — Fall 2020, Spring 2021
CSCI 551 – Elements of Artificial Intelligence — Spring 2020

TECHNICAL PROFICIENCY

Programming
Python*, HTML/CSS*, JavaScript*, Pytorch, TensorFlow

Software / Framework
PyTorch*, TensorFlow, OpenCV, Processing*, Matlab*, Mathematica, LaTeX*

(*) implies a sufficient level of expertise.

EMPLOYMENT

Amazon — Texas, Austin, USA
- Applied Scientist — Jan 2022 - Jan 2024

– Developed production-scale learning-to-rank models for Amazon Search, optimizing search relevance using large-scale user interaction data and behavioral signals.

– Formulated novel multi-objective optimization targets combining engagement, pricing, and fulfillment objectives to balance customer experience and business metrics, achieving an 8% CTR improvement and 0.6% revenue growth.

– Designed a reproducible machine learning experimentation platform integrating distributed data processing, feature generation, model training, inference, and automated evaluation, reducing experimentation cycles by over 90 % (1 week $\rightarrow$ 10 hours)

– Engineered scalable feature pipelines and performed extensive offline/online ablation studies to evaluate new ranking signals, driving measurable gains in search quality and model robustness.

Cisco — San Fransisco, USA
- Applied Scientist — July 2021 - Jan 2022

– Designed a Transformer-based sequential anomaly detection framework for large-scale data center network logs, leveraging Word2Vec embeddings to encode unstructured log events and capture temporal failure patterns for switch-level fault prediction.

– Built scalable preprocessing and feature engineering pipelines for network telemetry, enabling robust anomaly classification and comprehensive evaluation against traditional sequence modeling approaches.

Ernst & Young — Doha, Qatar
- Data Science Consultant — July 2017 - May 2019

– Developed statistical Probability of Default (PD) and Loss Given Default (LGD) stress-testing models for financial institutions, incorporating scenario-based risk modeling to assess credit portfolio resilience under adverse macroeconomic conditions.

– Performed rigorous credit risk model calibration and validation, including PD curve calibration, discriminatory power analysis (ROC/AUC, Gini), back-testing, and qualitative assessment of internal rating model performance to ensure regulatory compliance and predictive reliability.

– Architected and implemented a scalable Expected Credit Loss (ECL) estimation platform using R and SQL, integrating automated data processing, risk analytics, interactive dashboards, and reporting to support model governance and business decision-making.

– Designed and deployed a production-grade end-to-end stress-testing pipeline that automated model development, validation, and reporting workflows, earning **two organizational innovation awards for technical excellence and operational impact**