

SACHIT GAUDI

[in gaudisachit](#) | [✉ gaudisac@msu.edu](mailto:gaudisac@msu.edu) | [📍 EL, Michigan](#) | [☎ +1-517-944-7743](tel:+15179447743) | [🌐 sachit3022.github.io](https://sachit3022.github.io)




Education


Masters M.S - Computer Science **Michigan State University**, East Lansing, USA Jan 2023 - Expected Dec 2024
Bachelor of Technology **Indian Institute of Technology**, Guwahati, India Jul 2014 - May 2018

Academic Experience

Graduate Research - Human Analysis Lab Computer Vision, ML, NLP
Advisor [Prof. Vishnu Boddeti](#) [Links to my work on Privacy, Fairness, Optimization](#)

- **Fairness** - My research focus is improving generalization capability of model addressing challenges like imbalance, group shift, and memorization. Currently developing fairness-constrained optimization techniques.
- **Privacy** - Proposed solutions based on pruning, re-weighting, projections for removing private information from models.

Adversarial finetuning to mitigate bias in LLM generation    Python, huggingface

- Illustrated that bias in prompt translates to bias in model generation and proposed adversarial prompt tuning to identify harmful prompts. First to develop a stable adversarial fine-tuning framework for LLMs, leveraging LoRA adapter to reduce memory footprint, and Stochastic Weight Averaging (SWA) to enhance stability.
- Wrote a survey paper on adversarial techniques employed to mitigate harm in LLMs. Survey paper available .

Distributed training of LLMs   HPCC, C++, MPI, Python, OpenMP, CUDA

- Developed P-GPT algorithm - a gradient map-reduce parallelization strategy - for training large transformer models across multiple servers. P-GPT establishes optimal parameters for data scheduling, threading, and process parallelism tailored to specific network topologies achieving a parallel efficiency of **95%**.

Novel Domain Generalisation Algorithm   Python, PyTorch, sklearn


- Pattern Unrecognition - Addressed the problem of spurious correlation, developed a novel research solution by incorporating fairness constraints into LDA, yielding a substantial **8%** improvement in generalization accuracy.

Contrasting Learning based Image Search Python, PyTorch, Huggingface, MongoDB, Faiss

- Implemented a Triplet Pair Siamese Network with a base model of VGG16, trained using hinge loss. Achieved a **10% recall** while maintaining flexibility for training on search rank data.

Professional Experience

Reliance Jio Haptik  Mumbai, India
Machine Learning Engineer - Commerce AI Agent team *Apr 2021 - Dec 2022*

- Played a pivotal role in developing an AI commerce agent that revolutionized shopping by seamlessly integrating natural language interactions, eliminating button clicks and resulting in **50%** increase in conversion rate. [Media coverage](#) 
- Built the NLP pipeline featuring a **RAG** (Retrieval Augmented Generation) system. Extracted information from queries, queried ElasticSearch, and generated responses from a knowledge graph.
- Led research and implementation efforts in multi-task BERT-based transformer models for product labeling, named entity recognition (NER), and intent detection in JioMart Search (25M MAUs), resulting in a **10%** improvement in **recall**.
- Engineered a collaborative filtering based product personalization module. enhancing targeted recommendations, achieving significant improvement with **average first click rank** decreasing from 3.4 to **2.1**.

Software Engineer I *Jan 2020 - Apr 2021*

- Refactored monolith to microservices, employing async queues for inter-service communication. Orchestrated CI/CD pipelines for software and data delivery. Developed Git actions and wrote bash scripts triggering Jenkins deployments on Azure and AWS. resulting in a **30%** reduction in latency as well as contributing to a **99%** up-time guarantee.
- Developed a solution based on mediator pattern to integrate messages from all chat sources, to a central agent.
- Applied back-track algorithm to enhance Intent detection & utilized Tf-idf for improved semantic extraction.

Smatbot AI Agent  Hyderabad, India
Lead Backend Developer *Jan 2019 - Jan 2020*

- Led architecture and database design for the chatbot. Implemented database sharding, replication, and pushed efforts to make platform No-code, reaching a scale of **10 req/sec**. Contributed to securing a favorable exit for the investors.


Data Science Summer Research Associate Philadelphia, USA
Saint Joseph's University *May 2017 - Jul 2017*

- Summarized content and scores of hotel reviews, attributing sentiments to specific aspects (e.g., cleanliness). Conducted sentiment analysis for unstructured data and utilized feature importance to identify key aspects to improve occupancy.





Teaching

Michigan State University

Spring'24, Fall'23, Summer'23

- **Object Oriented programming:** Instructed 800 students in C++ language (CSE 232). Developed problem sets and delivered lectures on pointers, classes, IO, strings, objects, and recursion. My content is available [online](#) .
- **Database Systems:** Taught Locks and Object-Oriented Design. Guided students in hands-on creation of SQL database.

Coursework

- **Text generative models:** Implemented State Space models (Mamba) and Attention models (GPT-2) from scratch. Tuned performance of foundational models on translation task. Presented underlying learning of the models. 
- **Image generative models:** Studied Score, Diffusion, and VAE models. Implemented stochastic differential formulation of score and diffusion. Showed DDPM with variance schedule performs best on the 8-mode gaussian. 
- Showed Hourglass model with skip connections improves performance on **Semantic segmentation** over U-Net baseline 
- Explored techniques to address common challenges in large datasets, such as scale, multi-modality and imbalance. 

Technical Skills & Services

Programming	Python (PyTorch, TensorFlow, sklearn, CUDA, OpenCV), C++, Java, Shell Scripting
Databases & Search	PostgreSQL, MySQL, Redis, MongoDB, Solr, ElasticSearch
Tools & Technologies	Docker, Flask, Django, Git, Jenkins, RabbitMQ, Wandb, Jira
ML Concepts	Regression, Classification, Clustering, Boosting Trees, Gradient Boosting, Deep Learning.
Relevant Courses	Computer Vision, Graph Theory, Parallel Computing, Data Structures and Algorithms

Achievements

Best Employee: Received Big Hand award from Reliance for exemplary quarterly performance, youngest to do so.

Scholarship: Received Scholarship based on Merit from IIT, Kharagpur. (Top 10%).

Joint Entrance Examination: Secured an All India Rank of **2240** out of 1.35 million candidates. (Top 0.17%)

Extra curricular

Coding: Regularly solve coding challenges on Leetcode ([profile](#)), with 500+ problems solved and 8 unique solutions.

Running: Amateur runner, frequently completing 5K runs with a PB of 30:53, and a time of 1:11:50 for 10K runs.

Technical Blog: I maintain a [blog](#) where I strive explain deep learning concepts like convex optimization and diffusion models with graphical yet intuitive mathematical explanations.