

VISHNU BEJI

vishnubeji@gmail.com | +1 (347) 798 3743 | [Linkedin](#) | [Github](#) | [Google Scholar](#)

EDUCATION

New York University

Master of Science in Computer Science (Recipient of Merit-based scholarship)

May 2025

Coursework: Machine Learning, Deep Learning, Big Data, Operating Systems, Cloud Computing

CG: 3.96/4.00

Indian Institute of Technology Madras

Bachelor of Technology in Electrical Engineering

Jul 2020

SKILLS

Languages: Python, C, C++, R, SQL **Frameworks:** TensorFlow, Keras, PyTorch, NLTK, cuDNN, Pandas, LangChain, LlamaIndex, Transformers

Tools: Git, Docker, CUDA, Kubernetes, Spark, Kafka, MongoDB, AWS, GCS, EC2, S3, ECS, DynamoDB, Terraform

WORK EXPERIENCE

Courant Institute of Mathematical Sciences, NYU - New York City, NY

Jan 2024 - Present

Graduate Teaching Assistant

- Head Teaching Assistant and Tutor for Advanced Computer Vision and Advanced Machine Learning
- Designed self-contained programming and theoretical assignments on topics like Transformers, VAEs, GANs and Diffusion models

Lowe's - Charlotte, NC

May 2024 - Aug 2024

Machine Learning Intern

- Optimized Visual Search using **ConvNeXt** and **Vision Transformer** dual-tower **CLIP** to increase hit-rate from **75.05% to 83.21%**.
- Leveraged Google Cloud Platform and BigQuery to extract, preprocess, and handle large chunks of multi-modal data effectively to create a dataset of 290k image pairs, which were cleaned using Vision-based LLMs.
- Implemented search expansion and **cross-encoder**-based re-ranker post-retrieval to enhance search relevance.

CILVR group - advised by Prof. Saining Xie

Nov 2023 - May 2024

Research Assistant, Multimodal Learning for Data-Efficient Zero-Shot Object Recognition using LLMs

- Developed a **Feature Fusion** for **Multi-modal Large Language Models** that fuses visual and textual features into a shared semantic space, enhancing the model's ability to understand and recognize objects efficiently.
- Annotated a comprehensive dataset with rich semantic attributes, enabling the model to predict object attributes from both images and text, bridging the semantic gap and improving object recognition accuracy.

Oracle - Bangalore, India

Nov 2020 - Aug 2023

Senior Member of Technical Staff

Areas: **Distributed Systems, Data Structures, OS, Databases**

- Led the redesign of the Slice Management Layer (SLM), introducing "slicing" to improve query speed and performance while ensuring **99.9% system availability**
- Mentored and guided new hires on Database and systems architecture concepts, development tools and RDBMS bug fixing

Member of Technical Staff

- Enabled In-memory Transaction Private Journal to handle variable length bitmaps
- Refactored the hierarchical structure of SLM Catalogs (a set of metadata tables) residing at Level 2 of table abstraction to establish astute separation of logical and physical entities

Samsung Research - Bangalore, India

Jun 2019 - Aug 2019

Summer Intern

Areas: **NLP, LLM, Machine Learning, Data Structures**

- Augmented Bixby Search by developing **Intelligent Grouped Keywords** using **SMS data** to reduce query processing time by **20%**
- Optimized the Latent Dirichlet Allocation (**LDA**) based model with a self-developed algorithm for probabilistic topic modeling.
- Fine-tuned the **BERT model**, to craft topic-keyword clusters, resulting in a **40% enhancement** in content relevance

PUBLICATIONS

- Vishnu B, A. Sinha**, [Fast and Secure Routing Algorithms for Quantum Key Distribution Networks](#), International Conference on Communication Systems and Networks **COMSNETS 2022**.

PROJECTS

Infrapilot - DevOps Assistant using LLM

Oct 2024 - Present

- Created a **RAG**-based chatbot to automatically build and deploy CI/CD pipelines and other AWS resources as an AWS amateur.
- Using GPT-4o mini, Supabase Vector store, and Lambda functions, orchestrated a scalable multi-tenant cloud application on AWS.
- Dynamically generated terraform templates using OpenAI API calls and automated workflows using GitHub runners.

ComicGen - Winner of MongoDB Gen AI Hackathon, New York

Nov 2023 - Present

- Developed a **RAG**-based scene generation model to create comic-book-style renderings of fan theories and plot extensions.
- Using **Stable Diffusion 1.0** and **Mistral-7B-Instruct-v0.2**, generated comic strips in a user-specified illustration style.

Temporal Localisation for Action Detection on Streaming Video [\[code\]](#) -Prof Juan Rodriguez

Nov 2023 - Jan 2024

- Developed an object detection model for **Temporal Localisation** on **Youtube-8M Segment dataset** achieving **80.2% accuracy**.
- Used a **Context-Gated DBoF model** for temporal aggregation on rich static features from a pretrained Inception V3.