

# CHAITANYA CHAKKA

chaitanyachakkavsk@gmail.com | (857) 919-1799 | /in/chaitanya-chakka-vsk/ | github.com/LearningLeopard | chaitanya-chakka-portfolio.netlify.app/

## EDUCATION

### Boston University

Masters in Artificial Intelligence | GPA: 3.960/4.00

Sep 2024 - Exp. May 2026

### Birla Institute of Technology and Science Pilani Hyderabad Campus

Bachelor of Engineering in Computer Science, Minor in Data Science | CGPA: 3.76/4.00

Aug 2019 - May 2023

## SKILLS

**Languages:** Python, JavaScript (ES6), Go, Java, C++, C, SQL, Bash, Lua

**ML & GenAI:** Machine Learning, Deep Learning, Multimodal Learning, Transformer Models, Distributed Training, Prompt Engineering, LoRA, Diffusion Models, SFT, PEFT

**Vision:** Computer Vision, Vision-Language Models, VQA, VLN, Cross-Modal Attention, Image/Video, CLIP

**Frameworks:** PyTorch, PyTorch DDP, Hugging Face, Weights & Biases, OpenAI API, LangChain, OpenCV, NumPy, Pandas, Scikit-learn, Matplotlib

**Systems & Data:** AWS, Docker, Kubernetes, REST APIs, Kong, Git, Datadog, PostgreSQL, MongoDB, CI/CD, Linux, HPC Clusters

## EXPERIENCE

### Prospect33

Data Scientist Intern

June 2025 - Aug 2025

New York, NY

- Engineered a GPU-accelerated WebGL/React visualization system capable of rendering 8M+ datapoints in 3D using Deck.GL and binary GPU buffers while maintaining 60 FPS and <10s load time.
- Built an active-learning pipeline human-in-the-loop labeling to surface the top 1% most informative samples for model training.
- Integrated GPU-accelerated embedding pipelines (PCA, t-SNE, UMAP, autoencoders) enabling interactive exploration of model representations and improving downstream classification performance.

### Cashfree Payments

Software Development Engineer

June 2023 - June 2024

Bangalore, India

- Spearheaded integration of Kong API Gateway with Golang server wrappers across 30 teams serving 1500 TPS daily. Revamped the system with custom Lua plugins for team-specific use cases and deployed in Kubernetes with Postgre SQL backend.
- Introduced Twilio for Whatsapp and SMS notification channels via gateway routing, improving delivery rates by 20%.
- Approved and validated all API gateway changes, ensuring 100% uptime and high-quality deployments via integration tests.

## PROJECTS

### Layer-Residual Co-Attention Networks for VQA | Multimodal Attention, VQA, [Github](#)

May 2025 - Jul 2025

- Implemented a multimodal architecture integrating ResNet-152 visual encoders with GloVe-initialized LSTM language encoders.
- Built a full PyTorch training pipeline including distributed multi-GPU training using PyTorch Distributed Data Parallel (DDP) on university HPC clusters.
- Integrated experiment tracking with Weights & Biases to monitor model convergence, and hyperparameter experiments.

### Vagueness Analysis in Vision-Language Navigation | Vision Language Navigation(VLN) [Github](#)

Jan 2025 - Mar 2025

- Built a novel dual-metric framework (lexical and LLM-based) to quantify instruction vagueness across multiple VLN datasets.
- Analyzed vagueness affects on generalizability of SoTA VLN agents, explaining architectural factors behind stronger robustness.

### Optimizing LLM Question Generation for Enhanced Conversational AI | LLM, Question Answering, [Github](#)

Oct 2024 - Dec 2024

- Built a 3-module conversational QA pipeline that produces follow-up questions for missing context using an iterative correctness check before finalizing responses. Processed and fine-tuned on a 26k-sample immigration QA dataset.
- Delivered measurable gains (ROUGE-1 +0.01), demonstrating better answer quality after follow-up clarification.

## PUBLICATIONS

### Some Modalities Are More Equal Than Others: Decoding and Architecting Multimodal Integration in MLLMs

Accepted at CVPR, 2026 | Google Deepmind - Tianle Chen\*, **Chaitanya Chakka\***, Arjun Reddy Akula, Xavier Thomas, Deepti Ghadiyaram, [Website](#)

- Built MMA-Bench, a multimodal (audio-video-text) evaluation benchmark using the AudioSet ontology to diagnose cross-modal semantic misalignment in large multimodal models.
- Engineered a PyTorch evaluation pipeline combining black-box behavioral tests and attention-weight statistical analysis to diagnose modality attribution failures in multimodal transformers.
- Fine-tuned multimodal transformer models using LoRA adapters, improving modality-specific accuracy by 20-40% and reducing misleading-text errors by 54%.

### Improving Prompt Alignment in Vision Language Models: A Self-Learning Framework for Generative Models

Accepted at CVPR, 2025 - Tianle Chen, **Chaitanya Chakka**, Deepti Ghadiyaram. [Publication](#)

- Fine-tuned multiple open-source diffusion models (Stable Diffusion, SDXL, Flux, OmniGen), achieving +24% GPT score and +17% VQA score gains on our curated new benchmark called Object-State-Bench.
- Built a synthetic data pipeline to model absent and empty object states, using LLMs and LVLMs for recaptioning and filtering.
- Validated generalization to unseen objects via human studies, improving semantic alignment without loss of fidelity.

## ADDITIONAL EXPERIENCE

### Graduate Teaching Assistant | Dr. Andrew Wood | Foundations of Data Science - Boston University

Jan 2025 - Present

- Supported CS585 (image/video understanding: color, shading, stereo, motion; applications incl. face recognition, tracking, HCI) and Foundations of Data Science by running office hours, grading assignments/projects, and reinforcing core ML/stat topics (e.g., PCA, EM) through targeted feedback for 90+ students.