

CHAITANYA CHAKKA

chaitanyachakkavsk@gmail.com | (857) 919-1799 | www.linkedin.com/in/chaitanya-chakka-vsk/ | github.com/LearningLeopard

EDUCATION

Boston University

Masters in Artificial Intelligence | GPA: 3.960/4.00

Sep 2024 - Exp. May 2026

- Image Video and Computing
- Natural Language Processing
- Multimodal Machine Learning

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

SELECTED PUBLICATIONS

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

Under Review | Google Deepmind — Tianle Chen*, **Chaitanya Chakka***, Arjun Reddy Akula, Xavier Thomas, Deepti Ghadiyaram, [Website](#)

- Built **MMA-Bench**, a controlled audio-video-text semantic misalignment benchmark with Google's AudioSet ontology dataset.
- Developed a interpretability pipeline combining black-box tests and statistical analysis on attention weights(white-box).
- Applied **LoRA finetuning** on clean training samples, improving modality-specific accuracy by 20-40% and reducing misleading-text errors by 54%. Improved fundamental model understanding via cross-modal attention reallocation analysis.

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.

EXPERIENCE

Prospect33

Data Scientist Intern

June 2025 - Aug 2025

New York, NY

- Engineered a high-performance, low-latency WebGL/React interactive data visualization platform that renders upto **8 M+** points in 3-D (Deck.GL + binary GPU buffers); to sustain 60 FPS with < 10 s initial load.
- Built an active-learning loop, and human-in-the-loop labeling surfacing the top 1% most informative points though visual cues
- Added GPU-accelerated embeddings (PCA, t-SNE, UMAP, auto-encoder) so analysts can boost downstream model F1.

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.

SELECTED PROJECTS

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

May 2025 - Jul 2025

- Implemented the Layer-Residual Mechanism (LRM), integrating ResNet-152 features with GloVe-initialized LSTM text encoders.
- Engineered the full training pipeline (Adam + LR decay, BCE multi-label loss, checkpointing, logging) and achieved ~60% accuracy on VQAv2 with the Co-Stacking variant, matching expected architectural trends.

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.

SKILLS

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

ML & GenAI: Machine Learning, Multimodal Learning, LLMs, Prompt Engineering, LoRA, Diffusion Models, SFT, PEFT

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

Frameworks: PyTorch, Hugging Face, OpenAI API, LangChain, OpenCV, NumPy, Pandas, Scikit-learn, Matplotlib

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

ADDITIONAL EXPERIENCE

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

Jan 2025 - Present

- Teaching crucial math-intensive foundational topics in Data Science for a class of 90 students at Boston University.
- Delivered personalized feedback to 90 students, resulting in a 30% improvement in average assignment scores and a noticeable increase in class participation during lectures on EM and PCA.