

ADITYA V KALLAPPA

adityavk420@gmail.com | +91 7676636037 | Bengaluru, KA, India | [linkedin.com/in/aditya-kallappa/](https://www.linkedin.com/in/aditya-kallappa/)

OBJECTIVE

Innovative **Deep Learning Researcher** with hands-on experience in **NLP**, **Computer Vision (CV)**, and **Generative AI**. Specialized in developing and fine-tuning **Large Language Models (LLMs)**, custom tokenizers, and AI benchmarks for multilingual and multimodal applications. Skilled in **pretraining, alignment, distributed training (DeepSpeed, FSDP)**, and **efficiency optimization (PEFT, LoRA, Quantization)**. Passionate about advancing AI for Indic languages, optimizing LLM instruction following, reasoning, and alignment with human preferences to drive real-world impact.

TECHNICAL SKILLS

Languages: Python, C/C++ (Basics)

Tools & Libraries: PyTorch, CUDA, NumPy, Pandas, HuggingFace Transformers, Datasets

Key Expertise: Machine Learning (ML), Deep Learning (DL), LLMs, CV, NLP, Generative AI, Distributed Training, Linux, PEFT(LoRA), CUDA

WORK EXPERIENCE

Data Scientist 2, Krutrim SI Designs Ltd., Bengaluru, KA, India

April 2023 - Present

- **Krutrim-2 LLM:** Continually pretrained the Mistral-Nemo (MN) 12B model on high-quality English, Indic, Math, and Coding datasets to enhance support for Indian languages. The model outperforms MN-12B-Instruct across English, multilingual, and coding benchmarks, demonstrating significant gains in fluency, coherence, and reasoning [[krutrim-ai-labs/Krutrim-2-instruct](https://krutrim-ai-labs.github.io/Krutrim-2-instruct/)]
- **Synthetic Data & SFT:** Designed and implemented pipelines to generate **synthetic data** for Indic languages, ensuring data diversity and linguistic coverage. Performed Supervised Fine-Tuning (SFT) on Krutrim-2 LLM using curated high-quality English and Indic datasets, incorporating domain adaptation and task-specific optimizations. The model delivers best-in-class performance on Indic tasks, surpassing models **5-10x** its size in fluency, accuracy, and generalization [[Reference](#)]
- **Alignment & Preference Optimization:** Implemented Direct Preference Optimization (DPO) for English and Indic languages, aligning LLMs with human preferences to enhance response quality, cultural nuance, and multilingual robustness. Improved dialogue coherence, contextual accuracy, model self-identity, and instruction following for real-world applications.
- **Ongoing Work:** Enhancing model alignment with human preferences and improving Indic reasoning and user satisfaction

Data Scientist - 1, Krutrim SI Designs Ltd., Bengaluru, KA, India

July 2023 - March 2024

- **Krutrim-1 LLM:** Key contributor to India's first multilingual LLM, trained from scratch on **2.2T** tokens. Designed a robust pretraining pipeline, integrating custom architectures, **GQA**, and **ALiBi** [[krutrim-ai-labs/Krutrim-1-instruct](https://krutrim-ai-labs.github.io/Krutrim-1-instruct/)]
- **Large-Scale Data Processing:** Worked on large-scale data cleaning and preprocessing of petabytes of text data, ensuring high-quality inputs for model training. Used **CC-Net**, **RefinedWeb**, and custom filtering techniques to remove noise, deduplicate content, and enhance linguistic diversity, optimizing data efficiency, consistency, and relevance for training
- **Tokenizer Development:** Led the development of an in-house tokenizer optimized for English and Indic languages using **BPE algorithm** with **SentencePiece**, achieving a fertility score of **1.5** for English and **<2** for Indic languages [[Reference](#)]

Data Scientist - 1, Ola Cabs, Bengaluru, KA, India

July 2022 - June 2023

- **SFT & LLM Fine-Tuning:** Led SFT experiments using **DeepSpeed ZERO-3** on various LLMs for customer care automation
- **Cell Defect Detection:** Built a detection system with a custom model with **ResNet** architecture achieving **95%** accuracy
- **Traffic Sign Detection & Classification:** Delivered detection (~88%) and classification (~85%) models using YOLOv5

PUBLICATIONS & RESEARCH

Kallappa, A., Nagar, S., & Varma, G. (2023). **FInC Flow: Fast and Invertible $k \times k$ Convolutions for Normalizing Flows**. *VISAPP 2023 (Vol. 5)*, pp. 338-348. DOI: 10.5220/001187660000341.

BharatBench: Comprehensive Multilingual Multimodal Evaluations of Foundation AI models for Indian Languages (2025). [KrutrimAI Team] Technical Report available at: [BharatBench Report](#)

EDUCATION

Master of Science (By Research), Electronics and Communication Engineering, Specialization in **AI**

2019 - 2023

Bachelor of Engineering, Electronics and Communication Engineering

2013 - 2017

ADDITIONAL INFORMATION

- **Early IT Experience:** Gained hands-on experience at **Accenture**, working on IT systems, troubleshooting, and reporting.
- **Personal Interests:** Passionate chess enthusiast who enjoys analyzing and following games closely. Deeply interested in history and culture, exploring diverse traditions and historical narratives.