Surya Narayan Prasad

GitHub • LinkedIn • surya2019npd@gmail.com

Experience

Software Engineering Intern **Bengaluru, India**
Arista Networks *May 2025 – July 2025*

- Optimised FADT range handling for Gen4 hardware by dynamically identifying alpha via hardware detection, enhancing scalability.
- Ensured queue-threshold does not exceed limits, preventing head-of-line blocking.
- Engineered CLI for VOQ rate class mapping with robust socket-based agent communication and user-friendly configuration.
- Conducted manual testing across hardware generations to verify configuration consistency after reboot.
- Executed full development lifecycle: analysis, implementation, testing (BTest/PTest), reviews, and merge.

Skills

Languages: Python, C/C++, Java, Rust, Bash, SQL
Frameworks/Tools: Spring Boot, PyTorch, NumPy, Pandas, Scikit-learn, Git, Linux, Unix IPC
Core CS: Data Structures & Algorithms, OOP, Operating Systems, Computer Networks, TCP/IP, Routing & Switching, Memory Management
AI/ML: Transformers, Seq2Seq, Attention Mechanisms, NLP, Tokenization, LLM Agents

Education

Visvesvaraya National Institute of Technology, Nagpur **CGPA: 8.94**
B.Tech in Computer Science 2026
Delhi Public School (CBSE) **97.2%**
Class XII 2021
St. Francis School (ICSE) **95%**
Class X 2019

Projects

Unified Agent Runtime (amber-uar) *(GitHub)* *Rust, Unix IPC, JSON-RPC, LLMs*

- Engineered a high-performance Rust bridge connecting reasoning capabilities with a type-safe runtime via Unix IPC.
- Designed IPC routing and dynamic tool-calling mechanisms featuring cryptographic approval gates and persistent memory for secure autonomous agent execution.
- Pioneered the development of a specialized **.aos data format** optimized specifically for LLMs, minimizing serialization overhead and accelerating cross-process data exchange.

V-DeClip (Multi-Modal Video Alignment) *(GitHub)* *PyTorch, V-JEPA, MSRVTT, CV*

- Architected a Masked Multi-Component Gated Decomposition (MCGD) model in PyTorch, extracting semantically decomposed video embeddings via V-JEPA.
- Engineered end-to-end data alignment pipelines leveraging OpenCLIP and contrastive learning over the MSRVTT dataset for precise text-video temporal feature alignment.
- Optimized complex tensor memory allocations and matrix operations to significantly reduce bottlenecking during high-dimensional feature extraction.

Evalverse *(GitHub)* *LangChain, CrewAI, Flask, NLP*

- Built a modular LLM-powered framework for smart interviews and adaptive MCQ assessments, integrating voice-to-text, JSON filtering, and dynamic resume parsing.
- Orchestrated LangChain and CrewAI multi-agent workflows via Flask APIs to manage concurrent session states and real-time NLP evaluation metrics within a unified pipeline.

Recursive Subword Pruning Tokenizer *(GitHub)* *C++, NLP, Data Structures*

- Developed a custom high-performance C++ tokenizer utilizing Recursive Sub-word Pruning (RSPA), achieving 38x faster execution than GPT-4's tiktoken.
- Optimized memory-safe data structures and subword merging logic to minimize vocabulary redundancy while demonstrating superior morphological segmentation quality.

Joint Diagonalization Naive Bayes *(GitHub)* *ML, Linear Algebra, Classification*

- Formulated a research-focused ML classifier by applying joint diagonalization-inspired mathematical transforms to complex, high-dimensional feature spaces.
- Implemented a custom Naive Bayes probabilistic framework, successfully validating algorithmic performance and decision boundaries against standard baseline models.

Position of Responsibility

Project Head, IvLabs VNIT *Leadership, Events, Mentorship*

- Managed club workshops, speaker events, and hackathons; led student outreach and mentorship initiatives to strengthen campus technical learning culture.