

# SAI VIPUL MOHAN VELUDANDI

saifunny43@gmail.com | +91 6305088553 | Bangalore, Karnataka, India

LinkedIn | GitHub | Portfolio | LeetCode

## EDUCATION

**Indian Institute of Technology Palakkad (IIT PKD)**  
Bachelor of Technology in Computer Science and Engineering  
CGPA: 7.39/10

Palakkad, India  
Jul 2018 - Jun 2022

## PROFESSIONAL EXPERIENCE

**R&D Engineer, Software**  
Jul 2025 - Present

**Dagnosis** | Bangalore, India

- Architected a high-volume data migration strategy, moving **30TB** of video datasets from AWS S3/GCP to Backblaze B2, optimizing for a daily intake of **100GB** from 8 high-frame-rate 4K cameras.
- Eliminated **\$3,700/month** in egress costs and reduced storage overhead by **75%** by leveraging B2's free egress policy for ML pipeline data retrieval.
- Maintaining scalable full-stack solutions using **Rust**, **Python**, and **PHP**, while concurrently supporting team growth by conducting **10+ technical interviews** to screen R&D candidates.

**Systems Engineer**  
Aug 2022 - Jul 2025

**Subconscious Compute** | Bangalore, India

- Rewrote the legacy Python/FastAPI backend entirely in **Rust**, achieving an **80% reduction in latency** (from >500ms to <100ms) by implementing load balancing and efficient caching.
- Engineered a high-throughput ingestion pipeline handling **10,000+ RPM** across 100+ agents; optimized storage in **VictoriaMetrics** to handle **35GB/day** of time-series data per agent (decompressing 100MB payloads).
- Solely led the engineering of an on-premise infrastructure suite, containerizing the entire ecosystem (APIs, Queues, MDM) using **Docker** and **IaC** for multi-cloud deployment (AWS, GCP). Integrated license guarding across all services, directly unlocking a new enterprise licensing market.
- Developed the cross-platform Shepherd Agent and MDM solution in **Rust** and played a key role in team expansion by conducting **20+ technical interviews** for core engineering roles.

## PROJECTS

**Kiwis.club - AI-Powered Insights from your Email**  
LLMs, Self-Hosted AI, Golang

**Live Demo** · Source Code  
In Progress

- Developed a privacy-first platform that leverages AI to surface financial insights and upcoming payments from emails; designed a "Private by design" architecture where sensitive data never leaves the secure processing pipeline.
- Engineered a cost-efficient inference engine by self-hosting **Qwen2.5-7B (quantized)** via Docker on CPU, avoiding the projected **\$1,400/month** cost of external APIs while maintaining strict data isolation.
- Designed a distributed producer-consumer system in **Golang** to concurrently ingest emails, implementing keyword pre-filtering to reduce context window usage by **90%** and parsing data with near-zero hallucination.

**Multi-Modal Learning & Computer Vision**  
CS5007 Deep Learning, ID3801 Open Ended Lab

Academic Projects (IIT PKD)  
2021

- Image Description (CS5007):** Built an Encoder-Decoder (CNN-RNN) architecture for generating captions from images. Achieved competitive BLEU scores on Flickr8k. **Source Code**
- Face Mask Detection (ID3801):** Implemented a real-time detection system using MobileNetV2 and FaceNet during the COVID-19 lockdown, optimizing inference for edge devices. **Source Code**

## TECHNICAL SKILLS

|                       |   |
|-----------------------|---|
| <b>Languages</b>      | C, C++, Go, PHP, Python, Rust, SQL, TypeScript                                      |
| <b>Frameworks</b>     | Actix, Axum, FastAPI, Flask, Keras, Node.js, React, Scikit-learn, TensorFlow, VueJS |
| <b>Infrastructure</b> | Ansible, Argo CD, AWS, Docker, GCP, Grafana, Kubernetes, Prometheus, Terraform      |
| <b>Databases</b>      | ElasticSearch, MongoDB, MySQL, PostgreSQL, VictoriaMetrics                          |