Abinesh Mathivanan

+91 6379473054 abineshmathivanan31@gmail.com linkedin.com/in/abineshmathivanan
github.com/Abinesh-Mathivanan abinesh-mathivanan.vercel.app

Education

KIT-Kalaignarkarunanidhi Institute of Technology Expected June 202l
B.Tech in Artificial Intelligence and Data Science (CGPA: 8.74 / 10.00) *Coimbatore, TamilNadu*

- **Relevant Coursework:** Data Structures and Algorithms (C), Data Science and Intro to ML, OOPs (Python), Discrete Mathematics, Probability and Queuing Theory, Database Management Systems, Networking & TCP/IP

Experience

PhobosQ Dec 2024 – Apr 2025
Software Engineering Intern *Coimbatore, India*

- Engineered the Luna project interface using TypeScript, React, and ShadCN UI, orchestrating AI frameworks like Langchain and Gemini APIs.
- Implemented multi-threaded word document processing, achieving 200+ pages parsed in under 58 seconds and voice-enabled systems with $<$3 ms latency.
- Involved in building foundational models for native languages and Vision extraction for regional language scripts

Artizence Apr 2023 – Sep 2023
Machine Learning Intern *Lucknow, India*

- Implemented Computer Vision models such as Haar Cascade, YOLO v5, and YOLO v7 for real-time surveillance projects. Completed projects on text detection and extraction using hybrid SVM + HMM and enhanced the model accuracy by 20%.

Technical Skills

Languages: Python, C++, Go, Javascript, Typescript, Tile-lang, SQL
Frameworks: PyTorch, Triton, Jax, TensorFlow, Keras, Tailwind, Next.js, Astro.js
Databases:: MySQL, MongoDB, Postgres, ChromaDB, Cassandra
DevOps & Tools: Git, Github, Docker, Azure ML, Firebase, Supabase, Node.js, Bun, ONNX Runtime, Linux
Focused on: Inference Optimization, Foundational Model Engineering, Distributed Training, Building Frameworks, Engineering Networking architectures and XLA-Programming

Projects

DeepCode — LeetCode for ML | *Next.js*, *TypeScript*, *Python*, *Rust*, *Docker*, *Firebase*, *Google Bigtable*, *Cassandra*

- Built a ML problem-solving platform with a Next.js and a scalable backend with Python and Rust for code execution, containerized with Docker.
- Deployed using Firebase (backed by Google Bigtable) and Render, implementing distributed caching systems ensuring 99% availability for real-time inference. Achieved 1K+ API hit requests within the first 24 hours of launch.

Beens-MiniMax - LLM | *Python*, *PyTorch*, *HuggingFace*, *Triton*, *LaTeX*

- A 103-million-parameter SLM with hybrid attention mechanism (softmax & lightning) following MoE style architecture. Trained with Wikitext-103 for the base model and Ultrachat_200k for Instruction-SFT
- Trained in Kaggle 2x T4 GPUs for 15 GPU hours & Instruct-SFT trained for 6 GPU hours. Model weights and technical report was published in Github and Kaggle

go-torch | *Go*, *Python*, *OpenCL C*, *BLAS*

- A simple torch-like deep learning library written in Go. Implemented go-routines and BLAS to the core, performing 1024 x 1024 in 118ms
- Made the library 115x faster comparing to the first version and wrote custom intel kernels

in-love.js — JavaScript Library | *JavaScript*, *TypeScript*, *Node.js*, *npm*, *Git*, *Tailwind*

- Created an open-source JavaScript library for love-proposal sites, using node modules such as Babel.js, Webpack, and Puppeteer. Testing modules such as Mocha, Jest, and Chai were implemented for real-time testing across webs.
- Published on the npm official platform and crossed 100+ downloads within the first week of launch.