DHIVYA SREEDHAR

412-954-7892 ⋄ San Francisco Bay Area
dhivyasreedhar@gmail.com ⋄ linkedin.com/in/dhivya-sreedhar-03b541168/

EDUCATION

Carnegie Mellon University December 2025
Master of Science (MS) - Information Systems (Machine Learning & Natural Language Processing), Pittsburgh, PA

Relevant Coursework: Advanced Natural Language Processing, Deep Learning (PhD), Applications of LLMs, Generative AI
Leadership: Graduate Teaching Assistant – 11-785 Deep Learning (Flagship PhD Level CMU course - 400+ students)
Activities and societies: Women in Business, Swartz Center of Entrepreneurship, Corporate Startup Lab, AI Venture Studio
Winner: ProductWars 2024; Finalist: Hack-a-Startup 2025 (top 5)

Anna University May 2022
Bachelor of Engineering(BE) - Computer Science Engineering Chennai, India

WORK EXPERIENCE

Founding Engineer June 2026 – Present
dCortex *San Francisco Bay Area,USA*

- Led end-to-end research and development of a new AI product, now **live with airline customers**, secured funding at **Bessemer Venture Partners**.
- Developed custom constraint solvers, optimizers, and execution schedulers for autonomous agentic workflows, including computer-use agents
- Drove **daily accuracy and adoption improvements for crew reserve planning**: analyzed operational data, identified failure points with OCC staff, and implemented **same-day fixes**
- Built a LightGBM model forecasting daily crew reserve demand over a 30-day horizon using rostering data, weather, event calendars, and disruption feeds

Founding Engineer January 2026 – June 2026
Robot Toolworx (Stealth Startup) *New York, USA*

- Built a production multimodal RAG system for industrial clients — ingesting CAD drawings, P&ID schematics, and hand-drawn engineering diagrams that standard text pipelines could not parse
- Architected a cross-modal Knowledge Graph (extending LightRAG) with image nodes as first-class entities — four vector representations per image in Vertex AI, explicit part-number edges from drawing analysis for high-confidence retrieval over visual content
- Designed a multi-hop query engine (DPM) chaining KG traversal, graph expansion, and cross-document reference resolution in parallel - enabling answers that span multiple referenced documents rather than single-chunk retrieval
- Built an LLM-as-judge eval layer scoring outputs on relevance, fidelity, and utility; closed-loop production feedback continuously improves retrieval without manual intervention

Founding Engineer - hire #1 May 2025 – August 2025
Reclamation Factory (CMU Robotics Startup) *Pittsburgh, USA*

- Joined as **hire #1** and led a team of 5 across ML development, go-to-market, and product execution; built the acoustic-and-visual AI pilot from scratch to classify hard-to-sort recyclable plastic
- Trained AST/ViT models with cross-attention fusion on 20K samples, achieving 93.5% accuracy across six material classes; reduced Raspberry Pi inference from 150 ms to 30 ms using pruning, FP16, and ONNX
- Demonstrated the pilot to prospective customers through demos and technical talks, helping secure the **company's first paying customer, generate revenue, and build a pipeline of additional pilots**
- Co-filed a patent on the approach; used experimental results to inform a pivot in product strategy and articulate the technology's differentiation to investors

Software Developer August 2022 — August 2024
Zoho Corporation - Part of the Manage Engine - Log360 Cloud OD Team *Chennai, India*

- Designed high-throughput distributed data ingestion pipelines processing millions of log events, enabling real-time anomaly detection and behavioral analytics across globally distributed tenants
- Led the end-to-end development of a scalable, containerized search feature using **Docker**, **Spark**, and **EC2**, integrated with SQL databases. Achieved a **40% reduction in search latency** while enabling real-time analytics across globally distributed tenants in a high-availability environment

PROJECTS

emg2qwerty — *PyTorch, TensorFlow, Signal Processing* Led the development of a neuromusculoskeletal interface translating surface EMG signals into text input for AR/VR and spatial computing platforms. Achieved <5% CER and <30ms latency using a hybrid Conformer–Transformer with spectral feature extraction, self-attention, and CTC loss. Built a real-time beam search decoder with Flan-T5 and GPT-4 Turbo for autocorrection. Introduced EMG-specific augmentations (SpecAugment, RandomBandRotation, Temporal Jitter) and causal modeling for low-latency inference. Applied large-scale, high-dimensional time series analysis for robust cross-user generalization.

MyTorch — *Python, NumPy, PyTorch* Built a custom deep learning library from scratch with an Autograd engine for forward/backpropagation, loss functions, optimizers, linear/convolutional/recurrent layers, batch normalization, and pooling. Implemented MLPs, CNNs, LSTMs, RNNs, GANs, GNNs, and GRUs, showcasing deep learning and probabilistic modeling capabilities.

SKILLS

Programming/Scripting Languages: Java, Python, C++, C, C#, MySQL, PHP, Javascript, HTML5 / CSS3
Full Stack: FastAPI, Flask, React, Node.js, AWS, GCP (Vertex AI), Docker, Kubernetes, Kafka, Spark
AI/AI Agents: LangChain, LangGraph, RAG, Knowledge Graphs, MCP, LLM fine-tuning, context engineering, agentic design
ML: PyTorch, TensorFlow, HuggingFace Transformers, LoRA/PEFT, SigLIP, DINOv2, OpenCV, Scikit-learn