

# THILAK MOHAN

Palatine, Illinois | (847) 393-6601 | thilakcmjs@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

## EDUCATION

University of Maryland, College Park, MD

Aug '23 - May '25

**Masters of Science in Applied Machine Learning**

**3.9 GPA**

*Relevant Coursework:* Multimodal Foundational Models, Computer Vision (CV), Introduction to Optimization, Cloud Computing (AWS)

Vellore Institute of Technology, Vellore

June '19 - May '23

**Bachelors of Technology, Electronics and Communications Engineering**

**8.98/10 CGPA**

*Thesis:* Pedestrian Detection and Trajectory Prediction to minimize risk of accidents in Autonomous Cars using YOLOv8 and STGCNN

## SKILLS & CORE COMPETENCIES

- **Languages & Libraries:** Python, MATLAB, C++, R, SQL, Django, bash, zsh, Scikit-Learn, TensorFlow, Keras, PyTorch, OpenCV, Pandas, Seaborn, Matplotlib, Plotly, NLTK, spaCy, BeautifulSoup, SHAP, Streamlit, Tkinter, Selenium, Optuna, pytest
- **Tools & Platforms:** Firebase, AWS (ECS, Fargate, ECR, CodePipeline, CodeBuild, S3, SageMaker), Docker, Kubernetes, GitHub, Github Actions, Tableau, Jira, Hugging Face, PostgreSQL, nano, vim, Weights and Biases, ClearML, SLURM, SMTP

## WORK EXPERIENCE

PEP School V2 Montessori, **Founding AI engineer** | (Freelance) Firebase, MCP, RAG, Whisper API, React + Vite

July '25 – Present

- Deployed a SaaS as a solo developer across **3 branches (~100 teachers)**, **handling and working with 1000+ monthly voice/text notes**
- Engineered an LLM inference engine (“AI Coach”) with a systematic prompt library, few-shot prompt tuning, structured outputs, and evaluation framework to benchmark model performance and ensure consistent and aligned responses
- Built MCP-powered chatbot enabling teachers to use fuzzy natural language to execute in-app functions—an easier alternative to UI navigation

CDUS Trading LLC, **Data Science Intern**, Chicago, USA

May '24 - Aug '24

- Built an intraday MES futures trading engine in Python using prior-day features and a 1:1 risk-reward strategy; backtested over 12 months, consistently achieving ~22% annual profit margins while handling 3 trades/day
- Automated the pipeline via GitHub Actions (CI/CD) for pre-market data ingestion, feature engineering, inference, and SMTP alerts; added monthly retraining to counter data drift, pytest-based stepwise validation, and monitoring, improving reliability and reducing manual intervention by 90%.

University of Maryland, PRG Lab, **Graduate Research Assistant** | Guide: Dr. Yiannis Aloimonos

Mar '24 - Aug '24

- Transformed hand joint coordinates from local to global reference frames, improving spatial accuracy and action-understanding in egocentric videos
- Enhanced physical modeling by parametrizing action-based models, improving interpretability and prediction in complex motion sequences

University of Maryland, **Teaching Assistant** | Guides: Dr. Alejandra Mercado, Dr. Jerry Wu

Aug '24 - May '25

- Clarified ML and CV concepts for students during weekly TA hours, reinforcing my core understanding of the fundamentals

## PROJECTS

**GPT-2 Stripped: A Comparative Analysis** | PyTorch, GPU-training, DDP, Transformers

Sep '24 - Dec '24

- Built a GPT-2 model (**124M parameters**) **from scratch** and trained it on the **10-billion-token** FineWeb-Edu dataset using Distributed Data Processing (DDP) across **A5000 and A6000 GPUs** for ~50 hours (~20k epochs).
- Implemented 2024 attention (sparse and FLASH) and positional encoding (ROPE, KERPLE, FIRE) variants **from scratch in PyTorch**, achieving improved generalization on **HellaSwag (+10% accuracy vs baseline)**, providing insights into scalable transformer design.
- Leveraged **Weights & Biases** for continuous experiment tracking, attention map visualization, and gradient monitoring; implemented automated logging to detect and resolve training crashes, improving reliability of long-horizon runs.

**Taxi Demand Prediction Platform** | AWS (ECS, Fargate, ECR, CodePipeline), Docker, FastAPI, ClearML, Optuna

Feb '25 - May '25

- Built a hybrid LSTM-XGBoost (sklearn) forecaster for Taxi Demand using live weather/traffic data, improving RMSE and MAE by over 88%.
- Deployed FastAPI inference via Fargate with auto-scaling and CI/CD (CodePipeline, CodeBuild, Docker) to handle variable city-scale workloads
- Automated retraining, experiment tracking, and model versioning with ClearML Agents and Optuna, using IaC to streamline cloud deployment.

**Image Captioning** | Python, Tensorflow, Weights and Biases, Multimodal model

Mar '24 - May '24

- Unfroze and fine-tuned an image captioning model using CNNs and Transformers, leveraging pre-trained CNNs on the MS COCO dataset to achieve respectable performance in a constrained amount of time
- Tracked and evaluated model versions using Weights and Biases, optimizing performance with real-time monitoring

## HONORS

InfoChallenge '25 (Hackathon by **Ernst & Young**), University of Maryland; placed 1/8 teams, **won \$100 cash prize**

Mar '25

- Engineered a comprehensive parking management system for UMD with a Flutter-based user chatbot and admin console, implementing microservices architecture using Flask APIs to handle permit validation, lot management, and real-time availability tracking, streamlining operations for both users and administrators.

Datathon '24 (Hackathon hosted by **Deloitte**), University of Maryland; placed 6/53 teams, **won \$500 cash prize**

Mar '24

- Conducted a **strategic market assessment**, including **data analysis**, **market potential evaluation**, and **cost analysis for NBA expansion cities**.
- **Scraped and analyzed Reddit fan sentiment**, applying **clustering algorithms** to identify the best non-NBA cities for expansion.”
- Presented findings through Tableau visualizations and Canva to **6 Deloitte executives** and **~50 stakeholders**.