Mohammad Moazam

Mobile: +3375449986

Portfolio: https://mzmmoazam.github.io
Email: mzm.moazam@gmail.com

Github: https://github.com/mzmmoazam
LinkedIn: https://www.linkedin.com/in/mzmmoazam

PROFESSIONAL SUMMARY

Machine Learning Engineer specializing in LLMs and Computer Vision, with expertise in fine-tuning, efficient deployment on edge and cloud using advanced quantization techniques. Experienced in leading AI engineering teams and delivering production-ready solutions with measurable impact. Patent holder in automotive AI applications.

EXPERIENCE

- **Stellantis** — Paris, France
  Machine Learning Engineer (Full-time) — *Nov 2022 - Present*
  - **GenAI Tech Lead for Personal Assistant**: Leading a team of 6 AI engineers to build a proactive & personalized **Copilot** for automotive applications. Implemented custom evaluation framework using both classical NLP metrics and LLM-as-Judge approaches for continuous monitoring and improvement.
  - **Patents & IP**: Filed and received approval for 2 patents out of 5 submissions for GenAI use cases in Automobile domain:
    * "Method and device for providing vehicle operating information" (FR3153459A1)
    * "Method and device for providing information about an operating anomaly affecting a vehicle" (FR3153460A1)
  - **3D Head & Eye Pose**: Completed end-to-end project from data strategy to model optimization. Evaluated SOTA models and fine-tuned on synthetic and real-world data to optimize performance for automotive use cases. Implemented deployment pipeline for edge devices.

- **Boxy** — Paris, France
  Computer Vision Intern (Full-time) — *Feb 2022 - Sep 2022*
  - **Thesis Work**: Researched and implemented multi-perspective 3D pose estimation and tracking for real-time applications.
  - **Production Model & API**: Developed 2D pose estimation model with 94% accuracy and built scalable REST APIs for production environment using Flask. Implemented optimizations for real-time processing.
  - **MLOps & Automation**: Created data processing pipelines using Airflow, reducing manual processing time by 70%. Implemented CI/CD flows on GitLab for automated testing and deployment.

- **Laboratoire Hubert Curien** — St. Etienne, France
  Research Intern (Full-time) — *Jun 2021 - Sep 2021*
  - **Benchmark Performance**: Evaluated SOTA models and benchmarked performance against domain-specific datasets, creating performance analysis framework for future research.
  - **GAN Implementation**: Implemented Pix2Pix GAN with custom WGAN loss, resulting in 35% higher efficiency and improved image quality compared to baseline models.

- **Valuelabs** — Hyderabad, India
  Senior Software Engineer (Full-time) — *Jul 2019 - Nov 2020*
  - **Predictive Maintenance**: Developed fault detection and forecasting system using time-series analytics on historical data, achieving 87% prediction accuracy.
  - **Open-Source Contribution**: Created JsonSpark, a Python wrapper for PySpark that simplifies complex JSON data processing with 40% code reduction.

- **Yildiz Technical University** — Istanbul, Turkey
  Product Development Intern (Full-time) — *Apr 2018 - Jun 2018*
  - **Medical Imaging Web App**: Designed and developed web application for biomedical imaging analysis using JavaScript and Python backend.

EDUCATION

- **Université Jean Monnet & University of Eastern Finland** — France & Finland
  *Erasmus Mundus Joint Master Degree (PSRS); **Grade: B, Rank: 4*** — *2020 - 2022*
  - **Specialization:** Photonics and Machine Learning
- **Hindustan Institute of Technology & Science** — Chennai, India
  *Bachelor of Technology - Information Technology; **GPA: 9.44/10*** — *2015 - 2019*
  - **Key Courses:** Machine Learning, Artificial Intelligence, Data Structures, Algorithms, Operating Systems

Technical Skills

- **AI/ML**: LLMs (fine-tuning, quantization, evaluation), Computer Vision (pose estimation, tracking, segmentation), Deep Learning, Model Optimization

- **Languages**: Python (advanced), C++ (intermediate), JavaScript, Bash, Dart

- **Frameworks**: PyTorch, HuggingFace (Transformers, PEFT), OpenCV, Keras, Flask, RESTful APIs

- **MLOps & Tools**: Airflow, Docker, Git, CI/CD, Slurm, SQLite, Model Monitoring

- **Soft Skills**: Team Leadership, Project Management, Technical Writing, Public Speaking

Patents & Honors

- Stellantis recognizes as top talent in AI domain within the company. (2025)

- Patent: "Method and device for providing vehicle operating information" *[Explains the methodology to interface car info using LLMs]* (FR3153459A1, 2025)

- Patent: "Method and device for providing information about an operating anomaly affecting a vehicle" *[Explains how to troubleshoot faults from CAN signals using LLMs]* (FR3153460A1, 2025)

- Patent: "Method and device for providing information about an operating anomaly affecting a vehicle" [World wide patent] (WO2025062082A1, 2025)

- Stellantis Genius Award for best Patent on the topic "GenAI-based Vehicle Troubleshooting System" (2023)

- Ranked 4th in Master Degree (2020)

- IDEXLYON Scholarship recipient for Erasmus Mundus Joint Master Programme PSRS (2020)

- Erasmus+ Exchange Programme Scholarship (2018)

- Meritoriously passed Bachelor's with 9.44 CGPA (2019)

- Certificate of Merit for Best Bachelor's Research Project (2019)

Open Source Projects

- **OCR Flask App (2021)**: Web application with OCR capabilities via UI and REST APIs using Flask and Tesseract. github.com/mzmmoazam/ocr-flask-app

- **irisSeg (2020)**: Python package for iris and pupil segmentation using Daugman's integrodifferential operator. Available on PyPI with 500+ downloads. github.com/mzmmoazam/irisSeg

- **jsonSpark (2019)**: Python package that simplifies conversion of PySpark JSON dataframes into Python dictionaries, reducing code complexity by 40%. github.com/mzmmoazam/jsonSpark

- **Real-Time Face Recognition (2018)**: Tool that collects facial data via webcam and performs real-time recognition using TFLearn. github.com/mzmmoazam/face-recognition

Professional Development

- **Run LLM on Edge Devices**: University of Paris-Saclay & Pruna AI (2025) - Specialized training on optimizing and deploying LLMs on resource-constrained environments

- **Deep Learning with PyTorch**: Neural Style Transfer specialization with focus on GANs (2022)

- **Machine Learning Fundamentals**: Udacity Professional Track (2021)

Languages

- English (Proficient), Kashmiri (Native), French (Intermediate), Turkish (Basic), Hindi (Proficient), Urdu (Proficient)