

MEELAD DASHTI

Enschede, Netherlands

+31 6 1733 6317 meeladd7@gmail.com [linkedin.com/in/MeeladDashti](https://www.linkedin.com/in/MeeladDashti) github.com/meldashti

Education

University of Twente

Master Thesis – Hyperbolic Compositionality in Vision-Language Models

July 2025 - March 2026 Expected

Enschede, Netherlands

Politecnico di Torino

M.Sc. in Computer Engineering, Specialization: AI and Data Analytics (104/110 Expected)

2023 – Apr 2026 Expected

Turin, Italy

Yazd University

B.Sc. in Computer Engineering, Specialization: Software Engineering

2018 – 2022

Yazd

Experience

Zirak srl

AI/ML Engineer (Remote)

Feb 2025 – Present

Turin, Italy

- Architected a **production-focused AI meeting assistant** with multi-model transcription (WhisperX, MLX Whisper) and streaming inference, achieving **sub-2s latency** and 95% transcription accuracy.
- Built a real-time **LLM moderator** using **LangGraph** with safety controls, semantic deduplication, and human-in-the-loop review, improving interaction reliability and reducing repeated output by **80%**.
- Implemented a **Retrieval-Augmented Generation (RAG)** pipeline using ChromaDB (BGE-M3 embeddings) and cross-encoder reranking to enable context-aware reasoning over historical meeting data.
- Developed **agentic workflows** integrating multi-provider LLMs (Groq Llama-3.3-70B, LM Studio) with a 9-stage reasoning loop and **sub-10s end-to-end** processing.
- Containerized and deployed FastAPI microservices** with async WebSocket streaming, persistent state management, and typed interfaces bridging Python inference services with a React/Electron frontend.
- Migrating core components to **AWS** (OpenSearch, S3, ECS, Bedrock) to support **scalable inference**, vector retrieval, and **production deployment** pipelines.
- Built a **Talk-to-SQL agent** (Qwen-3-14B + LangGraph) with validation checkpoints, SQL safety checks, and human-in-the-loop approval for secure database interaction and structured outputs.

AROL Group & Politecnico di Torino

AI/ML Engineer (Industry Collaboration Project)

Jun 2024 – Feb 2025

Turin, Italy

- Fine-tuned** a LLaMA-3B model using **LoRA (Unsloth)** with 4/8-bit quantization for efficient training and inference, and built the supporting pipeline including scraping (Scrapy + Firecrawl), dataset creation, and supervised evaluation.
- Integrated Retrieval-Augmented Generation with **Pinecone** vector search to ground answers in internal documentation and improve factual accuracy.
- Deployed the chatbot on Hugging Face Spaces with GPU inference and a Gradio frontend, providing real-time querying, model versioning, and a hand-off ready deployment package for future website integration.

Pishgaman Innovation Accelerator

Software Engineer

Dec 2021 – Feb 2023

Yazd

- Developed and shipped an Android application using MVVM, Retrofit, and RESTful JSON APIs, enabling city-wide complaint reporting and streamlining citizen-municipality workflows.

Publications & Projects

Enhancing Software Maintainability through LLM-Assisted Code Refactoring (PROFES 2025, Springer LNCS)

- Developed a fully automated LLM-driven refactoring system using **LangGraph**, static analysis (Radon, CQI, Bandit), and iterative prompting, demonstrating measurable improvements in software maintainability and outlining integration pathways into CI/CD workflows.

Master Thesis – Hyperbolic Compositionality in Vision-Language Models | Research

Ongoing

- Designing and scaling **PyTorch** pipelines on a **GPU cluster** with **DINOv3-7B**, **CLIP**, and **Qwen-VL**, analyzing hyperbolic embeddings to improve compositional reasoning in vision-language models for multimodal AI.

Technical Skills

AI/ML & Cloud: PyTorch, TensorFlow, LangChain, LangGraph, RAG systems, vector DBs (Pinecone), LoRA/QLoRA fine-tuning, prompt engineering, computer vision, NLP, diffusion models, WhisperX, FasterWhisper, CLIP, HuggingFace, scikit-learn, AWS (S3, Lambda, ECS, OpenSearch, SageMaker, Bedrock, IAM, CloudWatch)
Software Engineering & Tools: Python, Java, C++, JavaScript, SQL, Kotlin, FastAPI, WebSockets, Docker, Git, REST APIs, Electron, React, MCP servers, pandas, NumPy