Recep Kaan Karaman

kaanrkaraman@gmail.com | +90 538 403 34 48 | LinkedIn | GitHub | Google Scholar | ORCID

Experience

Undergraduate Researcher (TÜBİTAK 2247-C STAR) December 2025 – Present

- Selected as a funded researcher under the TÜBİTAK (National Research Council of Turkey) 2247-C STAR program to contribute to a project on synthetic histopathology image generation
- Investigating generative modeling approaches (GANs, diffusion models) for high-fidelity and diagnostically relevant tissue image synthesis
- Building reproducible end-to-end experimentation workflows covering dataset management, model training, evaluation, and versioning

AI Engineer Working Student, Hagia Labs – Singapore (Remote) October 2025 – Present

- Developing agentic workflows for internal automation, focusing on multi-agent coordination and LLM-based task planning
- Built and deployed RAG systems for document retrieval and knowledge management, improving query accuracy by 38%
- Implemented AI-powered workflow automation solutions, reducing manual review time by 127% and improving process efficiency
- Designed and integrated LLM-based tools for enterprise applications across multiple industries

Team Lead — AI Engineer, Sema Aviation – Uludag University March 2025 – December 2025

- Led a 30-member interdisciplinary student team (mechanical, electrical, software) to design and deploy an autonomous UAV for TEKNOFEST 2026
- Implemented a lightweight YOLO-Seg + MobileSAM system to achieve reliable, continuous visual lock on small UAV targets.
- Built real time fault detection pipelines using Kalman filters and Gaussian Mixture Models for UAV telemetry
- Integrated YOLO and OpenCV for real time object tracking; deployed ONNX models on Jetson with sub-50ms latency

Awards & Scholarships

Research Grant Recipient, *Thinking Machines Lab, 2025* Tinker AI Research Grant Program

- Awarded a $5,000 Tinker AI Research Grant by Thinking Machines Lab (founded by former OpenAI CTO *Mira Murati*, ranked #9 on LinkedIn's Top Startups of 2025)
- Supported project: *Explainable UAV Anomaly Interpretation Using Large Language Models*, exploring LLM-based reasoning agents for explainable UAV anomaly interpretation and operator trust calibration
- Additional funding under review: **TÜBİTAK 2209-A Research Projects Support Program** for the same project

Funded Undergraduate Researcher, *TÜBİTAK*, 2025 TÜBİTAK 2247-C STAR Researcher Fellowship

- Selected for the competitive TÜBİTAK 2247-C STAR program, receiving national research funding to support work in AI-driven synthetic histopathology imaging

Publications & Preprints

Code2Doc: A Quality-First Curated Dataset for Code Documentation arXiv:2512.18748

- First author of a study introducing Code2Doc, a curated dataset of 13,358 high-quality function–documentation pairs across five programming languages, constructed via a multi-stage filtering and deduplication pipeline

RAG-Driven Data Quality Governance for Enterprise ERP Systems arXiv:2511.16700

- Co-author of a study proposing a multilingual RAG-based SQL generation and data-quality pipeline deployed on a 240k-record ERP system

Projects

Python Documentation RAG Assistant

- Developed a lightweight Retrieval-Augmented Generation (RAG) assistant for Python documentation using BeautifulSoup for large-scale web scraping (˜500 links in 5 minutes)
- Implemented semantic text chunking (1,200 characters with 200 overlap) and embedded chunks via transformer-based embeddings stored in Pinecone for vector retrieval
- Enhanced retrieval accuracy using a Cross-Encoder reranking stage, improving Hit@1 from 0.42 to 0.52 and top-3 accuracy to 77%
- Built a Streamlit interface to visualize the RAG pipeline and display generated answers using Ollama, integrating embedding retrieval, reranking, and generation steps

Convolutional Neural Network Library github.com/kaanrkaraman/cortex

- Developed a modular CNN framework in pure NumPy to deepen understanding of core deep learning mechanics (convolutions, pooling, backpropagation, optimizers)
- Achieved 89% accuracy on MNIST, demonstrating correctness of gradient propagation and optimization without external DL libraries

Education

Uludag University, B.Sc. in Computer Engineering – Bursa, Türkiye October 2023 – July 2027

- **GPA:** 3,36/4.00
- **Coursework:** Data Structures and Algorithms, Object-Oriented Programming, Automata Theory

Skills

Languages: Turkish (Native), English (C1), Dutch (A2)

GenAI & LLMs: Retrieval-Augmented Generation (RAG), LangChain, LangGraph, LoRA/QLoRA, Guardrails, OpenAI Embeddings, BM25, Multi-Agent Orchestration

Machine Learning: PyTorch, Scikit-learn, CNNs, NumPy, Pandas, OpenCV, YOLO, RF-DETR

MLOps & Tools: Azure, AWS, Google Cloud, Pinecone, Chroma, Airflow, n8n, FastAPI, Flask, Weights & Biases (W&B), MLFlow, ONNX, Git, GitHub CI, ROS 2

Programming Languages: Python, C++, C, TypeScript