

Ali Arda Fincan

+90 (539) 389 66 31 ardafincan@icloud.com
LinkedIn: /in/arda-fincan GitHub: @ardafincan HuggingFace: @aliarda

Experience

• ML ENGINEER, EARLY-STAGE AI AUDIO STARTUP – AUGUST 2026 - PRESENT (PART TIME)

- Fine-tuned a large open-source latent diffusion model for audio with LoRA adapters; built conditional generation models for a low-resource music domain.
- Designed the evaluation methodology: held-out splits by source and performer, fixed seeds and sampler settings across runs, FAD and CLAP as primary metrics, single-variable ablations.

• DEVOPS ENGINEER INTERN, PEAK GAMES – AUGUST 2026

- Built an internal AI agent in Go that updates cdk8s configuration across multiple AWS accounts and environments from natural-language prompts, replacing manual config edits across ~2000 services for the Cloud team.
- Deployed to dev and prod on AWS EKS as 5 containerized microservices; infrastructure defined with Terraform and cdk8s, caching on ElastiCache (Valkey).
- Wrote GitHub Actions pipelines for image builds and releases, with images stored in ECR.

• RESEARCH ENGINEER, MAGIBU AI – JUNE 2024 - PRESENT

- Pre-trained several ~50M-parameter language models in PyTorch on separate datasets; owned the full pipeline from data processing to training code and real-time monitoring on remote Linux hosts.
- Designed and implemented new experimental tokenization techniques. Reduced token count on Turkish text by 40% versus BPE at equal downstream accuracy.
- Built a semantic-similarity module and the API-based evaluation harness for closed models (GPT, Gemini, Claude) in the biggest Turkish LLM benchmark, TR-MMLU
- Co-authored 3 research papers as second author.

• CO-FOUNDER, COUTURE ISTANBUL – DEC 2023 - JUNE 2024

- Fashion brand using diffusion models to generate unique designs per customer.
- Built the image generation pipelines and integrated model outputs into the frontend.

Other Relevant Projects

- MCP Server for US Stock Markets; exposed live prices, company news, and analyst estimates to LLMs through Model Context Protocol.
- Legal Precedent Search - RAG pipeline over case law using embedding models, vectorDBs, PostgreSQL and Firestore.
- Contributed to an Open Source project called DocsGPT. Added request/token usage limiting for agents and origin whitelisting for API security. Also gained some experience with MongoDB.

Education

Yeditepe University, Bachelor's Degree in Computer Engineering - Ongoing (Grad. 2027)

Publications & Research

- Setting Standards in Turkish NLP: TR-MMLU for Large Language Model Evaluation – IEEE SIU 2025
- Tokenization Standards and Evaluation in NLP: A Comparative Analysis of Large Language Models on Turkish – IEEE SIU 2025
- Tokens with Meaning: A Hybrid Tokenization Approach for NLP
- **Peer Reviewer** – ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP), 2026