

Yağız Can AKAY *AI Engineer*

yagizakay11@gmail.com | linkedin.com/in/yagizcanakay | github.com/yahocan

Education

TOBB University of Economics and Technology, Bachelor's degree, Artificial Intelligence Engineering, Full Tuition Scholarship 09/2023 – 08/2027 | Ankara, Türkiye

Experience

Finlabs, Co-Founder / AI & Product Engineer 04/2026 – Present | Ankara, Türkiye

- Building a multi-source **financial intelligence platform** for **BIST and TEFAS** investors, unifying 4+ heterogeneous data streams (KAP disclosures, financial reports, price feeds, and news) into **technical, fundamental, and sentiment analysis** workflows.

Teknopar Industrial Automation, AI Engineering Intern 01/2026 – 04/2026 | Ankara, Türkiye

- Architected an **autonomous Text-to-SQL and Generative BI pipeline** to query massive industrial **PostgreSQL databases** under **strict on-premise compute constraints**; deployed quantized **Qwen3-14B** via local **vLLM**, leveraging **HyGE (Graph-Enhanced RAG)**, semantic schema mapping, and a **HITL** feedback loop to maximize query accuracy at minimal compute overhead.
- Engineered a Python-based **3D Digital Twin for industrial assembly simulation**, benchmarking generative assembly-planning frameworks (**ASAP, DiffAssemble**) and **Wan diffusion models** to automate mechanical assembly validation.
- Contributed to **Digital Thread** and **IIoT integration** in an **EU Horizon Europe R&D project**.

Alris, Co-Founder / Lead 01/2025 – 03/2026 | Ankara, Türkiye

- Led a **cross-functional team of 5** to build a **financial BI platform** combining fragmented data sources to **cut manual errors and speed up workflows**, securing **3rd Place nationally at TEKNOFEST 2025**.
- Architected a **multi-agent orchestrator** delegating to sub-agents for market data, web research, and **TCMB data via 3-tier RAG**, with **visualization and document agents** running Python in **E2B sandboxes** to render **interactive charts** and automate **Word, Excel, and PowerPoint** reports with embedded visuals, validated via **trajectory evaluations (LLM-as-a-judge)**.
- Built a **layout-aware multimodal retrieval pipeline** with **contextual chunk headers**, surfacing **tables and figures as visual source citations** with page references and **traceable reasoning**, while integrating **SPD-RAG deep research** and autonomous **news clustering with in-feed Q&A**.

Treomind, AI Engineering Intern 05/2025 – 08/2025 | Ankara, Türkiye

- Architected the core **multi-agent system** from scratch as the **sole AI engineer** for the MVP of **Promake.ai** (company later raised a **\$4M seed round**), enabling non-technical users to build **production-ready websites** via natural language.
- Built a **document intelligence pipeline** in **on-premises Docker** for enterprise PII redaction using **Azure AI Document Intelligence, GLiNER, and LLM validation**, guaranteeing zero data egress.
- Trained **DeepLabv3** on custom **CVAT-annotated datasets** to segment **sensitive zones** in claims for secure **LLM ingestion**.
- Engineered an **autonomous web agent** via an **MCP server** and **OpenAI Agents SDK** for local LLMs.

Publications & Research

SPD-RAG: Sub-Agent Per Document Retrieval-Augmented Generation 03/2026

arXiv:2603.08329 | Y.C. Akay

- Proposed a **hierarchical multi-agent RAG** framework assigning dedicated **sub-agents** to **isolated documents** via central coordination with **similarity map-reduce recursive synthesis**. Outperformed standard/agentive RAG baselines, achieving over **85% of full-context quality** at **38% API cost**.

MIT OASYS Lab — Recursive Language Models, Open Source Contributor 01/2026

- Resolved a critical recursive parsing defect disrupting **nested parenthesis structures** within the core RLM framework (Merged PR #75).

Projects

Early Fire & Smoke Detection

05/2026 – 08/2026

- Built an aspect-ratio-preserving detection pipeline for **6,500 images** using **640×640 letterboxing** and YOLO-to-xyxy remapping; characterized 4,812 training boxes via **IoU-distance k-medians clustering (k=6, 0.638 mean IoU)**.
- Achieved **0.931 mAP@0.5, 0.589 mAP@0.5:0.95, and 91.8% recall** on 1,300 held-out test images by fine-tuning **Faster R-CNN ResNet-50-FPN v2** via **staged unfreezing and MultiStepLR scheduling**; executed **7-axis sweeps** and rejected **SE attention, discriminative LR, and SWA**

Market-Driven Weak Supervision for Turkish Financial Sentiment

05/2026 – 08/2026

- Built an automated weak-supervision pipeline linking **176,740 BloombergHT articles (2010–2026)** to BIST 30 and ticker targets, deriving continuous and ternary labels via **event-study CAR metrics (120-day window, 5-day gap)**.
- Benchmarked **single- vs. dual-head Transformers (linear vs. MLP heads)** and **LoRA-adapted Qwen3.5-0.8B** against **zero-shot and TF-IDF baselines** on chronological splits; proved market-reaction labels fail to supervise sentiment (**MCC –0.007 to 0.013, AUROC $\approx$ 0.50, $R^2 \leq$ 0.0002**) due to **attribution and timing noise**.

Tsunamigenic Earthquake Classification

10/2025 – 12/2025

- Built a classification pipeline for **700 M6.5+ earthquakes (2001–2020, 154 tsunamis)**, mitigating catalog bias (**$r=0.65$ year artifact**, station coverage) via **NOAA/NCEI cross-referencing**, and engineering **5-km coastal distances**.
- Achieved **0.903 ROC-AUC with TabPFN** across a 7-model benchmark under 5-fold CV, outperforming tuned **LightGBM (0.828 ROC-AUC)**; F2 threshold tuning yielded **94.8% recall (0.773 F2, 5.2% false negative rate)**.

Skills

Programming Languages — Python, Java, SQL

ML & Frameworks — PyTorch, TensorFlow, Hugging Face, scikit-learn, vLLM, LangGraph, OpenCV

Backend, Data & Cloud — Docker, FastAPI, PostgreSQL, Git, Bash, Valkey/Redis, Microsoft Azure

Spoken Languages — Turkish (Native), English (Professional Proficiency), Spanish (Basic)

Interests

Music & Audio — Progressive Rock/Metal Guitar, Music Production & Film Scoring

Athletics & Fitness — Strength Training, Kickboxing