

MAHMUT SAFA ALICI

Software Engineer · AI Systems & Agentic Workflows · LLM Evaluation · Full-Stack Development

Istanbul, Turkey · alc.safa@gmail.com · +90 532 689 6344 · safalici.com · linkedin.com/in/safa-alici · github.com/Safalici

PROFESSIONAL SUMMARY

Software engineer with experience in production AI systems, LLM evaluation and benchmarking, and full-stack development. Recent work covers AI generation pipelines for documents, spreadsheets, presentations, and video; automated QA of generated output; agent orchestration and model selection; and TypeScript/Node.js backend reliability. 750+ AI benchmark and evaluation tasks completed across two engagements, plus hands-on industrial software work spanning ERP, factory-floor applications, and retrieval-based AI assistants.

WORK EXPERIENCE

Software Engineer | Apptribe

Aug 2026 – Oct 2026

- Developed and improved production AI generation pipelines and agent-based workflows for documents, spreadsheets, and presentations.
- Built automated evaluation and QA mechanisms to detect structural, visual, and content-related failures in generated output.
- Contributed to model selection and optimization, agent orchestration, and the reliability and lifecycle management of sandboxed execution.
- Investigated and improved video generation and editing pipelines and long-running backend workflows.
- Worked on TypeScript/Node.js backend services, cloud-based infrastructure, and deployment workflows.

AI Systems Evaluator / LLM Task Designer | Turing

Aug 2025 – Aug 2026

- Designed 500+ deterministic benchmark tasks for agentic AI systems across diverse software environments, including LibreOffice and Google Chrome, sustaining a pace of 10+ per week over nearly a year.
- Built automated evaluators with JSON schemas and Python logic, testing instruction–evaluator alignment and reducing false-failure paths.
- Identified evaluator loopholes and ambiguous success conditions, improving evaluator robustness through more deterministic success criteria.

LLM Developer – Prompt Engineering, Tooling & Evaluation | Outlier AI

May 2025 – May 2026

- Completed 250+ task-specific LLM development and evaluation assignments (4.5/5 average platform rating) across roughly six months of primary active work.
- Designed task-specific coding prompts and developed assignment-specific callable tools and functions for LLM tool-use scenarios, including weather- and calendar-oriented tasks.
- Evaluated AI-generated code and model outputs against correctness, quality, UI, and software-engineering criteria across JavaScript, TypeScript, React, HTML/CSS, and Node.js.

ERP Software Developer | Arden Sentetik Çuval

Jan 2024 – Feb 2025

- Built and deployed a full-stack ERP application that consolidated sales, purchasing, inventory, and financial tracking into a single production system.
- Automated VAT calculations, stock validation, and multi-item sales workflows, replacing repetitive manual checks and reducing pricing and inventory errors.
- Designed streamlined data-entry interfaces and operational dashboards on a SQLite-backed system.
- Extended the ERP into three factory-floor tablet applications for weaving, printing, and garment operations, sharing a common backend and enabling QR-based material traceability across production stages.

Software Developer | Ingreats

Mar 2024 – Nov 2024

- Co-developed a cross-platform culinary app (Android/iOS) with a real-time Firebase backend and an integrated multilingual AI assistant for recipe discovery and meal planning.

Web Application Developer | ERC Research

Feb 2024 – Apr 2024

- Built a document-processing web app with Turkish-language PDF extraction, Excel-to-PDF generation, and fuzzy-search retrieval.

EARLIER EXPERIENCE

Software Developer Intern | GetdeHealth

Aug 2023 – Nov 2023

Enhanced a full-stack health-data application with complex UI components and real-time data visualization.

Game Developer Intern | Digital Factory 34

Apr 2023 – Jun 2023

Contributed obstacle mechanics and performance optimizations to a production 3D endless-runner game.

Mobile Game Developer | Koç University, Economy Department

Feb 2023 – Mar 2023

Designed and led a two-player mobile research game with multiplayer mechanics and cloud-based behavioral data capture.

Game Developer (Jam) | Jingle Jam Game Jam

Dec 2022

Co-developed and shipped a playable Unity WebGL game within a 48-hour game-jam deadline.

SELECTED PROJECTS

Arden AI Assistant

In development

- Extending the Arden ERP and production-control system with an AI layer: converting roughly 8GB of technical documentation into a RAG/vector-based knowledge layer, with Qwen planned as the model layer.
- Developing tool calling so the assistant can query live operational data across the ERP and its factory-floor apps (weaving, printing, garment), alongside document retrieval.

Reel Generation Factory

In development

- Multi-stage AI orchestration system that generates short-form video from idea to rendered output: script, scene planning, visuals, voice/TTS, captions, and music.
- Provider-selection layer supporting both a cloud provider (Eachlabs) and local LLMs per capability, with persistent/resumable workflow state, cost/budget tracking, and a zero-cost partial-run test mode; multi-platform publishing planned next.

TECHNICAL SKILLS

Languages: Python, TypeScript, JavaScript (ES6+), C#, SQL, HTML5/CSS3

AI & LLM Engineering: LLM Evaluation & Benchmarking, Automated Evaluators, AI Agents & Orchestration, Prompt Engineering, RAG / Vector Retrieval, Tool & Function Calling, JSON Schema Design, TensorFlow

Backend & Infrastructure: Node.js, Flask, Firebase, SQLite, Jinja2, Cloud Services & Deployment, Sandboxed Execution

Frontend & Games: React, Pixi.js / WebGL, Unity, 2D/3D Game Mechanics, Multiplayer

Practices: OOP, SOLID, Design Patterns, Clean Code, Git

EDUCATION

Koç University

Istanbul, Turkey

Bachelor of Science in Computer Engineering