Yoav Hevroni

Tel Aviv, Israel
+972 52 870 1646 | YoavHevroni1@gmail.com | GitHub | LinkedIn

SUMMARY

Full-stack Developer with hands-on AI engineering experience across LLM applications and speech-to-text systems. Builds end-to-end products with TypeScript, React, Node.js, and Python, integrating OpenAI and Anthropic APIs. Experienced in multi-stage LLM pipelines, structured outputs, model evaluation, and human-in-the-loop workflows.

EXPERIENCE

Full-stack Developer **2023 – Present**

Self-employed, Tel Aviv, Israel

- Designed and built multiple end-to-end web applications, owning architecture, frontend, backend, and deployment.
- Developed AI-powered analysis, transcription, and research workflows using OpenAI and Anthropic APIs.
- Implemented asynchronous systems, real-time flows, and external integrations (APIs, scraping, data pipelines).

Full-stack Developer **2022 – 2023**

Cogniteam – Nimbus, Tel Aviv, Israel

- Built and maintained a production real-time application using React, TypeScript, Redux, and gRPC.
- Built reusable UI components and scalable design patterns with MUI5 and Styled-Components.
- Refactored legacy code and applied CI/CD practices to improve maintainability and release reliability.

SELECTED PROJECTS

GapHunter - AI Review Intelligence Tool

Next.js, TypeScript, Convex, OpenAI API, LangChain

- Built a multi-stage LLM pipeline for review classification, evidence-backed reports, and critic-guided revision, with task-specific prompts, JSON parsing, and source-linked citations.
- Implemented data cleaning, deduplication, token budgeting, and a 32-app evaluation suite integrated with GitHub Actions.

Transcripter - AI Transcription Platform

Next.js, TypeScript, Supabase, OpenAI Whisper

- Built asynchronous speech-to-text workflows with audio chunking, word-level timestamps, and editable transcripts.
- Integrated OpenAI, Deepgram, AssemblyAI, and Speechmatics into an evaluation framework comparing WER/CER, latency, and estimated cost against human-approved reference transcripts.

Trading Research Platform - LLM Strategy Assistant

Python, FastAPI, Anthropic Claude API

- Built an LLM assistant translating natural-language requests into strategy specifications, with versioned prompts, JSON Schema structured outputs, validation, corrective retries, and human approval.
- Implemented prompt caching, token usage tracking, request and concurrency limits, and per-request audit trails.

TECHNICAL SKILLS

Frontend	React, Next.js, TypeScript, Redux, TailwindCSS, MUI, Styled-Components, HTML, CSS
Backend	Python, FastAPI, Node.js, Express, REST APIs, gRPC, MongoDB, SQL, Supabase, Convex
Tooling	Git, GitHub Actions, Docker, Webpack, Vite
AI Engineering	OpenAI API, Anthropic API, LangChain, Whisper, Prompt Engineering, Structured Outputs, LLM Evaluation, ASR, WER/CER, Guardrails, Human-in-the-Loop
AI-Assisted Dev	Claude Code, GitHub Copilot; planning, implementation, testing, and refactoring

EDUCATION

B.Sc. in Mathematics & Computer Science
Ariel University 2017 – 2020

Full Stack Developer Bootcamp
Netcraft Academy Apr 2021 – Nov 2021

MILITARY SERVICE

Combatant, Kfir Brigade – Haruv Battalion 2009 – 2012

LANGUAGES

Hebrew (Native) | English (Fluent)