



# David Sheffer

Tech Lead & Full Stack Engineer

TypeScript · React · Angular · Node.js ·  
Python · SQL  
MCP · Agent orchestration · Local LLMs

BASED      Kiryat Tivon, North Israel  
EMAIL      dave.sheffer1@gmail.com  
LINKEDIN   linkedin.com/in/david-sheffer-b8b5a7132  
GITHUB      github.com/davesheffer  
SITE        dashdev.vercel.app

## PROFILE

Tech lead and full stack engineer. I'm the sole engineer behind a production multi-tenant SaaS platform at YLM Solutions and own it end to end, from SQL schema and Node.js services to Angular, React and Flutter clients.

I build the tooling that makes AI agents dependable in real codebases: Hunch, an open-source MCP server that gives coding agents a team's decision history; a multi-agent coding orchestrator; and experiments with local models, from llama.cpp inference to offline Whisper transcription.

*The best systems I've shipped came from good arguments with good people, then writing the code that settled them.*

## FEATURED OPEN SOURCE

**Hunch** · creator & maintainer

TypeScript · MCP SDK · Apache-2.0

Git-native engineering memory for AI coding agents, shipped as an MCP server and CLI on npm. It keeps agents from reintroducing decisions a team already rejected. It started as a solo project and now has outside contributors; I review their code and shape the roadmap with them.

- 28,000+ npm downloads since the June 2026 launch (8,400+ in the last 30 days), 170+ releases and 1,600+ commits.

[hunchmemory.com](https://hunchmemory.com)   [github.com/davesheffer/hunch](https://github.com/davesheffer/hunch)   npm [@davesheffer/hunch](https://www.npmjs.com/package/@davesheffer/hunch)

## SKILLS

|             |                                                                                                                          |
|-------------|--------------------------------------------------------------------------------------------------------------------------|
| Frontend    | TypeScript, React, Next.js, Angular, Vue.js, Flutter, Tailwind CSS, PWA, RTL / Hebrew UI                                 |
| Backend     | Node.js, Express, Fastify, Python, C++, REST APIs, JWT auth, background workers, PDF and email pipelines                 |
| Data        | SQL Server / Azure SQL, PostgreSQL, Prisma, Redis, SQLite, stored procedures, migrations, query tuning                   |
| AI & agents | MCP servers, multi-agent orchestration (Claude Code, Codex), llama.cpp, Whisper, BM25 hybrid search, LLM query expansion |
| Cloud & ops | Azure, Heroku, Docker, AWS S3, Firebase, GitHub Actions, backups and disaster recovery                                   |
| Languages   | Hebrew (native), English (native)                                                                                        |

## EXPERIENCE

**YLM Solutions**

Sep 2019 – Present · 7 yrs 1 mo

Yokne'am, Israel · Operations software for service businesses

Tech Lead & Full Stack Engineer

Mar 2026 – Present

Hybrid · Angular · Node.js · Express · Azure SQL · Flutter

Sole engineer on a production multi-tenant SaaS platform, from schema to deployment. I make every architectural call and write the code that backs it up: 1,000+ commits since March 2026.

- Led the full Angular migration of a live SPA with zero downtime.
- Build and run the Node.js / Express API on Azure SQL: JWT auth, stored procedures, migrations, query tuning, PDF generation and transactional email.
- Built a standalone notification service for scheduled push (Firebase) and email alerts, deployed and scaled separately from the API.
- Cut Azure and Heroku spend by 50% while strengthening backups and disaster recovery.
- Building the platform's Flutter app for field workers on Android, iOS and web, at feature parity with the web app.
- Shipped dashboards, chat-driven reporting, an RTL Hebrew PWA, QR flows and maps.

Frontend Developer

Aug 2023 – Feb 2026

On-site · React · Vue.js · JavaScript

- Migrated company forms from RDLC to JavaScript forms in Vue.js with JSON data, charts and complex calculations.
- Designed, built and launched an app for phone or email sign-in with chat inside the system. Led it from concept to deployment.
- Built internal React tools that automate barcode and QR generation, data formatting and image optimization.

Project Manager

HTML · CSS · WordPress · Wix

Turned customer requirements and work processes into system specs, and integrated technical data into company systems.

Worked with the development team on bugs and new features, and led QA testing.

Built and maintained the company's WordPress and Wix sites, writing the HTML and CSS for custom pages and layouts, optimized for SEO.

Sep 2019 – Aug 2023

dooble

Jan 2018 – Sep 2019 · 1 yr 9 mos

Project Manager

Ran WordPress projects from planning to launch across UX, design and development teams.

Earlier

Customer Support Engineer, אוטומציה החדשה · automated systems for the municipal sector

Nov 2016 – Jan 2018

QA Engineer, Fundtech (now D+H), Ramat Gan · system, integration and acceptance tests; Eurobank release upgrades; 24/7 production helpdesk

2000 – 2003

MORE OPEN SOURCE

Coding Orchestrator · creator

Python · Shell · MIT

An orchestrator and four subagent roles (scout, runner, builder, critic) for Claude Code and Codex. Bounded reading, commands and implementation go to cheaper models; design, judgment and verification stay with the orchestrator. Every agent reports its evidence and confidence, and risky changes get an adversarial review.

github.com/davesheffer/coding-orchestrator

adrkit · contributor

7 merged PRs · top outside contributor

Machine-readable architecture decision records (ADRs) with lint, CI and MCP.

Added stale-marker warnings: code that cites a superseded, rejected or deprecated ADR is flagged, with the live successor suggested.

Merged pull requests   adrkit.dev

Pull requests also submitted to litellm (proxy routing, dependency CVEs) and aider (cached-token cost accounting).

PROJECTS

PRODUCT PLATFORMS

WishScene · project lead, Coders Clan IL

Next.js · TypeScript · PostgreSQL · Vercel

Lead WishScene, an AI experience studio that turns one user into consistent photo, video and social content, for the Coders Clan IL developer community. Wrote the PRD, specs, technical design and roadmap, split the work into 18 scoped issues for clan developers, and built the pnpm monorepo and Vercel + Neon demo.

github.com/coders-clan/wishscene

WhatsApp automation platform · lead developer

Next.js · Prisma · PostgreSQL · Redis · S3

Multi-tenant messaging automation with a separate worker process for WhatsApp sessions, Redis rate limiting and S3 media storage. Main author, with 384 commits.

LOCAL MODELS & AI

Local LLM inference research

C++ · CUDA · Metal · Python

A reproducible llama.cpp patch series exploring Qwen3 mixture-of-experts routing, with CPU, CUDA and Metal kernels.

Local Whisper Transcriber

React · Python · Whisper

Offline speech-to-text: a React front end over a Python server running Whisper on the local machine, with no API keys and no audio leaving the device.

github.com/davesheffer/local-whisper-transcriber

Hybrid search lab

Python · BM25 · RRF

Catalog search research combining BM25 keyword retrieval, reciprocal rank fusion and LLM query expansion, with an evaluation harness and a demo server.