

Abhay Yemekar

Full-stack AI Engineer | Pune, Maharashtra, India | abhayyemekar.me

+91 9579065125 | abhay.s.yemekar@gmail.com | github.com/abhay-yemekar | linkedin.com/in/abhayyemekar

PROFILE

Software engineer with nearly 3 years of professional experience, now at Translytics building AI-powered features and backend services in Python for a supply-chain planning platform. I build AI products end to end (LLM apps, agents and RAG in Python and TypeScript) and treat model output as untrusted input: ground it, test it, log it. Outside work I ship open-source and published products, including HireLens (CLI on npm) and yDirect (Chrome Web Store).

EXPERIENCE

Software Engineer—Translytics, Onsite (Pune, MH) Mar 2026 – Present

- Build AI-powered features and backend services in Python for a supply-chain planning platform.
- Develop LLM-powered product features end to end, from design and implementation through QA and release.
- Design REST APIs and background jobs for data-heavy planning workflows.
- Improve how data moves in and out of the platform: imports, validation and exports.
- Contribute to the React and TypeScript product UI alongside the backend work.
- **Tech:** Python, Flask, SQLAlchemy, MySQL, Redis / RQ, pandas, LLM APIs, RAG, React, TypeScript, MUI, Docker.

Software Engineer—Tata Elxsi, Onsite (Pune, MH) Dec 2023 – Feb 2026

- Built Python tools and validation scripts following SDLC best practices (Git, Jira, unit and integration testing); automated 3+ reporting workflows, cutting manual effort by about 30%.
- Built a Python tool that converts legacy test cases from Excel, CSV, Word and plain text into Gherkin .feature files for BDD, with a web UI to upload, preview, edit and export.
- Contributed to embedded C / RTOS codebases using Git-based workflows.
- **GenAI Bootcamp:** built FastAPI applications with LLMs, RAG, LangChain and LlamaIndex, using Chroma and Postgres vector stores; also completed a Kafka bootcamp.
- **Client engagement: Mahindra & Mahindra, BE 6 and XEV 9E (Mar – Jun 2025)**
 - Wrote Python scripts to parse and analyse OBD diagnostic data and Diagnostic Trouble Codes (DTCs); ran ECU diagnostics over UDS (MDLR tool) to validate software configurations.
 - Automated Excel reports with OK / Not-OK status and OTA-impact tracking; did root-cause analysis with Centre of Excellence engineers and explored ML for predictive fault detection.

Intern—PHN Technology Pvt Ltd, Remote (Pune, MH) Jul 2023 – Oct 2023

- Built an end-to-end analytics pipeline for the ICC Men's T20 World Cup 2022: scraped all 45 matches from ESPNcricinfo with Bright Data, cleaned and linked them with pandas, and modelled a star schema.
- Delivered a 10-page Power BI report with DAX measures to pick a data-driven Playing XI (CricketStat Pro).

SDE Intern—BitMap Technology Pvt. Ltd., Hybrid (Pune, MH) Jan 2023 – May 2023

- Built a three-factor login for online banking (face recognition, email OTP, spoken passphrase) with liveness detection, using OpenCV with an LBPH model, a threshold-calibration tool, a threaded Tkinter UI, SQLite and SMTP.
- Demo results (~200 attempts): ~95% correct identification, ~88% of spoof attempts blocked, <50 ms per frame.

PROJECTS

HireLens: Glass-box AI Candidate Ranking | Nov 2025 – Present

hirelens-rosy.vercel.app | github.com/abhay-yemekar/hirelens | npmjs.com/package/hirelens

TypeScript | Next.js | Hono | PostgreSQL + pgvector | Drizzle | pnpm + Turborepo | Docker | OpenAPI 3.1

- Open-source hiring platform: turns a job description into an editable rubric and scores each resume per criterion, citing evidence spans at exact character offsets.
- Blind review, four-fifths-rule bias audit, hash-chained audit log and CI evals for accuracy and name-swap bias.

- Runs on Gemini, Claude, Groq, OpenRouter or local Ollama; v1.4.0 (Oct 2026) ships a CLI on npm (522 downloads in the 30 days to 11 Oct 2026) and Docker images on GHCR.

yDirect: Snippet Manager for Chrome | Jun 2026 – Present

chromewebstore.google.com/detail/ehkahipeoahnbfbahlkedgdcjefejihg

JavaScript | Chrome Manifest V3 | Firebase Auth | Firestore | Node.js

- Published on the Chrome Web Store (v1.3.3): save prompts, replies, templates and code snippets once, then search and copy them from Chrome's side panel without leaving the tab.
- No framework or build step: a service worker plus an offscreen document for Google sign-in; Firestore rules enforce access server-side, with a strict CSP and 16 Node test suites.

SearchScribe AI: Research-to-Draft Writing Studio | Nov 2025 – Present

searchscribe-ai.vercel.app | github.com/abhay-yemekar/SearchScribe_AI_Studio

Python | FastAPI | SQLAlchemy 2 | Alembic | Next.js 16 | React 19 | Gemini | Tavily | PostgreSQL

- Turns a topic into a researched, structured draft with Tavily web research and Gemini; six rewrite styles and an immutable version history, with sources kept through every version.
- The model never writes HTML: Jinja2 renders it and nh3 sanitises it. Runs on Vercel, Render and Neon Postgres; CI runs Ruff, mypy, pytest, ESLint, Vitest and Playwright.

ForecastGPT: Financial Outlook Agent | Oct 2025 – Present

forecastgpt-financial-outlook-agent.onrender.com | github.com/abhay-yemekar/forecastgpt-financial-outlook-agent

Python | FastAPI | LangChain | FAISS | pdfplumber | Ollama (LLaMA 3.2) | Redis / RQ | MySQL

- Reads Indian listed companies' results decks and earnings calls and writes a confidence-scored next-quarter outlook.
- RAG grounds every claim in a retrieved chunk; output is schema-enforced JSON (trends, risks, opportunities, confidence); local LLaMA 3.2 by default, OpenAI / Anthropic swappable; Redis / RQ jobs with MySQL logging.

SKILLS

- **Languages:** Python, TypeScript, JavaScript, SQL, C (embedded / RTOS), HTML/CSS
- **LLM apps & agents:** Claude, Gemini, OpenAI, Hugging Face, Ollama, LangChain, LangGraph, RAG, embeddings, vector search (FAISS, pgvector), evals, prompt engineering
- **Backend & APIs:** FastAPI, Flask, Node.js, Hono, SQLAlchemy, REST, OpenAPI / Swagger, JWT, SSE
- **Data & storage:** PostgreSQL, MySQL, Redis / RQ, Firebase, SQLite, Drizzle, Neon, Kafka (bootcamp)
- **Frontend & extensions:** React, Next.js, Tailwind CSS, Vite, MUI, Chrome extensions (Manifest V3)
- **ML & analytics:** pandas, NumPy, PyTorch, TensorFlow, OpenCV, Jupyter, Power BI (DAX, Power Query), Excel
- **DevOps & tools:** Docker, GitHub Actions (CI/CD), Vercel, Render, Sentry, pytest, Playwright, Git, Jira, Agile / Scrum, AWS / Azure (basics)

EDUCATION

- **B.Tech, Computer Science and Engineering**—SGGSIE&T, Nanded | 2019–2023 | CGPA 8.80/10; ranked 1st in my class in second year (9.88 SGPA)
- **Class XII**—Cambridge Junior College, Nanded | 2018–2019
- **Leadership:** Student Coordinator, Team Pragyaa (national socio-technical fest); Class Representative 2021–22; volunteer, Smart India Hackathon 2020

CERTIFICATIONS

- **AI Fluency: Framework & Foundations and Claude 101**—Anthropic
- **OCI 2025 Certified AI Foundations Associate**—Oracle—2025
- **Azure AI Essentials Professional Certificate**—Microsoft
- **Forward Program, Super Learner Certificate**—McKinsey.org—Oct–Dec'25
- **Software Engineer Certification**—HackerRank—Oct'25
- **Microsoft Python Development Professional Certificate**—Coursera—Feb'25
- **Career Essentials in GitHub** (Jul'25) and **in Generative AI** (Nov'24)—Microsoft & LinkedIn