

MUTHARASU A

Senior Software Engineer | Backend & AI Infrastructure

+91 93442 34157 | mutharasu.theofficial@gmail.com | Hyderabad, India | LinkedIn | GitHub

SUMMARY

Senior Software Engineer with 3 years building production backend systems where reliability is not optional. Architected a centralized Identity Provider (OAuth 2.0 / OIDC) serving 5 products, built real-time AI interview agents, and shipped codebase-intel, an open-source MCP server for AI coding agents (19 languages, 48 to 75% token reduction). **Three merged PRs in Typesense** (26K+ stars, C++): two production bug fixes and one ranking feature. I build the backend infrastructure that makes AI features actually work in production: auth, pipelines, real-time agents, and deployment.

TECHNICAL SKILLS

Languages: Python (Expert), C++, Java, JavaScript/TypeScript, SQL, Bash

Backend: FastAPI, Django, Spring Boot, Celery, Taskiq, Pydantic, asyncpg, Tortoise ORM, SQLAlchemy

Auth & Security: OAuth 2.0, OIDC, JWT (EdDSA), PKCE, SSO, token rotation, RBAC, rate limiting

AI & Real-Time Agents: MCP, LLM APIs, RAG, embeddings, pgvector, tree-sitter, real-time voice agents, barge-in, prewarming, prompt engineering

Systems Rigor: idempotency, reconciliation, partial-failure handling, backpressure, cascade revocation, audit logging

Databases: PostgreSQL, pgvector, Redis, Typesense, MySQL, SQLite, Elasticsearch

Data & Pipelines: Apache Airflow, PySpark, Celery/Redis worker pools, Apache Superset

APIs: REST, GraphQL, versioning, idempotency, Swagger/OpenAPI

Infra & DevOps: Docker, Kubernetes (GKE), AWS (EC2, S3), GCP, Terraform, Nginx, GitHub Actions, CI/CD, Git

Frontend & Desktop: React 19, Redux, Vite, TanStack Query, Electron | **Ways of working:** Agile/Scrum, code reviews, PyTest

EXPERIENCE

10xScale.ai, Hyderabad

Senior Software Engineer

2025 to Present

Centralized Identity Provider: auth for 5 products

- Architected an **OAuth 2.0 + OIDC** authorization server from scratch (FastAPI, 29 endpoints): JWT with EdDSA, PKCE, refresh token families with cascade revocation, and per-product token isolation. Replaced 3 fragmented systems, cut auth bugs ~70%, and enabled the org's first SSO
- Designed the **PostgreSQL** schema (26 tables, Snowflake IDs) with composite indexing and EXPLAIN ANALYZE tuning; **Redis** for sub-ms sessions and rate limiting. Timing-attack-safe, exponential backoff, zero user enumeration

InterviewPilot: AI-powered interview platform

- Built the platform where employers schedule **AI, AI+human, or human-only interviews** from a JD and weighted rubric. An AI agent runs the full interview to plan, adapts depth dynamically, and adheres to a strict time budget
- Engineered **reliability into the real-time voice agent**: barge-in handling and worker **prewarming** (replacing full-application crashes), graceful time-limit termination, and fixes for silent-agent and repeated-apology failure modes
- Built an **interview co-pilot agent** that assists human interviewers or takes over from the AI mid-interview; an **Electron desktop app** for candidates with in-app coding and system design, plus **ML-based proctoring**

CareerPilot: data & search platform

- Built **idempotent ETL pipelines** (Airflow + PySpark, 200K+ records); pgvector + Typesense hybrid search for a 10x latency improvement; deployed on **Kubernetes (GKE)** with KEDA autoscaling. Designed and ran **technical interviews for 16 candidates**

Software Engineer

2023 to 2025

- Built **Hire10x**, an AI recruitment platform, end to end: candidate import from LinkedIn, internal and local sources, embedding-based ranking, **Typesense** search, **GraphQL** APIs for JD management, and email plus WhatsApp outreach
- Owned the full stack: API design, PostgreSQL data modeling, React frontends, and CI/CD via GitHub Actions, Docker, and Nginx on AWS/GCP

OPEN SOURCE & IMPACT

Typesense (26K+ ★, C++), three merged PRs: two production bug fixes and one feature

- PR #2772 (bug):** /health probes timed out whenever bulk imports saturated the main thread pool, so orchestrators restarted nodes that were perfectly alive. Routed liveness checks onto the dedicated `meta_thread_pool`, a one-line change isolating probes from write load
- PR #3010 (bug):** Root-caused a **non-deterministic out-of-bounds read** in multi-clause sort scoring. Three identical requests returned three different orderings and never an error, because a shared cursor vector read past its allocation while CI stayed green. Fixed with per-clause cursor keying plus a regression test
- PR #3011 (feature):** Shipped a **sum mode for _eval**, cutting additive-ranking queries from **2ⁿ to n expressions**. Resolved parser ambiguity against existing filter syntax and handled `__int128` overflow

codebase-intel, MCP server for AI coding agents: dependency graphs, decision journals, quality contracts, and token-budgeted context assembly (**48 to 75% token reduction**). A 19-language code graph via tree-sitter, BFS traversal, and AI anti-pattern detection.

TEACHING

Taught Java, OOP, Spring Boot, and system design to **100+ students**; designed a microservices capstone. Able to explain complex technical concepts to technical and non-technical audiences alike.

EDUCATION

B.Sc. Computer Science

Government Arts College, Coimbatore