Richy Koshy

me@richykoshy.dev
Mumbai, India
linkedin.com/in/richy-koshy
github.com/richie-rk
richykoshy.dev

Python engineer: GenAI, backend and automation

I build RAG systems and automations that hold up in production.

Python engineer with **5+ years** of experience. For a US mortgage originator, owned delivery end to end of production GenAI, backend and automation systems: an **enterprise Azure RAG application** now used organization-wide for knowledge retrieval, a custom SDK for PRMI's loan origination platform, and **13 production automations** across compliance and investor operations.

Work experience

Software Developer

Primary Residential Mortgage, Inc. (PRMI)

US mortgage originator. Full technical ownership of MAX AI; sole engineer on the RPA automations.

- **MAX AI, enterprise RAG platform.** Python backend across three Azure services: FastAPI chat with NDJSON streaming, a Durable Functions ingestion pipeline, and Function App utility APIs. Azure AI Search hybrid retrieval (vector + semantic ranking) with HyDE as a fallback when recall is low, Azure OpenAI (GPT-4o, ada-002), Entra ID auth with multi-tenant access mapping in SQL Server, ingestion containerized via Azure Container Registry. **Now the primary knowledge-retrieval tool across PRMI.**
- **MAX AI front end.** Contributed React UI with MSAL / Microsoft Entra ID group-based access control, mapping internal and external user and admin groups to role-based permissions.
- **Final Docs refactor.** Rebuilt a monolithic automation into a config-driven, object-oriented dynamic-dispatch design (YAML → handler registry) with a pandas + SQLAlchemy reconciliation engine: keyed merges with rerun-safe, composite-key de-duplication and exception reports for ops review. **Cut per-investor onboarding ~75% and onboarded 4 new investors in 1 month.**

> *"PRMI now treats investor expansion as a business decision, not a dev-capacity constraint."*
> Corban Wells, VP of Technology Product Management, PRMI

- **Loan origination API SDK.** Python SDK wrapping the Encompass Cloud APIs (ICE Mortgage Technology's loan origination system): OAuth2 with token caching, concurrent paginated pipeline queries, safe loan updates (locks loans, skips active user-edit sessions), eFolder document CRUD, and disclosure tracking.
- **13 production RPA automations.** US mortgage / BFSI automations in Power Automate + Python, spanning ECOA adverse action, TRID, HMDA, MERS sync, final docs, HFA exceptions, Blend point-of-sale credit management, warehouse collateral, and suspense resolution across 9 investors. Includes 3 OCR pipelines (Tesseract + Azure Document Intelligence) extracting credit data from PDF reports.
- **Power BI data pipeline.** `Reporting DB → Blob Storage → Function App → SQL Server → Power BI dashboards`
- **Event-driven Smartsheet sync.** `Webhooks → Azure Service Bus → Function App → Smartsheet`, for branch partners.
- **Delivery ownership.** Led BRD and PRD gathering, produced flow diagrams, secured first-call sign-offs, coordinated UAT, and documented every delivered automation.

Skills

Python and frameworks

Python, FastAPI, asyncio, Multithreading, SQLAlchemy, Pydantic, pytest, Pandas

GenAI and RAG

Azure OpenAI, Azure AI Search, LangChain, LangGraph, Agno, Hybrid retrieval, Semantic ranking, HyDE, FAISS, BM25, Chunking and ingestion pipelines, Prompt engineering, RAG evaluation, VLM-based prompt-injection detection

Azure

Functions, Durable Functions, App Services, Container Apps, Container Registry, Service Bus, Blob / Table Storage, Key Vault, Managed Identity, Entra ID / MSAL, VNet integration, Bicep, Logic Apps

Data and integration

SQL Server, PostgreSQL, pyodbc, Encompass REST API, Blend API, Microsoft Graph (SharePoint, OneDrive), Smartsheet, Tesseract OCR, Azure Document Intelligence in Foundry Tools, Unstructured.io

APIs

REST API design, OAuth2, Webhooks, NDJSON streaming, Microservices

Automation and tooling

Power Automate, RPA, Playwright, Selenium, Docker, Git

US mortgage domain

ECOA, TRID, HMDA, MERS, Investor ops

Work experience, continued

Software Developer
Antz AI Solutions

- **HydroQ API (client: TechSpan Engineering, IoT).** Designed the FastAPI interface for the database CRUD operations of a multi-tenant remote water-quality monitoring platform, with a generic SQLAlchemy CRUD layer querying across PostgreSQL and MSSQL. **Replaced 2-4 hours of manual onboarding per site** with API-driven workflows; full endpoint test coverage.
- **CV-screening safeguard (internal).** Detects hidden-text prompt injection via VLM vs OCR mismatch, as an Agno agent tool.

Graduate Engineer Trainee
Goodyear South Asia Tyres

Industrial automation: PLC, SCADA, VFD, CTMS/BMS.

- Commissioned the Cut Tyre Measurement System (CTMS) and Bead Measurement System (BMS) alongside Rockwell and Honeywell engineers.
- Programmed and troubleshot PLCs and VFDs; SCADA and HMIs, including sensor calibration; Python (pylogix) scripts pulling drive data from Rockwell PLCs.

Education

BE, Electrical Engineering
University of Mumbai

CAD Intern

Awards

Best Paper
Intl. Conf. on Advanced Trends in Engineering

1st Place
Tantragyan National Project Competition

Reference

Corban Wells
VP of Technology Product Management, PRMI.
Recommendation letter on request.

> *"Conscientious ownership, earnest curiosity, and diligent follow-through."*

Read more

MAX AI deep dive
Grounded answers from the company's own documents, with citations.
`richykoshy.dev/max-ai`

Final Docs deep dive
Treat an investor as data, not code.
`richykoshy.dev/final-docs`

Recommendation letter
What the VP put in writing.
`richykoshy.dev/letter`