

Suraj Zala

Gandhinagar, Gujarat, India

hello@essnine.dev - github.com/essnine

Summary

I am proficient in Python, have experience in building and maintaining complex systems, working with teams spread out across the globe, for building both custom solutions or expanding product capabilities. I am open to learning new tools and languages, thrive while working in small teams with clear goals, and enjoy exploring and debugging all kinds of systems. I have worked on teams in several different verticals, mostly in a product-oriented position, from FinTech to GenAI. In my downtime, I enjoy reading, motorcycling, tinkering with computers, and long walks.

Experience

Senior Software Developer, Cognida.ai — Feb 2024 - Jan 2025

I spent the majority of my time at Cognida as a part of the Platform Development team, working on novel patterns and architectures for building and serving LLM-based applications in an effective manner, with the idea that systems or practices developed by this team would be adopted by the rest of the organization on both client projects and core product development. I also spent my first few months working on the Pod team, dealing with proposals and proof-of-concept implementations for customers that required added functionality. All in all, I was responsible for:

- building PoCs on top of Zuno.Assist, Cognida's chatbot product offering, using LLM-based design patterns and custom integrations for B2B customers, e.g. GraphRAG, DOCX processing, etc
- collaboratively architecting Zuno.Dx (still in development at the time of writing), Cognida's Agentic Automation Platform product offering, building features such as follow-up scheduling and email integrations for the platform's Agents
- standardizing developer tooling and practices for dev teams that used Python
- conducting technical interviews on a regular basis

Tools & Technologies: FastAPI, Temporal, SocketIO, gRPC, PostgreSQL, CouchDB, Redis, Docker, Kubernetes, Grafana, Azure, OpenAI, O365

Participant, Recurse Center — Sep 2023 - Dec 2023

I remotely participated in the Fall 2 2023 batch at the Recurse Center, a self-guided programmers' retreat.

Software Engineer, smartSense Solutions — Feb 2021 - Sep 2023

I was able to work on several different verticals during my time at smartSense, mostly as an IC on teams doing highly specific work in systems that were already built up and in use. Outside of the work in these teams, I was also responsible for training new hires in Python for backend web development. All in all, my time at smartSense saw me:

- helping design architecture for Foster, a smartSense product
- collaborating with the data science team to upgrade tools and train new joiners for a more rounded-out team structure
- prepared the first version for coding standards documentation for the Python team as part of the process for ISO certification
- working with a team collectively planning and conducting software development workshops as well as recruitment drives for universities

The projects/clients I was involved with are listed below.

Teamworks — June 2023 - Sep 2023

This was a short-term engagement as a contingent worker in a VC-funded startup based in the USA. I worked as a part of the Hub team focusing on two modules: Forms and News Feed. The components I worked with involved a large Flask+Postgres monolith, serving a React Native front-end over GraphQL for desktop web as well as both Android and iOS. The monolith with written in a CQRS pattern, and made use of Blinker signals and Promises extensively. In my role, I was responsible for:

- working on enhancements in the at-the-time barebones News Feed in Teamworks, adding APIs for features like embedded Tweet cards and rich content in user posts
- debugging and optimizing issues in the Forms component of the application, rewriting ORM code as well as using Promises in the application flow to reduce database load as well as overall query time to remain within acceptable thresholds
- isolating and debugging issues raised by customers in daily use

Tools & Technologies: Flask, GraphQL, PostgreSQL, Celery, Docker, Kubernetes, Datadog, AWS

BASF ReCoBo/CoSSBot — Nov 2021 - June 2023

This was a long-term engagement with BASF's Digitalization department. I was part of a team responsible for enhancing and maintaining CoSSBot (formerly ReCoBo), a Product and Compliance Information Search Platform. I spent the majority of my time working on two subsystems:

• **Document Indexing Automation:**

I designed and built a data connector pipeline system for ingesting data from several PIM systems into a document search engine, taking a process that was a partly manual exercise at the time from a week down to a few hours. This involved:

- using Flask and the Hera library for Argo Workflows to compose a DAG for parallel indexing of tens of thousands of regulatory and compliance documents, grouped into categories that were processed with different models with optional stages to run or skip on failure
- adding Cron-like scheduling options to the indexing API, configurable by BU admins from the application console
- setting up monitoring and alerting over email to process admins

• **Care360 Product Finder:**

I worked on a Product Search System for the Care360 department, a consolidation of several business units' product catalogs into a single e-commerce system with advanced filtering and natural language search support built on top of machine learning models and a graph database. This involved:

- designing the architecture for the search system, with a Backend-For-Frontend service consolidating queries to multiple ML model inference APIs, Neo4j instances, and the e-commerce backend
- working with stakeholders in the BUs to make sure our search interface was at par or better with their existing catalog applications, including filtering behaviour and interface elements for better UX
- discussing, setting, and meeting API performance goals to ensure reasonable round-trip-time for the majority of user queries using caching and minimal data retrieval from the databases

Tools & Technologies: Flask, Django, FastAPI, BentoML, MongoDB, MariaDB, Redis, Kafka, Locust, Docker, Kubernetes, Grafana, Argo Workflows, Azure DevOps, Azure Service Bus, Azure

CapitalFloat/Axio — April 2021 - Oct 2021

This was a medium-term engagement as a contingent worker. I was part of the team responsible for creating and maintaining the Decision Engine (DE) service, a runtime for executing credit policy application logic written as a DSL resembling Python code, processing all customer info as the final step before approving credit allocation/disbursal. In my role here, I was responsible for:

- implementing and deploying new credit products, collaborating heavily with the Data team to test and review policy impact using historical customer data
- working closely with the Product team to implement new data features for the Modelling component of DE for use in new credit products
- optimizing policy code execution to maintain RTT SLOs set by other services in the org as well as partner organizations and platforms

Tools & Technologies: Flask, Django, FastAPI, BentoML, PostgreSQL, AWS

Software Developer, RheoDigital — Jul 2018 - Nov 2020

RheoDigital was a bootstrapped startup working on chatbots for commercial customers. I worked on Rheo.bot, the core product, similar to Tawk.to and Intercom in terms of features and uses. It had deep integration with Facebook Messenger for chat and commerce, simple chat integration with Twitter, and user engagement automations for Facebook Pages and Instagram. During my time at Rheo, I was responsible for:

- implementing real-time chat and notification push using webhooks on Facebook and SocketIO for our custom embedded iframe Chat UI and PWA website
- moving the original management console service from a Yii2 implementation to Django
- implementing authentication using OAuth2 with Facebook and Google as well as a simple email+password flow
- implementing an adapter for GCP's DialogFlow APIs for intent, entity and agent management, allowing us to programmatically create, clone and manage Agents
- designing and implementing an event-based flow for handling user-bot and user-admin conversations, setting up a conversation admin page for Admin users to:
 - monitor ongoing conversations
 - gather stats for most frequently occurring flows
 - monitor alerts for common queries that were not covered by the existing flows
 - take over conversations from the chatbot seamlessly and hand them back to the bot, allowing for connection failures as well as multiple sessions per user
- designing and implementing the onboarding flow for customers, including the logic for mapping embedded iframes to organizations, creating and managing MongoDB databases on a per-org basis, and allow theming and testing chat flows as part of the setup process

Tools & Technologies: Flask, Eventlet, Django, MongoDB, Postgres, Heroku, DialogFlow, GCP, SocketIO

Education

Gujarat Technological University, BE in Information Technology, 2014-2018