

# Prachi Saibewar

Orlando, FL | +1 551-298-3766 | prachi.saibewar.cs@gmail.com | linkedin.com/in/prachi-saibewar | github.com/prachics

## Summary

Backend and AI engineer with 3+ years shipping production Python and C#/.NET services. Raised grounded answers from **79% to 92%** with a 300+ question evaluation set at UCF, added evaluation and guardrails to Rokt's generative ad engine, and cut real-time API latency by **29%**.

## Technical Skills

**Languages:** Python (asyncio), C#, TypeScript, SQL, Bash

**AI / LLM:** OpenAI & Anthropic APIs, RAG, embeddings, pgvector, tool calling, AI agents, prompt engineering, LLM evaluation & guardrails, LangChain, LangGraph, Pandas/NumPy, AI coding tools (Claude Code, Cursor)

**Backend & Data:** FastAPI, ASP.NET Core, REST APIs, OAuth2/JWT, Kafka, Redis, PostgreSQL, SQL Server, Oracle PL/SQL

**Cloud & Reliability:** AWS (Lambda, S3, RDS, ECS), Docker, Kubernetes, Terraform, Linux, Git, GitHub Actions, OpenTelemetry, pytest

## Experience

**Software Engineer** | Rokt (mParticle), USA

Sep 2025 – Aug 2026

*Python, FastAPI, C#/.NET, React/TypeScript, Redis, OpenTelemetry, AWS*

- Engineered Python LLM evaluation and guardrail workflows for Rokt's ACE generative ad engine, reducing rejected content generations from **12.4% to 4.8%** through structured outputs and automated quality checks
- Reduced p95 latency of real-time decisioning APIs by **29% (185 → 132 ms)** with Redis caching and parallel asyncio retrieval under production load
- Improved production reliability with OpenTelemetry tracing, Prometheus metrics, and Chronosphere alerting as part of the on-call rotation; detection time fell from **20 to 8 minutes** and resolution time from **60 to 25**
- Integrated identity-resolved and predictive customer attributes via C#/.NET APIs and event-driven pipelines, enforcing GDPR consent and CCPA opt-outs and expanding eligible customer signals by **23%**
- Owned 0→1 development of a React/TypeScript identity-tracing tool for the mParticle CDP, connecting React Query to C#/.NET APIs to visualize profile merges and cut investigation time by about **30%**

**AI Research Engineer** | University of Central Florida

Jun 2025 – Sep 2025

*Python, FastAPI, PostgreSQL/pgvector, LLM APIs, Next.js/TypeScript*

- Created a 300+ question evaluation set to tune retrieval, chunking, and prompts, raising grounded answers from **79% to 92%**
- Hardened model outputs with schema validation, retries, and human review: **98%+** valid responses, **35% fewer** manual corrections
- Built a Python/FastAPI RAG search service over PostgreSQL/pgvector for research papers and code, so lookups that took about **20 minutes** now take **under 5**
- Led end-to-end modernization of Chartalist.org, a NeurIPS 2022 dataset platform, from legacy PHP to Next.js/TypeScript and Node.js, delivering a searchable, responsive research interface with API integration

**Software Engineer** | Infosys, Remote

Jun 2021 – Jul 2023

*Python, C#/.NET, Oracle PL/SQL, SQL Server, REST APIs, Kafka, Kubernetes*

- Optimized Python and SQL pipelines for banking data transformation and validation, handling **20,000+** records per batch and cutting processing time from **90 to 35 minutes**
- Developed Python REST API integrations across 3 internal systems with JSON validation, retry handling, and structured logging, saving **5 hours per week** of manual cross-system data checks
- Built fault-tolerant Kafka integrations across 5 enterprise systems using idempotent consumers, retries, and dead-letter queues, reducing weekly reconciliation incidents from **25 to 6**
- Tuned Oracle PL/SQL stored procedures and database indexes, improving API response time by approximately **25%**

## Projects

**IncidentIQ: Incident-Support Agent** | Python, FastAPI, Pydantic, LangGraph, PostgreSQL/pgvector, Docker

- Engineered a stateful LangGraph incident-support agent with hybrid RAG, persistent investigation state, and approval-gated actions; achieved **0.80 Recall@5** on 12 labeled scenarios with **240** automated tests

**Geo-Enabled Points-of-Interest API** | Python, Django REST Framework, PostgreSQL/PostGIS, Docker

- Developed a geospatial REST API and map UI with PostGIS spatial indexes, benchmarked at **245 requests/sec** and **35 ms** median latency for 5 km queries

## Education & Certifications

**M.S. in Computer Science** | University of Central Florida | GPA: 3.9

Aug 2023 – May 2025

**B.E. in Computer Science** | Savitribai Phule Pune University

Aug 2016 – May 2020

**Certifications:** AWS Certified Solutions Architect – Associate | AWS Certified Cloud Practitioner