

SAKSHAM RAI

(408) 513-5080 | raisaksham2001@gmail.com | Portfolio - [🔗](#) | LinkedIn - [🔗](#) | GitHub - [🔗](#)

Experience

Software Engineer | **CreateBase** | **Python, Typescript, Supabase, AWS** *May 2025 – Present*

- Spearheading architecture from a single-user system to organizational management, allowing permission-based access to ~ **10,000+ songs, 500+ albums and \$100,000+ in royalty payouts**.
- Engineered the Story-to-Splits pipeline using intelligent routing that delegates complex narrative parsing to **Claude 3 (AWS Bedrock)** and **conversational tasks to GPT-4**, achieving a **145% increase in metadata extraction accuracy** and a **60% reduction in latency via context-aware model selection**.
- Engineered a 3-tier artist-identity resolution system (Selenium scraper, adapter layer, ranking service) with weighted algorithms (RapidFuzz/Metaphone) **achieving 95% accuracy across fragmented PRO databases**, eliminating **manual identity consolidation by 100% (an-industry-wide issue)**.
- Engineered an ETL data ingestion pipeline using PySpark, AWS Glue to clean and dedupe BWARM MLC parquet datasets, **hydrate song objects from 5+ tables (50M + records)** from S3 and finally upsert into MongoDB to enable ISRC-based lookup and unclaimed share calculation.
- Built an MLC data transformation algorithm that converts nested publisher hierarchies (3+ levels) into relational song-participant entries with **accurate split share across 10K+ Spotify tracks**.
- Built a rate-limited, batch-processing cron-architecture Supabase Edge Fn that uses the MLC Public Search API to fetch and populate an artist's Spotify song catalog with royalty data, for over 200+ albums.

Founding Engineer | **Gabriel AI** (#9 Product Hunt) | **Django, AWS, Node.js, Docker** *April 2025 – Dec 2025*

- Built and deployed a scalable delivery pipeline (AWS Lambda, Batch & EC2) **to queue and deliver 5,000 + personalized ringless voicemail campaigns in parallel** via DropCowboy, Telnyx, Elevenlabs API, S3 presigned URLs, with **99.99% reliability**.
- Decoupled the core campaign engine into a **public-facing SaaS API**, allowing B2B partnerships with external CRMs (Clay as of now) to leverage the existing **high-throughput delivery with 0% increase in infrastructure overhead**.

AI Software Engineer Intern | **AfyaChat** | **GPT-4, FAISS, Open AI Embeddings** *March – May 2024*

- Built an AI-powered eConsultation system, reducing specialist response times by 50% using GPT-4 & Flask.
- Built a RAG pipeline on **7+ years of historical medical data** for real-time AI-based specialist recommendations.

Software Engineer (x2 intern, x1 contract) | **County of SD** | **Java, MySQL, Node.js, PostgreSQL**
Sept 2022 – April 2025

- Engineered a real-time drug retrieval portal, **parsing & ingesting 50,000 + county records into PostgreSQL**
- Engineered a real-time opioid distribution portal for 13,000+ citizens via 21-mapped vending machines

Skills

- **Languages:** Python, JavaScript/Typescript, C++, Java, PHP, CSS/HTML
 - **Frameworks/Databases:** React, Next.js, Vue.js, Flask, MySQL, PostgreSQL, MongoDB, Firebase, Supabase
 - **Tools:** AWS ECS, EC2, S3, Amplify, Lambda, SQS, Docker, Render, Vercel, Postman, Git/Github
 - **LLM/AI-Tooling:** GPT-4, ElevenLabs, OpenAI Whisper & TTS, LangChain, LlamaIndex, FAISS Vector DB
-

Education

University of California San Diego, La Jolla, CA *October 2020 – June 2024*

- **Degree:** BS. Math-CS, Double Minor – Cognitive Science (ML Focus) and Entrepreneurship
- **Academic Achievements:** Departmental Provost Honors (x6), Sophomores Honors Program (x1), BlackStone Launchpad Fellowship
- **Relevant Coursework (GPA 3.7) :** Web-Client Technologies, Software Engineering, Algorithms and Data Structures, Object-Oriented Design, Systems Programming, Unsupervised Machine Learning, AI Algorithms