SHYAM PADIA

+1 (857)-294-9264 — shyam.padia930@gmail.com — Boston, MA
crypticsoul.dev — github.com/crizzy9 — linkedin.com/in/shyam-padia

Summary

AI technical lead building agent systems, retrieval and recommendation platforms, and large-scale data products. Combines hands-on architecture with engineering leadership to improve reliability, reduce costs, and turn model capabilities into practical workflows.

Education

Northeastern University, Khoury CCIS — *Boston, MA* — **Sept 2017 - Dec 2019**
Master of Science in Computer Science — GPA: 3.2/4

University of Mumbai, Fr. CRCE — *Mumbai, India* — **Aug 2012 - June 2016**
Bachelor of Engineering in Computer Engineering — GPA: 7.4/10

Work Experience

MosaicApp — AI Technical Lead — *San Francisco, CA* — **Aug 2026 - Present**

- Rearchitected Chello's AI execution layer, improving performance by 25% while reducing costs by 60%.
- Expanded Chello's Model Context Protocol (MCP) capabilities across connected systems, delivering 27 tools for agents to retrieve and update planning data through a shared interface.
- Redesigned agent caching to reuse prior work and reduce repeated processing within the LangChain harness.
- Extended the LangChain harness with subagents and task management to coordinate multi-step planning workflows.
- Delivered Slack and Microsoft Teams integrations, bringing Chello's planning capabilities into team collaboration workflows.
- Introduced query-complexity-based model routing to reduce token usage and inference cost.
- Enabled tiered billing for Chello by integrating Stripe and MonetizeNow.

Intuit, Personalization — Senior Software Engineer — *Mountain View, CA* — **Jan 2025 - Aug 2026**

- Built a LangChain agent orchestrator for RTB, on-call, and DevOps workflows.
- Architected unified user profiles with multi-source ingestion and real-time user embeddings on GPU-optimized EMR.
- Led a cross-regional team building a unified recommendation platform with Vespa, gRPC, and OpenSearch, replacing legacy systems.
- Developed an AI marketing agent with a six-person team using SymphonyAI, LangChain, and MCP, cutting campaign go-live time by 90%.
- Led five engineers building a bulk content system with dynamic templates and audience segmentation, increasing capacity by 1,400%.
- Built an LLM-powered content pipeline that automated marketing workflows and improved time-to-market by 75%.
- Created an MCP server for AI-powered marketing campaign creation, real-time content generation, and automated deployment.
- Defined and presented the personalization AI strategy across content generation, recommendations, predictive analytics, and workflow automation.

Intuit, Personalization — Software Engineer 2 — *Mountain View, CA* — **Jan 2021 - Jan 2025**

- Engineered a self-service analytics platform with PySpark, AWS EMR, and React, reducing incidents and impression loss by 40% and improving marketer productivity by 50%.
- Developed a RAG assistant with Hugging Face and LangChain for marketing content generation, reducing authoring time.
- Designed scalable PySpark data entities for real-time multi-source ingestion, eliminating approval bottlenecks.
- Implemented AWS Personalize on clickstream data with 20% lower API latency.
- Built a Kafka event pipeline for targeted offers that generated $10M in incremental revenue through dynamic audience targeting.
- Trained and deployed XGBoost and neural-network models for predictive business-health scoring, reducing investment risk by 23%.

Intuit, QuickBooks — Software Engineer 1 — *Mountain View, CA* — **Jan 2020 - Jan 2021**

- Built a predictive revenue-forecasting model using customer behavior to estimate the impact of dynamic pricing strategies.
- Engineered a segmentation pipeline with AWS Glue and PyArrow to identify eligible cohorts for targeted pricing campaigns.
- Integrated price simulation and campaign execution components, delivering $20M in cost savings.
- Developed a pricing management platform with React, Spring Boot, and AWS, increasing operational efficiency by 50%.
- Implemented a parallelized synthetic-testing framework with Cypress, preventing $10M in potential revenue loss.

Work Experience (continued)

Intuit, TurboTax Software Engineer Intern *San Diego, CA* **May 2019 - Aug 2019**

- Developed OCR document processing for W2 uploads in TurboTax onboarding, reducing tax preparation time by 20%.
- Integrated an NLP intent-classification decision engine, replacing a third-party API and reducing latency by 30%.

Intuit, TurboTax Software Engineer Co-op *San Diego, CA* **May 2018 - Dec 2018**

- Engineered a rule-based case-deduplication service for the Intuit Help Portal, improving response time by 45% and reducing duplicates by 7%.
- Developed Smart Fields with React and PHP to automate content publishing, saving 30 minutes in the publishing workflow and reducing support cases by 10%.
- Architected a serverless URL-redirect service with AWS Lambda and DynamoDB, improving routing efficiency by 35%.

OutreachCircle Software Developer *Bangalore, India* **Nov 2016 - May 2017**

- Integrated Elasticsearch contact matching for voter registration and social profiles, improving outreach by 10%.
- Developed a campaign outreach platform with Meteor.js and Angular, improving canvassing efficiency and voter turnout by 15%.

Technical Skills

Languages:	Python, Java, JavaScript, TypeScript, Go, Nix, Lua, SQL, Scala, C/C++
ML/AI:	PyTorch, TensorFlow, Hugging Face, LangChain, MCP, RAG, OpenCV, Vespa, NLTK, ScikitLearn, Keras, Gensim
Frameworks:	Spring Boot, React.js, HTMX, Astro.js, Django, FastAPI, Node.js, REST, gRPC, Expo, React Native
Big Data:	PySpark, Kafka, Hadoop, Pandas, PyArrow, Hive, Beam, Flink, EventBus, RabbitMQ
Cloud:	AWS (SageMaker, Lambda, Glue, Personalize, EMR, OpenSearch, Athena, Redshift), Docker, Kubernetes, NixOS, Nix Flakes, Home Manager
Databases:	PostgreSQL, DynamoDB, MySQL, SQLite, MongoDB, Redis, ElasticSearch, Vector Databases
Courses:	NLP, Reinforcement Learning, Robotics, Machine Learning, Artificial Intelligence

Projects

Lumino **Declarative infrastructure**

- Developed a YAML-driven NixOS platform using Flakes, Home Manager and sops-nix for reproducible desktop and service configuration.

Naavik **AI-powered career workflows**

- Built explainable job matching, tailored documents and reviewed application workflows with FastAPI, HTMX, PostgreSQL and Typst; V1 qualification ongoing.

Plano **Flight tracking — In development**

- Built Go/SQLite flight tracking with an Expo/React Native app and synthetic demo; live-source and device qualification ongoing.

Omnidirectional Autonomous Robot **Robotics & computer vision**

- Engineered a Raspberry Pi robot with a 6-DOF arm, inverse kinematics, OpenCV perception, voice commands and obstacle avoidance.

Reinforcement Learning for Robotic Manipulation **Transfer-learning research**

- Evaluated robot-arm transfer learning with PyTorch and OpenAI Gym, achieving 35% faster convergence and 28% improved task performance.

Multi-Track Music Generation **Transformer architectures**

- Developed four-track transformer models with global multi-headed attention, improving musical coherence by 45%.

Paragraph Embeddings with Variational Autoencoders **Natural language processing**

- Built TensorFlow VAE and LSTM sequence-to-sequence paragraph models on Gutenberg, improving semantic-similarity performance by 22%.

Fan Forge — Comic-Con Museum **Community software & team leadership**

- Led five engineers on an Express.js/AWS Lambda platform for community-curated Comic-Con Museum exhibit and artifact ideas.

Certifications

Machine Learning Professional Certificate - General Assembly **June 2022**

Apache PySpark Certification - Databricks **Dec 2021**

AWS Certified Solutions Architect - Associate **Oct 2019**