

JYOTIRMOY SINGH

✉ singhjyotirmoy17@gmail.com  [Jyotirmoy](#)  [Jyotirmoy17](#)  [Personal Website](#)

EDUCATION

Birla Institute of Technology and Science, Pilani, Hyderabad 09/2021 - 07/2026

BE (Hons.) Computer Science; MSc (Hons.) Economics | **GPA: 8.6/10**

Thesis (Advisor: Prof. Anand Rao, Carnegie Mellon University)

Fall 2025 - Spring 2026

Multi-Agent AI Orchestration: The System-3/0 Blueprint

- Designed the System-3/0 Blueprint: a domain-agnostic multi-agent orchestration framework with a System-3 meta-controller, System-0 constrained-decoding safety envelope, and a Dynamic Agent Registry formalised as a 6-tuple AgentSpec; proved 6 theorems, a lemma, and 2 design arguments covering safety, routing, and architectural necessity.
- Validated across 300 paired runs (150 per condition): reduced Cascading Failure Rate from 38.0% to 0.0% ($p < 0.001$, Cohen's $h = -1.328$) and cut mean token expenditure by $3.6\times$ ($5,724 \rightarrow 1,600$ tokens, $p < 0.001$); runs entirely on consumer hardware (24 GB RAM, no GPU) via Ollama.

EXPERIENCE

Joalent 

Aug 2025 - Present

Founder & CEO

Wyoming, USA

- Built the full AI stack from scratch: an active-learning OCR pipeline hitting 98% field-level accuracy and a conversational invoice agent at 96% accuracy for automated financial data ingestion.
- Integrated Stripe Connect, QuickBooks, and TaxBandits APIs with async webhooks to automate dunning, W-9 collection, and 1099 generation for US-based SMBs.
- Wrote the confidence-scoring system (field-level penalties + HITL fallback) that keeps error rates under 1% on financial documents in production.

Google, Google Cloud (Backup for GKE)

May 2025 - Aug 2025

Software Engineering Intern

Hyderabad, India

- Wrote the integration test suite for 7 asynchronous schedulers in GKE Backup using Go, gRPC, and GCP internal tooling, taking coverage from 0% to 42%.
- Caught 27+ critical bugs before production rollout through reusable test workflows designed for the scheduler codebase.
- Shipped 3 weeks ahead of schedule and built a generic polling mechanism and custom API clients in Go that replaced flaky async test patterns.

BITS Pilani x Impactsure 

Aug 2024 - Dec 2024

Research Lead (\$150K+ Funded)

On-Site

- Led a team of 4 engineers on a \$150K+ funded project building compact Small Language Models (SLMs) for low-latency deployment at a financial services firm.
- Designed Rail-Only architectures that cut training costs by 20% and infrastructure requirements by 30% compared to standard transformer fine-tuning.

PUBLICATIONS

Agentic Anomaly Detection (ORCA)

Manuscript under Preparation

- Built an edge-deployable anomaly detector combining MinGRU and ConSmax attention with a Gemini-powered Meta-Reasoner for real-time clinical synthesis on MIMIC-IV data.

Explainable AI for Predictive Healthcare

Target: Elsevier (Under Preparation)

- Compressed ML ensembles into single interpretable Decision Trees via a Glass-Box distillation pipeline, retaining 96% accuracy on MIMIC-IV clinical data.

Honors & Recognitions

- Invited to **OpenAI Emerging Talent Community** (2025 Cohort)
- Invited to **Y Combinator** Startup School
- Teaching Assistant, Machine Learning, BITS Pilani (Spring 2025)
- **National Level Gold Medalist** in Mathematics

TECHNICAL SKILLS

Languages: Python, Go, C++, Java, SQL

AI/ML: PyTorch, TensorFlow, LangChain, LangGraph, Ollama, SHAP, RAG, Multi-Agent Systems

DevOps & Systems: Docker, Kubernetes, gRPC, Protocol Buffers, Microservices, Distributed Systems

Cloud & APIs: GCP, Stripe Connect, QuickBooks API, Webhooks, REST/GraphQL