

INFANT SAM ANTON

+918637633059 • jsam2k4.pro@gmail.com • linkedin.com/in/jsam2k4/ • github.com/SamAnton-J

AI ENGINEER

A results-driven AI Engineer specializing in Generative AI and advanced prompt engineering. Expert in translating complex business requirements into high-impact AI solutions. Proven ability to architect sophisticated systems, from a GPT-4 chatbot that democratized data analytics for non-technical users, to rapidly mastering and deploying a full MERN stack and Web3 application. Blends strong AI research expertise with full-cycle product development and deployment.

EDUCATION

Bachelor of Computer Science and Engineering

St. Joseph's College of Engineering

Graduating May 2026

8.0 CGPA

PROFESSIONAL EXPERIENCE

Saipem India Projects Pvt. Ltd, Chennai: LLM Engineer (Intern)

Jun 2025 – Aug 2025

- **Architected a sophisticated prompt engineering strategy** that empowered non-technical users to generate complex DAX queries through natural language, democratizing data access.
- Developed a novel Power BI JSON schema to **constrain the LLM's output**, enforcing syntactical correctness and dramatically reducing query generation errors and hallucinations.
- Engineered a seamless integration with Power BI via XMLA, delivering a real-time, **self-service analytics tool that reduced dependency on dedicated data analysts**.

Posphene AI, Chennai: Deep Learning Researcher (Intern)

Sept 2024 – Mar 2025

- Fine-tuned a hybrid ResNet50-LSTM model, improving image sequence classification accuracy by 12%.
- Leveraged AWS Inferentia for inference acceleration, reducing latency by 40% and enabling real-time processing.
- Deployed the model with torch-neuronx, achieving 99.9% uptime and scalable production readiness.

ACADEMIC PROJECTS

DREACON: Detection & REcovery for Abused Content & Narratives

(AI Research & Development)

Engineered an AI framework to counter deepfake threats through identity defense, legal escalation, and victim recovery.

- Built a multi-stage deepfake detection pipeline using **ArcFace** embeddings and CNN-based facial recognition to match user-uploaded photos with social media content.
- Deployed **EfficientNet-B7** and **XceptionNet** on **FaceForensics++/DeeperForensics1.0** datasets for binary deepfake classification with real-time alerting.
- Integrated an LLM-powered chatbot using **advanced prompt engineering**, training it on global cyber laws (IT Act, GDPR) and wellness policies for personalized legal support.
- Used **ChromaDB** for fast embedding retrieval and face match indexing; frontend built with **React.js**, backend using **Django**, and deployed on **AWS Inferentia**.
- Enabled automated legal notice generation and real-time victim notifications upon detecting deepfake content across platforms like a cloned Instagram environment.

LinkedX: AI & Web3 Professional Networking Platform

(AI · MERN · Web3)

Architected and built a full-stack, dual-role marketplace integrating Web3 payments and AI-driven job matching to create a high-signal hiring ecosystem.

- Built a full **MERN stack** application (**MongoDB, Express.js, React, Node.js**) with role-based **JWT authentication** for 'Producer' and 'Freelancer' user types.
- Engineered an end-to-end **Web3 payment flow** using **Ethers.js** and **MetaMask** for on-chain job posting fees on the Sepolia testnet, with backend verification via the Etherscan API.
- Implemented two practical AI features: an NLP-based **skill extractor** to accelerate profile creation and a **job-to-freelancer matching algorithm** to provide a relevance 'Match Score'.
- Developed a comprehensive **Go-To-Market (GTM) strategy** outlining a 90-day user acquisition plan, monetization models, and KPIs, demonstrating full product ownership.
- Integrated real-time chat using **Socket.IO** and deployed the entire application on **Netlify** (frontend) and **Railway** (backend).

TECHNICAL SKILLS

AI/ML: Prompt Engineering, Machine Learning, Deep Learning, Generative AI, LLMs, NLP, Computer Vision, LangChain Agents, Neural Style Transfer (NST), Feature Engineering.

Languages: Python (ML/DL), JavaScript (Node.js/React), Java (DSA, OOP), C++, SQL, C.

Software Development: Development, Testing, Deployment, Debugging, Troubleshooting, Git/Version Control.

Web3 & Blockchain: On-Chain Payments, Wallet Integration (MetaMask), Ethers.js, Ethereum (Sepolia).

Backend: API Design (REST), Node.js, Express.js, Real-Time Communication (Socket.IO), Django, Flask.

Frontend: React.js, HTML, CSS, JavaScript (ES6+), TailwindCSS.

Databases: NoSQL (MongoDB), Relational (SQL), Vector DBs (ChromaDB, FAISS).

Data Visualization: Matplotlib, Seaborn, Power BI.

TOOLS & TECHNOLOGIES

Deployment & DevOps: AWS (EC2, Inferentia, Trainium), Docker, Netlify, Railway, Render, Heroku;

Development Tools: Git, GitHub, VS Code, PyCharm, Jupyter, IntelliJ IDEA, Postman;

AI Data Platforms: Cohere API, OpenAI API, Weights & Biases, OpenCV, Mongoose, LangChain, NeuronX, TensorRT;

Productivity AI: GitHub Copilot, ChatGPT, Perplexity, Google AI Studio, Blockbox, **Hoop.AI**, Lovable, Cursor, Replit AI, Tabnine, Codeium, Blackbox AI, CodeWhisperer.

PROFESSIONAL SKILLS

Strategic Thinking, Product Ownership, Problem Solving, Team Collaboration, Communication, Adaptability, Leadership, SWOT Analysis, Agile/Scrum.

CERTIFICATIONS

Advanced Learning Algorithms (DeepLearning.AI and Stanford University)

Covered TensorFlow, Model Development, Artificial Neural Networks, XGBoost, and Tree Ensembles.

Unsupervised Learning, Recommenders, Reinforcement Learning (DeepLearning.AI and Stanford University)

Covered Anomaly Detection, Collaborative Filtering, and Reinforcement Learning.

Machine Learning AI (PreplInsta)

Focused on ML Fundamentals, Advanced Algorithms, and Feature Engineering.

NLP and Deep Learning (PreplInsta)

Explored Sentiment Analysis, Named Entity Recognition, and Machine Translation.

The Complete Neural Networks Bootcamp: Theory, Applications (Udemy)

Covered neural networks, CNNs, RNNs, and practical deep learning with TensorFlow and PyTorch.

ACHEIVEMENTS

• **Built a custom VGG19-based feature extractor in PyTorch for Neural Style Transfer.**

• **Skillrack** Solved over 700+ problems.

• **Leetcode** Solved 300+ problems focused on DSA OOP.

• **HackerRank Badges:** C(4 star), Java(4 star), Python(4 star) all Silver.