

Shubham Goyal

Senior Data Scientist · GenAI & Agentic AI · Financial Crime Analytics

shubhamgoyal769@gmail.com
linkedin.com/in/shubham-goyal-0946b7127
github.com/shubham769 · shubhamgoyal.me
Abu Dhabi, United Arab Emirates

Senior Data Scientist with **7+ years** building production AI systems across banking, government, regulatory technology and enterprise platforms. Specialize in **GenAI and agentic AI, retrieval and document intelligence**, and **financial crime analytics** — from problem framing and architecture through model development, GPU serving, deployment and client delivery. Experienced in regulated and air-gapped environments where accuracy, explainability and auditability are requirements rather than features.

MEASURED IMPACT

£230K+ Fraud prevented per month	50% Fewer false AML alerts	90%+ Extraction accuracy	75% Less manual review	95% Entity match precision
--	--------------------------------------	------------------------------------	----------------------------------	--------------------------------------

EXPERIENCE

Senior Data Scientist — Presight.ai (G42) Sep 2023 – Present · Abu Dhabi, UAE
Data Scientist Sep 2023 – Mar 2024; promoted to Senior Data Scientist in April 2024.

- Built a **Natural Language → SDMX agent** for the UAE National Statistics Data Platform (FCSC program), letting analysts query structured national statistics conversationally; deployed to the demonstration environment.
- Worked on UAE SDMX data ingestion, conversational history, system integration API testing, Data Quality trigger APIs and evaluation of MCP-based service integrations; integrating an end-to-end CPI calculation workflow.
- Authored the wider **AI framework architecture** — LangGraph orchestration, MCP tool execution, GraphDB/RDF/SPARQL knowledge layer, FastAPI services, Redis and PostgreSQL state, human approval and provenance — and reviewed it with the solution architecture team.
- Delivered document intelligence for the ADAA audit and compliance program: OCR pipelines, government gazette analytics, contract and financial document analysis — **90%+ extraction accuracy** and **~75% less manual review**.
- Designed an **entity risk scoring framework** (quantitative, binary and count indicators) and conflict-of-interest detection using name matching and entity resolution at **~95% precision**.
- Built production RAG with citations and highlighted source passages — Qwen2.5 on **vLLM**, multilingual-e5 embeddings, FAISS, FastAPI, Streamlit, Docker — packaged with models baked into GPU containers for **air-gapped deployment**.
- Supervised team members on the ADAA engagement; regular client presentations, demonstrations and technical workshops.

Data Scientist — Barclays Apr 2022 – Aug 2023
Fraud & Financial Crime (Quantitative Analytics)

- Developed a bank-wide **mule detection** model identifying fraudulent and mule account activity; contributed to preventing **£230,000+ per month** in fraud losses.
- Improved transaction monitoring through feature engineering and threshold optimization, cutting **false-positive AML alerts by ~50%** while maintaining risk coverage.
- Delivered inside a global bank's model governance and control environment, working across compliance, audit, technology and financial crime teams; built ML pipelines and automated compliance reporting.

Applied Data Scientist — Tookitaki AI

Oct 2020 - Apr 2022 · Bengaluru, India

Cloud-based AML and transaction monitoring products for banks and payment providers

- Built ML-based **transaction monitoring** using a semi-supervised approach combining multi-dimensional unsupervised techniques, network analysis and supervised learning to detect complex structuring; deployed with multinational banks and payment providers across APAC, North America and Europe.
- Developed **name screening** models — multilayered NLP, fuzzy matching and entity resolution handling typos, spelling variants, titles and affixes — for sanctions, PEP, adverse entity, customer and transaction screening.
- Built customer risk scoring, crypto risk analytics and graph/network analytics; improved monitoring and investigation efficiency by **~60%**.
- Integrated models with client core banking systems and worked directly with product, engineering and compliance stakeholders.

Software Consultant, Data Science — Knoldus Inc.

Jan 2019 - Oct 2020 · Noida, India

- Helped lead the **Machinex AI/ML initiative** with a team of around four data scientists, delivering ML products across retail and resource analytics; **~95% client acceptance** on delivered solutions.
- Built real-time and distributed services in **Scala, Kafka, Akka, Lagom and Cassandra**, containerized with Docker and Kubernetes; predictive modeling with logistic regression, decision trees, random forest, SVM, KNN and XGBoost.

TECHNICAL SKILLS

GENAI & AGENTS	LLMs, RAG, AI agents, LangGraph, MCP, tool calling, Semantic Kernel, Haystack, prompt engineering, Qwen, Azure OpenAI, Hugging Face, Ollama
RETRIEVAL & KNOWLEDGE	FAISS, embeddings, multilingual-e5, semantic search, chunking & reranking, GraphDB, RDF, SPARQL, knowledge graphs, provenance
MACHINE LEARNING	Classification, regression, clustering, anomaly detection, semi-supervised learning, network/graph analytics, NLP, entity resolution, transformers, scikit-learn, PyTorch
FINANCIAL CRIME	AML, transaction monitoring, mule & fraud detection, name and sanctions screening, customer risk scoring, STR analytics, crypto risk
SERVING & MLOPS	vLLM, FastAPI, Docker, Kubernetes, CUDA/GPU inference, quantization, offline model packaging, air-gapped deployment; NVIDIA Triton and Ray Serve (evaluated)
DATA ENGINEERING	Python, SQL, PySpark, Spark, Hive, Databricks, Kafka; Scala, Akka, Lagom, Cassandra

EDUCATION & PUBLICATIONS

B.Tech, Computer Science & Engineering — Major in Data Science

Lovely Professional University · 2015 – 2019

Certifications

DeepLearning.AI TensorFlow Developer (Coursera) · Machine Learning, Stanford/Coursera · Python for Data Science, UC San Diego/edX · How Google Does ML · AWS ML · Spark for Scala Analytics

Novel XGBoost Tuned ML Model for Software Bug Prediction

IEEE · ICIEM 2020 — ieeexplore.ieee.org/document/9160152

Prediction on Diabetes Patients' Hospital Readmission Rates

ACM · ICAICR 2019 — dl.acm.org/doi/10.1145/3339311.3339349

Recognition

Rising Star Award (2020) · KNIME certified L1–L4 (2020) · Silver medal, top 2% — LANL Earthquake Prediction, Kaggle (2019) · upGrad ML Xpert, PG Diploma mentor