SRIJEET CHATTERJEE

AI Engineering Lead | Senior Machine Learning Engineer

London, UK | +44-7767978269 | champ.srijeet@gmail.com

srijeetchatterjee.com | LinkedIn | GitHub | Medium

PROFESSIONAL SUMMARY

Launched the first neural-network enforcement model on a platform with 2.5B+ monthly active users and shipped the first generative AI product into production at a Tier-1 investment bank.

AI Engineering Lead with 11+ years of experience driving enterprise-wide AI transformation. End-to-end ownership across the ML lifecycle—data strategy, model architecture, feature engineering, training, ML system design, A/B experimentation, deployment, and monitoring. Currently building production agentic AI platforms for media and advertising. Deep expertise in large-scale ML enforcement systems, LLM-powered applications, neural machine translation, and responsible AI. Strong theoretical foundations in deep learning, NLP, and classical ML, combined with hands-on engineering to deliver measurable business outcomes at scale.

CORE COMPETENCIES

ML & Deep Learning Fundamentals: Attention Mechanisms, Transformer Architectures, CNNs, LSTMs, Encoder-Decoder Networks, SVD, Classical ML (SVM, Decision Trees, Logistic Regression, KNN), Reinforcement Learning, Information Retrieval

NLP & Language Understanding: Seq-to-Seq Modelling, POS Tagging, Text Mining, NER, Neural Machine Translation, RAG

Algorithms & Computer Science: Data Structures & Algorithms (DSA), Complexity Analysis, Optimisation, Computer Vision, GANs

AI & ML Frameworks: PyTorch, Keras, Scikit-learn, OpenCV, Hugging Face Transformers, LoRA/QLoRA Fine-tuning

Agentic & Generative AI: LangChain, LangGraph, CrewAI, AWS Strands, AWS Bedrock Agent, AWS AgentCore, MCP, OpenAI API

Evaluation, Observability & Monitoring: Langfuse, MLFlow, DVC, AWS CloudWatch, AWS Lambda, Event-driven Hooks & Tracing, Custom Evaluation Pipelines

ML System Design & MLOps: End-to-end ML Pipelines, Feature Engineering, Model Serving, Inference Optimisation, A/B Experimentation Frameworks, Docker, Kubernetes, Jenkins, Databricks, CI/CD

Cloud & Infrastructure: AWS (Bedrock, Lambda, CloudWatch, AgentCore), Azure Cognitive Services, Git

Languages: Python, C++, Java, SQL, MATLAB, NumPy, Pandas, FastAPI, Flask, Streamlit

Data & Vector Stores: Qdrant, Chroma, FAISS, Oracle, Cloudant

Leadership: Cross-functional team leadership (6+ engineers), technical scoping, stakeholder alignment, data advocacy

PROFESSIONAL EXPERIENCE

Publicis Media   London, UK
AI Engineering Lead   *January 2026 – Present*

- Leading a team of 6 AI engineers to architect and deliver an enterprise-scale **Agentic AI platform** for programmatic media planning and optimisation, serving clients across search and e-commerce channels.
- Designed and deployed 5 specialised AI agents across 5 media channels, with 3 enterprise clients onboarded in the first 6 weeks and a pipeline of 15+ clients in progressive rollout.
- Engineered the complete agentic pipeline using AWS Strands framework (Bedrock) with AWS AgentCore as the managed runtime, enabling autonomous multi-step campaign orchestration and bid optimisation workflows.
- Built the full **AI Evaluation & Observability stack** using Langfuse for real-time agent tracing, AWS CloudWatch for infrastructure monitoring, and Lambda-based event hooks for end-to-end pipeline observability.
- Established **evaluation benchmarks, guardrails, and responsible AI protocols** for agentic behaviour, ensuring hallucination control, output safety, and deterministic performance under production traffic.

Meta (WhatsApp Integrity)   London, UK
Senior Software Engineer, Machine Learning   *April 2025 – December 2025*

- Designed and launched the **first neural-network-based enforcement classifier** on WhatsApp (2.5B+ MAU), targeting scam actors and reducing scammer-engaged conversations by 18%.

- Independently identified and pursued the initiative to curb **Victim-Initiated Scams**—contributing to >50% of all scams on the platform—by researching, prototyping, and validating enforcement models across two new operational contexts (disconnect events and wa.me entry points).
- Owned the full initiative lifecycle end-to-end: conducted impact analysis, trained classifiers operating on 300M+ disconnect events and 40M+ wa.me events daily, designed and executed 3-week A/B experiments, led experiment result evaluation with the team, and launched to production—delivering an 11% reduction in the SEC topline metric, eliminating ~3M Scam Engaged Conversations on a 7-day rolling average.
- Collaborated with the SMB (Small & Medium Business) team to extend enforcement models to business accounts, broadening the integrity coverage surface across WhatsApp.

UBS Investment Bank
London, UK

Generative AI Engineer, Director
December 2022 – March 2025

AI-Generated Hooks for Equity Research Reports

- Architected and delivered **UBS's first generative AI solution in production** within the Investment Banking group—a real-time hook generation engine for equity research reports serving Evidence Lab analysts across ~55,000 reports annually.
- Designed the end-to-end ML system: data pipeline, model selection, prompt engineering, inference serving, and A/B experimentation framework that demonstrated 60% higher client readership—a primary business KPI directly tied to Evidence Lab revenue.
- Accelerated enterprise-wide generative AI adoption while saving an estimated 500+ man-hours per year in analyst workflow efficiency.

Japanese–English Neural Machine Translation

- Led end-to-end data preparation and bilingual sentence alignment pipeline for a custom JP–EN translation model, including API development for Azure-based model training, serving, and deployment.
- Achieved a BLEU score of 0.54—a 46% improvement over the incumbent third-party (Sentience) NMT model—through systematic benchmarking against Azure Translator and custom architectures.

Deloitte
London, UK

Senior Machine Learning Engineer
June 2022 – December 2022

- Developed Talk-2-Tables and Talk-2-Docs: LLM-powered solutions enabling natural language interaction with domain-specific private documents and SQL databases using RAG with Chroma vector store.
- Designed evaluation criteria, security protocols, and explainability frameworks for enterprise LLM deployments.

IBM India Pvt Limited
Bangalore, India

Data Scientist III, Band 7A
July 2019 – March 2022

Cognitive Metrics Tool Development for AT&T

- Developed production web APIs for the Defect Removal Efficiency (SLA) metric, deployed within a CI/CD DevOps pipeline using Flask.

Acharya / AI-SME for AT&T (Seq-2-Seq Problem Modelling)

- Engineered an automated problem-solution extractor and multi-level ranking algorithm leveraging POS tagging, SVD-based Quality of Solution scoring, and domain-specific pattern recognition.
- Designed and deployed an encoder-decoder architecture for automated solution generation using ReLU-boosted LSTM, trained on historical ticket data and served via Flask API.

Tata Consultancy Services
Kolkata, India

Data Scientist, Advanced Analytics
March 2013 – July 2017

- Served as SME for eCRM and COPs modules; developed an offer conversion prediction model achieving 74.68% accuracy for the Cognitive Metrics initiative.

EDUCATION

Indian Institute of Technology (IIT) Delhi
Delhi, India

Master of Technology – Computer Science & Technology
August 2017 – June 2019

Techno India College of Technology
Kolkata, India

Bachelor of Technology – Electronics & Communications Engineering
August 2008 – July 2012

RESEARCH, PATENTS & PUBLICATIONS

U.S. Patent Filed – Business Language Processing using LoQoS Score & ReLU-boosted LSTM, IBM, November 2021. Filed with the U.S. Patent & Trademark Office.

Major Thesis – Brain Tumour Grading via Texture Features & ML, IIT Delhi, 2019. Supervised by Dr. Anup Singh & Dr. Sumeet Agarwal.

Minor Thesis – Image Super-Resolution using GANs, IIT Delhi, 2018. Supervised by Dr. Sumantra Dutta Roy.

Publication – "Services Recommendation Engine using Client's Demographics," IBM GTS Labs, 2018.

THOUGHT LEADERSHIP & COMMUNITY

- Technical Blog Author on Medium – Writing on deep learning, NLP, and applied AI/ML for a growing readership.
- Org-wide Deep Learning Instructor, IBM – Designed and delivered a 3-week intensive programme on deep learning fundamentals to 200+ participants (2020). Received IBM Extra Mile Award.
- Teaching Assistant, IIT Delhi – Conducted tutorials on SVMs, CNNs, GANs, Cache Memory, and Paging for Machine Intelligence & Learning and Computer Architecture courses (2017–2018).

SELECT ACHIEVEMENTS

- First Patent File Reward, IBM – December 2021
- All India Rank 344 in GATE CSE among ~130,000 candidates, 2017
- Best Team Award for outstanding organisational contribution, TCS, 2016
- KUDOS for Best Performer in Initial Learning Programme, TCS, 2013