

Aman Singh Thakur

Machine Learning Engineer - Agentic AI, RAG, and Foundational Research | 5+ Years

Seattle, WA, USA | +1-4134726405 | singh96aman.me | singh96aman@gmail.com | [linkedin.com/in/singh96aman](https://www.linkedin.com/in/singh96aman) | scholar.google.com

PROFESSIONAL EXPERIENCE

Amazon, Seattle, USA

July 2024 - Present

Machine Learning Engineer 2 | Agentic AI, RAG, Foundation Model Evaluation, Model Customization, Reinforcement Learning | [TechCrunch](#)

- *Opinionated RAG (Amazon Seller Assistant)*: Leading integration of opinion-aware retrieval into Amazon.com's Seller Assistant for approximately **10M selling partners**, grounding responses in Seller Central and Reddit while preserving representative sentiment; the underlying O-RAG method reduced distributional error by **18–48%** in offline benchmarks.
- *Foundation Model Evaluation & Multi-Turn RL*: Built SageMaker's model-evaluation service from the ground up, enabling base-vs-custom model comparison with LLM-as-judge, reward, pass@k, and trajectory metrics; extending the platform to serverless multi-turn reinforcement learning for multi-step tool-use and reasoning tasks.
- *Agent-Driven Model Customization*: Led end-to-end delivery of an Agents SDK workflow automating fine-tuning, evaluation, and deployment on SageMaker; demonstrated at re:Invent 2025 and scaled to **50+ customers**. Standardized the agent through A2A and MCP integration, supporting **10+ base models** across SFT, DPO, and RLVR.
- *Fraud Model Revamp (AWS Mechanical Turk)*: Overhauled the rule engine and retrained a PyTorch neural network to address data drift, improving fraud detection accuracy from **82% to 97%** across global marketplaces and defending against **\$25M+** in potential losses.

Goldman Sachs, Bengaluru, India

Oct 2020 - Jul 2022

Applied Scientist 2 | Applied ML, Statistical Modeling, Anomaly Detection, Python, SQL, Data Engineering

- *Credit Risk Modeling*: Built credit scoring models using hypothesis testing and feature selection (PCA, RFE, Spectral Embedding), achieving **95%+ accuracy** across **thousands of loan applicants**.
- *Risk Exposure & Anomaly Detection*: Developed quantitative models to monitor real-time firm exposure (**\$500Bn+**) across **100+ clients** and flag abrupt fluctuations indicative of anomalous trading behavior.
- *Regulatory Data Infrastructure*: Built concurrent Java, Python, and C++ pipelines validating **100+ data-quality contracts** and processing **20Bn+ daily data points** across data-lake, Sybase, DB2, and archival systems.

Morgan Stanley, Mumbai, India

Jul 2019 - Sep 2020

Software Engineer | NLP, Java, Spring, SQL, Data Pipelines

- *NLP Trade-Ingestion Platform*: Built Python-based tagging to extract tickers, buy/sell indicators, keywords, and quantities from CSV, XML, and email data, enabling onboarding of **50+ institutional clients**.

PUBLICATIONS & RESEARCH

Training Leaves Traces: Centered Residual Signatures for Language Model Lineage Verification

Aug 2026

Amazon, USA | Model Lineage, LLM Provenance, Weight-Space Analysis | Under submission - ACL ARR Main Track

- Introduces centered residual signatures for data-free weight-lineage verification, separating descendants from independent and distilled models while remaining invariant to checkpoint laundering.

WARP: Wasserstein-Aligned RAG for Population Opinions

Jun 2026

Amazon, Seattle, USA | RAG, NLP, Information Retrieval | Under submission - EMNLP 2026 Industry Track

- Aligns retrieved evidence with population opinion distributions, reducing distributional error by **45–81%** across **three domains** within **sub-second latency**.

Retrieval-Augmented Generation Must Move Beyond Factual Grounding to Represent Diverse Opinions

Apr 2026

Amazon, Seattle, USA | RAG, NLP, Information Retrieval | [Preprint \(arXiv:2604.12138\)](#)

- Proposed O-RAG to preserve corpus-level opinion distributions, reducing distributional error by 18–48% across seller forums and hotel reviews.

Judging the Judges: Evaluating Alignment and Vulnerabilities in LLMs-as-Judges

Jun 2024

UMass Amherst–Meta Collaboration, Remote, USA | NLP, LLM Evaluation, Generative AI | [ACM GEM² \(250+ citations\)](#)

- Designed benchmarking infrastructure for **1000+ A/B tests** across base and fine-tuned LLMs, measuring judge alignment with Kappa, ROUGE-L, BERTScore, and ranking metrics.

Baler - Deep Autoencoders for Scientific Data Compression

Oct 2023

Google Summer of Code 2023 · CERN-HSF / Lund University | PyTorch, Autoencoders, Spark | [ResearchGate](#)

- Added 2D/3D convolutional autoencoders to Baler, achieving **95%+ reconstruction accuracy** and **200%+ compression**; implemented error-bounded regularization.

Online Physics & ML Modeling

Sep 2023

SLAC National Accelerator Laboratory, Stanford | PyTorch, Go, Docker, Kubernetes, GPU, MLOps | [Poster](#)

- Migrated accelerator-simulation models from Keras to PyTorch and improved VectorDB querying, achieving a **10x speedup** over physics-simulation predictions.

EDUCATION

University of Massachusetts Amherst, USA

Sep 2022 - May 2024

Master of Science in Computer Science

CGPA: 3.9/4.0

Manipal Institute of Technology, India

Jun 2015 - Jun 2019

Bachelor of Technology in Computer Science Engineering

CGPA: 8.4/10

TECHNICAL SKILLS

- | | |
|------------------------------------|---|
| • ML & GenAI | PyTorch, RAG, Information Retrieval, LLM Evaluation, LLM-as-Judge, Reinforcement Learning |
| • Model Customization | SFT, DPO, RLVR, LoRA/PEFT, Agentic Workflows, Multi-Turn Evaluation |
| • Cloud & Agent Systems | AWS, SageMaker AI, Bedrock, Agents SDK, MCP, A2A, LangChain, Docker, Kubernetes |
| • Data & Languages | Python, Java, C/C++, SQL, Go, Spark, Flink, Kafka, Redshift, S3 |