Shivam Bhardwaj

Gurugram, India | bshivam.work@gmail.com | +91-9891741370 | LinkedIn:-bhardwajshivam- | GitHub:bhardwajshivam

Data science professional with 4+ years of experience building predictive, causal, and AI/ML systems from experimentation through production. Skilled in leading teams, designing scalable infrastructure, devloping business metrics and driving data informed decisions.

EXPERIENCE

YRISE — New Haven, CT
Research Associate (Data Scientist) — *Sep 2024 – Present*

- Lead end-to-end data science initiatives for cross-organization projects with MIT, Oxford, BRAC, KPLC, and the World Bank, partnering with research, engineering, and operations teams to define metrics and translate problems into scalable solutions.
- Built a secure RAG and knowledge-graph platform across 50+ research projects, integrating Dropbox, SurveyCTO, project metadata, and Neo4j to enable governed access to structured and unstructured research data.
- Designed and benchmarked a hybrid retrieval pipeline combining BM25/vector search, RRF, ColBERT reranking, and MMR, with conditional HyDE and graph retrieval; improved Precision@5 from 0.66 to 0.80 while evaluating faithfulness and latency.
- Developed an English-to-Cypher generation layer with schema validation and deterministic guardrails against Neo4j node labels, relationships, and properties, preventing invalid queries and unreliable retrieval.
- Productionized the platform on secure enterprise infrastructure managed by Yale IT using Docker, Kubernetes, and CI/CD; implemented MLflow-based monitoring for latency, token cost, retrieval quality, groundedness, and system reliability.
- Developed Python and ArcGIS pipelines combining population density, road-network accessibility, and monthly VIIRS nighttime light data to construct and validate 200+ clusters for randomized education and training interventions.
- Engineered ETL pipelines for 500GB+ of Landsat imagery to identify clusters, estimate socioeconomic characteristics, supporting stratification and baseline targeting.
- Built an XGBoost churn model using longitudinal attendance, engagement, and completion signals; achieved PR-AUC of 0.28 and captured ∼65% of eventual churners within the top 10% highest-risk participants.
- Translated churn probabilities into targeted retention interventions, achieving 6.5× Lift@10% over random targeting while reducing outreach to 10% of participants.
- Analyzed heterogeneous retention effects across 10 participant segments, identifying groups with ∼1.8× higher response to outreach and reducing cost per incremental retained participant by ∼20%.

Tufts Human Robot Interaction Lab — Medford, MA
AI Research Engineer — *Apr 2024 – Aug 2024*

- Fine-tuned Mistral Large using LoRA and model quantization, achieving a 5× inference speedup.
- Built AI editor application (React, FastAPI, Python, ChromaDB) integrating LLM based suggestions into user research workflows.
- Developed a human-in-the-loop feedback system across 10K+ LLM suggestions, converting acceptance into preference data.
- Applied reinforcement learning (preference optimization) on user feedback, improving suggestion acceptance by ∼18%.
- Built an optimization loop across acceptance, edit distance, latency, and task completion time, reducing analysis time by ∼20%.

Boston University School of Public Health — Boston, MA
Data Scientist — *Oct 2022 – Jan 2024*

- Built automated Python and Selenium ETL pipelines to scrape, clean, and normalize unstructured datasets from 2000+ documents.
- Developed unsupervised learning pipelines using Gaussian mixture models and latent Dirichlet allocation for clustering and dimensionality reduction of high-dimensional text data.
- Built a BERTopic-based NLP pipeline to identify interpretable themes across 1,000+ documents, achieving a 92% topic-coherence score and enabling comparative analysis across cities.

PROJECTS

Multi-Agent Restaurant Decision Assistant: Architected a multi-agent recommendation system using Google ADK, Ollama, Neo4j, GraphRAG, evaluation pipelines, and deterministic guardrails to generate evidence-grounded restaurant recommendations from structured Yelp data and unstructured customer reviews.

TinyLLM – Build, Train, Align, and Deploy an SLM in JAX: Built a 20M parameter decoder only Transformer in JAX and implemented GRPO based post training using preference/reward signals; improved evaluator preference win rate from 51% to 70% and reduced repetitive generations by 25%.

Cross-Sectional Attention for Asset Pricing: Developed an attention-based asset-pricing model using 153 firm characteristics and monthly U.S. equity returns from 1963–2019; implemented random Fourier features, neural factor models, and multi-head cross-sectional attention trained with an MSRR objective, achieving an out-of-sample Sharpe ratio of 2.1.

SKILLS

Programming Languages: Python, R, C++, Stata, SQL.
Data Science Frameworks: Pytorch, Pyspark, Tensorflow, Keras, Flask, CUDA, JAX, numpy, pandas.
Statistics: A/B Testing, Hypothesis Testing, Causal Inference, Difference-in-Differences, Fixed Effects Models, Power Analysis.
AI/LLM: LangChain, LangGraph, CrewAI, PEFT/LoRA, FAISS, Google ADK, ChromaDB, MLflow, Docker, Kubernetes.
Cloud: AWS (S3, Glue, DynamoDB, ECS/EKS, SageMaker, Bedrock), GCP (BigQuery, Vertex AI), Azure (AI Foundry, AKS)

EDUCATION

Boston University — Boston, MA
M.S, Computer Science (Applied Data Analytics) — *Sep 2022 – Jan 2024*
Relevant Coursework: [Advanced Machine Learning, Big Data, Data Mining, Analysis of Algorithms]

Netaji Subhas University of Technology — Delhi, India
B.Eng, MPAE (Automation Engineering) — *Aug 2018 – May 2022*
Relevant Coursework: [Mathematics I/II/III, Mathematical Statistics, Operations Research, Artificial Engineering]

PUBLICATIONS

- Detection of Open Metal Sites in Metal–Organic Frameworks Using Machine Learning — Springer Nature Singapore, 2023. (Link)
- Deep Learning Based Cryptanalysis of Salsa20 — InCACCT, 2023 (Accepted)