

Harmanan Gurvinder Kohli

✉ hkohli40@gmail.com | 📞 +91 98195 27168 | 🌐 github.com/Harmanankohli | 🔗 linkedin.com/in/harmanankohli/

Research Interests

LLM Systems, Retrieval-Augmented Generation, Agentic Workflows, Explainable AI, AI Security, Multimodal AI, Natural Language Processing

Education

Birla Institute of Technology and Science (BITS), Pilani – WILP

Apr 2022 - May 2024

M.Tech. in Data Science and Engineering

CGPA: 8.14/10

Relevant Coursework: Machine Learning, Deep Learning, NLP, Data Mining, Data Structures and Algorithms

SRM University, Haryana

Jul 2014 - May 2018

B.Tech. in Computer Science and Engineering

CGPA: 8.14/10

Relevant Coursework: Database Management Systems, Software Engineering, AI & Expert Systems, Data Mining, DSA

Research Experience

Master's Dissertation: Explainable AI for Predicting Malware Infection

Nov 2023 - Apr 2024

Graduate Researcher

- Developed a machine learning model on Microsoft Malware dataset (8.9M devices) achieving 66.7% accuracy and 66.4% recall.
- Applied EBM, SHAP, LIME to generate feature-level local and global explanations with uncertainty estimates.
- Focused on interpretability, stakeholder trust, and GDPR-aligned transparency for cybersecurity applications.
- Engineered features from 4.1GB, 83-field dataset using memory-optimized pipelines (Polars), improving preprocessing efficiency by 50%.

Launchpad.ai (via Fellowship.ai)

Jan 2024 - Mar 2024

Data Science Fellow

- Built a multimodal price discovery agent using Gemini Pro Vision, achieving 97% accuracy on internal tests.
- Designed scalable RAG pipeline + VectorDB + tool-calling agent for 100K+ SKUs, reducing manual lookup time by 50%.
- Implemented fallback logic: queries VectorDB for known items and SerpAPI for unseen items, enabling robust hybrid reasoning workflows.
- Engineered Python + Selenium scrapers to collect product data for model training and validation.

Capgemini

Aug 2025 - Present

LLM Systems Research / GenAI Consultant

- Built financial document intelligence system for Master Facility Agreements and Sanction Letters using Docling extraction + semantic chunking (LangChain) to enable long-context RAG retrieval.
- Designed a 2x2 VectorDB architecture (Working Capital vs Long-Term Finance × MFA vs Sanction Letter) to isolate semantic neighborhoods and reduce embedding interference.
- Implemented agentic orchestration using LangGraph for multi-step workflows (classification → routing → retrieval → synthesis) with deterministic control flow.
- Integrated Text-to-SQL using Bedrock Agents to enable hybrid structured + unstructured reasoning on financial data.
- Deployed on AWS Lambda + Bedrock, optimizing latency, cost and structured output quality.

Projects

- **Automated Classification of Electrical Product PDFs:** Streamlit + Apache Tika + NLP + XGBoost to classify product PDFs into Lighting, Fuses, Cables, or Others.
- **Census Income Data Clustering:** K-Means + PCA to segment customer demographics for targeted marketing.
- **Lyrics Generation using LSTM:** Deep learning model using LSTM-based RNN to generate song lyrics from seed text; deployed as interactive Streamlit web app.

Skills

Programming: Python, SQL

LLM & GenAI Systems: LangChain, LangGraph, Bedrock Agents, VectorDB, Embeddings, RAG, Tool Calling, Prompting, Docling, OCR

ML / DL: scikit-learn, TensorFlow, Keras, XGBoost, Pandas, Polars, Numpy, SHAP/LIME, EBM

Cloud & Deployment: AWS (Lambda, Bedrock, DynamoDB, CloudWatch), Azure, GCP, Streamlit
Analytics & Tools: Tableau, Selenium, SAP, Excel

Work Experience

Sterlite Technologies Oct 2024 - Apr 2025

Data Scientist

- Developed safety stock optimization model (Z-score, AMC, lead time, SD), reducing inventory holding costs and improving reorder accuracy.
- Applied Counterfactual Explanations (XAI) to RFQ classification, identifying actionable parameter shifts that improved deal conversion.
- Built POC with Gemini Pro Vision to extract structured data from scanned vendor invoices, streamlining AP workflows.

Tata Consultancy Services Jan 2020 - Jan 2024

Systems Engineer

- Delivered analytics & planning solutions within SAP ecosystem, improving forecasting, reporting, and operational decision workflows
- Reduced processing time from days → minutes via pipeline optimization in SAP BPC.
- Integrated spatial + financial + business data across SAP S4HANA, GIS, and Real Estate modules for geospatial analysis.

Open-Source Contributions

Open-Source Contributor at Scikit – Learn Library

- Added usage examples for `f_regression()` and `silhouette_score()` functions to improve documentation.
- PRs: `f_regression()` - #28104 and `silhouette_score()` - #28125

Awards, Honors & Certifications

- | | |
|--|-----------|
| • Microsoft — Azure AI Engineer Associate | June 2025 |
| • Acknowledged Contributor — The Hundred-Page Language Models Book | 2025 |
| • TCS: Best Team (2x), On-the-Spot (5x), Special Initiative Award | 2021–2023 |
| • Semi-Finalist — TechGig Code Gladiators | 2018 |