

Yash Hooda

Richmond, TX | 713-534-5609 | Work Authorization: US Citizen |
yash.hooda6@gmail.com | <https://www.linkedin.com/in/yash-hooda-384430242/> | <https://github.com/yashhooda1> |
<https://yashhooda1.github.io/yashhooda/>

EDUCATION

University of Texas at Dallas
Bachelor of Science, Computer Science

Richardson, TX
May 2024

WORK EXPERIENCE

Ternio Group — Healthcare Supply Chain Analytics

Data Engineer

Sugar Land, TX
Sep 2025 – Present

- Engineered automated Python + PySpark ETL pipelines for multi-hospital surgical instrument and sterilization datasets, producing clean, production-grade Lakehouse tables.
- Built scalable Microsoft Fabric dataflows integrating SQL, PowerQuery M, and Spark notebooks for anomaly detection, OR utilization, and compliance risk analysis.
- Developed real-time Power BI dashboards used by hospital leadership to monitor throughput, utilization, and multi-million-dollar compliance metrics.
- Improved data transformation efficiency and reduced manual reporting workload by 70%, replacing Excel and legacy processes with cloud-native automation.
- Optimized SQL queries and Lakehouse structures to improve performance across high-volume datasets (10M+ rows).
- Collaborated with compliance officers, data engineers, and analytics teams to drive data-backed decisions at scale.

Archrock – Natural Gas Compression Services

Post Graduate Data Engineer Intern

Houston, TX
May 2025 – Aug 2025

- Automated phishing simulation and employee training metrics tracking using Power BI dashboards, reducing reporting delays by 50%.
- Used Microsoft Fabric, Lakehouses, and Azure Synapse pipelines to process unstructured cybersecurity training data into clean, analyzable tables.
- Developed logic to evaluate phishing click rates, training effectiveness, and report data quality issues using Python and SQL.
- Collaborated with cybersecurity and IT leadership to present insights, contributing to safer internal communication protocols.

Century Instrument & Machine Company (CIMCO)

Data Analyst (Temporary Contractor)

Houston, TX
Feb 2025 – May 2025

- Supplier Analysis & Cost Optimization: Conducted a supplier performance analysis evaluating pricing, delivery times, and quality metrics across multiple vendors, identifying 10% potential cost savings and optimizing procurement strategies.
- Data Management & Reporting: Analyzed and cleaned 1,000+ rows of operational data, developing automated reports and dashboards to track machine shop productivity, cutting reporting time by 30%.
- Training & Documentation: Designed a comprehensive machinist training document, improving onboarding efficiency. Researched and presented four cost-effective training software solutions, leading to a 25% reduction in training software expenses.
- Business Intelligence Support: Provided data-driven insights that improved supplier negotiations, enhanced operational efficiency, and streamlined internal processes.

Outlier.AI

AI Trainer (Freelance Remote)

San Francisco, CA
May 2024 – Feb 2025

- Evaluated AI model outputs, chatbot responses, and Python/NumPy code for accuracy and quality.
- Provided feedback to improve model performance by over 90%.
- Participated in various NLP and AI prompt evaluation projects.

Parijat Controlware

Software Engineer Intern

Houston, TX
May 2023 – June 2023

- Developed automation logic using PLCs in a team environment.
- Applied object detection and machine learning using RoboFlow and Python.

ACADEMIC PROJECTS

HoodaAgents, Personal

Tools Used: Python, LangChain, OpenAI API, Streamlit, Tavily Search

- Designed and built a full-stack AI Agent framework capable of dynamic memory recall and personalized chatbot personas.
- Integrated LangChain and OpenAI API to handle task delegation, RAG, and persona-based conversation workflows.
- Deployed public demo with UI via Streamlit and TavilySearch; included Hindi language support and vector store memory.
- Positioned the project as a customizable agent stack for future client deployments under a potential startup brand "HoodaAgents".

January 2025 – June 2025

Liver Cancer Prediction, Personal

Tools Used: Python, Scikit-Learn, Pandas, NumPy, Matplotlib, Seaborn, EDA, Feature Engineering

February 2025

- Analyzed a dataset of liver disease patient records to develop a machine learning model predicting the likelihood of liver cancer.

- Conducted exploratory data analysis (EDA) to identify key risk factors influencing liver cancer development.
- Implemented multiple ML algorithms, optimizing model accuracy using Random Forest and Logistic Regression.
- Created visualizations and insights to interpret feature importance and assist in early detection of liver cancer.

Virtual TA Chatbot AI, UT Dallas

February 2024 – May 2024

Tools Used: Python, Rasa, NLP, Visual Studio Code, LLM

- Collaborated within a 6-member team to design and develop a Virtual TA chatbot for the UT Dallas student portal.
- Engineered and trained conversational intents using Rasa to handle queries related to campus events, student services, university history, and academic support.
- Applied NLP techniques and LLM integration to enhance user experience and ensure accurate, context-aware responses.
- Improved student engagement by delivering a functional AI chatbot capable of assisting users with university-related information in real-time.

TECHNICAL SKILLS

Data Engineering: PySpark, SQL, Microsoft Fabric, Power BI, Dataflows, ETL, Lakehouse Architecture

Programming: Python (Pandas, NumPy, PySpark), SQL, Java, C/C++, JavaScript, R

Cloud: Azure (Synapse, Fabric), Google Cloud, IBM Cloud

Big Data / ML: Spark, TensorFlow, Scikit-Learn, Data Pipelines, RAG, NLP

Tools: Git, VS Code, Jupyter, Streamlit, PowerQuery, DAX

Strengths: Analytical Thinking, Problem-Solving, Cross-Team Collaboration, Communication

AI Tools: OpenAI ChatGPT, Google Gemini, Meta Llama, Ollama, DeepSeek R-1, Grok-3, Microsoft CoPilot, Claude, Git CoPilot

Interpersonal Skills: Teamwork, Problem-Solving, Communication, Conflict Resolution, Active Listening, Emotional Intelligence, AI Prompt Engineering

CERTIFICATIONS

IBM Professional AI Engineering Certificate, Coursera

August 2024

TensorFlow Developer Certification, DeepLearning.AI

January 2024

IBM Professional Data Science Certification, Coursera

November 2024

Vanderbilt University AI Prompt Engineering Certification, Coursera

February 2025