Jay Singhvi

2066603609 | jay.singhvi@outlook.com | Portfolio: jay-singhvi.github.io/ | linkedin.com/in/jay-singhvi/ | github.com/jay-singhvi/

TECHNICAL SKILLS

- **Cloud & Infrastructure**: AWS (S3, EC2, EKS, Lambda, IAM, DynamoDB, SageMaker, Transcribe, CloudFormation, SDK, CLI), Terraform, Docker, Kubernetes, Git, CI/CD Pipelines (GitHub Actions), Team Foundation Version Control (TFS)
- **Data Engineering & ETL**: Databricks, Delta Lake, Apache Airflow, SQL Server Integration Services (SSIS), ETL Pipeline Development, Data Modeling (Fact/Dimension), Incremental Loading, Parallel Processing, Error Handling, Data Pipeline Monitoring, Automated Data Upload, System Analysis & Design, SDLC
- **Databases & Query Languages**: SQL Server, PostgreSQL, MySQL, NoSQL (MongoDB, DynamoDB), T-SQL, Stored Procedures, Triggers, User Defined Functions (UDF), Views, Query Optimization, Database Partitioning, Index Management
- **Big Data & Analytics Platforms**: Hadoop, MapReduce, Snowflake, Snowflake Cortex Analyst, Apache Kafka, Redshift, Data Warehousing, Data Marts, OLAP Cubes
- **Machine Learning & AI**: Machine Learning (Supervised/Unsupervised), Deep Learning, Transfer Learning, Ensemble Modeling, Neural Networks, Feature Engineering, A/B Testing, LLMs (OpenAI, Claude, Google), RAG, Random Forest, Decision Trees, GridSearchCV, Cross-validation, Hyperparameter Tuning
- **Programming & Development**: Python, PySpark, Go, Flask, RESTful APIs, OAuth 2.0, Bash Scripting, API Development, Object-Oriented Programming, Environment Management, CLI Development, Batch Processing, Docker Compose
- **Data Science Libraries & Tools**: LangChain, Pinecone, NumPy, Pandas, PyTorch, Scikit-learn, TensorFlow, Streamlit, Spark SQL, Google Colab, Jupyter Notebook, Hugging Face, Transformers, Prompt Engineering
- **Visualization & Reporting**: Matplotlib, Seaborn, Power BI, Tableau, Custom Dashboards, KPI Tracking, Real-time Data Visualization, Automated Reporting Solutions, Business Metrics Tracking

WORK EXPERIENCE

Research Assistant (Data Science) **Seattle University**, Seattle, WA Sep 2022 – Ongoing

- Architected and implemented HIPAA-compliant data pipelines for asthma patient research, leveraging transfer learning methodologies to engineer personalized ensemble prediction models achieving 88% accuracy in asthma onset forecasting - a 20% improvement over traditional classifiers and 12% enhancement compared to established neural network architectures.
- Secured comprehensive CITI Program certifications in research ethics and human subject protocols, establishing governance frameworks for sensitive medical data handling while authoring peer-reviewed research publications on novel machine learning approaches for healthcare analytics.
- Research Projects under review for publishing:
  - **Asthma Patient Research Project (South Korean Hospital Collaboration):**
    - Spearheaded advanced clustering analysis implementing diverse algorithms (K-means, DBSCAN, Affinity Propagation, BIRCH, Mean-Shift, OPTICS) to identify intricate patient cohort relationships while developing a comprehensive environmental analysis framework integrating weather and air quality data.
    - Implemented and optimized open-source Lag-Llama foundation model within a customized prediction framework to reconstruct missing values in time-series patient data, conducting rigorous comparative analysis against traditional imputation methods using multiple statistical metrics (MSE, RMSE, MAE, R^2, MAPE).
    - Performed systematic hyperparameter optimization for Lag-Llama context window configurations, effectively expanding usable training data segments by 37% while maximizing prediction accuracy on missing time-series values.
  - **Agricultural Computer Vision Project (Washington State Farmers Collaboration):**
    - Engineered sophisticated 3D visualization pipeline transforming 2D drone imagery into comprehensive volumetric models while implementing state-of-the-art YOLOv10 object detection architecture for automated plant counting with 89.8% accuracy.
    - Orchestrated comprehensive aerial data collection strategy utilizing drone technology and developed extensive training datasets through meticulous manual annotation of 2,000+ images with 7,000+ unique bounding boxes using open-source platforms (CVAT, Roboflow).
    - Conducted comparative performance analysis between YOLO (v8, v11, and v12) architectures, achieving superior 93.6% accuracy with YOLOv12 while integrating ML-Depth-Pro libraries for precise distance measurement and size estimation of detected fruits.

Data Engineer **Yardi Systems**, Dubai, UAE Apr 2019 - Jul 2022

- Architected sophisticated ETL pipelines using SSIS for Yardi's real estate BI module, implementing multi-source data extraction and complex transformations that reduced deployment costs and implementation time by 75%.
- Engineered enterprise-scale BI visualization ecosystems with real-time data streaming and interactive dashboards, while implementing data quality frameworks that reduced inconsistencies by 90% and enhanced organizational decision-making.
- Led cross-continental collaboration across 4 specialized teams in 3 time zones, managing BI transformation initiatives for 50+ Middle Eastern clients while maintaining 99.99% system uptime and establishing proactive monitoring infrastructure that reduced system downtime by 85%.

- Designed extensible data ingestion architectures supporting multiple data types (historical, event-based, batch) with incremental load strategies, accelerating customer onboarding by 60% and enhancing processing performance by 40% through parallel computation and SQL Server partitioning methodologies.
- Developed automated analytical reporting solutions and comprehensive database objects (Stored Procedures, Triggers, UDFs, optimized Indexes) using advanced T-SQL techniques, generating 20 hours weekly efficiency gains while systematically optimizing query performance through execution plan analysis.

Data Engineer **Yardi Systems**, Pune, India Nov 2016 - Mar 2019

- Engineered and deployed sophisticated ETL architectures using Yardi's proprietary frameworks and SSIS to transform lease approval workflows, generating 50% enhancement in system utilization metrics and securing $3M+ in critical revenue retention.
- Architected high-performance data warehousing ecosystems with dimensional modeling and fact table optimization to power real-time dashboards monitoring 100+ KPIs across property management domains, while implementing incremental data processing methodologies that reduced processing latency by 15%.
- Developed comprehensive Data Mart solutions and intuitive analytical interfaces enabling stakeholders to effectively access complex property management datasets, while establishing data governance standards with detailed documentation protocols and automated testing infrastructure.
- Mentored 4 junior engineers while implementing robust exception handling frameworks within ETL packages, instituting rigorous code review protocols and streamlining deployment processes to maintain exceptional system reliability metrics.
- Designed database performance optimization strategies through strategic indexing, query tuning, and sophisticated partitioning for large-scale data environments, while developing versatile integration frameworks using SSDT to process heterogeneous data sources including Flat Files, Excel, relational databases, and Raw File formats.

EDUCATION

MS, Computer Science (specialization in Data Science) **Seattle University**, Seattle, WA Sept 2022 – Jun 2024

Recipient of Seattle University's Dean's Honor Roll

Courses: Distributed Systems, Machine Learning, Big Data Analytics, AWS Cloud Computing, Artificial Intelligence

MS, Computer Applications **Symbiosis International University**, India July 2015 - Apr 2018

Courses: Python, Linux scripting, Data Structure Algorithms, Relational Database management, Data Mining and Warehousing

BS, Information Technology **University of Mumbai**, Mumbai, India Jun 2011 – Jan 2015

Courses: Database management, SQL, Linux, Data Warehousing

PUBLICATIONS & CERTIFICATIONS (*Research Papers:* github.com/jay-singhvi/publications)

- Published in DAWAK 2024 - **Incremental SMOTE with Control Coefficient for Classifiers in Data Starved Medical Applications**
- Accepted to be published in SAC_2025 - **A Retrieval-Augmented Framework for Meeting Insight Extraction**
- In peer-review for IEEE JBHI 2025 - **Hybrid Deep Learning using Transfer Learning as Feature Extractor in Env. Health Risk Prediction**
- CITI Program - Responsible Conduct of Research – Engineers | Human Subjects Research for IRB (Faculty, Staff, and Student) *(Other Certificates:* **linkedin.com/in/jay-singhvi/details/certifications/***)*

PROJECTS *(GitHub Portfolio*: github.com/jay-singhvi/*)*

Resonate AI Chatbot *(Tech Stack: Python, Transformers, LangChain, Pinecone, Hugging Face, LLM, RAG, AWS S3 & AWS Transcribe, Infra as code, NLP, QLoRA)*

- Deployed production-grade RAG system with semantic graph clustering, achieving 90% BERT similarity scores and 89% precision/recall metrics by implementing overlapping document chunking algorithms that preserved contextual integrity across boundaries.
- Engineered high-performance vector embedding layer using Pinecone that maintained 85% cosine similarity while optimizing dimensional reduction techniques to balance query performance and semantic accuracy for high-volume retrieval scenarios.
- Implemented comprehensive LLM evaluation framework across OpenAI, Anthropic, and Google models with standardized metrics tracking hallucination rates, factual accuracy, and response relevance while developing a semantic routing architecture that reduced costs by dynamically selecting appropriate models based on query intent.
- Fine-tuned Llama 2 (7B) using QLoRA techniques, reducing computational requirements by 70% compared to full parameter tuning through gradient checkpointing and mixed precision training, while constructing specialized datasets that enhanced response quality for enterprise use cases.
- Architected distributed inference system with intelligent caching that reduced average response latency by 65% while maintaining quality, establishing continuous deployment pipeline for iterative improvements and significantly enhancing user experience metrics.

AI-Agentic Synthetic Data Generation: *(Tech Stack: Python, Docker, Anthropic API, Claude AI, CSV manipulation, Environment management, CLI)*

- Architected a containerized AI system with specialized agent architecture, engineering analyzer and generator components that produce synthetic datasets with statistical fidelity to source distributions through a modular framework that decouples analysis from generation processes.

- Engineered robust error propagation mechanisms with comprehensive logging while implementing an advanced batch processing framework that dynamically adjusts parameters based on memory availability and CPU utilization, optimizing throughput through parallelized pipelines while maintaining data consistency.
- Leveraged Anthropic Claude 3.5 Sonnet through context-aware prompt engineering that preserved statistical properties, implementing adaptive prompting strategies that adjust instruction specificity based on data complexity and developing validation pipelines to verify synthetic data distributions.
- Engineered parameter validation, contextual help systems, and intelligent defaults alongside a secure API key management system with environment-based configuration and just-in-time authentication.
- Published containerized solution to Docker Hub with comprehensive documentation, CI testing, and versioned releases, facilitating widespread adoption while maintaining quality control over distributed components..

Serverless Employee Management System: *(Tech Stack: Python, AWS S3 & AWS DynamoDB, Docker, AWS EKS, RESTful APIs, OAuth 2.0, Microservices)*

- Architected cloud-native SaaS platform for enterprise workforce management using AWS and microservices architecture, implementing specialized DynamoDB data partitioning strategies optimized for access patterns that reduced operational costs by 40% while maintaining performance.
- Engineered multi-tiered storage architecture that intelligently routes data between hot and cold storage based on access frequency analysis, reducing costs while maintaining sub-second query performance through automated lifecycle policies without compromising availability.
- Developed comprehensive RESTful API ecosystem with OAuth 2.0 authentication, granular permissions, and rate-limiting alongside a robust versioning system with backward compatibility guarantees that protected client implementations.
- Implemented containerized deployment using Docker with multi-stage builds that reduced image sizes by 65%, configuring advanced Kubernetes deployments on AWS EKS with automated horizontal pod scaling and sophisticated health monitoring.
- Designed event-driven processing workflows using AWS Lambda with calibrated memory allocations and concurrency management that delivered 5x throughput improvements, engineering fault-tolerant dead-letter queue systems with intelligent retry mechanisms that ensured reliable processing during service disruptions.

Personalized Marketing Campaign Optimizer: *(Tech Stack: Python, Scikit-learn, Pandas, Matplotlib, Seaborn, SMOTE, GridSearchCV, Machine Learning)*

- Architected marketing campaign optimization system using ensemble machine learning techniques (Decision Tree, KNN, Random Forest) that achieved 86% prediction accuracy in identifying high-conversion customer segments while developing custom evaluation metrics aligned with business objectives.
- Engineered model interpretation methodologies translating complex algorithmic patterns into actionable marketing insights alongside advanced class imbalance mitigation strategies combining SMOTE and Random Under Sampling that improved minority class prediction without compromising accuracy.
- Designed robust feature engineering pipelines transforming raw customer data into predictive indicators with compound features capturing complex behavioral patterns while implementing sophisticated EDA workflows with custom visualizations that revealed previously undetected patterns across market segments.
- Engineered automated data quality systems with statistical outlier identification and anomaly detection algorithms that substantially improved pipeline integrity prior to model training.
- Implemented comprehensive model optimization protocols combining stratified cross-validation with hyperparameter tuning using GridSearchCV and custom scoring functions, establishing automated workflows that systematically identified optimal configurations for production deployment.

SQL Query Assistant using Snowflake Cortex Analyst: (Tech Stack: AWS S3, Python, Snowflake, Streamlit, SQL, LLM, Snowflake Cortex LLM)

- Architected advanced SQL query generation system leveraging Snowflake Cortex Analyst that transforms natural language questions into optimized SQL queries through a sophisticated conversational interface for non-technical users.
- Engineered comprehensive semantic model framework with YAML configurations defining logical tables, dimensions, measures, and custom expressions that enabled high-precision query generation through detailed metadata mapping of complex database structures.
- Designed production-grade Streamlit chatbot with robust error handling, request management, and session state capabilities creating an intuitive experience for natural language query composition while maintaining secure connections throughout user sessions.
- Implemented advanced data ingestion workflows for multiple revenue datasets with optimized ETL processes and precise data type handling that ensured integrity across the analytical ecosystem while engineering API integration with Snowflake Cortex Analyst's REST endpoints.
- Designed verified query repository system capturing validated SQL patterns that enhanced accuracy and performance by leveraging proven structures for similar natural language inputs while implementing intelligent caching mechanisms that significantly reduced API call frequency.