

Jay Singhvi

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

PROFILE

Applied AI engineer shipping production LLM systems into regulated enterprise workflows — automations live across Accounting, Tax, Legal, HR, Digital Marketing, and Asset Management at a 900-employee REIT. Governing design principle: arithmetic and reconciliation stay in deterministic code; the model is confined to extraction and classification. Outcomes include ~1 FTE reduction, ~\$100K use-tax penalty exposure resolved, and the #1 most-used internal GPT company-wide.

WORK EXPERIENCE

Senior AI Analyst (Applied AI Engineering) **Welltower Inc**, Dallas, TX Aug 2025 – Present

- Shipped production AI systems adopted across six departments — **ChatGPT** custom **GPTs**, **Streamlit** applications, and **n8n** workflow automation — owning each end-to-end from department requirements through **UAT**, deployment, logging, and failure monitoring.
- Architected the shared automation platform every department workflow runs on — a queue orchestrator with a reusable sub-workflow library, atomic retry, and centralized error recovery, so adding a new department automation requires no orchestrator change.

Data Scientist (Research Assistant) **Seattle University**, Seattle, WA Sep 2022 – Aug 2025

- Predicted asthma onset at **88%** accuracy — **20%** better than traditional methods — with transfer-learning and **PySpark** ensemble models on sparse medical data, improving forecasting accuracy by **12%** over standard neural networks, on HIPAA-compliant **Docker** and **AWS** pipelines.

Technical Consultant **Yardi Systems**, Dubai, UAE Apr 2019 – Jul 2022

- Cut implementation time and deployment cost **75%** for real estate **BI** modules with **SSIS** and **SQL Server** ETL pipelines, and accelerated customer onboarding **60%** with multi-format ingestion (batch, real-time, historical) partitioned for **40%** faster processing.
- Reduced data inconsistencies by **90%** with real-time streaming **BI** dashboards and saved **20** hours weekly by automating reporting in **T-SQL** stored procedures, triggers, and tuned indexes.
- Delivered end-to-end BI for **20+** UAE real estate clients - requirements through build, testing, and client-validated UAT - then drove adoption through C-suite and property-staff walkthroughs, turning delivered dashboards into the reporting system leadership ran the business on.

Software Engineer **Yardi Systems**, Pune, India Nov 2016 – Mar 2019

- Secured **\$3M+** in revenue retention and lifted system usage **50%** with **SSIS** ETL systems that streamlined lease approval workflows.
- Powered real-time dashboards across **100+ KPIs** with a dimensionally modeled data warehouse and data marts — **15%** faster processing via incremental loading — and set the governance standards, automated testing, and query tuning behind them.

PROJECTS (Welltower Inc)

Insurance Requirements Extraction & Compliance Benchmarking — Legal

- Replaced manual clause-by-clause insurance review in acquisition diligence for the Legal team — document-intelligence workflows read purchase and sale agreements and extract every requirement with its dollar limits, then a deterministic benchmark engine flags each limit green, yellow, or red against historical deal statistics.
- Made every extracted limit auditable to its source clause — a three-stage **LLM** chain on **Amazon Bedrock** handling deal context, clause extraction with verbatim quotes and page citations, then normalization of free-text limits onto standard coverage types, each stage grounded with anti-hallucination rules.

Webb — Enterprise HR Assistant

- Cut HR support tickets by **85%** by migrating the company-wide HR assistant from a legacy **Slack** bot to a **ChatGPT** custom **GPT** - **700** users (**78%** of organization), **200** monthly active users, **100+** daily queries, and the **#1** most-used custom **GPT** company-wide for six months.
- Held the assistant secure through emerging adversarial use across successive prompt revisions — injection red-flag patterns, identity lock, silent refusal of system-prompt disclosure, web search narrowed to checklist-gated cases, and the connector scoped to user identification.
- Engineered out the silent-wrong-answer failure mode — a location gate withholds answers until the employee's region is known, routing only accurate information, and barring extrapolation from partial documents.
- Shipped every revision defect-free by gating release on a regression suite and a 10-day Human Capital hallucination audit — cases pinning both correct and specific wrong responses across policy, regional routing, refusals, and injection attempts.
- Built the AI governance layer on the **OpenAI** Compliance API, landing conversation data in **Snowflake** with a review frontend and human-in-the-loop flagging, now HR's standing weekly review.

Use Tax Compliance Automation

- Resolved approximately **\$100K** in use tax penalty exposure and established the standing control preventing recurrence across all operating jurisdictions — automating use tax determination for state, city, and county codes for **3,000+** invoices monthly across **1,500+** properties.
- Eliminated a hallucination risk unacceptable in a compliance calculation — rebuilt an **Azure** live-search-per-line-item design as an **n8n** workflow grounded in **Thomson Reuters** tax code data: a reference-data lookup handles known commodities, unmatched lines escalate to a higher-capability model returning a cited proposal, and line-total mismatches route to human review.

Property Site Map — Data-Driven Rent Setting

- Built a custom **GPT** generating review-ready **Excel** site maps for unit-level rent setting — a controlled A/B test across two structurally identical halves of one property showed generated pricing beat the incumbent on occupancy, rental income, and resident preference; asset management adopted it as the standing pricing framework.

TalkFurther Call Center Analytics

- Recovered **15%** additional leads from dropped conversations and drove a **3%** revenue increase across **50+** properties — replaced a third-party batch transcript vendor with an in-house AI analysis system moving **3,000+** calls to daily internal review, now property managers' standing **KPI** via **Power BI** dashboards linking call topics to occupancy and conversion.

Performance View Chat — Natural Language to SQL → Claude Plugin

- Built a natural-language-to-SQL system for the Investment team across **1,500+** properties, removing their dependency on the data team for ad-hoc analysis — later retiring the **Streamlit** UI for a **Claude** plugin pairing a **Snowflake** connector with skills encoding the team's output formats and report structures, so analysis runs directly in chat.

Scope Creation Tool — Capital Projects Scope Generation

- Replaced manual per-project drafting for the capital projects team with a guided generation tool that turns a project manager's answers into complete construction scope-of-work documents — now used by every project manager on the team.
- Kept every generated spec traceable to precedent and correct for its site — grounded generation in a **1,022-item** scope corpus and **16** templates across **41** trade categories, conditioned per trade on a **2,384-property** master dataset, then an automated climate and wind pass rewrites specs for the building's environment.
- Preserved **Word** numbering, styles, and run-level formatting through an **LLM** rewrite — a similarity-based merge re-aligns generated text onto the original paragraph objects and keeps the original wherever confidence is low.

Journal Entry Tie-Out Automation

- Recovered **~1** FTE annually by cutting payroll tie-out from **8** people at **6** hours weekly to a single exception reviewer at **4–5** hours — implementing payroll cash tie-out, payroll journal entries, and suspense item clearing as **n8n** production workflows, after documenting **25–30** undocumented processes across **5–6** departments as machine-readable specifications.
- Designed reconciliation workflows structurally incapable of mutating the general ledger —proposes changes (missing chart-of-accounts entries, mapping corrections) for human approval rather than applying them, making an incorrect automated inference impossible to post.

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

Resonate AI Chatbot: (Tech Stack: Python, LangChain, Transformers, FAISS, Pinecone, QLoRA, AWS Transcribe)

- Reached **90%** BERT similarity and **89%** precision/recall with a **LangChain** RAG system — semantic graph clustering and overlapping chunking that preserved context across boundaries, over a **FAISS** and **Pinecone** embedding layer holding **85%** cosine similarity.
- Reduced response latency by **35%** with a distributed inference and caching layer, and cut model cost via semantic routing across **OpenAI**, **Anthropic**, and **Google** models — evaluated on hallucination rate and response accuracy, with **Llama 2 (7B) QLoRA** fine-tuning cutting compute requirements by **50%**.

TECHNICAL SKILLS

- **AI & Machine Learning:** Python, PyTorch, RAG, LLM evaluation, embeddings, vector databases (FAISS, Pinecone), prompt engineering, guardrail and anti-hallucination design, OpenAI (GPT, custom GPTs, Compliance API), Anthropic Claude (plugins, skills, MCP), Amazon Bedrock, LangChain, Hugging Face, Scikit-learn
- **Data & Analytics:** Snowflake (Cortex Analyst), SQL, T-SQL, PostgreSQL, PySpark, Spark SQL, Pandas, NumPy, ETL pipeline development, SSIS, dimensional modeling, data warehousing, incremental loading, query optimization
- **Cloud & Infrastructure:** AWS (S3, SageMaker, Transcribe, Bedrock), Docker, Git, GitHub Actions CI/CD, RESTful APIs
- **Applications & Reporting:** Streamlit, n8n, Power BI, Tableau, KPI dashboards, Excel

EDUCATION

MS, Computer Science (specialization in Data Science)	Seattle University , Seattle, WA	Sept 2022 – Jun 2024
MS, Computer Applications	Symbiosis International University , India	July 2015 – Apr 2018
BS, Information Technology	University of Mumbai , Mumbai, India	Jun 2011 – Jan 2015

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](#)
- Published in SAC 2025 - [A Retrieval-Augmented Framework for Meeting Insight Extraction](#)
- 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/)