

ASHISH GAUTAM

Atlanta, GA 30341 | (256) 288-6948 | ashish.gm74@gmail.com | [LinkedIn](#) | [GitHub](#) | silentashish.com

PROFESSIONAL SUMMARY

Full-Stack and Machine Learning Engineer with 7+ years of experience shipping production web platforms and AI systems across NASA-funded research, e-commerce, and fintech. Expert in Python (Django, FastAPI), React/TypeScript, and AWS, with deep hands-on experience in LLM applications, RAG pipelines, LangGraph agent orchestration, vector databases, and parameter-efficient fine-tuning (LoRA/DoRA). Published researcher (70+ citations) completing an M.S. in Computer Science. Proven quantifiable impact: 30% faster data delivery at NASA, 35% lower API latency at Houzz, 25% infrastructure cost reduction via Kubernetes autoscaling.

TECHNICAL SKILLS

- Languages & Frameworks:** Python, TypeScript, JavaScript, Go, SQL, Django, Django REST Framework, FastAPI, Flask, Node.js, React, Next.js, GraphQL
- AI/ML & GenAI:** PyTorch, TensorFlow, scikit-learn, Hugging Face Transformers, LLM Fine-Tuning (LoRA, DoRA, Unsloth), RAG, LangChain, LangGraph, PydanticAI, Model Context Protocol (MCP), OpenCV, LSTM, CNNs
- Vector & Graph Databases:** Weaviate, ChromaDB, pgvector, Neo4j, Knowledge Graphs, Semantic Search, Embeddings
- Cloud & Infrastructure:** AWS (EC2, S3, RDS, Lambda, SageMaker, ECS, ECR, ALB/ELB, Glue, Redshift, Step Functions, CloudWatch), GCP (BigQuery), Docker, Kubernetes, KEDA
- MLOps & DevOps:** Terraform, Ansible, Jenkins, GitHub Actions, ArgoCD, CI/CD, Blue-Green Deployments, Apache Airflow, Apache Kafka
- Data & Observability:** PostgreSQL, Redis, PgBouncer, PySpark, Pandas, NumPy, Prometheus, Grafana, Loki, OpenSearch

PROFESSIONAL EXPERIENCE

NASA IMPACT

Aug 2024 – Present

Research Engineer (ML & Backend) | Part-Time

Huntsville, AL

- Drove a 30% increase in data delivery speed to clients through the implementation of the new Tasking Management System, significantly improving data ingestion and request handling for the NASA CSDA program.
- Initiated and built an agent-based data search project leveraging LLMs with Retrieval-Augmented Generation (RAG) using LangChain and LangGraph, enabling users to intelligently search and retrieve specific data from a massive STAC metadata catalog.
- Developed a RAG-based automatic metadata file generator for the CSDA program, utilizing local LLM models with Weaviate and ChromaDB as vector databases and maintaining a knowledge graph for NASA Spatiotemporal Asset Catalog compliance.
- Managed scalable infrastructure across AWS (S3, EC2, RDS, Lambda, Step Functions) to ensure a robust and highly available system using technologies including Next.js and FastAPI.
- Implemented data ingestion pipelines for new vendors, including Satellogic and Maxar DEM, within the CSDA system.
- Developed and pioneered the new NASA CSDA tasking management system, utilizing a Django/Airflow backend and a React and TypeScript frontend to manage commercial satellite data requests and ingestion.
- Utilized Django, Apache Airflow, FastAPI, LangChain, LangGraph, Weaviate, ChromaDB, Amazon RDS, Amazon S3, Amazon EC2, Grafana, and Prometheus to support project development and infrastructure management.
- Project: [Data Acquisition Request System](#)

Houzz

Apr 2023 – Jul 2024

Software Engineer (ML & Backend) | Remote

Palo Alto, CA

- Boosted website performance by 20% by leading the migration of a legacy PHP-based system to a modern Python/Django backend with Server-Side Rendered (SSR) architecture, reducing API response times by 35%.
- Developed an automated background removal pipeline for importing product images from external sources, leveraging OpenCV and deep learning models (U-Net) to process 5,000+ images daily with 95% accuracy.
- Architected and managed heavy AWS infrastructure including EC2, ECS, S3, RDS (PostgreSQL), and CloudFront, ensuring 99.9% uptime across production services.
- Implemented automated CI/CD pipelines using Jenkins with blue-green deployment strategies, reducing deployment downtime by 90% and enabling zero-downtime releases.
- Configured KEDA (Kubernetes Event-Driven Autoscaling) for dynamic workload scaling, reducing infrastructure costs by 25% during off-peak hours.
- Managed infrastructure as code using Ansible and Terraform, provisioning and maintaining Amazon RDS instances, load balancers, and auto-scaling groups across multiple environments.
- Project: [Houzz Moodboard](#)

Event Horizon Partners

May 2021 – Feb 2023

Software Engineer (ML & Backend) | Remote

Seattle, WA

- Developed a quantitative analytical trading platform while strictly adhering to Agile methodologies, ensuring timely and high-quality deliverables.
- Integrated Python with Django and the Django REST Framework for reliable backend services, managing SQL databases and emphasizing cross-functional collaboration.
- Built responsive, intuitive interfaces using React.js on the frontend, integrating Plotly.js for data visualization and Context APIs for robust state management.
- Streamlined platform releases by implementing Git version control and CI/CD pipelines to enable smooth deployments to AWS, including Lambda functions for event-driven processes.
- Implemented comprehensive observability using Prometheus + Grafana for metrics, Loki + Grafana for log aggregation, and AWS CloudWatch for infrastructure monitoring across the entire trading platform.

- Utilized Django, React.js, Plotly, and scikit-learn to enhance platform functionality, resulting in improved data processing and visualization capabilities.
- Project: [Trading Platform](#)

Diagonal Software

Mar 2019 – May 2021

Software Engineer (Full-Stack Development) | Remote

Hessen, Germany

- Led the backend development of the ListInfo hotel booking platform using Python, Django, and Django REST Framework, architecting RESTful APIs that served 100K+ daily requests.
- Deployed and managed the full application stack on AWS using EC2, ALB, ELB, and ECR for containerized deployments, achieving 99% uptime.
- Designed and implemented CI/CD pipelines for automated testing, building, and deploying Docker containers to AWS ECS, reducing release cycles by 50%.
- Optimized PostgreSQL database performance through query tuning, indexing strategies, and connection pooling, resulting in a 40% reduction in average query execution time.
- Project: [ListInfo](#)

PERSONAL PROJECTS

Grafana Copilot — Natural-Language Dashboard Generator | [GitHub](#)

- Built an AI-powered Grafana app plugin that generates complex dashboards from natural-language prompts, eliminating hand-authored panel JSON; architected as a React/TypeScript frontend, secure Go proxy protecting API credentials, and a headless Python/FastAPI agent.
- Orchestrated multi-step LLM workflows with LangGraph and PydanticAI, using Model Context Protocol (MCP) servers to autonomously discover OpenSearch data sources and write dashboards directly to Grafana.
- Engineered low-latency WebSocket token streaming to the browser plus a database-backed config layer with hot-swappable LLM providers (OpenAI, Anthropic, Ollama) and PostgreSQL-persisted conversation history.

CSDA Data Search Agent with Knowledge Graph | [GitHub](#)

- Built a conversational agent (FastAPI, Chainlit, LangGraph, PydanticAI) that lets scientists search and download NASA Earth-observation datasets in plain English, using local LLMs and embeddings via Ollama to translate ambiguous prompts into precise STAC search filters.
- Designed a dual-database retrieval architecture — PostgreSQL + pgvector for auth and conversational state recall, Neo4j for relationship-aware semantic catalog retrieval — with a real-time Mapbox panel over WebSockets visualizing spatial queries and dataset heatmaps.
- Containerized the backend with Docker Compose, integrating NASA Earthdata OAuth for secure asset downloads and Logfire for end-to-end agent observability.

Swochalit — LSTM-Based Algorithmic Trading Bot | [GitHub](#)

- Automated slow, emotion-driven manual trading by engineering a time-series pipeline (Python, TensorFlow/Keras) that cleans OHLCV market history, derives technical-indicator features, and trains stacked LSTM networks to forecast short-horizon price movement.
- Implemented a rule-based execution layer converting model forecasts into buy/sell/hold signals, backtested against held-out historical data before deployment.

Nepali Poem Generator — Fine-Tuned LLM for a Low-Resource Language | [GitHub](#)

- Curated and preprocessed a corpus of historic Nepali poems (Devanagari script), then fine-tuned Google Gemma with Unsloth using parameter-efficient LoRA adapters on a single consumer GPU.
- Produced a model generating stylistically coherent Nepali verse from short prompts, demonstrating practical low-resource-language adaptation of open-weight LLMs.

RESEARCH & PUBLICATIONS

Total citations: 70+

- **Mitigating Reconstruction Loss in Neural Compression for Remote Sensing Downstream Tasks Using Weight-Decomposed Low-Rank Adaptation** — M.S. Thesis, 2026 ([link](#)). Applied DoRA adapters to a neural codec's synthesis transform, no retraining, no bitstream change, via label-free distillation from a segmentation teacher, improving downstream consistency mIoU from 0.643 to 0.723 (+8.04 pp) at unchanged PSNR/MS-SSIM, outperforming JPEG, JPEG 2000, and WebP at low bitrates.
- **Comparative Analysis of Multiple Deep CNN Models for Waste Classification** — 2020 ([arXiv](#)). Benchmarked deep CNN architectures for automated waste sorting on TrashNet plus a self-collected dataset; fine-tuned ResNet-18 to **87.8% validation accuracy** and deployed it in a smart dustbin prototype that physically segregates waste.

EDUCATION

University of Alabama in Huntsville

Aug 2024 – May 2026

Master of Science, Computer Science

Huntsville, AL

- Coursework: Cloud Computing, Machine Learning, Advanced Operating Systems, Data Structures & Algorithms, Big Data

Institute of Engineering, Pulchowk Campus

Mar 2015 – Mar 2019

Bachelor of Engineering, Electronics and Communications Engineering

Kathmandu, Nepal

- Coursework: Artificial Intelligence, Big Data Technology, Object-Oriented Programming, Digital Signal Processing

CERTIFICATIONS

[Meta Front-End Developer](#) | [Meta React Native Specialization](#) | [Django Application Development with SQL and Databases](#)