

Shubhankar Goje

AWS | Applied AI | Intelligent Automation

jobs.shubhankar.goje@gmail.com • Seattle, WA

shubhankargoje.com • linkedin.com/in/shubhankargoje • github.com/shubhankargoje

PROFESSIONAL SUMMARY

Data and AI professional delivering AWS analytics, generative AI applications, and agentic workflow automation for Amazon Career Choice. Experience integrating cloud services for incident diagnosis, application and ETL remediation, automated testing, and human-reviewed delivery. Built self-service analytics and conversational AI solutions with SQL validation, access controls, and audit logging. Master's in Data Science.

TECHNICAL SKILLS

AWS: Amazon Bedrock Agents, Lambda, Redshift / Redshift Data API, S3, Glue jobs and crawlers, CloudWatch, QuickSight, API Gateway, IAM, Secrets Manager.

AI & Automation: Generative AI, multi-agent orchestration, tool integration, Playwright MCP, Kiro Crew, root-cause analysis, iterative remediation, automated QA, RAG, LangChain, LLM evaluation, PEFT / LoRA.

Engineering Practices: Requirements documentation, root-cause analysis, incident troubleshooting, smoke testing, alpha/beta validation, code-review preparation, human approval, access control, audit logging.

Development & Data: Python, SQL, Java, R, HTML, XML, ETL pipelines, event schemas, data quality, Docker, Airflow, pandas, NumPy, PyTorch, TensorFlow, Hugging Face, scikit-learn.

PROFESSIONAL EXPERIENCE

Data Analyst | Amazon Career Choice

2024 – Present

Seattle, WA | Contract engagement through Fiduciary Tech

Amazon Career Choice is a global prepaid education and skills-training benefit that has served 300,000+ employees through 600+ education partners. Build analytics, self-service AI, and workflow automation for the technology team, translating program needs into data products covering eligibility, participation, completion, job placement, and sentiment.

- Reduced incident turnaround from 24 to 3 hours (87.5%) by building an agentic workflow that investigates tickets and CloudWatch alarms across Lambda, Glue jobs and crawlers, and Redshift ETL, then develops and tests application and pipeline fixes.
- Delivered fixes for 83 incidents that passed code review without revisions or blocking comments in 2026 by automating root-cause investigation, fix-specific test suites, iterative alpha/beta validation, and code-review submissions with change rationale and testing documentation; retained human approval for production deployment.
- Reduced QA execution time by 87.5% by defining web-analytics event payload requirements and building a Playwright MCP agent workflow to exercise frontend interactions, compare emitted events against the schema, and iteratively develop and test fixes.
- Saved an estimated 8–24 hours of team effort per week for a 15-person technology team by integrating ticket investigation, service diagnostics, fix development, test execution, and technical documentation into a repeatable agent workflow.
- Established analytics validation across eight interaction categories—page visits, clicks, agent messages, agent actions, uploads, validations, filters, and actions—by documenting event payloads, session/event/ingestion identifiers, timestamps, geographic attributes, and alarm requirements for agent-driven QA.
- Reduced analytics support effort by 40–60 hours per month by building an AI chatbot over curated North Star tables, enabling nontechnical program managers to independently query program- and school-level performance.
- Standardized reporting for seven priority-zero business metrics by engineering approximately 3,000 lines of SQL to reconcile conflicting definitions and power automated AWS pipelines, QuickSight dashboards, and monthly KPI reports.
- Integrated employee sentiment into a seven-metric performance framework by building a Qualtrics pipeline that automated lifecycle-based survey targeting, scheduling, transformation, record matching, deduplication, failure alerts, and reporting.
- Informed the planned replacement of an underperforming AI recommender through analysis across 16 countries by using S3, Glue, Redshift, SQL, and QuickSight to identify low click-through rates, negative feedback, and weak participation.
- Enabled performance comparisons across 600+ education partners by developing school-level reporting on student support, data-update timeliness, participation, completion, and employee experience.
- Guided geographic marketing decisions across 16 countries and 40+ career paths by analyzing relationships between program promotion, participation, and job transitions at country, state, city, and site levels.
- Expanded analytics coverage from 14 to 16 countries by translating Japan and Mexico launch requirements into country-specific data transformations and QuickSight reporting.

SELECTED AI & AUTOMATION PROJECTS

Movie Chatbot for Structured Data | AWS Agentic AI

2026

Built a self-service natural-language interface to structured movie data using Amazon Bedrock Agents, Python, Lambda, Redshift Serverless, API Gateway, and a web frontend. Integrated user confirmation, read-only access, query safeguards, and execution diagnostics into the request lifecycle.

- Achieved 98% answer correctness in project evaluation by implementing a four-agent architecture that coordinates natural-language requests, SQL generation, query execution, and result explanation over structured movie data.
- Achieved 100% SQL execution success in project evaluation by integrating Bedrock action groups with a Python Lambda and the Redshift Data API, with SELECT-only SQL validation, query-cost checks, row limits, and timeout cancellation.
- Established four specialized agent roles with one execution-authorized agent by separating orchestration, SQL generation, execution, and explanation, and requiring user confirmation before query execution.
- Enabled execution troubleshooting through four diagnostic fields—statement ID, execution time, row count, and correlation ID—by implementing S3 JSONL audit logs and a frontend debug panel, with read-only database access and IAM-controlled service permissions.

Talk Like | Conversational AI and RAG | CU Boulder

2024 – Present

Developed a personalized conversational AI combining Persona, an LLM fine-tuned to emulate individual speech patterns, with CortexCache, a retrieval-augmented memory framework for changing schedules, preferences, and relationships.

- Achieved a G-Eval score of 0.93 for contextual accuracy by engineering CortexCache, a RAG framework that incorporates dynamic schedules, preferences, and social relationships into conversational responses.
- Supported 20+ user profiles and context windows of up to 5,000 tokens by integrating APIs with CortexCache for real-time management of personalized conversational context.
- Achieved a BLEU score of 0.21 for personalized response generation by fine-tuning Llama 3 on nine years of chat data using parameter-efficient fine-tuning and LoRA.

Multi-Evidence Natural Language Inference | CU Boulder

2023

Built and evaluated models that classify clinical-trial statements as entailment or contradiction, addressing numerical reasoning, specialized terminology, and evidence retrieval across report sections.

- Achieved a peak F1 score of 66.74% on biomedical entailment and contradiction classification by fine-tuning Llama 2 and PubMedBERT using PEFT and LoRA for evidence interpretation in breast cancer clinical-trial reports.

Unmanned Aerial Vehicle | Predictive Models and Automation

2019 – 2021

Developed an autonomous package-delivery system combining predictive control, navigation, and image processing; the work resulted in a published Indian patent application and two research publications.

- Improved flight stability by 70%, measured through control precision metrics, by redesigning a cascaded PID controller and implementing a custom regression model for gain-factor prediction.
- Reduced flight time by 54% by developing a deep-learning navigation algorithm using PyTorch and TensorFlow for autonomous flight.
- Improved delivery accuracy by 30% by training a Haar cascade in Python to process imagery from an onboard camera.

Lung CT Segmentation | Applied Machine Learning

2021 – 2022

Developed neural-network methods to automate lung segmentation and remove extraneous information from medical imagery.

- Achieved 87% overall accuracy in 2D medical-image segmentation by developing a context encoder network with dense atrous convolution and residual multi-kernel pooling blocks.
- Improved segmentation performance on complex lung-tissue images by 20% by integrating a U-Net architecture with a pretrained ResNet block to preserve spatial information.

EDUCATION

Master's in Data Science | University of Colorado Boulder

2022 – 2024

GPA: 3.85/4.0 | Machine Learning, Natural Language Processing, Statistical Analysis, Data Mining

BE in Electronics and Telecommunication Engineering | PICT

2018 – 2022

Honors in Data Science | GPA: 3.71/4.0

PUBLICATIONS & PATENT

Patent application published: Unmanned Aerial Vehicle (UAV) and Navigation Method Thereof, Indian Application No. 202221010717 (2022).

Publications: Path Planning and Obstacle Avoidance for UAV in an Uncharted Environment (IJRASET, 2021); Cascaded PID Control System for UAV with Gain Factor Prediction Using ML (IJRASET, 2021).