

# Giovanni Lunetta

Technology LDP Analyst

 [glunetta02@gmail.com](mailto:glunetta02@gmail.com)  Pequannock, NJ  [Github](#)  [Medium](#)  [LinkedIn](#)  [My Website](#)

## PROFESSIONAL SUMMARY

---

Data Scientist with 2+ years of experience in healthcare and commercial analytics, focused on machine learning, advanced analytics, and AI applications. Proven track record delivering measurable business impact: built GenAI document automation with 50% token reduction, automated pipelines saving 200+ hours annually, and attributed multi-million dollars in AWS costs at 85%+ coverage. Strong skills in Python, AWS, machine learning (XGBoost, clustering, statistical modeling), with experience in LLMs and GenAI systems, product ownership, and cross-functional collaboration.

## WORK EXPERIENCE

---

**Technology LDP Analyst on Data (Data & Decision Science),  
Johnson & Johnson (MedTech R&D)**

06/2024 – Current | Raritan, NJ

- **GenAI Document Automation (Document AI/RAG):** Built GenAI-powered automation for Clinical Evaluation Report (CER) extraction with semantic search and schema validation, achieving **50% token reduction** through optimized document embedding strategies and LLM cleaning agents. Implemented production document intelligence system with quality assurance workflows.
- **Pre-Surgical Configuration Analytics (Preference Modeling):** Designed and validated unsupervised preference models (Fuzzy C-Means + rule-based logic), achieving **67% surgical classification** (+17 points over baseline) across **10 iterative refinements** with **4 independent feature sets** (silhouette score **0.66**). Delivered comprehensive Tableau dashboard (6+ iterations) adopted for clinical decision support and commercial strategy.
- **AWS Cost Optimization (FinOps):** Architected **3-Lambda event-driven pipeline** automating cost data management –eliminating **200+ hours annually** of manual processing. Achieved **multi-million dollar cost attribution** across enterprise AWS infrastructure at **85%+ coverage** with self-service dashboards enabling data-driven optimization decisions.
- **Persistency Model (ML Pipeline Validation):** Validated end-to-end **14-node Kedro ML pipeline** for biologic drug persistency modeling (**1.3M+ records processed**, **XGBoost**, **ROC-AUC 0.717**). Conducted comprehensive pipeline review including architecture documentation, feature analysis, and scalability assessment across **multiple therapeutic indications** using advanced claims analytics.
- **HCP Segmentation & Treatment Analytics:** Performed unsupervised clustering analysis on healthcare provider data with **90+ engineered features** using **K-Means and Fuzzy C-Means** (silhouette score 0.42), validated with **ANOVA and Kruskal-Wallis** tests. Developed treatment-pattern analytics identifying priority healthcare providers for commercial targeting, delivering actionable HCP personas and insights adopted by field teams.
- **Microsoft 365 Copilot Program (Product Operations):** Managed enterprise pilot program, **provisioning 877 users** with automated lifecycle workflows (Python + Excel macros) that **reduced processing time by 50%**. Built daily reporting system serving as single source of truth for program leadership, enabling data-driven license allocation decisions.

**Technology Summer Intern (Data Science),  
Johnson & Johnson (Janssen Pharmaceuticals)**

05/2023 – 08/2023 | Titusville, NJ

- Developed a machine learning model using healthcare claims data to predict patient drug eligibility 6 months in advance, optimizing doctor outreach targeting and pharmaceutical processes.

## EDUCATION

---

**M.S. in Data Science, University of Connecticut**

09/2023 – 05/2024 | Storrs, CT

**B.A in Stats & Poli Sci, Math (Minor), University of Connecticut**

09/2020 – 05/2023 | Storrs, CT

## PROJECTS

---

### AI Portfolio Assistant (Self-Evaluating LLM System)

- Architected production conversational AI using **GPT-4o-mini** and **Gemini 2.0 Flash** featuring **automated quality control, function calling for lead capture, and self-improving feedback loop**. Implemented **evaluation retry loop** with dual-model assessment (generation vs. evaluation) and JSONL-based analytics for continuous improvement. Deployed via **Gradio at giovannilunetta.com** ↗.

### DateBack (macOS Application)

- Built production macOS desktop app (**Electron + Python**) processing **Snapchat exports (50GB+)** with **EXIF date restoration, batched cloud streaming, and Apple code signing**. Demonstrates full-stack engineering and professional distribution workflows (**notarization, auto-updates, DMG installer**) with 100% offline-first architecture.

### Hack4Health — Surgical Intelligent Assistant

- Created voice-activated, offline-capable AI assistant supporting surgical teams in the OR with real-time guidance and automation. **Placed Top 3 in MedTech category; selected for Proof-of-Concept funding.**

## SKILLS

---

- **GenAI & LLM Engineering:** OpenAI API, LangChain, LangGraph, RAG architectures, hybrid retrieval (vector + BM25), Postgres + pgvector (HNSW), schema-validated extraction, prompt engineering, token optimization, Microsoft Copilot, GitHub Copilot
- **Machine Learning & Modeling:** Python, R, XGBoost, neural networks, Kedro pipelines, semi-supervised learning (co-training, self-training), clustering (Fuzzy C-Means, K-Means), PCA, evaluation (CV, AUC/PR, silhouette, SHAP), ANOVA, Kruskal-Wallis, feature engineering
- **Data Engineering & Cloud:** AWS (Lambda, S3, Athena, Redshift, IAM, CloudShell), event-driven pipelines, ETL automation, SQL (Postgres), Databricks (fundamentals), schema design, query optimization
- **Analytics & BI:** Tableau (Desktop Specialist), Power BI (semantic models, Q&A/synonyms), Excel (macros, automation), claims analytics, HCP segmentation, treatment-pattern analysis
- **Product & Leadership:** Product ownership, agile delivery, change management, stakeholder coordination, cross-functional collaboration, technical documentation

## CERTIFICATES

---

AI Engineer Agentic Track: The Complete Agent & MCP Course, AWS Certified Cloud Practitioner, Tableau Desktop Specialist, Harvard/edX - Data Science Professional Certificate; CS50x ↗, DeepLearning.AI – ML Specialization, Prompt Engineering, LangChain, Diffusion Models, Babson Exec Ed - J&J Design Thinking, Databricks Fundamentals Accreditation