

GLEICE CHAVES

Data Scientist | Applied AI

Los Angeles, CA | gchaves@usc.edu | [linkedin.com/in/gleiceschaves](https://www.linkedin.com/in/gleiceschaves) | U.S. Permanent Resident, no sponsorship needed

SUMMARY

Data scientist graduating from USC with an M.S. in Applied Data Science (Dec 2026), with industry experience in cybersecurity, healthcare, and finance and skills in Python, SQL, machine learning, and LLM evaluation. At PlayStation, built an LLM-assisted entity-matching pipeline for 186 organizations; at Keck Medicine of USC, built semantic models and machine learning forecasting, and raised Copilot answer accuracy **35%**. Two-time hackathon winner, ready to build models teams trust, full-time from January 2027.

EDUCATION

University of Southern California M.S., Applied Data Science	Aug 2024 – Dec 2026 (expected)
University of California, Irvine Certificate Program, Data Science, CS/Math	Jan 2023 – Aug 2024
FIAP (Brazil) B.S., Business Analytics	Jan 2019 – Dec 2022

EXPERIENCE

Sony Interactive Entertainment (PlayStation) <i>Data Science Intern, Global Vulnerability Management</i>	San Diego, CA May 2026 – Present
- Built an entity-matching framework that combines AI agents, probabilistic scoring, and deterministic rules to resolve machine identities across security tools (Qualys, CrowdStrike, Wiz) - Designed it as an hourly Snowflake pipeline sized for 800M+ records per run, giving 186 organizations one stable ID per machine and an 80-column asset inventory - Tested the LLM matching step like a classifier and reached 95%+ accuracy on calibration (two models vote, a third breaks ties) - Made every run auditable and reversible with 278 versioned rules, 56 data quality gates, SCD2 lineage, and anomaly alerts	
Keck Medicine of USC <i>Data Science Intern</i>	Los Angeles, CA Mar 2025 – Present
- Co-designed a time-series forecast of ICU bed capacity so a pilot team can plan beds and speed up discharges - Built a no-show risk model (logistic regression, XGBoost) and tuned its thresholds so clinics know when to double-book - Designed an automated AWS machine learning pipeline with leakage-safe features and per-clinic model calibration	
<i>Analytics Engineer Intern</i>	Mar 2025 – Apr 2026
- Raised Power BI Copilot answer accuracy 35% by building a 100+ question golden set to test and tune it - Co-designed one shared Power BI semantic model (SSAS Tabular, 150+ DAX measures) that replaced legacy SSRS finance reports	
USC Institute for Creative Technologies <i>Research Assistant</i>	Los Angeles, CA Mar 2025 – May 2026
Studied why people over-trust or under-trust AI-written code and developed the Trust Orbit Model of six trust cues	
Diagnostics of America <i>Analytics Engineer</i>	São Paulo, Brazil Jun 2021 – Dec 2022
- Merged scattered finance and operations data into one GCP database (Alteryx, Databricks) for 700+ clinics and labs - Rebuilt company-wide FP&A reports on that database, cutting the monthly cycle from 4-6 days to ~5 hours	
Decathlon Brasil <i>Marketing Analytics Intern</i>	São Paulo, Brazil Dec 2019 – Dec 2020
- Raised ad ROI 29% by analyzing Instagram and Google Ads spend and moving budget to what worked - Built Google Analytics reports on traffic and conversions, and paired them with market research to guide digital strategy	

PROJECTS

DischargeAgent, Multi-Agent Discharge Coordination Winner, Keck Hackathon	Apr 2026
Built a six-agent LangGraph system (React, FastAPI, AWS Bedrock AgentCore) with a capacity-forecasting agent; targets 15%+ more early discharges and 25% fewer delays	
Document Intelligence for Keck Winner, Keck Hackathon	Nov 2025
Designed an eight-node LangGraph pipeline that turns scanned documents into structured data, keeping PHI on premises	

SKILLS

- Modeling & Statistics:** Python (pandas, NumPy), scikit-learn, XGBoost, time-series forecasting, anomaly detection, logistic regression, survival analysis, entity resolution, model calibration
- ML & AI:** Large language models (LLMs), LLM evaluation (golden sets, LLM-as-judge), RAG, hybrid retrieval (BM25, Pinecone), LangGraph, LangChain, agent orchestration, prompt engineering, MLflow, responsible AI (Fairlearn, bias checks)
- Data & Cloud:** SQL, dbt, data modeling, SCD2, data quality, Snowflake (Tasks, Streams, Time Travel, Cortex AI), Databricks, AWS (S3, Redshift, Lambda, SageMaker, Bedrock), GCP, Alteryx, Streamlit
- BI & Domains:** Power BI, DAX, SSAS Tabular, Microsoft Fabric, vulnerability management, threat intelligence, healthcare