HELEN ZILLAH SATCHWELL

Boulder, CO | 720-490-8811 | helensatchwell@gmail.com | LinkedIn

Professional Summary

Bilingual linguist and technical lead with an M.A. and a background in pedagogy, specializing in translating complex processes into structured documentation and operational guidelines for GenAI applications. As a Quality Lead at AWS Bedrock, I designed an audit system that moved a QA feedback quality baseline from 47% to 96.7%. I approach problems with a whiteboard mindset, mapping out the logic to ensure human-in-the-loop workflows remain accurate and consistent.

Technical Skills

- **Linguistic Data Science**: LLM evaluation, RAG evaluation, adversarial testing & red-teaming, guardrails analysis, framework design, and vendor calibration. Identifying ambiguity and intent recognition.
- **Analysis & Logic:** Multi-dimension evaluation models, Inter-Annotator Agreement (IAA) metrics (Cohen's Kappa, F1), statistical significance testing ($p < 0.05$), and sampling methodologies using statsmodels and sklearn.
- **Python**: Loading, transforming, and writing annotated datasets (JSON, CSV, TXT). Data preprocessing and postprocessing with Pandas and NumPy. NLP analysis with NLTK and spaCy. Statistical analysis with statsmodels and sklearn.
- **Tools** & Platforms: AWS (S3, SageMaker), SQL, JIRA, Excel (pivot tables, charts, data visualization), Wiki/Confluence, Appen.
- **Enablement**: Technical writing, SOP development, scripting for data transformation, annotation workflow design, and cross-functional stakeholder communication.

Professional Experience

Quality Lead (QL)- Machine Learning Data Linguist III
Amazon Web Services – Bedrock | Sept 2025 – Jan 2026 (Team-wide Restructure)

- Designed and owned the audit framework for the internal linguist QA team, writing Python scripts to load annotated datasets (JSON/CSV), automate performance metrics calculations, and write structured results to files for leadership review.
- Raised quality from 47% to 96.7% by designing a QA feedback guide and standardizing rubrics, achieved through cross-functional calibration sessions, coaching vendors and MLDL team, and collaborating with Data Science to refine guidelines.
- Ran inter-rater reliability analysis (Cohen's Kappa) via statsmodels) across 250k data points, flagging low-agreement batches and exporting metrics to CSV for leadership reporting.
- Built and executed RAG evaluation workflows, assessing 8 MLDLs across 360 questions with statistical significance testing ($p < 0.05$) across Legal, Sports, and Enterprise domains, enabling consistent quality validation at scale.
- Led red-teaming and guardrails analysis, analyzing complex bypass scenarios across six harm categories and quantifying a 67.2% guardrail bypass success rate, driving targeted model improvement recommendations.
- Monitored high-stakes safety dimensions including Harmlessness, maintaining high consensus (typically >0.9) through targeted coaching and deep dives into low-agreement batches.
- Engineered data pipelines in Python (Pandas) to load, transform, and partition annotated data across JSON/CSV/TXT formats based on metadata, writing processed outputs to structured files for text versus code-based ML workflows.
- Implemented automated vendor tracking system in Appen to monitor AHT and quality metrics across nine dimensions, maintaining 10–15% QA coverage while improving process efficiency through automated reporting.
- Authored technical reports for leadership, applying NLP preprocessing toolkits to surface recurring patterns in model failure and user intent recognition.

Machine Learning Data Linguist II & III
Amazon Web Services – Transcribe / Bedrock | May 2022 – Sept 2025

- Led training and onboarding for global vendor teams and internal linguist cohorts. Simplified technical guidelines into structured reference materials to reduce annotation ambiguity.
- Wrote Python scripts for generating performance reports, transforming annotated datasets, and identifying data trends across large-scale annotation workflows.
- Conducted targeted adversarial testing to detect model vulnerabilities and partnered with Data Science teams on workflow and guideline improvements.
- Led MLDL team in testing internal SageMaker UI for task processing, providing feedback on tool functionality and enabling improvements to the internal annotation platform.

Independent Contractor
QA, Content, and Learning Enablement | 2019 – Present

- Perform linguistic and functional QA for global product launches to verify cultural and technical accuracy. Currently contracted with micro1 as an AI Trainer.
- Developed activities and assessments for Study.com that reinforced core linguistic concepts.

Spanish Lecturer
University of Colorado Denver | 2015 – 2018

- Designed undergraduate Spanish courses and interactive online lessons, focusing on student comprehension and the real-world application of linguistic concepts.
- Taught Spanish I and II, breaking down complex grammar and partnering with faculty to create adaptive instructional materials.

Leadership and Community Impact

Secretary, Board of Directors (HOA)
Boulder, CO | 2022 – Present

- Provide administrative oversight for a multimillion-dollar annual budget for a 448-unit residential community.
- Revised community rules and regulations to remove ambiguity and ensure alignment with current state law.
- Designed a standardized enforcement procedure for fair and structured violation handling.
- Translate complex legal and financial decisions into clear policy updates and community newsletters.

Education

- M.A. Spanish Linguistics – University of Colorado Boulder
- B.A. Linguistics – University of Colorado Boulder
- B.A. English Pedagogy – Universidad Nacional de Andrés Bello (Santiago, Chile)

Languages

- English: Native; Spanish: Near Native; Portuguese: Professional working proficiency

References

- Mickael Dahlen: AWS Manager | (929) 2231-2204
- Elvira Magomedova: AWS Manager | elvira.magomedora@gmil.com
- Daisy Carlesso: AWS Manager | dccapuani@gmail.com
- Ariel Krawczyk: HOA Board President | ariel.powderhorn@gmail.com