



Ariya Palmer

Senior Life Sciences Knowledge Engineer

☎ (919) 555-7316 ✉ ariya.palmer@example.com

🌐 linkedin.com/example/ariyapalmer 📍 Raleigh, NC 27601

SUMMARY

Senior life sciences knowledge engineer with six years building production grade clinical and biomedical datasets for **AI enabled applications**. Graduate training in biomedical sciences supports practical work across clinical, regulatory, and scientific data environments. Translates complex requirements into **task taxonomies**, output schemas, controlled vocabularies, annotation guidance, **gold standard examples**, and measurable model behaviors. Leads fine tuning dataset design, statistical analysis, data cleaning, preprocessing, provenance, licensing, and **human in the loop evaluation**. Partners closely with scientific experts, data scientists, machine learning engineers, and data engineers across proof of concept and production delivery. Hands on experience includes Generative AI, LLMs, MCP based agentic workflows, AWS, Azure DevOps, CI/CD, MLOps, GxP documentation, and production monitoring. Brings disciplined collaboration and domain judgment that can support reliable knowledge assets and useful life sciences products.

EXPERIENCE

Senior Life Sciences Knowledge Engineer

Veridian Bioinformatics 📅 March 2024 - Present 📍 Raleigh, NC

Owns knowledge engineering for clinical evidence and regulatory intelligence products, guiding dataset architecture, evaluation, agentic prototypes, and auditable delivery with scientific and technical partners.

- Defined clinical task taxonomies, **output schemas**, vocabularies, and **gold standard examples** for fine tuning.
- Trained 18 subject matter expert graders through calibration exercises and traceable annotation guidance.
- Diagnosed LLM failure modes through targeted slices and **human in the loop evaluation** rounds.
- Translated domain feedback into prompt, retrieval, schema, and tool call changes with **data science** partners.
- Maintained provenance, licensing, GxP, regulatory, and **intellectual property** records for **auditable** datasets.
- Connected grounded language models with approved services through MCP enabled agentic prototypes.

Knowledge Engineer, Life Sciences AI

BluePeak Clinical Data 📅 July 2022 - February 2024 📍 Durham, NC

Developed clinical and medical content datasets and cloud workflows, connecting structured publishing sources, evaluation practices, and controlled delivery from experimentation through monitoring.

- Built supervised **fine tuning datasets** from structured publishing sources and curated clinical examples.
- Applied **statistical analysis**, cleaning, and **preprocessing** to large biomedical datasets with reproducible transformation records.
- Resolved ambiguous clinical concepts with scientific and regulatory experts through labeling schema reviews.
- Created evaluation sets covering routine cases, edge conditions, and known model failure patterns.
- Implemented S3, SageMaker, Bedrock, ECS, and **API Gateway** workflows for controlled AI experimentation.
- Added **Azure DevOps** CI/CD checks for data artifacts, model handoffs, and **production monitoring** traceability.

Biomedical Data Scientist

Carolina Translational Informatics Lab 📅 June 2020 - June 2022 📍 Chapel Hill, NC

STRENGTHS

- ♥ **Domain translation**
Turned clinical and regulatory questions into schemas and examples that gave technical teams clearer model targets.
- 🔍 **Evaluation judgment**
When models failed on edge cases, created focused slices and grader reviews that made next steps easier.
- 👥 **Scientific collaboration**
Worked through ambiguous labeling decisions with experts, and became a trusted partner during difficult reviews.
- 📁 **Audit discipline**
Kept provenance and licensing records connected to dataset decisions, giving reviewers a dependable evidence trail.
- 💡 **Practical innovation**
Built MCP prototypes with technical partners, turning promising agent ideas into testable life sciences workflows.

SKILLS

Life Sciences Data Clinical Data

Fine Tuning Data Generative AI

LLMs Agents MCP

Agentic Workflows

Annotation Guidelines

Task Taxonomy Output Schemas

Human in Loop Evaluation

[Statistical Analysis](#) [Data Cleaning](#)

[MLOps](#)

LANGUAGES

English Native

Spanish Intermediate

MY CAREER



● Senior Life Sciences Knowledge Engineer at Veridian Bioinformatics (2.5 Years)

● Knowledge Engineer, Life Sciences AI at BluePeak Clinical Data (1.6 Years)

● Biomedical Data Scientist at Carolina Translational Informatics Lab (2 Years)

Curated translational research data and early machine learning studies, partnering with investigators and technical staff on analysis ready structures, reproducibility, and scientific review.

- Curated **clinical** research datasets across laboratory, phenotype, and study metadata for translational analyses.
- Applied **reproducible** cleaning and statistical checks to missingness, coding, outliers, and source discrepancies.
- Developed biomedical classification **proofs of concept** with documented assumptions, evaluation methods, and limitations.
- Created structured publishing templates and metadata conventions that improved reuse across research teams.
- Converted investigator questions into data specifications, evaluation plans, and interpretable analytical deliverables.
- Versioned provenance, transformations, and model experiments for auditable scientific review.

PROJECTS

Clinical Evidence Agent Workflow 📅 2025

Built an MCP connected workflow that linked grounded language models with approved knowledge services for clinical evidence tasks, creating a monitored path from proof of concept to controlled evaluation.

Biomedical Dataset Quality Framework 📅 2023

Created reusable quality checks for cleaning, preprocessing, provenance, and evaluation across large biomedical datasets, helping analysts compare model iterations with consistent evidence.

LEADERSHIP & AWARDS

- Recognized by scientific and technical peers for improving evaluator calibration and model feedback quality.
- Selected to lead cross functional proofs of concept connecting life sciences knowledge assets with agentic workflows.
- Received leadership recognition for maintaining auditable dataset provenance across regulated AI development.

EDUCATION

Master of Science in Biomedical Sciences

University of North Carolina at Chapel Hill 🎓 GPA: 4.0 📅 2020 📍 Chapel Hill, NC

Coursework: Biomedical research, clinical data analysis, biostatistics, molecular biology

Bachelor of Science in Biology

North Carolina State University 🎓 GPA: 4.0 📅 2018 📍 Raleigh, NC

Coursework: Cell biology, genetics, biochemistry, research methods

CERTIFICATIONS

- AWS Certified Machine Learning Specialty 📅 2025
- Azure DevOps and CI/CD for Machine Learning 📅 2024

TECHNICAL SKILLS

- **Life Sciences Data:** Clinical data, Regulatory data, Biomedical datasets
- **Generative AI:** LLMs, Fine tuning, Model behavior
- **Dataset Design:** Input output pairs, Edge cases, Failure modes
- **Taxonomy and Schemas:** Task taxonomies, Output schemas, Controlled vocabularies
- **Annotation Systems:** Annotation guidelines, Labeling, Grader calibration
- **Evaluation Methods:** Human in the loop, Evaluation sets, Statistical analysis
- **Data Engineering:** Data cleaning, Preprocessing, Structured publishing
- **AWS Machine Learning:** S3, SageMaker, Bedrock
- **AWS Application Services:** ECS, API Gateway, Cloud deployment
- **DevOps and MLOps:** Azure DevOps, CI/CD, Production monitoring

- **Agentic Systems:** MCP, Tool calls, Agentic workflows
- **Governance and Compliance:** GxP, Dataset provenance, Licensing

PROFESSIONAL AFFILIATIONS

- Life Sciences Knowledge Engineering community contributor, sharing practices across scientific and technical teams.
- Biomedical data science collaborator, building relationships across research, analytics, and machine learning groups.

ADDITIONAL INFORMATION

Work Status : Authorized to work in United States. No sponsorship required.

REFERENCES

AVAILABLE ON REQUEST