

Megan Kuo

U.S. Citizen, No sponsorship required | Cupertino, CA
[personal website](#) | [linkedin](#) | [github](#) | mkuo22@wisc.edu | +1-650-289-8993

Education

University of Wisconsin–Madison

Bachelor of Science in Data Science

Madison, WI

Sept 2023 – May 2026 (Expected)

Relevant Coursework: Database Management, Big Data Systems, Data Science Programming, Machine Learning, Artificial Intelligence, Probability Theory, Linear Algebra, Time Series

Technical Skills

Languages: Python, SQL, Java, JavaScript, R, HTML/CSS

Backend: REST APIs (Flask), Linux, Docker, Git, CI/CD, schema validation, unit testing

Data Engineering: dbt, DuckDB, Airbyte, Spark, Snowflake, data modeling, ETL, OpenSearch/Elasticsearch

Cloud & DevOps: AWS (S3, DynamoDB), GCP, Docker, Linux, orchestration (Airflow-style DAGs), Kubernetes

Data Science: NumPy, Pandas, scikit-learn, experimentation/A-B testing, time-series analysis

Visualization & Tools: Streamlit, React, Matplotlib, ggplot2, dplyr, OpenCV

Projects

EDGAR Document Search – *Python, DuckDB, dbt, OpenSearch, Docker, GitHub Actions, Streamlit* Jan 2026

- Built an end-to-end ETL pipeline ingesting SEC EDGAR filings into a DuckDB warehouse and dbt models, improving data correctness by adding 54 automated data-quality tests.
- Enabled low-latency retrieval across 8,000+ documents by deploying an OpenSearch index and orchestrating scheduled refresh jobs (Airflow-style DAGs) via GitHub Actions.
- Improved answer traceability by shipping a RAG-style Streamlit UI that retrieves cited filing chunks via OpenSearch and optionally generates responses with Gemini.

AI-Powered Spelling/Essay Grading Web App – *Flask, OCR, AWS S3, DynamoDB* Aug 2025

- Built a backend service that ingests binary uploads (handwritten images/PDF scans), runs OCR + scoring, and outputs structured JSON for downstream analytics and reprocessing.
 - Reduced grading inconsistencies by implementing a Flask REST API with deterministic preprocessing and schema validation for OCR extraction and grading-rule execution.
 - Improved debugging throughput by computing error breakdowns and grading-quality metrics, enabling analytics views for OCR/labeling performance.
-

Work Experience

UW–Madison Computer Sciences

Undergraduate Teaching Assistant – CS320 (Data Science Programming II)

Madison, WI

Jan 2025 – Present

Python, Git, Data Structures, Basic ML, A/B Testing, Web Scraping, Data Visualization

- Supported 400+ students by debugging Python/SQL code and explaining core data-structure concepts, improving student success through systematic troubleshooting and edge-case reasoning.
 - Improved code quality and reproducibility by reviewing assignments, reinforcing clean abstractions, and teaching Git-based workflows and testing habits.
-

Research & Publication

phylobar: multiresolution compositional barplots in omics studies

May 2025 – Present

R, JavaScript, React, Python | Mentor: [Prof. Kris Sankaran](#) | [bioRxiv](#) | [github](#) | [website](#)

Accepted March 2026

- Built an open-source visualization package under Bioconductor review with documented APIs and reproducible examples, standardizing exploratory workflows for high-dimensional omics data.
- Improved reusability by designing modular data-processing utilities and plotting interfaces for compositional data and hierarchical metadata across multiple datasets.
- Enabled interactive exploration by implementing tree-driven drill-down + stacked-bar views and supporting standardized SVG exports for reproducible reporting.