

# Qiao-Yun (Katharina) Cheng

qiaoyc2@uw.edu • (206)823-6913 • <http://www.linkedin.com/in/qiaoyc>

## EDUCATION

### University of Washington

Seattle, WA

*Bachelor of Science in Informatics, Concentration: Data Science*

June 2026

Annual Dean's List (2022 – 2025)

Cumulative GPA: 3.81/ 4.0

**Relevant Coursework:** Data Programming | Databases & Data Modeling | Database Design & Management | Advanced Methods in Data Science | Data Structure & Algorithms | Introduction to Search & Recommendation Systems | Introduction to Machine Learning | Introduction to Artificial Intelligence | Text Mining | Client-Side Development | Server-Side Development | IOS Mobile Development | Product & Information Systems Management

## SKILLS

**Technical:** R, Python, Java, SQL, HTML, CSS, D3, JavaScript, Swift, Tableau, Figma, DevOps, Azure, MongoDB

**Languages:** Bilingual (Mandarin, English)

## WORK EXPERIENCE

### University of Washington

Seattle, WA

*Teaching Assistant*

Sep 2024 - Jun 2025

- Conduct lab sessions for INFO 103, facilitate discussions on social media history, privacy, consent, and free speech
- Grade assignments and provide feedback to students on their performance, ensuring clarity and accuracy
- Support students applying concepts to automation projects like Bluesky, Reddit, or Discord bot

### Svim

Seattle, WA

*UI/UX Designer*

Jun 2024 - Dec 2024

- Designed and launched Svim's first website, providing a user-friendly interface for global participants
- Acquired in-depth knowledge of the underlying logic and trading patterns for Web3 online trading platforms
- Able to independently complete tasks on time and deliver high-quality results consistently

## RESEARCH EXPERIENCE

### Social Media Algorithms (X Community Notes)

Seattle, WA

*Researcher*

Sep 2025 - now

- Analyzed large-scale Community Notes data by cleaning and joining tables to compare latent factor/intercept behavior
- Built an interpretation pipeline for latent representations using K-means clustering in 3-factor space and iterative experiments to balance interpretability and overfitting
- Implemented LLM-assisted cluster/factor annotation with controlled positive/negative examples and repeated runs to assess stability; validated labels via manual “smell-check” (73% clear, 87% partial match across sampled notes)
- Executed topic prediction locally and visualized topic- and cluster-based structure in interactive 3D (Plotly); evaluated confidence-threshold sensitivity to balance coverage and label reliability

## RELATIVE EXPERIENCE

### Study Buddy

Seattle, WA

*Project Lead & Full-Stack Developer*

Oct 2025 - Dec 2025

- Led a 4-person team and produced the system architecture + schema + endpoint documentation for a UW-only study-partner platform with verified login and profiles
- Built homepage + invitation workflow and implemented session middleware to resolve current user identity
- Co-built real-time chat (WebSockets + database-backed chat history) for matched users

### Human-AI Content Moderation Study

Seattle, WA

*Research Lead (Study Design & Survey Ops)*

Apr 2025 - Jun 2025

- Framed hypothesis comparing baseline vs. bias-aware AI; defined metrics for error-rate and subgroup-fairness deltas
- Designed a survey with 20+ operationalized variables; encoded responses into an analysis-ready schema.
- Synthesized 10+ peer-reviewed papers into measurable constructs and a power-aware sampling plan.
- Delivered a codebook + data model enabling rapid EDA and subgroup analyses for the final poster/readout.

### **GameLensDB (NoSQL Video Game Metadata & Trends)**

**Seattle, WA**

*Data Engineer (Web Scraping & MongoDB)*

May 2025 - Jun 2025

- Scraped MobyGames with Selenium + BeautifulSoup, exporting JSON and loading 1,100+ titles into MongoDB
- Standardized 8 key attributes across 1.1k+ MongoDB docs, improving filterability and downstream analyses
- Designed aggregation pipelines to surface genre/platform trends, developer/publisher activity, and score vs. business model, packaged into concise takeaways

### **Climate TRACE Greenhouse Gas Data Warehouse**

**Seattle, WA**

*Database Designer & ETL Developer*

Apr 2025 - May 2025

- Designed a star-schema dimensional model and loaded it into Azure to support fast OLAP analysis
- Built ETL to ingest/clean 54 country-level CSVs (70k+ sites), mapped ISO codes to regions, and automated bulk load
- Authored advanced SQL (window function, CUBE, top-k sector) for analytics to surface policy-ready insights
- Delivered dashboards/figures that highlighted persistent top emitters to inform mitigation focus areas

### **Weather Visualization Dashboard**

**Seattle, WA**

*Data Analyst*

Feb 2025 - Mar 2025

- Developed an interactive weather trends dashboard using D3.js and JavaScript, enabling users to compare daily or monthly average temperatures between global cities from 2014 to 2015.
- Integrated dropdown menus and responsive charts for dynamic data filtering and enhanced user interactivity.
- Cleaned and transformed raw historical weather data from Weather Underground

### **Evaluating Comment-Post Relatedness in the Climate Reddit Community**

**Seattle, WA**

*LLM Prompting Engineering (Data Science and Modeling)*

Nov 2024 - Dec 2024

- Implement a binary text classification model to evaluate the relatedness of comments to post titles in the Climate subreddit, using Logistic Regression, BERT embeddings, and LLM prompting
- Trained and fine-tuned large language models (chatgpt3.5) with both zero-shot and few-shot prompting to predict the relevance of comments in climate-related Reddit posts
- Crawled the subreddit and structured data around 600 Reddit posts, facilitated the training of models

### **Recipe Finder: Recipe Discovery and Recommendation Platform**

**Seattle, WA**

*Search and Recommendation System Designer & Front-End Developer*

Nov 2024 - Dec 2024

- Designed a platform with information retrieval capabilities, leveraging information crawled from the internet
- Utilize a mapping function to suggest related queries and improve user experience over time
- Integrated PyTerrier and the BM25 ranking model to ensure efficient and relevant recipe retrieval

### **Music Industry Analytics Project**

**Seattle, WA**

*Database Developer*

Jan 2024 - Mar 2024

- Developed an SQL database for analyzing data from music streaming platforms such as Spotify
- Designed schema and queries to extract insights on artist popularity and music trends
- Managed real-time data integration and queries in a simulated production environment

### **Opioid Overdose Research Project: WA Drug-related Deaths Analysis**

**Seattle, WA**

*Data Analyst*

Oct 2023 – Dec 2023

- Utilized data cleaning and merging techniques for comprehensive analyses
- Segmented data by various demographic factors to determine specific impacts on different communities
- Created visualizations including bar charts, line graphs, and heat maps, to communicate findings

## **LEADERSHIP EXPERIENCE**

**Chinese Entrepreneur Networking (CEN), Event Planning Director**

Oct 2022 - Mar 2025

- Developed and coordinated networking activities and competitions that more than 100 participants joined per year, aimed at career exploration and professional growth
- Created activities that focus on precise groups with deeper knowledge to explore and plan for a career