

Yueran (Hannah) Sun

Master's Student in Data Science · University of Washington

Seattle, WA — yuerans@uw.edu — sun770311.github.io — Google Scholar — LinkedIn

My research focuses on how AI can help people find, understand, and learn from complex information. I study AI-assisted tools for scientific reading and information accessibility, as well as machine learning methods for discovering patterns and relationships in large-scale cultural heritage collections. I evaluate these systems through computational analyses and user studies measuring outcomes such as comprehension and long-term retention.

Keywords: Human-Computer Interaction · Computational Humanities · Information Accessibility · Information Retrieval · Digital Libraries · AI Safety · Responsible AI · AI & Society

EDUCATION

University of Washington, Seattle, WA

M.S., Data Science

GPA: 4.0/4.0

Sep 2025 – Mar 2027 (Expected)

University of Michigan, Ann Arbor, MI

B.S., Data Science · Minor: Mathematics · *University Honors*

GPA: 3.71/4.0

Aug 2021 – May 2025

PUBLICATIONS

C = Conference S = In Submission

- [C.1] Kumar, A., Glazko, K., **Sun, Y.**, Harniss, M., Wang, L. L., & Mankoff, J. (2026). *Beyond Readability Metrics: Plain Language Priorities in Disability Advocacy Organizations*. In *Proceedings of the 2026 ACM Conference on Fairness, Accountability, and Transparency (FAccT '26)*, 7864–7892.
- [S.1] Buccalon, B., **Sun, Y.**, & Lee, B. C. G. (under review). *Viral Images: Identifying Reprintings within 1.5 Million Photographs in Chronicling America*. Submitted to *Computational Humanities Research 2027 (CHR 2027)*.

RESEARCH EXPERIENCE

University of Washington – Language Accessibility Research Lab

Research Collaborator | *Primary Collaborator: Anukriti Kumar* | *PI: Prof. Lucy Lu Wang*

Sep 2025 – Present

Seattle, WA

Productive Friction for Scientific Reading

- Designing a mixed-methods evaluation framework to investigate how AI-assisted productive-friction interfaces influence scientific paper comprehension and long-term retention.
- Reviewed prior HCI reading systems to develop the experimental design, hypotheses, and mixed-effects analysis plan for a three-condition longitudinal within-subject study.
- Developing immediate and delayed comprehension assessments with rubric-based scoring to evaluate surface, textbase, and situation-model learning outcomes across reading conditions.

AI-Assisted Plain-Language Adaptation

- Evaluated 33 pairs of original and plain-language documents using 28 readability and linguistic metrics, identifying gaps between metric-based evaluation and practitioner-defined accessibility standards.
- Analyzed practitioner evaluations of AI-generated plain-language documents to identify limitations in accuracy, information completeness, definitional support, and document structure.
- Proposed design principles for AI-assisted plain-language tools that center human oversight, transparency, and verification while preserving practitioner and community expertise.

University of Washington – Lab for Computing Cultural Heritage

Research Collaborator | *PI: Prof. Benjamin Charles Germain Lee*

Sep 2025 – Present

Seattle, WA

- Developed an unsupervised computer vision pipeline using CLIP embeddings and DBSCAN to identify clusters of reprinted photographs across 1.5 million historical newspaper images from the Newspaper Navigator dataset.
- Built an API-driven image retrieval and validation workflow to surface clustered photographs on demand, enabling interactive inspection without downloading the full image dataset.
- Applied zero-shot CLIP classification to quantify semantic distributions across image clusters.

Humanities Data Science Summer Institute (American Institute of Physics Team) Jun – Jul 2026

Graduate Research Assistant | *PI: Prof. Benjamin Charles Germain Lee*

Seattle, WA

- Mentored a team of four undergraduate researchers to improve discovery across archival collections held by the American Institute of Physics, including 1,700+ oral history interviews with physicists.

- Visualized relationships among physicists, laboratories, and institutions by transforming oral history transcripts into interactive knowledge graphs using GLiNER-based named entity recognition and relation extraction.
- Identified unnamed collaborators by connecting relationship-based references made by interviewees to external knowledge bases, strengthening attribution across historical scientific archives.

MobiDrop (Zhejiang) Co., Ltd.

Nov 2023 – May 2024

Undergraduate Research Assistant | *Mentor: Dr. Guantao Zheng*

Remote

- Pretrained scGPT, a Transformer-based foundation model for single-cell gene expression modeling, on a curated dataset of 300,000 human blood cells from the CELLxGENE repository.
- Automated GPU training with Bash scripts for job submission, checkpointing, and experiment tracking.

INDUSTRY EXPERIENCE

Ternary

Jun – Aug 2025

Software Engineer Intern, Product Delivery Team | *Supervisor: Kyle Rattet*

Remote

- Integrated Meta Prophet into the cloud cost prediction API to complement the existing linear regression baseline with time-series forecasting.
- Engineered backend integrations between Go and Python services with schema validation, automated testing, and CI pipelines; deployed Terraform-managed staging environments on Google Cloud.
- Enhanced frontend client dashboards with interactive forecasting visualizations, including prediction intervals and breakdowns of the highest-cost services.

University of Michigan Multidisciplinary Design Program (ProQuest Team)

Jan – Dec 2024

Student Developer | *Mentor: Prof. Sindhu Kutty* | *Industry Sponsors: John Dillon, Dan Hepp* Ann Arbor, MI

- Developed an automated document segmentation pipeline to extract article-level structure, including titles, bylines, body text, and reading order, improving indexing and search of historical newspapers.
- Built an MLOps pipeline integrating large language models, Gaussian mixture models, and computer vision for text classification, article boundary detection, and noise identification.
- Validated pipeline performance through A/B testing, reducing segmentation costs by 50% relative to manual workflows while increasing throughput to 12 pages per minute.

Pachira Information Technology (Hengqin) Co., Ltd.

May – Aug 2024

NLP Algorithm Intern | *Supervisor: Dr. Chenglong Ma*

Zhuhai, China

- Enhanced the RAG pipeline powering an in-car voice assistant for Toyota vehicles.
- Developed and deployed a warning-light query handling module through iterative testing in a vehicle simulator.
- Implemented multi-intent recognition and refusal logic to improve response accuracy and user safety.

TEACHING EXPERIENCE

University of Michigan - Department of Mathematics

Jan – May 2025

Grader, MATH 417: Matrix Algebra I | *Supervisor: Prof. Ruiyuan Chen*

Ann Arbor, MI

Co-developed rubrics with the instructor and graded weekly assignments for 30 students.

University of Michigan - Math Learning Center

Sep – Dec 2023

Undergraduate Tutor | *Supervisor: Irina Arakelian*

Ann Arbor, MI

Tutored students in one-on-one and small-group settings across calculus, linear algebra, and differential equations.

LEADERSHIP & SERVICE

Develop for Good

Remote

Program Lead

Mar – Aug 2026

Overseeing a portfolio of four interdisciplinary student volunteer teams, coordinating project execution, client communication, and mentor engagement across nonprofit software development initiatives.

Product Manager, *PainUSA Team*

Oct 2025 – Feb 2026

Led five undergraduate student volunteers to prototype a web application for PainUSA, a nonprofit founded by Stanford University physicians, featuring an interactive clinician lookup map and chronic pain resources.

SKILLS

Research Methods: Mixed-Methods Evaluation, Experimental Design, Statistical Modeling, NLP, Computer Vision, Information Retrieval

Programming: Python, R, SQL, Bash, JavaScript/TypeScript, Go, C/C++

ML & Data: PyTorch, scikit-learn, Pandas, NumPy, OpenCV, XGBoost, LightGBM, LangChain, Dask

Systems & Platforms: Git, Docker, GitHub Actions, Linux, GCP, AWS, Terraform, FastAPI, React, Node.js, MongoDB, IBM Db2