

Mackenzie Hanh Cramer

*Fun Fact: I am on in the top 1% of players on the 2048 global leaderboard and have crocheted 4 blankets
mackenzie.hanh@gmail.com | <https://mackenziehanhcramer.github.io/>*

EDUCATION

University of California, Berkeley

School of Information

Master of Information Management and Systems

Berkeley, California

August 2025 - Present

University of California, Berkeley

GPA 3.72

- BA in Cognitive Science, Honors Student
- BA in Data Science, Domain Emphasis in Cognition
- Semester Abroad Spring 2024 at Yonsei University in Seoul, South Korea

Berkeley, California

August 2021 - December 2024

WORK EXPERIENCE

Teaching Assistant, Grading and Assignments Lead

Course: Data, Inference, and Decisions

- Lead discussion sections of 30-90 students and hosted office hours
- Created assignments and helped make exams, posted resources on website
- Maintained and shared assignment resources across 3 Github repositories
- Assigned staff to grading and oversaw student's regrade requests

Berkeley, California

August 2025 - Present

Bancroft Library Student III

Under Lorna Kirwan, Bancroft Public Services

- Working circulation desk and registration desk to assist patrons
- Process and page materials for duplication and permissions requests
- Aid curators for pulling items for exhibits, deliver items to and from preservation for marking

Berkeley, California

May 2025 - Present

Emotect AI Intern

Under David Jiang, Mark Barger Elliot

- Constructed 4-stage annotation task and guide which was employed by the team for AI Model
- Design for optimizing model precision using Microsoft Excel, Google Sheets, Google Docs, and Microsoft Word Docs

San Francisco, California

June 2025 - July 2025

Bancroft Library Student II

Under Iris Donovan, Bancroft Public Services

- Worked with programs ALMA and AEON in tracking and identifying patron requests
- Leveraged Microsoft Excel spreadsheets and Microsoft Word for data entry and reports
- Handling materials in UC Berkeley archives - paging, barcoding, retrieving, reshelving, and organizing various different items
- Trained 5+ new employees and was responsible for scheduling personal work shifts

Berkeley, California

July 2022 - May 2025

Reader

Course: Natural Language Processing

- Creating homework and coding assignments in Python
- Communicated during office hours for students and offered feedback towards questions concerning Natural Language Processing (NLP), assignments, and course information
- Formulated organizational system using Google Suite, specifically Google Sheets

Berkeley, California

January 2025 - May 2025

Undergraduate Research Assistant

Under David Bamman, NLP Research

Berkeley, California

February 2023 - December 2024

- Applied NLP techniques in Python to analyze datasets, identify patterns and new insights within the fields of digital humanities and cultural analytics
- Annotated various datasets, scrutinized data guidelines to increase clarity and reproducibility
- Engineered visualizations and conducted analyses to monitor performance trends and identify patterns

UCEAP Intern

Under Hwanho Noh at Barun ICT

Seoul, Korea

March 2024 - June 2024

- Researched and pitched article ideas regarding data privacy and security
- 3 columns concerning subjects including AI chatbots modeled after deceased people, low-resource languages and lack of AI accessibility, and the privacy concerns of bioinformatics

Summer Data Analytics Intern

Under Stephen Zander at Evolent Health

Remote, United States

June 2023 - August 2023

- Pioneered prediction model for time between biopsy and cancer treatment using Python
- Inspected insurance and membership claims using SQL, extracted Time Series measurements
- Collaborated with stakeholders such as health care experts, data scientists, and business professionals

Data Analytics Intern

Institute for Youth in Policy

Remote, United States

January 2023 - July 2023

- Produced data analytics for the data-education division, with focuses on identifying and measuring polarity of different political topics and potential improvements on high school curriculums
- Visualized data from workshops using Python and JASP for comprehensive statistical analysis.
- Gave feedback on others' visualizations and suggested improvements for clarity and minimized bias

Data Science Mentee

Undergraduate Laboratory of Data Science

Berkeley, California

September 2021 - June 2022

- Aided in production of prediction model for number of COVID-19 cases in the US by sourcing data and clean data to extract trends
- Wrote, produced, and organized poster that we presented at poster convention at conclusion of the program and semester

PROJECTS

Honors Thesis: Analyzing Descriptions in Novels

Mackenzie Cramer

Berkeley, California

August 2024 - December 2024

- Extracted characters information from 5,360 novels with Named Entity Recognition (NER) and NLP
- Mapped characters' descriptions onto their roles to identify common descriptions in varying genres
- Produced 20+ page report alongside visualizations and trend explanations

Youtube Apology Analysis

Mackenzie Cramer, Sabrina Baur

Berkeley, California

August 2023 - December 2023

- Employed NLP techniques in order to analyze patterns that arise among Youtube Apology videos
- Used Python while employed NLTK, sequence alignment, coreference chains, and other techniques
- Sourced, collected, and cleaned dataset and produced visualizations and analyses of trends

PUBLICATIONS

Measuring the Stories in Contemporary Songs

David Bamman, Sabrina Baur, Mackenzie Hanh Cramer, Anna Ho, Thomas McEnaney

August 2025

https://osf.io/preprints/socarxiv/5ys9c_v1

Abstract: Lyric poetry--the poetry of song--is often defined in opposition to narrative. In this work, we examine this relationship by carrying out an empirical study to measure the degree of narrativity present in contemporary songs, using a dataset of popular (Billboard Hot 100) and prestigious (Grammy-nominated) songs spanning 1960-2024. While we might expect the 1960s (with ballad-driven

folk singers like Joan Baez, Bob Dylan and Simon & Garfunkel) to be a high-water mark for narrativity, we find the opposite: narrativity has been steadily increasing over this period, largely due to the rise of the strongly narrative genres of hip hop and rap. We also find that it is a marker of prestige for country music, with Grammy-award nominated "Best Country" songs displaying significantly higher narrativity rates than non-nominated songs from the same album.

Multimodal Conversation Structure Understanding

Kent K. Chang, Mackenzie Hanh Cramer, Anna Ho, Ti Ti Nguyen, Yilin Yuan, David Bamman

May 2025

<https://arxiv.org/pdf/2505.17536>

Abstract: Conversations are usually structured by roles -- who is speaking, who's being addressed, and who's listening -- and unfold in threads that break with changes in speaker floor or topical focus. While large language models (LLMs) have shown incredible capabilities in dialogue and reasoning, their ability to understand fine-grained conversational structure, especially in multi-modal, multi-party settings, remains underexplored. To address this gap, we introduce a suite of tasks focused on conversational role attribution (speaker, addressees, side-participants) and conversation threading (utterance linking and clustering), drawing on conversation analysis and sociolinguistics. To support those tasks, we present a human annotated dataset of 4,398 annotations for speakers and reply-to relationship, 5,755 addressees, and 3,142 side-participants.

Speak, Memory: An Archeology of Books Known to Chat-GPT/GPT-4

Kent K. Chang, Mackenzie Cramer, Sandeep Soni, David Bamman

October 2023

<https://aclanthology.org/2023.emnlp-main.453/>

Abstract: In this work, we carry out a data archaeology to infer books that are known to ChatGPT and GPT-4 using a name cloze membership inference query. We find that OpenAI models have memorized a wide collection of copyrighted materials, and that the degree of memorization is tied to the frequency with which passages of those books appear on the web. The ability of these models to memorize an unknown set of books complicates assessments of measurement validity for cultural analytics by contaminating test data; we show that models perform much better on memorized books than on non-memorized books for downstream tasks. We argue that this supports a case for open models whose training data is known.

VOLUNTEERING

Ethnic Studies Curriculum Advisor

With Ethnic Studies Course Curriculum

Santa Rosa, California

January 2021- March 2021

- Worked on committee focused on High School English Literature Ethnic Studies course, to be deployed by Santa Rosa City School District
- Pitched topics to teach, promoting texts and books to be taught to students
- Worked with teachers, other students, and aided in course development

SCHOLARSHIPS

Redwood Empire Chinese Association

Sonoma, California

March 2021

Awarded to graduating high school seniors of Asian descent for excellent academic performance and community service, with plans to pursue post-secondary education.

ADDITIONAL WORK

SMASH Panel

Hosted by Anoop Kaur and Maggie Melone-Echiburú

Berkeley, California

July 2024

Panel Member for SMASH (Summer Math And Science Honors) Academy at UC Berkeley Campus and discussed admissions, taking Computer Science courses, and advice for applying and networking