

Ian Duke

[Email](#) | [LinkedIn](#) | [GitHub](#) | San Francisco, CA

EDUCATION

Master of Science in Data Science - GPA: 3.9

University of San Francisco

July 2023 - June 2024

San Francisco, CA

Bachelor of Arts in Government and Psychology - GPA: 4.0

Cornell University

August 2018 - May 2020

Ithaca, NY

Cum Laude with Distinction in All Subjects

EXPERIENCE

Data Scientist, Racial and Economic Justice

October 2023 - September 2024

American Civil Liberties Union of Northern California

San Francisco, CA

- Automated processes to analyze unstructured video data from over 500 police traffic stops – employing natural language processing and machine learning to classify videos relevant to litigation (88% accuracy, 83% recall).
- Built a data pipeline that associates more than 3,500 computer aided dispatch (CAD) reports with corresponding body camera videos by integrating pretrained computer vision models, temporal analyses, and data storage methods.
- Utilized locally-deployed, private transformer models, including large language models, to generate over 1,000 transcripts and conduct in-depth analysis of sensitive data.
- Regularly created data visualizations for office staff and explained methods to non-technical legal professionals.

Investigative Specialist, Special Litigation

April 2021 - October 2022

The Public Defender Service for the District of Columbia

Washington, DC

- Directed investigations for post-conviction resentencing motions by collecting records, conducting more than 50 interviews per year, and developing reports analyzing qualitative data – resulting in release of multiple clients serving extreme sentences.
- Supported launch of a police accountability database by conducting exploratory data analysis on and cleaning approximately 50GB of data, preparing it for successful migration into a web application.
- Mentored 6 interns and consulted with staff attorneys and mitigation specialists on a weekly basis to formulate data collection plans.

Data Analyst and Thesis Researcher

March 2019 - May 2020

Cornell Social Perception and Intergroup Inequality Lab

Ithaca, NY

- Partnered with Dr. Amy Krosch in pioneering a reverse correlation experiment investigating the impact of crime-specific stereotypes on eyewitness processing and recall.
- Created and executed a comprehensive data collection methodology involving acquisition of responses from over 1,025 participants..

HONORS AND AWARDS

- **Winner of Annual USF Entrepreneurship with AI Competition (2024):** Won a competition among 90+ data science master's students presenting an end-to-end application tracking federal prison population rates and predicting daily risk of future lockdown.
- **Schmidt Family Foundation Fellowship (2024):** Received a competitive \$8,500 grant for conducting research at the intersection of human rights and data science
- **Einhorn Discovery Grant (2020):** Awarded \$2,000 to fund research conducted in Cornell's Social Perception and Intergroup Inequality Laboratory.
- **Society for Personality and Social Psychology's (SPSP) Diversity Undergraduate Registration Award (2020):** Granted free attendance to SPSP's annual conference and invited to present research on crime type stereotypicality.

PROJECTS

- **Body Camera Video and Computer Aided Dispatch Report Preprocessing Pipeline – [GitHub](#)**
Developed an innovative pipeline leveraging pretrained machine learning models to transcribe body camera footage, extract video timestamps, and accurately associate body camera files with unique computer aided dispatch reports.
- **Prison Operations Snapshot Application – [User Interface](#) | [GitHub](#)**
Designed and implemented an application to automatically scrape and store daily population and operational data from 122 federal prison websites. Applied time series methods and trained a custom random forest model to forecast population trends and assess future lockdown risk with 93% accuracy.
- **User-Friendly Text Embedding Manager – [Demo Video](#) | [GitHub](#)**
Engineered a Python-based Text Embedding Manager application leveraging pre-trained machine learning models, empowering ACLU's non-technical staff to conduct efficient analysis of large textual datasets.