Isabel Silva Corpus
Brooklyn, New York | isc36@cornell.edu | https://isabelsilvacorpus.github.io

EDUCATION

Cornell University Ann S. Bowers College of Computing and Information Science August 2024 - Present
Ph.D. in Information Science Ithaca, NY
Committee: Mor Naaman (chair), Allison Koenecke (co-chair), Angelina Wang

Yale University August 2017 - May 2021
B.S. in Statistics and Data Science with Distinction New Haven, CT
Honors: Distinction in Statistics and Data Science (2021)

RESEARCH AREAS

Human-AI Interaction, Socio-technical Audits, Algorithmic Fairness, Computational Social Science

SELECTED PUBLICATIONS

- Lex Beattie, **Isabel Corpus**, Lucy Lin, and Praveen Ravichandran. 2025. Understanding Sensitive Attribute Association Bias in Recommendation Embedding Algorithms. ACM Transactions on Recommender Systems https://doi.org/10.1145/3777548
- Navin Kumar, **Isabel Corpus**, Meher Hans, et al. "COVID-19 vaccine perceptions in the initial phases of US vaccine roll-out: an observational study on reddit." *BMC Public Health* 22, no. 1 (2022): 1-14. http://dx.doi.org/10.1186/s12889-022-12824-7

WORKING PAPERS

- **Isabel Corpus**, Eric Gilbert, Allison Koenecke, and Mor Naaman. "Introducing AI to an Online Petition Platform Changed Outputs but not Outcomes." arXiv preprint arXiv:2511.13949 (2025). https://arxiv.org/pdf/2511.13949

PROFESSIONAL EXPERIENCE

Spotify, **Data Scientist II** New York, NY Jul. 2021 – Dec. 2023

Algorithmic Responsibility

- Developed and applied methodology for disentangling social factors in high dimensional embeddings, producing forthcoming research and productionization of new metrics within Spotify to assess fairness in recommendation systems.
- Evaluated Spotify recommendation systems for representation and allocation bias. Audited interpretability tooling and methodology (e.g. SHAP, TFMA fairness indicators) using Spotify ML use cases, presented recommendations for best practices.
- Led a project to understand partisan podcasts. Applied SVM's and PCA to linearly separate high-dimensional representations of partisan podcasts, constructed a graph and created metrics to quantify diversity in listening patterns.

Machine Learning Platform

- Productionized and maintained data models using Google BigQuery and dbt for reproducible analyses.
- Prototyped ML models in coordination with teams across Spotify. Designed and led workshops attended by 300+ Spotify employees, teaching principles of machine learning and driving development of ML for novel use cases across Spotify.
- Developed strategy for ML governance at Spotify through interviewing engineers, liaising across legal and technical teams, and developing tracking for machine learning models.

INTERNSHIPS

Action Research Partners: Research Intern, Brooklyn, NY Summer 2019

- Evaluated national child welfare systems by writing SPSS and R scripts to identify trends in welfare negligence using datasets with records of tens of millions of individuals. Presented findings to members of the Administration for Children's Services.

InsideSchools: Data Analyst Intern, New York, NY Summer 2018

- Combined qualitative interviews and tours of schools with quantitative survey data to report on school metrics and write school reviews for InsideSchools.

Human Nature Lab: Technical Lead & Research Assistant, New Haven, CT Fall 2020 – Jan. 2021

- Created a novel Reddit dataset about COVID-19 vaccine hesitancy and modeled Reddit comments with LDA topic modeling. Organized and instrumented web scraping, data cleaning, topic modeling, data visualization in Python. Culminated in publication.

- Led technical project management through task delegation, timeline creation and organization, facilitating communication between technical and non-technical researchers.

STEM PERL Lab: Research Assistant, New Haven, CT Spring 2020
- Contributed to research on educational best practices through data cleaning and analysis of large data sets using R. Analyzed survey data and streamlined the data cleaning process for the lab by introducing data exploration techniques.

TEACHING EXPERIENCE

Yale Statistics & Data Science Dept: Undergraduate Learning Assistant & Course Coordinator Jan. 2019-Jan. 2021
- Managed 20 teaching assistants and course affairs for an introductory statistics class of 400+ students. Improved the efficiency of the course by delegating tasks, creating timeline for course affairs, and designing a course roadmap.
- Spent 10 hours per week grading problem sets and holding weekly office hours.
- Courses taught: S&DS 10X: Introduction to Statistics, S&DS 230: Data Exploration and Analysis, S&DS 365: Data Mining and Machine Learning

SERVICE

Broadening Participation
- Committee member: Broadening Participation in Information Science (Spring 2025)
- Organizer and panelist: Breaking into Research Panel with Cornell URMC x NSBE (Spring 2025)
- Volunteer reviewer for Cornell Student-Applicant Reading Program (SARP) (Fall 2024)
- Organizer: Spotify for Women in Tech Mentorship Program (80+ participants). Organized career development events and matched mentors to mentees.

Mentorship
- Latinas in Tech
- Spotify's early career program.

Reviewing
- Reviewer for ACM CHI 2026.
- Co-reviewer for European chapter of ACL 2023 SRW.

Undergraduate
- Member of Yale Statistics Student Advisory Committee
- Co-founder and co-captain of the women's club basketball team at Yale.

TALKS & POSTERS

- CODE@MIT (Fall 2025)
- Yale New Directions in Social Algorithms Research (Fall 2025)
- Cornell Artificial Intelligence, Policy, and Practice (AIPP) Initiative (Spring 2025)

AWARDS

Dean's Excellence and Hopper-Dean Fellowship (2024)
Cornell University Ann S. Bowers College of Computing and Information Science

SKILLS

- Python, R, dbt, Github, SQL
- Technical Coursework: Discrete Math, Probability Theory, Theory of Statistics, Data Mining and Machine Learning, Linear Algebra, Neural Networks and Language
- Socio-technical Coursework: The Structure of Information Networks, Engineering Societal Systems, Non-Ideal Algorithmic Fairness