

Abby Williams, PhD

New York, NY | abbylwms@outlook.com | 570-351-1233 | [LinkedIn](#) | [GitHub](#)

PROFILE

Senior data scientist with production deployment experience and a track record of shipping end-to-end data systems. Designed and deployed Meridian, a career intelligence platform trained on 75,000+ cross-sectional records using ensemble classification, SHAP explainability, and a custom AWS EC2 deployment serving 5,000+ active users with a 30% reduction in advising cycle time. Led a team of 10+ analysts, managed 5 large-scale survey programs annually, and published peer-reviewed methodology in the Journal of Experimental Political Science and Cambridge University Press. Brings deep domain expertise in predictive modeling, large-scale behavioral analysis, and workforce analytics.

TECHNICAL SKILLS

Languages: Python, R, SQL

ML / Modeling: scikit-learn, XGBoost, Random Forest, K-means, logistic regression, KNN, PCA, EFA, interrupted time series, SHAP explainability

NLP: TextBlob, VADER, sentiment scoring, open-text feature engineering

Data & Visualization: pandas, NumPy, matplotlib, Tableau, ArcGIS, Qualtrics, SPSS, STATA, Jupyter

Deployment: Streamlit, Docker, AWS (S3, EC2), GCP (BigQuery, AI Platform)

Methods: Survey design, longitudinal analysis, mixed methods, usability research, behavioral segmentation

WORK EXPERIENCE

New York University

New York, NY

Senior Data Scientist & Assistant Dean (Director-level, 10+ reports)

Sep 2022 - Present

- Designed and deployed GSAS Meridian end-to-end — owned feature engineering, model selection, validation, and AWS EC2 deployment for a career intelligence platform now serving 5,000+ users with a 30% reduction in advising cycle time.
- Reduced career advising cycle time 30% by replacing manual, inconsistent advising workflows with a model-driven persona classification system that scales across cohorts without staff intervention.
- Led NLP pipeline development on 3,000+ open-text survey responses using TextBlob and VADER sentiment scoring, mapped polarity patterns to ML persona clusters for mixed-methods validation, and surfaced actionable insights that drove a 12% improvement in satisfaction scores.
- Produced four validated behavioral career personas from KNN-based segmentation using RIASEC composites, satisfaction metrics, and motivational factors as feature inputs.
- Led interrupted time series analysis on a national doctoral survey (n=329) measuring pandemic impact on career aspirations, applying composite index creation and statistical weighting; published in Journal of Experimental Political Science with full replication package on Harvard Dataverse.
- Managed 5 large-scale annual survey programs end-to-end: instrument design, sampling strategy, data collection, cleaning, and automated reporting pipelines delivering insights to senior leadership and external funders.
- Built and led a team of 10+ analysts across applied ML, data engineering, and program analytics, setting technical standards, reviewing model outputs, and delivering AI initiatives across institutional workflows.
- Deployed a CV-to-Resume conversion tool (Streamlit/AWS) and a PhD Internships Database (Streamlit/AWS) serving a combined 15,000+ users annually, both built on structured and unstructured data pipelines.

Assistant Director, Fellowships & Data Analytics

Mar 2016 - Aug 2022

- Applied predictive analytics and targeted outreach to grow Fulbright award rate 200%, contributing to outcomes tied directly to a \$3M grant renewal; secured \$18M+ in total competitive funding across Mellon, NSF, and other major funders.
- Built automated data tracking infrastructure that cut fellowship processing time 50%, reducing manual data handling and improving reporting accuracy across concurrent award programs.
- Designed the office's first real-time analytics dashboard in Tableau, integrating Fulbright pipeline data for senior staff; first structured data visibility tool of its kind in the office.
- Built the Public Humanities Certificate program, inc \$2.5M initial grant, curriculum design, 4x enrollment growth, NYU top 10 nationally in Fulbright winners over 10 years.
- Led the Luce Foundation scholarship review as team leader, managing 4 reviewers and delivering finalist recommendations to Luce senior staff

Program Administrator

Jan 2015 - Feb 2016

- Ran user research with 150+ PhD users, applied findings to redesign service delivery, reducing in-person visits 75%.
- Streamlined dissertation submission for 300+ doctoral candidates annually; built budgeting systems improving financial oversight accuracy.
- Built a custom fellowship application portal in SurveyMonkey Apply, migrating the office from Google Forms to a fully automated end-to-end process, reducing manual handling and creating a structured, trackable applicant pipeline for the first time.

Handshake AI / Feather (OpenAI)

AI Training Specialist |

November 2025 - April 2026

- Contributed domain expertise in higher education and PhD career pathways to AI model training and evaluation for career development applications.

SELECTED PROJECTS

Meridian Career Intelligence Platform | New York University | 2024 - 2026

End-to-end ML system built on 75,000+ harmonized longitudinal survey records. Feature engineering from RIASEC composites, satisfaction metrics, and motivational factors. Model suite: K-means (silhouette=0.37), 3-class XGBoost (F1=0.58, precision=0.60, recall=0.57), Random Forest, logistic regression. SHAP explainability layer for advisor-facing model transparency. Streamlit dashboard deployed on AWS EC2. Serves 5,000+ students. 30% reduction in advising cycle time. Manuscript targeting Journal of Higher Education. <https://gsasmeridian.com/>

Sentiment-Augmented Persona Validation | New York University | 2024

NLP pipeline on 3,000+ open-text survey responses and 60+ interview transcripts using TextBlob and VADER. Mapped polarity and subjectivity scores to ML persona clusters for mixed-methods validation. Built automated stakeholder dashboards linking sentiment trends to personas and program interventions.

COVID-19 PhD Career Impact Study | NYU / Harvard Dataverse | 2020 - 2022

Interrupted time series analysis on national survey data (n=329) measuring pandemic-driven shifts in doctoral career aspirations. Methods: composite index creation, statistical weighting, longitudinal panel design. Published Journal of Experimental Political Science (2022). Full replication package on Harvard Dataverse. doi:10.1017/XPS.2020.34

NSF/Mellon Career Pathways Dashboard | New York University | 2023 - 2024

Integrated ML persona predictions with demographic and geospatial data across Tableau, ArcGIS, and GCP BigQuery. Visualizations directly informed \$1.2M+ in program funding allocation decisions.

EDUCATION

PhD, Higher Education (Machine Learning, UX/UI) | New York University | New York, NY | May 2026

Dissertation: "Designing PhD Career Personas Using Cluster Analysis"

MA, International Education | New York University | New York, NY | Dec 2016

BA, Chinese and French | University of Pittsburgh | Pittsburgh, PA | May 2011

PUBLICATIONS

- Williams, A. & Williams, F. (2024). How to Thrive Beyond Academia: Addressing Five Common Challenges for Humanities PhDs. Public Humanities (Cambridge University Press).
- Haas, N., Gureghian, A., Jusino Díaz, C., & Williams, A. (2022). Through Their Own Eyes: The Implications of COVID-19 for PhD Students. Journal of Experimental Political Science, 9(1), 1-21. doi:10.1017/XPS.2020.34
- Presenter, Modern Language Association (MLA) Annual Convention - PhD career development and workforce analytics.
- Panelist, Chronicle of Higher Education - PhD career pathways and academic transitions (with Stacy Hartman)

