Jeremy Springman, PhD
Research Assistant Professor
University of Pennsylvania
jrspringman.github.io

I am an applied researcher passionate about creating large-scale data solutions to inform strategy and decision-making. As a Research Assistant Professor running major research projects in a social science research lab, I have a decade of experience with cutting-edge data science and six years of experience overseeing data governance and management of large-scale, sensitive data in a university research lab.

As an applied researcher, I have been extremely successful at winning highly competitive research grants. However, recent terminations of major federal research initiatives have cut-off opportunities for high-impact collaboration between external researchers and government agencies. I am now looking to transfer my technical and management skills to informing strategy and innovation within a university setting. I have set down roots in Philadelphia and planned for it to be a permanent home. Princeton's campus is a short commute from my home. Joining Princeton would allow me to keep my family in place while serving in a role that is well-suited to my interests and skills.

I have spent the last decade using causal inference and machine learning to inform high-stakes decision-making by policymakers in the U.S. government. From designing and analyzing complex experimental studies to delivering rigorous, data-driven insights that inform strategic decisions, I have cultivated the skills and experience required to help Haus's customers answer mission-critical questions and expand Haus's product suite.

Over the past six years, I have managed and mentored an interdisciplinary team of twelve researchers and data scientists. Under my leadership, this team has planned and executed many large-scale projects, ensuring that the findings are both rigorous and actionable. Since 2022, I have been awarded nearly $3.5 million as a principal investigator on both federal and private research grants, carrying projects from conceptualization and proposal through design, implementation, and the dissemination of results to the scientific and policy communities. My most recent awards include a $1.6m Department of Defense Minerva Research Initiative grant to support interdisciplinary research to model the impact of climate change on adaptation, conflict, and migration and a $900,000 Agency for International Development grant to maintain and expand a large-scale webscraping and machine learning text classification pipeline to monitor and forecast political instability in more than 60 countries.

Working at a university has also given the opportunity to teach and mentor students at the undergraduate and graduate level. My academic career has taught me the importance of mentorship, and I plan to continue fostering data science best practices by guiding junior colleagues and building a collaborative culture that values creativity and rigor.

My research has required close attention data governance practices and policies. Since 2019, I have led the construction and maintenance of a large, highly multilingual corpus of high-quality online news from developing countries. This corpus includes more than 100 million articles published in more than 40 languages and stored in a document-oriented NoSQL database program on university servers. Management of this novel media corpus has required establishing security procedures for remote access and automated back-ups with university IT administrators, determining data access policies for permanent project staff, temporary undergraduate and graduate research fellows, and faculty affiliates within and outside our university. As this data was collected from public sources

but is unlicensed, research with this data has required setting careful policies for collaboration and data sharing, often requiring consultation with university lawyers on international Intellectual Property law.

I have also managed the collection and usage of sensitive data from both individuals and U.S. government partners in developing countries, including activists and NGOs in highly repressive countries. I have been responsible for securing approval from university ethical review boards and determining access for datasets including personal identifiable data, sensitive financial documents and records, and long-form open-ended responses on topics that could expose respondents to significant risk.

As a researcher, I have deep experience leading data science projects commissioned by stakeholders to generate actionable insights. These projects have given me years of experience using advanced methods to address novel design and measurement challenges and translating scientific research into action. I have given invited talks to hundreds of policymakers across more than a dozen agencies in five countries. My experience working closely with policymakers and stakeholders prepares me to collaborate with Princeton's Data Strategy & Innovation team to build data-driven decision-making pipelines.

I have technical proficiency in statistical programming (R and Python), document preparation and visualization (markdown, Quarto, and Shiny), version control (git, GitHub, and renv), programming online surveys (Javascript and html), and have managed teams using high-performance computing environments, NoSQL document-stage databases, and core machine learning Python libraries (scikit-learn, XGBoost, LightGBM, PyTorch, Optuna).

As a post-doctoral research associate, I co-founded a high-profile project to build a public-facing early warning system forecasting political instability in emerging markets. Managing a team of data scientists and PhD students, I led the construction of a research infrastructure to continually update a large, highly multilingual corpus of high-quality media and use deep learning translation models and fine-tuned transformers and LLMs to extract information from text. This project's funding was renewed over six years and attracted support from the U.S. Department of Defense, Agency for International Development, and private donors. My lab used data from this project publish more than a dozen commissioned policy reports, develop public-facing data dashboards averaging 150 hours of monthly active usage time by decision-makers, and attracted 1,200 unique users from 70 countries over a 10 month period, with more than 800 policymakers signing up to receive our monthly reports.

In addition to generating academic publications and contributing to our theoretical understanding of the social world, my research has directly informed policy. As a research associate at Duke University, I led an evaluation that provided rigorous evidence that a U.S. government foreign aid program was failing to accomplish it's goals, resulting in the program's cancellation and a more effective use of government resources. In another project, I worked with policymakers at the U.S. Agency for International Development to design a conjoint survey experiment that answered pressing questions about the strategic behavior of non-profits receiving U.S. support in highly repressive countries. Based on the findings, I produced a policy brief recommending changes to how the agency supports local partners in repressive countries and presented the recommendations to high-level officials. A third project used insights from behavioral psychology and economics to design and evaluate an intervention to increase civic engagement and social cohesion among youth in highly polarized countries. This resulted in a highly effective intervention that my partners are seeking to scale across developing countries.