TRAVIS AARON HOPPE

Condensed Resume (Federal CV available)
travis.hoppe@gmail.com
PhD Physics
(775) 287-4033

PROFESSIONAL EXPERIENCE

Associate Director of Data Science and Analytics
CDC / National Center for Health Statistics (NCHS)
February 2022 - Present

- Developed and led three organizations within the agency: NCHS Innovation, NCHS Data Science Community of Practice, and the EDAV Best Practices group to build community, foster communication, and develop innovative practices across CDC.
- Developed the first set of model standards for AI/ML within the agency for trustworthy, responsible, and ethical usage. Aligned with existing standards including the NIST AI RMF and HHS Trustworthy AI.
- Harmonized metadata usage for all NCHS datasets published on data.cdc.gov. Developed and implemented standards for tagging datasets around Social Determinants of Health (SDoH).
- Aligned NCHS with the agency data modernization goal of cloud integration. Coordinated policy, governance, and legal issues around the usage of CIPSEA protected data. Lead Tiger Team to identify gaps in existing cloud infrastructure for AI/ML needs.
- Piloted many innovative data science projects around the Center including homomorphic encryption, text-to-speech transcription, natural language processing, and bibliometrics.

Chief Data Scientist
CDC / National Center for Health Statistics (NCHS)
September 2021 - February 2022

- Lead for NCHS Data Modernization Initiative: *Increase Use, Discoverability, and Access to NCHS Data*. Established pilot projects and worked with stakeholders to create statements of work and business needs.
- Served on CDC's response to the Executive Order on AI and the NCHS Data Science Strategic Plan. Presented to Board of Scientific Council, NCHS All-Hands, and delivered subject matter talks on Bias in AI, advances in Natural Language Processing, bibliometrics, and more.
- Organized Center-wide metadata standards and built a custom ontology using evidence based sources: publications, web searches, and market research.
- Developed new methodology to study free text responses from the Research and Development Survey (RANDS), including non-response detection and zero-shot learning objectives.
- Implemented PII detection processes for restricted microlevel data.

Senior Data Scientist / Portfolio Analyst
NIH / DPCPSI / OPA contracted under Lexical Intelligence
February 2016 - February 2020

- Scientific team leader for a novel inter-agency government blockchain to detect grant duplication with minimal shared data. Coordinated research, oversaw design, and developed protocols within the NIH and NSF teams.
- Developed new analytic tools to process the text of NIH grants and publications using distributional embeddings (word2vec) and transformers (BERT). Tools were deployed for analysis presented to NIH senior leadership, Congress, and publications in high-ranking journals.
- Architected and productionized machine learning models for classification, regression, outlier detection, and language modeling. Creator and maintainer of several open-source tools used internationally in the scientific community.
- Analyzed grant and publication portfolios, evaluating metrics such as clinical impact, technological impact, and award rates to build quantitative comparisons between various populations.

Postdoctoral Fellowship (IRTA) at National Institutes of Health
Research Scientist, April 2014 - February 2016

Postdoctoral Fellowship at National Institutes of Health
Research Scientist, August 2011 - April 2014

COMMITTEES SERVED

2021-Present **National Science and Technology Council (NSTC) subcommittee on AI/ML,** CDC representative

2021-Present **Federal Committee on Statistical Methodology (FCSM),** Board Member

2021 **National Science and Technology Council (NSTC): Epidemic Modeling and Forecasting Fast Track Action Committee (FTAC),** Contributor: Plan to Advance Data Innovation

2021 **Health and Human Services (HHS): Open Data Task Force,** Committee member

SELECTED PUBLICATIONS

Policy

2023 **Dark citations to Federal resources and their contribution public health**, *Jessica Keralis, Juan Albertorio-Díaz, & Travis Hoppe*, Frontiers in Research Metrics and Analytics

2023 **Application of a Novel Machine Learning Technique in a Bibliometric Analysis of Health Disparities Articles**, *Pascaline Ezouah, Bao-Ping Zhu, & Travis Hoppe*, Manuscript in preparation

2019 **Topic Choice Contributes to Lower Rate of NIH Awards to African-American/Black Scientists**, *Travis Hoppe, Aviva Litovitz, Kristine Willis, Rebecca Meseroll, Matthew Perkins, B. Ian Hutchins, Alison Davis, Michael Lauer, Hannah Valantine, James Anderson, & George Santangelo*, Science Advances

Data Science

2023 **Model Release: Semi-Automated Nonresponse Detection for Surveys (SANDS)**, *Kristen Cibelli Hibben, Zachary Smith, Ben Rogers, Valerie Ryan, Travis Hoppe*, Hugging Face

2023 **Semi-Automated Nonresponse Detection for Open-text Survey Data**, *Kristen Cibelli Hibben, Zachary Smith, Ben Rogers, Valerie Ryan, Travis Hoppe*, Under review: Social Science Computer Review

2023 **Prediction of transformative breakthroughs in biomedical research**, *Matthew T. Davis, B. Ian Hutchins, Brad Busse, Payam Meyer, Grant Jones, Travis A. Hoppe, Kristine A. Willis, Abbey Zuehlke, Rebecca A. Meseroll, and George M. Santangelo*, Manuscript in preparation

2022 **Predicting causal citations without full text**, *Travis Hoppe, Salsabil Arabi, Ian Hutchins*, bioRxiv

2020 **The Pile: An 800GB Dataset of Diverse Text for Language Modeling**, *Leo Gao, Stella Biderman, Travis Hoppe, et al.*, arXiv

RECENT CONFERENCES

2023 **Health Datapalooza,** Panelist: Federal Health, Responsible & Trustworthy AI: From Principles to Practice

2023 **Conference on Statistical Practice,** Speaker: Examination of Dark Citations of Federal Information and their Contribution to Research

2022 **General Services Administration AI Community of Practice,** Invited Speaker: Detecting Non-response and PII in Web-based Surveys

2021 **Federal Privacy Summit: Washington DC,** Panel Chair: Automated PII detection

2021 **Federal Committee on Statistical Methodology: Washington DC,** Presentation: Short communication as a medium: Is Engagement a substitute for efficacy?

2021 **Federal Committee on Statistical Methodology: Washington DC,** Presentation coauthor: Analysis of Open-text Time Reference Web Probes on a COVID-19 Survey

AWARDS

2022 **NCHS Director's Office Individual Merit Award,** Data modernization and workforce coordination around data science

2022 **NCHS Director's Office Group Merit Award,** Release of two trillion accelerometer data points to the public from the Physical Monitoring Team

2017 **NIH Office of the Director's Honor Award,** Outstanding support for the Grants Support Index & Next Generation Research Initiative Analytical Team