TRAVIS AARON HOPPE

AI Leader, Curriculum Vitae

travis.hoppe@gmail.com

PhD Physics

(775) 287-4033

HIGHLIGHTS

- Federal expert in artificial intelligence (AI), machine learning (ML), and data science. Leader and innovator at the White House, coordinating policy agendas around research and development with partners in the National Science Foundation and the President's Council of Advisors on Science and Technology.

- Led multiple teams across CDC to develop and implement cutting-edge policies, guidance, and strategies for AI/ML adoption, including conversational AI (Chat GPT), and infrastructure needs to support AI/ML implementation. Extensive experience in presenting to federal and industry venues on topics around Trustworthy AI, implementation, and development.

- Respected and well cited author in AI/ML and biomedical research to address questions on multiple topics including: disparities in NIH funding (Topic Choice: 400 citations, key evidence used by the Congressional Black Caucus), open citations (NIH Open Citation Collection: 70 citations), and a fundamental AI dataset for training large language models (The Pile: 390 citations).

PROFESSIONAL EXPERIENCE

Assistant Director of AI Research and Development
White House Office of Science and Technology Policy (OSTP)
February 2024 - January 2025 (detailed from CDC)

- **Set national vision for AI R&D** for major White House initiatives including the National AI Research Resource (NAIRR), the National Secure Data Service (NSDS), the National Science and Technology Council (NSTC) subcommittee AI and ML, and the OMB Memorandum (M-24-10), "Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence"

- **Built coalitions** across the Federal government by connecting State, Tribal, Local, and Territorial governments through many engagements via the White House Tech Policy Network, including the National League of Cities, the American Public Human Services Association, and the Council of State and Territorial Epidemiologists

- **Managed, prioritized, and directed resources** for projects within agencies including the Federal AI use-case inventory, the Networking and Information Technology Research and Development (NITRD) AI program repository, and the Chief AI Officer's Council working groups.

Associate Director for Data Analytics and Data Science (ADDADS)
Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS)
December 2023 - Present (on detail to OSTP)
ST 1560

Senior Service Fellow
May 2022 - December 2023
RG GS-15/2 1530, 40 Hours/Week

- **Led organizational change** across CDC by bringing generative AI to all staff, the first Federal agency to do so. Accomplished results by leading 15 pilots programs across the Centers, coordinating financial, human, and IT resources, cleaning the program through cybersecurity, legal, and labor concerns.

- **Modernized** NCHS data systems through cloud adoption through a multi-year plan. Aligned and coordinated the plan with CDC's data modernization initiative (DMI) to help build the foundation for data sharing across all levels of public health. Coordinated policy, governance, and legal issues around the usage of Confidential Information Protection and Statistical Efficiency Act (CIPSEA) protected data. Worked with Office of Management and Budget (OMB) to adopt CIPSEA guidance and lead NCHS as an early cloud adopter for statistical agencies.

- **Piloted** innovative AI, ML, and data science projects to help identity, analyze, and report on emerging public health data. Projects include an item non-response detection model for survey text which resulted in the first AI model released by NCHS, privacy enhancing technologies (PETs) like homomorphic encryption and privacy preserving record linkage, text-to-speech transcription (which resulted in a 10-fold improvement from prior methods), and the creation of a bibliometrics dashboard for reporting.

- **Built community** by starting and leading three different inclusive organizations within the agency: NCHS Innovation, NCHS Data Science Community of Practice, and the agency-wide EDAV Best Practices group. Organizations helped build community, foster communication, and spur innovative practices across CDC.

- **Implemented** 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 and the organizational of CDC and NCHS.

- **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).

- **Served** as an agency representative for National Science and Technology Council (NSTC) subcommittee on AI/ML and regular presenter at HHS AI and GSA AI communities of practice. Active committee member of the Federal Committee on Statistical Methodology (FCSM) co-leading a Federal level metadata initiative.

Senior Service Fellow: Chief Data Scientist
Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS)
September 2020 - May 2022
Title 42 (GS-14/5 equivalent), 40 Hours/Week

- **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.

- **Advised** 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.

- **Developed and coordinated** Center-wide metadata standards and built a custom ontology using evidence based sources: publications, web searches, and market research.

- **Researched** 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 to allow privacy preserving research and development.

Senior Data Scientist / Portfolio Analyst
National Institutes of Health (NIH) / Division of Program Coordination, Planning, and Strategic Initiatives (DPCPSI) / Office of Portfolio Analysis (OPA) contracted under Lexical Intelligence
February 2016 - February 2020
40 Hours/Week

- 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 National Science Foundation (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.

- Trained and mentored junior staff in natural language processing (NLP) and machine learning.

- Analyzed grant and publication portfolios, evaluating metrics such as clinical impact, technological impact, and award rates to build quantitative comparisons between various populations.

- Restored historical texts from books and generated new structured data from free text. Expanded NIH grant coverage by thirty years from archival documents. Cross-linked publications to an NIH application's biographical sketch and literature cited. Data used internally with the NIH for analysis on racial disparity, topic analysis, mentorship, and grant efficacy.

Postdoctoral Fellowship (IRTA) at National Institutes of Health
Research Scientist

April 2014 - February 2016
40 Hours/Week

- Researched novel integration schemes for molecular dynamics simulations (MDS). Developed protein models for tertiary structure prediction from primary sequence.

- Designed and managed high-performance computing models on the NIH supercomputer, Biowulf. First to investigate containerized solutions for MDS using a graphics processing card.

- Worked in collaboration with experimentalists to test and validate models.

- Mentored post-baccalaureate researchers, guiding them in their research and professional development.

Postdoctoral Fellowship at National Institutes of Health
Research Scientist

August 2011 - April 2014
40 Hours/Week

- Developed multi-scale theoretical and computational models to study protein folding, structure, and protein-protein. Derived hard-sphere models to account for crowding in biomolecular simulations and potentials to model anisotropic charge distributions.

- Managed large-scale parallel projects (1000+ cores) to simulate the cellular environment.

EDUCATION

2011	**Doctor of Philosophy, Physics** Drexel University *On the Role of Entropy in the Protein Folding Process*, Thesis
2008	**Master of Science, Physics** Drexel University
2005	**Bachelor of Science, Physics** University of Nevada
2005	**Bachelor of Science, Mathematics** University of Nevada

COMMITTEES SERVED

2021-Present	**National Science and Technology Council (NSTC) subcommittee on AI/ML,** Co-chair (2024-present), CDC representative (2021-2024)
2021-Present	**Federal Committee on Statistical Methodology (FCSM),** Board Member
2022-2024	**Building Trust and FAIRness into the Process for Finding and Using Government Data (Chief Data Officers Council and FCSM),** Co-lead
2023	**Utility and Risks to CDC of Conversational Artificial Intelligence (AI) Technologies like Chat GPT (CDC),** Tiger Team Lead
2023	**Supporting the U.S. Public Health Workforce, President's Council of Advisors on Science and Technology (PCAST),** External Expert
2022	**Analytics and Machine Learning Implementation within CDC's Cloud Environment (CDC),** Tiger Team Lead
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

2024	**Semi-Automated Nonresponse Detection for Open-text Survey Data**, *Kristen Cibelli Hibben, Travis Hoppe, et al.*, Social Science Computer Review
2023	**Predicting causal citations without full text**, *Travis Hoppe, et al.*, Proceedings of the National Academy of Sciences of the United States of America
2023	**A Framework for Data Quality: Case Studies**, *Lisa Mirel, Travis Hoppe, et al.*, Federal Committee on Statistical Methodology
2023	**Dark citations to Federal resources and their contribution public health**, *Jessica Keralis, Travis Hoppe, et al.*, Frontiers in Research Metrics and Analytics
2019	**Topic Choice Contributes to Lower Rate of NIH Awards to African-American/Black Scientists**, *Travis Hoppe, et al.*, Science Advances
2020	**The Pile: An 800GB Dataset of Diverse Text for Language Modeling**, *Leo Gao, Travis Hoppe, et al.*, arXiv
2019	**The NIH Open Citation Collection: A public access, broad coverage resource**, *Ian Hutchins, Travis Hoppe, et al.*, PLoS Biology

AWARDS, CONFERENCES, & PANELS

Available upon request and online Federal Curriculum Vitae.