TRAVIS AARON HOPPE

Federal Resume

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
40 hours per week
GS-15/5 $153,434/year base compensation with high-performance bonus

- 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
40 hours per week
Title 42 (GS-14/5 equivalent) $138,866/year base compensation with high-performance bonus

- 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
40 hours per week
$135,000/year base compensation, yearly bonus of $4,000

- 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.

- 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 per week
$48,000/year

- 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.

Postdoctoral Fellowship at National Institutes of Health
Research Scientist

August 2011 - April 2014
40 hours per week
$46,167/year

- 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.

Teaching Assistantship / Curriculum Designer
September 2005 - May 2011
35 hours per week
$28,000/year
Teaching Assistant (Drexel)

- Organized, taught, and ran 22 undergraduate courses.

- Personally restructured the entire computational component for physics majors by transitioning from FORTRAN to Python.

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,** 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
2021	**CDC: Information Technology and Data Governance (ITDG),** Committee member

Machine learning / AI: Natural Language Processing (NLP), Convolutional Neural Networks (CNN), Generative Adversarial Networks (GANs), Transformers (BERT, GPT), pyTorch, Tensorflow, word2vec

Programming & Database: Python, C++, JavaScript, SQL, NoSQL (MongoDB, Elasticsearch), Cloud Infrastructure (AWS, Azure, Heroku), Containerization (Docker, Kubernetes)

Project Management: Experienced team leader for analysis, code, and design. Trained and mentored staff.

Public Trust Clearance: Level 5, valid from 2018-2023

PUBLICATIONS

Policy

2023 **Dark citations to Federal resources and their contribution public health**, *Jessica Keralis, Juan Albertorio-Díaz, & Travis Hoppe*, Manuscript in preparation for the Proceedings of the National Academy of Sciences

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

2019 **The NIH Open Citation Collection: A public access, broad coverage resource**, *Ian Hutchins, Kirk Baker, Matthew Davis, Mario Diwersy, Ehsanul Haque, Robert Harriman, Travis Hoppe, Stephen Leicht, Payam Meyer, George Santangelo*, PLoS Biology

2017 **Additional support for RCR: A validated article-level measure of scientific influence**, *Ian Hutchins, Travis Hoppe, Rebecca Meseroll, James Anderson, & George Santangelo*, PLoS Biology

Protein Topology and Interactions

2019 **Non-specific Interactions Between Macromolecular Solutes in Concentrated Solution: Physico-Chemical Manifestations and Biochemical Consequences**, *Travis Hoppe & Allen Minton*, Frontiers in Molecular Biosciences

2016 **Incorporation of Hard and Soft Protein-Protein Interactions into Models for Crowding Effects in Binary and Ternary Protein Mixtures**, *Travis Hoppe & Allen Minton*, Journal of the Physical Chemistry B

2015 **Dependence of Internal Friction on Folding Mechanism**, *Wenwei Zheng, David De Sancho, Travis Hoppe & Robert B. Best*, Journal of the American Chemical Society

2015 **An equilibrium model for the combined effect of macromolecular crowding and surface adsorption on the formation of linear protein fibrils**, *Travis Hoppe, Allen Minton*, Biophysical Journal

2013 **A simplified representation of anisotropic charge distributions in proteins**, *Travis Hoppe*, Journal of Chemical Physics

2013 **Singular Value Decomposition of the Radial Distribution Function for Hard Sphere and Square Well Potentials**, *Travis Hoppe*, PLoS ONE

2010 **Protein Folding with Implicit Crowders: A Study of Conformational States Using the Wang-Landau Method**, *Travis Hoppe, Jian-Min Yuan*, Journal of Physical Chemistry B

2009 **Entropic flows, crowding effects, and stability of asymmetric proteins**, *Travis Hoppe, Jian-Min Yuan*, Physical Review E

Graph theory

2014 **Integer sequence discovery from small graphs**, *Travis Hoppe, Anna Petrone*, Discrete Applied Mathematics

Experimental Modeling

2014 **Programmable Nanoscaffolds that Control Ligand Display to a G-Protein Coupled-Receptor in Membranes allow Dissection of Multivalent Effects**, *Andrew Dix, Daniel Appella, Travis Hoppe, et al.*, Journal of the American Chemical Society

2014   **Quantification of plasma HIV RNA using chemically engineered peptide nucleic acids**, *Chao Zhao, Daniel Appella, Travis Hoppe, et al.*, Nature Communications

2008   **The importance of EBIT data for Z-pinch plasma diagnostics**, *A S Safronova, Travis Hoppe, et al.*, Canadian Journal of Physics

2006   **Spectroscopic and Imaging Study of Combined W and Mo-pinches at 1 MA-pinch Generators**, *Alla Safronova, Travis Hoppe, et al.*, IEEE Transactions on Plasma Science

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

2016   **Biophysical Society: Los Angeles, CA,** Poster: Coevolutionary signal enhancement

2015   **Biophysical Society: Baltimore, MD,** Seminar: Mean-field lattice-model IDPs, Binding Affinity & Specificity

2014   **Advances in Enhanced Sampling Algorithms: Telluride, CO,** Seminar: Topological considerations in the Wang-Landau algorithm

2013   **Biophysical Society: Philadelphia, PA,** Seminar: Coarse-grained Electrostatic Models for Protein Solutions

2010   **Biophysical Society: San Francisco, CA,** Poster: Wang-Landau Density of States in Crowded Protein Environments

2009   **Drexel University Libraries' Communication Symposium: The Hidden Costs of Scholarly Communication: Philadelphia, PA,** Invited Panel Member

2009   **Biophysical Society: Boston, MA,** Poster: Exhaustive Properties of Simple Lattice Peptides

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	**Office of the Director's Honor Award,** Outstanding support for the Grants Support Index & Next Generation Research Initiative Analytical Team	
2014	**Top Presentation Award,** Institution-wide recognition during the NIDDK Annual Conference.	
2010	**Research Assistant Grant,** Competitive grant from Drexel Physics Department on the basis of outstanding research and teaching.	
2010	**Student Research Achievement Award (SRAA),** Top poster at the Biophysical Society 2010 meeting.	
2009	**Department Research Award (Senior Division),** Given by the Drexel Physics Department, this award recognized a high proficiency in both original research and synthesis of results into publications.	
2008	**Department Research Award (Junior Division),** Restricted to the first two years of study, the junior division award was awarded for early achievements in research.	
2007	**Teaching Assistant of the Year,** Recognition by Drexel University as the top Teaching Assistant in the College of Arts and Sciences.	

TEACHING EXPERIENCE: DREXEL

2011	**PHYS 305**, Computational Physics II*
2010	**PHYS 304,** Computational Physics I*
	PHYS 160, Introduction to Scientific Computing*
	PHYS 305, Computational Physics II*
2009	**PHYS 304**, Computational Physics I*
	PHYS 160, Introduction to Scientific Computing*
	DSP 099, Dragon Summer Program: Remedial Mathematics
	PHYS 100, Preparation for Engineering Studies
	PHYS 305, Computational Physics II*
2008	**PHYS 304**, Computational Physics I*
	PHYS 102, Fundamentals of Physics II*
	PHYS 115, Contemporary Physics III*
	PHYS 114, Contemporary Physics II*
2007	**PHYS 113**, Contemporary Physics I*
	PHYS 102, Fundamentals of Physics II, Lab
	PHYS 115, Contemporary Physics III*
	PHYS 114, Contemporary Physics II*
2006	**PHYS 113**, Contemporary Physics I*
	TDEC 101, Fundamentals of Physics I, Lab
	TDEC 103, Fundamentals of Physics III*
	TDEC 102, Fundamentals of Physics II*
2005	**TDEC 101**, Fundamentals of Physics I*

*Developed new curricula and modernized the Computational Physics, Contemporary Physics and Introduction to Scientific Computing courses at Drexel.