Daniel Parshall, Ph.D.

https://www.linkedin.com/in/danparshall 302.455.7299
https://www.github.com/danparshall parshall.dan@gmail.com
Washington, DC

Summary

I'm a maestro of finding the golden needle of insight buried within the enormous haystacks of data. I love solving technical problems of any sort, but I'm especially interested in **causal inference**, **differential privacy**, and **machine learning** (and these days, the **policy and safety questions** that AI deployment raises). I'm an **extrovert** with a knack for **explaining complex topics** in simple terms to clients, executives, or members of Congress. My Ph.D. training was in **condensed-matter physics**, but I'm also an armchair **economist doing impact analysis** and, lately, the founder of an **AI policy institute**. Most of all, **I get things done.**

Employment

Principal & Founder, Canary Institute, Washington DC (2026 – Present)
I founded the **Canary Institute** (canaryinstitute.ai) as the public face for my **AI Policy** work in **safety and governance**. It seems AI is moving fast enough that policymakers need translations of what the research community is finding into language Hill staff can act on.

My current focus is **AI compute verification**, the technical foundation that lets agreements between otherwise-distrustful parties actually hold. I published an explainer of Cankaya's "Glass Perimeter" low-trust verification architecture (2026-08-28); I'm extending Avenhaus and Krieger's **inspection-game** theorems into the time-dependent scenario for AI (this year's frontier compute is next year's average); I've done simulations estimating that at current utility metering resolution a covert distributed training run can hide down to about 16 accelerators drawing 27 kW per site, and how to spot those using kHz-sampling power monitors.

In the Institute's first few months I **cold-walked roughly fifty Congressional offices** with fact sheets, leave-behinds, and offers of **technical briefings**. Several offices have subsequently taken positions I'd advocated for. I co-authored "Measuring AI's Economic Reach", introducing the Cognitive–Deployment–Regulatory (CDR) framework to **disentangle three dimensions** that prior labor-impact estimates had been silently collapsing; applied across 23,850 task–activity pairs and 923 occupations, the framework finds that 50.5% of U.S. economy-weighted labor time currently meets all three accessibility criteria for AI.

The work above can be found at canaryinstitute.ai/blog/.

Independent Consultant (2017 – Present)

As a Fellow with the Corda Democracy Fellowship (Analogy Group), I prototyped a "LobbyView for all 50 states." The project uses a combination of **Generative AI and traditional ML** to extract lobbying disclosure data from state portals and statutes, producing per-state chains of influence (principal $\rightarrow$ lobbyist $\rightarrow$ lawmaker $\rightarrow$ bill) for WI, NY, and OH. I also designed a pipeline that reads state statutes to evaluate them on a 181-row cell-typed schema based on rubrics in prior art.

I worked with RadVac (Rapid Deployment Vaccine Collaborative) trying to begin clinical trials of a novel open-source intranasal peptide vaccine for preventing the spread of SARS-COV-2. I transformed a 90-

page whitepaper into a 6-page **policy brief**, and **presented core ideas to top government officials** in Latin America.

I worked with researchers from the Inter-American Development Bank (IADB) to identify family connections based on records of citizen information. This work was intended to support **identifying tax fraud and corruption**, with the goal of open-source release. I mentored a research assistant who was accepted as a doctoral student in Economics at both Harvard and MIT.

Principal Data Scientist, Lakeside Software, Remote (2023 – 2025)
My mission at Lakeside was building **"AI that speaks IT"**: bringing visibility to the Digital Employee Experience using one of the richest endpoint telemetry datasets in the industry. Each endpoint collects ~10,000 data points every 15 seconds, distributed across millions of devices, totaling **petabytes of structured data**. That made the technical challenges of machine learning considerable (and really, really fun to solve).

I worked at every level of the DS stack, from **architecting our data science infrastructure**, designing custom ML models, and implementing high-quality production code. I worked closely with Engineering and DevOps to build something **robust, scalable, and cost-efficient**. I trained and mentored staff, providing guidance about code architecture and data science fundamentals (as well as broad career advice).

I also collaborated with **Product, Sales, and Marketing** to identify what customers actually needed, and to shape strategy accordingly. My background in **science communication** was especially useful for providing easy-to-understand explanations of what goes into the machine learning sausage, even to non-technical audiences.

Senior Machine Learning Scientist, Whip Media Group, Los Angeles CA (2022)
Our overall project was a "meta-recommendation engine" for suggesting which titles would be appropriate for which streaming platforms and countries.

While there I extended methods from **explainable AI** to understand the **model stability and feature importance** of the neural network recommendation engine, and **sanity-checks to build confidence** in the model output. I also identified user location from phone and IP data (not trivial given that many users have VPNs for bypassing local restrictions on content availability), and used Bayesian smoothing for identifying trends in user behavior at different geographic levels. Finally, I made numerous improvements to our methods for defining content affinity, user similarity, and title metadata.

Senior Data Scientist, Hobsons/PowerSchool, Arlington VA (2019 – 2021)
My mission at Hobsons was to improve the college, career, and life readiness of students throughout their education journey.

One of the largest projects I worked on was overhauling the customer onboarding for our student-retention predictive analytics. Ultimately we **reduced staff time by 90%**, and also **decreased customer onboarding time by 50%**. Along the way I upgraded the prediction engine to use XGBoost (which **reduced model training time by 80%**), and to provide **explainable AI models** by using SHAP. I **automated delivery** of those reports using LaTeX to generate professional documents.

In other projects I architected a method which allowed predictive models to share results without violating PII restrictions, by using **differential privacy**. I implemented a **generative adversarial network** (GAN) to create synthetic data that could be shared with external collaborators. I developed new metrics to define the similarity between colleges, and high schools. On the **thought leadership** team, I used **causal inference** to conduct an impact analysis of how the Naviance product improved college attendance rates.

Improved data cleanup pipelines by using anomaly detection to find and remove bad records, and in the process **improved the accuracy** of important metrics by 2% (e.g. high school graduation, college persistence). Used **natural-language processing** for a variety of tasks such as matching names of users across datasets, and parsing resumes to identify awards. Provided **mentorship** and advice to staff working on **recommendation engine** for helping students discover possible career paths.

Senior Multidisciplinary Systems Engineer, MITRE, McLean VA (2016 – 2019)
I worked in the Center for Advanced Aviation Systems Design (CAASD), consulting for the FAA. I used **petabyte-scale data** to study safety and efficiency within the National Airspace System (NAS).

While there, I **fostered a culture of continual improvement**. I trained and mentored other analysts, encouraging them to think beyond just their current project, and to extend tools to be useful to other folks within the organization (including creation of reusable analytics software). I organized talks with case studies about analyses. I became the **leading expert in the core datasets** used. I was the official steward of these datasets, which involved teaching classes, serving on the advisory council, and **coordinating with stakeholders** to improve accessibility for everyone. I worked closely with the software developers to identify flaws in the data, and make improvements in the algorithms. I led the effort to improve documentation practices, which decreased engineer onboarding time.

While at CAASD I did a fair amount of **causal analysis**, in order to determine the impact of unusual events on the NAS and identify ways to improve optimization. These events ranged from changes in standardized flight routes, all the way up to airspace closures due to space launches. To make these comparisons, I needed to establish what *would have* happened on a "normal day", so I **defined similarity metrics** to quantify the similarity between days. I used ideas from graph theory to define traffic similarity, and from laser physics to define weather similarity. These metrics **improved accuracy by over 50%**.

I worked on several projects which used machine learning. This ranged from simple problems like classification of aircraft categories (commercial, military, etc.) based on callsign, supervised learning using radar data to pinpoint an aircraft's "wheels up" time, unsupervised learning to identify helipad locations, and building a **variational autoencoder** to suggest possible flight path detours to air traffic controllers. I additionally performed **simulation modeling** and **risk analysis** to estimate the probability of near-collisions during simultaneous runway operations.

Physicist / Instrument Scientist, NIST, Gaithersburg MD (2013 – 2016)
At the National Institute for Standards & Technology I worked in the equivalent of a tenure-track role at the NIST Center for Neutron Research.

Neutron science is now starting to collect datasets large enough to reach the "big data" size. In my own research I developed **new algorithms to take advantage of multi-terabyte data**, by looking for redundant data in order to improve the statistical power and tease out new results, and used statistical tests to conduct hypothesis testing.

In addition to my own research, I also collaborated with other scientists studying novel materials such as superconductors, photovoltaics, and thermoelectrics. I helped them plan and conduct their own experiments using the equipment I helped maintain (a triple-axis neutron spectrometer). I also presented our research methods and results both to other scientists, and to the general public. I also mentored a high-school student, who has since gone on to complete his own Ph.D. in Physics at MIT.

Education

2010 Ph.D. Physics, University of Tennessee, Knoxville
2004 M.S. Applied Physics, University of South Florida
2002 B.A. Physics, University of South Florida (*cum laude in cursu honorum*)

Continuing Education

2026	BlueDot Impact	Biosecurity
2026	BlueDot Impact	Frontier AI Governance
2026	BlueDot Impact	AGI Strategy
2016	The Data Incubator	Data Science Fellowship
2016	National Academy of Sciences	Learning from Big Data
2015	Stanford University	Machine Learning (Andrew Ng)

Awards

2026 Best in Show "Hacking the Think Tank" Hackathon (Foundation for American Innovation)

Teaching ($\sim$10k hours in a variety of settings)

1996 – 2001	Museum of Science & Industry	Science demonstrator
2002 – 2004	University of South Florida	General physics
2004 – 2006	University of Tennessee	General physics
2017	The Data Incubator	Machine learning & TensorFlow
2016 – 2018	MITRE Institute	Statistics and analysis with aviation data
2024 – 2025	Lakeside Software	Employee training on data science methods

Languages

English native
Spanish fluent (CEFR B2)

Keywords for your Applicant Tracking System

Tools Spark, Hadoop, SQL, Python, Pandas, scikit-learn, XGBoost, jupyter notebook, PyTorch, TensorFlow by Google, Matlab, linux, git, command line, AWS by Amazon, Databricks, large language models (LLMs), generative AI

Techniques A/B testing, linear regression, logistic regression, multiple regression, supervised learning, unsupervised learning, classification, customer segmentation, recommendation system, anomaly detection, natural language processing, differential privacy, generative adversarial neural network, variational autoencoder, geospatial, time series, graph theory, random forest, predictive modeling, simulation, risk analysis, causal inference, explainable AI (SHAP), Bayesian smoothing, Bilateral Constitutional AI (BCAI), constitutional AI, value alignment, RLHF, mechanism design, task-level exposure analysis, occupational classification (O*NET), CDR framework

Skills mentoring, leadership, teaching, hypothesis testing, Bayesian statistics, client-facing, thought leadership, data analysis, AI policy, AI governance, AI safety, science communication, public speaking, lobbying, policy briefs, multi-stakeholder governance