

Trey Saddler

📍 Durham, NC | 📞 (248) 919-8739 | ✉ treyosaddler@gmail.com | 🌐 treysaddler |
📧 tosaddler | 🌐 treysaddler.com

Senior Data Scientist & AI Engineer | LLMs, RAG, Cloud Infrastructure, Scientific Computing

Professional Summary

Data scientist and AI engineer with 10+ years building scientific computing, data, and AI systems at the National Institutes of Health. Most recently, as subject-matter expert for NIEHS's CAMERA platform, I led its LLM development and AWS cloud strategy and set technical direction for a two-person engineering team moving it to a NIH-compliant, infrastructure-as-code environment. Earlier, I won a \$150,000 NIH award to bring LLMs into the cloud at NIEHS and turned it into a shared LLM interface and API, RAG pipelines over vector databases, and agentic data-extraction workflows that developers used on active research projects. My background in toxicogenomics and HPC means I can talk to the scientists and still ship the production system. Tribal college graduate in environmental health and advocate for Indigenous data sovereignty.

Core Skills

LLMs & RAG	Large language models (LLMs) RAG architectures Agentic workflows Prompt engineering Embeddings & semantic search Transformer architectures
LLM Tooling	LangChain LangGraph LlamaIndex LangFuse promptfoo LLM APIs
Vector / Graph DB	Chroma FAISS pgvector Graph databases Knowledge graphs
Cloud & MLOps	AWS Infrastructure as code CI/CD Docker Apptainer Posit Team Linux administration
HPC & Data	Slurm NIH Biowulf Globus Spark Hadoop Reproducible pipelines (targets, renv)
Languages & Apps	Python R Bash FastAPI Flask Plumber Shiny Quarto
Science Domain	Toxicogenomics BMDEpress Cytoscape Ingenuity Pathway Analysis FAIR data practices

Professional Experience

Dynanet Corporation, supporting NIEHS

COMPUTER AND INFORMATION RESEARCH SCIENTIST (SME)

Durham, NC (Remote)

Dec 2025 - Sep 2026

- Led AI/LLM development and AWS cloud strategy for the CAMERA platform (camera.niehs.nih.gov), setting the technical roadmap for its infrastructure and AI capabilities.
- Directed a full-stack developer and a database architect through cloud migration, performance tuning, and AI/ML integration.
- Drove CAMERA's migration to a secure, NIH-compliant AWS environment built on infrastructure as code, CI/CD, and container orchestration, making deployments repeatable and auditable.
- Built LLM-powered search over a structured knowledge base, with spell correction, synonym handling, and semantic query parsing so users find relevant studies without exact keyword matches.
- Automated generation and upkeep of search filters and facets with LLMs, cutting the manual curation work behind each release.
- Refined PubMed search strategies and literature-identification workflows that feed CAMERA's periodic review cycles.
- Designed model-results visualizations adapted from EASA templates for CAMERA users.

Division of Translational Toxicology, National Institute of Environmental Health Sciences

DATA SCIENTIST

Durham, NC

May 2022 - Jul 2025

- Won a \$150,000 NIH NOSI award (NOT-OD-23-070) to evaluate LLMs and AI in the cloud, and used it to stand up NIEHS's internal LLM service.
- Launched an internal LLM interface and API built on open-source models and tooling, avoiding vendor lock-in and enabling collaboration across NIH institutes; trained developers to apply it on active research projects.
- Built RAG pipelines over large volumes of unstructured scientific text, generating embeddings for semantic search in Chroma, FAISS, and pgvector to improve retrieval quality for downstream LLM applications.
- Engineered agentic workflows with LangChain and LlamaIndex to extract structured data from disparate datasets.
- Authored the NIEHS Scientific Developer's Guide, the onboarding reference for scientific developers on pipelines, HPC, reproducibility, and FAIR data storage, transfer, and sharing.
- Deployed and monitored analytical pipelines on NIH Biowulf and the NIEHS Slurm cluster using Apptainer, Docker, Globus, and Bash.

Office of Data Science, National Institute of Environmental Health Sciences

DATA SCIENTIST

Durham, NC

Aug 2018 - May 2022

- Architected and administered the institute's Posit Team platform, giving scientists a self-service path to publish R Markdown, Jupyter, and Quarto notebooks, Shiny apps, dashboards, and FastAPI, Flask, and Plumber APIs.
- Built reproducible analysis pipelines with targets, renv, Quarto, and Docker so analyses stay rerunnable and maintainable over time.
- Ran the NIEHS GitHub organization and coached researchers on version control, semantic versioning, and package management (pip, conda, renv, npm).
- Developed interactive Shiny, plotly, and ggplot2 applications for exploring scientific data.
- Taught workshops on Cytoscape network visualization and gene co-expression analysis.
- Administered CentOS, Ubuntu, and Rocky Linux servers hosting internal scientific deployments.

National Toxicology Program, National Institute of Environmental Health Sciences

POSTBACCALAUREATE IRTA RESEARCH FELLOW

Durham, NC

Sep 2015 - Aug 2018

- Co-developed and documented BMDExpress 2.0, NTP's tool for benchmark-dose analysis of genomic data; trained users and presented it at scientific conferences.
- Performed toxicogenomic analysis of gene expression data with Ingenuity Pathway Analysis, Cytoscape, BMDExpress 2.0, R, PubChem APIs, DrugMatrix, and ToxFX.
- Automated toxicology report generation with R Markdown, ggplot2, and drake (now targets).

Education

Salish Kootenai College

B.S., ENVIRONMENTAL HEALTH

Pablo, MT, USA

Jun 2015

Early Research & Internships

US EPA, NIEHS, and Salish Kootenai College

RESEARCH INTERNSHIPS, LAB WORK, AND TUTORING

MT, WA, CA, NC

2011 - 2015

- **EPA Region 9 ITEP Intern** (San Francisco, 2015): tribal indoor air quality.
- **EPA GRO Fellow, Region 10** (Seattle, 2014): NPDES stormwater and hatchery permitting.
- **NIEHS/NTP Summer Intern** (Durham, 2013): effects of TBBPA on human uterine cells, Molecular Pathology Unit.
- **Salish Kootenai College** (2011-2015): measured mercury, arsenic, and selenium in environmental and biological samples (EPA Methods 7473, 1630); tutored math, biology, and chemistry.

Publications

- Borrel, A., Mansouri, K., Nolte, S., Saddler, T., Conway, M., Schmitt, C., & Kleinstreuer, N. C. (2020). InterPred: a webtool to predict chemical autofluorescence and luminescence interference. *Nucleic Acids Research*, 48(W1), W586–W590. <https://doi.org/10.1093/nar/gkaa378>
- Castro, L., Liu, J., Yu, L., Burwell, A. D., Saddler, T. O., Santiago, L. A., Xue, W., Foley, J. F., Staup, M., Flagler, N. D., Shi, M., Birnbaum, L. S., & Dixon, D. (2021). Differential receptor tyrosine kinase phosphorylation in the uterus of rats following developmental exposure to tetrabromobisphenol A. *Toxicology Research and Application*, 5, 239784732110471. <https://doi.org/10.1177/23978473211047164>
- Jokinen, M., Morgan, D., Price, H., Herbert, R., Saddler, T., & Dixon, D. (2017). Immunohistochemical Characterization of Sarcomas in Trp53+/- Haploinsufficient Mice. *Toxicologic Pathology*, 45(6), 774–785.
- Phillips, J., Svoboda, D., Tandon, A., Patel, S., Sedykh, A., Mav, D., Kuo, B., Yauk, C., Yang, L., Thomas, R., Gift, J., Allen Davis, J., Olszyk, L., Alex Merrick, B., Paules, R., Parham, F., Saddler, T., Shah, R., & Auerbach, S. (2019). BMD Express 2: Enhanced transcriptomic dose-response analysis workflow. *Bioinformatics*, 35(10), 1780–1782.
- Ramaiahgari, S., Auerbach, S., Saddler, T., Rice, J., Dunlap, P., Sipes, N., Devito, M., Shah, R., Bushel, P., Merrick, B., Paules, R., & Ferguson, S. (2019). The power of resolution: Contextualized understanding of biological responses to liver injury chemicals using high-throughput transcriptomics and benchmark concentration modeling. *Toxicological Sciences*, 169(2), 553–566.
- Svoboda, D., Saddler, T., & Auerbach, S. (2019). An Overview of National Toxicology Program's Toxicogenomic Applications: DrugMatrix and ToxFX. *Challenges and Advances in Computational Chemistry and Physics*, 30, 141–157.