John Bollenbacher

Contact

jmbollenbacher@gmail.com ✉

jbollenbacher.github.io 🌐

jmbollenbacher in

Languages & Tools

Python (primary),
SQL, R (familiar)

pandas, sklearn, pytorch, nltk,
statsmodels, Ollama,
SetFit/BERT, etc.

Linux systems, Bash, Git

High Performance Computing
(HPC)

Data Science

Natural Language Processing
(NLP/NLU)

Generative AI / LLMs,
Task Automation & RAG

Social Media Analysis

Geospatial Data Science

Classification, Clustering,
Nonlinear Regression

Time Series &
Point Process Methods

Linear Models

Causal Inference

Bayesian Statistics

Model Validation
& Model Selection

Communication & Leadership

Grant & Contract Proposals

Project Management

Technical Reports
& Academic Papers

Presentations & Demos

Data Visualization

Education

May 2024	**PhD** Informatics & Complex Systems; minor in Statistics	Indiana University
May 2019	**MS** Informatics	Indiana University
May 2016	**BS** Physics; minor in Computer Science	Georgia Institute of Technology

Selected Experience

2022 - Present — **Research Data Scientist** — RTI International
• Conducting quantitative research in various domains including public health and media studies, in collaboration with domain experts.
• Providing analytical support for strategic decision making by Federal agencies and private clients, including through data dashboards.
• Specializing in NLP, causal inference, and generative AI / LLMs.
• Co-Authoring $8+ million of successful contract and grant proposals.

2016 - 2022 — **Research Assistant & Assistant Instructor** — Indiana University
Analyzed large social media datasets and modeling & forecasting social systems (2017-2019; 2021-2022). Taught an ethics of technology course (2019 - 2022) and a discrete math course (2016 - 2017).

Summer 2021 — **Data Scientist** — GeniusMesh
Used LinkedIn data to analyze Executive MBA's career paths to make career decision recommendations to others. Worked with a development team to deploy a client-facing data dashboard to production.

Selected Projects & Research

2022 - Present — **Using LLMs to Accelerate Qualitative Research**
Developing Large Language Model (LLM) apps to solve problems for research teams. Use cases include: plain language writing, qualitative coding of texts, text cluster naming, document-based question answering (RAG), information extraction from PDF documents, NER, and others.

2024 - Present — **Supporting NSF Program Evaluation & Improvements**
Leading the analytics team on a multimillion dollar contract supporting National Science Foundation's program evaluation work, and related program improvements. Conducting analysis of program evaluation reports and processes, website user experience data, and communications strategy outcomes to support decision making these areas.

2020 - 2023 — **Measuring Offline Effects of Online Social Media**
Developed methods for causally linking online content to offline outcomes in public health and politics. Showed that antivaccine Tweets lead to reduced vaccine uptake and increased deaths during the COVID pandemic. Showed that the UK Parliament's attention follows public attention on social media. Analyzed datasets of tens of billions of short texts.