John Bollenbacher

Contact

jmbollenbacher@rti.org
jbollenbacher.github.io
jmbollenbacher

Languages & Tools

Python (primary)
SQL, R, Matlab (familiar)
pandas, sklearn, scipy, pytorch, spacy, etc.
Linux Systems, Bash, Git
HPC & Cloud Computing

Data Analysis & ML

Natural Language Processing (NLP/NLU)
Generative AI & Task Automation
Causal Inference
Linear Models
Social Network Analysis
Classification, Clustering, Nonlinear Regression
Dimensionality Reduction & Manifold Learning
Time Series & Point Process Methods
Bayesian Statistics
Modeling & Simulation
Model Validation & Model Selection

Communications

Technical Reports & Academic Papers
Presentations & Demos
Data Visualization
Teaching & Lectures

Education

Dec. 2023	**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
Applying data science methods to conduct quantitative research in collaboration with domain experts in subject areas including public health, environmental science, and media studies. Specializing in NLP, causal inference, and task automation with generative AI.

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

Summer 2021 — **Data Scientist** — GeniusMesh
Performed data science and NLP tasks to analyze career paths of Executive MBAs and make career decision recommendations. Worked with a development team to deploy a client-facing data dashboard to production.

Selected Research & Projects

2022 - Present — **Using LLMs to Facilitate Qualitative Research**
Developed Large Language Model (LLM) apps to solve problems for several different research teams. Specific applications included plain language writing, qualitative coding of texts, text cluster naming, document-based question answering, and cleaning texts and tables extracted from PDFs.

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 adopts topics that appeared first on social media.

2017 - 2019 — **Modeling Online Conversations and User Behavior**
Used social media data and machine learning models to forecast user behaviors and the structure of online conversations. Modeled user conversations on Reddit, Twitter virality, and the GitHub ecosystem including repo activity and user behavior.