Cody C. Baker, Ph.D.

codycbakerphd.com | cody.c.baker.phd@gmail.com | cody-c-baker | github.com/CodyCBakerPhD

SUMMARY

Experienced data scientist and software engineer with 12 years spent designing, conducting, and sharing results of complex research. Possesses a diverse set of software skills, including DevOps, full-stack app development, performance benchmarking, ETL, and cloud deployments.

EDUCATION

2015 – 2020 **Ph.D.** | Applied and Computational Mathematics and Statistics | *University of Notre Dame*

2011 – 2015 **B.S.** | Applied Mathematics and Cognitive Science | *University of Evansville*

EXPERIENCE

2025 – Present **Research Software Engineer** | *Center for Open Neuroscience* | *Dartmouth College* | Hybrid

- Maintaining and enhancing several NIH data archives that store neuroscientific data by standardizing dataset organization across multiple modalities and providing usage analytics, including geolocation, through parsing terabytes of raw S3 log files.

2020 – 2024 **Research Software Engineer** | *CatalystNeuro* | Remote

- Created and maintained 12+ data processing pipelines for neuroscience labs, allowing their data to flow from acquisition to sharing in a seamless fashion.
- Curated a total of 256 TB of high-value datasets to NIH data archives on behalf of various research groups.
- Managed the company's cloud resources (Amazon Web Services; AWS), including storage, compute resources, and identity access management (IAM).

2016 – 2020 **Research Assistant** for the Neural Computing Group at *University of Notre Dame*

SELECTED PROJECTS

2026 – Present **HEARTH** | *Lead Developer*

- Built the Hub for Exploratory Analysis, Resource-sharing, Testing, and Hypothesis-generation (HEARTH), a suite of web apps for the Brain Behavior Quantification and Synchronization (BBQS) consortium and its data on the EMBER Archive.

2024 – Present **S3 Log Extraction** | *Lead Developer*

- Created a Python package for the fast extraction of access summary data from Amazon S3 server logs, originally developed for the DANDI Archive and steadily being rolled out to others.

2023 – 2024 **NWB GUIDE** | *Lead Developer*

- Developed an intuitive user interface for file management using interactive validation and real-time suggestions, streamlining the process for data submission to NIH archives.

2020 – 2024 **NeuroConv** | *Lead Developer*

- Established an automated data conversion tool capable of reading more than 40 distinct data formats used by neurophysiology acquisition devices in order to automatically write to the NeurodataWithoutBorders (NWB) standard.

SKILLS

Coding languages: Python, MATLAB, Typescript, Javascript, HTML, LaTeX, R, C, C++, Java, CSS

Math, stats, & ML: Multivariate regression, classification, cross-validation, clustering, mixture models, dimensionality reduction, ensemble methods, deep learning, reinforcement learning, time series analysis, anomaly detection, computer vision

DevOps: Git, GitHub Actions, Docker, unit & integration testing, user & quality testing, Storybook, Chromatic, Puppeteer, Codecov

SELECTED PUBLICATIONS

2025 C. J. Gillon, **C. Baker**, R. Ly, E. Balzani, B. W. Brunton, M. Schottdorf, S. Ghosh, and N. Dehghani. "Open data in neurophysiology: Advancements, solutions & challenges." *eNeuro, vol. 12, no. 11.*

H. Mayorquin, **C. Baker**, P. Adkisson-Floro, S. Weigl, A. Trappani, L. Tauffer, O. Rübel, and B. Dichter. "NeuroConv: Streamlining neurophysiology data conversion to the NWB standard." *Python in Science Conference, 2025.*

2020 **C. Baker**, V. Zhu, and R. Rosenbaum. "Nonlinear stimulus representations in neural circuits with approximate excitatory-inhibitory balance." *PLoS computational biology, vol. 16, no. 9, e1008192, 2020.*