Cody C. Baker

✉ cody.c.baker.phd@gmail.com   in cody-c-baker   🌐 codycbakerphd.com
⬛ github.com/CodyCBakerPhD

ⓘ Summary

Experienced data scientist and software engineer with 12 years spent designing, conducting, and sharing results of complex research. Worked with faculty, staff, and students from dozens of top universities to achieve excellence in the preparation and performance of research-related tasks, including data management, hypothesis formulation, and mathematical modeling.

Possesses a diverse set of software skills, including DevOps, full-stack app development, performance benchmarking, ETL, and cloud deployments. Enjoys hosting workshops and hackathons to train users in state-of-the-art technologies and best practices. Familiar with a large number of data formats, with a strong concentration on updating local file standards (e.g., binary, CSV, JPG, etc.) to cloud-performant ones (HDF5, Zarr).

⚙ Skills

Coding (expert)	🔖 Python, MATLAB
Coding (proficient)	🔖 R, C, C++, Logo, LaTeX
Coding (experienced)	🔖 Java, JavaScript, CSS, HTML, SQL
Machine learning	🔖 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, linting
Libraries	🔖 NumPy, SciPy, pandas, sklearn, OpenCV, FFmpeg, timeit, airspeed velocity, Matplotlib, Plotly, Flask, PyInstaller, Pydantic, Boto3
Frameworks	🔖 AWS (IAM, S3, EC2, Batch, DynamoDB), Electron, Tk, REDIS, REST, GraphQL, Jupyter, Sphinx, pre-commit
Data formats	🔖 HDF5, Zarr, JSON, YAML, XML, Markdown, RST, Parquet, TSV, XLSX, MP4, AVI, TIFF, PNG, JPEG, MP3, WAV, PKL, MAT
Administration	🔖 GitHub, AWS, PyPI, Condaforge, DockerHub, ReadTheDocs
Business & team skills	🔖 Agile, scrum, research, mentoring, training, client consultations, drafting scopes of work, budgeting, grant writing

🎓 Education

2015 – 2020	🔖 **Ph.D.** | Applied and Computational Mathematics and Statistics | *University of Notre Dame* Doctoral thesis: "Second-order moments of activity in large neural network models" Relevant coursework: machine learning, Bayesian statistics, network science, stochastic analysis, time series analysis, partial differential equations, nonlinear dynamics
2011 – 2015	🔖 **B.S.** | Applied Mathematics and Cognitive Science | *University of Evansville* Minors in Computer Science, Philosophy, Psychology, and Neuroscience Senior thesis: "Crime prediction models applied to the city of Evansville" Honors thesis: "Goodness of fit metrics for joint probability distributions" Relevant coursework: calculus, linear algebra, probability, statistics, discrete & combinatorics, numerical analysis, real analysis, mathematical physics, algorithms & data structures, cryptography, machine learning, symbolic logic, cognitive psychology, neurophysiology, advanced neuroscience, philosophy of mind, philosophy of science

♕ Key Achievements

Data	🔖 Published over **256** TB of neuroscience data to NIH archives, allowing anyone in the world to test scientific hypotheses.
Education	🔖 Trained **200+** professionals through **6** workshops and hackathons, disseminating the practices of open data.
Software	🔖 Developed software used by more than **50** research groups globally, integrating with daily data pipelines and practices.
Standardization	🔖 Standardized metadata of **40** data formats used by scientific devices, creating a universal language for comparing results.
Publications	🔖 Published **6** papers and **8** conference posters, spreading the word about open-source software and public data.
Pipelines	🔖 Updated **10** data analysis pipelines to leverage cloud streaming capabilities to improve speed and robustness.

◔ Experience

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

· Developed and maintained 5 repositories of open-source software by ensuring proper functionality of automated testing suites, documentation, tutorials, and demos.

· Created and maintained 12+ data processing pipelines for neuroscience labs, allowing their data to flow from acquisition to sharing in a seamless fashion.

· Personally 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).

· Handled user interactions within the research community by offering technical support and resolving issues or feature requests in a timely manner.

· Facilitated user education across various platforms by running sessions at multiple conferences and workshops, increasing user adoption and effective system utilization.

· All software related in some way to the facilitation of terabyte-scale data management, analysis, and visualization for the field of neurophysiology.

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

· Reconciled the computational properties of biologically realistic neural networks with artificial machine learning models (such as those used across computer vision) through complex mathematical theory and stochastic biophysical simulations, with results communicated through 3 journal publications and a presentation at the high-impact COSYNE conference (top 4% of abstracts accepted).

· Collaborated with several top experimentalists in neuroscience as a trainee in the NeuroNex program, which focused on understanding how neural function emerges from underlying structure.

2015 – 2017   🔖 **Teaching Assistant** for Applied Mathematical Methods at *University of Notre Dame*

· Ran 4 tutorial sections for 112 students, aiding their study of course concepts to achieve a 96% satisfaction rate.

· Graded homework assignments and exams. Filled in main lectures for the professor as needed.

2015 - 2016   🔖 **Data Analyst** for *JRM Environmental Inc*

· Analyzed geological data from water samples in conjunction with the Indiana Department of Environmental Management to issue compliance permits for Indianapolis regulation standards.

Summer 2015   🔖 **Technical Editor** for *Penguin Random House*

· Reviewed and corrected over 300 pages of "An Idiot's Guide to Algebra II".

Summer 2014   🔖 **Research Intern** for the BITLab at *Michigan State University*

· Explored statistical effects of algorithmic curation (the use of automated filtering mechanisms in the delivery and display of information) by measuring properties of simulated models of social networks, with results presented at the MIDSURE conference.

◕ Experience (continued)

Summer 2013   ▮ **Research Intern** | Bioinformatics | *University of Kansas*

· Examined information diffusion through large-scale simulations of G-protein signaling mechanisms using a high-performance super-computing cluster (HPC), then presented results at an undergraduate symposium.

⚗ Highlighted Projects

Below are the most recent highlights of my major work.
Explore a full listing of all my projects on LinkedIn.

2023 – 2024   ▮ **Lead developer** of *NWB GUIDE*

· Developed an intuitive user interface for file management using interactive validation and real-time suggestions, streamlining the process for data submission to NIH archives.
· Ensured an extremely robust testing suite involving multiple levels of integration and user interactions emulated using Puppeteer to enhance reliability and long-term maintainability.
· Tracked and documented dozens of hands-on user tests to refine the user experience.

2022 - 2024   ▮ **NeuroConv** | *Lead Developer*

· Led the development of an automated data conversion tool capable of reading more than 40 distinct data formats used by neurophysiology experiment devices in order to automatically write to the NeurodataWithoutBorders (NWB) standard.
· Designed universal APIs which transparently handled each layer of complexity to simplify the tasks of tagging, grouping, metadata transcription, temporal alignment, asset linking, buffering, chunking, and compression.
· Implemented a cloud deployment system to run large-scale, off-site, batched conversions through Amazon Web Services (AWS).

2022 – 2024   ▮ **Lead developer** of *NWB Inspector*

· Created a command line tool used by the NIH data archive to validate all data uploads, which ensures all submissions are provided automated suggestions for metadata improvements that enhance data findability and reuse.
· Mirrored the design, style, and functionality of linting tools such as flake8, pydocstyle, and ruff.

目 Publications

Journal Articles

1   C. J. Gillon et al., "Open Data In Neurophysiology: Advancements, Solutions & Challenges," *ArXiv*, arXiv–2407, 2024.

2   J. Magland, J. Soules, C. Baker, and B. Dichter, "Neurosift: DANDI exploration and NWB visualization in the browser," *Journal of Open Source Software*, vol. 9, no. 97, p. 6590, 2024.

3   M. Hawrylycz et al., "A guide to the BRAIN Initiative Cell Census Network data ecosystem," *PLoS biology*, vol. 21, no. 6, e3002133, 2023.

4   C. Baker, E. Froudarakis, D. Yatsenko, A. S. Tolias, and R. Rosenbaum, "Inference of synaptic connectivity and external variability in neural microcircuits," *Journal of computational neuroscience*, vol. 48, pp. 123–147, 2020.

5   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.

6   C. Baker, C. Ebsch, I. Lampl, and R. Rosenbaum, "Correlated states in balanced neuronal networks," *Physical Review E*, vol. 99, no. 5, p. 052 414, 2019.

Conference Presentations

1   C. Baker, G. Flynn, R. Ly, O. Ruebel, and B. Dichter, "NWB GUIDE: Simplifying the conversion of neurophysiology data to NWB format," in *Society for Neuroscience (SfN)*, 2024.

2   C. Baker, S. Weigl, H. Mayorquin, and B. Dichter, "NeuroConv: Automated conversion of neurophysiology data to NWB format," in *Society for Neuroscience (SfN)*, 2023.

3   C. Baker, V. Zhu, and R. Rosenbaum, "Nonlinear computations in semi-balanced networks," in *Computational and Systems Neuroscience (COSYNE)*, contributed talk (top 4% of abstracts), 2020.

4   C. Baker and R. Rosenbaum, "Inferring connectivity and latent input covariance from spike train correlations," in *Computational and Systems Neuroscience (COSYNE)*, poster (top 43% of abstracts), 2018.

Workshop Sessions

1   **Baker, Cody** and G. Viejo, "Artificial intellgence, machine learning, computing, and visualization in neuroinformatics," in *Open Data in Neuroscience (ODIN)*, Massachusetts Institute of Technology, 2023.