

Jennifer Williams

✉ jlw1@alumni.cmu.edu

🌐 <http://github.com/jennwilliams>

🏠 jennwilliams.github.io

📍 Los Angeles, California

Highlights

- Expertise building ML solutions specializing in Natural Language Processing (NLP), including Large Language Models (LLMs), multi-modal learning, deep learning, unsupervised and supervised learning
- Designed, deployed, evaluated, and optimized ML models, improving experiences for thousands of internal users, and customer satisfaction, in collaboration with cross-functional teams
- Extensive experience developing ML models across diverse data types—including text, audio, images, videos, biosignals, timeseries, and tabular data—demonstrating versatility in solving complex problems
- Proposed and implemented novel ML frameworks and causal inference tools (i.e., Bayesian networks) enabling actionable insights and scientific discovery from complex real-world datasets

Technical Skills

- Python (Scikit-learn, Pandas, NumPy, SciPy, Matplotlib), PyTorch, TensorFlow, Java, R, MATLAB, Bash, Cloud Platforms (AWS, GCP), Version Control (Git), Container Platforms, Database Systems (SQL)

Education

PhD, School of Computer Science, Carnegie Mellon University (CMU), 2016 - 2022

Advisor: Dr. Leila Wehbe (Machine Learning Department)

Thesis: Modeling individual differences in language processing in the presence and absence of disease

Thesis Committee: Tom Mitchell, Timothy Verstynen, Ashok Panigrahy

Cancer and Systems Biology EU-USA Atlantis Dual Degree Program, 2013 - 2016

MS in Natural Science, *Roswell Park, University at Buffalo*

MSc in Integrated Systems Biology, *University of Luxembourg*

BS in Biology, *Canisius University*, Magna Cum Laude, 2009 - 2013

Experience

Senior Machine Learning Scientist, CVS Health, April 2023 - present

- Developed and deployed an internal conversation agent leveraging Retrieval-Augmented Generation (RAG) with hybrid search, driving up employee productivity by 38% and customer satisfaction by 2%
- Designed and implemented evaluation framework for Generative AI system, enabling deployment
- Presented technical demonstrations for products I co-developed to groups of up to 1,600 colleagues and Senior VPs, showcasing the strategic impact of ML solutions on business objectives
- Led a self-organizing Agile team in creating a tool for deploying ranking models, achieving a remarkable reduction in time-to-deployment to 20 minutes, enhancing operational efficiency
- Co-founded “ML Lunch and Learn” lecture series and co-led “Generative AI Training” workshops, driving upskilling across technical teams, non-technical teams, and leadership

Selected

Publications

- **Same cause; different effects in the brain.**
M. Toneva*, J. Williams* (co-first), A. Bollu, C. Dann, L. Wehbe. *Causal Learning and Reasoning (CLear)* 2022
- **Behavior measures are predicted by how information is encoded in an individual's brain.**
J. Williams, L. Wehbe. *arXiv* 2021 (in submission)
- **Discriminative subtyping of lung cancers from histopathology images via contextual deep learning.**
B.J. Lengerich, M. Al-Shedivat, A. Alavi, J. Williams, S. Labbaki, E.P. Xing. *medRxiv* 2020

Relevant Courses

- Machine Learning
- Statistical Methods in Machine Learning
- Intermediate Statistics
- Probabilistic Graphical Models

Leadership and Service

- Co-organizer of the ICLR workshop *ICBINB Deep Learning* and the “*brAln*” seminar series at CMU
- Reviewer for Nature Scientific Reports, ICLR, NeurIPS (WiML, Learning from Time Series for Health), IJCAI, ICML (New in ML), ML4H, ECCB
- Mentored CS student (now PhD at Princeton)

Awards

- International Conference on Machine Learning (ICML) Travel Award
- Top 10 Reviewer Machine Learning for Health (ML4H)
- Invited to attend Machine Learning Summer School (MLSS)

Selected Research Projects

Deep learning to integrate multi-modal data

- Question: Can sample-specific models, similarly to subject matter experts, effectively integrate multi-modal data for accurate classification?
- Method Innovation: Inspired by contextual deep learning, created sample-specific multi-modal models for lung cancer classification, by adapting Contextual Explanation Networks (CENs) (AI-Shedivat et al., JMLR 2020). Integrated both imaging and transcriptomic data into the classification models.
- Scientific Discovery: Sample-specific multi-modal models increase classification accuracy and capture the heterogeneity of biological processes underlying lung cancer.
- Paper: [medRxiv](#) (DOI: [10.1101/2020.06.25.20140053](#))

Disambiguating language processing with causality

- Question: Why do neural network derived features of language predict large parts of the brain well? Do these brain zones process the features similarly or differently?
- Method Innovation: Developed a causal inference framework, that includes two new metrics, to provide insights beyond current brain mapping techniques. Specifically, the framework enables researchers to infer if a complex (multivariate and high dimensional) stimulus, such as language, affects two brain zones similarly.
- Scientific Discovery: Real-world language stimuli (i.e., videos) do not affect all parts of the brain's language network similarly.
- Paper: CLeaR 2022 ([arXiv 2202.10376](#))
- Code: [github.com/brainML/stim-effect](#)

Modeling individual differences for personalized insights

- Question: Can individual differences in how information is encoded in the brain predict behavior?
- Method Innovation: Built on insights from two sub-fields of neuroscience (brain mapping and behavioral neuroscience), to create the first machine learning framework to identify individual differences in brain encoding and test if these differences predict behavior.
- Scientific Discoveries:
 - Individual differences in brain encoding can predict behavioral variability.
 - Advised researchers to optimize their choice of neuroimaging task and feature-space for their behavior of interest.
- Paper: [arXiv 2112.06048](#)
- Code: [github.com/brainML/great-apes](#)

Fellowships

- **Digital Health Fellowship** – Center for Machine Learning and Health (2020 - 2021)
Full tuition and stipend for 12 months and \$3,000 for research-related expenses
- **NIH T32 Training Grant** – National Institute of Biomedical Imaging and Bioengineering (2017 - 2019)
Full tuition and stipend for 2 years and \$6,000 for research-related expenses
- **CanSys MS Scholarship** – Atlantis EU-USA Training Program (2013 - 2015)
Stipend for 12 months