Jennifer Williams

✉ jlw1@alumni.cmu.edu
○ http://github.com/jennwilliams
⌂ jennwilliams.github.io

Highlights

- Deep expertise in building Machine Learning (ML) solutions specializing in Natural Language Processing (NLP), encompassing Large Language Models (LLMs), unsupervised, supervised, and deep learning
- Collaborated with cross-functional teams to develop ML models, showcasing versatility in crafting solutions for various data types including natural language, sensor data, imaging data, and tabular data
- Proposed, implemented, and validated novel ML frameworks to analyze imaging timeseries data
- Conceptualized and developed causal inference tools (i.e., Bayesian networks) fostering scientific discovery in complex datasets

Education

PhD in Computational Biology, *School of Computer Science, Carnegie Mellon University*, 2016 - 2022
Advisor: Dr. Leila Wehbe (Machine Learning Department)
Thesis: Modeling individual differences in the brain 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 College,* Magna Cum Laude*,* 2009 - 2013

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*
- **LSD1 dual function in mediating epigenetic corruption of the vitamin D signaling in prostate cancer.**
  S. Battaglia, E. Karasik, B. Gillard, J. Williams, T. Winchester, M.T. Moser, D.J. Smiraglia, B.A. Foster. *Clinical Epigenetics 2017*

Experience

Senior Machine Learning Scientist, *CVS Health*, since April 2023

- Orchestrated and led a self-organizing Agile team to develop a tool for deploying ranking models on Google Cloud Platform (GCP), achieving a remarkable reduction in time-to-deployment to 20 minutes
- Presented technical demonstrations for co-developed products to diverse audiences, ranging from small groups to gatherings of up to 1,600 colleagues, including Senior VPs
- Designed and developed a scalable TensorFlow personalization model framework, enhancing ML solution efficiency

Computing Skills

- Python (Scikit-learn, Pandas, NumPy, SciPy, Matplotlib), Java, R, MATLAB, Bash, Cloud Platforms, Git, Container Platforms, Database Systems (SQL)

Relevant Courses

- Machine Learning
- ABCDE of Statistical Methods in Machine Learning
- Intermediate Statistics
- Probabilistic Graphical Models
- Cognitive Neuroscience

Leadership and Service

- Co-founder ML Platform Lunch and Learn Series
- Reviewer for IJCAI, Nature Scientific Reports, WiML and Learning from Time Series for Health NeurIPS Workshops, New in ML NeurIPS and ICML Workshops, ML4H Conference, ECCB,
- Mentored Undergraduate Computer Science student (currently PhD student 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)
- CMU Provost Conference Award

Selected Research Projects

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

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 multimodal models for lung cancer classification, by adapting Contextual Explanation Networks (CENs) (Al-Shedivat et al., JMLR 2020). Integrated both imaging and transcriptomic data into the classification models.
- Scientific Discovery: Sample-specific multimodal models increase classification accuracy and capture the heterogeneity of biological processes underlying lung cancer.
- Paper: medRxiv (DOI: 10.1101/2020.06.25.20140053)

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