Calvin McCarter

http://calvinmccarter.com/

ABOUT ME
AI researcher and engineer, with
▷ first-author papers in NeurIPS and AISTATS,
▷ extensive open-source software contributions, and
▷ experience training neural networks at fast-paced startups.

Boston, MA
mccarter.calvin@gmail.com
(616) 272-0909

EDUCATION

Carnegie Mellon University, Pittsburgh, PA

Ph.D. in Machine Learning **August 2013 - May 2019**

▷ Advisor: Seyoung Kim GPA: 3.80

▷ Selected Courses: Probabilistic Graphical Models, Convex Optimization, Statistical Machine Learning, Foundations of Machine Learning Theory

University of Michigan, Ann Arbor, MI

Bachelor of Science in Engineering **August 2009 - May 2013**

▷ Major: Computer Science, Minor: Mathematics GPA: 3.98

▷ Selected Courses: Operating Systems, Computer Architecture, Database Systems, Numerical Methods, Linear Algebra, Theoretical Statistics

PUBLICATIONS

CompArch
CompBio
ML

R Hanson, D Martin, C McCarter, J Paulson, "If Loud Aliens Explain Human Earliness, Quiet Aliens Are Also Rare." *The Astrophysical Journal (APJ)*, 2021.

LE Fernandes, . . . , C McCarter, et al., "Real-world Evidence of Diagnostic Testing and Treatment Patterns in US Breast Cancer Patients with Implications for Treatment Biomarkers from RNA-sequencing Data." *Clinical Breast Cancer*, 2020.

C McCarter, J Howrylak, S Kim, "Learning Gene Networks Underlying Clinical Phenotypes Using SNP Perturbations", *PLOS Computational Biology*, 2020.

C McCarter and S Kim, "Large-Scale Optimization Algorithms for Sparse Conditional Gaussian Graphical Models", *AISTATS*, 2016.

K Kandasamy, C McCarter. "Penalised Additive Least Squares Models for High Dimensional Nonparametric Regression and Function Selection", *Large-Scale Kernel Learning Workshop @ ICML*, 2015.

C McCarter and S Kim, "On Sparse Gaussian Chain Graph Models", *NeurIPS*, 2014.

S Moon, C McCarter, YH Kuo, "Active learning with partially featured data", *Proceedings of the 23rd International Conference on World Wide Web (WWW)*, 2014.

C McCarter, D Kletter, H Tang, K Partyka, Y Ma, S Singh, J Yadav, M Bern, B Haab, "Prediction of Glycan Motifs Using Quantitative Analysis of Multi-lectin Binding", *Proteomics Clinical Applications*, vol: 7, issue: 9-10, 2013.

D Chatterjee, C McCarter, V Bertacco, "Simulation-based Signal Selection for State Restoration in Silicon Debug", *ICCAD*, 2011.

UNPUBLISHED
MANUSCRIPTS

C McCarter, "Towards Backwards-Compatible Data with Confounded Domain Adaptation", arXiv preprint arXiv:2203.12720, *Under Review*, 2022.

A Basumallik, D Bunandar, N Dronen, L Levkova, C McCarter, L Nair, D Walter, D Widemann, "Adaptive Block Floating-Point for Analog Deep Learning Hardware", *Under Review.*

PATENTS

D Bunandar, C McCarter, A Basumallik, "Improving the accuracy of analog linear processor." US Provisional Patent 63/287,219 (2021).

J Michuda, . . . , C McCarter, et al., "Systems and methods for multilabel cancer classification." US Patent App. 17/150,992 (2021).

Independent Research   February 2022 - Present

Invented domain adaptation method which conditions on confounders, and wrote and submitted paper. Combining this with methods for cross-lingual word embeddings in low-resource languages, and then applying it to systems biology. Explored founding a startup bringing vector-space representation learning to biology, by enabling the alignment of embeddings across biological contexts.

Lightmatter, ML Scientist   January 2021 - February 2022

Researched ways to accelerate deep learning inference on photonic AI accelerator. Explored finetuning strategies to ensure model accuracy despite hardware noise and quantization. Helped guide development of next generation of hardware to improve noise robustness.

Tempus Labs, ML Scientist   June 2019 - January 2021

Created and validated a new batch effect correction method, which was deployed on the Tempus RNA-seq pipeline as the source-of-truth for all clinical AI models and pharma data deliveries. Developed a new topic model for gene expression deconvolution in metastatic cancers. Explored network learning methods and graph neural nets for gene expression networks and chromosomal rearrangement graphs.

Carnegie Mellon University, PhD Student   August 2013 - May 2019

Developed novel sparse graphical models and scalable optimization algorithms for disease systems biology. Used statistical learning to discover the gene regulatory networks which explain the effect of genetic variation on clinical traits.

Van Andel Research Institute, Research Intern   Summer 2013

Worked under the supervision of Brian Haab to apply feature selection method in pancreatic cancer biomarker discovery and to validate method on proteomics database.

Google, Software Engineering Intern   Summer 2012

Worked on server backend for Google Flight Search, developing functionality for international results for live Flight Search queries.

University of Michigan, Research Assistant   2011

Worked under the supervision of Valeria Bertacco and Debapriya Chatterjee to develop post-silicon validation method. Designed and implemented parallel algorithm in CUDA.

ConDo   https://github.com/calvinmccarter/condo-adapter
Toolbox for Confounded Domain Adaptation.   [author]

onnx2pytorch   https://github.com/ToriML/onnx2pytorch
Converts ONNX models to PyTorch.   [main contributor]

PerturbNet   https://github.com/SeyoungKimLab/PerturbNet
Learns multi-omic gene regulatory networks.   [author]

MLPerf Inference   https://github.com/mlcommons/inference
Deep learning benchmark. [memory-efficient pyramidal encoder for RNN-Transducer]

matrix-completion   https://github.com/tonyduan/matrix-completion
Classical matrix completion.   [incremental singular-vector thresholding]

PyTorch   https://github.com/pytorch/pytorch
Deep learning framework.   [added LazyInstanceNorm]

nanopq   https://github.com/matsui528/nanopq
Product quantization (PQ) and optimized PQ.   [eigenvalue allocation initialization]

Python (PyTorch, TensorFlow, pandas, NumPy, Numba), Matlab, C++, C, CUDA

PRESENTATIONS	*Transcriptome background tissue correction in metastatic cancers using a correlated composition admixture model.* American Association for Cancer Research (AACR), Annual Meeting 2020. *An efficient algorithm for learning a gene network underlying clinical phenotypes under SNP perturbations.* Genome Informatics meeting at Cold Spring Harbor Labs, November 2017.

TEACHING

Probabilistic Graphical Models (Teaching Assistant, CMU)	**Spring 2016**
Introduction to Machine Learning (Teaching Assistant, CMU)	**Fall 2015**

ACTIVITIES AND PROFESSIONAL SERVICE

Elicitation of latent knowledge (ELK) award contest	**February 2022**

AI Alignment Research Center (research proposal received honorable mention).

Paper reviewing	**June 2016 - Present**

Reviewer for *NeurIPS*, *IEEE Internet of Things*, *Statistics and Computing*, and *SciPy*.

University of Pittsburgh Biomedical Informatics Training Program	**2017**

Mentor to undergraduate research intern through iBRIC program.

Middle school science fair judging	**2015-2020**

Science fair judge for PA Junior Academy of Science and Chicago Public Schools.

Machine Learning Department Student Research Symposium	**2014**

Member of organizing committee. Created website and helped plan symposium.

English Language Institute Conversation Circle Program	**2011 - 2013**

Group leader of conversation circle for ESL students at University of Michigan.

University of Michigan Robocup (Robot Soccer) Team	**2009 - 2012**

Member and team leader (2010-2011). Developed computer vision subsystem.