Zachariah Carmichael, Ph.D.

zjcarmichael@gmail.com | +1.860.853.0228 | zcarmichael.com ⌐

Education

University of Notre Dame

Ph.D. in Computer Science & Engineering
May 2024 | Notre Dame, IN
GPA: 4.0

Rochester Institute of Technology

M.S. in Computer Engineering
May 2019 | Rochester, NY
GPA: 4.0 | *magna cum laude*

B.S. in Computer Engineering
May 2019 | Rochester, NY
GPA: 3.7 | *magna cum laude*

Skills

AI & ML

Explainable AI • Deep Learning • Computer Vision • Machine Learning • Trustworthy AI • Responsible AI • Interpretable ML • Convolutional Neural Networks (CNNs) • Deep Neural Networks (DNNs) • Large Language Models (LLMs) • Generative AI • Transformers • Image Processing Algorithms

Programming

Python • C/C++ • LaTeX • Bash/Shell • Julia • Java • Verilog • VHDL • JS/HTML/CSS

Software

PyTorch • TensorFlow • Keras • scikit-learn • OpenCV • SymPy • Git • SVN • Docker • Singularity • Matplotlib/Seaborn • MATLAB • ray • Django • ROS • SQL • Azure • AWS

OS

Arch Linux • Ubuntu/Debian • CentOS • NixOS • Linux • Windows

Other

System admin • Project management • Web dev

Links

⌐ **GitHub** | @ craymichael
⌐ **LinkedIn** | @ craymichael
⌐ **Google Scholar**

Service

- Computer Vision Foundation
- NeurIPS
- IEEE/CVF CVPR

Experience

Meta Platforms, Inc. | Research Scientist
July 2024 - Present | Menlo Park, CA

- Researched and developed engineer- and user-facing prompt explanation algorithms for generative AI models, including Meta AI and Llama.
- Researched and developed algorithms for estimating the influence of training data on generative AI models, including Meta AI and Llama.
- Developed an algorithm that accelerates model debugging by $10\times$ (more than $1 million annual cost reduction) via architecture-aware perturbation-based neuron attribution.
- Maintained and contributed to the open-source Captum library for model and data interpretability.

University of Notre Dame | Graduate Research Assistant
August 2019 - May 2024 | Notre Dame, IN

- Proposed a joint predictive and generative model for intrinsically interpretable classification. By combining concepts in explainable AI (prototypical neural networks) and generative AI (normalizing flows), we achieve **state-of-the-art accuracy, density estimation, calibration quality, and interpretability** across computer vision benchmarks. This work is under review.
- Researched and implemented an **intrinsically interpretable neural network** architecture to open the AI black box that is compatible with vision, language, and tabular data. My **patent-pending** approach based on prototypical part neural networks quantifiably **increased the interpretability of state-of-the-art convolutional and transformer neural networks by up to 25×** on fine-grained image classification tasks. This research was presented at WACV'24.
- Proposed a highly effective **defense for explainers against adversarial attacks** to identify malicious auditees at an **80% identification rate** and recover faithful explanations with **99% fidelity**. The results were presented at my talk at **AAAI'23**.
- Uncovered the **infidelity of post hoc explainable AI (XAI)** methods through rigorous research and analysis, exposing the limitations of black box interpretation in data-driven decision-making processes. Surprisingly, **popular XAI methods show up to a 70% accuracy decrease** on both synthetic and real-world scenarios. The results will be presented at a conference this year.
- Developed an efficient **open-source** symbolic framework in **Python** that enables researchers to study feature attribution, interaction effects, and explanations of arbitrarily complex scientific or data-driven models.
- **Computer Vision Research Lab** ⌐ member and advised by **Prof. Walter J. Scheirer** ⌐. Defended by Ph.D. dissertation in March 2024.

Hewlett Packard Enterprise Labs | Research Associate Intern
May 2023 - March 2024 | Milpitas, CA

- Developed a **3D image to image convolutional neural network** surrogate for a computational fluid dynamics solver to improve **data center sustainability**, achieving **2,000× speedup**. The surrogate is combined with online **reinforcement learning** for the optimization of the **carbon footprint** of data centers.
- Researched and developed **reinforcement learning** methods for the evaluation and enhancement of **natural adversarial robustness** in deep neural networks for vision. The framework identifies distortion-specific errors in models **up to 60% better** than the prior state-of-the-art.

Mitsubishi Electric Research Lab | PhD Research Intern
June 2022 - September 2022 | Boston, MA

- Conducted original research on intrinsically **human-interpretable AI** for (weakly supervised) vision tasks under supervision of **Dr. Mike Jones** ⌐ that led to a **patent** submission.
- Uncovered and **mitigated a fundamental shortcoming** of prototypical part neural networks (often applied to **high-stakes domains** including biomedical imaging) that can produce **highly misleading explanations** – the **solution** ⌐ that I researched and developed **improves interpretability by up to 25×**.
- Helped run a reading group for the state-of-the-art in computer vision that comprised over 20 researchers.

Lawrence Livermore National Lab | Graduate Student Intern
May 2021 - August 2021 | Remote

- Proposed a novel algorithm (XNAS) based on representation disentanglement for the optimization of the **accuracy-interpretability Pareto front via multi-objective neural architecture search (NAS)**. XNAS discovers architectures that are **3× better at error detection and 10% better calibrated**. Presented the research at **AutoML'23**.
- Scaled XNAS to a cluster of >**100 nodes** using `Ray` and **asynchronous algorithm design** with near-linear speedup.
- Employed a deep learning **object detection pipeline** (MaskRCNN and YOLO) for asteroids in Zwicky Transient Facility (ZTF) difference image data with detection **accuracy** >**90%**.

Neuromorphic AI Lab | Research Scientist
August 2019 - May 2021 | University of Texas at San Antonio, San Antonio, TX

- Maintained **cross-team collaborations** with epidemiologists and demographers in the epidemiological modeling of COVID-19 infectious spread. Researched and developed a **recurrent neural-compartmental epidemiological model** for forecasting COVID-19 spread in real-time. The research was presented at **IJCAI'20**.
- Developed and **deployed a real-time online dashboard to production** for Texas state showcasing case data and forecasts.
- Maintained a **cross-team collaboration** with medical professionals in an IRB-approved study of Alzheimer's patients in underrepresented communities in Texas. Researched methods for **deep learning applications to Alzheimer's detection** with EEG data rather than MRI due to cost and accessibility.

Neuromorphic AI Lab | Graduate Research Assistant
January 2018 - August 2019 | Rochester Institute of Technology, Rochester, NY

- The culmination of these works won me the **RIT Outstanding M.S. Thesis Award** for my thesis, "Towards Lightweight AI: Leveraging Stochasticity, Quantization, and Tensorization for Forecasting."
- Developed the **register-transfer-level (RTL) design** of generic bit width multiply-and-accumulate (MAC) components for the posit numerical system. Researched the accuracy-energy-latency trade-off of **model compression** via **low-precision arithmetic** and software-hardware co-design and development of a neural network on FPGAs. The implementation achieved up to **10% less accuracy degradation** than standard IEEE floating point with comparable power and latency. The now well-cited papers were presented and published at the venues **DATE'19 and CoNGA'19**.
- Improved **compute and memory efficiency** of neural networks for **time series forecasting** upwards of **95% in size and training speed** using randomness and model compression for **resource-constrained devices**.
- Researched the **Pareto front of design characteristics** of echo state networks for forecasting of spatiotemporal sequences with **state-of-the-art results** on forecasting tasks, (e.g., beating the previous Piano-midi.de 1-step ahead forecasting best by 1%).

Publications
H-Index: 12 | I10-Index: 12 | Citations: 518   †Paper | §Oral | ‡Poster

† **Z. Carmichael**, T. Redgrave, D. Gonzalez, W. J. Scheirer. "**This Probably Looks *Exactly* Like That: An Invertible Prototypical Network**." In *Proceedings of the European Conference on Computer Vision*, Milan, Italy, 2024. arXiv⬀

Z. Carmichael. "**Explainable AI for High-Stakes Decision-Making**." Ph.D. Dissertation, Department of Computer Science and Engineering, University of Notre Dame, 2024. Curate ND⬀

†‡ **Z. Carmichael**, S. Lohit, A. Cheerian, M. Jones, W. J. Scheirer. "**Pixel-Grounded Prototypical Part Networks**." In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, Waikoloa, Hawaii, USA, 2024. arXiv⬀

†‡ S. Sarkar, A. Ramesh Babu, S. Mousavi, **Z. Carmichael**, V. Gundecha, S. Ghorbanpour, R. Luna Gutierrez, A. Guillen, A. Naug. "**Benchmark Generation Framework with Custom Distortions for Image Classifier Robustness**." In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, Waikoloa, Hawaii, USA, 2024. arXiv⬀

†§‡ **Z. Carmichael**, W. J. Scheirer. "**Unfooling Perturbation-Based Post Hoc Explainers**." In *Proceedings of the AAAI Conference on Artificial Intelligence*, Washington D.C., USA, 2023. AAAI

Press ↗ | arXiv ↗

†‡ **Z. Carmichael**, W. J. Scheirer. "**How Well Do Feature-Additive Explainers Explain Feature-Additive Predictors?**." In *Proceedings of the NeurIPS Workshop XAI in Action: Past, Present, and Future Applications*, New Orleans, LA, 2023. arXiv ↗

S. Sarkar, A. Naug, **Z. Carmichael**, A Guillen, V. Gundecha, R. Luna, L. D. Kashyap, S. Ghorbanpour, D. Markovikj, S. Mousavi, A. Ramesh-Babu. "**CFD Surrogates for Data Center Sustainability Using 3D Convolutional Models**." *Preprint* (Under Review), –, 2023. arXiv ↗

S. Sarkar, A. Naug, **Z. Carmichael**, V. Gundecha, A Guillen, R. Luna, A. Ramesh-Babu. "**Enhancing Data Center Sustainability with a 3D CNN-Based CFD Surrogate Model**." In *Proceedings of the NeurIPS Workshop Tackling Climate Change with Machine Learning*, New Orleans, LA, 2023. arXiv ↗

†§‡ **Z. Carmichael**, T. Moon, S. A. Jacobs. "**Learning Debuggable Models Through Multi-Objective Neural Architecture Search**." *International Conference on Automated Machine Learning (AutoML) Workshop*, Potsdam/Berlin, Germany, 2023. arXiv ↗

S. J. Abraham, K. D. G. Maduranga, J. Kinnison, **Z. Carmichael**, J. D. Hauenstein, W. J. Scheirer. "**HomOpt: A Homotopy-Based Hyperparameter Optimization Method**." *Preprint* (Under Review), –, 2023. arXiv ↗

†§‡ W. Theisen, D. Gonzalez, **Z. Carmichael**, T. Weninger, W. J. Scheirer. "**Motif Mining: Finding and Summarizing Remixed Image Content**." In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, Hawaii, USA, 2023. arXiv ↗

†‡ J. Takeshita, **Z. Carmichael**, R. Karl, T. Jung. "**TERSE: Tiny Encryptions and Really Speedy Execution for Post-Quantum Private Stream Aggregation**." In *EAI International Conference on Security and Privacy in Communication Networks (SecureComm)*, Kansas City, USA, 2022. IACR Cryptology ePrint Archive ↗

†§ S. Abraham, **Z. Carmichael**, S. Banerjee, R. VidalMata, A. Agrawal, M. N. Al Islam, W. Scheirer, J. Cleland-Huang. "**Adaptive Autonomy in Human-on-the-Loop Vision-Based Robotics Systems**." In *1st Workshop on AI Engineering – Software Engineering for AI (WAIN'21)*, Remote, 2021. arXiv ↗

†‡ H. Langroudi, V. Karia, **Z. Carmichael**, A. Zyarah, T. Pandit, J. L. Gustafson, D. Kudithipudi. "**ALPS: Adaptive Quantization of Deep Neural Networks With GeneraLized PositS**." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Remote, 2021. CVF Open Access ↗

† N. Soures, D. Chambers, **Z. Carmichael**, A. Daram, D. P. Shah, K. Clark, L. Potter, D. Kudithipudi. "**SIRNet: Understanding Social Distancing Measures with Hybrid Neural Network Model for COVID-19 Infectious Spread**." In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI) Disease Computational Modeling Workshop*, Remote, 2020. IJCAI DCM ↗ | arXiv (out-of-date) ↗

‡ —. "**—**." In *Proceedings of the International Conference on Machine Learning (ICML) Machine Learning for Global Health Workshop*, Remote, 2020. Poster ↗

†§ **Z. Carmichael**, D. Kudithipudi. "**Stochastic Tucker-Decomposed Recurrent Neural Networks for Forecasting**." In *IEEE Global Conference on Signal and Information Processing (GlobalSIP 2019)*, Ottawa, Canada, 2019. IEEE Xplore® ↗

†§ **Z. Carmichael**. "**Towards Lightweight AI: Leveraging Stochasticity, Quantization, and Tensorization for Forecasting**." Master's Thesis (*Won the 2019 RIT Outstanding Master's Thesis Award*), Department of Computer Engineering, Rochester Institute of Technology, 2019. RIT Scholar Works ↗

†§ **Z. Carmichael**, H. F. Langroudi, C. Khazanov, J. Lillie, J. L. Gustafson, D. Kudithipudi. "**Deep Positron: A Deep Neural Network Using the Posit Number System**." In *Proceedings of the IEEE Conference and Exhibition on Design, Automation and Test in Europe (DATE)*, Florence, Italy, March 25-29, 2019. IEEE, 1421–1426. IEEE Xplore® ↗ | arXiv ↗

†§ **Z. Carmichael**, H. F. Langroudi, C. Khazanov, J. Lillie, J. L. Gustafson, D. Kudithipudi. "**Performance-Efficiency Trade-off of Low-Precision Numerical Formats in Deep Neural Networks**." In *Proceedings of the ACM Conference for Next Generation Arithmetic (CoNGA)*, Singapore, 2019. ACM DL ↗ | arXiv ↗

H. F. Langroudi, **Z. Carmichael**, J. L. Gustafson, D. Kudithipudi. "***Cheetah*: Mixed Low-Precision Hardware & Software Co-Design Framework for DNNs on the Edge**." *arXiv*, Preprint, 2019. arXiv ↗

†§H. F. Langroudi, **Z. Carmichael**, J. L. Gustafson, D. Kudithipudi. "**PositNN Framework: Tapered Precision Deep Learning Inference for the Edge**." In *Proceedings of the Twelfth IEEE Space Computing Conference (SCC 2019* ↗*)*, Pasadena, CA, July 30-August 1, 2019. IEEE, 53–59. IEEE Xplore® ↗

†§**Z. Carmichael**, H. Syed, D. Kudithipudi. "**Analysis of Wide and Deep Echo State Networks for Multiscale Spatiotemporal Time Series Forecasting**." In *ACM International Conference Proceedings Series (ICPS) of the Neuro Inspired Computational Elements (NICE) Workshop*, Albany, NY, 2019. ACM DL ↗ | arXiv ↗

†§**Z. Carmichael**, B. Glasstone, F. Cwitkowitz, K. Alexopoulos, R. Relyea, R. Ptucha. "**Autonomous Navigation Using Localization Priors, Sensor Fusion, and Terrain Classification**." In *Proceedings of IS&T International Symposium on Electronic Imaging: Autonomous Vehicles and Machines*, San Francisco, CA, 2019. Ingenta Connect ↗

†‡**Z. Carmichael**, H. Syed, S. Burtner, D. Kudithipudi. "**Mod-DeepESN: Modular Deep Echo State Network**." In *Annual Conference on Cognitive Computational Neuroscience*, Philadelphia, PA, 2018. CCN (out-of-date) ↗ | arXiv ↗

Other Publications

Z. Carmichael. "**Red Flags in the AI Executive Order about "Dual Use" Models**." *Digital Spirits Substack*, Online, 2023. DigitalSpirits ↗

Z. Carmichael. "**Demystifying ChatGPT and Other Large Language Models**." *Digital Spirits Substack*, Online, 2023. DigitalSpirits ↗

Z. Carmichael. "**Noncompliance in Algorithmic Audits and Defending Auditors**." *Medium*, Online, 2023. Medium ↗

Grants & Fellowships

NSF Graduate Fellowships Research Program (GRFP) Honorable Mention
2020

University of Notre Dame Jack and Mary Ann Remick Fellowship in Engineering 2019-2024

University of Notre Dame Kilgallon Family Graduate Fellowship 2019-2024

Honors & Awards

RIT Outstanding M.S. Thesis Award 2019
Thesis: "Towards Lightweight AI: Leveraging Stochasticity, Quantization, and Tensorization for Forecasting"

UTSA Best Poster: Fundamental Research in AI (Ph.D.) 2019
Poster: "Cheetah: Mixed Low-Precision Hardware & Software Co-Design Framework for DNNs on the Edge"

RIT KGCOE Dean's List 2014-2019

RIT Presidential Scholarship 2014-2019

RIT BS/MS Tuition Award 2014-2019

RIT Excellence in Computing 2014

Projects

Black & White Video Enhancement: Amahl & the Night Visitors 2022
video super-resolution • video colorization • video denoising • computer vision • deep learning

NFL Betting App with Betting AI 2020
deep learning • probability • web development • databases

CNNs for Loop-Closure Detection in vSLAM Systems 2018-2019

SLAM • autonomous driving • convolutional neural networks

Autonomous Golf Cart – "Tiger Taxi" 2018

SLAM • autonomous driving • convolutional neural networks • ROS • timed-elastic-band navigation and control

Segmentation of Histopathological Images Using U-Net 2018

biomedical imaging • convolutional neural networks • image to image translation • image segmentation • morphological operations

Other Experience

Computer Vision Foundation | Web Manager

September 2019 - June 2024 | Remote

- Position funded my PhD
- Rewrote, audited, & actively maintain CVF Open Access ⬀ to better serve papers, talks, posters, & other open content from the CVPR, ICCV, ECCV, & WACV conferences to *500,000+* monthly visitors
- Automated synchronization of CVF COVE ⬀ computer vision datasets & arXiv erratum retrieval with Open Access
- Discovered & mitigated several SQL security vulnerabilities

University of Notre Dame | Graduate Teaching Assistant

August 2019 - May 2020 | Notre Dame, IN

- Courses: *Advanced Topics in Machine Learning (ML)* (20+ students | Graduate CS); *Theory of Computing* (30+ students | Upper-level undergraduate CS)
- Taught students core CS & ML concepts in office hours, scoped assignments, held review sessions, & graded exams

Plexus Corp. | Digital Engineering Intern

June 2017 - August 2017 | Raleigh, NC

- Carried out RTL design of FPGA-agnostic module for evaluation of FPGA cooling systems, validated all test cases with digital engineering team
- Developed embedded software for a battery testing unit using the FRDM-K64F dev board, validated design & integration with mechanical, electrical, & software teams

CUBRC, Inc. | Research/Software Engineering Intern

June 2016 - December 2016 | Cheektowaga, NY

- Developed a machine learning framework to model surgery risk, patient mortality, & other analytics using `TensorFlow` & `scikit-learn` with automatic model search & hyperparameter optimization
- Worked with customers in the design of electronic health record-unifying database & interface

Membership

Institute of Electrical and Electronics Engineers (IEEE**) Student** 2018 - Now

Tau Beta Pi – The Engineering Honor Society ($\mathrm{TB\Pi}$**)** 2018 - Now

The National Society of Leadership and Success ($\mathrm{\Sigma A\Pi}$**)** 2018 - Now

Sigma Xi Nomination ($\mathrm{\Sigma\Xi}$**)** 2020