Landon Butler

landonbutler.github.io ◇ landonb@berkeley.edu

Research Interests

My research focuses on developing trustworthy machine learning, with an emphasis on methods that interpret and explain the complex decision-making processes of foundation models, using techniques from signal processing and game theory.

Education

University of California, Berkeley
Ph.D. candidate in Electrical Engineering and Computer Science 2027
Advised by Prof. Kannan Ramchandran

University of Pennsylvania
M.S.E. in Data Science 2022
Thesis: Convolutional Learning on Multigraphs
Advised by Prof. Alejandro Ribeiro

University of Pennsylvania
B.S.E. in Systems Engineering 2022
Concentration: Artificial Intelligence and Data Science
Minors: Computer Science, Mathematics

Recent Publications

ProxySPEX: Inference-Efficient Interpretability via Sparse Feature Interactions in LLMs
NeurIPS, 2025 (**Spotlight**)
Landon Butler*, Abhineet Agarwal*, Justin Singh Kang*, Yigit Efe Erginbas, Kannan Ramchandran, Bin Yu

SPEX: Scaling Feature Interaction Explanations for LLMs
ICML, 2025
Justin Singh Kang*, Landon Butler*, Abhineet Agarwal*, Yigit Efe Erginbas, Ramtin Pedarsani, Kannan Ramchandran, Bin Yu

Learning to Understand: Identifying Interactions via the Mobius Transform
NeurIPS, 2024
Justin Singh Kang, Yigit Efe Erginbas, Landon Butler, Ramtin Pedarsani, Kannan Ramchandran

Fellowships

NSF Graduate Research Fellowship 2022
Littlejohn Fellowship, *University of Pennsylvania* 2021

Internships

Ph.D. Machine Learning Intern at Apple, Summer 2025
Researched machine learning for communication systems

Ph.D. Software Engineering Intern at Uber AI, Summer 2024
Researched text embedding models for use in search and relevance tasks

Teaching

Graduate Student Instructor, *University of California, Berkeley EECS Department*
- Signals and Systems, Fall 2024

Teaching Assistant, *University of Pennsylvania ESE Department*
- Statistics for Data Science, Spring 2021, Summer 2021
- Graph Neural Networks, Fall 2021
- Foundations of Data Science, Fall 2021

Publications

Conference Papers

1. *ProxySPEX: Inference-Efficient Interpretability via Sparse Feature Interactions in LLMs*
   NeurIPS, 2025 (**Spotlight**)
   Landon Butler*, Abhineet Agarwal*, Justin Singh Kang*, Yigit Efe Erginbas, Kannan Ramchandran, Bin Yu
2. *SPEX: Scaling Feature Interaction Explanations for LLMs*
   ICML, 2025
   Justin Singh Kang*, Landon Butler*, Abhineet Agarwal*, Yigit Efe Erginbas, Ramtin Pedarsani, Kannan Ramchandran, Bin Yu
3. *Learning to Understand: Identifying Interactions via the Mobius Transform*
   NeurIPS, 2024
   Justin Singh Kang, Yigit Efe Erginbas, Landon Butler, Ramtin Pedarsani, Kannan Ramchandran
4. *Non Commutative Convolutional Signal Models in Neural Networks: Stability to Small Deformations*
   ICASSP, 2024
   Alejandro Parada-Mayorga, Landon Butler, and Alejandro Ribeiro
5. *Learning with Multigraph Convolutional Filters*
   ICASSP, 2023
   Landon Butler, Alejandro Parada-Mayorga, and Alejandro Ribeiro
6. *Democratizing Aviation Emissions Estimation: Development of an Open-Source, Data-Driven Methodology*
   ICRAT, 2022
   Andy Eskenazi, Landon Butler, Arnav Joshi, and Megan Ryerson
7. *Learning Connectivity for Data Distribution in Robot Teams*
   IROS, 2021
   Ekaterina Tolstaya, Landon Butler, Daniel Mox, James Paulos, Vijay Kumar, and Alejandro Ribeiro

Journal Publications

1. *Convolutional Learning on Multigraphs*
   IEEE Transactions on Signal Processing, 2023
   Landon Butler, Alejandro Parada-Mayorga, and Alejandro Ribeiro
2. *Convolutional Filtering and Neural Networks with Non-Commutative Algebras*
   IEEE Transactions on Signal Processing, 2023
   Alejandro Parada-Mayorga, Landon Butler, and Alejandro Ribeiro
3. *Equitable Optimization of U.S. Airline Route Networks*
   Computers, Environment and Urban Systems, 2023
   Andy Eskenazi, Arnav Joshi, Landon Butler, and Megan Ryerson