Juhan Bae

CONTACT
INFORMATION

✉ baejuhan21@gmail.com   ⏻ @pomonam
🏠 juhanbae.com   🐦 @juhan_bae

EDUCATION   **University of Toronto**, Toronto, ON, Canada

Ph.D., Computer Science   Sep. 2019 - Mar. 2025
- Advisor: Roger Grosse
- Thesis: *Beyond Gradients: Using Curvature Information for Deep Learning*

B.Sc. Hons., Computer Science and Statistics   Sep. 2015 - Nov. 2019
- Graduated with High Distinction (cGPA = 3.98/4.00)
- Focus in Artificial Intelligence

PEER-
REVIEWED
CONFERENCE
PUBLICATIONS

[1] **Accelerating Neural Network training: An Analysis of the AlgoPerf Competition**
Priya Kasimbeg, Frank Schneider, Runa Eschenhagen, Juhan Bae, Chandramouli Shama
Sastry, Mark Saroufim, Boyuan Feng, Less Wright, Edward Yang, Zachary Nado, Sourabh
Medapati, Philipp Hennig, Michael Rabbat, George Dahl.
International Conference on Learning Representations (ICLR 2025), Singapore.

[2] **What Kind of Pretraining Data Do Large Language Models Rely on When Doing Reasoning?**
Laura Ruis, Maximilian Mozes, Juhan Bae, Siddhartha Kamalakara, Dwaraknath Gnaneshwar, Acyr Locatelli, Robert Kirk, Tim Rocktaschel, Edward Grefenstette, Max Bartolo.
International Conference on Learning Representations (ICLR 2025), Singapore.

[3] **Influence Functions for Scalable Data Attribution in Diffusion Models**
Bruno Mlodozeniec, Runa Eschenhagen, Juhan Bae, Alexander Immer, David Krueger,
Richard Turner.
International Conference on Learning Representations (ICLR 2025), Singapore.
Oral Presentation — Acceptance Rate = 207/11672 ≈ 1.8%

[4] **Training Data Attribution via Approximate Unrolled Differentiation**
Juhan Bae, Wu Lin, Jonathan Lorraine, Roger Grosse.
Advances in Neural Information Processing Systems (NeurIPS 2024), Vancouver, Canada.

[5] **Can We Remove the Square-Root in Adaptive Gradient Methods?**
Wu Lin, Felix Dangel, Runa Eschenhagen, Juhan Bae, Richard Turner, Alireza Makhzani.
International Conference on Machine Learning (ICML 2024), Vienna, Austria.

[6] **Efficient Parametric Approximations of Neural Network Function Space Distance**
Nikita Dhawan, Sheldon Huang, Juhan Bae, Roger Grosse.
International Conference on Machine Learning (ICML 2023), Hawaii, USA.

[7] **Multi-Rate VAE: Train Once, Get the Full Rate-Distortion Curve**
Juhan Bae, Michael Zhang, Michael Ruan, Eric Wang, So Hasegawa, Jimmy Ba, Roger
Grosse.
International Conference on Learning Representations (ICLR 2023), Kigali, Rwanda.
Oral Presentation — Acceptance Rate = 91/4956 ≈ 1.8%

[8] **If Influence Functions are the Answer, Then What is the Question?**
Juhan Bae, Nathan Ng, Alston Lo, Marzyeh Ghassemi, Roger Grosse.
Advances in Neural Information Processing Systems (NeurIPS 2022), Louisiana, USA.

[9] **Amortized Proximal Optimization**
Juhan Bae*, Paul Vicol*, Jeff Z. HaoChen, Roger Grosse. (*: Equal Contributions)
Advances in Neural Information Processing Systems (NeurIPS 2022), Louisiana, USA.

[10] **Analyzing Monotonic Linear Interpolation in Neural Network Loss Landscapes**
James Lucas, Juhan Bae, Michael Zhang, Stanislav Fort, Richard Zemel, Roger Grosse.
The International Conference on Machine Learning (ICML 2021), Virtual.

[11] **Delta-STN: Efficient Bilevel Optimization for Neural Networks using Structured Response Jacobians**
Juhan Bae, Roger Grosse.
Advances in Neural Information Processing Systems (NeurIPS 2020), Virtual.

[12] **Fast 6DoF Pose Estimation with Synthetic Textureless CAD model for Mobile Applications**
Bowen Chen, Juhan Bae, Dibyendu Mukherjee.
International Conference on Image Processing (ICIP 2019), Taipei, Taiwan.

PEER-
REVIEWED
WORKSHOP
PUBLICATIONS

[13] **Influence Functions for Scalable Data Attribution in Diffusion Models**
Bruno Mlodozeniec, Runa Eschenhagen, Juhan Bae, Alexander Immer, David Krueger, Richard Turner.
Attributing Model Behavior at Scale (NeurIPS 2024 Workshop), Vancouver, Canada.

[14] **Using Large Language Models for Hyperparameter Optimization**
Michael Zhang, Nishkrit Desai, Juhan Bae, Jonathan Lorraine, Jimmy Ba.
Foundation Models for Decision Making (NeurIPS 2023 Workshop), Louisiana, USA.

[15] **Monotonic Linear Interpolation of Neural Network Parameters**
James Lucas, Juhan Bae, Michael Zhang, Richard Zemel, Jimmy Ba, Roger Grosse.
Optimization for Machine Learning (NeurIPS 2020 Workshop), Virtual.

[16] **Eigenvalue Corrected Noisy Natural Gradient**
Juhan Bae, Guodong Zhang, Roger Grosse.
Bayesian Deep Learning (NeurIPS 2018 Workshop), Montreal, Canada.

[17] **Learnable Pooling Methods for Video Classification**
Sebastian Kmiec, Juhan Bae, Ruijian An.
Large-Scale Video Understanding (ECCV 2018 Workshop), Munich, Germany.
Oral Presentation

TECHNICAL
REPORTS

[18] **Training Data Attribution (TDA): Examining Its Adoption & Use Cases**
Deric Cheng, Juhan Bae, Justin Bullock, David Kristofferson. 2025.

[19] **Studying Large Language Model Generalization with Influence Functions**
Roger Grosse*, Juhan Bae*, Cem Anil*, Nelson Elhage, Alex Tamkin, Amirhossein Tajdini, Benoit Steiner, Dustin Li, Esin Durmus, Ethan Perez, Evan Hubinger, Kamilė Lukošiūtė, Karina Nguyen, Nicholas Joseph, Sam McCandlish, Jared Kaplan, Samuel Bowman (*: Equal Contributions). 2023.

[20] **Benchmarking Neural Network Training Algorithms**
George Dahl*, Frank Schneider*, Zachary Nado*, Naman Agarwal*, Chandramouli Shama Sastry[†], Philipp Hennig[†], Sourabh Medapati[†], Runa Eschenhagen[†], Priya Kasimbeg[†], Daniel Suo[†], Juhan Bae[†], Justin Gilmer[†], Abel Peirson[†], Bilal Khan[†], Rohan Anil[†], Mike Rabbat[†], Shankar Krishnan[‡], Daniel Snider[‡], Ehsan Amid[‡], Kongtao Chen[‡], Chris Maddison[‡], Rakshith Vasudev[‡], Michal Badura[‡], Ankush Garg[‡], Peter Mattson[‡] (*, †, ‡: Equal Contributions). 2023.

PREPRINTS

[21] **What's Your Data Worth to GPT? LLM-Scale Data Valuation with Influence Functions**
Sang Keun Choe, Hwijeen Ahn*, Juhan Bae*, Kewen Zhao*, Minsoo Kang, Youngseog Chung, Adithya Pratapa, Willie Neiswanger, Emma Strubell, Teruko Mitamura, Jeff Schneider, Eduard Hovy, Roger Grosse, Eric P. Xing (*: Equal Contributions). 2024.

[22] **Spectral-factorized Positive-definite Curvature Learning for NN Training**
Wu Lin, Felix Dangel, Runa Eschenhagen, Juhan Bae, Richard Turner, Roger Grosse. 2024.

PROFESSIONAL EXPERIENCES	**Anthropic** *Resident*	Feb. 2023 - Aug. 2023
	Microsoft Research *Research Intern*	Jun. 2021 - Aug. 2021
	Epson Research *Software Developer - Algorithm Research*	May. 2017 - Mar. 2019

GRANTS AND AWARDS	Vector Institute Research Grant	2020 - 2024
	Expert Reviewer at ICML	2021
	Top Reviewer at NeurIPS	2020, 2024
	Top Reviewer at ICML	2020
	Faculty of Arts & Science Fellowship	2020 - 2023
	St.Michael's College Silver Medal	2019
	St.Micheal's College Scholarship	2017, 2018
	Dean's List Scholar	2016 - 2019

TEACHING   **University of Toronto**, Toronto, ON

Instructor

• TUSK (Machine Learning Software Foundations)	2022
• CSC311 (Introduction to Machine Learning)	2020

Teaching Assistant

• CSC2547 (AI Alignment)	2024
• CSC110 (Foundations of Computer Science I)	2023, 2021
• CSC2702 (Technical Entrepreneurship)	2022
• HLP101 (Undergraduate CS Course Help Centre)	2022
• CSC2541 (Neural Network Training Dynamics)	2022
• STA314 (Statistical Methods for Machine Learning I)	2021
• CSC320 (Introduction to Visual Computing)	2021
• CSC412 (Probabilistic Learning and Reasoning)	2020
• CSC165 (Mathematical Expression and Reasoning for CS)	2019, 2016

Vector Institute, Toronto, ON

Teaching Assistant

• AI Certificate: Deep Learning 2	2020
• AI Certificate: Deep Learning 1	2019

PATENTS   [23] **Methods and Systems for Training an Object Detection Algorithm using Synthetic Images**
Dibyendu Mukherjee, Bowen Chen, Juhan Bae
US11107241B2, 2021

SERVICES   *Journal Reviewer*

- Transactions on Machine Learning Research (TMLR)

Conference Reviewer

- Conference on Neural Information Processing Systems (NeurIPS)
- Society for Artificial Intelligence and Statistics (AISTATS)
- International Conference on Learning Representations (ICLR)
- International Conference on Machine Learning (ICML)
- Association for the Advancement of Artificial Intelligence (AAAI)

Workshop Reviewer

- Distribution Shifts (NeurIPS)
- Attributing Model Behavior at Scale (NeurIPS)
- Tiny Papers Showcase Day (ICLR)

INVITED TALKS	[1] Training Data Attribution with Unrolled Differentiation. Data Attribution Reading Group. Illinois, USA. 2024.
	[2] LLM-Scale Data Valuation with Influence Functions. xAI. California, USA. 2024
	[3] Tutorial on Training Data Attribution. Inria Soda. Rocquencourt, France. 2024.
	[4] LLM-Scale Data Valuation with Influence Functions. Google Brain. Massachusetts, USA. 2024
	[5] Studying Large Language Model Generalization with Influence Functions. Future of Life Institute. California, USA. 2024.
	[6] Studying Large Language Model Generalization with Influence Functions. Guest Lecture for Stanford CS329. California, USA. 2023.
	[7] Studying Large Language Model Generalization with Influence Functions. AI Safety Reading Group (University of Toronto). Ontario, Canada. 2023.
	[8] Studying Large Language Model Generalization with Influence Functions. Mechanistic Interpretability Reading Group (University of Ottawa). Remote Presentation. 2023.
	[9] Multi-Rate VAE: Train Once, Get the Full Rate-Distortion Curve. Vector Institute. Ontario, Canada. 2022.
	[10] Tutorial on Neural Architecture Search (NAS). University of Toronto & Vector Institute. Remote Presentation. 2020.

SUPERVISIONS		
Alston Lo (UofT undergraduate → Ph.D. student at MIT)		2022
Eric Wang (UofT undergraduate → Software developer at Clearpath)		2022
Michael Ruan (UofT undergraduate → Software developer at AMD)		2022
Yuchen Wang (UofT undergraduate → Masters student at Stanford)		2020