

SHENGYANG SUN

ssydasheng@gmail.com Toronto, ON, Canada
[Personal Page](#), [Google Scholar](#), [Linkedin](#), [Github](#)

EXPERIENCE

Present: NVIDIA Remote, Canada
Applied Scientist, Deep Learning Oct.2023 - present

- Conduct applied research on the alignment of large language models.

Amazon Toronto, Canada
Applied Scientist II Aug.2022 - Oct.2023

- Utilize large language models (LLM) to retrieve relevant ads at a massive scale.
- Develop fast predictive models for early-stage ads selection using LLM embeddings.

DeepMind Remote, Canada
Research Scientist Intern Jun. 2021 - Oct. 2021

- Lead a research project focused on efficient memory management in online systems to enable continuously learning new abilities without forgetting old ones.

Google AI Beijing, China
Research Scientist Intern Jun. 2018 - Sept. 2018

- Worked on calibrating language models in machine translation, specifically examining their ability to assess uncertainty in predictions and developing enhancements.

Duke University Durham, USA
Research Scholar Jul. 2016 – Aug. 2016

- Combined neural networks with Bayesian models to quantify predictive uncertainty.

EDUCATION

University of Toronto (Department of Computer Science) Toronto, ON, Canada
PhD, Advised by Roger Grosse Sept. 2017 – Nov. 2022

Tsinghua University (Department of Electronic Engineering) Beijing, China
Bachelor of Engineering Sept. 2013 – Jul. 2017

AWARDS

- Top Reviewers NeurIPS2020, ICLR2021
- Borealis AI Global Fellowship Award 2019
- Connaught New Researcher Award 2017
- Connaught International Scholarship (University-wide 20) 2017-2022
- First Class Undergraduate Scholarship 2014, 2015, 2016
- Silver Medal of Chinese Mathematics Olympics 2012

Python, Tensorflow, PyTorch, Jax, PySpark, AWS Tools

TALKS

-
- Structured Inter-domain Inducing points @Gaussian process seminar, 2022
 - Information-theoretic Online Memory Selection @ Tsinghua University, 2021
 - Harmonic variational Gaussian processes @ AI TIME, 2021
 - A Benchmark for Predictive Posterior Correlations @ SIAM-UQ Symposium, 2020
 - Kernel Implicit Variational Inference @ Tsinghua University, 2018
 - Functional variational Bayesian neural networks @ Google Toronto, 2018

PUBLICATIONS

[Google Scholar](#)

Peer-reviewed Publications and Preprints

1. **Shengyang Sun** et al., Nemotron-4 340B Technical Report. 2024. <https://arxiv.org/pdf/2406.11704>
2. Gerald Shen, Zhilin Wang, Olivier Delalleau, Jiaqi Zeng, Yi Dong, Daniel Egert, **Shengyang Sun**, Jimmy Zhang, Sahil Jain, Ali Taghibakhshi, Markel Sanz Ausin, Ashwath Aithal, Oleksii Kuchaiev. NeMo-Aligner: Scalable Toolkit for Efficient Model Alignment. COLM 2024.
3. **Shengyang Sun**, Daniele Calandriello, Huiyi Hu, Ang Li, Michalis Titsias. Information-theoretic Online Memory Selection for Continual Learning. ICLR 2022.
4. Jimmy Ba, Murat A Erdogdu, Marzyeh Ghassemi, **Shengyang Sun**, Taiji Suzuki, Denny Wu, Tianzong Zhang. Understanding the Variance Collapse of SVGD in High Dimensions. ICLR 2022.
5. **Shengyang Sun**, Jiaxin Shi, Andrew Gordon Wilson, Roger Grosse. Scalable Variational Gaussian Processes via Harmonic Kernel Decomposition. ICML 2021.
6. **Shengyang Sun***, Jiaxin Shi*, Roger Grosse. Neural Networks as Inter-domain Inducing Points. AABI 2021 Symposium.
7. Chaoqi Wang*, **Shengyang Sun***, Roger Grosse. Beyond Marginal Uncertainty: How Accurately can Bayesian Regression Models Estimate Posterior Predictive Correlations? AISTATS 2021 [Oral Presentation].
8. Jun Yang*, **Shengyang Sun***, Daniel Roy. Fast-rate PAC-Bayes Generalization Bounds via Shifted Rademacher Processes. NeurIPS 2019.
9. **Shengyang Sun***, Guodong Zhang*, Jiaxin Shi*, Roger Grosse. Functional variational Bayesian neural networks. ICLR 2019.
10. James Lucas, **Shengyang Sun**, Richard Zemel, Roger Grosse. Aggregated Momentum: Stability Through Passive Damping. ICLR 2019.
11. **Shengyang Sun**, Guodong Zhang, Chaoqi Wang, Wenyuan Zeng, Jiaman Li, and Roger Grosse. Differentiable compositional kernel learning for Gaussian processes. ICML 2018.
12. Guodong Zhang*, **Shengyang Sun***, Roger Grosse. (2017). Natural Gradient as Stochastic Variational Inference. ICML 2018.
13. Jiaxin Shi, **Shengyang Sun**, Jun Zhu. (2017). A Spectral Approach to Gradient Estimation for Implicit Distributions. ICML 2018.
14. Jiaxin Shi*, **Shengyang Sun***, Jun Zhu. (2017). Kernel Implicit Variational Inference. ICLR 2018.
15. Jiaxin Shi, Jianfei Chen, Jun Zhu, **Shengyang Sun**, Yucen Luo, Yihong Gu, Yuhao Zhou. ZhuSuan: A library for Bayesian deep learning. <https://arxiv.org/abs/1709.05870>.

16. **Shengyang Sun**, Changyou Chan and Lawrence Carin. (2016). Learning Structured Weight Uncertainty in Bayesian Neural Networks. AISTATS 2017.

Note: * represents equal contribution.