Dustin Tran

Research Scientist
Google Brain
Mountain View, CA

trandustin@google.com
http://www.dustintran.com/

Education

Ph.D. Computer Science, Columbia University Advisors: David M. Blei, Andrew Gelman	2016–2020
Ph.D. Statistics, Harvard University (transferred) M.S. Computational Science & Engineering, Harvard University Advisor: Edoardo M. Airoldi	2014–2015
B.A. (Hon.) Mathematics, Statistics, University of California, Berkeley	2010–2014

Employment

Research Scientist Google Brain	2018–
Research Intern Google	Oct 2017 – Jan 2018
Research Intern OpenAI	May 2017 – Oct 2017

Awards

John M. Chambers Statistical Software Award (for Edward)	2018
Google Ph.D. Fellowship in Machine Learning ($34,000 + tuition/fees)	2017–2020
Columbia SEAS Fellowship (Full funding)	2016–20
Adobe Research Fellowship ($10,000)	2016
LinkedIn Economic Graph Challenge	2015
Harvard GSAS Fellowship (Full funding)	2015
Dorothea Klumpke Roberts Prize in Mathematics	2014
Regents' and Chancellor's Scholarship (Full funding)	2010–2014
Cal Alumni Leadership Scholarship ($2,500)	2010

Publications

PREPRINTS

1. J. Wei, J. Wei, Y. Tay, **D. Tran**, A. Webson, Y. Lu, X. Chen, H. Liu, D. Huang, D. Zhou, and others. Larger language models do in-context learning differently. 2023.

2. M. Dehghani, J. Djolonga, B. Mustafa, P. Padlewski, J. Heek, J. Gilmer, A. Steiner, M. Caron, R. Geirhos, I. Alabdulmohsin, R. Jenatton, L. Beyer, M. Tschannen, A. Arnab, X. Wang, C. Riquelme, M. Minderer, J. Puigcerver, U. Evci, M. Kumar, S. van Steenkiste, G. F. Elsayed, A. Mahendran, F. Yu, A. Oliver, F. Huot, J. Bastings, M. P. Collier, A. Gritsenko, V. Birodkar, C. Vasconcelos, Y. Tay, T. Mensink, A. Kolesnikov, F. Pavetić, **D. Tran**, T. Kipf, M. Lučić, X. Zhai, D. Keysers, J. Harmsen, and N. Houlsby. Scaling vision transformers to 22 billion parameters. 2023.

3. J. U. Allingham, J. Ren, M. W. Dusenberry, J. Z. Liu, X. Gu, Y. Cui, **D. Tran**, and B. Lakshminarayanan. A simple zero-shot prompt weighting technique to improve prompt ensembling in text-image models. 2023.

4. **D. Tran**, J. Liu, M. W. Dusenberry, D. Phan, M. Collier, J. Ren, K. Han, Z. Wang, Z. Mariet, H. Hu, N. Band, T. G. J. Rudner, K. Singhal, Z. Nado, J. van Amersfoort, A. Kirsch, R. Jenatton, N. Thain, H. Yuan, K. Buchanan, K. Murphy, D. Sculley, Y. Gal, Z. Ghahramani, J. Snoek, and B. Lakshminarayanan. Plex: Towards reliability using pretrained large model extensions. 2022.

5. Z. Nado, N. Band, M. Collier, J. Djolonga, M. W. Dusenberry, S. Farquhar, Q. Feng, A. Filos, M. Havasi, R. Jenatton, G. Jerfel, J. Liu, Z. Mariet, J. Nixon, S. Padhy, J. Ren, T. G. J. Rudner, F. Sbahi, Y. Wen, F. Wenzel, K. Murphy, D. Sculley, B. Lakshminarayanan, J. Snoek, Y. Gal, and **D. Tran**. Uncertainty Baselines: Benchmarks for uncertainty & robustness in deep learning. 2021.

6. J. Lee, **D. Tran**, O. Firat, and K. Cho. On the discrepancy between density estimation and sequence generation. 2020.

7. J. Nixon, M. W. Dusenberry, L. Zhang, G. Jerfel, and **D. Tran**. Measuring calibration in deep learning. 2018.

8. M. Hoffman, P. Sountsov, J. V. Dillon, I. Langmore, **D. Tran**, and S. Vasudevan. Neutra-lizing bad geometry in Hamiltonian Monte Carlo using neural transport. 2019.

9. **D. Tran**, Y. Burda, and I. Sutskever. Feature-matching auto-encoders. 2017.

10. J. Dillon, I. Langmore, **D. Tran**, E. Brevdo, S. Vasudevan, D. Moore, B. Patton, A. Alemi, M. Hoffman, and R. Saurous. TensorFlow Distributions. 2017.

11. **D. Tran**, A. Kucukelbir, A. B. Dieng, M. Rudolph, D. Liang, and D. M. Blei. Edward: A library for probabilistic modeling, inference, and criticism. 2016.

12. **D. Tran**, F. J. R. Ruiz, S. Athey, and D. M. Blei. Model criticism for Bayesian causal inference. 2016.

13. **D. Tran**, P. Toulis, and E. M. Airoldi. Stochastic gradient descent methods for estimation with large data sets. 2016.

JOURNAL ARTICLES

14. E. Nalisnick, P. Smyth, and **D. Tran**. A brief tour of deep learning from a statistical perspective. *Annual Review of Statistics and Its Application*, 10(1):219–246, 2023.

15. J. Z. Liu, S. Padhy, J. Ren, Z. Lin, Y. Wen, G. Jerfel, Z. Nado, J. Snoek, **D. Tran**, and B. Lakshminarayanan. A simple approach to improve single-model deep uncertainty via distance-awareness. *Journal of Machine Learning Research*, 24(42):1–63, 2022.

16. M. Havasi, J. Snoek, **D. Tran**, J. Gordon, and J. M. Hernández-Lobato. Sampling the variational posterior with local refinement. *Entropy*, 23(11):1475, 2021.

17. A. Vehtari, A. Gelman, T. Sivula, P. Jylanki, **D. Tran**, S. Sahai, P. Blomstedt, J. P. Cunningham, D. Schiminovich, and C. P. Robert. Expectation propagation as a way of life: A framework for Bayesian inference on partitioned data. *Journal of Machine Learning Research*, 21(17):1–53, 2020.

18. **D. Tran** and D. M. Blei. Comment, "Fast approximate inference for arbitrarily large semiparametric regression models via message passing". *Journal of the American Statistical Association*, 112(517):156–158, 2017.

19. A. Kucukelbir, **D. Tran**, R. Ranganath, A. Gelman, and D. M. Blei. Automatic differentiation variational inference. *Journal of Machine Learning Research*, 18(14):1–45, 2017.

CONFERENCE ARTICLES

20. J. U. Allingham, F. Wenzel, Z. E. Mariet, B. Mustafa, J. Puigcerver, N. Houlsby, G. Jerfel, V. Fortuin, B. Lakshminarayanan, J. Snoek, and others. Sparse MoEs meet efficient ensembles. *Transactions on Machine Learning Research*, 2022.

21. V. Fortuin, M. Collier, F. Wenzel, J. Allingham, J. Liu, **D. Tran**, B. Lakshminarayanan, J. Berent, R. Jenatton, and E. Kokiopoulou. Deep classifiers with label noise modeling and distance awareness. *Transactions on Machine Learning Research*, 2022.

22. M. Minderer, J. Djolonga, R. Romijnders, F. Hubis, X. Zhai, N. Houlsby, **D. Tran**, and M. Lucic. Revisiting the calibration of modern neural networks. In *Neural Information Processing Systems*, 2021.

23. N. Band, T. G. J. Rudner, Q. Feng, A. Filos, Z. Nado, M. W. Dusenberry, G. Jerfel, **D. Tran**, and Y. Gal. Benchmarking Bayesian deep learning on diabetic retinopathy detection tasks. In *Neural Information Processing Systems*, 2021.

24. A. Karandikar, N. Cain, **D. Tran**, B. Lakshminarayanan, J. Shlens, M. C. Mozer, and B. Roelofs. Soft calibration objectives for neural networks. In *Neural Information Processing Systems*, 2021.

25. Y. Wen, G. Jerfel, R. Muller, M. W. Dusenberry, J. Snoek, B. Lakshminarayanan, and **D. Tran**. Combining ensembles and data augmentation can harm your calibration. In *International Conference on Learning Representations*, 2021.

26. M. Havasi, R. Jenatton, S. Fort, J. Z. Liu, J. Snoek, B. Lakshminarayanan, A. M. Dai, and **D. Tran**. Training independent subnetworks for robust prediction. In *International Conference on Learning Representations*, 2020.

27. F. Wenzel, J. Snoek, **D. Tran**, and R. Jenatton. Hyperparameter ensembles for robustness and uncertainty quantification. In *Neural Information Processing Systems*, 2020.

28. J. Z. Liu, Z. Lin, S. Padhy, **D. Tran**, T. Bedrax-Weiss, and B. Lakshminarayanan. Simple and principled uncertainty estimation with deterministic deep learning via distance awareness. In

Neural Information Processing Systems, 2020.

29. M. Mladenov, C.-w. Hsu, V. Jain, E. Ie, C. Colby, N. Mayoraz, H. Pham, **D. Tran**, I. Vendrov, and C. Boutilier. Demonstrating principled uncertainty modeling for recommender ecosystems with RecSim NG. In *ACM Conference on Recommender Systems*, 2020.

30. M. W. Dusenberry, G. Jerfel, Y. Wen, Y. Ma, J. Snoek, K. Heller, B. Lakshminarayanan, and **D. Tran**. Efficient and scalable Bayesian neural nets with rank-1 factors. In *International Conference on Machine Learning*, 2020.

31. Y. Wen, **D. Tran**, and J. Ba. Batchensemble: An alternative approach to efficient ensemble and lifelong learning. In *International Conference on Learning Representations*, 2020.

32. M. W. Dusenberry, **D. Tran**, E. Choi, J. Kemp, J. Nixon, G. Jerfel, K. Heller, and A. Dai. Analyzing the role of model uncertainty in electronic health records. In *ACM Conference on Health, Inference, and Learning*, 2020.

33. **D. Tran**, K. Vafa, K. K. Agrawal, L. Dinh, and D. Poole. Discrete flows: Invertible generative models for discrete data. In *Neural Information Processing Systems*, 2019.

34. **D. Tran**, M. W. Dusenberry, D. Hafner, and M. van der Wilk. Bayesian layers: A module for neural network uncertainty. 2019.

35. D. Hafner, **D. Tran**, A. Irpan, T. Lillicrap, and J. Davidson. Noise contrastive priors for functional uncertainty. In *Uncertainty in Artificial Intelligence*, 2019.

36. **D. Tran**, M. D. Hoffman, D. Moore, C. Suter, S. Vasudevan, A. Radul, M. Johnson, and R. A. Saurous. Simple, distributed, and accelerated probabilistic programming. In *Neural Information Processing Systems*, 2018.

37. N. Shazeer, Y. Cheng, N. Parmar, **D. Tran**, A. Vaswani, P. Koanantakool, P. Hawkins, H. Lee, M. Hong, C. Young, R. Sepassi, and B. Hechtman. Mesh-TensorFlow: Deep learning for supercomputers. In *Neural Information Processing Systems*, 2018.

38. M. D. Hoffman, M. Johnson, and **D. Tran**. Autoconj: Recognizing and exploiting conjugacy without a domain-specific language. In *Neural Information Processing Systems*, 2018.

39. N. Parmar, A. Vaswani, J. Uszkoreit, L. Kaiser, N. Shazeer, A. Ku, and **D. Tran**. Image Transformer. In *International Conference on Machine Learning*, 2018.

40. Y. Wen, P. Vicol, J. Ba, **D. Tran**, and R. Grosse. Flipout: Efficient pseudo-independent weight perturbations on mini-batches. In *International Conference on Learning Representations*, 2018.

41. **D. Tran** and D. M. Blei. Implicit causal models for genome-wide association studies. In *International Conference on Learning Representations*, 2018.

42. **D. Tran**, R. Ranganath, and D. M. Blei. Hierarchical implicit models and likelihood-free variational inference. In *Neural Information Processing Systems*, 2017.

43. A. B. Dieng, **D. Tran**, R. Ranganath, J. Paisley, and D. M. Blei. Variational inference via χ upper bound minimization. In *Neural Information Processing Systems*, 2017.

44. **D. Tran**, M. D. Hoffman, R. A. Saurous, E. Brevdo, K. Murphy, and D. M. Blei. Deep probabilistic programming. In *International Conference on Learning Representations*, 2017.

45. R. Ranganath, J. Altosaar, **D. Tran**, and D. M. Blei. Operator variational inference. In *Neural Information Processing Systems*, 2016.

46. R. Ranganath, **D. Tran**, and D. M. Blei. Hierarchical variational models. In *International Conference on Machine Learning*, 2016.

47. **D. Tran**, M. Kim, and F. Doshi-Velez. Spectral M-estimation with application to hidden Markov models. In *Artificial Intelligence and Statistics*, 2016.

48. P. Toulis, **D. Tran**, and E. M. Airoldi. Towards stability and optimality in stochastic gradient descent. In *Artificial Intelligence and Statistics*, 2016.

49. **D. Tran**, R. Ranganath, and D. M. Blei. The variational Gaussian process. In *International Conference on Learning Representations*, 2016.

50. **D. Tran**, D. M. Blei, and E. M. Airoldi. Copula variational inference. In *Neural Information Processing Systems*, 2015.

Software

1. Uncertainty Baselines   2020

2. Robustness Metrics   2020

3. Bayesian Layers: A module for neural network uncertainty   2018
   D. Tran, M. Dusenberry, M. van der Wilk, D. Hafner.

4. Mesh-TensorFlow: Deep learning for supercomputers   2018
   N. Shazeer, Y. Cheng, N. Parmar, **D. Tran**, A. Vaswani, P. Koanantakool, P. Hawkins, H. Lee, M. Hong, C. Young, R. Sepassi, B. Hechtman.

5. Edward2: Simple, distributed, and accelerated probabilistic programming   2018
   D. Tran, M. D. Hoffman, D. Moore, C. Suter, S. Vasudevan, A. Radul, M. Johnson, and R. A. Saurous.

6. Tensor2Tensor: Library of deep learning models and datasets   2017

7. Observations: A one-line API for loading standard data sets in machine learning   2017
   D. Tran.

8. Edward: A library for probabilistic modeling, inference, and criticism   2016
   D. Tran, A. Kucukelbir, A.B. Dieng, D. Liang, M. Rudolph, and D.M. Blei.

9. sgd: An R package for large-scale estimation   2015
   D. Tran, P. Toulis, and E.M. Airoldi.

10. Stan: A platform for statistical modeling and high-performance statistical computation   2012
    A. Gelman, B. Carpenter, M. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo,

P. Li, A. Riddell, M. Inacio, J. Arnold, M. Morris, R. Trangucci, R. Goedman, B. Lau, J. Gabry, A. Kucukelbir, R. Grant, **D. Tran**, K. Sakrejda, A. Vehtari, R. Lei, S. Weber.

Teaching

1. Teaching Assistant | Columbia University   2016
   STAT/CS 6509: Foundations of Graphical Models

2. Teaching Fellow | Harvard University   2015
   AM 205: Advanced Scientific Computing–Numerical Methods

3. Teaching Assistant | University of California, Berkeley   2013
   MATH 10B: Methods in Calculus, Statistics, Combinatorics

4. Teaching Assistant | University of California, Berkeley   2011
   MATH 128A: Numerical Analysis

Professional Service

PROGRAM COMMITTEE

Senior Area Chair: Artificial Intelligence and Statistics   2023

Area Chair: Artificial Intelligence and Statistics   2019–2022

Area Chair: Association for the Advancement of Artificial Intelligence   2020–

Area Chair: Neural Information Processing Systems   2019–

Area Chair: International Conference on Learning Representations   2020–

Area Chair: International Conference on Machine Learning   2019–

Senior Program Committee: International Joint Conferences on Artificial Intelligence   2020–

WORKSHOP ORGANIZATION

Symposium: Advances in Approximate Bayesian Inference   2019

Symposium: Advances in Approximate Bayesian Inference   2018

UAI Workshop: Uncertainty in Deep Learning   2018

NIPS Workshop: Advances in Approximate Bayesian Inference   2017

ICML Workshop: Implicit Generative Models   2017

NIPS Workshop: Advances in Approximate Bayesian Inference   2016

NIPS Workshop: Advances in Approximate Bayesian Inference   2015

MENTORING

Kelly Buchanan (Google Student Researcher, 2022)

James Allingham (Google Brain Intern, 2021, 2022)

Frances Hubis (Google AI Resident, 2021)

Archit Karandikar (Google Software Engineer, 2021)

Jeremy Nixon (Google Software Engineer, 2020–2021)

Yeming Wen (Google Brain Intern, 2019–2020)

Aditya Grover (Google AI Resident, 2019)

Jason Lee (Google Brain Intern, 2019–2020)

Ghassen Jerfel (Google Student Researcher, 2019–2021)

Kumar Krishna Agrawal (Google AI Resident, 2019)

Michael W. Dusenberry (Google AI Resident, 2018–2020)

Keyon Vafa (Google Brain Intern, 2018)

Danijar Hafner (Google Brain Intern, 2018)

Akshay Khatri (M.S. Columbia University, 2017)