

# David Kristjanson Duvenaud

---

## PROFESSIONAL EXPERIENCE

**Fable Therapeutics** August 2022 – present  
*Co-Founder*. Advising on the use of generative models for *in silico* therapeutics design.

**Anthropic** June 2023 – October 2024  
*Team Lead*, Alignment Evaluations. Led a team of 5 to design and build a set of evaluations to fulfill Anthropic's Responsible Scaling Policy commitments regarding evaluation integrity. [paper link]

**Cohere** January 2021 – June 2023  
*Advisor*. Helped direct research projects relating to personalizing LLMs.

**Radical Ventures** Dec 2019 – June 2023  
*Advisor*. Did technical due diligence on early-stage machine learning startups.

**Google Brain Toronto** 2020 – 2022  
*Visiting Researcher*. Contributed to meta-learning research projects, JAX, and the Dex language.

**ElementAI** November 2016 – December 2019  
*Faculty Fellow*. Advised industrial R & D efforts.

**Vector Institute** September 2017 – present  
*Co-Founder and Faculty Member*

**University of Toronto** July 2016 – present  
*Associate Professor*, Computer Science and Statistical Sciences  
Canada Research Chair in Generative Models

**Harvard School of Engineering and Applied Sciences** Sept 2014 – June 2016  
*Postdoctoral Fellow*, Intelligent Probabilistic Systems group  
Worked with Prof. Ryan P. Adams on Bayesian optimization, deep learning, molecular modeling, and variational inference.

**Max Planck Institute for Intelligent Systems** Summer 2012  
*Visiting Researcher*, Schölkopf group  
Worked with Phillip Hennig on stochastic quasi-Newton optimization, model-based ordinary differential equation solvers, and nonparametric inference methods.

**Google Research** Summers 2010 and 2011  
*Software Engineering Intern*, Video Content Analysis team  
Used machine vision to solve YouTube video classification problems at scale. Contributed to Dist-Belief, a close-to-the-metal distributed deep learning framework, and precursor to TensorFlow.

**Invenia** 2006 – 2008  
*Cofounder*  
Co-founded a machine learning research consulting company. Recruited, trained and supervised five research assistants, plus consultants. Wrote, presented and was awarded several research grants. Led two research contracts applying machine learning methods to energy forecasts. These projects led to the deployment of automated forecasting systems for several major utilities.

**Canadian Army Reserve** 2005 - 2010  
*Trooper* (armoured reconnaissance), British Columbia Regiment & Fort Garry Horse

## EDUCATION

**University of Cambridge, Machine Learning Group** 2010 – 2014

Ph.D., Engineering

Advisors: Carl Rasmussen and Zoubin Ghahramani

Thesis: Automatic model construction with Gaussian processes

**University of British Columbia, Laboratory for Computational Intelligence** 2008 – 2010

M. Sc., Computer Science

Advisor: Kevin P. Murphy

Thesis: Multiscale conditional random fields for machine vision

**University of Manitoba**

2001 – 2006

B. Sc. Hons., Computer Science. First class honours.

## GRANTS AND AWARDS (CAD)

CS-Can Outstanding Early Career Computer Science Researcher Award: \$1,000	2025
Inaugural Schwartz Reisman Chair: \$500,000	2024
Sloan Research Fellowship: \$150,000	2022
ICFP Distinguished Paper Award (top 4 papers in conference)	2021
ICML Outstanding Paper Honorable Mention (top 4 papers in conference)	2021
Ontario Early Researcher Award: \$140,000	2021
CIFAR AI Chair: \$750,000	2021
Google faculty award: \$45,000	2019
NeurIPS Best paper award (top 4 papers in conference)	2018
Samsung research gift: \$67,250	2018
Connaught New Researcher Grant: \$9,700	2021
NVIDIA Compute the Cure research grant: \$250,000	2017
Tier II Canada Research Chair: \$500,000	2017
NSERC Discovery Grant: \$140,000	2017

## PREPRINTS AND TECHNICAL REPORTS

1. Sharma, M., McCain, M., Douglas, R., & Duvenaud, D. (2026). Who’s in charge? disempowerment patterns in real-world LLM usage. *arXiv preprint arXiv:2601.19062*.
2. Lowe, R., Edelman, J., Zhi-Xuan, T., Klingefjord, O., Hain, E., Duvenaud, D., ... others (2025). Full-stack alignment: Co-aligning AI and institutions with thicker models of value. In *2nd workshop on models of human feedback for ai alignment*.
3. Kulveit, J., Douglas, R., Ammann, N., Turan, D., Krueger, D., & Duvenaud, D. (2025a). Gradual disempowerment: Systemic existential risks from incremental AI development. *arXiv preprint arXiv:2501.16946*.
4. Greenblatt, R., Denison, C., Wright, B., Roger, F., MacDiarmid, M., Duvenaud, D., ... others (2024). Alignment faking in large language models. *arXiv preprint arXiv:2412.14093*.
5. Benton, J., Wagner, M., Christiansen, E., Anil, C., Perez, E., Srivastav, J., ... Duvenaud, D. (2024). Sabotage evaluations for frontier models. *arXiv preprint arXiv:2410.21514*.
6. Richter-Powell, J., Thiede, L., Asparu-Guzik, A., & Duvenaud, D. (2023). Sorting out quantum Monte Carlo. *arXiv preprint arXiv:2311.05598*.
7. Lorraine, J., & Duvenaud, D. (2017). Stochastic hyperparameter optimization through hypernetworks. In *NeurIPS workshop on meta-learning*.

- [61] Kulveit, J., Douglas, R., Ammann, N., Turan, D., Krueger, D., & Duvenaud, D. (2025b). Position: Humanity faces existential risk from gradual disempowerment. In *Proceedings of the 42nd international conference on machine learning*. Vancouver, Canada. (Poster #40107)
- [60] Shysheya, A., Bronskill, J., Requeima, J., Siddiqui, S. A., Gonzalez, J., Duvenaud, D., & Turner, R. E. (2025). Jolt: Joint probabilistic predictions on tabular data using llms. *arXiv preprint arXiv:2502.11877*.
- [59] Anil, C., Durmus, E., Sharma, M., Benton, J., Kundu, S., Batson, J., ... Duvenaud, D. (2024). Many-shot jailbreaking. In *Neural information processing systems*.
- [58] Sharma, M., Tong, M., Korbak, T., Duvenaud, D., Askell, A., Bowman, S. R., ... Perez, E. (2024). Towards understanding sycophancy in language models. In *International conference on learning representations*.
- [57] Requeima, J., Bronskill, J., Choi, D., Turner, R. E., & Duvenaud, D. (2024). Llm processes: Numerical predictive distributions conditioned on natural language. In *Thirty-eighth conference on neural information processing systems*.
- [56] Johnson, D. D., Tarlow, D., Duvenaud, D., & Maddison, C. J. (2024). Experts don't cheat: Learning what you don't know by predicting pairs. In *International conference on machine learning*.
- [55] Sharma, M., Tong, M., Korbak, T., Duvenaud, D., Askell, A., Bowman, S. R., ... Perez, E. (2024). Towards understanding sycophancy in language models. In *International conference on learning representations*.
- [54] Choi, D., Shavit, Y., & Duvenaud, D. (2023). Tools for verifying neural models' training data. In *Neural information processing systems*.
- [53] Vicol, P., Lorraine, J. P., Duvenaud, D., & Grosse, R. B. (2023). On implicit regularization in overparameterized bilevel optimization. In *International conference on machine learning*.
- [52] Lorraine, J., Acuna, D., Vicol, P., & Duvenaud, D. (2022). Complex momentum for learning in games. In *Artificial intelligence and statistics*.
- [51] Xu, W., Chen, R. T., Li, X., & Duvenaud, D. (2022). Infinitely deep bayesian neural networks with stochastic differential equations. In *Artificial intelligence and statistics*.
- [50] Raghu, A., Lorraine, J. P., Kornblith, S., McDermott, M. B., & Duvenaud, D. (2021). Meta-learning to improve pre-training. In *Neural information processing systems*.
- [49] Grathwohl, W., Swersky, K., Hashemi, M., Duvenaud, D., & Maddison, C. J. (2021). Oops I took a gradient: Scalable sampling for discrete distributions. In *International conference on machine learning*. Outstanding Paper Honorable Mention (top 4 papers)
- [48] Paszke, A., Johnson, D., Duvenaud, D., Vytiniotis, D., Radul, A., Johnson, M., ... Maclaurin, D. (2021). Getting to the point. index sets and parallelism-preserving autodiff for pointful array programming. In *International conference on functional programming*. Distinguished Paper Award (top 4 papers)
- [47] Raghu, A., Raghu, M., Kornblith, S., Duvenaud, D., & Hinton, G. (2021). Teaching with commentaries. In *International conference on learning representations*.
- [46] Grathowhl, W., Kelly, J., Hashemi, M., Norouzi, M., Swersky, K., & Duvenaud, D. (2021). No MCMC for me: Amortized sampling for fast and stable training of energy-based models. In *International conference on learning representations*.
- [45] Tonekaboni, S., Joshi, S., Campbell, K., Duvenaud, D., & Goldenberg, A. (2020). What went wrong and when? instance-wise feature importance for time-series models. In *Neural information processing systems*.

- [44] Kelly, J., Bettencourt, J., Johnson, M. J., & Duvenaud, D. (2020). Learning differential equations that are easy to solve. In *Neural information processing systems*.
- [43] Grathwohl, W., Wang, K.-C., Jacobsen, J.-H., Duvenaud, D., & Zemel, R. (2020). Learning the Stein discrepancy for training and evaluating energy-based models without sampling. In *International conference on machine learning*.
- [42] Li, X., Chen, R. T. Q., Wong, T.-K. L., & Duvenaud, D. (2020). Scalable gradients for stochastic differential equations. In *Artificial intelligence and statistics*.
- [41] Lorraine, J., Vicol, P., & Duvenaud, D. (2020). Optimizing millions of hyperparameters by implicit differentiation. In *Artificial intelligence and statistics*.
- [40] Grathwohl, W., Wang, K.-C., Jacobsen, J.-H., Duvenaud, D., Norouzi, M., & Swersky, K. (2020). Your classifier is secretly an energy based model and you should treat it like one. In *International conference on learning representations*.
- [39] Luo, Y., Beatson, A., Norouzi, M., Zhu, J., Duvenaud, D., Adams, R. P., & Chen, R. T. (2020). SUMO: Unbiased estimation of log marginal probability for latent variable models. In *International conference on learning representations*.
- [38] Chen, R. T. Q., & Duvenaud, D. (2019). Neural networks with cheap differential operators. In *Neural information processing systems*.
- [37] Chen, R. T. Q., Rubanova, Y., & Duvenaud, D. (2019). Latent ODEs for irregularly-sampled time series. In *Neural information processing systems*.
- [36] Liao, R., Li, Y., Song, Y., Wang, S., Hamilton, W., Duvenaud, D., ... Zemel, R. (2019). Efficient graph generation with graph recurrent attention networks. In *Neural information processing systems*.
- [35] Chen, R. T., mann, J., Duvenaud, D., & Jacobsen, J.-H. (2019). Residual flows for invertible generative modeling. In *Neural information processing systems*.
- [34] Ethayarajh, K., Duvenaud, D., & Hirst, G. (2019a). Towards understanding linear word analogies. In *Association for computational linguistics*.
- [33] Ethayarajh, K., Duvenaud, D., & Hirst, G. (2019b). Understanding undesirable word embedding associations. In *Association for computational linguistics*.
- [32] Chen, R. T., mann, J., Duvenaud, D., & Jacobsen, J.-H. (2019). Residual flows for invertible generative modeling. In *Neural information processing systems*.
- [31] Behrmann, J., Grathwohl, W., Chen, R. T. Q., Duvenaud, D., & Jacobsen, J.-H. (2019). Invertible residual networks. In *International conference on machine learning*. Oral presentation.
- [30] Grathwohl, W., Chen, R. T. Q., Bettencourt, J., Sutskever, I., & Duvenaud, D. (2019). Ffjord: Free-form continuous dynamics for scalable reversible generative models. *International Conference on Learning Representations*. Oral presentation.
- [29] MacKay, M., Vicol, P., Lorraine, J., Duvenaud, D., & Grosse, R. (2019). Self-tuning networks: Bilevel optimization of hyperparameters using structured best-response functions. In *International conference on learning representations*.
- [28] Chang, C.-H., Creager, E., Goldenberg, A., & Duvenaud, D. (2019). Explaining image classifiers by adaptive dropout and generative in-filling. In *International conference on learning representations*.
- [27] Fulton, L., Modi, V., Duvenaud, D., Levin, D. I., & Jacobson, A. (2019). Latent-space dynamics for reduced deformable simulation. In *Computer graphics forum* (Vol. 38, pp. 379–391).

- [26] Chen, R. T. Q., Rubanova, Y., Bettencourt, J., & Duvenaud, D. (2018). Neural ordinary differential equations. *Neural Information Processing Systems*. Best paper award.
- [25] Chen, R. T. Q., Li, X., Grosse, R., & Duvenaud, D. (2018). Isolating sources of disentanglement in variational autoencoders. *Neural Information Processing Systems*. Oral Presentation.
- [24] Cremer, C., Li, X., & Duvenaud, D. (2018). Inference suboptimality in variational autoencoders. *International Conference on Machine Learning*.
- [23] Zhang, G., Sun, S., Duvenaud, D., & Grosse, R. (2018). Noisy natural gradient as variational inference. *International Conference on Machine Learning*.
- [22] Grathwohl, W., Choi, D., Wu, Y., Roeder, G., & Duvenaud, D. (2018). Backpropagation through the void: Optimizing control variates for black-box gradient estimation. In *International conference on learning representations*.
- [21] Gomez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernandez-Lobato, J. M., Sanchez-Lengeling, B., Sheberla, D., . . . Aspuru-Guzik, A. (2018). Automatic chemical design using a data-driven continuous representation of molecules. *American Chemical Society Central Science*.
- [20] Schulz, E., Tenenbaum, J. B., Duvenaud, D., Speekenbrink, M., & Gershman, S. J. (2017). Compositional inductive biases in function learning. *Cognitive psychology*, 99, 44–79.
- [19] Roeder, G., Wu, Y., & Duvenaud, D. (2017). Sticking the landing: Simple, lower-variance gradient estimators for variational inference. In *Neural information processing systems*.
- [18] Wei, J. N., Duvenaud, D., & Aspuru-Guzik, A. (2016). Neural networks for the prediction of organic chemistry reactions. *ACS Central Science*, 2(10), 725–732.
- [17] Johnson, M. J., Duvenaud, D., Wiltschko, A., Datta, S., & Adams, R. P. (2016). Composing graphical models with neural networks for structured representations and fast inference. In *Neural information processing systems*.
- [16] Shulz, E., Tenenbaum, J. B., Duvenaud, D., Speekenbrink, M., & Gershman, S. J. (2016). Probing the compositionality of intuitive functions. In *Neural information processing systems*.
- [15] Gómez-Bombarelli, R., Aguilera-Iparraguirre, J., Hirzel, T. D., Duvenaud, D., Maclaurin, D., Blood-Forsythe, M. A., . . . Aspuru-Guzik, A. (2016). Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. *Nature materials*, 15(10), 1120.
- [14] Duvenaud, D., Maclaurin, D., & Adams, R. (2016). Early stopping as nonparametric variational inference. In *Artificial intelligence and statistics* (pp. 1070–1077).
- [13] Huang, C.-Z. A., Duvenaud, D., & Gajos, K. Z. (2016). Chordripple: Recommending chords to help novice composers go beyond the ordinary. In *Intelligent user interfaces* (pp. 241–250).
- [12] Duvenaud, D., Maclaurin, D., Aguilera-Iparraguirre, J., Gómez-Bombarelli, R., Hirzel, T., Aspuru-Guzik, A., & Adams, R. P. (2015). Convolutional networks on graphs for learning molecular fingerprints. In *Neural information processing systems*.
- [11] Maclaurin, D., Duvenaud, D., & Adams, R. P. (2015, July). Gradient-based hyperparameter optimization through reversible learning. In *International conference on machine learning*.
- [10] Schober, M., Duvenaud, D., & Hennig, P. (2014). Probabilistic ODE solvers with Runge-Kutta means. In *Neural information processing systems*. Oral presentation.
- [9] Lloyd, J. R., Duvenaud, D., Grosse, R., Tenenbaum, J. B., & Ghahramani, Z. (2014). Automatic construction and natural-language description of nonparametric regression models. In *Association for the advancement of artificial intelligence (aaai)*.

- [8] Duvenaud, D., Rippel, O., Adams, R. P., & Ghahramani, Z. (2014). Avoiding pathologies in very deep networks. In *Artificial intelligence and statistics*.
- [7] Anna Huang, C.-Z., Duvenaud, D., Arnold, K., Partridge, B., W. Oberholtzer, J., & Z. Gajos, K. (2014, 02). Active learning of intuitive control knobs for synthesizers using Gaussian processes. In (p. 115-124).
- [6] Tomoharu Iwata, Z. G., David Duvenaud. (2013). Warped mixtures for nonparametric cluster shapes. In *Uncertainty in artificial intelligence* (p. 311-319).
- [5] Duvenaud, D., Lloyd, J. R., Grosse, R., Tenenbaum, J. B., & Ghahramani, Z. (2013). Structure discovery in nonparametric regression through compositional kernel search. In *International conference on machine learning* (pp. 1166–1174).
- [4] Osborne, M. A., Duvenaud, D., Garnett, R., Rasmussen, C. E., Roberts, S. J., & Ghahramani, Z. (2012). Active learning of model evidence using Bayesian quadrature. In *Neural information processing systems*.
- [3] Huszár, F., & Duvenaud, D. (2012). Optimally-weighted herding is Bayesian quadrature. In *Uncertainty in artificial intelligence* (pp. 377–385). Oral presentation.
- [2] Duvenaud, D., Nickisch, H., & Rasmussen, C. E. (2011). Additive Gaussian processes. In *Neural information processing systems* (pp. 226–234).
- [1] Duvenaud, D., Marlin, B., & Murphy, K. (2011). Multiscale conditional random fields for semi-supervised labeling and classification. In *Proceedings of the 8th Canadian conference on computer and robot vision* (pp. 371–378). IEEE Computer Society.

WORKSHOP  
PUBLICATIONS

- [9] Nado, Z., Snoek, J., Grosse, R., Duvenaud, D., Xu, B., & Martens, J. (2018). Stochastic gradient langevin dynamics that exploit neural network structure. In *International conference on learning representations workshop track*.
- [8] Killoran, N., Lee, L. J., DeLong, A., Duvenaud, D., & Frey, B. J. (2017). Generating and designing DNA with deep generative models. In *NeurIPS workshop on machine learning in computational biology*.
- [7] Cremer, C., Morris, Q., & Duvenaud, D. (2017). Reinterpreting importance-weighted autoencoders. *International Conference on Learning Representations Workshop Track*.
- [6] Duvenaud, D., & Adams, R. P. (2015). Black-box stochastic variational inference in five lines of python. *NeurIPS Workshop on Black-box Learning and Inference*.
- [5] Altieri, N., & Duvenaud, D. (2015). Variational inference with gradient flows. In *NeurIPS workshop on advances in approximate bayesian inference*.
- [4] Maclaurin, D., Duvenaud, D., Johnson, M. J., & Adams, R. P. (2015). Autograd: Reverse-mode differentiation of native python. *ICML workshop on Automatic Machine Learning*.
- [3] Grosse, R., & Duvenaud, D. (2014). Testing Markov-chain Monte Carlo code. In *NeurIPS workshop on software engineering for machine learning*.
- [2] Swersky, K., Duvenaud, D., Snoek, J., Hutter, F., & Osborne, M. (2013). Raiders of the lost architecture: Kernels for Bayesian optimization in conditional parameter spaces. In *NeurIPS workshop on bayesian optimization*.
- [1] Duvenaud, D., Eaton, D., Murphy, K., & Schmidt, M. (2010). Causal learning without DAGs. In *Journal of machine learning research workshop and conference proceedings* (Vol. 6, pp. 177–190).

## PATENTS

Aspuru-Guzik, A., Gomez-Bombarelli, R., Hirzel, T.D., Aguilera-Iparraguirre, J., Adams, R.P., Maclaurin, D., and Duvenaud, D. Organic light-emitting diode materials. WO2015175678

## INVITED TALKS

University of Toronto CS Distinguished Lecture Series	October 2025
Forethought Institute, Oxford	July 2025
Toronto AI Safety Meetup	July 2025
Absolutely Interdisciplinary Conference, Toronto	May 2025
GovAI internal research seminar series (remote)	May 2025
Inter-Math-AI Special Lecture (remote)	April 2025
Google Deepmind, AGI and Society reading group (remote)	April 2025
Think7 Summit, Centre for International Governance Innovation	April 2025
NeurIPS Workshop on Multimodal LLMs	December 2024
AE Global Summit on Open Problems in AI	December 2024
CHAI Alignment workshop	October 2024
CIFAR Deep Learning Summer School	July 2024
University of Toronto Student AI Conference	January 2023
Cambridge University, Computational and Biological Learning Lab	December 2022
Secondmind Ltd.	December 2022
Distinguished Lecture, Berkeley Computer Science Department	October 2022
Research Club, University of Toronto Schools (high school) (remote)	April 2022
NeurIPS workshop: Programming Languages and Neurosymbolic Systems (remote)	December 2021
Schwartz-Riesman Institute Seminar Series (remote)	November 2021
ICCV Workshop on Neural Architectures: Present and Future (remote)	October 2021
Keynote: KDD Workshop on Mining and Learning from Time Series (remote)	August 2021
ICML Workshop on Time Series (remote)	July 2021
Oxford University, StatML Centre for Doctoral Training Seminar (remote)	July 2021
Centre for Mathematics and Algorithms for Data, University of Bath (remote)	July 2021
Microsoft Research AutoML Lecture Series (remote)	May 2021
Flatiron Institute, Center for Computational Mathematics	May 2021
ICLR Workshop on Deep Learning for Simulation (remote)	April 2021
University College London, DeepMind/ELLIS CSML Seminar Series (remote)	February 2021
NeurIPS Europe meetup on Bayesian Deep Learning (remote)	December 2020
NeurIPS Workshop: Beyond Backpropagation (remote)	December 2020
NeurIPS Tutorial: Deep Implicit Layers (remote)	December 2020
Toronto Machine Learning Summit (remote)	November 2020
University of Amsterdam Machine Learning Seminar (remote)	November 2020
University of Washington, Applied Math seminar series (remote)	October 2020
ODSC West Virtual Conference (remote)	October 2020
University of Toronto, Computer Science Student Union Seminar (remote)	October 2020
University of Pennsylvania, Applied Math Colloquium Series (remote)	October 2020
Tenth International Workshop on Climate Informatics (remote)	September 2020
Symposium on Sparse Recovery and Machine Learning, SIAM Annual Meeting (remote)	July 2020
World AI Conference, Beijing (remote)	July 2020
Institute for Advanced Study, Princeton University (remote)	April 2020
Guest Lecture, Yale University (remote)	March 2020
Deep Structures Workshop, Aalto University, Finland	December 2019
NeurIPS Retrospectives Workshop	December 2019
NeurIPS Workshop on Learning Meaningful Representations of Life	December 2019
NeurIPS Communications Practicum	December 2019
University of British Columbia	December 2019
Toronto Machine Learning Summit	November 2019
Fields Institute, Conference on Data Science	November 2019
MIT CSAIL Machine Learning Seminar	October 2019
Google Brain, Cambridge, Massachusetts	October 2019

Harvard University, Data to Actionable Knowledge Group	October 2019
Broad Institute, Models, Inference & Algorithms Initiative	October 2019
Gatsby Computational Neuroscience Unit, University College London	May 2019
Oxford Undergraduate Maths Society	May 2019
NVIDIA Research Toronto	December 2018
Symposium on Advances in Approximate Bayesian Inference, Montréal	December 2018
Canada-UK Colloquium on AI	November 2018
Toronto Machine Learning Summit	November 2018
CIFAR Deep Learning Summer School	July 2018
University of Oxford, Robotics Research Group	July 2018
Microsoft Research Cambridge	July 2018
Google Deepmind	July 2018
ICML Workshop on Credit Assignment in Reinforcement Learning	July 2018
Google Brain, San Francisco	June 2018
UC Berkeley, Center for Human-Compatible AI	June 2018
NeurIPS Workshop on Machine Learning for Molecules and Materials	December 2017
Toronto Machine Learning Summit	November 2017
Montréal Deep Learning Summit	October 2017
Simons Institute, Workshop on Machine Learning	May 2017
Data Learning and Inference Meeting (DALI)	April 2017
Google Brain, Mountain View	February 2017
University of Waterloo, Computational Mathematics Colloquium	January 2017
NeurIPS Workshop on Automatic Differentiation	December 2016
NeurIPS Workshop on Optimizing the Optimizers	December 2016
American Chemical Society National Meeting, Machine Learning Workshop	August 2016
OpenAI	April 2016
University of Toronto, Department of Computer Science	March 2016
University of British Columbia, Department of Computer Science	March 2016
New York University, Computer Science Department	February 2016
Princeton University, Department of Computer Science	February 2016
Université de Montréal, Institute for Learning Algorithms	February 2016
Cambridge University, Computational and Biological Learning Lab	February 2016
Twitter Cortex	February 2016
NeurIPS Workshop on Probabilistic Integration	December 2015
MIT Media Lab, Laboratory for Social Machines	November 2015
UMass Amherst, Machine Learning and Friends Lunch	November 2015
Broad Institute, Stat Math Reading Club	November 2015
Brown University, Scientific Computing Group	November 2015
University of Toronto, Machine Learning Group	October 2015
Microsoft Research Cambridge	July 2015
University of Oxford, Robotics Research Group	July 2015
University of Oxford, Future of Humanity Institute	July 2015
Google DeepMind	July 2015
Cambridge University, Computational and Biological Learning Lab	July 2015
ICML Workshop on Automatic Machine Learning	July 2015
Conference on Bayesian Nonparametrics	June 2015
Boston Machine Learning Meetup	February 2015
Harvard Society for Mind, Brain and Behavior	December 2014
Sheffield University, Deep Probabilistic Models Workshop	October 2014
MIT CSAIL, Clinical Decision Making Group	October 2014
London Machine Learning Meetup	June 2014
University of Oxford, Future of Humanity Institute	January 2014
University of Toronto, Machine Learning Group	January 2014
University of Oxford, Robotics Research Group	April 2013

Microsoft Research Cambridge	March 2013
Sheffield University, Institute for Translational Neuroscience	February 2013
NeurIPS Workshop on Confluence between Kernel Methods and Graphical Models	December 2012
NeurIPS Workshop on Probabilistic Numerics	December 2012
ICML Workshop on RKHS and Kernel-based methods	July 2012
University of Washington, Statistics Department	January 2011
DeepMind Technologies	November 2011
University of Manitoba, Machine Learning Course guest lecture	February 2008
University of Manitoba, Numerical Physics Course guest lecture	April 2007

SERVICE	Lead Organizer, Workshop on Post-AGI Economics, Culture, and Governance	December 2025
	Lead Organizer, Workshop on Post-AGI Civilizational Equilibria	July 2025
	Scientific Advisor, LawZero	June 2025 - present
	Scientific Advisor, Apollo Research	April 2025 - present
	Safe and Secure AI Advisory Group for Canadian Ministry of Science (ISED)	2025
	Co-Director, AI Safety Foundation	2024 - present
	Co-Chair, Schwartz Reisman Institute for Technology and Society	2024 - present
	Faculty Affiliate, Schwartz Reisman Institute for Technology and Society	2020, 2021, 2022, 2023
	Co-organizer, NeurIPS workshop on Deep Learning and Differential Equations	2022
	Senior Area Chair, Neural Information Processing Systems (NeurIPS)	2020, 2021
	Area Chair, Intl. Conference on Learning Representations (ICLR)	2017, 2018, 2019, 2020, 2021
	Area Chair, International Conference on Machine Learning (ICML)	2017, 2018, 2019, 2021
	Judge, ProjectX Machine Learning Research Competition	2020
	Sponsorships Chair, Uncertainty in Artificial Intelligence (UAI)	2020
	Area Chair, Neural Information Processing Systems (NeurIPS)	2017, 2018, 2019
	Area Chair, Artificial Intelligence and Statistics (AISTATS)	2017, 2018
	Co-organizer, NeurIPS Workshop on Aligned Artificial Intelligence	2017
	Area Chair, Association for the Advancement of Artificial Intelligence (AAAI)	2017
	Co-organizer, NeurIPS Workshop on Reliable Machine learning	2016

REVIEWING	International Conference on Machine Learning (ICML)	2013, 2014, 2015, 2016, 2025
	Journal of Machine Learning Research (JMLR)	2012, 2013, 2015, 2018, 2019, 2020, 2021
	Wellcome Trust Grants	2021
	Nature Communications	2020
	Neural Computation	2020
	ICML Workshop on Invertible Neural Networks and Normalizing Flows	2020
	Proceedings of the National Academy of Sciences of the United States of America (PNAS)	2020
	NeurIPS Workshop Proposals	2019
	European Physical Journal C	2019
	Workshop on Language for Inference (LAFI)	2019
	Journal of Chemical Information and Modeling	2018
	Science	2018
	Computer Graphics and Interactive Techniques (SIGGRAPH)	2018
	Nature	2017
	American Chemical Society Central Science (ACS)	2017, 2108
	IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)	2012, 2017
	International Joint Conferences on Artificial Intelligence (IJCAI)	2016
	International Conference on Learning Representations (ICLR)	2016
	Neural Information Processing Systems (NeurIPS)	2013, 2014, 2015, 2016
	Artificial Intelligence and Statistics (AISTATS)	2014, 2015
	Statistics and Computing	2013, 2014, 2015

PRESS COVERAGE AND APPEARANCES	Interview on <i>80,000 hours podcast</i>	Forthcoming
--------------------------------	------------------------------------------	-------------

Interview on <i>Existential Hope</i> podcast	Sept 25, 2025
Interview on the Forethought Institute's <i>ForeCast</i> podcast	Forthcoming
Op-ed: Can democracy survive the end of human employment? <i>The Economist</i>	Sept 18th, 2025
Interview on <i>The Trajectory</i> podcast	June 6, 2025
Op-ed: Can we stop AI making humans obsolete? <i>The Guardian</i>	May 4th, 2025
Interview on <i>Doom Debates</i> podcast	April 18, 2025
Sabine Hossenfelder's summary of Gradual Disempowerment	Feb 11, 2025
Zvi Mowshowitz's review of Gradual Disempowerment	Feb 05, 2025
Interview on the AI X-risk Research Podcast	March 1, 2025
Profile on Geoff Hinton and AI Risk. <i>Toronto Star</i>	November 20, 2023
Interview on <i>Minds Almost Meeting</i> podcast	May 31, 2023
Interview on the FLI open letter. <i>BBC Newsnight</i>	April 3, 2023
Interview on the FLI open letter. <i>Metro Morning on CBC Radio</i>	March 31, 2023
AI makes predictions about random events like market trades. <i>VentureBeat</i>	May 7, 2020
Profile: The Chosen Few. <i>Report on Business Magazine</i>	February 25, 2019
New AI Method Wins Coveted NeurIPS Award. <i>Psychology Today</i>	January 14, 2019
NeurIPS 2018 Best Paper Team: "Math Is Forever" <i>Synched Newsletter</i>	December 21, 2018
A radical new neural network design. <i>MIT Technology Review</i>	December 12, 2018
Interview: <i>Talking Machines</i> podcast	Sept 8, 2018
Should Artificial Intelligence Copy the Human Brain? <i>Wall Street Journal</i>	August 4, 2018
Interview: New frontiers in deep learning research. <i>In Context Podcast</i>	March 29, 2018
Interview: <i>This week in machine learning &amp; AI</i> podcast	January 15, 2018
Is AI riding a one-trick pony? <i>MIT Tech. Rev.</i>	September 29, 2017
Finally, a way to halt Canada's 'brain drain'. <i>Globe and Mail.</i>	July 10, 2017
Software dreams up new molecules in quest for wonder drugs. <i>MIT Tech. Rev.</i>	November 3, 2016
'Artificial brain' aces undergrad organic chemistry test. <i>Chemistry World</i>	October 17, 2016
The hunt for tomorrow's diodes is tangled up in blue. <i>Wall Street Journal</i>	August 19, 2016
'Molecular Tinder' may change the game for OLED screens. <i>Techcrunch</i>	August 8, 2016
The growing influence of statisticians. <i>Phys-org</i>	June 4, 2015
The Automatic Statistician and electrified meat. <i>Talking Machines</i> podcast	March 26, 2015
Automating the data scientists. <i>MIT Technology Review</i>	February 13, 2015
How machines learned to think statistically. <i>Significance magazine</i>	February 3, 2015
Google is funding an artificial intelligence for data science. <i>Yahoo! News</i>	December 2, 2014