David M. Blei

Professor
Columbia University
Department of Computer Science
Department of Statistics
Columbia Data Science Institute
New York, NY

david.blei@columbia.edu
http://www.cs.columbia.edu/∼blei
CV Updated: April 13, 2024

Education

- Ph.D. Computer Science, University of California Berkeley, 2004
  Advisor: Michael Jordan
- B.Sc. (Honors) Computer Science and Mathematics, Brown University, 1997
  Advisor: Leslie Kaelbling

Employment

- Professor, Departments of Statistics and Computer Science, Columbia University, 2014–
- Associate Professor, Department of Computer Science, Princeton University, 2011–2014
- Assistant Professor, Department of Computer Science, Princeton University, 2006–2011
- Postdoctoral Fellow, Department of Machine Learning, Carnegie Mellon University, 2004–2006
  Advisor: John Lafferty

Awards

- Lenfest Distinguished Faculty Award, 2024
- NeurIPS Test of Time Award, 2021
- KDD Test of Time Award, 2021
- Simons Investigator, 2019
- Guggenheim Fellowship, 2017
- Fellow of the Institute of Mathematical Statistics, 2017
- ICML Test of Time Award, 2016
- Presidential Award for Outstanding Teaching, Honorable Mention, 2016
- Fellow of the Association of Computing Machinery, 2015
- SIGIR Test of Time Award Honorable Mention, 2015
- ACM Prize in Computing, 2013
- Blavatnik Award for Young Scientists: Faculty Winner, 2013
- Presidential Early Career Award for Scientists and Engineers (PECASE), 2011

- Office of Naval Research Young Investigator Award, 2011
- Alfred P. Sloan Fellowship, 2010
- E.L. Keyes Jr. Emerson Electric Co. Faculty Award, 2008
- National Science Foundation CAREER Award, 2008
- Microsoft New Faculty Fellowship Finalist, 2007
- Microsoft Research Award, 2007
- Google Research Award, 2006, 2007, 2010
- Princeton Engineering Commendation List for Outstanding Teaching, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013
- U.C. Berkeley C.V. Ramamoorthy Distinguished Research Award, 2006
- Microsoft Research Graduate Fellowship, 2002
- Berkeley Micro-Electronics Fellowship, 1999
- Sigma Xi Scientific Honor Society, 1997

Publications

Journal Articles

1. S. He, Y. Jin, A. Nazaret, L. Shi, X. Chen, S. Rampersaud, B. Dhillon, I. Valdez, L. Friend, J. Fan, Y. Park, R. Mintz, Y. Lao, D. Carrera, K. Fang, K. Mehdi, M. Rohde, J. McFaline-Figueroa, D. Blei, K. Leong, A. Rudensky, G. Plitas, and E. Azizi. Starfysh integrates spatial transcriptomic and histologic data to reveal heterogeneous tumor–immune hubs. *Nature Biotechnology*, 2024.

2. M. Yin, C. Shi, Y. Wang, and D. Blei. Conformal sensitivity analysis for individual treatment effects. *Journal of the American Statistical Association*, 19(545); 122–135, 2024.

3. K. Vafa, E. Palikot, T. Du, A. Kanodia, S. Athey, and D. Blei. CAREER: A foundation model for labor sequence data. *Transactions on Machine Learning Research*, 2024.

4. G. Moran, D. Blei, and R. Ranganath. Holdout predictive checks for Bayesian model criticism. *Journal of the Royal Statistical Society, Series B*, 86(1):194–214, 2024.

5. G. Moran, J. Cunningham, and D. Blei. The posterior predictive null. *Bayesian Analysis*, 18(4):194–214, 2023.

6. C. Zheng, K. Vafa, and D. Blei. Revisiting topic-guided language models. *Transactions on Machine Learning Research* 2023.

7. L. Zhang, D. Blei, and C. Naesseth. Transport score climbing: Variational inference using forward KL and adaptive neural transport. *Transactions on Machine Learning Research*, 2023.

8. Y. Wang, D. Sridhar, and D. Blei. Adjusting machine learning decisions for equal opportunity and counterfactual fairness. *Transactions on Machine Learning Research*, 2023.

9. C. Modi, Y. Li, and D. Blei. Reconstructing the universe with variational self-boosted sampling. *Journal of Cosmology and Astroparticle Physics*, 059, 2023.

10. L. Zhang, Y. Wang, M. Schuemie, D. Blei, and G. Hripcsak. Adjusting for indirectly measured confounding using large-scale propensity score. *Journal of Biomedical Informatics*, 134, 2022.

11. D. Sridhar and D. Blei. Causal inference from text: A commentary. *Science Advances*, 8(42), 2022.

12. G. Moran, D. Sridhar, Y. Wang, and D. Blei. Identifiable deep generative models via sparse decoding. *Transactions on Machine Learning Research*, 2022.

13. A. Miller, L. Anderson, B. Leistedt, J. Cunningham, D. Hogg, and D. Blei. Mapping interstellar dust with Gaussian processes. *Annals of Applied Statistics*, 16(4):2672–2692, 2022.

14. W. Tansey, C. Tosh, and D. Blei. A Bayesian model of dose-response for cancer drug studies. *Annals of Applied Statistics*, 16(2):680–705, 2022.

15. D. Sridhar, H. Daume, and D. Blei. Heterogeneous supervised topic models for text prediction. *Transactions of the Association for Computational Linguistics*, 10:732–745, 2022.

16. W. Tansey, V. Veitch, H. Zhang, R. Rabadan, and D. Blei. The holdout randomization test for feature selection in black box models. *Journal of Computational and Graphical Statistics*, 31(1):151–162, 2022.

17. W. Tansey, K. Li, H. Zhang, S. Linderman, D. Blei, R. Rabadan, and C. Wiggins. Dose-response modeling in high-throughput cancer drug screenings: An end-to-end approach. *Biostatistics*, 23(2):643–665, 2022.

18. J. Loper, D. Blei, J. Cunningham, and L. Paninski. A general linear-time inference method for Gaussian processes on one dimension. *Journal of Machine Learning Research*, 22(234):1–36, 2021.

19. R. Donnelly, F. Ruiz, D. Blei, and S. Athey. Counterfactual inference for consumer choice across many product categories. *Quantitiative Marketing and Economics*, 19:369–407, 2021.

20. W. Tansey, Y. Wang, R. Rabadan, and D. Blei. Double empirical Bayes testing. *International Statistical Review*, 88, 2020.

21. A. Dieng, F. Ruiz, and D. Blei. Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*, 8:439–453, 2020.

22. F. Ruiz, S. Athey, and D. Blei. SHOPPER: A probabilistic model of consumer choice with substitutes and complements. *Annals of Applied Statistics*, 14(1):1–27, 2020. **Best of Annals of Applied Statistics, JSM 2020**

23. Y. Wang and D. Blei. The blessings of multiple causes. *Journal of the American Statistical Association*, 114(528):1574–1596, 2019 (with discussion and rejoinder). **Editor-selected JSM discussion paper, 2019**

24. Y. Wang and D. Blei. Frequentist consistency of variational Bayes. *Journal of the American Statistical Association*, 114:527, 1147–1161, 2019.

25. Y. Wang, A. Miller, and D. Blei. Comment: Variational autoencoders as empirical Bayes. *Statistical Science*, 34(2):229-233, 2019.

26. H. Levitin, J. Yuan, Y. Cheng, F. Ruiz, E. Bush, J. Bruce, P. Canoll, A. Iavarone, A. Lasorella, D. Blei, and P. Sims. De novo gene signature identification from single-cell RNA-seq with hierarchical Poisson factorization. *Molecular Systems Biology*, 15(e8557), 2019.

27. C. Wang and D. Blei. A general method for robust Bayesian modeling. *Bayesian Analysis*, 13(4):1163–1191, 2018.

28. J. Manning, X. Zhu, T. Willke, R. Ranganath, K. Stachenfeld, U. Hasson, D. Blei, and K. Norman. A probabilistic approach to discovering dynamic full-brain functional connectivity patterns. *NeuroImage*, 180:243–252, 2018.

29. R. Ranganath and D. Blei. Correlated random measures. *Journal of the American Statistical Association*, 113(521):417–430, 2018.

30. S. Athey, D. Blei, R. Donnelly, F. Ruiz, and T. Schmidt. Estimating heterogeneous consumer preferences for restaurants and travel time using mobile location data. *AEA Papers and Proceedings*, 108:64–67, 2018.

31. D. Blei. Expressive probabilistic models and scalable method of moments. *Communications of the ACM*, 61(4):84, 2018.

32. A. Gerow, Y. Hu, J. Boyd-Graber, D. Blei, and J. Evans. Measuring discursive influence across scholarship. *Proceedings of the National Academy of Sciences*, 115(13):3308–3313, 2018.

33. S. Linderman and D. Blei. A Discussion of "Nonparametric Bayes modeling of populations of networks". *Journal of the American Statistical Association*, 112(520):1543–1547, 2018.

34. S. Mandt, M. Hoffman, and D. Blei. Stochastic gradient descent as approximate Bayesian inference. *Journal of Machine Learning Research*, 18:1–35, 2017.

35. D. Blei and P. Smyth. Science and data science. Proceedings of the National Academy of Sciences, 114(33):8689–8692, 2017.

36. D. Blei, A. Kucukelbir, and J. McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518): 859–877, 2017.

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

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

39. P. Gopalan, W. Hao, D. Blei, and J. Storey. Scaling probabilistic models of genetic variation to millions of humans. *Nature Genetics*, 48 (1587–1590), 2016.

40. D. Blei. Comment on "Improving and evaluating topic models and other models of text." *Journal of the American Statistical Association*, 111(516):1408–1410, 2016.

41. D. Mimno, D. Blei, and B. Engelhardt. Posterior predictive checks to quantify lack-of-fit in admixture models of latent population structure. *Proceedings of the National Academy of Sciences*, 112(26), 2015.

42. A. Perotte, R. Ranganath, J. Hirsch, D. Blei, and N. Elhadad. Risk prediction for chronic kidney disease progression using heterogeneous electronic health record data and time series analysis. *Journal of the American Medical Informatics Association*, 22 (4), 2015.

43. J. Paisley, C. Wang, D. Blei, and M. Jordan. A nested HDP for hierarchical topic modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37 (2), 2015.

44. G. Polatkan, M. Zhou, L. Carin, D. Blei, and I. Daubechies. A Bayesian nonparametric approach to image super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37 (2), 2015.

45. S. Gershman, P. Frazier, and D. Blei. Distance dependent infinite latent feature models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37 (2), 2015.

46. D. Blei. Build, Compute, Critique, Repeat: Data Analysis with Latent Variable Models. *Annual Review of Statistics and Its Application*, 1 203–232, 2014.

47. S. Gershman, D. Blei, K. Norman, and P. Sederberg. Decomposing spatiotemporal brain patterns into topographic latent sources. *NeuroImage*, 98:91–102, 2014.

48. J. Manning, R. Ranganath, K. Norman, and D. Blei. Topographic factor analysis: A Bayesian model for inferring brain networks from neural data. *PLoS ONE*, 9(5), 2014.

49. P. Gopalan and D. Blei. Efficient discovery of overlapping communities in massive networks. *Proceedings of the National Academy of Sciences*, 110 (36) 14534–14539, 2013.

50. M. Hoffman, D. Blei, C. Wang, and J. Paisley. Stochastic variational inference. *Journal of Machine Learning Research*, 14:1303–1347, 2013.

51. C. Wang and D. Blei. Variational inference in nonconjugate models. *Journal of Machine Learning Research*, 14:1005–1031, 2013.

52. P. DiMaggio, M. Nag, and D. Blei. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics*, 41:6, 2013.

53. D. Blei. Topic modeling and digital humanities. *Journal of Digital Humanities*, 2(1), 2013.

54. D. Blei. Comment on "Multinomial inverse regression for text analysis. *Journal of the American Statistical Association*, 108 (503) 771–772, 2013.

55. B. Chen, G. Polatkan, G. Sapiro, D. Blei, D. Dunson, L. Carin. Deep learning with hierarchical convolutional factor analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35 (8), 2013.

56. J. Paisley, C. Wang and D. Blei. The discrete infinite logistic normal distribution. *Bayesian Analysis*, 7(2):235–272, 2012.

57. D. Blei. Probabilistic topic models. *Communications of the ACM*, 55(4):77–84, 2012.

58. S. Gershman and D. Blei. A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, 56:1–12, 2012.

59. D. Blei and P. Frazier. Distance dependent Chinese restaurant processes. *Journal of Machine Learning Research*, 12:2461–2488, 2011.

60. L. Hannah, D. Blei and W. Powell. Dirichlet process mixtures of generalized linear models. *Journal of Machine Learning Research*, 12:1923–1953, 2011.

61. S. Gershman, D. Blei, F. Pereira, and K. Norman. A topographic latent source model for fMRI data. *NeuroImage*, 57:89–100, 2011.

62. D. Blei, L. Carin, and D. Dunson. Probabilistic topic models. *Signal Processing*, 27(6):55–65, 2010.

63. D. Blei, T. Griffiths, and M. Jordan. The nested Chinese restaurant process and Bayesian nonparametric inference of topic hierarchies. *Journal of the ACM*, 57(2):1–30, 2010.

64. J. Chang and D. Blei. Hierarchical relational models for document networks. *Annals of Applied Statistics*, 4(1), 2010.

65. S. Gershman, D. Blei, and Y. Niv. Context, learning and extinction. *Psychological Review*, 117(1):197–209, 2010.

66. E. Airoldi, D. Blei, S. Fienberg, and E. Xing. Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, 9:1981–2014, 2008.

67. D. Blei and J. Lafferty. A correlated topic model of Science. *Annals of Applied Statistics*, 1(1):17–35, 2007.

68. D. Blei and S. Fienberg. Discussion of model-based clustering for social networks. *Journal of the Royal Statistical Society, Series A*, 170:332, 2007.

69. J. McAuliffe, D. Blei, and M. Jordan. Nonparametric empirical Bayes for the Dirichlet process mixture model. *Statistics and Computing*, 16(1):5–14, 2006.

70. Y. Teh, M. Jordan, M. Beal, and D. Blei. Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581, 2006.

71. D. Blei, K. Franks, M. Jordan, and S. Mian. Statistical modeling of biomedical corpora: Mining the Caenorhabditis Genetic Center Bibliography for genes related to life span. *BMC Bioinformatics*, 7(250), 2006.

72. D. Blei and M. Jordan. Variational inference for Dirichlet process mixtures. *Journal of Bayesian Analysis*, 1(1):121–144, 2005.

73. K. Barnard, P. Duygulu, N. de Freitas, D. Forsyth, D. Blei, and M. Jordan. Matching words and pictures. *Journal of Machine Learning Research*, 3:1107–1135, 2003.

74. D. Blei, A. Ng, and M. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, January 2003.

Conference Articles

75. A. Nazaret, C. Shi, and D. Blei. On the misspecification of linear assumptions in synthetic controls. In *Artificial Intelligence and Statistics*, 2024.

76. Y. Park and D. Blei. Density uncertainty layers for reliable uncertainty estimation. In *Artificial Intelligence and Statistics*, 2024.

77. C. de Bacco, Y. Wang, and D. Blei. A causality-inspired plus-minus model for player evaluation in team sports. In *Causal Learning and Reasoning*, 2024.

78. A. Feder, Y. Wald, C. Shi, S. Saria, and D. Blei. Causal-structure driven augmentations for text OOD generalization. In *Neural Information Processing Systems*, 2023.

79. C. Modi, R. Gower, C. Margossian, Y. Yao, D. Blei, and L. Saul. Variational inference with Gaussian score matching. In *Neural Information Processing Systems*, 2023.

80. N. Scherrer, C. Shi, A. Feder, and D. Blei. Evaluating the moral beliefs encoded in LLMs. In *Neural Information Processing Systems*, 2023.

81. J. von Kügelgen, M. Besserve, W. Liang, L. Gresele, A. Kekić, E. Bareinboim, D. Blei, and B. Schölkopf. Nonparametric identifiability of causal representations from unknown interventions. In *Neural Information Processing Systems*, 2023.

82. L. Wu, B. Trippe, C. Naesseth, D. Blei, and J. Cunningham. Practical and asymptotically exact conditional sampling in diffusion models. In *Neural Information Processing Systems*, 2023.

83. C. Zheng, C. Shi, K. Vafa, A. Feder, D. Blei. An invariant learning characterization of controlled text generation. In *Association for Computational Linguistics*, 2023.

84. Z. Wang, R. Gao, M. Yin, M. Zhou, and D. Blei. Probabilistic conformal prediction using conditional random samples. In *Artificial Intelligence and Statistics*, 2023.

85. A. Nazaret and D. Blei. Variational inference for infinitely deep neural networks. In *International Conference on Machine Learning*, 2022.

86. S. Menon, D. Blei, and C. Vondrick. Forget-me-not! Contrastive critics for mitigating posterior collapse. In *Uncertainty in Artificial Intelligence*, 2022.

87. C. Shi, D. Sridhar, V. Misra, and D. Blei. On the assumptions of synthetic control methods. In *Artificial Intelligence and Statistics*, 2022.

88. D. Sridhar, C. D. Bacco, and D. Blei. Estimating social influence from observational data. In *Causal Learning and Reasoning*, 2022.

89. Y. Wang, D. Blei, and J. Cunningham. Posterior collapse and latent variable non-identifiability. In *Neural Information Processing Systems*, 2021.

90. Y. Park, S. Lee, G. Kim, and D. Blei. Unsupervised representation learning via neural activation coding. In *International Conference on Machine Learning*, 2021.

91. Y. Wang and D. Blei. A proxy variable view of shared confounding. In *International Conference on Machine Learning*, 2021.

92. A. Moretti and L. Zhang and C. Naesseth and H. Venner and D. Blei and I. Pe'er. Variational combinatorial sequential Monte Carlo methods for Bayesian phylogenetic inference. In *Uncertainty in Artificial Intelligence*, 2021.

93. C. Shi, V. Veitch, and D. Blei. Invariant representation learning for treatment effect estimation. *Uncertainty in Artificial Intelligence*, 2021.

94. L. Wu, A. Miller, L. Anderson, G. Pleiss, D. Blei, and J. Cunningham. Hierachical inducing point Gaussian process for inter-domian observations. In *Artificial Intelligence and Statistics*, 2021.

95. A. Schein, K. Vafa, D. Sridhar, V. Veitch, J. Quinn, J. Moffet, D. Blei, and D. Green. Assessing the effects of friend-to-friend texting on turnout in the 2018 US midterm elections. In *The Web Conference*, 2021.

96. K. Vafa, Y. Deng, D. Blei, and A. Rush. Rationales for sequential predictions. In *Empirical Methods in Natural Language Processing*, 2021.

97. C. Naesseth, F. Lindsten, D. Blei. Markovian score climbing: Variational inference with KL(p||q). In *Neural Information Processing Systems*, 2020.

98. Y. Wang, D. Liang, L. Charlin, and D. Blei. Causal inference for recommender systems. In *ACM Conference on Recommender Systems*, 2020.

99. K. Vafa, S. Naidu, and D. Blei. Text-based ideal points. In *Association for Computational Linguistics*, 2020.

100. V. Veitch, D. Sridhar, and D. Blei. Adapting text embeddings for causal inference. In *Uncertainty in Artificial Intelligence*, 2020.

101. C. Shi, D. Blei, and V. Veitch. Adapting neural networks for the estimation of treatment effects. In *Neural Information Processing Systems*, 2019.

102. V. Veitch, Y. Wang, and D. Blei. Using embeddings to correct for unobserved confounding in networks. In *Neural Information Processing Systems*, 2019.

103. Y. Wang and D. Blei. Variational Bayes under model misspecification. In *Neural Information Processing Systems*, 2019.

104. A. Schein, S. Linderman, M. Zhou, D. Blei, and H. Wallach. Poisson-randomized gamma dynamical systems. In Neural Information Processing Systems, 2019.

105. L. Zhang, Y. Wang, A. Ostropolets, J. Mulgrave, D. Blei, and G. Hripcsak. The medical deconfounder: Assessing treatment effects with electronic health records. In *Machine Learning for Health Care*, 2019.

106. V. Veitch, M. Austern, W. Zhou, D. Blei, and P. Orbanz. Empirical risk minimization and stochastic gradient descent for relational data. In *Artificial Intelligence and Statistics*, 2019.

107. A. Dieng, Y. Kim, A. Rush, and D. Blei. Avoiding latent variable collapse with generative skip models. In *Artificial Intelligence and Statistics*, 2019.

108. F. Ruiz, M. Titsias, A. Dieng, and D. Blei. Augment and reduce: Stochastic inference for large categorical distributions. *In International Conference on Machine Learning*, 2018.

109. A. Dieng, R. Ranganath, J. Altosaar, and D. Blei. Noisin: Unbiased regularization for recurrent neural networks. *In International Conference on Machine Learning*, 2018.

110. W. Tansey, Y. Wang, D. Blei, and R. Rabadan. Black box FDR. *In International Conference on Machine Learning*, 2018.

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

112. M. Rudolph and D. Blei. Dynamic embeddings for language evolution. *In International World Wide Web Conference*, 2018.

113. J. Altosaar, R. Ranganath, and D. Blei. Proximity variational inference. *In Artificial Intelligence and Statistics*, 2018.

114. C. Naesseth, S. Linderman, R. Ranganath, and D. Blei. Variational sequential Monte Carlo. *In Artificial Intelligence and Statistics*, 2018.

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

116. L. Liu, F. Ruiz, and D. Blei. Context selection for embedding models. *In Neural Information Processing Systems*, 2017.

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

118. M. Rudolph, F. Ruiz, and D. Blei. Structured embedding models for grouped data. *In Neural Information Processing Systems*, 2017.

119. A. Kucukelbir, Y. Wang, and D. Blei. Evaluating Bayesian models with posterior dispersion indices. *In International Conference on Machine Learning*, 2017.

120. Y. Wang, A. Kucukelbir, and D. Blei. Robust probabilistic modeling with Bayesian data reweighting. *In International Conference on Machine Learning*, 2017.

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

122. C. Naesseth, F. Ruiz, S. Linderman, and D. Blei. Reparameterization gradients through acceptance- rejection sampling algorithms. *In Artificial Intelligence and Statistics*, 2017. **Best Student Paper Award.**

123. S. Linderman, M. Johnson, A. Miller, R. Adams, D. Blei, and L. Paninski. Bayesian learning and inference in recurrent switching linear dynamical systems. In *Artificial Intelligence and Statistics*, 2017.

124. A. Chaney, H. Wallach, M. Connelly, and D. Blei. Detecting and Characterizing Events. *In Empirical Methods in Natural Language Processing*, 2016.

125. F. Ruiz, M. Titsias, D. Blei. The generalized reparameterization gradient. *In Neural Information Processing Systems*, 2016.

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

127. M. Rudolph, F. Ruiz, S. Mandt, and D. Blei. Exponential family embeddings. *In Neural Information Processing Systems*, 2016.

128. R. Ranganath, A. Perotte, N. Elhadad, and D. Blei. Deep survival analysis. *Machine Learning for Health Care*, 2016.

129. D. Liang, J. Altosaar, L. Charlin, and D. Blei. Factorization meets the item embedding: Regularizing matrix factorization with item co-occurrence. *In ACM Conference on Recommendation Systems*, 2016.

130. F. Ruiz, M. Titsias, and D. Blei. Overdispersed black-box variational inference. In *Uncertainty in Artificial Intelligence*, 2016.

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

132. A. Schein, M. Zhou, D. Blei, and H. Wallach. Bayesian Poisson Tucker decomposition for learning the structure of international relations. In *International Conference on Machine Learning*, 2016.

133. S. Mandt, M. Hoffman, and D. Blei. A variational analysis of stochastic gradient algorithms. In *International Conference on Machine Learning*, 2016.

134. D. Tran, R. Ranganath, and D. Blei. The variational Gaussian process. In *International Conference on Learning and Representation*, 2016.

135. D. Liang, L. Charlin, J. McInerney, D. Blei. Modeling user exposure in recommendation. In *International World Wide Web Conference*, 2016.

136. M. Rudolph, J. Ellis, and D. Blei. Objective variables for probabilistic revenue maximization in second-price auctions with reserve. In *International World Wide Web Conference*, 2016.

137. S. Mandt, J. McInerney, F. Abrol, R. Ranganath, and D. Blei. Variational tempering. In *Artificial Intelligence and Statistics*, 2016.

138. D. Tran, D. Blei, and E. Airoldi. Variational inference with copula augmentation. In *Neural Information Processing Systems*, 2015.

139. A. Kucukelbir, R. Ranganath, A. Gelman, and D. Blei. Automatic variational inference in Stan. In *Neural Information Processing Systems*, 2015.

140. J. McInerney, R. Ranganath, and D. Blei. The population posterior and Bayesian inference on streams. In *Neural Information Processing Systems*, 2015.

141. L. Charlin, R. Ranganath, J. McInerney, and D. Blei. Dynamic Poisson factorization. In *ACM Conference on Recommendation Systems*, 2015.

142. A. Chaney and D. Blei and T. Elassi-Rad. A probabilistic model for using social networks in personalized item recommendation. In *ACM Conference on Recommendation Systems*, 2015.

143. P. Gopalan, J. Hofman, and D. Blei. Scalable recommendation with hierarchical Poisson factorization. In *Uncertainty in Artificial Intelligence*, 2015.

144. R. Ranganath, A. Perotte, N. Elhadad, and D. Blei. The survival filter: Joint survival analysis with a latent time series. In *Uncertainty in Artificial Intelligence*, 2015.

145. A. Kucukelbir and D. Blei. Population empirical Bayes. In *Uncertainty in Artificial Intelligence*, 2015.

146. A. Schein, J. Paisley, D. Blei, and H. Wallach. Bayesian Poisson tensor factorization for inferring multilateral relations from sparse dyadic event counts. In *Knowledge Discovery and Data Mining*, 2015.

147. M. Hoffman and D. Blei. Structured stochastic variational inference. In *Artificial Intelligence and Statistics*, 2015.

148. R. Ranganath, L. Tang, L. Charlin, and D. Blei. Deep exponential families. In *Artificial Intelligence and Statistics*, 2015.

149. N. Houlsby and D. Blei. A filtering approach to stochastic variational inference. In *Neural Information Processing Systems*, 2014.

150. S. Mandt and D. Blei. Smoothed gradients for stochastic variational inference. In *Neural Information Processing Systems*, 2014.

151. P. Gopalan, L. Charlin, and D. Blei. Content based recommendations with Poisson factorization. In *Neural Information Processing Systems*, 2014.

152. R. Ranganath, S. Gerrish, and D. Blei. Black box variational inference. In *Artificial Intelligence and Statistics*, 2014.

153. P. Gopalan, F. Ruiz, R. Ranganath, and D. Blei. Bayesian nonparametric Poisson factorization for recommendation systems. In *Artificial Intelligence and Statistics*, 2014.

154. M. Rabinovich and D. Blei. The inverse regression topic model. In *International Conference on Machine Learning*, 2014.

155. P. Gopalan, C. Wang and D. Blei. Modeling overlapping communities with node popularities. In *Neural Information Processing Systems*, 2013.

156. D. Kim, P. Gopalan, D. Blei, and E. Sudderth. Efficient online inference for Bayesian nonparametric relational models. In *Neural Information Processing Systems*, 2013.

157. R. Ranganath, C. Wang, D. Blei, and E. Xing. An adaptive learning rate for stochastic variational inference. In *International Conference on Machine Learning*, 2013.

158. P. Gopalan, D. Mimno, S. Gerrish, M. Freedman, and D Blei. Scalable inference of overlapping communities. In *Neural Information Processing Systems*, 2012.

159. S. Gerrish and D. Blei. How they vote: Issue-adjusted models of legislative behavior. In *Neural Information Processing Systems*, 2012.

160. C. Wang and D. Blei. Truncation-free online variational inference for Bayesian nonparametric models. In *Neural Information Processing Systems*, 2012.

161. J. Paisley, D. Blei and M. Jordan. Variational Bayesian inference with stochastic search. In *International Conference On Machine Learning*, 2012.

162. D. Mimno, M. Hoffman and D. Blei. Sparse stochastic inference for latent Dirichlet allocation. In *International Conference On Machine Learning*, 2012.

163. S. Gershman, M. Hoffman and D. Blei. Nonparametric variational inference. In *International Conference On Machine Learning*, 2012.

164. A. Chaney and D. Blei. Visualizing topic models. In *International AAAI Conference on Weblogs and Social Media*, 2012.

165. J. Paisley, D. Blei, and M. Jordan. Stick-breaking beta processes and the Poisson process. In *Artificial Intelligence and Statistics*, 2012.

166. S. Ghosh, A. Ungureanu, E. Sudderth, and D. Blei. A Spatial distance dependent Chinese restaurant process for image segmentation. In *Neural Information Processing Systems*, 2011.

167. C. Wang and D. Blei. Collaborative topic modeling for recommending scientific articles. In *Knowledge Discovery and Data Mining*, 2011. **Best Student Paper Award. 2021 KDD Test of Time Award.**

168. D. Mimno and D. Blei. Bayesian checking for topic models. In *Empirical Methods in Natural Language Processing*, 2011.

169. S. Gerrish and D. Blei. Predicting legislative roll call from text. In *International Conference on Machine Learning*, 2011. **Distinguished Application Paper Award.**

170. J. Paisley, D. Blei, and L. Carin. Variational inference for stick-breaking beta process priors. In *International Conference on Machine Learning*, 2011.

171. J. Paisley, C. Wang and D. Blei. The discrete infinite logistic normal distribution for mixed-membership modeling. In *Artificial Intelligence and Statistics*, 2011. **Notable Paper Award.**

172. C. Wang, J. Paisley and D. Blei. Online variational inference for the hierarchical Dirichlet process. In *Artificial Intelligence and Statistics*, 2011.

173. M. Hoffman, D. Blei, and F. Bach. On-line learning for latent Dirichlet allocation. In *Neural Information Processing Systems*, 2010. **2021 NeurIPS Test of Time Award.**

174. L. Hannah, W. Powell, and D. Blei. Nonparametric density estimation for stochastic optimization with an observable state variable. In *Neural Information Processing Systems*, 2010.

175. D. Blei and P. Frazier. Distance dependent Chinese restaurant processes. In *International Conference on Machine Learning*, 2010.

176. S. Gerrish and D. Blei. A language-based approach to measuring scholarly impact. In *International Conference on Machine Learning*, 2010.

177. M. Hoffman, D. Blei, and P. Cook. Bayesian nonparametric matrix factorization for recorded music. In *International Conference on Machine Learning*, 2010.

178. S. Williamson, C. Wang, K. Heller, and D. Blei. The IBP compound Dirichlet process and its application to focused topic modeling. In *International Conference on Machine Learning*, 2010.

179. L. Hannah, D. Blei, and W. Powell. Dirichlet process mixtures of generalized linear models. In *Artificial Intelligence and Statistics*, 2010.

180. A. Lorbert, D. Eis, V. Kostina, D. Blei, and P. Ramadge. Exploiting covariate similarity in sparse regression via the pairwise elastic net. In *Artificial Intelligence and Statistics*, 2010.

181. J. Li, C. Wang, Y. Lim, D. Blei, and L. Fei-Fei. Building and using a semantivisual image hierarchy. In *Computer Vision and Pattern Recognition*, 2010.

182. S. Cohen, D. Blei, and N. Smith. Variational inference for adaptor grammars. In *North American Chapter of the Association for Computational Linguistics*, 2010.

183. C. Wang and D. Blei. Decoupling sparsity and smoothness in the discrete hierarchical Dirichlet process. In *Neural Information Processing Systems*, 2009.

184. C. Wang and D. Blei. Variational inference for the nested Chinese restaurant process. In *Neural Information Processing Systems*, 2009.

185. R. Socher, S. Gershman, A. Perotte, P. Sederberg, D. Blei, and K. Norman. A Bayesian analysis of dynamics in free recall. In *Neural Information Processing Systems*, 2009.

186. J. Chang, J. Boyd-Graber, S. Gerrish, C. Wang, and D. Blei. Reading tea leaves: How humans interpret topic models. In *Neural Information Processing Systems*, 2009. **Honorable Mention: Best Student Paper Award.**

187. J. Chang, J. Boyd-Graber, and D. Blei. Connections between the lines: Augmenting social networks with text. In *Knowledge Discovery and Data Mining*, 2009.

188. J. Boyd-Graber and D. Blei. Multilingual topic models for unaligned text. In *Uncertainty in Artificial Intelligence*, 2009.

189. J. Chang and D. Blei. Relational topic models for document networks. In *Artificial Intelligence and Statistics*, 2009.

190. C. Wang, B. Thiesson, C. Meek, and D. Blei. Markov topic models. In *Artificial Intelligence and Statistics*, 2009.

191. M. Hoffman, D. Blei, and P. Cook. Finding latent sources in recorded music with a shift-invariant HDP. In *International Conference on Digital Audio Effects*, 2009.

192. M. Hoffman, D. Blei, and P. Cook. Easy as CBA: A simple probabilistic model for tagging music. In *International Conference on Music Information Retrieval*, 2009. **Best Student Paper Award.**

193. M. Hoffman, P. Cook, and D. Blei. Bayesian spectral matching: Turning young MC into MC hammer via MCMC sampling. In *International Computer Music Conference*, 2009.

194. C. Wang, D. Blei, and L. Fei-Fei. Simultaneous image classification and annotation. In *Computer Vision and Pattern Recognition*, 2009.

195. I. Mukherjee and D. Blei. Relative performance guarantees for approximate inference in latent Dirichlet allocation. In *Neural Information Processing Systems*, 2008.

196. J. Boyd-Graber and D. Blei. Syntactic topic models. In *Neural Information Processing Systems*, 2008.

197. E. Airoldi, D. Blei, S. Fienberg, and E. Xing. Mixed membership stochastic blockmodels. In *Neural Information Processing Systems*, 2008.

198. C. Wang, D. Blei, and D. Heckerman. Continuous time dynamic topic models. In *Uncertainty in Artificial Intelligence (UAI)*, 2008.

199. M. Hoffman, D. Blei, and P. Cook. Content-based musical similarity computation using the hierarchical Dirichlet process. In *International Conference on Music Information Retrieval*, 2008.

200. M. Hoffman, P. Cook, and D. Blei. Data-driven recomposition using the hierarchical Dirichlet process hidden Markov model. In *International Computer Music Conference*, 2008.

201. M. Dudik, D. Blei, and R. Schapire. Hierarchical maximum entropy density estimation. In *Proceedings of the 28th International Conference on Machine Learning*, 2007.

202. W. Li, D. Blei, and A. McCallum. Nonparametric Bayes pachinko allocation. In *The 23rd Conference on Uncertainty in Artificial Intelligence*, 2007.

203. D. Kaplan and D. Blei. A computational approach to style in American poetry. In *IEEE Conference on Data Mining*, 2007.

204. D. Blei and J. McAuliffe. Supervised topic models. In *Neural Information Processing Systems*, 2007.

205. J. Boyd-Graber, D. Blei, and X. Zhu. A topic model for word sense disambiguation. In *Empirical Methods in Natural Language Processing*, 2007.

206. D. Blei and J. Lafferty. Correlated topic models. In *Neural Information Processing Systems*, 2006.

207. D. Blei and J. Lafferty. Dynamic topic models. In *International Conference on Machine Learning*, 2006. **ICML 2016 Test of Time Award.**

208. T. Griffiths, M. Steyvers, D. Blei, and J. Tenenbaum. Integrating topics and syntax. In *Neural Information Processing Systems*, 2005.

209. D. Blei and M. Jordan. Variational methods for the Dirichlet process. In *International Conference on Machine Learning*, 2004.

210. D. Blei and M. Jordan. Modeling annotated data. In *ACM SIGIR Conference on Research and Development in Information Retrieval*, 2003. **SIGIR Test of Time Award (Honorable Mention).**

211. D. Blei, T. Griffiths, M. Jordan, and J. Tenenbaum. Hierarchical topic models and the nested Chinese restaurant process. In *Neural Information Processing Systems*, 2003. **Best Student Paper Award.**

212. D. Blei, A. Ng, and M. Jordan. Latent Dirichlet allocation. In *Neural Information Processing Systems*, 2002.

213. D. Blei, J. Bagnell, and A. McCallum. Learning with scope, with application to information extraction and classification. In *Uncertainty in Artificial Intelligence*, 2002.

214. D. Blei and P. Moreno. Topic segmentation with an aspect hidden Markov model. In *ACM SIGIR conference on Research and Development in Information Retrieval*, 2001.

Book Chapters

215. E. Airoldi, D. Blei, E. Erosheva, and S. Fienberg. Introduction to Mixed Membership Models and Methods. In *Handbook of Mixed-Membership Models and Their Applications*. Chapman & Hall/CRC, 2014.

216. S. Williamson, C. Wang, K. Heller, and D. Blei. Nonparametric mixed membership models using the IBP compound Dirichlet process. In K. Mengerson, C. Robert, and D. Titterington, editors, *Mixture Estimation and Applications*. John Wiley and Sons, 2011.

217. D. Blei and J. Lafferty. Topic models. In A. Srivastava and M. Sahami, editors, *Text Mining: Classification, Clustering, and Applications*. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series, 2009.

218. E. Airoldi, D. Blei, S. Fienberg, and E. Xing. Combining stochastic block models and mixed membership for statistical network analysis. In *Statistical Network Analysis: Models, Issues and New Directions*, Lecture Notes in Computer Science, pages 57–74. Springer-Verlag, 2007.

219. D. Blei, A. Ng, and M. Jordan. Hierarchical Bayesian models for applications in information retrieval. In J. Bernardo, J. Berger, A. Dawid, D. Heckerman, A. Smith, and M. West, editors, *Bayesian Statistics 7*, volume 7, pages 25–44. Oxford University Press, 2003.

Edited Volumes

220. E. Airoldi, D. Blei, E. Erosheva, and S. Fienberg, editors. *Handbook of Mixed-Membership Models and Their Applications*. Chapman and Hall/CRC, 2014.

221. E. Airoldi, D. Blei, S. Fienberg, A. Goldenberg, E. Xing, and A. Zheng, editors. *Statistical Network Analysis: Models, Issues and New Directions*. Lecture Notes in Computer Science. Springer-Verlag, 2007.

Awarded Grants

1. *New Directions in Bayesian Model Criticism* (PI). National Science Foundation. $225K. (2023-2026)

2. *New Directions in Probabilistic Deep Learning: Exponential Families, Bayesian Nonparametrics and Empirical Bayes* (PI). National Science Foundation. $700K. (2021-2024)

3. *Modern Probabilistic Models for Modern Deep Learning* (PI). Office of Naval Research. $400K. (2020-2023).

4. *Interpretable and Robust Artificial Intelligence Software* (PI). Air Force Office of Scientific Research. $1M. (2018-2022)

5. *TRIPODS: From Foundations to Practice of Data Science and Back* (Co-PI). National Science Foundation. $500K. (2017-2020)

6. *Next-Generational Variational Methods: Active Inference, Streaming Inference, and Assessing Model Fitness* (PI). Office of Naval Research. $1M. 2015-2020.

7. *Extracting Mathematical Knowledge from the Scientific Literature: Statistical Machine Learning Models and Tools* (Co-PI). Sloan Foundation. $900K. 2015-2018.

8. *Deciphering the Cortex: Circuit Inference from Large-Scale Brain Activity Data* (Co-PI). Defense Advanced Research Projects Agency. $1M. 2015-2016.

9. *Estimating Multidimensional Influence in Science and Scholarship* (PI). Templeton Foundation. $100K. 2014-2015.

10. *The Next Generation of Probabilistic Programming: Massive Data, Data Streams, and Model Diagnostics* (PI). Defense Advanced Research Project Agency. $1.8M. 2013-2017.

11. *BIGDATA: Discovery and Social Analytics for Large-Scale Scientific Literature* (Co-PI). National Science Foundation. $1M. 2013-2015.

12. *Scalable Topic Modeling: Online Learning, Diagnostics, and Recommendation* (PI). Office of Naval Research. $510K. 2011-2014.

13. *Text, Neuroimaging, and Memory: Unified Models of Corpora and Cognition* (PI). National Science Foundation. $730K. 2010-2013.

14. *Non-Parametric Bayesian Analysis of Heterogeneous Data* (PI). Air Force Office of Scientific Research. $360K. 2009-2012.

15. *Dynamic and Supervised Topic Models for Literature-Based Discovery* (PI). Office of Naval Research. $300K. 2008-2011.

16. *CAREER: New Directions in Probabilistic Topic Models* (PI). National Science Foundation. $550K. 2008-2013.

17. *Interactive Discovery and Semantic Labeling of Patterns in Spatial Data* (Co-PI). National Science Foundation. $500K. 2009-2012.

Courses

- Causality, Field Experiments, and Machine Learning, Spring 2024
- Probabilistic Models and Machine Learning, Fall 2023
- Applied Causality, Spring 2023
- Probabilistic Models and Machine Learning, Fall 2022
- Applied Causality, Spring 2021
- Foundations of Graphical Models, Fall 2020
- Representation Learning: A Probabilistic Perspective, Spring 2020
- Foundations of Graphical Models, Fall 2019
- Applied Causality, Spring 2019
- Foundations of Graphical Models, Fall 2018
- Applied Causality, Spring 2017
- Foundations of Graphical Models, Fall 2016
- Probabilistic Models of Discrete Data, Spring 2016
- Foundations of Graphical Models, Fall 2015
- Truth in Data, Spring 2015
- Foundations of Graphical Models, Fall 2014
- Interacting with Data, Spring 2014
- Advanced Methods in Probabilistic Modeling, Fall 2013
- Interacting with Data, Spring 2012
- Advanced Methods in Probabilistic Modeling, Fall 2011
- Foundations of Probabilistic Modeling, Fall 2010
- Truth in Data, Fall 2009
- Foundations of Probabilistic Modeling, Spring 2009
- Interacting with Data, Spring 2008
- Bayesian Nonparametrics, Fall 2007
- Interacting with Data, Spring 2007

- Introduction to Artificial Intelligence, Fall 2006
- Approximate Inference in Graphical Models, Spring 2006

Professional Activities

- **Editor-in-Chief**
  - Journal of Machine Learning Research (2018–2023)
- **Senior Program Committee**
  - Causal Learning and Representations (2022)
  - International Conference on Machine Learning, 2015 Program Co-chair
  - International Conference on Machine Learning (2008, 2009, 2010, 2011, 2012, 2013, 2014, 2016, 2017)
  - Neural Information Processing Systems (2009, 2010, 2014)
  - Bayesian Nonparametrics (2017)
  - Artificial Intelligence and Statistics (2008, 2012, 2014)
- **Associate Editor and Editorial Board**
  - Journal of Machine Learning Research (2008–2018)
  - Chapman Hall Series on Computer Science and Data Analysis (2008–2019)
  - International Machine Learning Society Board (2013–2023)
  - Journal of the American Statistical Association (2017–2018)
  - IEEE Transactions on Pattern Analysis and Machine Intelligence (2014–2016)
  - Statistics and Computing (2009–2013)
- **Journal Reviewing and Editorial Board**
  - Proceedings of the National Academy of Science
  - Science Magazine
  - Journal of Machine Learning Research
  - Journal of the American Statistical Association
  - Journal of the Royal Statistical Society
  - Foundations and Trends in Machine Learning
  - Annals of Applied Statistics
  - Bayesian Analysis
  - Statistics and Computing
  - Machine Learning Journal
  - Journal of Artificial Intelligence Research
  - IEEE Transactions on Pattern Analysis and Machine Intelligence
  - IEEE Transactions on Neural Networks
  - IEEE Transactions on Audio, Speech, and Language Processing
  - International Journal on Very Large Data Bases
  - ACM Transactions on Knowledge Discovery from Data

- **Conference Reviewing**
  - Neural Information Processing Systems (2005, 2006, 2007, 2008, 2011, 2012, 2013)
  - Artificial Intelligence and Statistics (2005, 2007, 2010)
  - International Conference on Machine Learning (2006, 2007)
  - Uncertainty in Artificial Intelligence (2005, 2006, 2007)
  - Association of Computational Linguistics (2008)
  - Empirical Methods in Natural Language Processing (2007)
  - Association of Artificial Intelligence (2007)
  - International Joint Conference on Artificial Intelligence (2005)
  - SIGIR Conference on Information Retrieval (2005)
  - Knowledge Discovery and Data Mining (2005)
- **Columbia University**
  - Director of Graduate Studies, Data Science Institute
  - Executive Committee of the Data Science Institute
  - Education Committee of the Data Science Institute
  - Senior Search Committee of the Data Science Institute
  - Digital Humanities Task Force
  - Internal Review Committee for the Economics Department
  - Vision committee (Statistics)
  - Search committee (Statistics and Computer Science)
  - Data Science committee (Statistics)
  - PhD committee (Statistics and Computer Science)
  - Distinguished lecture committee (Computer Science)
- **Princeton University**
  - Executive Committee for the Committee on Statistical Studies
  - Faculty Advisory Committee on Athletics and Campus Recreation
  - Advisor to Computer Science A.B. classes of 2009 and 2010
  - Program in Applied and Computational Mathematics
  - Princeton Institute for Computational Science and Engineering, Associated Faculty
  - Center for Information Technology Policy, Affiliated Faculty
  - Princeton Neuroscience Institute, Affiliated Faculty
- **Organizing**
  - Machine Learning in NYC, talk series (2019-present)
  - "Deep Generative Models for Highly Structured Data" (ICLR, 2019)
  - "Implicit Probabilistic Models" (ICML, 2017)
  - "Computational Challenges in Machine Learning" (Simons Institute, 2017)
  - "Mathematical Analysis of Cultural Expressive Forms" (IPAM, 2016)
  - "Advanced in Approximate Bayesian Inference" (NIPS, 2015)
  - "Advances in Variational Inference" (NIPS, 2014)
  - "Topic Models: Computation, Application, and Evaluation" (NIPS, 2013)

  - "Statistics and Machine Learning at Princeton" (Princeton University, 2011)
  - "Applications of Topic Modeling" (NIPS, 2008)
  - "Statistical Network Analysis" (ICML, 2005)
  - "Syntax and Semantics" (NIPS, 2003)
- **Current Ph.D. Students**
  - Casey Bradshaw
  - Sweta Karlekar
  - Achille Nazaret
  - Yookoon Park
  - Claudia Shi
  - Bohan Wu
  - Luhuan Wu
  - Carolina Zheng
- **Current Postdoctoral Fellows**
  - Amir Feder
  - Alessandro Grande
  - Andrew Jesson
  - Brian Trippe
  - Eli Weinstein
- **Former Ph.D. Students**
  - Linying Zhang (2023); Assistant Professor, Washington University
  - Keyon Vafa (2023); Postdoctoal Fellow, Harvard University
  - Alessandro Grande (2022); Postdoctoral Fellow, Memorial Sloan Kettering
  - Dustin Tran (2020); Research Scientist, Google
  - Yixin Wang (2020); Assistant Professor, University of Michigan
  - Adji Dieng (2020); Assistant Professor, Princeton University
  - Jaan Aaltosar (2020) ; Postdoctoral fellow, Columbia University
  - Maja Rudolph (2018); Research Scientist, Bosch Research
  - Rajesh Ranganath (2017); Assistant Professor, New York University
  - Allison Chaney (2016); Assistant Professor, Duke University
  - Prem Gopalan (2014); Voleon Capital
  - Sean Gerrish (2012); Data Scientist, Google
  - Samuel Gershman (2012); Professor, Harvard University
  - Gungor Polatkan (2012); Data Scientist, Twitter
  - Chong Wang (2012); Senior Research Scientist, Apple
  - Jonathan Chang (2011); Data Scientist, Facebook
  - Matthew Hoffman (2010); Research Scientist, Google
  - Lauren Hannah (2010); Apple Computer
  - Jordan Boyd-Graber (2009); Associate Professor, University of Maryland
- **Former Postdoctoral Fellows**

  - Laurent Charlin; Assistant Professor, University of Montreal
  - Alp Kucukelbir; Fero Labs, New York NY
  - Jeremy Manning; Assistant Professor, Dartmouth College
  - Stephan Mandt; Associate Professor, University of California Irvine
  - James McInerney; Netflix, San Francisco CA
  - David Mimno; Professor, Cornell University
  - Scott Linderman; Assistant Professor, Stanford University
  - Liping Liu; Assistant Professor, Tufts University
  - Gemma Moran; Assistant Professor, Rutgers University
  - Andrew Miller; Apple Computer
  - Christian Naesseth; Assistant Professor, University of Amsterdam
  - John Paisley; Associate Professor, Columbia University
  - Kriste Krstovski; Associate Research Scientist, Columbia University
  - Jackson Loper; Postdoctoral fellow, University of Michigan
  - Francisco Ruiz; Research Scientist, DeepMind
  - Aaron Schein; Assistant Professor, University of Chicago
  - Dhanya Sridhar; Assistant Professor, University of Montreal
  - Wesley Tansey; Principal Investigator, Memorial Sloan-Kettering
  - Victor Vietch; Assistant Professor, University of Chicago
  - Mingzhang Yin; Assistant Professor, University of Florida

- **Professional Memberships**
  - Association of Computing Machinery
  - Institute for Mathematical Statistics
  - American Statistical Association
  - Bernoulli Society
  - International Society of Bayesian Analysis

- **Advising and Consulting**
  - Visiting Researcher, Flatiron Institute (2017-present)
  - Scientific Advisor, CNN and Warner Media (2020-present)
  - Scientific Advisor, Fero Labs (2017–present)
  - Scientific Advisor, Gamalon Technologies (2015–2022)
  - Scientific Advisor, Base10 (2017–2020)
  - Scientific Advisor, Canopy (2017–2020, acquired by CNN)
  - Scientific Advisor, Liftlighter (2015–2020)
  - Scientific Advisor, Undecidable Labs (2015–2016, Acquired by Google)
  - Scientific Advisor, Recruit Artificial Intelligence Laboratories (2015–2017)
  - Scientific Advisor, MyRoll (2015–2016)
  - Scientific Advisor, VoxGov (2014–2016)
  - Steering Committee, Declassification Engine (2014–2016)
  - Scientific Advisor, Applied Communications Sciences (2012–2013)
  - Consulting Researcher, Microsoft Research (2013, 2014)

– Scientific Advisor, Chomp (2011–2012, Acquired by Apple)

Invited Talks

- **Year 2024**
  1. Statistics Seminar, Stanford
  2. Neyman Seminar, Berkeley
- **Year 2023**
  1. Keynote Speaker, Temple Data Science Conference
  2. Invited Lecture, Berkeley CLIMB Seminar
  3. Distinguished Lecture, Berkeley Institute for Data Science
  4. Distinguished Lecture, UCSD
  5. Machine Learning Seminar, UCLA
  6. Machine Learning Seminar, Cornell Tech
  7. StatLearn Conference, Montpelier France, Keynote
  8. Special Quarter on Causality, Institut Henri Poincare, Paris France, Keynote
  9. INRIA Paris Machine Learning Seminar
  10. Mathematical and Computational Methods in Cancer and Biology
- **Year 2022**
  11. Columbia Applied Probability Seminar
  12. University of Connecticut Statistics 60th Anniversary, Keynote
  13. Columbia Statistics PhD Student Seminar
  14. Princeton Workshop on Synthetic Controls
  15. Memorial Sloan Kettering Machine Learning Seminar
- **Year 2021**
  16. NeurIPS 2021 Test of Time Award Talk
  17. NeurIPS 2021 Workshop on Causality and Machine Learning
  18. CERN, the European Organization for Nuclear Research
  19. EMNLP Workshop on Causality and Text, Keynote
  20. Political Methodology Seminar, Columbia University
  21. Joint Statistical Meetings, Invited Lecture on Latent Variables in Causality
  22. Flatiron Institute Algorithms Conference
- **Year 2020**
  23. Simons Foundation Conference on Mathematical and Physical Sciences
  24. IMS/AMS Conference of Foundations of Data Science, Invited Tutorial
  25. Online Causal Inference Seminar
  26. Melbourne Data Science Institute
  27. Joint Statistical Meetings, Invited Overview Lecture
  28. Conference on Learning Theory (COLT), Keynote

29. BayesComp, Plenary Lecture
30. University of Michigan Statistics Seminar
31. The Flatiron Institute, Machine Learning seminar
32. The Flatiron Institute, Flatiron seminar
33. University of Massachusetts, Distinguished Data Science Lecture
34. Institute for Advanced Study, Seminar

- **Year 2019**

35. Duke University, Statistics seminar
36. University of Amsterdam, Colloquium
37. Linkoping University, Colloquium
38. Johns Hopkins University, MINDS seminar
39. Columbia Workshop on Trustworthy AI workshop, Keynote
40. Mathematics of Data Science seminar, New York University
41. Columbia Program on Economic Research mini-course
42. The Flatiron Institute CSB Colloquium
43. Joint Statistical Meetings, JASA Discussion Paper Lecture
44. Microsoft Research, AI Distinguished Lecture
45. The Flatiron Institute Seminar
46. Oberwolfach Workshop on Foundations and New Horizons in Causal Inference
47. Harvard University, Joint Statistics/Computer Science Colloquium
48. Flatiron Workshop on ML+Physics, Invited Talk
49. University of Chicago, Statistics Colloquium
50. University of British Columbia, Constance van Eeden Lecture
51. Montreal Institute of Learning and Automation
52. Northeastern University, Distinguished Lecture
53. Annual Meeting of the AAAS, Symposium on Machine Learning and Statistics
54. Two Sigma, Research Seminar
55. Machine Learning Summer School, Stellenbosch, South Africa

- **Year 2018**

56. NeurIPS Workshop on Causal Inference
57. Columbia University Statistics Seminar
58. Yale University Seminar on Statistics and Data Science
59. NYU Tandon School of Engineering, Seminar on Modern Artificial Intelligence
60. Duke University, Statistics Seminar
61. International Conference on Probabilistic Programming, Keynote
62. Nature Conference on Big Data and Cancer, Keynote
63. Ideas42, Research seminar
64. Cornell Tech University, Distinguished Lecture
65. Goldman Sachs, Invited talk
66. Google New York, Research seminar
67. Columbia Center for Computational Social Sciences, Invited talk

68. International Meeting of the Psychometric Society, Keynote
69. Machine Learning Summer School, Buenos Aires, Argentina
70. Microsoft Research New York, MSR Research Seminar
71. Department of Biomedical Informatics Retreat, Invited talk
72. Princeton University, Machine Learning Colloquium
73. IBM Research, Machine Learning Seminar
74. Artificial Intelligence and Statistics, Keynote
75. D.E. Shaw, Research Seminar

- **Year 2017**

76. University of California Los Angeles, Statistics Seminar
77. European Association of Computational Linguistics, Keynote Lecture
78. University of Toronto, AI Institute Lecture
79. University of Michigan, Data Science Seminar
80. University of Michigan, Statistics Seminar
81. University of Edinburgh, Distinguished Lecture
82. Workshop for Young European Statisticians, Invited Tutorial
83. Conference on Big Data and Marketing, Invited Tutorial
84. Simons Institute Workshop on Computational Challenges in Machine Learning, Invited Talk
85. Santa Fe Institute, Invited Talk
86. New York University, Statistics Seminar

- **Year 2016**

87. University of California Berkeley, Neyman Seminar
88. University of California Berkeley, Focused Research Group
89. University of Tokyo, Machine Learning Seminar
90. Stonybrook University Computer Science, Distinguished Lecture
91. Keynote Speaker: Recruit Tokyo Data Science Conference
92. Columbia University, Sustainable Development Seminar
93. AT&T Research, Statistics Seminar
94. Columbia University, Biostatistics Seminar
95. Broad Institute (Cambridge, MA)
96. MIT Statistics Seminar (Cambridge, MA)
97. IPAM Workshop on the Mathematical Analysis of Cultural Expressive Forms (Los Angeles, CA)
98. AIG Data Science Seminar (New York, NY)
99. Isaac Newton Institute Workshop on the Limits of Graph Statistics (Cambridge, UK)
100. Cambridge University Machine Learning Seminar
101. Microsoft Research Conference on Machine Learning and Economics (New York, NY)
102. University of Illinois Distinguished Lecture in Computer Science
103. University of Chicago conference: "Machine Learning: What's In It For Economics?"
104. Sandia National Laboratories Distinguished Lecture (Albuquerque, NM)
105. Latent Variables 2016, Plenary Speaker (Colombia, SC)

106. Spotify (New York, NY)
107. Princeton Day of Statistics (Princeton, NJ)
108. Two Sigma (New York, NY)
109. Invited NIPS Tutorial on Variational Inference (Barcelona, Spain)
110. NIPS Workshop on Bayesian Deep Learning (Barcelona, Spain)
111. NIPS Workshop on Causal Inference (Barcelona, Spain)

- **Year 2015**

112. 2015 Joint Statistical Meetings, Session on Large-Scale Inference, Seattle
113. Center for Mathematical Research, Montreal Canada
114. Princeton Conference on Text Analysis and the Social Sciences
115. Facebook Artificial Intelligence Research, New York City
116. 10th Conference on Bayesian Nonparametric Statistics
117. Amazon Inc., Machine Learning @ Amazon
118. University of Chicago, Statistics Colloquium
119. University of Connecticut, Statistics Colloquium
120. Brown University, Computer Science Colloquium
121. Brown University, Applied Mathematics Colloquium
122. Rutgers University, Innovations in Statistics and Data Analysis
123. Office of Naval Research, Naval Future Force
124. Microsoft Research, Data Science Seminar
125. NIPS Workshop on Black Box Inference
126. Indiana University, Distinguished Lecture
127. New York University, Text as Data Series
128. Columbia University, IGERT Distinguished Speaker
129. University of Massachusetts, Data Science Distinguished Lecture
130. University of Massachusetts, Machine Learning Colloquium
131. Johns Hopkins University, Applied Mathematics Seminar

- **Year 2014**

132. Keynote Speaker, IBM Research Colloquium on Cognitive Computing (Haifa)
133. Keynote Speaker, IBM Research Machine Learning Seminar (Haifa)
134. Keynote Speaker, DIMACS Mixer
135. George Mason University, Distinguished Lecture
136. University of Washington, Distinguished Lecture
137. Simons Foundation, Frontiers of Data Science
138. Microsoft Research, Redmond
139. Keynote speaker, Uncertainty in Artificial Intelligence
140. Data, Society, and Inference Seminar at Stanford University
141. IPAM Workshop on “Stochastic Gradient Methods”

- **Year 2013**

142. Andresseen-Horowitz Academic Summit (Menlo Park, California)

143. Applied Communications Sciences (New Jersey)
144. Bloomberg LLC Distinguished Lecture (New York, New York)
145. City University of New York Computer Science Colloquium
146. Columbia University Data Sciences Institute
147. Duke University Machine Learning Seminar
148. Google Tech Talk (Mountain View, California)
149. INRIA Machine Learning Colloquium (Paris, France)
150. Microsoft Research New York
151. NIPS Workshop on "Probabilistic Modeling of Big Data" (Stateline, Nevada)
152. Stanford University Statistics Seminar
153. Temple University Computer Science Colloquium
154. Xerox Research 20th Anniversary Distinguished Lecture (Grenoble, France)

- **Year 2012**

155. Carnegie-Mellon University Machine Learning Department
156. Harvard University Computer Science Colloquium
157. Harvard University Machine Learning Seminar
158. International Conference on Machine Learning, Invited Tutorial (Edinburgh, Scotland)
159. Johns Hopkins University Computer Science Colloquium
160. Machine Learning Summer School, Invited Lecture Series (Kyoto, Japan)
161. Massachusetts Institute of Technology
162. Jamon Lecture (Stateline, Nevada)
163. New York University Machine Learning Seminar
164. Purdue University Computer Science Colloquium
165. University of California San Diego Computer Science Colloquium
166. University of Texas Austin Statistics Seminar

- **Year 2011**

167. Boston University Computer Engineering Colloquium
168. Conference on Political Methodology (Princeton, New Jersey)
169. Duke University Statistics Seminar
170. IBM Watson (Yorktown Heights, NY)
171. ISBA Workshop on Bayesian Nonparametrics (Veracruz, Mexico)
172. Joint Statistical Meetings (Miami, Florida)
173. Knowledge Discovery and Data Mining, Invited Tutorial (San Diego, California)
174. New York University Statistics Seminar
175. Stanford University Machine Learning Seminar
176. Stanford University Statistics Seminar
177. University of California Berkeley Neyman Seminar
178. University of Chicago
179. University of Pennsylvania Computer Science Colloquium
180. University of Tennessee Computer Science Colloquium
181. Yale University Applied Mathematics

182. Yale University Statistics

- **Year 2010**

183. Columbia University Computer Science Colloquium
184. Cornell University Computer Science Colloquium
185. Duke University Statistics Seminar
186. Educational Testing Service Seminar (Princeton, New Jersey)
187. Institute for Pure and Applied Mathematics (Los Angeles, California)
188. New York Academy of Sciences Machine Learning Symposium
189. New York Machine Learning Meetup
190. New York University Computer Science Colloquium

- **Year 2009**

191. Carnegie Mellon University Machine Learning Seminar
192. Carnegie Mellon University Statistics Seminar
193. Center for Discrete Mathematics and Theoretical Computer Science (Rutgers)
194. Columbia University Statistics Seminar
195. Machine Learning Summer School (Cambridge, England)
196. New Directions in Analyzing Text as Data (Cambridge, Massachusetts)
197. Rutgers University Statistics Seminar

- **Before 2009**

198. AAAI Spring Symposium (2002)
199. BAE Systems (2008)
200. Brown University (2005)
201. Carnegie Mellon University (2003)
202. Columbia University (2007)
203. Cornell University (2007)
204. Center for Discrete Mathematics and Theoretical Computer Science (2008)
205. Educational Testing Services (2006)
206. Duke University (2006)
207. Google Research (2004)
208. Google Research (2006)
209. Google Research (2007)
210. The Hebrew University (2008)
211. IBM Almaden (2002)
212. Institute for Pure and Applied Mathematics (2006)
213. Johns Hopkins University (2006)
214. Joint Statistical Meetings (2006
215. Massachusetts Institute of Technology (2003)
216. Massachusetts Institute of Technology (2007)
217. Massachusetts Institute of Technology (2008)
218. Microsoft Research Redmond (2007)

219. Nature Publishing Group (2008)
220. New York University Computer Science (2005)
221. Princeton University Computer Science (2005)
222. Rutgers University (2007)
223. University of California Irvine (2007)
224. University of California Los Angeles (2005)
225. University of California San Diego (2005)
226. University of California Santa Cruz (2005)
227. University of Illinois Champagne-Urbana (2006)
228. University College London (2008)
229. University of Cambridge (2008)
230. University of Connecticut (2006)
231. University of Connecticut (2007)
232. University of Massachusetts Amherst (2004)
233. University of Pennsylvania (2006)
234. University of Toronto (2003)
235. Xerox PARC (2002)

Popular Press

- "An Interview with Jessica Bruder and David Blei." *The Believer.* June 6, 2018.
- "Avalanches of Words, Sifted and Sorted." *The New York Times.* March 24, 2012.
- "Organising the Web: The Science of Science." *The Economist.* April 28, 2011.
- "Statistical Time Travel Helps to Answer What-Ifs." *Wall Street Journal*. November 12, 2009.