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: January 14, 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

- 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
Publications

Journal Articles


Conference Articles


140. 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.


BOOK CHAPTERS


EDITED VOLUMES


AWARDED GRANTS


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


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


Courses

- 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**  

- **Senior Program Committee**  
  - Causal Learning and Representations (2022)  
  - International Conference on Machine Learning, 2015 Program Co-chair  
  - International Conference on Machine Learning  
  - Bayesian Nonparametrics (2017)  

- **Associate Editor and Editorial Board**  
  - International Machine Learning Society Board (2013–present)  

- **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**
– Association of Artificial Intelligence (2007)
– International Joint Conference on Artificial Intelligence (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 chair (Statistics)
– Data Science committee (Statistics)
– PhD committee (Statistics)
– PhD committee (Computer Science)
– Distinguished lecture committee (Computer Science)
– Search 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

- Workshop organizing
  - “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)
  - “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
  - Luhuan Wu
  - Bohan 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); Postdoctoral Fellow, Harvard University
  - 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**

– Scientific Advisor, CNN (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)
– Steering Committee, Declassification Engine (2014–2016)
– Scientific Advisor, Applied Communications Sciences (2012–2013)
– Consulting Researcher, Microsoft Research (2013, 2014)

**Invited Talks**

**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
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
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
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)
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 Inference
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
  193. Center for Discrete Mathematics and Theoretical Computer Science (Rutgers)
  194. Columbia University Statistics Seminar
  196. New Directions in Analyzing Text as Data (Cambridge, Massachusetts)
  197. Rutgers University Statistics Seminar

- **Before 2009**
  198. AAAI Spring Symposium (2002)
211. IBM Almaden (2002)
216. Massachusetts Institute of Technology (2007)
217. Massachusetts Institute of Technology (2008)
218. Microsoft Research Redmond (2007)
222. Rutgers University (2007)
226. University of California Santa Cruz (2005)
228. University College London (2008)
235. Xerox PARC (2002)

**Popular Press**