Hanna Wallach

http://dirichlet.net
hanna@dirichlet.net

Microsoft Research
641 Avenue of the Americas, 7th Floor
New York, NY 10011

Hanna Wallach develops machine learning methods for analyzing the structure, content, and dynamics of social processes. Her work is inherently interdisciplinary: she collaborates with political scientists, sociologists, and journalists to learn how organizations work in practice by analyzing publicly available interaction data such as public record email networks, document dumps, press releases, meeting transcripts, and news articles. Hanna's research has had broad impact in machine learning, natural language processing, and the nascent field of computational social science. In 2010 her work on infinite belief networks won the best paper award at the AISTATS conference, in 2014 she was named one of Glamour magazine's "35 Women Under 35 Who Are Changing the Tech Industry," and in 2015 she was profiled in O'Reilly's book on "Women in Data." She is the recipient of several NSF grants, an IARPA grant, and a grant from the Office of Juvenile Justice and Delinquency Prevention. Hanna is committed to increasing diversity and has worked for over a decade to address the underrepresentation of women in computing. She founded two projects—the first of their kind—to increase women's involvement in free and open source software development: Debian Women and the GNOME Women's Summer Outreach Program. She also founded the annual Women in Machine Learning Workshop, which is now in its eleventh year.

Education

Doctor of Philosophy Cantab., 2002–2008.
Cavendish Laboratory, University of Cambridge.
Thesis: *Structured Topic Models for Language*.
Advisors: David MacKay (official) and Fernando Pereira (acting; University of Pennsylvania).

Master of Science, 2001–2002. Graduated with distinction.
School of Informatics, University of Edinburgh.
Thesis: *Efficient Training of Conditional Random Fields*.
Advisor: Miles Osborne.

Bachelor of Arts (Hons.) Cantab., 1998–2001. Graduated with 1st class honors.
Computer Laboratory, University of Cambridge.
Thesis: *Visual Representation of Computer-Aided Design Constraints*.
Advisor: Alan Blackwell.

Professional Experience

Senior Researcher, September 2015–present.
Researcher, January 2014–September 2015.
Microsoft Research New York City.

Adjunct Associate Professor, January 2016–present.
Assistant Professor, September 2010–January 2016. (On leave January 2014–January 2016.)
College of Information and Computer Sciences, University of Massachusetts Amherst.

Software Engineer, June–August 2010.

Advisory Board Member, June 2010–present.
Convertro, Inc.
Development of machine learning techniques for cookieless visitor session tracking.

Senior Postdoctoral Research Associate, November 2008–August 2010.
Postdoctoral Research Associate, July 2007–November 2008.
Department of Computer Science, University of Massachusetts Amherst.
Advisor: Andrew McCallum.
Research on machine learning for structured and unstructured data, with an emphasis on developing statistical topic models for science and innovation policy analysis and thematic community discovery. Day-to-day management of a fourteen-person research group in McCallum's sabbatical absence, including organization of weekly lab meetings and provision of on-site advising of MS and PhD students (July 2009–July 2010).

Research Assistant, June 2003–July 2007.
Department of Computer and Information Science, University of Pennsylvania.
Research on conditional random fields, latent variable models for splice site data, and Bayesian methods for text analysis, including language modeling and statistical topic modeling.

Research Assistant, August 2005–May 2006.
Department of Social Anthropology, University of Cambridge.
Part-time work on gender dimensions of free and open source software development. Organization of a one-day workshop on understanding and increasing women's involvement in free software.

Research Assistant, July–September 2001.
Computer Laboratory, University of Cambridge.
Research on human–computer interaction, focusing on programming by example systems, visual program representation, direct manipulation, and automatic inference of regular expressions.

Intern and Research Assistant, July–September 2000.
Altera Corporation and Computer Laboratory, University of Cambridge.
Implementation of encryption algorithms on field-programmable gate arrays.

Working Papers

J. ben-Aaron, M. Denny, B. Desmarais, and **H. Wallach**. Transparency by conformity: A field experiment evaluating openness in local governments. 32 pages.

A. Boydstun, **H. Wallach**, and D. Young. Who's laughing now? Applying text analysis to humor in Federal Open Market Committee meetings. 36 pages. Presented at the Sixth Annual Text as Data Conference, 2015; the Annual Meeting of the American Political Science Association, 2015; the Midwest Political Science Assocation Conference, 2015; the Annual Meeting of the American Political Science Association PoliInformatics Research Challenge Short Course, 2014.

A. Chaney, **H. Wallach**, and D. Blei. Who, what, when, where, and why? a computational approach to understanding historical events using State Department cables. 8 pages. Presented at the Sixth Annual Text as Data Conference, 2015; the Ninth Annual Machine Learning Symposium, 2015.

M. Denny, J. ben-Aaron, B. Desmarais, and **H. Wallach**. Reading between the emails: Gendered patterns of communication in local government. 24 pages. Presented at the Sixth Annual Text as Data Conference, 2015; the Midwest Political Science Assocation Conference, 2015.

J. N. Matias and **H. Wallach**. Modeling gender discrimination by audiences of online news. 5 pages. Presented at the Computation + Journalism Conference, 2015.

L. Hannah and **H. Wallach**. Summarizing topics: From word lists to phrases. 9 pages. Presented at the Joint Statistical Meetings, 2015; Innovation and Application at Columbia University, 2015.

M. Denny, J. ben-Aaron, **H. Wallach**, and B. Desmarais. Topic-conditioned hierarchical latent space models for text-valued networks. 46 pages. Presented at the International Conference on Computational Social Science, 2015; the Southern Political Science Association Meeting, 2015; KDD at Bloomberg: Data Frameworks Track, 2014; the Society for Political Methodology Thirty-First Annual Summer Meeting, 2014; the Seventh Annual Political Networks Conference, 2014; the Fourteenth Annual Northeast Political Methodology Meeting, 2014; the Midwest Political Science Association Conference, 2014.

M. Denny, B. O'Connor, and **H. Wallach**. A little bit of NLP goes a long way: Finding meaning in legislative texts with phrase extraction. 28 pages. Presented at the Midwest Political Science Assocation Conference, 2015.

J. Grimmer, R. Shorey, **H. Wallach**, and F. Zlotnik. A class of Bayesian semiparametric topic models for political texts. 43 pages. Presented at the Midwest Political Science Association Conference, 2012; the Society for Political Methodology Twenty-Eighth Annual Summer Meeting, 2011.

R. Shorey, **H. Wallach**, and B. Desmarais. Toward a framework for the large-scale textual and contextual analysis of government information declassification patterns. 12 pages. Presented at the Workshop on Computational and Online Social Science, 2012; the DataGotham Conference, 2012; the Second Annual Text as Data Conference, 2011.

Book Chapters

H. Wallach. Computational social science: Toward a collaborative future. In R. Alvarez, editor, *Computational Social Science: Discovery and Prediction*. Cambridge University Press, 2016 (forthcoming).

Peer-Reviewed Journal and Conference Publications

G. Bissias, B. Levine, M. Liberatore, B. Lynn, J. Moore, **H. Wallach**, and J. Wolak. Characterization of contact offenders and child exploitation material trafficking on five peer-to-peer networks. *Child Abuse and Neglect*, 2015.

A. Schein, J. Paisley, D. Blei, and H. **Wallach**. Bayesian Poisson tensor factorization for inferring multilateral relations from sparse dyadic event counts. In *Proceedings of the Twenty-First ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2015. Also presented at the Tenth Conference on Bayesian Nonparametrics, 2015.

F. Guo, C. Blundell, **H. Wallach**, and K. Heller. The Bayesian echo chamber: Modeling social influence via linguistic accommodation. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 2015. Also presented at the Tenth Conference on Bayesian Nonparametrics, 2015; the Fifth Annual Text as Data Conference, 2014.

W. Mason, J. Wortman Vaughan, and **H. Wallach**. Computational social science and social computing. *Machine Learning*, 95(3):257–260, 2014.

S. Counts, M. De Choudhury, J. Diesner, E. Gilbert, M. Gonzalez, B. Keegan, M. Naaman, and **H. Wallach**. Computational social science: CSCW in the social media era. In *Proceedings of the Companion Publication of the Seventeenth ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 105–108, 2014. (Selected for panel discussion.)

K. Krstovski, D. Smith, **H. Wallach**, and A. McGregor. Efficient nearest neighbor search in the probability simplex. In *Proceedings of the Fourth International Conference on the Theory of Information Retrieval*, 2013.

P. Krafft, J. Moore, B. Desmarais, and **H. Wallach**. Topic-partitioned multinetwork embeddings. In *Advances in*

Neural Information Processing Systems Twenty-Five, 2012. Also presented at the Workshop on Information in Networks, 2012; the Third New Directions in Analyzing Text as Data Conference, 2012; the Fifth Annual Political Networks Conference, 2012.

A. Bakalov, A. McCallum, **H. Wallach**, and D. Mimno. Topic models for taxonomies. In *Proceedings of the Twelfth ACM/IEEE-CS Joint Conference on Digital Libraries*, 2012. (Selected for talk.)

D. Mimno, **H. Wallach**, M. Leenders, E. Talley, and A. McCallum. Optimizing semantic coherence in topic models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2011. (Selected for talk.)

E. Talley, D. Newman, D. Mimno, B. Herr II, **H. Wallach**, G. Burns, M. Leenders, and A. McCallum. Database of NIH grants using machine-learned categories and graphical clustering. *Nature Methods*, 8(6):443–444, 2011.

R. Adams, **H. Wallach**, and Z. Ghahramani. Learning the structure of deep, sparse graphical models. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 2010. (Won best paper award.)

H. Wallach, S. Jensen, L. Dicker, and K. Heller. An alternative prior process for nonparametric Bayesian clustering. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 2010.

H. Wallach, D. Mimno, and A. McCallum. Rethinking LDA: Why priors matter. In *Advances in Neural Information Processing Systems Twenty-Two*, 2009. (Selected for spotlight.)

D. Mimno, **H. Wallach**, J. Naradowsky, D. Smith, and A. McCallum. Polylingual topic models. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 880–889, 2009. (Selected for talk.)

H. Wallach, I. Murray, R. Salakhutdinov, and D. Mimno. Evaluation methods for topic models. In *Proceedings of the Twenty-Sixth International Conference on Machine Learning*, 2009. (Selected for talk.)

M. Dredze, **H. Wallach**, D. Puller, T. Brooks, J. Caroll, J. Magarick, J. Blitzer, and F. Pereira. Intelligent email: Aiding users with artificial intelligence. In *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, NECTAR Track*, pages 1524–1527, 2008. (Selected for talk.)

M. Dredze, **H. Wallach**, D. Puller, and F. Pereira. Generating summary keywords for emails using topics. In *Proceedings of the 2008 International Conference on Intelligent User Interfaces*, pages 199–206, 2008. Also presented at the Women in Machine Learning Workshop, 2007.

H. Wallach. Topic modeling: Beyond bag-of-words. In *Proceedings of the Twenty-Third International Conference on Machine Learning*, 2006. (Selected for talk.)

A. Blackwell and **H. Wallach**. Diagrammatic integration of abstract operations into software work contexts. In *Proceedings of Diagrammatic Representation and Inference: Second International Conference*, 2002. (Selected for talk.)

Peer-Reviewed Workshop Publications

A. Schein, M. Zhou, D. Blei, and **H. Wallach**. Modeling topic-partitioned homophily and stochastic equivalence in dyadic events data. In *Proceedings of the Neural Information Processing Systems Workshop on "Networks in the Social and Information Sciences"*, 2015 (forthcoming).

J. Miller, B. Bettancourt, A. Zaidi, **H. Wallach**, and R. Steorts. Microclustering: When cluster sizes grow sublinearly with the size of the data set. In *Proceedings of the Neural Information Processing Systems Workshop on "Bayesian Nonparametrics: The Next Generation"*, 2015 (forthcoming). Also presented at G70: A Celebration of Alan Gelfand's Seventieth Birthday, 2015.

L. Hannah and **H. Wallach**. Topic summarization: From word lists to phrases. In *Proceedings of the Neural Information Processing Systems Workshop on "Modern Machine Learning and Natural Language Processing"*, 2014.

A. Schein, J. Paisley, D. Blei, and **H. Wallach**. Inferring polyadic events with Poisson tensor factorization. In *Proceedings of the Neural Information Processing Systems Workshop on "Networks: From Graphs to Rich Data"*, 2014.

F. Guo, C. Blundell, **H. Wallach**, and K. Heller. The Bayesian echo chamber: Modeling influence in conversations. In *Proceedings of the Neural Information Processing Systems Workshop on "Networks: From Graphs to Rich Data"*, 2014.

A. Schein, J. Moore, and **H. Wallach**. Inferring multilateral relations from dynamic pairwise interactions. In *Proceedings of the Neural Information Processing Systems Workshop on "Frontiers of Network Analysis: Methods, Models, and Applications"*, 2013. Also presented the Computational Social Science Society Conference, 2013.

A. Passos, **H. Wallach**, and A. McCallum. Correlations and anticorrelations in LDA inference. In *Proceedings of the Neural Information Processing Systems Workshop on "Challenges in Learning Hierarchical Models: Transfer Learning and Optimization"*, 2011.

D. Mimno, **H. Wallach**, J. Naradowsky, D. Smith, and A. McCallum. Polylingual topic models. In *Proceedings of the Learning Workshop*, 2009.

H. Wallach, I. Murray, R. Salakhutdinov, and D. Mimno. Evaluation methods for topic models. In *Proceedings of the Learning Workshop*, 2009. (Selected for talk.)

D. Mimno, **H. Wallach**, and A. McCallum. Gibbs sampling for logistic normal topic models with graph-based priors. In *Proceedings of the Neural Information Processing Systems Workshop on "Analyzing Graphs"*, 2008. (Selected for talk.)

H. Wallach, C. Sutton, and A. McCallum. Bayesian modeling of dependency trees using hierarchical Pitman-Yor priors. In *Proceedings of the ICML/UAI/COLT Workshop on "Prior Knowledge for Text and Language"*, 2008. (Selected for talk.)

M. Dredze and **H. Wallach**. User models for email activity management. In *Proceedings of the Fifth International Workshop on Ubiquitous User Modeling*, 2008. (Selected for talk.)

D. Mimno, **H. Wallach**, and A. McCallum. Community-based link prediction with text. In *Proceedings of the Neural Information Processing Systems Workshop on "Statistical Models of Networks"*, 2007.

H. Wallach. Topic modeling: Beyond bag-of-words. In *Proceedings of the First Annual North East Student Colloquium on Artificial Intelligence*, 2006. (Selected for talk.)

H. Wallach. Topic modeling: Beyond bag-of-words. In *Proceedings of the Neural Information Processing Systems Workshop on "Bayesian Methods for Natural Language Processing"*, 2005. (Selected for talk.)

H. Wallach, M. Allan, and D. Harries. The Debian new maintainer process: History and aims. In *Proceedings of the Sixth Annual International Debian Developers' Conference*, 2005. (Selected for talk.)

H. Wallach. Efficient training of conditional random fields. In *Proceedings of the Sixth Annual Computational Linguistics Research Colloquium*, 2003. (Selected for talk.)

Other Publications and Technical Reports

H. Wallach. Big data, machine learning, and the social sciences: Fairness, accountability, and transparency. `https://medium.com/@hannawallach/big-data-machine-learning-and-the-social-sciences-927a8e20460d`, 2014. Over 20,000 views to date.

H. Wallach. The benefits of double-blind review. `http://people.cs.umass.edu/~wallach/publications/wallach13benefits.pdf` and

`http://hunch.net/?p=2656`, 2013.

Logistic Aggression. Seven things I wish I'd known about derby. *Derby Life*, 2013.
`http://www.derbylife.com/articles/2013/03/seven_things_i_wish_id_known_about_derby`.

M. Dredze and **H. Wallach**. How to be a successful PhD student.
`http://www.cs.umass.edu/~wallach/how_to_be_a_successful_phd_student.pdf`, 2012.

H. Wallach. Evaluation metrics for hard classifiers.
`http://www.inference.phy.cam.ac.uk/hmw26/papers/evaluation.ps`, 2004.

H. Wallach. Conditional random fields: An introduction. Technical Report MS-CIS-04-21, University of Pennsylvania, 2004.

Selected Press and Media Appearances

Guest host for episode one of season two, Talking Machines (podcast), 2016.

Interviewed for "Dataclysm," Science for the People (radio show and podcast). 2015.

Profiled in "Changing Lives," Newnham College (University of Cambridge) Newsletter. 2015.

Profiled in "Women in Data Science" by Cornelia Lévy-Bencheton and Shannon Cutt, O'Reilly. 2015.

Interviewed for "Using Models in the Wild and Women in Machine Learning," Talking Machines (podcast). 2015.

Interviewed for "Season Preview," Talking Machines (podcast). 2015.

Research discussed in Inside Microsoft Research, TechNet Blogs. 2014.

Profiled in "35 Women Under 35 Who Are Changing the Tech Industry" by Donna Fenn, Glamour Magazine. 2014.

Research discussed in "Pencils and Pixels," Columbia Journalism Review, 2014.

Interviewed (individually plus four-person panel discussion) in "The History and Future of the Neural Information Processing Systems (NIPS) Conference," The Science Network. 2011.

Editorial. Mapping the money. *Nature Methods*, 7(6):437, 2011.

Featured in "Anatomy of a Geek Desktop," Linux Format.

Current and Recent Grants

PI with Bruce Desmarais (UMass Amherst; co-PI), "Organizational Responsiveness to Open Outside Input: A Modeling Approach based on Statistical Text and Network Analysis," National Science Foundation Directorate for Computer and Information Science and Engineering, Information and Intelligent Systems. 09/01/2013–08/31/2016. $479,628.

PI with David Jensen (UMass Amherst; co-PI) and Andrew McCallum (UMass Amherst; co-PI), "Foresight and Understanding from Scientific Exposition," Raytheon BBN Technologies (prime to Intelligence Advanced Research Projects Agency, Office of Incisive Analysis). 09/01/2011–08/31/2016, $5,940,330.

Co-PI with Marc Liberatore (UMass Amherst; PI), Brian Levine (UMass Amherst; co-PI), Thomas Kerle (Fox Valley

Technical College; co-PI), Janice Wolak (University of New Hampshire; co-PI), "RoundUp Predictive Tool (RPT) Project," Office of Juvenile Justice and Delinquency Prevention, Fiscal Year 2011 Child Protection Research Program. 10/01/2011–09/30/2014, $618,426.

PI with Andrew McCallum (UMass Amherst; co-PI) and Fiona Murray (MIT Sloan School of Management; co-PI), "Understanding the Diversity of Science," National Science Foundation Directorate of Social, Behavioral and Economic Sciences, Science of Science and Innovation Policy. 05/15/2010–04/30/2016, $286,939.

Co-PI with Jennifer Wortman Vaughan (University of California Los Angeles; PI), "Workshop for Women in Machine Learning," National Science Foundation Directorate for Computer and Information Science and Engineering, Information and Intelligent Systems. 09/01/2010–08/31/2012.

PI with Andrew McCallum (UMass Amherst; co-PI), "Topic Modeling Analysis of NIH Grant Proposals," National Institute of Neurological Disorders and Stroke/National Institutes of Health. 09/28/2010–09/27/2011, $160,000.

Teaching Experience

Co-instructor, Data Science Summer School, Microsoft Research, New York City.
Eight-week-long data science summer school intended to increase diversity in computer science, Summer 2014.

Co-instructor, Department of Computer Science, University of Massachusetts Amherst.
Graduate seminar (non-credit) on machine learning for data science, Fall 2012 and Spring 2013.

Instructor, Department of Computer Science, University of Massachusetts Amherst.
CMPSCI 691BM, "Bayesian Methods for Text," Fall 2012.

Instructor, Department of Computer Science, University of Massachusetts Amherst.
CMPSCI 240, "Reasoning Under Uncertainty," Spring 2012.

Guest Lecturer (1 lecture), Department of Computer Science, University of Massachusetts Amherst.
CMPSCI 191A, lecture title: "Machine Learning, Predictive Text, and Topic Models," Fall 2011.

Instructor, Department of Computer Science, University of Massachusetts, Amherst.
Unofficial graduate course on topic modeling and Bayesian text analysis (thirteen students), Fall 2011.

Instructor, Department of Computer Science, University of Massachusetts Amherst.
CMPSCI 791SS, "Computational Social Science," Spring 2011.

Guest Lecturer (2 lectures), Department of Computer Science, University of Massachusetts Amherst.
CMPSCI 691GM, "Graphical Models," Spring 2011.

Instructor, Department of Computer Science, University of Massachusetts Amherst.
CMPSCI 240, "Reasoning Under Uncertainty," Fall 2010.

Guest Lecturer (1 lecture), Department of Computer Science, University of Massachusetts Amherst.
CMPSCI 689, "Machine Learning," Fall 2009.

Instructor, Department of Computer Science, University of Massachusetts, Amherst.
Unofficial course on topic modeling (six students; five graduate, one undergraduate), Fall 2009.

Supervisor (Teaching Assistant), Computer Laboratory, University of Cambridge.
Five undergraduate courses: "Information Theory and Coding," "Probability," "Human–Computer Interaction," "Computer Systems Modeling," and "Continuous Mathematics," 2002–2003.

Doctoral Committees

Abigail Jacobs, University of Colorado at Boulder. Advisor: Aaron Clauset. TBD.

Michael Paul, Johns Hopkins University. Advisor: Mark Dredze. 2015.

Viet-An Nguyen, University of Maryland. Advisors: Jordan Boyd-Graber and Philip Resnik. 2015.

Marc Maier, University of Massachusetts Amherst. Advisor: David Jensen. 2014.

Sean Gerrish, Princeton University. Advisor: David Blei. 2013.

Graduate Advisees and Research Supervised

Allison Chaney (Princeton PhD student in Computer Science; advisor: David Blei). Statistical topic Modeling and Network Analysis for Understanding Historical Events Using US State Department Cables. 2015–present.

Richard Guo (Duke University PhD student in Computer Science; advisor: Katherine Heller). Inferring Influence Networks using Multivariate Hawkes Processes and Dynamic Language Models. 2014–2015.

Matthew Denny (Penn State PhD student in Political Science; co-advisor Bruce Desmarais). Statistical Text and Network Analysis of Local Government Communication Networks. 2014–present.

Rick Freedman (UMass Amherst PhD student; advisor: Schlomo Zilberstein). Synthesis Project. 2013–2014.

Aaron Schein (UMass Amherst MS/PhD student). Bayesian Tensor Factorization. 2012–present.

Jingyi Guo (UMass Amherst MS student; co-advisor: Brian Levine). Statistical Analyses of Peer-to-Peer File Sharing Networks for Child Protection. 2012–2013.

Juston Moore (UMass Amherst MS/PhD student). Detecting Anomalies in Collaborative Networks. 2011–2014.

Ravali Pochampally (UMass Amherst PhD student). Modeling Emergence in Scientific Exposition. 2011–2012.

Melissa Frechette (UMass Amherst MS student). Generation and Curation of Gene Patent Data. 2011–2012.

Peter Krafft (UMass Amherst MS student). Modeling Government Email Networks. 2011–2012.

Aaron Schein (UMass Amherst MA student in Linguistics; co-advisor: Brian Dillon, UMass Amherst Linguistics). Machine Learning Analysis of Medical School Recommendation Letters. 2011–2012.

Laura Sevilla (UMass Amherst PhD student; advisor: Erik Learned-Miller). Synthesis Project. 2011–2012.

Kriste Krstovski (UMass Amherst PhD student; advisor: David Smith). Synthesis Project. 2011–2012.

Meagan Day (UMass Amherst MS student). Enhancing the Usability of Statistical Topic Models. 2010–2012.

Rachel Shorey (UMass Amherst MS student). Probabilistic Topic Modeling for Political Text Data. 2010–2011.

Anton Bakalov (UMass Amherst MS/PhD student; co-advisor: Andrew McCallum). Probabilistic Models for Analyzing Emergence and Evolution of Topics over Time. 2009–2012.

Undergraduate Advisees and Research Supervised

Nathanial May (UMass Amherst). Probabilistic Modeling of Text and Networks. Fall 2011.

Rebecca Knowles (Haverford College; co-advisor: Mark Dredze, Johns Hopkins University). Statistical Topic Models for Modeling Linguistic Register. Summer 2011.

Xinlei Chen (Zhejiang University; co-advisor: Jennifer Wortman Vaughan, University of California Los Angeles). Modeling Questions and Answers with Polylingual Topic Models. Summer 2011.

Nicolas Ioannou (UMass Amherst). Polylingual Modeling of Open Source Communities. Spring 2011.

Peter Kritikos (UMass Amherst). Vertex Nomination. Spring 2011.

Jessica Ray (UMass Amherst). Cookieless Visitor Session Tracking. Spring 2011.

Meagan Day (UMass Amherst; co-advisor: Andrew McCallum). Application of Natural Language Processing Techniques to the Organization of Biomedical Research Literature. 2009–2010.

Neal Parikh (University of Pennsylvania; co-advisors: Mark Dredze and Sarah Kaplan, University of Pennsylvania). Statistical Topic Modeling for Carbon Nanotechnology Patent Data. Summer 2008.

Danny Puller (University of Pennsylvania; co-advisors: Mark Dredze and Fernando Pereira, University of Pennsylvania). Email Keyword Summarization Using Topic Models. Summer 2007.

Invited Talks and Presentations

Topic-Partitioned Network Structure. NIPS Workshop on "Networks in the Social and Information Sciences." 2015.

The Past, Present, and Future of Women in Machine Learning (opening address). WiML Workshop. 2015.

The Bayesian Echo Chamber. NYU. 2015.

Learning in the Sunshine: Analysis of Local Government Email. DataPoint conference. 2015.

Learning in the Sunshine: Analysis of Local Government Email. Data Science for Social Good. 2015.

Panel discussion. ICML Workshop on "Fairness, Accountability, and Transparency in Machine Learning." 2015.

Computational Social Science. Invited tutorial (one of six) at ICML. 2015.

Lessons from Computational Social Science (keynote presentation). NICAR. 2015.

The Ethics of Machine Learning. ACM Queue Editorial Board Meeting. 2015.

Big Data, Machine Learning, and the Social Sciences. Data Driven NYC. 2015.

Big Data, Machine Learning, and the Social Sciences. NIPS Workshop on "Fairness, Accountability, and Transparency in Machine Learning." 2014.

The Bayesian Echo Chamber. Stanford University. 2014.

The Case for Hierarchical Bayesian Latent Variable Models for Text. Stanford University. 2014.

Panel discussion. Computation + Journalism Conference. 2014.

Statistical Topic Models. Strata NYC. 2014.

The Bayesian Echo Chamber. Text as Data Conference. 2014.

Textual Analysis of Government Declassification Patterns. Columbia. 2014.

Learning in the Sunshine: Analysis of Local Government Email Corpora. Duke University. 2014.

Perspectives in Machine Learning (Q&A session with students). Duke University. 2014.

Learning in the Sunshine: Analysis of Local Government Email Corpora. KDD at Bloomberg. 2014.

Discussant for "Mirrors for Princes and Sultans: Advice on the Art of Governance in the Medieval Christian and Islamic Worlds". Society for Political Methodology Annual Summer Meeting. 2014.

Panel discussion. Social Dynamics and Personal Attributes in Social Media Workshop at ACL. 2014.

Social Data Science (round table discussion). MSR Social Research Workshop. 2014

Social Science Makes Big Data Better (panel discussion). TechFest (Microsoft Research). 2014.

Investigative Journalism + Topic Modeling. Bit-by-Bit at the Brown Institute for Media Innovation. 2014.

Computational Social Science: CSCW in the Social Media Era (panel discussion). CSCW. 2014.

A History of the Women in Machine Learning Workshop (opening address). WiML Workshop. 2013.

Challenges and Successes of Collaboration (panel discussion). Atlanta Workshop on CSS. 2013.

Machine Learning for Complex Social Processes (keynote address). Atlanta Workshop on CSS. 2013.

How to be a Successful Graduate Student and Transition to a Great Job (five-person panel discussion). Mid-Atlantic Student Colloquium on Speech, Language and Learning. 2013.

Four Challenges for Computational Social Science. MSR New England 5^{th} Anniversary Symposium. 2013.

Discussant for "Measuring Ideological Proportions in Political Speeches". New Directions in Text as Data. 2013.

Statistical Models for Complex Social Processes. Facebook. 2013.

Statistical Models for Complex Social Processes. Microsoft Reseach New York City. 2013.

Statistical Models for Complex Social Processes. Google New York. 2013.

Machine Learning for Complex Social Processes. Etsy. 2013.

Machine Learning for Complex Social Processes. Microsoft Research New England. 2013.

The Benefits of Double-Blind Review. ICML Workshop on Peer Reviewing and Publishing Models. 2013.

Textual Analysis of Government Declassification Patterns. Declassification Engine Conference. 2013.

Machine Learning for Complex Social Processes. New England Machine Learning Day. 2013.

Machine Learning for Complex Social Processes. Sunlight Labs. 2012.

Statistical Topic Models. National Security Agency. 2012.

Machine Learning for Complex Social Processes. Johns Hopkins University. 2012.

Vertex Nomination. Human Language Technology Center of Excellence, Johns Hopkins University. 2012.

Statistical Topic Models for Science and Innovation Policy. Joint Statistical Meetings. 2011.

Statistical Topic Models for Computational Social Science. Johns Hopkins University. 2011.

Women in Free/Open Source Software Development. Johns Hopkins University. 2011.

Statistical Topic Models for Computational Social Science. University of Chicago. 2011.

Statistical Topic Models for Computational Social Science. Mount Holyoke College. 2011.

Statistical Topic Models for Science and Innovation Policy. Williams College. 2010.

Statistical Topic Models for Science and Innovation Policy. UMass Lowell. 2010.

Text Analysis for Science and Innovation Policy. New Directions in Text Analysis. 2010.

Women in Free/Open Source Software Development. Politics of Open Source. 2010.

Recruiting/Retaining Women in Free Software Projects (panel discussion). LibrePlanet. 2010.

Topic Models: Priors, Stop Words and Languages. Brown University. 2010.

Topic Models: Priors, Stop Words and Languages. University of Edinburgh. 2010.

Topic Modeling. NIPS Workshop on "Applications for Topic Models: Text and Beyond." 2009.

Bayesian Models for Dependency Parsing Using Pitman-Yor Priors. NIPS Workshop on "Unsupervised Latent Variable Models for Speech and Language." 2008.

Machine Learning, Predictive Text, and Topic Models. University of Baltimore. 2007.

Topic Modeling: Beyond Bag-of-Words. UMass Amherst. 2006.

Topic Modeling: Beyond Bag-of-Words. University College London. 2006.

Women in Free and Open Source Software Development. University of Pennsylvania. 2005.

The Debian Women Project. Libre Software Meeting. 2005.

The Debian Women Project. University of Cambridge. 2005.

Women in Free Software. Free and Open Source Developers' European Meeting. 2005.

Other Talks and Presentations

Statistical Topic Models. Human Language Technology Center of Excellence, Johns Hopkins University. 2012.

Statistical Machine Learning Analysis of Debian Mailing Lists. DebConf. 2010.

Computational Papyrology. Media in Transition International Conference. 2009.

Dasher: Information-Efficient Text Entry. Grace Hopper Conference. 2006.

Women in Free Software Development: Findings from FLOSSPOLS. FOSDEM. 2006.

The Debian Women Project. FOSDEM. 2005.

Honors and Awards

35 Women Under 35 Who Are Changing the Tech Industry, Glamour Magazine, 2014.

Flex Grant for Teaching/Faculty Development, UMass Amherst, 2010.

Best Paper Award for "Learning the Structure of Deep, Sparse Graphical Models," AISTATS, 2010.

Studentship, Engineering and Physical Sciences Research Council, UK, 2002.

Entrance Research Studentship, Newnham College, University of Cambridge, 2002.

Best MSc Student in Cognitive Science, University of Edinburgh, 2002.

Studentship, Economic and Social Sciences Research Council, UK, 2001.

MISYS Award for the Best Computer or Computer Software Student, National Science, Engineering and Technology Student of the Year Awards, UK, 2001.

Outstanding Part II Dissertation, Computer Laboratory, University of Cambridge, 2001.

Laurie Hart Memorial Prize, Newnham College, University of Cambridge, 2001.

Jemima Clough Prize, Newnham College, University of Cambridge, 2001.

Helen Gladstone Prize, Newnham College, University of Cambridge, 2000.

Letitia Chitty Award for Engineering, Newnham College, University of Cambridge, 1999.

Professional Service and Outreach

(For talks and presentations, please see "Invited Talks and Presentations" above.)

Tutorials Co-chair, NIPS, 2016.

Senior Program Committee Member, ICML, 2016.

Board Member (elected), International Machine Learning Society, 2015–present.

Organizer, NIPS Workshop on "Bayesian Nonparametrics: The Next Generation," 2015.
One-day workshop, consisting of talks, posters, and panel discussion, for researchers in Bayesian nonparametrics.

Founding Board Member, Text as Data Association, 2015–present.

Co-founder and Organizer, Women in Machine Learning NYC monthly lunch, 2015.

Organizer, ICML Workshop on "Fairness, Accountability, and Transparency in Machine Learning," 2015.
One-day workshop, consisting of talks, panel discussion, and dinner, for researchers addressing interdisciplinary questions regarding fairness, accountability, and transparency in machine learning.

Guest Editor, ACM Queue, Special Issue on the Ethics of Machine Learning and Experimentation, 2015–present.

Invited Participant, conference on "Algorithmic Transparency in the Media," hosted by the Tow Center, 2015.

Proposal Reviewer, Magic Grants, Brown Institute for Media Innovation, 2015.

Steering Committee Member, Computation + Journalism Conference, 2015–present.

Senior Program Committee Member, ICML, 2015.

Guest Editor, Machine Learning Journal, Special Issue on Computational Social Science, Volume 95, Issue 3, 2014.

Senior Program Committee Member, ICWSM, 2014.

Organizer, NIPS Workshop on "Topic Models: Computation, Application, and Evaluation," 2013.
One-day workshop, consisting of talks, posters, panel discussion and dinner, for researchers working on topic modeling. Over 100 participants. 26 accepted papers.

Principal Investigator, "User Feedback Regarding Open Peer Reviewing at `openreview.net`," 2013.
Collaborative human subjects research on ICLR 2013's open peer reviewing model. Design and deployment of a user survey to gather data on author and reviewer experiences, concerns about bias, etc.

Executive Board Member, Women in Machine Learning Workshop, 2012–present.

Organizer, NIPS Workshop on "Computational Social Science and the Wisdom of Crowds," 2011.
One-day workshop, consisting of talks, posters, panel discussion and dinner, for researchers undertaking interdisciplinary work on computational social science. 35 paper submissions.

Grant Proposal Review Panelist, NSF, February and December 2011.

Invited Participant, conference on "The Future of Patent Data," hosted by the USPTO, 2011.

Senior Program Committee Member, NIPS, 2011.

Mentor, Women in Machine Learning mentoring program, 2010, 2012, 2014, 2015.

Organizer, NIPS Workshop on "Computational Social Science and the Wisdom of Crowds," 2010.
One-day workshop, consisting of talks, posters, panel discussion and dinner, for researchers undertaking interdisciplinary work on computational social science. 45 paper submissions.

Invited Participant, conference on "FOIA 2.0," hosted by the USDA, 2010.

Session Chair, EMNLP, 2010.

Best Paper Award Committee Member, EMNLP, 2010.

Advisory Board Member, GNOME Outreach Program for Women, 2010–2011.

Executive Board Chair (elected), Women in Machine Learning Workshop, 2009–2012.

Organizer, NIPS Workshop on "Applications for Topic Models: Text and Beyond," 2009.
One-day workshop, consisting of talks, posters, panel discussion and dinner, for researchers working on topic modeling. Around 100 participants. Dinner sponsor: PASCAL.

Project Leader, collaborative project with Nature Publishing Group, 2009–2010.
Statistical topic model analysis of research literature provided by Nature Publishing Group.

Project Leader, collaborative project with Harvard Medical School Library, 2009–2011.
Comprehensive study of new trends in multidisciplinary biomedical research at Harvard University.

Founder and Organizer, topic modeling dinner at NIPS, 2007–2009.

Co-founder and Organizer, Women's Summer Outreach Program, GNOME Project, 2006.
Two-month-long outreach program, in which six female students undertook mentored GNOME-related software development projects. Sponsors: GNOME Foundation, Google Inc.

Co-founder and Organizer, Workshop for Women in Machine Learning, 2006.
One day workshop, consisting of talks, spotlights and posters, for just under 100 female faculty, research scientists, postdoctoral researchers, and students in machine learning. Sponsors: CRA-W, Google Inc., ITA Software, Microsoft Research, NSF (#IIS-0647431), PASCAL, and University of Pennsylvania.

Co-founder and Leader, Debian Women project, 2004–2008.
Project to encourage the participation of women in Debian GNU/Linux development, offering tutorials, mentoring, a mailing list, and an Internet Relay Chat channel.

Package Maintainer, Debian GNU/Linux project, 2004–2008.

Session Chair, NESCAI, 2006.

Organizer, Machine Learning Lunch, University of Pennsylvania, 2004–2005.

President, University of Cambridge Computing Society, 2003–2004.

Reviewer or Program Committee Member, ACL (2010 and 2012), AISTATS (2011–2012), COLING (2010), ICML (2009–2011), IJCAI (2011), NESCAI (2006), NIPS (2009), ACL Workshop on "Language Technology and Computational Social Science" (2014), Computation + Journalism Conference (2014–2015), NIPS Workshop on Bayesian Nonparametrics: The Next Generation" (2015), Data For Good Exchange (2015).

Reviewer, Annals of Applied Statistics (2010), IEEE Signal Processing (2010), Journal of Computing Science and Engineering (2009), Journal of Machine Learning Research (2006–present), Topics in Cognitive Science (2009).

Departmental and University Service and Outreach

Co-organizer, Microsoft Research New York City diversity reading group, 2015–present.

Organizer, Microsoft Research New England and New York City Annual Retreat, 2014.

Guest Speaker on "How to Be a Successful Graduate Student," Professionalism Seminar, UMass Amherst, 2011.

Member (elected), Executive Committee, UMass Amherst, 2011–2012.

Member (elected), Annual Faculty Review Committee, UMass Amherst, 2011–2012.

Guest Speaker, Women in Engineering Career Day, Women in Engineering Program, UMass Amherst, 2011.

Co-organizer, UMass Amherst Computational Social Science Initiative Seminar Series, 2010–present.
New, cross-departmental weekly seminar series on computational social science at UMass Amherst. Up to 65 attendees each week from departments throughout the university. The series has brought in speakers from the New York Times, Princeton, UCLA, Columbia, University of Michigan, UC Irvine, Yahoo!, Harvard, Northeastern, Stanford, and UC San Diego. Attained sponsorship from Yahoo! for the seminar series.

Member, Undergraduate Recruiting and Diversity Committee, UMass Amherst, 2010–2011.

Member, Faculty Recruiting Committee (Machine Learning), UMass Amherst, 2010–2011.