

Byron C. Wallace

Appointments Assistant Professor
College of Computer and Information Science, Northeastern University

Assistant Professor (adjunct)
Dept. of Health Services and Public Policy, Brown University

Research Interests

My research is in natural language processing (NLP) and machine learning (ML) / data science, with an emphasis on applications in health informatics. In short, there are too few biomedical experts to make sense of the torrents of available health-related data and I am interested in mitigating this problem via novel computational approaches. For example, a large ongoing project of mine aims to semi-automate aspects of the evidence-based medicine pipeline via innovative ML methods. Another example of this research motif is my work developing statistical models of patient-doctor communication, ultimately with the aim of improving it. Furthermore, I have multiple ongoing projects that use NLP to analyze data from the web relevant to the patient experience. More generally, I work on core NLP and ML problems, e.g.: structured and unstructured classification techniques; semi-supervised learning; learning from imbalanced data; and interactive learning to make the best possible use of human time and expertise in hybrid human/computer systems. Most recently, I have been working on convolutional neural models for textual data. In general, I tend to be most excited by interdisciplinary research that motivates technical questions by way of interesting applications.

Education

- 2012 **Ph.D. Computer Science**, Tufts University, Medford, MA
Advisor Carla E. Brodley
- Ph.D. Thesis Machine Learning in Health Informatics: Making Better use of Domain Experts. *Selected as The Runner-Up for the 2013 ACM SIGKDD Doctoral Dissertation Award.*
- Committee Carla Brodley (Tufts), Roni Khardon (Tufts), Anselm Blumer (Tufts), Thomas A. Trikalinos (Brown) and Jaime Carbonell (CMU)
Members
- 2006 **B.S. Computer Science**, University of Massachusetts at Amherst, Amherst, MA
Minor in Philosophy

Professional Experience/Prior Appointments

- 2014–2016 **Assistant Professor**, School of Information, University of Texas, Austin, TX
- 2014–2016 **Assistant Professor (by courtesy)**, Department of Computer Science, University of Texas, Austin, TX
- 2012–2014 **Assistant Professor (Research)**, Center for Evidence-based Medicine
Brown University, Providence, RI
- 2008–2012 **Research Computer Scientist**, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, MA
- 2007–2008 **Research Assistant (in the machine learning group)**, Department of Computer Science, Tufts University, Medford, MA

Summer 2006 **Software Engineer**, IBM, Westford, MA

2002–2006 **Research Assistant (in the *RIPPLEs group*)**, Department of Computer Science, UMass, Amherst, MA

Publications

Refereed Conference Publications

A preamble (mostly for non-computer scientists): for historical reasons, conferences are a main venue of research dissemination in computer science. Conference proceedings comprise full-length articles and are peer-reviewed (usually by 2-4 reviewers). Acceptance rates for top-tier conference range from 10-30%.

1. Ye Zhang, Matt Lease, and Byron C. Wallace. Active Discriminative Text Representation Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 3386–3392. Association for the Advancement of Artificial Intelligence (AAAI), 2017.
2. An Thanh Nguyen, Byron C. Wallace, Junyi Jessy Li, Ani Nenkova, and Matthew Lease. Aggregating and Predicting Sequence Labels from Crowd Annotations. In *Proceedings of the Association for Computational Linguistics (ACL)*, pages 299–309. Association for Computational Linguistics (ACL), 2017.
3. Iain Marshall, Joël Kuiper, Edward Banner, and Byron C. Wallace. Automating biomedical evidence synthesis: Robotreviewer. In *Proceedings of the Association for Computational Linguistics (ACL), System Demonstrations*, pages 7–12. Association for Computational Linguistics, 2017.
4. Ye Zhang, Matthew Lease, and Byron C. Wallace. Exploiting domain knowledge via grouped weight sharing with application to text categorization. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 155–160. Association for Computational Linguistics, 2017.
5. Gaurav Singh, Iain Marshall, James Thomas, John Shawe-Taylor, and Byron C. Wallace. A neural candidate-selector architecture for automatic structured clinical text annotation. *International Conference on Information and Knowledge Management (CIKM)*, 2017.
6. Jette Henderson, Ryan Bridges, Joyce C. Ho, Byron C. Wallace, and Joydeep Ghosh. PheKnow-Cloud: A Tool for Evaluating High-Throughput Phenotype Candidates using Online Medical Literature. In *Proceedings of the AMIA Joint Summits on Translational Science*. American Medical Informatics Association Medical (AMIA), 2017. *Recipient of the 2017 AMIA Distinguished Clinical Research Informatics Paper Award.*
7. Silvio Amir, Glen Coppersmith, Paula Carvalho, Mario J. Silva, and Byron C. Wallace. Quantifying Mental Health from Social Media with Neural User Embeddings. In *Proceedings of Machine Learning in Health Care (MLHC)*. Journal of Machine Learning Research (JMLR) Conferences Track, 2017.
8. Zhiguo Yu, Byron C. Wallace, Todd Johnson, and Trevor Cohen. Retrofitting Concept Vector Representations of Medical Concepts to Improve Estimates of Semantic Similarity and Relatedness. In *Proceedings of MEDINFO*. International Medical Informatics Association (IMIA), 2017.
9. An T. Nguyen, Byron C. Wallace, and Matthew Lease. A Correlated Worker Model for Grouped, Imbalanced and Multitask Data. In *Proceedings of The Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 537–546. Association for Uncertainty in Artificial Intelligence Press, 2016.
10. Ye Zhang, Stephen Roller, and Byron C. Wallace. MGNC-CNN: A Simple Approach to Exploiting Multiple Word Embeddings for Sentence Classification. In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 1522–1527. Association for Computational Linguistics (ACL), 2016.
11. Silvio Moreira, Byron C. Wallace, Hao Lyu, Paula Carvalho, and Mário J. Gaspar da Silva. Modelling Context with User Embeddings for Sarcasm Detection in Social Media. In *Proceedings of the Conference on Computational Natural Language Learning (CoNLL)*, pages 167–177. Association for Computational Linguistics (ACL) / Special Interest Group on Natural Language Learning (SIGNLL), 2016.
12. An T. Nguyen, Matthew Halpern, Byron C. Wallace, and Matthew Lease. Probabilistic Modeling for Crowdsourcing Partially-Subjective Ratings. In *Proceedings of The Conference on Human Computation and Crowdsourcing (HCOMP)*, pages 149–158. Association for the Advancement of Artificial Intelligence (AAAI), 2016.

13. Ye Zhang, Iain J. Marshall, and Byron C. Wallace. Rationale-Augmented Convolutional Neural Networks for Text Classification. In *Proceedings of Empirical Methods in Natural Language Processing (EMNLP)*, pages 795–804. Association for Computational Linguistics (ACL), 2016.
14. Mengqi Jin, Hongli Li, Christopher H. Schmid, and Byron C. Wallace. Using Electronic Medical Records and Physician Data to Improve Information Retrieval for Evidence-Based Care. In *Proceedings of the IEEE International Conference on Healthcare Informatics (ICHI)*. Institute of Electrical and Electronics Engineers (IEEE), 2016.
15. An T. Nguyen, Byron C. Wallace, and Matt Lease. Combining Crowd and Expert Labels Using Decision Theoretic Active Learning. In *AAAI Conference on Human Computation and Crowdsourcing (HCOMP)*, pages 120–129. Association for the Advancement of Artificial Intelligence (AAAI), 2015.
16. Finale Doshi-Velez, Byron C. Wallace, and Ryan P. Adams. Graph-Sparse LDA: a topic model with structured sparsity. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 2575–2581. Association for the Advancement of Artificial Intelligence (AAAI), 2015.
17. Byron C. Wallace, Do Kook Choe, and Eugene Charniak. Sparse, contextually informed models for irony detection: Exploiting user communities, entities and sentiment. In *Proceedings of the Association for Computational Linguistics (ACL)*, pages 1035–1044. Association for Computational Linguistics (ACL), 2015.
18. Iain J. Marshall, Joël Kuiper, and Byron C. Wallace. Automating risk of bias assessment for clinical trials. In *Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Health Informatics (BCB)*, pages 88–95. Association for Computing Machinery (ACM), 2014. *Selected for inclusion in a forthcoming special issue of the IEEE Journal of Biomedical and Health Informatics.*
19. Kelly H. Moran, Byron C. Wallace, and Carla E. Brodley. Discovering better AAAI keywords via clustering with community-sourced constraints. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 1265–1271. AAAI, 2014.
20. Byron C. Wallace, Do Kook Choe, Laura Kertz, and Eugene Charniak. Humans require context to infer ironic intent (so computers probably do, too). In *Proceedings of the Association for Computational Linguistics (ACL)*, pages 512–516. Association for Computational Linguistics (ACL), 2014.
21. Byron C. Wallace, Issa J. Dahabreh, Thomas A. Trikalinos, M. Barton Laws, Ira B. Wilson, and Eugene Charniak. Identifying differences in physician communication styles with a log-linear transition component model. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI)*, pages 1314–1320. AAAI, 2014.
22. Joël Kuiper, Iain J. Marshall, Byron C. Wallace, and Morris A. Swertz. Spa: a web-based viewer for text mining in evidence based medicine. In *Proceedings of the European Conference on Machine Learning (ECML)*, pages 452–455. Springer, 2014.
23. Byron C. Wallace, Thomas A. Trikalinos, M. Barton Laws, Ira B. Wilson, and Eugene Charniak. A generative joint, additive, sequential model of topics and speech acts in patient-doctor communication. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1765–1775, Seattle, Washington, USA, October 2013. Association for Computational Linguistics (ACL).
24. Byron C. Wallace and Issa J. Dahabreh. Class probability estimates are unreliable for imbalanced data (and how to fix them). In *Proceedings of the International Conference on Data Mining (ICDM)*, pages 695–704. Institute of Electrical and Electronics Engineers (IEEE), 2012. *Selected as one of the ‘best of ICDM-2012’.*
25. Byron C. Wallace, Kevin Small, Carla E. Brodley, Joseph Lau, and Thomas A. Trikalinos. Deploying an interactive machine learning system in an evidence-based practice center: abstrackr. In *Proceedings of the ACM SIGHIT International Health Informatics Symposium*, pages 819–824. Association for Computing Machinery (ACM), 2012.
26. Byron C. Wallace. Multiple narrative disentanglement: Unraveling infinite jest. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 1–10. Association for Computational Linguistics (ACL), 2012.
27. Byron C. Wallace, K. Small, Carla E. Brodley, and Thomas A. Trikalinos. Class imbalance, redux. In *Proceedings of the International Conference on Data Mining (ICDM)*, pages 754–763. Institute of Electrical and Electronics Engineers (IEEE), 2011.
28. Kevin Small, Byron C. Wallace, Carla E. Brodley, and Thomas A. Trikalinos. The constrained weight space SVM: Learning with ranked features. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 754–763, 2011.
29. Byron C. Wallace, Kevin Small, Carla E. Brodley, and Thomas A. Trikalinos. Who should label what? instance allocation in multiple expert active learning. In *Proceedings of the International Conference on Data Mining (SDM)*, pages 176–187. SIAM, 2011.

30. Byron C. Wallace, Kevin Small, Carla E. Brodley, and Thomas A. Trikalinos. Active learning for biomedical citation screening. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 173–182. Association for Computing Machinery (ACM), 2010.
31. Byron C. Wallace, Kevin Small, Carla E. Brodley, Joseph Lau, and Thomas A. Trikalinos. Modeling Annotation Time to Reduce Workload in Comparative Effectiveness Reviews. In *Proceedings of the 1st ACM International Health Informatics Symposium (IHI)*, pages 28–35. Association for Computing Machinery (ACM), 2010.

Journal Articles

32. Michael L. Mortensen, Gaelen P. Adam, Thomas A. Trikalinos, Tim Kraska, and Byron C. Wallace. An exploration of crowdsourcing citation screening for systematic reviews. *Research Synthesis Methods*, 2017.
33. Byron C. Wallace, Anna Noel-Storr, Iain J. Marshall, Aaron M. Cohen, Neil R. Smalheiser, and James Thomas. Identifying reports of Randomized Controlled Trials (RCTs) via a hybrid machine learning and crowdsourcing approach. *Journal of the American Medical Informatics Association (JAMIA)*, 2017.
34. Byron C. Wallace, Marc J. Lajeunesse, George Dietz, Issa J. Dahabreh, Thomas A. Trikalinos, Christopher H. Schmid, and Jessica Gurevitch. OpenMEE: Intuitive, open-source software for meta-analysis in ecology and evolutionary biology. *Methods in Ecology and Evolution*, 2017.
35. Ye Zhang, Erin Willis, Michael J Paul, Noémie Elhadad, and Byron C Wallace. Characterizing the (perceived) newsworthiness of health science articles: A data-driven approach. *JMIR Medical Informatics*, 4(3):e27, 2016.
36. Byron C. Wallace, Joël Kuiper, Aakash Sharma, Mingxi (Brian) Zhu, and Iain J. Marshall. Extracting pico sentences from clinical trial reports using supervised distant supervision. *Journal of Machine Learning Research*, 17(132):1–25, 2016.
37. Zhiguo Yu, Elmer Bernstam, Trevor Cohen, Byron C. Wallace, and Todd R. Johnson. Improving the utility of MeSH terms using the TopicalMeSH representation. *Journal of Biomedical Informatics*, 61:77–86, 2016.
38. Iain J. Marshall, Joël Kuiper, and Byron C. Wallace. RobotReviewer: Evaluation of a System for Automatically Assessing Bias in Clinical Trials. *Journal of the American Medical Informatics Association (JAMIA)*, 23(1):193–201, 2016.
39. Iain J. Marshall, Joël Kuiper, and Byron C. Wallace. Automating Risk of Bias Assessment for Clinical Trials. *Journal of Biomedical and Health Informatics (JBHI)*, 19(4):1406–1412, 2015. (An invited journal version of our 2014 ACM-BCB paper of the same the title.).
40. Byron C. Wallace, M. Barton Laws, Kevin Small, Ira B. Wilson, and Thomas A. Trikalinos. Automatically annotating topics in transcripts of patient-provider interactions via machine learning. *Medical Decision Making*, 34(4):503–512, 2014. *Highlighted in an editorial piece entitled “From Text Tagging to Decision Support” by HP Lehmann, MDM, 2014 (<http://mdm.sagepub.com/content/34/4/414.extract>).*
41. Byron C. Wallace and Issa J. Dahabreh. Improving class probability estimates for imbalanced data. *Knowledge and Information Systems (KAIS)*, 41:33–52, 2014.
42. Byron C. Wallace, Michael J. Paul, Urmimala Sarkar, Thomas A. Trikalinos, and Mark Dredze. A large-scale quantitative analysis of latent factors and sentiment in online doctor reviews. *Journal of the American Medical Informatics Association (JAMIA)*, 21:1098–1103, 2014.
43. Byron C. Wallace. Computational irony: A survey and new perspectives. *Artificial Intelligence Review*, 43(4):1–17, 2013.
44. Byron C. Wallace, Issa J. Dahabreh, Christopher H. Schmid, Joseph Lau, and Thomas A. Trikalinos. Modernizing the systematic review process to inform comparative effectiveness: tools and methods. *Journal of Comparative Effectiveness Research*, 2(3):273–282, 2013.
45. Carla E. Brodley, Umaa Rebbapragada, Kevin Small, and Byron C. Wallace. Challenges and opportunities in applied machine learning. *Artificial Intelligence Magazine*, 33(1):11–24, 2012.
46. Byron C. Wallace, Issa J. Dahabreh, Thomas A. Trikalinos, Joseph Lau, Paul Trow, and Christopher H. Schmid. Closing the gap between methodologists and end-users: R as a computational back-end. *Journal of Statistical Software*, 49(5):1–15, 6 2012.
47. Byron C. Wallace, Kevin Small, Carla E. Brodley, Joseph Lau, Christopher H. Schmid, Lars Bertram, Christina M. Lill, Joshua T. Cohen, and Thomas A. Trikalinos. Toward modernizing the systematic review pipeline in genetics: efficient updating via data mining. *Genetics in Medicine*, 14(7):663–669, 2012.
48. R.D. Whitaker, S. Pember, B.C. Wallace, C.E. Brodley, and D.R. Walt. Single cell time-resolved quorum responses reveal dependence on cell density and configuration. *Journal of Biological Chemistry*, 286(24):21623–21632, 2011.

49. PJ Castaldi, MH Cho, M Cohn, F Langerman, S Moran, N Tarragona, H Moukhachen, R Venugopal, D Hasimja, E Kao, BC Wallace, CP Hersh, S Bagade, L Bertram, EK Silverman, and TA Trikalinos. The copd genetic association compendium: a comprehensive online database of copd genetic associations. *Human Molecular Genetics*, 19(3):526–534, 2010.
50. Byron C. Wallace, Thomas A. Trikalinos, Joseph Lau, Carla E. Brodley, and Christopher H. Schmid. Semi-automated screening of biomedical citations for systematic reviews. *BMC Bioinformatics*, 11(1):55+, 2010.
51. Byron C. Wallace, Christopher H. Schmid, Joseph Lau, and Thomas A. Trikalinos. Meta-analyst: software for meta-analysis of binary, continuous and diagnostic data. *BMC Medical Research Methodology*, 9(1):80+, 2009.

Refereed Workshop/Symposium Publications and Technical Reports

52. Gaurav Singh, Iain Marshall, James Thomas, and Byron C. Wallace. Identifying diagnostic test accuracy publications using a deep model. *CLEF eHealth*, 2017.
53. Byron C. Wallace and Michael J. Paul. ‘Jerk’ or ‘Judgemental’? Patient perceptions of male versus female physicians in online reviews. In *Proceedings of the International Workshop on the World Wide Web and Public Health Intelligence (W3PHI)*. AAAI, 2017.
54. Ryan Bridges, Jette Henderson, Joyce Ho, Byron C Wallace, and Joydeep Ghosh. Automated Verification of Phenotypes using PubMed. In *Proceedings of the Workshop on Methods and Applications in Healthcare Analytics at ACM BCB*, 2016.
55. Yalin Sun, Shengwei Wang, Pengxiang Cheng, Hao Lyu, Iain Marshall, and Byron C. Wallace. Crowdsourcing Information Extraction for Biomedical Systematic Reviews. In *Human Computation and Crowdsourcing (HCOMP) Works-in-Progress*, 2016.
56. Elisa Ferracane, Iain Marshall, Byron C. Wallace, and Katrin Erk. Leveraging coreference to identify arms in medical abstracts: An experimental study. In *The International Workshop on Health Text Mining and Information Analysis at EMNLP*, 2016.
57. Zhiguo Yu, Trevor Cohen, Todd R. Johnson, Byron C. Wallace, and Elmer Bernstam. Retrofitting Word Vectors of MeSH Terms to Improve Semantic Similarity Measures. In *The International Workshop on Health Text Mining and Information Analysis at EMNLP*, 2016.
58. Matthew Lease, Cormack V. Gordon, An Thanh Nguyen, Thomas A. Trikalinos, and Byron C. Wallace. Systematic Review is e-Discovery in Doctor’s Clothing. In *Proceedings of the Medical Information Retrieval (MedIR) Workshop at the International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2016.
59. Yu Zhiguo, Byron C. Wallace, and Todd R. Johnson. Healthcare data analytics challenge. In *Proceedings of the International Conference on Healthcare Informatics (ICHI)*. IEEE, 2015. (*Describes our submission to the challenge, which won first place.*)
60. Ye Zhang and Byron C. Wallace. A sensitivity analysis of (and practitioners’ guide to) convolutional neural networks for sentence classification. *arXiv preprint arXiv:1510.03820*, 2015.
61. Byron C. Wallace, Michael J. Paul, and Noemie Elhadad. What predicts media coverage of health science articles? In *Proceedings of the International Workshop on the World Wide Web and Public Health Intelligence (W3PHI)*. AAAI, 2015.
62. Byron C. Wallace, Issa J. Dahabreh, Kelly H. Moran, Carla E. Brodley, and Thomas A. Trikalinos. Active Literature Discovery for Scoping Evidence Reviews: How Many Needles are There? In *Proceedings of the KDD Workshop on Data Mining for Healthcare (KDD-DMH)*, 2013.
63. Michael J. Paul, Byron C. Wallace, and Mark Dredze. What Affects Patient (Dis)satisfaction? Analyzing Online Doctor Ratings with a Joint Topic-Sentiment Model. In *Proceedings of the AAAI Workshop on Expanding the Boundaries of Health Informatics Using AI (HIAI)*. AAAI, 2013.
64. Byron C. Wallace, Kevin Small, Carla E. Brodley, and Thomas A. Trikalinos. Active learning for biomedical citation screening. In *Proceedings of the 2010 Northeastern Student Conference on Artificial Intelligence (NESCAI)*, 2010.

Book Chapters

65. Byron C. Wallace, Issa J. Dahabreh, Christopher H. Schmid, Joseph Lau, and Thomas A. Trikalinos. Modernizing evidence synthesis for evidence-based medicine. In Robert A. Greenes, editor, *Clinical Decision Support (Second Edition)*, chapter 12. Wiley, 2013.

Commentaries and Editorials

66. Jenna Wiens and Byron C. Wallace. Editorial: special issue on machine learning for health and medicine. *Machine Learning Journal*, 2015.
67. Byron C. Wallace and Iain J. Marshall. Invited Response to: Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Cochrane Methods*, 2015.
68. Darius Braziuinas David L. Buckeridge Heriberto Cuayáhuitl Nina Dethlefs Markus Endres Amir-massoud Farahmand Mark Fox Lutz Frommberger Sam Ganzfried Yolanda Gil Sébastien Guillet Lawrence E. Hunter Arnav Jhala Kristian Kersting George Konidaris Freddy Lecue Sheila McIlraith Sriraam Natarajan Zeinab Noorian David Poole Rémi Ronfard Alessandro Saffiotti Arash Shaban-Nejad Biplav Srivastava Gerald Tesaro Rosario Uceda-Sosa Guy Van den Broeck Martijn van Otterlo Byron C. Wallace Paul Weng Jenna Wiens Jie Zhang Stefano V. Albrecht, André M. S. Barreto. Reports of the AAAI 2014 Conference Workshops. *Artificial Intelligence Magazine*, 2015.
69. Julian Elliott, Ida Sim, Jessica Thomas, Nancy Owens, Gordon Dooley, Jacob Riis, Byron Wallace, James Thomas, Anna Noel-Storr, Gabriel Rada, Caroline Struthers, Tracey Howe, Harriet MacLehose, Linn Brandt, Ilkka Kunnamo, and Chris Mavergames. #CochraneTech: Technology and the Future of Systematic Reviews. *The Cochrane Library*, 2014.
70. Byron C. Wallace and Thomas A. Trikalinos. Invited Response to: Applications of text mining within systematic reviews. *Cochrane Methods*, 2012.

Selected Invited Talks

Note that this is non-exhaustive and that these are in addition to the talks accompanying the conference publications above.

1. Expediting Clinical Evidence Synthesis via Machine Learning, Natural Language Processing and Crowdsourcing. *Keynote at the 4th International Symposium on Systematic Review and Meta-Analysis of Laboratory Animal Studies*, 08/24/2017, 2017.
2. Expediting Clinical Evidence Synthesis via Machine Learning, Natural Language Processing and Crowdsourcing. *Weill Cornell Medicine, Machine Learning in Medicine series*, 07/14/2017, 2017.
3. Expediting Clinical Evidence Synthesis via Machine Learning, Natural Language Processing and Crowdsourcing. *MIT (MEDG Reading Group)*, 06/02/2017, 2017.
4. *Invited panel member*: Data mining, clinical medicine, and medical research: The current state and future of the nexus. *Statistics and Data Mining (SDM)*, 4/28/2017, 2017.
5. Expediting Clinical Evidence Synthesis via Machine Learning, Natural Language Processing and Crowdsourcing. *UMass, Worcester Medical School (BioNLP group)*, 01/05/2017, 2017.
6. Data Science, Social Media and Health. *Promoting Research in Social Media and Health Symposium (PRISM Health Symposium)*, UCSF, San Francisco, CA, 12/02/2016, 2016.
7. Some recent work on Convolutional Neural Networks (CNNs) for text classification. *Raytheon BBN Technologies (BBN)*, Cambridge, MA, 11/22/2016, 2016.
8. Expediting Clinical Evidence Synthesis via Machine Learning, Natural Language Processing and Crowdsourcing. *Amazon Machine Learning Lecture Series*, Seattle, WA, 11/17/2016, 2016.
9. Expediting Clinical Evidence Synthesis via Machine Learning, Natural Language Processing and Crowdsourcing. *Evidence Synthesis: Current Practices and Future Possibilities (Panel at IEEE ICHI)*, Chicago, IL, 10/5/2016, 2016.
10. Expediting Clinical Evidence Synthesis via Machine Learning, Natural Language Processing and Crowdsourcing. *Penn Institute for Computational Science (PICS) Symposium on Emerging Paradigms in Scientific Discovery*, Philadelphia, PA, 10/6/2016, 2016.
11. Open Challenges in Automating Clinical Evidence Synthesis. *International Conference on the Automation of Systematic Reviews*, Philadelphia, PA, 10/3/2016, 2016.

12. Some recent work on Convolutional Neural Networks (CNNs) for text classification. *MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA*, 10/11/2016, 2016.
13. Some recent work on Convolutional Neural Networks (CNNs) for text classification. *University of Lisbon, Portugal*, 6/22/2016, 2016.
14. Automating Evidence Synthesis via Machine Learning and Natural Language Processing. *Machine Learning and Data Analytics Symposium (MLDAS), Doha, Qatar*, 3/15/2016, 2016.
15. Automating evidence synthesis via machine learning and natural language processing. *Symposium on Text Mining, Bristol University*, 11/11/2015, 2015.
16. Automating evidence synthesis via machine learning and natural language processing. *Department of Biostatistics, Brown University*, 10/19/2015, 2015.
17. Automating evidence synthesis via machine learning and natural language processing. *Department of Computer and Information Science, University of Pennsylvania*, 3/31/2015, 2015.
18. Automating evidence synthesis. *Department of Biomedical Informatics, University of Texas at Houston*, 10/1/2014, 2014.
19. Can machine learning help us improve physician communication? *Department of Biomedical Informatics, Columbia University, NYC*, 5/15/2014, 2014.
20. Automating the systematic review: Tools and methods. *National Institute of Environmental Health Sciences, Durham, North Carolina*, 4/30/2014, 2014.
21. Machine Learning in Evidence-Based Medicine: Managing the Clinical Data Deluge. *University of Groningen, Netherlands*, 4/24/2014, 2014.
22. Can Machine Learning Help us Improve Physician Communication? *MIT Lincoln Lab, Lexington, MA*, 11/1/2013, 2013.
23. Can Machine Learning Help us Improve Physician Communication? *Data Management Lab at Boston University, Boston, MA*, 11/8/2013, 2013.
24. Semi-Automating Systematic Reviews: Text mining, NLP and machine learning. *#CochraneTech Symposium: Technology and the future of the systematic review, Quebec City, Canada*, 9/2013, 2013.
25. Statistical Models of Patient-Doctor Communication. *Meaningful Use of Complex Medical Data (MUCMD), LA, CA*, 8/16/2013, 2013.
26. Machine learning in health informatics: Making better use of domain experts (doctoral dissertation awards session presentation). *The 19th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Chicago, IL*, 8/12/2013, 2013.
27. Machine learning in systematic reviews: Making better use of domain expertise. *Statistical Analysis of "Big Data" Group, Brown University, Providence, RI*, 11/09/2012, 2012.
28. Better models, less effort: Active learning and dual supervision. *Department of Computer Science at UMass, Boston*, 03/07/2012, 2012.

Academic Honors & Professional Awards

- Our paper, "PheKnow-Cloud: A Tool for Evaluating High-Throughput Phenotype Candidates using Online Medical Literature" received the *Distinguished Clinical Research Informatics Paper Award* at the AMIA Joint Summits; joint work with Jette Henderson, Ryan Bridges, Joyce Ho and Joydeep Ghosh. 2017.
- Our system won the *Healthcare Data Analytics* at the IEEE International Conference on Healthcare Informatics (ICHI); joint work with Zhiguo Yu and Todd Johnson. 2015.
- Recipient of the *ACM SIGKDD Dissertation Award (Runner-Up)*. Association for Computational Machinery (ACM). 2013.
- Recipient of the *Outstanding Graduate Researcher at the Doctoral Level* award. Tufts University. 2012.
- Recipient of the *Outstanding Student in the Area of Systems in Computer Science* award. UMass, Amherst. 2006.
- Recipient of the 2005-2006 *Jonathon Edwards Philosophy Essay Prize*. UMass, Amherst. 2006.

- Recipient of the 2005-2006 *Gerald F. Scanlon Student Employee of the Year Award*.
UMass, Amherst. 2006.

Research Support (Grants)

Grant Title Hybrid Approaches to Optimizing Evidence Synthesis via Machine Learning and Crowdsourcing (1R03HS025024-01)
Funder Agency for Healthcare Research and Quality (AHRQ)
Role Principal Investigator
Program Small Research Grants; Special Emphasis on Innovative Methods Research to Increase the Utility of Systematic Reviews
Period 9/30/2016 – 9/29/2017
Amount ~\$100,000
Collaborators Ani Nenkova (Consultant)

Grant Title Crowdsourcing Mark-up of the Medical Literature to Support Evidence-Based Medicine and Develop Automated Annotation Capabilities (1UH2CA203711-01)
Funder National Institutes of Health/National Cancer Institute
Role Principal Investigator (PD/PI)
Program Big Data to Knowledge (BD2K)
Period 6/1/2016 – 6/1/2018
Amount ~\$311,371
Collaborators MPIs Ani Nenkova and Zachary Ives (UPenn)

Grant Title Semi-Automating Data Extraction for Systematic Reviews (1R01LM012086-01A1)
Funder National Institutes of Health/National Library of Medicine
Role Principal Investigator (PD/PI)
Period 9/20/2015 – 8/31/2019
Amount ~\$1.3 million total
Collaborators MPIs Iain J. Marshall (King's College), Thomas Trikalinos (Brown), Randolph Bias (UT Austin)

Grant Title Influencing Cervical Cancer Prevention and Detection Online through Social Media (R01CA178875)
Funder National Institutes of Health/National Cancer Institute
Role Investigator
Period 9/2014 – 9/2017
Amount ~\$1.3 million total; UT portion ~\$92,000
Collaborators MPIs Urmimala Sarkar (UCSF) and Damon Centola (UPenn)

Grant Title Expression and Recognition of Irony in Multicultural Social Media (UTAP-EXPL/EEI-ESS/0031/2014)
Funder FCT
Role Investigator
Period 1/2015 – 12/31/2015
Amount UT portion: ~\$34,000

Collaborators Principal Investigators Paula Carvalho, Investigators Mário J. Silva and Bruno Martins (all collaborators at University of Lisbon)

Grant Title Sociolinguistically Informed Natural Language Processing: Automating Irony Detection (W911NF1410442)

Funder Army Research Office (ARO)

Role Principal Investigator

Program Mathematical Sciences Division

Period 9/2013 – 9/2017

Amount ~\$400,000

Collaborators Co-Investigators Eugene Charniak, Thomas Trikalinos and David Beaver

Grant Title Scaling Evidence-Based Medicine via Automation and Crowdsourcing

Funder Brown University

Role Principal Investigator

Program Seed Funding

Period 5/2014 – 8/2015

Amount ~\$80,000

Collaborators Tim Kraska, Ugur Cetintemel and Thomas Trikalinos

Other support I have twice received support from the Amazon AWS Cloud Credits for Research program. Once in 2014 (for \$2,000) and again in 2016 (for \$4,500)

Past Projects (Grants) I Have Worked on

Grant Title Making Advanced Statistical Tools Accessible for Quantitative Research Synthesis and Discovery in Ecology and Evolutionary Biology (DBI-1262442)

Funder National Science Foundation (NSF)

Role Principal Investigator

Program Collaborative research: ABI Development: Making Advanced Statistical Tools Accessible for Quantitative Research Synthesis and Discovery in Ecology and Evolutionary Biology.

Period 5/2013 – 5/2016

Amount ~\$500,000

Collaborators Co-PI Jessica Gurevitch (SUNY), Co-PI Marc Lajeunesse, Investigators Thomas Trikalinos and Christopher H. Schmid (Brown)

Grant Title Manual and Automatic Analysis of Patients' Values and Preferences Using Seton HC-AHPS Surveys

Funder Seton/CHASP

Role Principal Investigator

Period 8/1/2015 – 9/1/2016

Amount ~\$50,000

Collaborators Co-PIs Kenneth Fleischmann and Bo Xie

Grant Title R01 HS 018494: Semi-automating Citation Screening for Systematic Reviews

Funder AHRQ
Amount \$1.2 million
Period 12/01/2009–11/30/2012
Role Key Investigator
Collaborators Principal Investigator Thomas A. Trikalinos, Investigator Joseph Lau

Grant Title R01 HS 018574: Modernizing Meta-Analysis to Facilitate Comparative Effectiveness Reviews
Funder AHRQ
Amount \$1.2 million
Period 12/01/2009–11/30/2012
Role Key Investigator
Collaborators Principal Investigator Christopher H. Schmid, Investigator Thomas A. Trikalinos, Investigator Joseph Lau

Service

Community

I co-chaired the *Machine Learning in Healthcare* (MLHC) conference in 2016 and am doing so again in 2017.

I guest edited a special issue in the Machine Learning Journal (MLJ) on applications in health (alongside Jenna Wiens).

I am co-chairing The 3rd International Workshop on the World Wide Web and Public Health Intelligence (W3PHI) 2016.

I co-organized the workshop *Modern Artificial Intelligence in Health Analytics (MAIHA)* at AAAI 2014, along with Jenna Wiens, David Kale and Finale Doshi-Velez.

I co-organized the workshop *Can Cognitive Scientists Help Computers Recognize Irony?* at CogSci 2014, in collaboration with Laura Kertz.

Grant/Proposal Reviewing

I served on review panels for the National Institutes of Health (NIH) in academic years 2015-2016 and 2016-2017.

I served on review panels for the National Science Foundation (NSF) in academic years 2013-2014 and 2014-2015.

I have reviewed proposals for the Army Research Office (ARO); mathematical sciences division (academic year 2014-2015).

Peer Reviewing &etc.

I am an action editor for the Machine Learning Journal (MLJ).

I am on the editorial board of Research Synthesis Methods (RSM), and have served as the Computational Tools and Methods Editor from Summer 2017.

I will be an senior program committee member for the 2018 conference of the *Association for the Advancement of Artificial Intelligence (AAAI)*. I was an area chair for 2016 Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL), a senior program committee member for the 2014 conference of the *Association for the Advancement of Artificial Intelligence (AAAI)* and a senior program committee member for *Knowledge Discovery and Databases (KDD)* in 2015.

Conference program committees on which I have served (or am serving): *2017 International Conference Machine Learning (ICML)*, *2017 International Joint Conference on Artificial Intelligence (IJCAI)*, *2017 Association for Computational Linguistics (ACL)*, *2017 World Wide Web conference (WWW)*, *2017 European Association for Computational Linguistics (EACL)*, *2017 Association for the Advancement of Artificial Intelligence (AAAI)*, *2016 Neural Information Processing (NIPS)*, *2016 Association for Computational Linguistics (ACL)*, *2016 International Conference on Machine Learning (ICML)*, *2016 Knowledge Discovery and Databases (KDD)*, *2016 AAAI Conference on Human Computation and Crowdsourcing (HCOMP)*, *2016 International Joint Conference on Artificial Intelligence (IJCAI)*, *2015 Association for the Advancement of Artificial Intelligence (AAAI)*, *2015 Neural Information Processing (NIPS)*, *2015 European Conference on Machine Learning (ECML)*, *2015 American Medical Informatics Association (AMIA)*, *2015 International Conference on Machine Learning (ICML)*, *2015 AAAI WWW and Public Health Intelligence (W3PHI)*, *2014 American Medical Informatics Association (AMIA)*, *2014 Knowledge Discovery and Databases (SIGKDD) Workshop on Health Informatics*, *2014 European Conference on Machine Learning (ECML)*, *2014 International Conference on Data Mining (ICDM)*, *2013 European Conference on Machine Learning (ECML)*, *2013 Association for the Advancement of Artificial Intelligence (AAAI)*, *2013 meeting of the American Medical Informatics Association (AMIA)*, *2013 IEEE Conference on Big Data*, *2012 Knowledge Discovery and Databases (SIGKDD) Workshop on Health Informatics*, *2012 ACM International Conference on Information and Knowledge Management (CIKM)*

Other conferences for which I have peer-reviewed (as an external reviewer) *2011 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, *2011 Conference on Empirical Methods in Natural Language Processing (EMNLP)*

Journals for which I have reviewed: *BMC Bioinformatics*, *BMC Medical Informatics and Decision Making*, *BMC Medical Research Methodology*, *BMC Research Notes*, *Bioinformatics*, *Current Bioinformatics*, *Data Mining and Knowledge Discovery*, *IEEE Transactions on Knowledge and Data Engineering*, *International Journal of Epidemiology*, *Journal of the American Medical Association (JAMA)*, *Journal of the American Medical Informatics Association (JAMIA)*, *Journal of Biomedical Informatics (JBI)*, *Journal of Clinical Epidemiology (JCE)*, *Journal of Machine Learning Research (JMLR)*, *Journal of Machine Learning and Cybernetics (JMLC)*, *Journal of Medical Internet Research (JMIR)*, *Journal of Systematic Reviews*, *Knowledge and Information Systems (KAIS)*, *Language Resources and Evaluation*, *Omega: The International Journal of Management Science*, *Research Synthesis Methods (RSM)*

Advising/Thesis Committees

I have served on the following PhD committees:

David Inouye. Primary advisor: Inderjit Dhillon (UT AUSTIN). Other committee members: Pradeep Ravikumar (UT Austin), Raymond Mooney (UT Austin), Qixing Huang (UT Austin). Thesis title: *Appropriate, Accessible and Appealing Probabilistic Graphical Models*. 2017.

David Batista. Primary advisor: Mario Gaspar da Silva (Universidade de Lisboa/INESC-ID). Other committee members: Paulo Quaresma (Universidade de Évora), David Manuel Martins de Matos (Universidade de Lisboa). Thesis title: *Large-Scale Semantic Relationship Extraction for Information Discovery*. 2016.

Hyun Joon Jung (now a Data Scientist at Apple). Primary advisor: Matt Lease. Other committee members: Paul Bennett (Microsoft Research), Ken Fleischmann (UT Austin), Raymond Mooney (UT Austin). Thesis title: *Temporal Modeling of Crowd Work Quality for Quality Assurance in Crowdsourcing*. 2015.

Louise Millard (now a Research Associate in Data Mining at Bristol), Bristol University. Co-supervisors: Julian Higgins and Peter Flach (Bristol University). Other committee members: William Hollingworth (Bristol University). Thesis title: *Towards data-intensive epidemiology: Explorations in systematic reviews and causal inference*. 2015.

Memberships

I am a member of the *Society for Research Synthesis Methodology* (SRSM), an interdisciplinary society interested in quantitative methods for information synthesis. SRSM membership is by invitation only.

Teaching

I taught Artificial Intelligence in Spring 2017: <http://www.byronwallace.com/ccis4100>.

I led a seminar class on machine learning in health in the College of Computer and Information Science at Northeastern in fall 2016: <http://www.byronwallace.com/ccis7180>.

I developed and taught the new course *Applied Data Mining* within the iSchool at UT Austin in fall 2014 and repeated it in spring 2015 and spring 2016. I taught *Database Management* in fall 2015.

I was co-head instructor of CS150 AIH: *Artificial Intelligence in Health Informatics* alongside Dr. Kevin Small in the Computer Science department at Tufts University. Fall 2011. <http://www.cs.tufts.edu/comp/150AIH/>