

Byron C. Wallace

Research interests

My research is in natural language processing and machine learning, with an emphasis on applications in health informatics.

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.
- 2006 **B.S. Computer Science**, University of Massachusetts at Amherst, Amherst, MA
Minor in Philosophy

Appointments & experience

- 9/2016 – present **Assistant Professor**
College of Computer and Information Science, Northeastern University
Assistant Professor (Adjunct)
Department of Health Services and Public Policy, Brown University
- 9/2014 – 8/2016 **Assistant Professor**
School of Information, University of Texas at Austin
Assistant Professor (Courtesy)
Department of Computer Science, University of Texas at Austin
- 6/2012 – 8/2014 **Assistant Professor (Research)**
Health Services and Public Policy, Brown University
- 9/2008 – 6/2012 **Research Computer Scientist**
Institute for Clinical Research and Health Policy Studies, Tufts Medical Center
- 9/2007 – 9/2008 **Research Assistant**
The *machine learning group*, Department of Computer Science, Tufts University
- 9/2002 – 6/2006 **Research Assistant**
RIPPLEs *group*, Department of Computer Science, UMass Amherst

Academic honors & professional awards

- 2018 Recipient of the annual Early Career Award from the *Society for Research Synthesis Methodology (SRSM)*.
- 2018 NSF CAREER Award: *Structured Scientific Evidence Extraction: Models and Corpora*.
- 2018 Acknowledged as a *Top Reviewer* for ACL 2018 (<https://acl2018.org/2018/07/02/top-reviewers/>)
- 2017 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.
- 2015 Our system won the *Healthcare Data Analytics Challenge* at the IEEE International Conference on Healthcare Informatics (ICHI); joint work with Zhiguo Yu and Todd Johnson.
- 2014 Inducted as a member of the *Society for Research Synthesis Methodology (SRSM)*, an interdisciplinary group of researchers interested in quantitative methods for information synthesis; membership is by invitation only.
- 2013 Recipient of the *ACM SIGKDD Dissertation Award (Runner-Up)*. Association for Computational Machinery (ACM).
- 2012 Recipient of the *Outstanding Graduate Researcher at the Doctoral Level* award. Tufts University.
- 2006 Recipient of the *Outstanding Student in the Area of Systems in Computer Science* award. UMass, Amherst.
- 2006 Recipient of the *Jonathon Edwards Philosophy Essay Prize*. UMass, Amherst.
- 2006 Recipient of the *Gerald F. Scanlon Student Employee of the Year Award*. UMass, Amherst.

Publications

Conference publications

These do not include workshop publications, which are listed in a separate section (Workshop & symposium publications), below.

1. Benjamin Nye, Jessy Li, Roma Patel, Yinfei Yang, Iain Marshall, Ani Nenkova, and Byron C. Wallace. A Corpus with Multi-Level Annotations of Patients, Interventions and Outcomes to Support Language Processing for Medical Literature. In *Proceedings of the Conference of the Association for Computational Linguistics (ACL)*, pages 197 – 207, 2018.
2. An Thanh Nguyen, Aditya Kharosekar, Matthew Lease, and Byron C. Wallace. An Interpretable Joint Graphical Model for Fact-Checking from Crowds. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2018.
3. An Thanh Nguyen, Byron C. Wallace, and Matthew Lease. Believe it or not: Designing a Human-AI Partnership for Mixed-Initiative Fact-Checking. In *Proceedings of the ACM User Interface Software and Technology Symposium (UIST)*, 2018. forthcoming.
4. Sarthak Jain, Edward Banner, Marshall Iain J. van de Meent, Jan-Willem, and Byron C. Wallace. Learning Disentangled Representations of Texts with Application to Biomedical Abstracts. In *Proceedings of Empirical Methods in Natural Language Processing (EMNLP)*, 2018. forthcoming.
5. Gaurav Singh, James Thomas, Iain J. Marshall, John Shawe-Taylor, and Byron C. Wallace. Structured Multi-Label Biomedical Text Tagging via Attentive Neural Tree Decoding. In *Proceedings of Empirical Methods in Natural Language Processing (EMNLP)*, 2018. forthcoming.
6. Roma Patel, Yinfei Yang, Iain Marshall, Ani Nenkova, and Byron C. Wallace. Syntactic Patterns Improve Information Extraction for Medical Search. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 371–377, 2018.

7. 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)*, pages 1519–1528, 2017.
8. Ye Zhang and Byron C. Wallace. A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification. *International Joint Conference on Natural Language Processing (IJCNLP)*, pages 253–263, 2017.
9. Ye Zhang, Matthew 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.
10. 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.
11. 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 (ACL), 2017.
12. Ye Zhang, Matthew Lease, and Byron C. Wallace. Exploiting Domain Knowledge via Grouped Weight Sharing with Application to Text Categorization. In *Proceedings of the Association for Computational Linguistics (ACL)*, pages 155–160. Association for Computational Linguistics (ACL), 2017.
13. 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*, pages 149 – 157. American Medical Informatics Association Medical (AMIA), 2017. *Recipient of the 2017 AMIA Distinguished Clinical Research Informatics Paper Award.*
14. 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.
15. 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*, pages 657 – 661. International Medical Informatics Association (IMIA), 2017.
16. 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.
17. 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.
18. 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.
19. 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.
20. 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.
21. Mengqi Jin, Hongli Li, Christopher Schmid, and Byron C. Wallace. Using Electronic Medical Records and Physician Data to Improve Information Retrieval for Evidence-Based Care. In *Proceedings of the International Conference on Healthcare Informatics (ICHI)*, pages 61–64. Institute of Electrical and Electronics Engineers (IEEE), 2016.
22. An T. Nguyen, Byron C. Wallace, and Matthew 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.
23. 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.
24. 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.

25. 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 special issue of the IEEE Journal of Biomedical and Health Informatics.*
26. 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. Association for the Advancement of Artificial Intelligence (AAAI), 2014.
27. 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.
28. 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. Association for the Advancement of Artificial Intelligence (AAAI), 2014.
29. 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.
30. 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. Association for Computational Linguistics (ACL), 2013.
31. 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'.*
32. 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.
33. 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.
34. 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.
35. Kevin Small, Byron C. Wallace, Carla E. Brodley, and Thomas A. Trikalinos. The Constrained Weight Space SVM: Learning with Ranked Reatures. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 754–763, 2011.
36. 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. Society for Industrial and Applied Mathematics (SIAM), 2011.
37. 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.
38. 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

39. Iain J. Marshall, Anna Noel-Storr, Joël Kuiper, James Thomas, and Byron C. Wallace. Machine learning for identifying randomized controlled trials: An evaluation and practitioner's guide. *Research Synthesis Methods*, pages 1–12, 2018.
40. Jette Henderson, Junyuan Ke, Joyce C. Ho, Joydeep Ghosh, and Byron C. Wallace. PIVET: A scaled phenotype evidence generation framework using online medical literature. *Journal of Medical Internet Research (JMIR)*, 2018. *forthcoming.*
41. 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*, 8(3):366–386, 2017.

42. 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)*, pages 1165–1168, 2017.
43. James Thomas, Anna Noel-Storr, Iain Marshall, Byron C. Wallace, Steven McDonald, Chris Mavergames, Paul Glasziou, Ian Shemilt, Anneliese Synnot, Tari Turner, et al. Living Systematic Reviews: 2. Combining Human and Machine Effort. *Journal of Clinical Epidemiology*, pages 31–37, 2017.
44. Kezban Dilek Onal, Ye Zhang, Ismail Sengor Altingovde, Md Mustafizur Rahman, Pinar Karagoz, Alex Braylan, Brandon Dang, Heng-Lu Chang, Henna Kim, Quinten McNamara, Aaron Angert, Edward Banner, Vivek Khetan, Tyler McDonnell, An Thanh Nguyen, Dan Xu, Byron C. Wallace, Maarten de Rijke, and Matthew Lease. Neural Information Retrieval: At the End of the Early Years. *Information Retrieval*, pages 1–72, 2017.
45. 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*, 8:941–947, 2017.
46. 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.
47. Byron C. Wallace, Joël Kuiper, Aakash Sharma, Mingxi (Brian) Zhu, and Iain Marshall. Extracting PICO Sentences from Clinical Trial Reports using Supervised Distant Supervision. *Journal of Machine Learning Research*, 17(132):1–25, 2016.
48. 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.
49. 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.
50. 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.).
51. 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>).
52. Byron C. Wallace and Issa J. Dahabreh. Improving class probability estimates for imbalanced data. *Knowledge and Information Systems (KAIS)*, 41:33–52, 2014.
53. 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.
54. Byron C. Wallace. Computational irony: A survey and new perspectives. *Artificial Intelligence Review*, 43(4):1–17, 2013.
55. Byron C. Wallace, Issa J. Dahabreh, Chistopher 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.
56. 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.
57. 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.
58. 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.
59. RD Whitaker, S Pember, BC Wallace, CE Brodley, and DR Walt. Single Cell Time-resolved Quorum Responses Reveal Dependence on Cell Density and Configuration. *Journal of Biological Chemistry*, 286(24):21623–21632, 2011.
60. 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.
61. 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.

62. 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.

Workshop/symposium publications & technical reports

63. Gaurav Singh, Iain Marshall, James Thomas, and Byron C. Wallace. Identifying Diagnostic Test Accuracy Publications using a Deep Model. *CLEF eHealth*, 2017.
64. 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)*. Association for the Advancement of Artificial Intelligence (AAAI), 2017.
65. Sarthak Jain, Xun Peng, and Byron C. Wallace. pages 72–75, 2017.
66. 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.
67. 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*. Association for the Advancement of Artificial Intelligence (AAAI), 2016.
68. 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*. Association for Computational Linguistics (ACL), 2016.
69. 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*. Association for Computational Linguistics (ACL), 2016.
70. 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.
71. 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.*)
72. 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)*. Association for the Advancement of Artificial Intelligence (AAAI), 2015.
73. 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.
74. 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)*. Association for the Advancement of Artificial Intelligence (AAAI), 2013.
75. 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

76. 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 & editorials

77. Jenna Wiens and Byron C. Wallace. Editorial: Special Issue on Machine Learning for Health and Medicine. *Machine Learning Journal*, pages 1–3, 2015.

78. Byron C. Wallace and Iain J. Marshall. Invited Response: Using Text Mining for Study identification in Systematic Reviews: a Systematic Review of Current Approaches. *Cochrane Methods*, 2015.
79. 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.
80. Byron C. Wallace and Thomas A. Trikalinos. Invited Response: Applications of text mining within systematic reviews. *Cochrane Methods*, 2012.

Selected invited talks & panels

Note that this is non-exhaustive and that these are in addition to the talks accompanying the conference publications above. I do not include job talks.

1. Keynote: Making Sense of the Biomedical Literature via ML and NLP. *The Joint Workshop on Bibliometric-enhanced IR and NLP for Digital Libraries at SIGIR*, 7/12/2018, 2018.
2. Training Neural NLP Models in Minimally Supervised Settings. *New England Machine Learning Day (NEML)*, 5/7/2018, 2018.
3. NLP in literature mining: Semi-automated quality assessment. *IBTN Workshop on Intelligent Interventions*, 5/26/2018, 2018.
4. Making Sense of the Biomedical Literature via Machine Learning and Natural Language Processing. *Open Data Science Conference*, 5/4/2018, 2018.
5. *Invited panel member*: Explainable AI models for health informatics (other panelists: David Sontag, Ziad Obermeyer, Anil Jain). *Northeast Computational Health Summit (NECHS)*, 4/27/2018, 2018.
6. 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.
7. Expediting Clinical Evidence Synthesis via Machine Learning, Natural Language Processing and Crowdsourcing. *Weill Cornell Medicine, Machine Learning in Medicine series*, 07/14/2017, 2017.
8. Expediting Clinical Evidence Synthesis via Machine Learning, Natural Language Processing and Crowdsourcing. *MIT (MEDG Reading Group)*, 06/02/2017, 2017.
9. *Invited panel member*: Data mining, clinical medicine, and medical research: The current state and future of the nexus (other panelists: Genevera Allen, Elmer V Bernstam, Chris Jermaine, Olivier Lichtarge, Srinivasan Parthasarathy, Jimeng Sun). *Statistics and Data Mining (SDM)*, 4/28/2017, 2017.
10. Expediting Clinical Evidence Synthesis via Machine Learning, Natural Language Processing and Crowdsourcing. *UMass, Worcester Medical School (BioNLP group)*, 01/05/2017, 2017.
11. 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.
12. Some recent work on Convolutional Neural Networks (CNNs) for text classification. *Raytheon BBN Technologies (BBN)*, Cambridge, MA, 11/22/2016, 2016.
13. Expediting Clinical Evidence Synthesis via Machine Learning, Natural Language Processing and Crowdsourcing. *Amazon Machine Learning Lecture Series*, Seattle, WA, 11/17/2016, 2016.
14. 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.
15. 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.
16. Open Challenges in Automating Clinical Evidence Synthesis. *International Conference on the Automation of Systematic Reviews*, Philadelphia, PA, 10/3/2016, 2016.
17. Some recent work on Convolutional Neural Networks (CNNs) for text classification. *MIT Computer Science and Artificial Intelligence Laboratory*, Cambridge, MA, 10/11/2016, 2016.
18. Some recent work on Convolutional Neural Networks (CNNs) for text classification. *University of Lisbon, Portugal*, 6/22/2016, 2016.
19. Automating Evidence Synthesis via Machine Learning and Natural Language Processing. *Machine Learning and Data Analytics Symposium (MLDAS)*, Doha, Qatar, 3/15/2016, 2016.

20. Automating evidence synthesis via machine learning and natural language processing. *Symposium on Text Mining, Bristol University, 11/11/2015, 2015.*
21. Automating evidence synthesis via machine learning and natural language processing. *Department of Biostatistics, Brown University, 10/19/2015, 2015.*
22. Automating evidence synthesis via machine learning and natural language processing. *Department of Computer and Information Science, University of Pennsylvania, 3/31/2015, 2015.*
23. Automating evidence synthesis. *Department of Biomedical Informatics, University of Texas at Houston, 10/1/2014, 2014.*
24. Can machine learning help us improve physician communication? *Department of Biomedical Informatics, Columbia University, NYC, 5/15/2014, 2014.*
25. Automating the systematic review: Tools and methods. *National Institute of Environmental Health Sciences, Durham, North Carolina, 4/30/2014, 2014.*
26. Machine Learning in Evidence-Based Medicine: Managing the Clinical Data Deluge. *University of Groningen, Netherlands, 4/24/2014, 2014.*
27. Can Machine Learning Help us Improve Physician Communication? *MIT Lincoln Lab, Lexington, MA, 11/1/2013, 2013.*
28. Can Machine Learning Help us Improve Physician Communication? *Data Management Lab at Boston University, Boston, MA, 11/8/2013, 2013.*
29. 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.*
30. Statistical Models of Patient-Doctor Communication. *Meaningful Use of Complex Medical Data (MUCMD), LA, CA, 8/16/2013, 2013.*
31. 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.*
32. 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.*
33. Better models, less effort: Active learning and dual supervision. *Department of Computer Science at UMass, Boston, 03/07/2012, 2012.*

Research Support (Grants)

A few notes on the below

- *I **bold** the titles of grants on which I am Principal Investigator (PI) or Co-/M- PI.*
- *Dollar amounts are best estimates of total award figures. For active awards, these are projections. In cases where grants involve multiple PIs (Co-PIs or M-PIs), I note my estimated portion parenthetically.*

Grant Title **Neural Models for Text: Improving Efficiency, Interpretability and Accuracy**
 Funder Army Research Office (ARO)
 Role Principal Investigator (PI)
 Period 5/15/2018 – 4/30/2021
 Amount ~415k (total); ~366k (Wallace)
 Collaborators Investigator Matt Lease UT Austin

Grant Title **CAREER: Structured Scientific Evidence Extraction: Models and Corpora (NSF 1750978)**
 Funder National Science Foundation (NSF)
 Role Principal Investigator (PI)
 Program CISE/III
 Period 7/1/2018 – 6/30/2023

Amount ~550k

Grant Title Measuring, Mining, and Understanding Communication Behaviors: Markers for Quality Healthcare (VA HX002289)

Funder VA Health Services Research and Development Service

PI Thomas Houston (UMass Medical School)

Role Investigator

Period 1/1/2018 – 1/1/2021

Amount ~\$1.1m (total); ~\$32k (Wallace)

Grant Title **Semi-Automating Data Extraction for Systematic Reviews (R01LM012086)**

Funder National Institutes of Health/National Library of Medicine

Role Project Director/Principal Investigator (Contact) (PD/PI)

Collaborators MPIs Iain J. Marshall (King's College), Thomas Trikalinos (Brown), Randolph Bias (UT Austin)

Period 9/20/2015 – 8/31/2019

Amount ~\$1.18M (total); ~\$625k (Wallace)

Grant Title **Crowdsourcing Mark-up of the Medical Literature to Support Evidence-Based Medicine and Develop Automated Annotation Capabilities (UH2CA203711)**

Funder National Institutes of Health/National Cancer Institute

Role Project Director/Principal Investigator (Contact) (PD/PI)

Collaborators MPIs Ani Nenkova and Zachary Ives (UPenn)

Program Big Data to Knowledge (BD2K)

Period 8/1/2016 – 9/29/2018

Amount ~\$240k (total); ~\$105k (Wallace)

Grant Title **Hybrid Approaches to Optimizing Evidence Synthesis via Machine Learning and Crowdsourcing (R03HS025024)**

Funder Agency for Healthcare Research and Quality (AHRQ)

Role Principal Investigator (PI)

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 ~\$100k

Grant Title **Neural Embeddings for Inferring Mental Health Status from Twitter Streams**

Funder Qntfy (<https://qntfy.com/>)

Role Principal Investigator (PI)

Period 5/2017 (this was an unstructured gift to support the work.)

Amount ~\$20k

Grant Title **Manual and Automatic Analysis of Patients' Values and Preferences Using Seton HCAHPS Surveys**

Funder Seton Medical Center/UT Center for Health and Social Policy (CHASP)

Role Co-Principal Investigator (Co-PI)

Collaborators Co-PIs Kenneth Fleischmann and Bo Xie (UT)

Period 8/2015 – 9/2016
Amount ~\$50k (total); ~\$10k (Wallace)

Grant Title Expression and Recognition of Irony in Multicultural Social Media (UTAP-EXPL/EEI-ESS/0031/2014)

Funder FCT

Role Investigator

PI Paula Carvalho (INESC-ID Lisboa & Universidade Europeia)

Collaborators Investigators Mário J. Silva and Bruno Martins (University of Lisbon)

Period 1/2015 – 1/2016

Amount ~\$35k (Wallace)

Grant Title Influencing Cervical Cancer Prevention and Detection Online through Social Media (R01CA178875)

Funder National Institutes of Health (NIH)/National Cancer Institute (NCI)

Role Investigator

PI MPIs Urmimala Sarkar (UCSF) and Damon Centola (UPenn)

Period 9/2014 – 9/2017

Amount ~\$1.3m (total); ~\$92k (Wallace)

Grant Title **Scaling Evidence-Based Medicine via Automation and Crowdsourcing**

Funder Brown University

Role Principal Investigator (PI)

Collaborators Tim Kraska, Ugur Cetintemel and Thomas Trikalinos (Brown)

Program Seed Funding (Internal to Brown)

Period 8/2014 – 8/2015

Amount ~\$80k

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

Funder Army Research Office (ARO)

Role Principal Investigator (PI)

Collaborators Investigators Eugene Charniak, Thomas Trikalinos and Laura Kertz (Brown)

Program Mathematical Sciences Division

Period 9/2013 – 9/2016

Amount ~\$300k

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 Co-Principal Investigator (Co-PI)

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

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

Period 5/1/2013 – 5/1/2016

Amount ~\$915k (total); ~\$500k (Wallace)

Other support I have twice received support from the Amazon AWS Cloud Credits for Research program. Once in 2014 (for \$2k) and again in 2016 (for \$4.5k)

Teaching

Northeastern

- Fall 2017 CS2500, Fundamentals I (Undergraduate)
Enrollment: 40
Mean overall instructor teaching effectiveness rating: 4.4/5.0
- Spring 2017 CS4100, Artificial Intelligence (Undergraduate)
Course website and syllabus: <http://www.byronwallace.com/ccis4100>
Mean overall instructor teaching effectiveness rating: 5.0/5.0
- Fall 2016 CS7180, Special Topics: Machine Learning in Health (Graduate)
Course website and syllabus: <http://www.byronwallace.com/ccis7180>
Mean overall instructor teaching effectiveness rating: 5.0/5.0

UT Austin

- Spring 2016 INF385T, Applied Data Mining (Advanced Undergraduate/Graduate)
Mean overall instructor rating: 4.9/5.0
- Fall 2015 INF385M, Database Management (Advanced Undergraduate/Graduate)
Mean overall instructor rating: 4.9/5.0
- Spring 2015 INF385T, Applied Data Mining (Advanced Undergraduate/Graduate)
Mean overall instructor rating: 4.2/5.0
- Fall 2014 INF385T, Applied Data Mining (Advanced Undergraduate/Graduate)
Mean overall instructor rating: 3.9/5.0

Advising & thesis committees

Advising

Current Northeastern CS PhD students

- 5/2016 – *present* Ben Nye
9/2016 – *present* Sarthak Jain
9/2016 – *present* David Lowell

Northeastern CS undergraduates working with me

- 1/2017 – *present* Derek Schuster
9/2017 – *present* Eric Lehman

CS PhD Students at UT Austin

9/2015 – *present* Ye Zhang (co-advised with Matthew Lease)
9/2014 – *present* An Than Nguyen (co-advised with Matthew Lease)

PhD committees

Current/ongoing

Student Jesse Anderton
University Northeastern
Thesis title *Scalable Ordinal Embedding to Model Text Similarity*
Advisor Javed Aslam
Other committee members Fernando Diaz (NYU), David Smith (Northeastern)

Student Maryam Aziz
University Northeastern
Thesis title *Pure Exploration for the Infinitely-Armed Bandit Models in Fixed-Confidence and Fixed-Budget settings*
Advisor Javed Aslam
Other committee members Emilie Kaufmann (INRIA), Jonathan Ulman (Northeastern)

Completed

Student David Inouye
Defense year 2017
University UT Austin
Thesis title *Appropriate, Accessible and Appealing Probabilistic Graphical Models*
Advisor Inderjit Dhillon (UT Austin)
Other committee members Pradeep Ravikumar (UT Austin), Raymond Mooney (UT Austin), Qixing Huang (UT Austin)

Student David Batista
Defense year 2016
Thesis title *Large-Scale Semantic Relationship Extraction for Information Discovery*
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)

Student Hyun Joon Jung
Defense year 2015
University UT Austin
Thesis title *Temporal Modeling of Crowd Work Quality for Quality Assurance in Crowdsourcing*
Advisor Matthew Lease
Other committee members Paul Bennett (Microsoft Research), Ken Fleischmann (UT Austin), Raymond Mooney (UT Austin)

Student Louise Millard
Defense year 2015
University University of Bristol

Thesis title *Towards data-intensive epidemiology: Explorations in systematic reviews and causal inference*

Co-Advisors Julian Higgins and Peter Flach

Other committee members James Thomas (UCL)

Service

Conference organization

MLHC Co-chair of the Machine Learning in Healthcare *Machine Learning in Healthcare* (MLHC) conference in 2016 & 2017. In 2017 (this past summer), we hosted the event here at Northeastern.

W3PHI Co-organizer of The International Workshop on the World Wide Web and Public Health Intelligence (W3PHI) – a workshop co-located with AAAI – 2016 and 2017. Currently co-chairing for 2018 as well.

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

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

Grant & proposal reviewing

Review panel service National Institutes of Health (NIH) in academic years 2015-2016, 2016-2017 and 2017-2018. National Science Foundation (NSF) in academic years 2013-2014 and 2014-2015.

Ad-hoc review I have reviewed proposals remotely for the National Science Foundation (NSF) in academic year 2018, and for the Army Research Office (ARO), mathematical sciences division in academic year 2014-2015.

Peer reviewing & etc.

Action editor Machine Learning Journal (MLJ); 2016-present.

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

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

Senior program committee or area chair I am serving as an area chair (AC) for NIPs 2018. I served as a senior program committee (SPC) member for *Association for the Advancement of Artificial Intelligence (AAAI)* in 2018 and 2014, and for *Knowledge Discovery and Databases (KDD)* in 2015. I was an area chair for *Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)* in 2016.

Program committees (active & past)

2018 International Conference Machine Learning (ICML); International Joint Conference on Artificial Intelligence (IJCAI); Neural Information Processing (NIPS); International Conference on Learning Representations (ICLR); Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)

2017 Association for the Advancement of Artificial Intelligence (AAAI); Association for Computational Linguistics (ACL); European Association for Computational Linguistics (EACL); International Conference Machine Learning (ICML); Neural Information Processing (NIPS); International Joint Conference on Artificial Intelligence (IJCAI); World Wide Web conference (WWW)

- 2016 Association for the Advancement of Artificial Intelligence (AAAI); Association for Computational Linguistics (ACL); Neural Information Processing (NIPS); International Conference on Machine Learning (ICML); Knowledge Discovery and Databases (KDD); AAAI Conference on Human Computation and Crowdsourcing (HCOMP); International Joint Conference on Artificial Intelligence (IJCAI)
- 2015 Association for the Advancement of Artificial Intelligence (AAAI); American Medical Informatics Association (AMIA); European Conference on Machine Learning (ECML); International Conference on Machine Learning (ICML); Neural Information Processing (NIPS)
- 2014 American Medical Informatics Association (AMIA); European Conference on Machine Learning (ECML); International Conference on Data Mining (ICDM); Knowledge Discovery and Databases (SIGKDD) Workshop on Health Informatics)
- 2013 Advancement of Artificial Intelligence (AAAI); American Medical Informatics Association (AMIA); IEEE Conference on Big Data (ICBD)
- 2012 International Conference on Information and Knowledge Management (CIKM); Knowledge Discovery and Databases (SIGKDD) Workshop on Health Informatics

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 Medical Internet Research (JMIR), Journal of Systematic Reviews, Knowledge and Information Systems (KAIS), Language Resources and Evaluation (LREC), Research Synthesis Methods (RSM)