

Mario Boley, PhD – Curriculum Vitae, October 2024

Senior lecturer at Department of Information Systems, University of Haifa
Senior lecturer (adjunct) at Department of Data Science and AI, Monash University, Melbourne
mboley@is.haifa.ac.il, mario.boleymonash.edu

Research Areas

Interpretable Machine Learning, Statistical Learning Theory, Discrete Optimisation, Materials Informatics

Education

PhD, Summa Cum Laude, Computer Science, University of Bonn, 2007–2011
Diplom with Distinction (combined BA and MA), Computer Science, University of Bonn, 2002–2007

Affiliation History

2024–today University of Haifa, Department of Information Systems, Senior Lecturer
2018–today Monash University, Melbourne, Department of Data Science and AI, Senior Lecturer
(tenured until Aug 2024, now adjunct)
2017–2018 Max Planck Institute for Informatics, Saarbrücken, Post-doctoral Fellow
2015–2017 Fritz Haber Institute of the Max Planck Society (Materials Science), Berlin, Post-doctoral Fellow
2012 Technion, Israel Institute of Technology, Haifa, Visiting Fellow (hosted by Assaf Schuster)
2011 University of Antwerp, Visiting Fellow (hosted by Bart Goethals)
2011–2015 University of Bonn, Department of Information Systems and AI, Post-doctoral Fellow
2007–2014 Fraunhofer Institute for Intelligent Analysis and Information Systems, St. Augustin, Research Scientist

Publication Statistics

55 peer reviewed publications in diverse venues (e.g., Theoretical Computer Science, NeurIPS, Nature Communications)
h-index 20, i10-index 31, total citations 1542

Awards

2018 Best-Paper-Award IEEE Int. Conference on Data Mining (ICDM)
2017 Highlights of 2017 Collection New Journal of Physics
2009 Distinguished-Paper-Award SIAM Int. Conference on Data Mining (SDM)
2008 Best-Student-Paper-Award IEEE Int. Conference on Data Mining (ICDM)
2007 Distinguished-Paper-Award Int. Workshop on Mining and Learning with Graphs (MLG)

Grants

2023–2026 *Optimal design of metering systems for intelligent water networks*, Australian Research Council (ARC) and South East Water Ltd (SEW), EUR 131,254, co-leader (one of two Chief Investigators, CIs)
2023 *Interpretable machine learning for reliable polymerization predictions across monomers*
Monash Data Futures Seed Grant, EUR 30,629, co-leader (one of two CIs)
2022–2026 *Advanced Manufacturing with 2D Materials*, Australian Research Council (ARC) IH210100025, EUR 2,700,000 contributor, AI Methods for Materials Engineering (one of eleven CIs)
2021–2023 *Rethinking the Data-driven Discovery of Rare Phenomena*, ARC DP210100045, EUR 279,360 leader (Principal Chief Investigator, PCI, of three CIs)
2020–2021 *Statistical Modelling of Vibration Sensors and Water Leaks*, SEW, EUR 50,000, leader (one of two CIs)
2018 *Machine learning for the design of next generation aluminium alloys*
Monash Infrastructure Interdisciplinary Seed Fund, EUR 29,285, co-leader (one of four CIs)
2012–2015 *Well-behaved pattern mining through sampling*, German Research Foundation, GA161512-1, EUR 500,000 proposal and project leader (delegate of CI)

Leadership Roles

2022–2024 Deputy Director of Research, Department of Data Science and AI, Monash University
2021–2024 Initiative Lead, Materials Informatics at Monash, Faculties of IT, Engineering, and Science
2018–2022 Graduate Research Coordinator, Faculty of IT, Monash University

Teaching Experience

- 2024 *Introduction to Programming and Computers*, Bachelor of Information Systems at University of Haifa
2022–2024 *Machine Learning*, Master of Data Science and Master of AI at Monash Uni.
2019–2021 *Introduction to Algorithms and Programming in Python*, Bachelor of Computer Science at Monash Uni.
first-year core unit with over 1,000 enrolments per offering
2017 *Subgroup Discovery*, Master of Computer Science at Saarland University
2013 *Online Learning*, Master of Computer Science at University of Bonn
2010–2011 *Pattern Discovery*, Master of Computer Science at University of Bonn

Senior Community Service

- 2023 General Chair Data Science and AI Summit at Monash
2022–2025 Senior PC AAAI Conference on Artificial Intelligence
2020–2024 Editorial Board Data Mining and Knowledge Discovery Journal
2020 Senior PC Europ. Conf. on ML and Principles and Practice of KDD (ECMLPKDD)

Program Committees

- 2019–2024 International Conference on Neural Information Processing systems (NeurIPS)
2019–2024 International Conference on Machine Learning (ICML)
2021–2022 SIAM International Conference on Data Mining (SDM)
2016–2019 Europ. Conf. on ML and Principles and Practice of Knowl. Discovery in Databases (ECMLPKDD)
2018 ACM International Conference on Information and Knowledge Management (CIKM)

Journal Reviewer

- 2020–2023 Machine Learning
2019–2022 Data Mining and Knowledge Discovery
2019 Statistical Analysis and Data Mining
2018 SIAM Journal on Discrete Mathematics

Funding Agency Reviewer

- 2023 Israeli Science Foundation, *Personal Research Grants*
2021–2022 Australian Research Council, *Discovery Project, Discovery Early Career Researcher Award*

Invited Plenary and Keynote Talks

- 2024 AI³ Symposium on Materials Theory, Driven by Aphrodite, Ab Initio Computations, and Artificial Intelligence, Paphos, Cyprus *From Prediction to Action: Critical Role of Performance Estimation for ML-Driven Discovery*
2022 IRIS Adlershof Workshop on Modeling Materials at Realistic Time Scales, Berlin
Discovering Exceptional Double Perovskites through Active and Reinforcement Learning
2022 VinAI (prominent AI research company in Vietnam) Research Seminar Series
Statistical Rule Learning for Materials Science
<https://www.youtube.com/watch?v=7sGAIRfNwnE>
2020 FAIR-DI Conference on a FAIR Data Infrastructure for Materials Genomics, Virtual
Identifying Domains of Applicability of Machine Learning Models for Materials Science
<https://www.youtube.com/watch?v=N0o26y8e3Gc>
2018 2nd Intern. Workshop on Formal Concept Analysis for Knowledge Discovery, Moscow
Keynote: *From Concept Enumeration to Constrained Optimization*
2018 NOMAD Summer School, Lausanne, *Subgroup Discovery for Materials Science*
<https://www.youtube.com/watch?v=4tCFpRbNLHo>
2017 NOMAD Summer School, Berlin, *Interpretable Modelling for Materials Science*
2016 IRISA INRIA Symposium on Instant and Interactive Data Mining, Rennes
Keynote: *From Case Studies to High-Throughput CTFs—Addressing the Evaluation Bottleneck in KDD Research*
2015 CECAM Workshop on Big Data of Materials Science—Critical Next Steps, Lausanne
Interpretable Local Modeling for Data-driven Science

Completed Research Students

- Yiwen Lu PhD, Monash University, 2024 (submitted). *Non-parametric Variable Selection for Interpretable Classification Models*
Shu Tew PhD, Monash University, 2024. *Bayesian Shrinkage Methods*

Yun Zhao *for Linear Regression*. Currently post-doctoral fellow at Monash University.
 PhD, Monash University, 2024. *Statistical Machine Learning Methods for Modelling, Imaging, and Monitoring the Brain*. Currently post-doctoral fellow at University of Sydney.
 Marzie Ghorbani PhD, Monash University, 2024. *Accelerated computational discovery and design of novel magnesium alloys by machine learning*. Currently post-doctoral fellow at Deakin University.
 Maurice Ntahobari Master by Research, Monash University, 2023. *Machine Learning for Epileptic Seizure Prediction*
 Panagiotis Mandros PhD, Saarland University, 2021. *Information-theoretic variable selection*.
 Currently post-doctoral fellow at Harvard University.
 Michael Kamp PhD, University of Bonn, 2019. *Model Aggregation and Federated Learning*.
 Currently independent research group leader at Ruhr University, Bochum.
 Sandy Moens PhD, Antwerp University, 2017. Co-supervised by me at Bonn University, *Randomised rule learning*.
 Currently innovation manager at ENSEK Benelux.

Current Research Students

Fan Yang PhD, Monash University, 2024 (planned). *Interpretable Machine Learning by Design*
 Neil Liu PhD, Monash University, 2025 (planned). *Spatial adaptivity and smoothness for sequential black-box optimisation problems*
 Simon Teshuva PhD, Monash University, 2025 (planned). *Symbolic Regression via Sparse Linear Model Identification*
 Jack Teng PhD, Monash University, 2025 (planned). *Physics-Informed Machine Learning*
 Rehan Mendis PhD, Monash University, 2025 (planned). *Optimisation and Statistical Modelling for Water Leak Detection*
 Shahrzad Bezhadi PhD, Monash University, 2026 (planned). *Interpretable Representation Learning for Rule Ensembles*

Languages

German mother tongue
 English native level
 Hebrew proficient

Journal Articles

1. T. See, D. Zhang, M. Boley, and D. Chalmers. Graph neural network based molecular property prediction with patch aggregation. *Journal of Chemical Theory and Computation*, 2024 (accepted)
2. S. Bauer, P. Benner, T. Bereau, V. Blum, M. Boley, C. Carbogno, R. Catlow, G. Dehm, S. Eibl, R. Ernstorfer, et al. Roadmap on data-centric materials science. *Modelling and Simulation in Materials Science and Engineering*, 32(6):063–301, 2024
3. M. Ghorbani, M. Boley, P. Nakashima, and N. Birbilis. An active machine learning approach for optimal design of magnesium alloys using bayesian optimisation. *Nature Scientific Reports*, 14(1):8299, 2024
4. M. Ghorbani, M. Boley, P. Nakashima, and N. Birbilis. A machine learning approach for accelerated design of magnesium alloys. Part A: Alloy data and property space. *Journal of Magnesium and Alloys*, 11(10):3620–3633, 2023
5. M. Ghorbani, M. Boley, P. Nakashima, and N. Birbilis. A machine learning approach for accelerated design of magnesium alloys. Part B: Regression and property prediction. *Journal of Magnesium and Alloys*, 11(11):4197–4205, 2023
6. Y. Lu, D. Yalcin, P. J. Pigram, L. D. Blackman, and M. Boley. Interpretable machine learning models for phase prediction in polymerization-induced self-assembly. *Journal of Chemical Information and Modeling*, 63:3288–3306, 2023
7. E. Van de Reydt, N. Marom, J. Saunderson, M. Boley, and T. Junkers. A predictive machine-learning model for propagation rate coefficients in radical polymerization. *Polymer Chemistry*, 14:1622–1629, 2023
8. Y. Zhao, F. Luong, S. Teshuva, A. Pelentritou, W. Woods, D. Liley, D. F. Schmidt, M. Boley, and L. Kuhlmann. Improved neurophysiological process imaging through optimisation of kalman filter initial conditions. *International Journal of Neural Systems*, 2023
9. Y. Zhao, M. Boley, A. Pelentritou, P. J. Karoly, D. R. Freestone, Y. Liu, S. Muthukumaraswamy, W. Woods, D. Liley, and L. Kuhlmann. Space-time resolved inference-based neurophysiological process imaging: Application to resting-state alpha rhythm. *NeuroImage*, 263:119592, 2022
10. H. Kulik, T. Hammerschmidt, J. Schmidt, S. Botti, M. A. Marques, M. Boley, M. Scheffler, M. Todorović, P. Rinke, C. Oses, et al. Roadmap on machine learning in electronic structure. *Electronic Structure*, 4(2):023004, 2022
11. C. Sutton, M. Boley, L. M. Ghiringhelli, M. Rupp, J. Vreeken, and M. Scheffler. Identifying domains of applicability of machine learning models for materials science. *Nature communications*, 11(1):1–9, 2020
12. P. Mandros, M. Boley, and J. Vreeken. Discovering dependencies with reliable mutual information. *Knowledge and Information Systems*, 62(11):4223–4253, 2020
13. M. Boley, B. R. Goldsmith, L. M. Ghiringhelli, and J. Vreeken. Identifying consistent statements about numerical data with dispersion-corrected subgroup discovery. *Data Mining and Knowledge Discovery*, 31(5):1391–1418, 2017
14. B. R. Goldsmith, M. Boley, J. Vreeken, M. Scheffler, and L. M. Ghiringhelli. Uncovering structure-property relationships of materials by subgroup discovery. *New Journal of Physics*, 19(1):013031, 2017
15. E. Spyropoulou, T. De Bie, and M. Boley. Interesting pattern mining in multi-relational data. *Data Mining and Knowledge Discovery*, 28(3):808–849, 2014
16. M. Boley, T. Horváth, A. Poigné, and S. Wrobel. Listing closed sets of strongly accessible set systems with applications to data mining. *Theoretical computer science*, 411(3):691–700, 2010
17. M. Boley and H. Grosskreutz. Approximating the number of frequent sets in dense data. *Knowledge and information systems*, 21(1):65–89, 2009
18. M. Boley, T. Horváth, and S. Wrobel. Efficient discovery of interesting patterns based on strong closedness. *Statistical Analysis and Data Mining*, 2(5-6):346–360, 2009

Rank-A* Conference Papers

19. F. Yang, P. Le Bodic, M. Kamp, and M. Boley. Corrective orthogonal boosting for simpler additive rule ensembles. In *Artificial Intelligence and Statistics (AISTATS)*, pages 1117–1125. PMLR, 2024
20. S. Yu-Tew, M. Boley, and D. F. Schmidt. Bayes beats cross validation: Fast and accurate ridge regression via expectation maximization. *Advances in Neural Information Processing Systems (NeurIPS)*, 37:19749–19768, 2023
21. H. Petzka, M. Kamp, L. Adilova, C. Sminchisescu, and M. Boley. Relative flatness and generalization. *Advances in Neural Information Processing Systems (NeurIPS)*, 34:18420–18432, 2021
22. P. Mandros, D. Kaltenpoth, M. Boley, and J. Vreeken. Discovering functional dependencies from mixed-type data. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1404–1414, 2020
23. J. Kalofolias, M. Boley, and J. Vreeken. Discovering robustly connected subgraphs with simple descriptions. In *2019 IEEE International Conference on Data Mining (ICDM)*, pages 1150–1155. IEEE, 2019
24. P. Mandros, M. Boley, and J. Vreeken. Discovering reliable correlations in categorical data. In *2019 IEEE International Conference on Data Mining (ICDM)*, pages 1252–1257. IEEE, 2019
25. P. Mandros, M. Boley, and J. Vreeken. Discovering reliable dependencies from data: Hardness and improved algorithms. In *18th IEEE International Conference on Data Mining (ICDM)*, pages 317–326. IEEE, 2018
26. M. Kamp, M. Boley, O. Missura, and T. Gärtner. Effective parallelisation for machine learning. *Advances in Neural Information Processing Systems (NeurIPS)*, 30:6477–6488, 2017
27. J. Kalofolias, M. Boley, and J. Vreeken. Efficiently discovering locally exceptional yet globally representative subgroups. In *17th IEEE International Conference on Data Mining (ICDM)*, pages 197–206. IEEE, 2017
28. P. Mandros, M. Boley, and J. Vreeken. Discovering reliable approximate functional dependencies. In *The 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 355–363. ACM, 2017
29. M. Boley, S. Moens, and T. Gärtner. Linear space direct pattern sampling using coupling from the past. In *The 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 69–77. ACM, 2012
30. M. Boley, C. Lucchese, D. Paurat, and T. Gärtner. Direct local pattern sampling by efficient two-step random procedures. In *The 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 582–590. ACM, 2011
31. M. Boley and H. Grosskreutz. A randomized approach for approximating the number of frequent sets. In *Eighth IEEE International Conference on Data Mining (ICDM)*, pages 43–52. IEEE, 2008

Rank-A Conference Papers

32. M. Boley, S. Teshuva, P. Le Bodic, and G. I. Webb. Better short than greedy: Interpretable models through optimal rule boosting. In *Proceedings of the 2021 SIAM International Conference on Data Mining (SDM)*, pages 351–359. SIAM, 2021
33. K. Budhathoki, M. Boley, and J. Vreeken. Discovering reliable causal rules. In *Proceedings of the 2021 SIAM International Conference on Data Mining (SDM)*, pages 1–9. SIAM, 2021
34. M. Kamp, S. Bothe, M. Boley, and M. Mock. Communication-efficient distributed online learning with kernels. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases, (ECML PKDD)*, volume 9852 of *LNCS*, pages 805–819. Springer, 2016
35. M. Kamp, M. Boley, and T. Gärtner. Beating human analysts in nowcasting corporate earnings by using publicly available stock price and correlation features. In *SIAM International Conference on Data Mining (SDM)*. SIAM, 2014
36. M. Kamp, M. Boley, D. Keren, A. Schuster, and I. Sharfman. Communication-efficient distributed online prediction by dynamic model synchronization. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD)*, volume 8724 of *LNCS*, pages 623–639. Springer, 2014
37. M. Kamp, C. Kopp, M. Mock, M. Boley, and M. May. Privacy-preserving mobility monitoring using sketches of stationary sensor readings. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD)*, volume 8190 of *LNCS*, pages 370–386. Springer, 2013

38. M. Boley, T. Gärtner, and H. Grosskreutz. Formal concept sampling for counting and threshold-free local pattern mining. In *SIAM International Conference on Data Mining (SDM)*, pages 177–188. SIAM, 2010
39. S. Vembu, T. Gärtner, and M. Boley. Probabilistic structured predictors. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 557–564. AUAI Press, 2009
40. M. Boley and H. Grosskreutz. Non-redundant subgroup discovery using a closure system. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD)*, 5781:179–194, 2009
41. M. Boley, T. Horváth, A. Poigné, and S. Wrobel. Efficient closed pattern mining in strongly accessible set systems. *11th European Conference on Knowledge Discovery in Databases (PKDD)*, 4702:382–389, 2007

Other Conference and Peer-reviewed Workshop Papers

42. Y. Zhao, D. Graydon, M. Boley, Y. Liu, P. J. Karoly, M. Cook, and L. Kuhlmann. Inference-based time-resolved chaos analysis of brain models: application to focal epilepsy. In *27th International Conference on Information Fusion (FUSION)*. IEEE, 2024
43. Y. Zhao, M. Boley, A. Pelentritou, W. Woods, D. Liley, and L. Kuhlmann. Inference-based time-resolved stability analysis of nonlinear whole-cortex modeling: application to xenon anaesthesia. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1–4. IEEE, 2023
44. M. Boley and A. Kariryaa. On the intuitiveness of common discretization methods. In *ACM SIGKDD Workshop on Interactive Data Exploration and Analytics (IDEA)*, 2016
45. M. Boley, M. Krause-Traudes, B. Kang, and B. Jacobs. Creed—scalable and repeatable extrinsic evaluation for pattern discovery systems by online user studies. In *ACM SIGKDD Workshop on Interactive Data Exploration and Analytics (IDEA)*, 2015
46. D. Trabold, M. Boley, M. Mock, and T. Horváth. In-stream frequent itemset mining with output proportional memory footprint. In *LWA 2015 Workshops: KDML, FGWM, IR, and FGDB*, volume 1458 of *CEUR Workshop Proceedings*, pages 93–104. CEUR-WS.org, 2015
47. S. Moens, M. Boley, and B. Goethals. Providing concise database covers instantly by recursive tile sampling. In *17th International Conference on Discovery Science (DS)*, volume 8777 of *LNCS*, pages 216–227. Springer, 2014
48. S. Moens and M. Boley. Instant exceptional model mining using weighted controlled pattern sampling. In *13th International Symposium on Intelligent Data Analysis (IDA)*, volume 8819 of *LNCS*, pages 203–214. Springer, 2014
49. M. Kamp, M. Boley, M. Mock, D. Keren, and A. Schuster. Adaptive communication bounds for distributed online learning. In *7th NIPS Workshop on Optimization for Machine Learning*, 2014
50. E. Spyropoulou, T. De Bie, and M. Boley. Mining interesting patterns in multi-relational data with n-ary relationships. In *16th International Conference on Discovery Science (DS)*, volume 8140 of *LNCS*, pages 217–232. Springer, 2013
51. M. Boley, M. Mampaey, B. Kang, P. Tokmakov, and S. Wrobel. One click mining: interactive local pattern discovery through implicit preference and performance learning. In *Proceedings of the ACM SIGKDD Workshop on Interactive Data Exploration and Analytics (IDEA)*, pages 27–35. ACM, 2013
52. H. Grosskreutz, M. Boley, and M. Krause-Traudes. Subgroup discovery for election analysis: a case study in descriptive data mining. In *13th International Conference on Discovery Science (DS)*, volume 6332 of *LNCS*, pages 57–71. Springer, 2010
53. M. Boley and T. Gärtner. On the complexity of constraint-based theory extraction. In *12th International Conference on Discovery Science (DS)*, volume 5808 of *LNCS*, pages 92–106. Springer, 2009
54. M. Boley. On approximating minimum infrequent and maximum frequent sets. In *10th International Conference on Discovery Science (DS)*, volume 4755 of *LNCS*, pages 68–77. Springer, 2007
55. M. Boley. Intelligent pattern mining via quick parameter evaluation. In *NSF Symp. on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation*, 2007

Other

56. M. Boley and S. Wrobel. Methods and apparatuses for iterative data mining, Mar. 19 2020. US Patent App. 16/691,690
57. M. Boley. *The Efficient Discovery of Interesting Closed Pattern Collections*. PhD thesis, Universitäts- und Landesbibliothek Bonn, 2011

Submitted

58. Z. Liu, M. Boley, F. Luong, and D. F. Schmidt. Improving random forests by smoothing. *AAAI Conference on Artificial Intelligence (AAAI, submitted)*, 2025

In Preparation

59. S. Dey, M. Kamp, R. P. Xian, M. Boley, J. Hatrick-Simpers, T. J. Jacobsson, and C. Sutton. Calibrating design choices for halide perovskite solar cells using interpretable machine learning. *Progress in Photovoltaics (targeted)*, 2024
60. S. Y. Tew, M. Boley, and D. Schmidt. Efficient adaptive horseshoe regression with grouping. *Computational and Graphical Statistics (targeted)*, 2024
61. T. See, D. Zhang, M. Boley, and D. Chalmers. Layer-to-layer knowledge mixing for graph neural network based quantum chemical property prediction. *Journal of Chemical Theory and Computation (targeted)*, 2024
62. Y. Zhao, N. Tsuchiya, M. Boley, Y. Liu, P. J. Karoly, A. Pelentritou, W. Woods, D. Liley, and L. Kuhlmann. Cortical excitability, connectivity and stability correlates of global consciousness levels. *Nature Communications (targeted)*, 2024