Mario Boley, PhD – Curriculum Vitae, 4 Aug 2025

PART A: CURRICULUM VITAE

1. PERSONAL DETAILS

Position Associate Professor at Department of Information Systems, University of Haifa
Senior Lecturer (adjunct) at Department of Data Science and AI, Monash University, Melbourne
Phone 04-8288364 (Office)
055-3145074 (Cellular)
Email `mboley@is.haifa.ac.il`, `mario.boley@monash.edu`
Webpage `https://marioboley.github.io/`

2. SUMMARY OF CONTRIBUTIONS TO SCIENCE

Interpretable Machine Learning Developed novel methods for learning rule-based models that are accurate and directly interpretable, eliminating the need for post hoc explanations typical of black-box approaches such as deep learning. These models enable transparent decision-making and are particularly valuable in scientific and high-stakes domains where trust, auditability, and insight into the reasoning process are essential. The underlying discrete optimisation algorithms combine mathematical order theory [C1, C3–4, E14–15], randomisation [C2, E1–3], and the branch-and-bound paradigm [C6, E5, E9, E13, E22, E23].

Statistical Learning Theory Developed methods that improve the reliability and comprehensibility of statistical models by focusing on the most informative variables, avoiding overfitting, and ensuring generalisation beyond the training data. This includes automatic variable selection based on information theory [C7, E4, E7, E10], reliable parameter estimation through Bayesian shrinkage methods [E12, E16], and novel observable proxy measures for model generalisation [E11, E6, E19]. These techniques contribute to the broader goal of building models that are not only accurate but also trustworthy and transparent.

Materials Informatics Established a new paradigm for interpretable machine learning in materials science [C5, C8, C9], enabling data-driven prediction and optimisation of material properties. This work supports the characterisation and rational design of materials with bespoke functionalities—critical for addressing grand challenges such as sustainable energy, environmental resilience, and advanced manufacturing. It has been applied to a wide range of materials including polymers [C12, C13, C19], magnesium alloys [C14, C15, C16], and recently also double perovskites [C17] and metal-organic frameworks.

3. HIGHER EDUCATION

2002–2007	Diplom with Distinction (combined BA and MA), Computer Science, University of Bonn
2007–2011	PhD, Summa Cum Laude, Computer Science, University of Bonn
2011–2015	Post-doctoral Studies, University of Bonn, Department of Information Systems and AI
2011	University of Antwerp, Visiting Fellow (hosted by Bart Goethals)
2012	Technion, Israel Institute of Technology, Haifa, Visiting Fellow (hosted by Assaf Schuster)
2015–2017	Post-doctoral Studies, Fritz Haber Institute of the Max Planck Society (Materials Science), Berlin
2017–2018	Post-doctoral Studies, Max Planck Institute for Informatics, Saarbrücken

4. ACADEMIC RANKS AND TENURE IN INSTITUTES OF HIGHER EDUCATION

2018–2022	Lecturer, Monash University, Melbourne, Department of Data Science and AI
2022–2024	Senior Lecturer, Monash University, Melbourne, Department of Data Science and AI
2024–today	Senior Lecturer (adjunct), Monash University, Melbourne, Department of Data Science and AI
2024–today	Associate Professor, University of Haifa, Department of Information Systems

5. OFFICES IN ACADEMIC ADMINISTRATION

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

6. SCHOLARLY POSITIONS AND ACTIVITIES OUTSIDE THE UNIVERSITY

Senior Community Service

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

Program Committees

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)
2019–2025	International Conference on Machine Learning (ICML)
2019–2025	International Conference on Neural Information Processing systems (NeurIPS)
2021–2022	SIAM International Conference on Data Mining (SDM)

Journal Reviewer

2018	SIAM Journal on Discrete Mathematics **JCR IF 2024: 1.0; Mathematics 129/483 (Q2)**
2019	Statistical Analysis and Data Mining **JCR IF 2024: 2.1; Computer Science, Artificial Intelligence 80/204 (Q2)**
2019–2025	Data Mining and Knowledge Discovery **JCR IF 2024: 4.3; Computer Science, Artificial Intelligence 67/204 (Q2)**
2020–2023	Machine Learning **JCR IF 2024: 2.9; Computer Science, Artificial Intelligence 101/204 (Q2)**

Funding Agency Reviewer

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

7. CONFERENCES / INVITED LECTURES / SEMINARS

Keynote Lectures

2016 IRISA INRIA Symposium on Instant and Interactive Data Mining, Rennes
From Case Studies to High-Throughput CTFs—Addressing the Evaluation Bottleneck in KDD Research

2018 2nd Intern. Workshop on Formal Concept Analysis for Knowledge Discovery, Moscow
From Concept Enumeration to Constrained Optimization

Other Invited Conference Talks

2015 CECAM Workshop on Big Data of Materials Science—Critical Next Steps, Lausanne
Interpretable Local Modeling for Data-driven Science

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=NOo26y8e3Gc`

2022 IRIS Adlershof Workshop on Modeling Materials at Realistic Time Scales, Berlin
Discovering Exceptional Double Perovskites through Active and Reinforcement Learning

2024 AI^3 Symposium on Materials Theory, Driven by Aphrodite, Ab Initio Computations, and AI, Paphos, Cyprus
From Prediction to Action: Critical Role of Performance Estimation for ML-Driven Discovery

Invited Talks at Seminars and Summer Schools

2011 University of Antwerp, Department of Computer Science, *Local Patterns—Listing, Counting, and sampling*

2013 Technion, Haifa, Department of Computer Science, *Communication Efficient Online Prediction in Dynamic Distributed Environments*

2015 KU Leuven, Department of Computer Science, *Creedo—Scalable and Repeatable Extrinsic Evaluation for Pattern Discovery Systems*

2017 NOMAD Summer School, Berlin, *Interpretable Modelling for Materials Science*

2018 NOMAD Summer School, Lausanne, *Subgroup Discovery for Materials Science*

https://www.youtube.com/watch?v=4tCFpRbNLHo
2021 Monash University, Materials Science & Engineering Seminar
Machine Learning Models for Materials Science
2022 VinAI (prominent AI research company in Vietnam) Research Seminar Series
Statistical Rule Learning for Materials Science
https://www.youtube.com/watch?v=7sGAIRfNwnE

8. RESEARCH GRANTS

Years	Role	Other Researchers	Title	Funded by	Amount EUR	Pub.
2012–2015	Proposal Main Author, Co-PI	T. Gärtner S. Wrobel (Co-PIs)	Well-behaved pattern mining through sampling	German Research Foundation (GA161512-1)	500,000 (jointly)	C4, E2, E3, E29, E30, E35, E32, E33, F2
2018	Proposal Co-Author, Co-PI	N Birblis P Nakashima Geoff Webb (Co-PIs)	Machine learning for the design of next generation aluminium alloys	Monash Infrastruct. Interdisciplinary Seed Fund	29,285 (jointly)	C14, C15, C16
2020–2021	Proposal Main Author, PI	D. Taniar M. Kamp (CIs)	Statistical Modelling of Vibration Sensors and Water Leaks	South East Water Ltd (SEW)	50,000	
2021–2024	Proposal Main Author, PI	W Buntine D Schmidt L Kuhlmann (CIs)	Rethinking the Data-driven Discovery of Rare Phenomena	Australian Research Council (DP210100045)	279,360	C9, C10, C11, C13, C17, C20, E12, E13, G7
2022–2026	Proposal Co-Author, CI	M Majumder (PI) N Medhekar (CI) ten more CIs	Research Hub for Advanced Manufacturing with 2D Materials	Australian Research Council (IH210100025)	2,700,000 (131,254 share)	
2023	Proposal Co-Author Co-PI	T Junkers (Co-PI)	Interpretable machine learning for reliable polymerization predictions across monomers	Monash Data Futures Seed Grant	30,629 (15,315 share)	C12
2023–2026	Co-PI	G Tack (Co-PI)	Optimal design of metering systems for intelligent water networks	Australian Research Council and South East Water Ltd (SEW)	131,254 (jointly)	G4

9. SCHOLARSHIPS, AWARDS AND PRIZES

2007 Distinguished-Paper-Award — Int. Workshop on Mining and Learning with Graphs (MLG)
2008 Best-Student-Paper-Award — IEEE Int. Conference on Data Mining (ICDM)
2009 Distinguished-Paper-Award — SIAM Int. Conference on Data Mining (SDM)
one of seven papers selected out of 355 submitted papers (105 accepted)
2017 Highlights of 2017 Collection — New Journal of Physics
2018 Best-Paper-Award — IEEE Int. Conference on Data Mining (ICDM)

10. TEACHING

University Courses taught as Lecturer

2010–2011 *Pattern Discovery*, Master of Computer Science at University of Bonn
2013 *Online Learning*, Master of Computer Science at University of Bonn
2017 *Subgroup Discovery*, Master of Computer Science at Saarland University
2019–2021 *Introduction to Algorithms and Programming in Python*, Bachelor of Computer Science at Monash Uni.
first-year mandatory course with over 1,000 enrolments per offering
2022–2024 *Machine Learning*, Master of Data Science and Master of AI at Monash Uni.
2024–2025 *Introduction to Computers and Programming*, Bachelor of Information Systems at University of Haifa

	first-year mandatory course with around 150 enrolments per offering
2025	*Data Science Lab*, Bachelor of Information Systems and Bachelor of Data Science at University of Haifa
2026	*Machine Learning under Uncertainty*, Masters of Information Systems at University of Haifa (planned)

Gradudate Research Students, Completed

Sandy Moens	PhD, Antwerp University, 2017. Co-supervised at Bonn University, *Randomised rule learning.* Currently innovation manager at ENSEK Benelux.
Michael Kamp	PhD, University of Bonn, 2019. *Model Aggregation and Federated Learning.* Post-doctoral fellow at Monash University, 2019–2021 Currently Associate Professor for Machine Learning and Artificial at TU Dortmund.
Panagiotis Mandros	PhD, Saarland University, 2021. *Information-theoretic variable selection.* Currently post-doctoral fellow at Harvard University.
Maurice Ntahobari	Master by Research, Monash University, 2023. *Machine Learning for Epileptic Seizure Prediction*
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.
Yun Zhao	PhD, Monash University, 2024. *Statistical Machine Learning Methods for Modelling, Imaging, and Monitoring the Brain.* Currently post-doctoral fellow at University of Sydney.
Shu Tew	PhD, Monash University, 2024. *Bayesian Shrinkage Methods for Linear Regression.* Currently post-doctoral fellow at Monash University.
Yiwen Lu	PhD, Monash University, 2024. *Non-parametric Variable Selection for Interpretable Classification Models*
Fan Yang	PhD, Monash University, 2025. *Interpretable Machine Learning by Design*

Graduate Research Students, Ongoing

Neil Liu	PhD, Monash University, 2025 (planned). *Smoothing for Interpretable Modelling and Uncertainty Quantification*
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*

11. MISCELLANEOUS

Languages

German	mother tongue
English	native level
Hebrew	proficient

PART B: PUBLICATIONS

GENERAL NOTES AND STATISTICS

My publication record spans both computer science, particularly machine learning (ML) and knowledge discovery from databases (KDD), and the physical and chemical Sciences (particularly materials science). In computer science, especially in ML and KDD, top-tier peer-reviewed conferences (e.g., NeurIPS or KDD) are considered primary publication venues and are more competitive and influential than even the best journals in the field. In the physical and chemical sciences, the norm is publication in peer-reviewed journals.

While author order may be affected by specific circumstances, in both fields one tends to list:

- first the researchers who primarily carried out the work, typically PhD student(s),
- followed by their day-to-day supervisors, often post-doctoral fellow or junior faculty,
- and lastly main supervisors, often group or grant leaders.

These are general tendencies. For an accurate assessment, author contribution statements and title notes should be considered where present.

H-index	22 (Source: Google Scholar)
Citation count	1848 (Source: Google Scholar)
Publication count	58 (peer-reviewed)
ORCID	`https://orcid.org/0000-0002-0704-4968`
Google Scholar	`https://scholar.google.com/citations?user=0jlBueMAAAAJ&hl=en`

A. PH.D. DISSERTATION

Title	The Efficient Enumeration of Closed Pattern Collections
Date of submission	2010
Numer of pages	131
Language	English
Name of supervisor	Prof. Stefan Wrobel
University	University of Bonn
Publications	C1, C2, C3, E1, E14, E15, E17, E25, E26, E27

B. SCIENTIFIC BOOKS (REFEREED)

None

C. ARTICLES IN REFEREED JOURNALS

1. **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

   JCR IF 2019 [2009 N/A]: 1.396; Computer Science, Artificial Intelligence 105/137 (Q4), Computer Science, Interdisciplinary Application 83/109 (Q4)

2. **M. Boley** and H. Grosskreutz. Approximating the number of frequent sets in dense data. *Knowledge and information systems*, 21(1):65–89, 2009

   JCR IF 2009: 2.211; Computer Science, Artificial Intelligence 27/103 (Q2), Computer Science, Information Systems 24/116 (Q1)

3. **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

   JCR IF 2010: 0.838; Computer Science, Theory & Methods 58/97 (Q3)

4. 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

   JCR IF 2014: 1.987; Computer Science, Artificial Intelligence 41/123 (Q2); Computer Science, Information Systems 25/139 (Q1)

5. 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

   JCR IF 2017: 3.579; Physics, Multidisciplinary 11/78 (Q1)

6. **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

   JCR IF 2017: 2.481; Computer Science, Artificial Intelligence 40/132 (Q2); Computer Science, Information Systems 53/148 (Q2)

7. #P. Mandros, **M. Boley**, and J. Vreeken. Discovering dependencies with reliable mutual information. *Knowledge and Information Systems*, 62(11):4223–4253, 2020

   JCR IF 2020: 2.822; Computer Science, Information Systems 79/161 (Q2); Computer Science, Artificial Intelligence 65/139 (Q2)

8. 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

   JCR IF 2020: 14.92; Multidisciplinary Sciences 4/72 (Q1)

9. 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

   JCR IF 2023 [2022 N/A]: 2.9; Chemistry, Physical 100/178 (Q3); Materials Science, Multidisciplinary 221/439 (Q3); Physics, Condensed Matter 34/79 (Q2)

10. #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

    JCR IF 2022: 5.7; Neurosciences 50/272 (Q1); Neuroimaging 1/14 (Q1); Radiology, Nuclear Medicine & Medical Imaging 17/135 (Q1)

11. #Y. Zhao, F. Luong, #T. 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

    JCR IF 2023: 6.6; Computer Science, Artificial Intelligence 33/197 (Q1)

12. 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

    JCR IF 2023: 4.1; Polymer Science 27/95 (Q2)

13. #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

    JCR IF 2023: 5.7; Computer Science, Interdisciplinary Applications 28/170 (Q1); Chemistry, Medicinal 10/72 (Q1); Chemistry, Multidisciplinary 61/231 (Q2); Computer Science, Information Systems 34/250 (Q1)

14. #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

    JCR IF 2023: 15.8; Metallurgy & Metallurgical Engineering 1/90 (Q1)

15. #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

    JCR IF 2023: 15.8; Metallurgy & Metallurgical Engineering 1/90 (Q1)

16. #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

    JCR IF 2023: 3.8; Multidisciplinary Sciences 25/134 (Q1)

17. 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

    JCR IF 2023: 1.9; Materials Science, Multidisciplinary 302/439 (Q3); Physics, Applied 117/179 (Q3)

18. #T. J. See, D. Zhang, **M. Boley**, and D. K. Chalmers. Graph neural network-based molecular property prediction with patch aggregation. *Journal of Chemical Theory and Computation*, 20(20):8886–8896, 2024

    JCR IF 2023: 5.7; Chemistry, Physical 52/178 (Q2); Physics, Atomic, Molec. & Chemical 5/40 (Q1)

19. F. Herb, **M. Boley**, and W.-K. Fong. Machine learning outperforms humans in microplastic characterization and reveals human labelling errors in FTIR data. *Journal of Hazardous Materials*, 487:136989, 2025

    JCR IF 2023: 12.2; Engineering, Environmental: 4/81 (Q1), Environmental Sciences 12/358 (Q1)

20. #Y. Zhao, D. B. Grayden, **M. Boley**, Y. Liu, P. J. Karoly, M. J. Cook, and L. Kuhlmann. Cortical stability and chaos during focal seizures: insights from inference-based modeling. *Journal of Neural Engineering*, 2025 (accepted)

    JCR IF 2023: 3.7; Engineering, Biomedical: 46/123 (Q2), Neurosciences 98/310 (Q2)

D. ARTICLES OR CHAPTERS IN SCIENTIFIC BOOKS (REFEREED)

None

E. ARTICLES IN CONFERENCE PROCEEDINGS

Rank-A* Conference Papers

1. **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

2. **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

3. **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

4. #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

5. 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

6. #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

7. #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

8. #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

9. 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

10. #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

11. 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

12. #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

13. #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

Rank-A Conference Papers

14. **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

15. **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

16. 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

17. **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

18. [#]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

19. [#]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

20. [#]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

21. [#]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

22. 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

23. **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

24. [#]F. Yang, P. Le Bodic, and **M. Boley**. Gradient boosting versus mixed integer programming for sparse additive modeling. In *European Joint Conference on Machine Learning and Principles of Knowledge Discovery from Databases (ECMLPKDD)*, 2025 (accepted)

Other Conference and Peer-reviewed Workshop Papers

25. **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

26. **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

27. **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

28. 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

29. **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

30. 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

31. [#]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

32. #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

33. #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

34. 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

35. **M. Boley**, M. Krause-Traudes, B. Kang, and B. Jacobs. Creedo—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

36. **M. Boley** and A. Kariryaa. On the intuitiveness of common discretization methods. In *ACM SIGKDD Workshop on Interactive Data Exploration and Analytics (IDEA)*, 2016

37. #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

38. #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

F. OTHER SCIENTIFIC PUBLICATIONS

1. **M. Boley**. *The Efficient Discovery of Interesting Closed Pattern Collections*. PhD thesis, Universitäts-und Landesbibliothek Bonn, 2011

2. **M. Boley** and S. Wrobel. Methods and apparatuses for iterative data mining, Mar. 19 2020. US Patent App. 16/691,690

G. SUBMITTED FOR PUBLICATION

1. R. Zichao, C. Zihao, L. Felix, C. Saumil, L. H. T. Nhan, Z. Zhiling, G. Kevin, A. Abdullah, B. K. Huyen, I. Theo, N.-D. Tung, P. H. Hieu, L. Dung, S. Joachim, P. V. B. T., C. Jennifer, B. Christian, **M. Boley**, E. G. Laurent, and Y. Omar. AI-driven reticular synthesis of zeolitic imidazolate framework crystals. ***Revise & Resubmit at Nature Synthesis***

2. #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. ***Revise & Resubmit at Nature Communications Biology***

3. #T. Teshuva, D. F. Schmidt, and **M. Boley**. Computation-cost aware iterative correlation learning

4. #R. Mendis, D. Bergmann, **M. Boley**, J. Crook, and G. Tack. Modelling in-situ vibration sensor sensitivity in water networks

5. #S. Yu-Tew, **M. Boley**, and D. Schmidt. Efficient adaptive horseshoe regression with grouping

6. #S. Behzadimanesh, P. Le Bodic, and **M. Boley**. Interpretable representation learning for additive rule ensembles

7. #Z. Liu, **M. Boley**, F. Luong, and D. F. Schmidt. Improving random forests by smoothing

H. FUTURE RESEARCH PLANS

In the near to mid-term, my research focuses on advancing trustworthy machine learning methods and their applications through three interconnected directions:

Improving the accuracy of interpretable models A major challenge in trustworthy AI is closing the accuracy gap between black-box models—such as those in deep learning—and inherently intelligible models like logical rule ensembles. This gap stems largely from the computational complexity of learning discrete rule structures, which contrasts with the efficiency of continuous optimisation in black-box systems. Two of my ongoing PhD projects, aim to address this issue [G6, E24] through using continuous convex optimisation as sub-routine in discrete model fitting. In addition, I am currently preparing a comprehensive interdisciplinary review article [I1] that consolidates rule learning methods across machine learning, knowledge discovery, and statistics, aiming to unify perspectives and stimulate progress toward more scalable and expressive interpretable models.

Enhancing uncertainty quantification for decision support Reliable quantification of the uncertainty associated with model predictions is critical to act on them with confidence. For the currently most common machine learning modes it is hard to provide this quantification, because they lack an integrated approach to represent epistemic uncertainty, the uncertainty stemming from insufficient training data. To address this, I am collaborating with Daniel Schmidt, a world-renowned expert in Bayesian statistic from Monash University, to integrate the Bayesian approach into interpretable modelling. In one line of work, we are aiming to re-imagine interpretable generalised additive models in the Bayesian framework [G5]. In another, we start with Bayesian Gaussian process regression and aim to equip it with the interpretable semantics and flexibility of rule based models [G7]. Both aim to provide a coherent combination of uncertainty quantification and an interpretable model syntax—–laying the foundation for model-informed actions in sensitive domains.

Translating trustworthy machine learning into scientific discovery Even with intelligible models and calibrated uncertainty estimates, a key barrier to practical adoption remains: the lack of principled guidance on how to translate these outputs into effective decisions. In domains such as materials science, this is particularly evident in the choice of *acquisition strategies*—algorithms that use model outputs to determine which experiments to perform next. Their variety and complexity often hinder the reliable application of theoretical advances. I am addressing this currently through two representative studies on durable photovoltaics [I2] and metal organic frameworks [G1]. The first is conducted jointly with the group of Matthias Scheffler at Humboldt University, Berlin, a world leader in computational materials science. The second is conducted with the group of Omar Yaghi at University of California in Berkeley, the world leader and pioneer in the synthesis of metal organic frameworks. Both studies systematically examine how acquisition functions interact with uncertainty estimates to enable data-efficient discovery.

I. PUBLICATIONS IN PREPARATION

1. **M. Boley**. Gradient boosting, subgroup discovery, and closed itemsets: Fast additive rule ensemble learning for interpretable machine learning. *Nature Machine Intelligence*

2. F. Luong, **M. Boley**, D. Schmidt, L. Foppa, and M. Scheffler. AI-enabled property optimization: Critical role of uncertainty quantification and acquisition function