Annie En-Shiun Lee
| Curriculum Vitæ

✉ annie.lee@cs.toronto.edu   •   Citizenship:Canadian

Education and Training

Education

Doctor of Philosophy   **University of Waterloo**
Systems Design Engineering, Centre for Pattern Analysis and Machine Intelligence   *2009–2014*
Supervisors: Andrew K. C. Wong, Distinguished Professor Emeritus, and Daniel Stashuk, Professor

Masters of Mathematics   **University of Waterloo**
Computer Science, Bioinformatics and Algorithms Group   *2006–2008*
Supervisor: Ming Li, Canada Research Chair in Bioinformatics

Joint Honours Bachelor of Mathematics with Co-operative Education   **University of Waterloo**
Computer Science; Combinatorics & Optimization   *1999–2004*

MITACS Workshops: Foundation of Project Management I and II, Team Building Excellence, Proactive and Practical Communication

Selected Grants, Scholarships, and Awards

Fellowships, Prizes, Grants

- NSERC ENGAGE; Contextual Linking and Datasource Triggers; Fei Chiang; Data Science Lab; Computing and Software; McMaster University; 2018.
- NSERC ENGAGE; Domain Adaptation and Transfer Learning; Masoud Makrehchi; Sci Lab; Computer and Software Engineering; University of Ontario Institute of Technology; 2018.
- NSERC ENGAGE; Quality Metric and Visualization; Christopher Collins; Canada Research Chair in Linguistic Information Visualization; ViaLab; Computer Science, UOIT; 2017.
- Visiting Scientific Researcher; Statistical Inference, Learning, and Models for Big Data; Nancy Reid; Canada Research Chair in Statistical Theory and Applications; Fields Institute; 2015.
- MITACS Accelerate; Predicting Unbalanced Health Data; Yang Wang; Pattern Discovery Technologies; 2012-2013.

International, National, and Provincial Awards

- Summer Institute in Taiwan; NSERC and National Science Council in Taiwan; Tun-Wen Pai; Biomedical Informatics Lab; Ocean's University; Summer 2012.
- Post Graduate Scholarship, NSERC, 2009-2012.
- Ontario Graduate Scholarship (OGS), Ontario Ministry of Training, Colleges and Universities, 2007 and 2009.

Other Awards

- Canadian Association for University Continuing Education Program Award, Non-credit Programming over 48 hours, 2020.

Relevant Industry and Research Experiences

Masters of Applied Computing Assistant Professor (CLTA Teaching Stream) November 2020–Current
Department of Computer Science, University of Toronto Toronto, Ontario, Canada
○ Take leadership on enhancing the academic quality of the MScAC program through admissions and recruitment, evaluating company research project submissions, academic advising for MScAC students, and evaluating student's final research report submissions.

Natural Language Understanding Research Scientist October 2018–September 2020
Machine Learning and Development Team, Stradigi AI Montreal, Ontario, Canada
○ Explored transfer learning for client projects and adversarial neural architecture through evaluating insights from attending external conference and internal discussions with the team's six expert areas, adapting models from latest literature and code repositories, and conducting model optimizing and parameter tuning.
○ Produced the state-of-the-art results for transfer learning by combining multi-task meta-learning techniques with neural interpretability of neural language model thereby advancing the boundary for text classification.

Lead Machine Learning Research Scientist August 2015–September 2018
Data Science Research and Development Team, VerticalScope Inc. Toronto, Ontario, Canada
○ Executed the end-to-end sentiment analysis project by performing exploratory analysis, selecting appropriate machine learning algorithms from literature survey, designing model based on performance evaluation and iterative improvement to produce data pipeline and visualization dashboard.
○ Supervised and advised research engineers and students by initiating daily team scrums, providing constructive feedback, establishing goals and deadlines, assist in blockers, and advocate personal and career goals.

Statistical Visiting Scientific Researcher January –June 2015
Statistical Inference, Learning, and Models for Big Data, Fields Institute Toronto, Ontario, Canada
○ Co-authored statistical summary review of latest research development with international colleagues as a result of participating in statistical workshops given by world-leading experts in various big data subject areas.

Artificial Intelligence Research Associate March–September 2014
Computer Science and Engineering, Chinese University Of Hong Kong Hong Kong, China
○ Extended real-world applications of novel patented algorithm through strategic partnering with international influencers by guiding interdisciplinary collaboration, planning research goals and group objectives, monitoring project deadline and team progress, and reviewing publication quality.

Pattern Analysis & Machine Intelligence Research Assistant September 2009–February 2014
Systems Design Engineering, University of Waterloo Waterloo, Ontario, Canada
○ Led undergraduate, masters, and PhD students of the research team by guiding students in research projects that led to implementation, application, and publications of novel methods resulting in a Masters thesis.
○ Pioneered novel algorithm through the commercialization process by convincing the university to patent, creating value proposition, conducting market assessment, pitching to industry partners, applying for grant applications, and participating in entrepreneurship program.

Teaching Experiences

Machine Learning Instructor May 2018–Current
Applied Machine Learning and Lifecycle, York University Toronto, Ontario, Canada
○ Won 2020 National Program Award for the certificate program by improving upon student feedback, ensuring smooth transition with previous and preceding courses, maintaining program objectives and quality with program manager.
○ Managed more than a dozen student projects by motivating real-world application, breaking complex problem down into milestones, and monitoring progress to ensure project completion.

Machine Intelligence Teaching Assistant

Winter 2011, 2012, 2013

Systems Design Engineering, University of Waterloo *Waterloo, Ontario, Canada*

- Published conference paper on e-learning by initiating new pedagogy teaching technique through prompting interactive student mobile responses during hands-on tutorials and code demonstrations.
- Advised student course projects and fourth year design projects on using machine intelligence in real-world problem, including a Kaggle machine learning competition and a start-up company at DMZ/MaRS.

Algorithms and Data Structures Teaching Assistant

Spring 2010, 2011

Mechatronics Engineering, University of Waterloo *Waterloo, Ontario, Canada*

- Nominated for Distinguished Teaching Assistant by students and support staff by engaging students through face-to-face instruction especially over midterm and final exam review sessions.
- Co-authored textbook by developing course content and creating evaluation (i.e. assignments, midterms and finals); thereby earning an excellent teaching assistant ranking from the course professor.

Administrative Service Experiences

Data Mining Co-Chair

March 2016–October 2018

Broadening Participation of Data Mining, ACM SIGKDD 2017 *Halifax, Nova Scotia, Canada*

- Orchestrated the mentoring program by inviting academic mentors, communicating with team and chairs, planned multiple activities to successfully conduct speed introductions, mentoring round tables, and panelist session.
- Managed different aspects of the workshop organization, including reviewing scholarship applications, pursuing industry sponsorship and grants, and preparing partnership package.

Machine Learning Workshop Co-Organizer

February 2010–February 2012

Women in Machine Learning, NeurIPS 2010 *Vancouver, British Columbia, Canada*

- Accomplished largest conference attendance to date by driving the presence of under-represented minorities at an international conference through networking with alumni, word-of-mouth advertising, and posting to special groups.
- Launched volunteer system strategically by opening registration with preferred tasks, matching tasks with talents, and assigning activities based on skills.

Others

Reviewer/Program Committee: ACM SIG Computer Science Education 2013-2021, Women in Machine Learning Workshop Unworkshop co-hosted with ICML 2020, Workshop on Broadening Participation of Data Mining co-hosted with SIGKDD 2017, Women in Machine Learning Workshop co-hosted with NeurIPS 2010/2018

Supervisor: CSML1030 Machine Learning Capstone 2020 - Pattern Analysis on Biosequence, Fields Summer Undergraduate Research Program 2018 – Big Data Extraction, 2nd and 4th year design projects, undergraduate research assistant (5 times)

Organizer: Broadening Participation of Data Mining (Mentoring Co-Chair) co-hosted with ACM SIG-KDD 2017 in Halifax, Women in Machine Learning (Co-Organizer) co-hosted with NIPS 2010 in Vancouver, Mandarin Chinese Christian Fellowship (Chair)

Instructor: Learn Easy 2020 - Artificial Intelligence and Machine Learning for Kids/Tweens/Teens, National Learning Code 2017-2019 - Introduction to Artificial Intelligence and Machine Learning, Ladies Learning Code: Data Insights with Python

Speaker/Panellist: Windsor-Essex DevFest 2019, AI with the Best 2016 & 2018, Sentiment Symposium 2017, Canadian Applied and Industrial Mathematics Society Annual Meeting 2017, Toronto Machine Learning Summit 2017, Fields Industry Lunch Seminar 2017, Conference on Big Data and Information Analytics 2017, Ladies who Tech - Filling the Gap 2017

Language Proficiency: English (Native Proficiency), Mandarin (Professional Working Proficiency)

Publication List

Technical Reports

1. **En-Shiun Annie Lee**, Surangika Ranathuga, Ravi Shekhar, Marjana Skenduli, Mehreen Alam, and Rishemjit Kuar. A review of low-resource neural machine translation. *in progress*, 2021.

2. Mehrdad Valipour, **En-Shiun Annie Lee**, Jaime R Jamacaro, and Carolina Bessega. Unsupervised transfer learning via bert neuron selection. *arXiv preprint arXiv:1912.05308*, 2019.

3. Mennatallah El-Assady, Gabrielle C. Perez-Dias, Kyle Noon, Christopher Collins, and **En-Shiun Annie Lee**. Forumvis: Navigating and analyzing massive online communications through user-specific scoring. 2019.

4. Afsaneh Towhidi, Mahboubeh Ahmadalinezhad, **En-Shiun Annie Lee**, and Masoud Makrehchi. Unsupervised domain adaptaion for named entity recognition. 2019.

5. **En-Shiun Annie Lee**. Training of template-specific weighted energy function for sequence-to-structure alignment. Technical report, University of Waterloo, 2008. MS Thesis, University of Waterloo.

6. **En-Shiun Annie Lee**. The impact of foreign political relations on health policies in chinese medicine. Technical report, University of Toronto, 2006. History of Medicine Course Essay (Awarded the Pauline M.H. Mazumdar Prizes in the History of Medicine 2006).

Books and Book Chapters

7. **En-Shiun Annie Lee**, Peiyuan Zhou, and Andrew K. C. Wong. Pattern discovery and disentanglement for protein binding complex detection and aligned pattern cluster analysis. In *Bioinformatics*, page in progress. exon publications, 2021.

8. **En-Shiun Annie Lee**, Peiyuan Zhou, and Andrew K. C. Wong. Discovering, summarizing, and refining patterns representations of protein functional domains that are compressed, flexible, and robust. In *Bioinformatics*, page in progress. exon publications, 2021.

9. Alexander Wong, James R. Wallace, **En-Shiun Annie Lee**, Xiao Yu Wang, Victor Cheung, and Abhishek Kumar. *Data Structures and Algorithms in a Nutshell*. University of Waterloo, 2012.

10. Andrew K. C. Wong, Dennis Zhuang, Gary CL Li, and **En-Shiun Annie Lee**. Pattern discovery and recognition in sequences. In *Pattern Recognition, Machine Intelligence and Biometrics*, pages 29–59. Springer, 2011.

Peer Reviewed Conferences Papers

11. **En-Shiun Annie Lee**, Karthik Kuber, Hashmat Rohian, and Sean Woodhead. Creating a problem-based machine learning curriculum for the flipped online setting. In *ACM Special Interest Group on Computer Science Education*, 2021.

12. SoHyun Park, Afsaneh Fazly, **En-Shiun Annie Lee**, Brandon Seibel, Wenjie Zi, and Paul Cook. Classifying out-of-vocabulary terms in a domain-specific social media corpus. In *International Conference on Language Resources and Evaluation (LREC'16)*, pages 2971–2975, 2016.

13. Antonio Sze-To, Sanderz Fung, **En-Shiun Annie Lee**, and Andrew K. C. Wong. Predicting protein-protein interaction using co-occurring aligned pattern clusters. In *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 55–60, 2015.

14. **En-Shiun Annie Lee**, Man-Hon Wong Antonio Ho-Yin Sze-To, Kwong-Sak Leung, Terrence Chi-Kong Lau, and Andrew K. C. Wong. Discovering protein-dna binding cores by aligned pattern clustering. In *Bioinformatics and Biomedicine (BIBM)*.

15. **En-Shiun Annie Lee**, Sanderz Fung, Antonio Ho-Yin Sze-To, and Andrew K. C. Wong. Confirming biological significance of co-occurrence clusters of aligned pattern clusters. In *2013 IEEE International Conference on Bioinformatics and Biomedicine*, pages 422–427, 2013.

16. **En-Shiun Annie Lee** and Andrew K. C. Wong. Classifying proteins by amino acid variations of sequential patterns. In *Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics*, page 276. ACM, 2013.

17. **En-Shiun Annie Lee**, Dennis Zhuang, and Andrew K. C. Wong. Aligning discovered patterns from protein family sequences. In *Pattern Recognition in Bioinformatics*, pages 243–254. Springer, 2012.

18. **En-Shiun Annie Lee**. Applying activity theory of mobile learning to context-aware smartphones. In *World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, volume 2012, pages 1092–1098, 2012.

19. **En-Shiun Annie Lee** and Andrew K. C. Wong. Identifying protein binding functionality of protein family sequences by aligned pattern clusters. In *Bioinformatics and Biomedicine (BIBM), 2012 IEEE International Conference on*, pages 1–6. IEEE, 2012.

20. Andrew K. C. Wong, Dennis Zhuang, Gary CL Li, and **En-Shiun Annie Lee**. Discovery of non-induced patterns from sequences. In *IAPR International Conference on Pattern Recognition in Bioinformatics*, pages 149–160. Springer, 2010.

Peer Reviewed Workshop Papers...

21. **En-Shiun Annie Lee**, Richie Wenjie Zi, Afsaneh Fazly, Brandon Seibel, and Anderson De Andrade. Unsupervised aspect extraction from free-form conversations. *Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM 2017) in conjunction with SIGKDD 2017*, 2017.

22. **En-Shiun Annie Lee**, Fiona J Whelan, Dawn ME Bowdish, and Andrew K. C. Wong. Characterizing amino acid variations of scavenger receptors by class information gain. In *Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics*, page 818. ACM, 2013.

23. Sanderz Fung, **En-Shiun Annie Lee**, and Andrew K. C. Wong. Comparing two algorithms for clustering aligned pattern clusters. *Mining Data Semantics in Heterogeneous Information Networks Workshop (MDS'2013), in conjunction with SIGKDD 2013*, 2013.

24. Sanderz Fung, **En-Shiun Annie Lee**, and Andrew K. C. Wong. Revealing protein structures by co-occurrence clustering of aligned pattern clusters. In *Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics*, page 869. ACM, 2013.

25. Pei-Yuan Zhou, **En-Shiun Annie Lee**, and Andrew K.C. Wong. Regrouping of pattern clusters to reveal characteristics of distinct classes and related classes. In *2013 IEEE International Conference on Bioinformatics and Biomedicine*, pages 55–61, Shanghai, China, December 2013.

26. **En-Shiun Annie Lee**, Franky Kin-Wai Yeung, and Tzu-Yang Ben Yu. Variable categorization and modelling: A novel adversarial approach to mobile location-based advertising. In *AAAI Workshop on Intelligent Techniques for Web Personalization and Recommender Systems (ITWP 2012)*, pages 43–47, 2012.

27. **En-Shiun Annie Lee** and Andrew K. C. Wong. Synthesizing aligned random pattern digraphs from protein sequence patterns. In *Bioinformatics and Biomedicine Workshops (BIBMW), 2011 IEEE International Conference on*, pages 178–185. IEEE, 2011.

28. **En-Shiun Annie Lee** and Andrew K. C. Wong. Summarizing patterns by suffix tree matching. In *5th Canadian Student Conference on Biomedical Computing and Engineering*, page 80, May 2011.

Peer Reviewed Journal Papers

29. Pei-Yuan Zhou, **En-Shiun Annie Lee**, Antonio Ho-Yin Sze-To, and Andrew K.C. Wong. Revealing subtle functional subgroups in class a scavenger receptors by pattern discovery and disentanglement of aligned pattern clusters. *Proteome Science*, 14(2):254–263, 2017.

30. **En-Shiun Annie Lee**, Ho-Yin Antonio Sze-To, Man-Hon Wong, Kwong-Sak Leung, Terrence Chi-Kong Lau, and Andrew KC Wong. Discovering protein-dna binding cores by aligned pattern clustering. *IEEE/ACM transactions on computational biology and bioinformatics*, 14(2):254–263, 2015.

31. **En-Shiun Annie Lee**, Ho-Yin Sze-To, Andrew K. C. Wong, and Daniel Stashuk. Unsupervised pattern discovery in biosequences using aligned pattern clustering. *SM Journal of Bioinformatics and Proteomics*, 1(2):1008, 2016.

32. Beate Franke, Jean-François Plante, Ribana Roscher, **En-shiun Annie Lee**, Cathal Smyth, Armin Hatefi, Fuqi Chen, Einat Gil, Alexander Schwing, Alessandro Selvitella, Michael M. Hoffman, Roger Grosse, Dieter Hendricks, and Nancy Reid. Statistical inference, learning and models in big data. *International Statistical Review*, 84(3):371–389, 2016.

33. Antonio Sze-To, Sanderz Fung, **Lee, En-Shiun Annie**, and Andrew K. C. Wong. Prediction of protein–protein interaction via co-occurring aligned pattern clusters. *Methods*, 110:26–34, 2016.

34. **En-Shiun Annie Lee**, Fiona J. Whelan, Dawn M. E. Bowdish, and Andrew K. C. Wong. Partitioning and correlating subgroup characteristics from aligned pattern clusters. *Bioinformatics*, 32(16):2427–2434, 2016.

35. **En-Shiun Annie Lee**, Sanderz Fung, Ho-Yin Sze-To, and Andrew K. C. Wong. Discovering co-occurring patterns and their biological significance in protein families. *BMC Bioinformatics*, 15(12):S2, 2014.

36. Andrew K. C. Wong and **En-Shiun Annie Lee**. Aligning and clustering patterns with refinement to reveal protein functionality of sequences. *IEEE/ACM transactions on computational biology and bioinformatics*, 11(3):548–560, 2014.

37. **En-Shiun Annie Lee** and Andrew K. C. Wong. Ranking and compacting binding segments of protein families using aligned pattern clusters. *Proteome Science*, 11(Suppl 1):S8, 2013.

38. Andrew K. C. Wong, Dennis Zhuang, Gary CL Li, and **En-Shiun Anne Lee**. Discovery of delta closed patterns and noninduced patterns from sequences. *Knowledge and Data Engineering, IEEE Transactions on*, 24(8):1408–1421, 2012.

39. Andrew G. Woolley, **En-Shiun Annie Lee**, and Fuzhong Zhang. sgal: a computational method for finding surface exposed sites in proteins suitable for cys-mediated cross-linking. *Bioinformatics*, 22(24):3101–3102, 2006.

Patent

40. Andrew K. C. Wong and **En-Shiun Annie Lee**. Aligning and clustering sequence patterns to reveal classificatory functionality of sequences, April 2013. United States Patent Application 20160070854.