Lihong Li

Senior Principal Applied Scientist
Amazon.com, Inc.
Seattle, WA, USA

llh@amazon.com
lihongli.cs@gmail.com
https://lihongli.github.io

RESEARCH INTERESTS

My core research interest is in machine learning for interactive, agentic systems that maximizes a utility function by taking actions, which is in contrast to prediction-oriented machine learning like supervised learning. I am working on **large language models**, **reinforcement learning**, **contextual bandits**, and related areas. My work has been applied to industrial recommendation, Web search, advertising, and conversational systems at leading companies.

EDUCATION

01/2005 – 05/2009	Ph.D.	Computer Science, Rutgers University, USA
09/2002 – 07/2004	M.Sc.	Computing Science, University of Alberta, Canada
09/1998 – 07/2002	B.Eng.	Computer Science and Technology, Tsinghua University, China

RESEARCH & INDUSTRY EXPERIENCE

10/2020 – present	Senior Principal Applied Scientist at Amazon (Seattle, WA)
10/2017 – 10/2020	Research Scientist at Google (Kirkland, WA)
06/2012 – 10/2017	Senior, Principal, Senior Principal Researcher at Microsoft Research (Redmond, WA)
09/2010 – 06/2012	Research Scientist at Yahoo! Research (Santa Clara, CA)
06/2009 – 08/2010	Postdoctoral Scientist at Yahoo! Research (Santa Clara, CA)
06/2008 – 08/2008	Research Intern at AT&T Shannon Labs (Florham Park, NJ)
05/2007 – 08/2007	Research Intern at Yahoo! Research (New York, NY)
05/2006 – 08/2006	Engineering Intern at Google (New York, NY)
01/2005 – 05/2009	Graduate Research Assistant at the Rutgers University (New Brunswick, NJ)
09/2002 – 07/2004	Research Assistant at the University of Alberta (Edmonton, AB, Canada)

SELECTED AWARDS

2023	Seoul Test of Time Award
2011	Yahoo! Super Star Team Award (highest team achievement award in the company)
2011	Notable Paper Award, AISTATS
2011	Best Paper Award, WSDM
2008	Best Student Paper Award, ICML

TEACHING/ADVISING EXPERIENCE

2018 – 2020	Supervised research interns at Google
2013–2017	Supervised student interns at Microsoft Research Projects on reinforcement learning, multi-armed bandits, imitation learning and Web search
2010/2011	Supervised student interns at Yahoo! Labs Projects on anomaly detection in distributed file systems, large-scale prediction models in advertising, and news ranking
Spring 2009	Guest lecturer for a graduate-level course at the Rutgers University Taught the least-squares policy iteration (LSPI) algorithm in the course "Learning and Sequential Decision Making".
09/2007 – 12/2007	Co-organizer for a graduate seminar at the Rutgers University Compiled reading materials, arranged weekly meetings, and presented papers for "Planning in Learned Environments" (w/ Michael Littman).
05/2005 – 08/2005	Organizer for a graduate seminar at the Rutgers University Compiled reading materials, arranged weekly meetings, presented papers, and invited an external speaker for "Abstractions and Hierarchies for Learning and Planning".
09/2002 – 07/2004	Teaching Assistant at the University of Alberta Taught seminar sessions and graded homework for the undergraduate course on discrete mathematics: "Formal Systems and Logic in Computing Science".

PROFESSIONAL ACTIVITIES

- Conference Organization
  - Area Chair, Senior Area Chair, and/or Senior Program Committee Member
    - AAAI Conference on Artificial Intelligence (AAAI): 2017–2019
    - International Conference on Artificial Intelligence and Statistic (AISTATS): 2017, 2019
    - International Conferences on Learning Representations (ICLR): 2019, 2022–2025
    - International Conferences on Machine Learning (ICML): 2012–2017, 2019–2025
    - International Joint Conferences on Artificial Intelligence (IJCAI): 2011, 2016, 2017
    - Annual Conferences on Neural Information Processing Systems (NIPS/NeurIPS): 2014, 2017–2021, 2023–2025
    - Reinforcement Learning Conference (RLC): 2024
  - Workshop Co-chairs
    - Reinforcement Learning Competition (ICML/UAI/COLT'09 Workshop)
    - PASCAL2 Exploration & Exploitation Challenge (ICML'12 Workshop)
    - Large-Scale Online Learning and Decision-Making Workshop (Cumberland Lodge, 2012)
    - IEEE BigData Workshop (DC, USA, 2014)
    - WWW Workshop on Offline and Online Evaluation of Web-based Services (Florence, Italy, 2015)
    - SIAM Conference on Optimization — Algorithms for Reinforcement Learning Minisymposium (Vancouver, Canada, 2017)
    - AI Frontiers (San Jose, CA, USA, November 2017)
    - From "What If" to "What Next" (NIPS'17 Workshop)
    - Reinforcement Learning for Real Life (ICML 2018 & 2021 workshops)
    - Optimization Foundations of Reinforcement Learning (NeurIPS 2019 workshop)
    - Simons Institute Workshop on Deep Reinforcement Learning (September, 2019)
    - Offline Reinforcement Learning (NeurIPS 2020 workshop)
  - Workshop program committee member
    - Planning and Learning in A Priori Unknown or Dynamic Domains, IJCAI 2005
    - Abstraction in Reinforcement Learning, ICML/UAI/COLT 2009
    - Bayesian Optimization, Experimental Design and Bandits, NIPS, 2011
    - AdML: Online Advertising Workshop, ICML 2012

    - Bayesian Optimization & Decision Making, NIPS 2012
    - Exploration in Reinforcement Learning, ICML 2018
- Editorial Services
  - Acting Editor, Transaction on Machine Learning Research (TMLR), since 2022
  - Associate Editor, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), since 2019
  - Guest editor, Machine Learning Journal Special Issue on "Reinforcement Learning for Real Life", 2019
- Tutorials
  - "Offline Evaluation and Optimization for Interactive Systems: A Practical Guide", at the *8th International Conference on Web Search and Data Mining (WSDM)*, Shanghai, China, February, 2015.
  - "Neural Approaches to to Conversational AI", with Jianfeng Gao and Michel Galley, at the 56th Annual Meeting of the Association for Computational Linguistics (ACL), Melbourne, Australia, July, 2018.
  - "Neural Approaches to to Conversational AI", with Jianfeng Gao and Michel Galley, at the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), Ann Arbor, MI, USA, July, 2018.
  - "A Tutorial on Policy Gradient Methods", with Lin Xiao, at the SIAM Conference on Optimization, 2021.
- Referee for funding agencies
  - Natural Sciences and Engineering Research Council of Canada (NSERC)
  - United States-Israel Binational Science Foundation (BFS)
- Referee for journals
  - ACM Transactions on Intelligent Systems and Technology
  - ACM Transactions on Knowledge Discovery from Data
  - Advances in Complex Systems
  - Artificial Intelligence
  - Artificial Intelligence Communications
  - Computer Speech and Language
  - Data Mining and Knowledge Discovery
  - IEEE Journal of Selected Topics in Signal Processing
  - IEEE Transactions on Automatic Control
  - IEEE Transactions on Knowledge and Data Engineering
  - IEEE Transactions on Neural Networks
  - IEEE Transactions on Wireless Communications
  - Journal of Artificial Intelligence Research
  - Journal of Computer Science and Technology
  - Journal of Machine Learning Research
  - Journal of Selected Topics in Signal Processing
  - Machine Learning
  - Mathematics of Operations Research
  - Nature Machine Intelligence
  - Neural Computation
  - Neurocomputing
  - Statistical Science
- Referee for conferences (including services as area chair and senior program committee member):
  - AAAI (AAAI Conferences on Artificial Intelligence): 2006, 2008, 2010, 2016 (Demo), 2017 (SPC)
  - ACML (Asian Conferences on Machine Learning): 2018
  - AISTATS (International Conferences on Artificial Intelligence and Statistics): 2011, 2017 (SPC)
  - ALT (International Conferences on Algorithmic Learning Theory): 2015
  - COLT (Annual Conferences on Learning Theory): 2010, 2011, 2012, 2015
  - ECML (European Conferences on Machine Learning): 2009
  - KDD (ACM SIGKDD Conferences on Knowledge Discovery and Data Mining): 2012
  - ICML (International Conferences on Machine Learning): 2009–2011, 2012–2017 (AC)

  - IJCAI (International Joint Conferences on Artificial Intelligence): 2007, 2011 (SPC), 2015, 2016 (SPC)
  - NIPS (Annual Meetings on Neural Information Processing Systems): 2008–2013, 2014 (AC)
  - STOC (ACM Symposium on Theory of Computing): 2014
  - UAI (Annual Conferences on Uncertainty in Artificial Intelligence): 2010, 2012, 2016
  - UbiComp (International Conferences on Ubiquitous Computing): 2011
  - WSDM (ACM International Conferences on Web Search and Data Mining): 2012, 2013
  - WWW (International Conferences on World Wide Web): 2012
- Open source and dataset contributions
  - Vowpal Wabbit: an open source project started with John Langford and Alexander L. Strehl for fast online learning in large-scale prediction problems. [link]
  - Yahoo! Front Page Today Module User Click Log Dataset: the first large-scale real-life dataset that supports unbiased evaluation of multi-armed bandit algorithms (with help from Wei Chu). [link]
- Misc
  - Google AI: "Off-policy estimation for infinite-horizon reinforcement learning", April 17, 2020. [link]
  - Amazon Science: "Decisions, decisions: Lihong Li's Amazon Ads reinforcement learning research", Jan 19, 2022. [link]
  - Grainger Foundation Frontiers of Engineering Symposium of the National Academy of Engineering, Univ of Colorado, September, 2023.
  - Panel on "Bridge Theory and Practice" at NSF Workshop on Reinforcement Learning, Boston, MA, January, 2025.

INVITED TALKS

The talks are grouped into several clusters based on their topics; the actual contents vary over time.

- Large Language Model Post-training
  - Keynote at AI+Robotics Research Symposium, Austin, TX. March, 2025.
- Off-policy Learning and Counterfactual Evaluation
  - RecSys Workshop on Causality, Counterfactuals, Sequential Decision-Making & Reinforcement Learning, Seattle, WA, USA. September, 2022.
  - School of Electrical Engineering and Computer Science, Oregon State University. November, 2021.
  - AI for Economics Seminar. December, 2020.
  - Department of Computer Science, University of California, Los Angeles, CA, USA. October, 2020.
  - Department of Statistics and Data Science, Yale University, USA. September, 2020.
  - Department of Computing Science, University of Alberta, Edmonton, AB, Canada. October, 2019.
  - Department of Statistics, Purdue University, West Lafayette, IN, USA. April, 2019.
  - Department of Computer Science, University of Texas, Austin, TX, USA. September, 2018.
  - Graduate School of Business, Stanford University, CA, USA. May, 2017.
  - Oxford University, Oxford, UK. November, 2015.
  - Google DeepMind, London, UK. November, 2015.
  - AdTech LA Meetup, Santa Monica, CA, USA. October, 2015.
  - UW CSE MSR Summer Institute, Union, WA, USA. August, 2015.
  - INRIA SequeL, Lille, France. December, 2014.
  - Criteo, Paris, France. December, 2014.
  - Department of Computing Science, University of Alberta, Edmonton, AB, Canada. November, 2014.
  - KDD Workshop on User Engagement Optimization, New York, NY, USA. August, 2014.
  - AAAI Workshop on Sequential Decision-Making with Big Data, Quebéc City, QC, Canada. July, 2014.
  - Microsoft Research Latin American Faculty Summit, Viña del Mar, Chile. May, 2014.
  - IEEE Information Theory and Application (ITA) Workshop, San Diego, CA, USA. February, 2014.
  - Distinguished Faculty and Graduate Student Seminar, Department of Statistics, University of Michigan, Ann Arbor, MI, USA. February, 2014.
- Reinforcement Learning via an Optimization Lens

  - Google DeepMind, Edmonton, AB, Canada. October, 2019
  - Simons Institute Workshop on Emerging Challenges in Deep Learning. Berkeley, CA, USA. August, 2019.
  - Mathematics of Data and Decisions seminar, University of California, Davis, CA. October, 2018.
  - International Symposium on Mathematical Programming (ISMP), Bordeaux, France. July, 2018.
  - INFORMS International Conference, Taipei, Taiwan. June, 2018.
  - Machine Learning Theory Workshop, Peking University, Beijing, China. June, 2018.
  - Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ, USA. March, 2018.
  - Google Machine Learning Day, Beijing, China. March, 2018.
  - Department of Electrical Engineering, Stanford University, Palo Alto, CA, USA. February, 2018.
  - Google Brain, Montreal, QC, Canada. September, 2017.
  - New York University, New York, NY, USA. May, 2017.
  - Simons Institute, Berkeley, CA, USA. February, 2017.
- Reinforcement Learning for Conversational Systems
  - NIPS Workshop on "Wordplay: Reinforcement and Language Learning in Text-based Games", Montreal, QC, Canada. December, 2018.
  - Google Brain, Montreal, QC, Canada. September, 2017.
  - ICML Workshop on "Interactive Machine Learning and Semantic Information Retrieval", Sydney, AU. August, 2017.
  - Multidisciplinary Conference on Reinforcement Learning and Decision Making (RLDM), Ann Arbor, MI, USA. June, 2017.
  - Korea Advanced Institute of Science and Technology, Korea. June 2017.
  - Sungkyunkwan University, Suwon, Korea. June 2017.
  - ACML Workshop on Reinforcement Learning, Hamilton, NZ. November, 2016.
  - Global AI Conference, Shanghai, China. November, 2016.
- Overview of Reinforcement Learning
  - Foster School of Business, University of Washington, Seattle, USA. 2022.
  - AI for Everyone Workshop Series, Google Beijing, China. April, 2018.
  - Department of Computer Science and Technology, Tsinghua University, Beijing, China. April, 2018.
  - Algorithms for Reinforcement Learning Minisymposium, SIAM Conference on Optimization, Vancouver, BC, Canada. May, 2017.
- Machine Learning in the Bandit Setting: Algorithms, Evaluation, and Case Studies
  - KDD Workshop on Multi-Armed Bandits and Reinforcement Learning (co-presenter: Yi Liu), 2021.
  - Department of Computer Science, University of South California, Los Angeles, CA, USA. October, 2015.
  - Department of Computer Science, Purdue University, West Lafayette, IN, USA. April, 2014.
  - Joint Statistical Meetings (Statistics in Marketing Track), Montreal, QC, Canada. August, 2013.
  - Tenth National Symposium of Search Engine and Web Mining, Beijing, China. May 2012.
  - Microsoft Research Asia, Beijing, China. May 2012.
  - Department of Machine Intelligence, Peking University, Beijing, China. May 2012.
  - Department of Computer Science and Technology, Tsinghua University, Beijing, China. May 2012.
  - Department of Computer Science, University of California, Los Angeles, CA, USA. May 2012.
  - Department of Computer Science & Engineering, University of California, San Diego, CA, USA. May 2012.
  - Department of Computer Science, University of California, Irvine, CA, USA. May 2012.
  - Google Research, Mountain View, CA, USA. April 2012.
  - Microsoft Research, Redmond, WA, USA. April 2012.
  - Adobe Advanced Technology Labs, San Jose, CA, USA. April 2012.
  - Microsoft Research, Mountain View, CA, USA. April 2012.
  - Department of Computer Science, Virginia Tech, Blacksburg, VA, USA. February 2012.
  - Department of Computer Science, Johns Hopkins University, MD, USA. February 2012.
  - Technicolor Research Center, Palo Alto, CA, USA. February 2012.
  - Department of Computing Science, University of Alberta, Edmonton, AB, Canada. June 2011.
  - Industrial Affiliates Annual Conference, Department of Statistics, Stanford University, USA. May 2011.

With Deepak Agarwal and Bee-Chung Chen.
  - Microsoft Sillicon Valley Center, Mountain View, CA, USA. March 2011.
  - Artificial Intelligence Center, SRI International, Menlo Park, CA, USA. April 2010.
- Vowpal Wabbit for Extremely Fast Machine Learning
  - GraphLab Workshop on Big Learning, San Francisco, CA, USA. July, 2012.
  - First data mining meetup on large-scale machine learning algorithms, San Francisco, CA, USA. August 2011.
- A Unifying Framework for Computational Reinforcement Learning Theory
  - ICML Workshop on Planning and Acting with Uncertain Models, Bellevue, WA, USA. June 2011.
  - Department of Computing Science, University of Alberta, Edmonton, AB, Canada. June 2011.
  - Yahoo! Research, Sunnyvale, CA, USA. April 2009.
  - Google Research, New York, NY, USA. April 2009.
  - Yahoo! Research, New York, NY, USA. January 2009.
  - Reasoning and Learning Laboratory, McGill University, McGill, QC, Canada. May 2008.
  - DARPA Information Processing Technology meeting, Arlington, VA, USA. February 2008.
  - AT&T Shannon Labs, Florham Park, NJ, USA. January 2008.
- Sparse Online Learning via Truncated Gradient
  - Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA. November 2009.
  - eBay Research Labs, San Jose, CA, USA. April 2009.
  - Department of Information Analysis & Management, NEC Laboratories America, Cupertino, CA, USA. April 2009.
  - Text Analysis and Machine Learning Group, University of Ottawa, Ottawa, ON, Canada. May 2008.
- Others
  - Guest lecture on "batch reinforcement learning" at Stanford CS234 (Prof. E. Brunskill), March 2022.

PUBLICATIONS

Journal Papers

(J1) *L. Li*: A perspective on off-policy evaluation in reinforcement learning. *Frontiers of Computer Science*, 13(5):911–912, 2019. (Invited paper)

(J2) M. Dudík, D. Erhan, J. Langford, and *L. Li*: Doubly robust policy evaluation and optimization. In *Statistical Science*, 29(4):485–511, 2014.

(J3) J. Bian, B. Long, *L. Li*, T. Moon, A. Dong, and Y. Chang: Exploiting user preference for online learning in Web content optimization systems. In *ACM Transactions on Intelligent Systems and Technology*, 5(2), 2014.

(J4) T. Moon, W. Chu, *L. Li*, Z. Zheng, and Y. Chang: Refining recency search results with user click feedback. In *ACM Transactions on Information Systems*, 30(4), 2012.

(J5) J. Langford, *L. Li*, P. McAfee, and K. Papineni: Cloud control: Voluntary admission control for Intranet traffic management. In *Information Systems and e-Business Management*, 10(3):295–308, 2012.

(J6) *L. Li*, M.L. Littman, T.J. Walsh, and A.L. Strehl: Knows what it knows: A framework for self-aware learning. In *Machine Learning*, 82(3):399–443, 2011.

(J7) *L. Li* and M.L. Littman: Reducing reinforcement learning to KWIK online regression. In the *Annals of Mathematics and Artificial Intelligence*, 58(3–4):217–237, 2010.

(J8) J. Langford, *L. Li*, J. Wortman, and Y. Vorobeychik: Maintaining equilibria during exploration in sponsored search auctions. In *Algorithmica*, 58(4):990–1021, 2010.

(J9) A.L. Strehl, *L. Li*, and M.L. Littman: Reinforcement learning in finite MDPs: PAC analysis. In the *Journal of Machine Learning Research*, 10:2413–2444, 2009.

(J10) E. Brunskill, B.R. Leffler, *L. Li*, M.L. Littman, and N. Roy: Provably efficient learning with typed parametric models. In the *Journal of Machine Learning Research*, 10:1955–1988, 2009.

(J11) J. Langford, *L. Li*, and T. Zhang: Sparse online learning via truncated gradient. In the *Journal of Machine Learning Research*, 10:777–801, 2009.

(J12) T.J. Walsh, A. Nouri, *L. Li*, and M.L. Littman: Planning and learning in environments with delayed feedback. In the *Journal of Autonomous Agents and Multi-Agent Systems*, 18(1):83–105, 2009.

(J13) *L. Li*, V. Bulitko, and R. Greiner: Focus of attention in reinforcement learning. In the *Journal of Universal Computer Science*, 13(9):1246–1269, 2007.

Refereed Conference Papers

(C1) Z. Wei, W. Yao, Y. Liu, W. Zhang, Q. Lu, L. Qiu, C. Yu, P. Xu, C. Zhang, B. Yin, H. Yun, and *L. Li*: WebAgent-R1: Training web agents via end-to-end multi-turn reinforcement learning. In *the 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2025.

(C2) Q. Zhang, L. Qiu, I. Hong, Z. Xu, T. Liu, S. Li, R. Zhang, Z. Li, *L. Li*, B. Yin, C. Zhang, J. Chen, H. Jiang, and T. Zhao: Self-rewarding PPO: Aligning large language models with demonstrations only. In *the 2nd Conference on Language Modeling (COLM)*, 2025.

(C3) X. Chen, J. Hu, C. Jin, *L. Li*, and L. Wang: Understanding domain randomization for sim-to-real transfer. In *the 10th International Conference on Learning Representations (ICLR)*, 2022.

(C4) Z. Tang, Y. Duan, S. Zhu, S. Zhang, and *L. Li*: Estimating long-term effects from experimental data. In *the 16th ACM Conference on Recommender Systems (RecSys)*, Industry Track, 2022.

(C5) C. Xiao, Y. Wu, T. Lattimore, B. Dai, J. Mei, *L. Li*, Cs. Szepesvari, and D. Schuurmans: On the optimality of batch policy optimization algorithms. In *the 38th International Conference on Machine Learning (ICML)*, 2021.

(C6) J. Hu, X. Chen, C. Jin, *L. Li*, and L. Wang: Near-optimal representation learning for linear bandits and linear RL. In *the 38th International Conference on Machine Learning (ICML)*, 2021.

(C7) X. Chen, J. Hu, C. Jin, *L. Li*, and L. Wang: Efficient reinforcement learning in factored MDPs with application to constrained RL. In *the 9th International Conference on Learning Representations (ICLR)*, 2021.

(C8) W. Zhang, D. Zhou, *L. Li*, and Q. Gu: Neural Thompson sampling. In *the 9th International Conference on Learning Representations (ICLR)*, 2021.

(C9) A. Bennett, N. Kallus, *L. Li*, and A. Mousavi: Off-policy evaluation in infinite-horizon reinforcement learning with latent confounders. In *the 24th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2021.

(C10) J. Mei, C. Xiao, B. Dai, *L. Li*, Cs. Szepesvari, and D. Schuurmans: Escaping the gravitational pull of softmax. In *Advances in Neural Information Processing Systems 33 (NeurIPS)*, oral, 2020.

(C11) B. Dai, O. Nachum, Y. Chow, *L. Li*, Cs. Szepesvari, and D. Schuurmans: CoinDICE: Off-policy confidence interval estimation. In *Advances in Neural Information Processing Systems 33 (NeurIPS)*, spotlight, 2020.

(C12) M. Yang, O. Nachum, B. Dai, *L. Li*, D. Schuurmans: Off-policy evaluation via the regularized Lagrangian. In *Advances in Neural Information Processing Systems 33 (NeurIPS)*, 2020.

(C13) J. Wen, B. Dai, *L. Li*, and D. Schuurmans: Batch stationary distribution estimation. In *the 37th International Conference on Machine Learning (ICML)*, 2020.

(C14) D. Zhou, *L. Li*, and Q. Gu: Neural contextual bandits with UCB-based exploration. In *the 37th International Conference on Machine Learning (ICML)*, 2020.

(C15) B. Kveton, M. Zaheer, Cs. Szepesvari, *L. Li*, M. Ghavamzadeh, and C. Boutilier: Randomized exploration in generalized linear bandits. In *the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2020.

(C16) R. Zhang, B. Dai, *L. Li*, and D. Schuurmans: GenDICE: Generalized offline estimation of stationary values. In *the 8th International Conference on Learning Representations (ICLR)*, 2020.

(C17) Z. Tang, Y. Feng, *L. Li*, D. Zhou, and Q. Liu: Doubly robust bias reduction in infinite horizon off-policy estimation. In *the 8th International Conference on Learning Representations (ICLR)*, 2020.

(C18) A. Mousavi, *L. Li*, Q. Liu, and D. Zhou: Black-box off-policy estimation for infinite-horizon reinforcement learning. In *the 8th International Conference on Learning Representations (ICLR)*, 2020.

(C19) O. Nachum, Y. Chow, B. Dai, and *L. Li*: DualDICE: Behavior-agnostic estimation of discounted stationary distribution corrections. In *Advances in Neural Information Processing Systems 32 (NeurIPS)*, spotlight, 2019.

(C20) Y. Feng, *L. Li*, and Q. Liu: A kernel loss for solving the Bellman equation. In *Advances in Neural Information Processing Systems 32 (NeurIPS)*, 2019.

(C21) C. Dann, *L. Li*, W. Wei, and E. Brunskill: Policy certificates: Towards accountable reinforcement learning. In the *36th International Conference on Machine Learning (ICML)*, 2019.

(C22) H. Dong, J. Mao, T. Lin, C. Wang, *L. Li*, and D. Zhou: Neural logic machines. In the *7th International Conference on Learning Representations (ICLR)*, 2019.

(C23) Q. Liu, *L. Li*, Z. Tang, and D. Zhou: Breaking the curse of horizon: Infinite-horizon off-policy estimation. In *Advances in Neural Information Processing Systems 31 (NeurIPS)*, spotlight, 2018.

(C24) K.-S. Jun, *L. Li*, Y. Ma, and J. Zhu: Adversarial attacks on stochastic bandits. In *Advances in Neural Information Processing Systems 31 (NeurIPS)*, 2018.

(C25) Y. Ma, K.-S. Jun, *L. Li*, and J. Zhu: Data poisoning attacks in contextual bandits. In *the 9th Conference on Decision and Game Theory for Security (GameSec)*, 2018.

(C26) D. Tang, X. Li, J. Gao, C. Wang, *L. Li*, and T. Jebara: Subgoal discovery for hierarchical dialogue policy learning. In *the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2018.

(C27) B. Dai, A. Shaw, *L. Li*, L. Xiao, N. He, Z. Liu, J. Chen, and L. Song: SBEED: Convergent reinforcement learning with nonlinear function approximation. In *the 35th International Conference on Machine Learning (ICML)*, 2018.

(C28) Y. Chen, *L. Li*, and M. Wang: Scalable bilinear π learning using state and action features. In *the 35th International Conference on Machine Learning (ICML)*, 2018.

(C29) B. Dai, A. Shaw, N. He, *L. Li*, and L. Song: Boosting the actor with dual critic. In *the 6th International Conference on Learning Representations (ICLR)*, 2018.

(C30) Z. Lipton, X. Li, J. Gao, *L. Li*, F. Ahmed, and L. Deng: Efficient dialogue policy learning with BBQ-networks. In *the 32nd AAAI Conference on Artificial Intelligence (AAAI)*, 2018.

(C31) J. Chen, C. Wang, L. Xiao, J. He, *L. Li*, and L. Deng: Q-LDA: Uncovering latent patterns in text-based sequential decision processes. In *Advances in Neural Information Processing Systems 30 (NIPS)*, 2017.

(C32) B. Peng, X. Li, *L. Li*, J. Gao, A. Celikyilmaz, S. Lee, K.-F. Wong: Composite task-completion dialogue system via hierarchical deep reinforcement learning. In *the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2017.

(C33) *L. Li*, Y. Lu, and D. Zhou: Provably optimal algorithms for generalized linear contextual bandits. In *the 34th International Conference on Machine Learning (ICML)*, 2017.

(C34) S. Du, J. Chen, *L. Li*, L. Xiao, and D. Zhou: Stochastic variance reduction methods for policy evaluation. In *the 34th International Conference on Machine Learning (ICML)*, 2017.

(C35) B. Dhingra, *L. Li*, X. Li, J. Gao, Y.-N. Chen, F. Ahmed, and L. Deng: Towards end-to-end reinforcement learning of dialogue agents for information access. In *the 55th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2017.

(C36) E. Parisotto, A. Mohamed, R. Singh, *L. Li*, D. Zhou, and P. Kohli: Neuro-symbolic program synthesis. In *the 5th International Conference on Learning Representations (ICLR)*, 2017.

(C37) X. Li, Y.-N. Chen, *L. Li*, J. Gao, A. Çelikyilmaz: End-to-End task-completion neural dialogue systems. In *the 8th International Joint Conference on Natural Language Processing (IJCNLP)*, 2017.

(C38) T.K. Huang, *L. Li*, A. Vartanian, S. Amershi, and J. Zhu: Active learning with oracle epiphany. In Advances in Neural Information Processing Systems 29 (NIPS), 2016.

(C39) J. He, M. Ostendorf, X. He, J. Chen, J. Gao, *L. Li*, and L. Deng: Deep reinforcement learning with a combinatorial action space for predicting and tracking popular discussion threads. In *the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2016.

(C40) C.-Y. Liu and *L. Li*: On the Prior Sensitivity of Thompson Sampling. In *the 27th International Conference on Algorithmic Learning Theory (ALT)*, 2016.

(C41) J. He, J. Chen, X. He, J. Gao, *L. Li*, L. Deng, and M. Ostendorf: Deep reinforcement learning with a natural language action space. In *the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2016.

(C42) N. Jiang and *L. Li*: Doubly robust off-policy value evaluation for reinforcement learning. In *the 33rd International Conference on Machine Learning (ICML)*, 2016.

(C43) S. Agrawal, N. R. Devanur, and *L. Li*: An efficient algorithm for contextual bandits with knapsacks, and an extension to concave objectives. In *the 29th Annual Conference on Learning Theory (COLT)*, 2016.

(C44) M. Zoghi, T. Tunys, *L. Li*, D. Jose, J. Chen, C.-M. Chin, and M. de Rijke: Click-based hot fixes for underperforming torso queries. In *the 39th International ACM SIGIR Conference on Research and Development in*

Information Retrieval (SIGIR), 2016.

(C45) J. He, J. Chen, X. He, J. Gao, *L. Li*, L. Deng, and M. Ostendorf: Deep reinforcement learning with an unbounded action space. In *the International Conference on Learning Representations (ICLR), Workshop Track*, 2016.

(C46) *L. Li*, R. Munos, and Cs. Szepesvári: Toward minimax off-policy value estimation. In *the 18th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2015.

(C47) *L. Li*, S. Chen, J. Kleban, and A. Gupta: Counterfactual estimation and optimization of click metrics in search engines: A case study. In *the 24th International Conference on World Wide Web (WWW), Companion*, 2015.

(C48) *L. Li*, J. Kim, and I. Zitouni: Toward predicting the outcome of an A/B experiment for search relevance. In *the 8th International Conference on Web Search and Data Mining (WSDM)*, 2015.

(C49) *L. Li*, H. He, and J.D. Williams: Temporal supervised learning for inferring a dialog policy from example conversations. In the *IEEE Spoken Language Technology Workshop (SLT)*, 2014.

(C50) A. Agarwal, D. Hsu, S. Kale, J. Langford, *L. Li*, and R.E. Schapire: Taming the monster: A fast and simple algorithm for contextual bandits. In *the 31st International Conference on Machine Learning (ICML)*, 2014.

(C51) E. Brunskill and *L. Li*: PAC-inspired option discovery in lifelong reinforcement learning. In *the 31st International Conference on Machine Learning (ICML)*, 2014.

(C52) E. Brunskill and *L. Li*: Sample complexity of multi-task reinforcement learning. In *the 29th Conference on Uncertainty in Artificial Intelligence (UAI)*, 2013.

(C53) M. Dudík, D. Erhan, J. Langford, and *L. Li*: Sample-efficient nonstationary-policy evaluation for contextual bandits. In *the 28th Conference on Uncertainty in Artificial Intelligence (UAI)*, 2012.

(C54) *L. Li*, W. Chu, J. Langford, T. Moon, and X. Wang: An unbiased offline evaluation of contextual bandit algorithms with generalized linear models. In *Journal of Machine Learning Research - Workshop and Conference Proceedings 26: On-line Trading of Exploration and Exploitation 2*, 2012.

(C55) V. Navalpakkam, R. Kumar, *L. Li*, and D. Sivakumar: Attention and selection in online choice tasks. In *the 20th International Conference on User Modeling, Adaptation and Personalization (UMAP)*, 2012

(C56) H. Wang, A. Dong, *L. Li*, Y. Chang, and E. Gabrilovich: Joint relevance and freshness learning From clickthroughs for news search. In *the 21st International Conference on World Wide Web (WWW)*, 2012.

(C57) O. Chapelle and *L. Li*: An empirical evaluation of Thompson sampling. In *Advances in Neural Information Processing Systems 24 (NIPS)*, 2011.

(C58) M. Dudík, J. Langford, and *L. Li*: Doubly robust policy evaluation and learning. In *the 28th International Conference on Machine Learning (ICML)*, 2011.

(C59) W. Chu, M. Zinkevich, *L. Li*, A. Thomas, and B. Tseng: Unbiased online active learning in data streams. In *the 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2011.

(C60) D. Agarwal, *L. Li*, and A.J. Smola: Linear-time algorithms for propensity scores. In *the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2011.

(C61) A. Beygelzimer, J. Langford, *L. Li*, L. Reyzin, and R.E. Schapire: Contextual bandit algorithms with supervised learning guarantees. In *the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2011. **Co-winner of the Notable Paper Award.**

(C62) W. Chu, *L. Li*, L. Reyzin, and R. Schapire: Linear contextual bandit problems. In *the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2011.

(C63) *L. Li*, Wei Chu, John Langford, and Xuanhui Wang: Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms. In *the 4th ACM International Conference on Web Search and Data Mining (WSDM)*, 2011. **Winner of the Best Paper Award.**

(C64) A.L. Strehl, J. Langford, *L. Li*, and S. Kakade: Learning from logged implicit exploration data. In *Advances in Neural Information Processing Systems 23 (NIPS)*, spotlight, 2011.

(C65) M. Zinkevich, M. Weimer, A.J. Smola, and *L. Li*: Convergence rates of parallel online learning via stochastic gradient descent. In *Advances in Neural Information Processing Systems 23 (NIPS)*, 2011.

(C66) T. Moon, *L. Li*, W. Chu, C. Liao, Z. Zheng, and Y. Chang: Online learning for recency search ranking using real-time user feedback. In *the 19th ACM Conference on Information and Knowledge Management (CIKM)*, 2010.

(C67) *L. Li*, W. Chu, J. Langford, and R.E. Schapire: A contextual-bandit approach to personalized news article recommendation. In *the 19th International Conference on World Wide Web (WWW)*, 2010.

(C68) Y. Xie, Y. Zhang, and *L. Li*: Neuro-fuzzy reinforcement learning for adaptive intersection traffic signal control. In *the Annual Meeting of Transportation Research Board (TRB)*, 2010.

(C69) *L. Li*, J.D. Williams, and S. Balakrishnan: Reinforcement learning for spoken dialog management using least-squares policy iteration and fast feature selection. In *the 10th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, 2009.

(C70) C. Diuk, *L. Li*, and B.R. Leffler: The adaptive k-meteorologists problem and its application to structure learning and feature selection in reinforcement learning. In *the 26th International Conference on Machine Learning (ICML)*, 2009.

(C71) J. Asmuth, *L. Li*, M.L. Littman, A. Nouri, and D. Wingate: A Bayesian sampling approach to exploration in reinforcement learning. In *the 25th International Conference on Uncertainty in Artificial Intelligence (UAI)*, 2009.

(C72) *L. Li*, M.L. Littman and C.R. Mansley: Online exploration in least-squares policy iteration. In *the 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2009.

(C73) L. Langford, *L. Li*, and T. Zhang: Sparse online learning via truncated gradient. In *Advances in Neural Information Processing Systems 21 (NIPS)*, spotlight, 2009.

(C74) *L. Li*: A worst-case comparison between temporal difference and residual gradient. In *the 25th International Conference on Machine Learning (ICML)*, 2008.

(C75) *L. Li*, M.L. Littman, and T.J. Walsh: Knows what it knows: A framework for self-aware learning. In *the 25th International Conference on Machine Learning (ICML)*, 2008. **Best Student Paper Award. Google Student Award at the New York Academy of Sciences Symposium on Machine Learning, 2008.**

(C76) R. Parr, *L. Li*, G. Taylor, C. Painter-Wakefield, and M.L. Littman: An analysis of linear models, linear value function approximation, and feature selection for reinforcement learning. In *the 25th International Conference on Machine Learning (ICML)*, 2008.

(C77) E. Brunskill, B.R. Leffler, *L. Li*, M.L. Littman, and N. Roy: CORL: A continuous-state offset-dynamics reinforcement learner. In *the 24th Conference on Uncertainty in Artificial Intelligence (UAI)*, 2008.

(C78) *L. Li* and M.L. Littman: Efficient value-function approximation via online linear regression. In *the 10th International Symposium on Artificial Intelligence and Mathematics (AI&Math)*, 2008.

(C79) J. Wortman, Y. Vorobeychik, *L. Li*, and J. Langford: Maintaining equilibria during exploration in sponsored search auctions. In *the 3rd International Workshop on Internet and Network Economics (WINE)*, 2007.

(C80) T.J. Walsh, A. Nouri, *L. Li*, and M.L. Littman: Planning and learning in environments with delayed feedback. In *the 18th European Conference on Machine Learning (ECML)*, LNCS 4701, 2007.

(C81) R. Parr, C. Painter-Wakefield, *L. Li*, and M.L. Littman: Analyzing feature generation for value-function approximation. In *the 24th International Conference on Machine Learning (ICML)*, 2007.

(C82) A.L. Strehl, *L. Li*, E. Wiewiora, J. Langford, and M.L. Littman: PAC model-free reinforcement learning. In *the 23rd International Conference on Machine Learning (ICML)*, 2006. **Best Student Poster Award winner at the New York Academy of Sciences Symposium on Machine Learning, 2006.**

(C83) A.L. Strehl, *L. Li*, and M.L. Littman: Incremental model-based learners with formal learning-time guarantees. In *the 22nd Conference on Uncertainty in Artificial Intelligence (UAI)*, 2006.

(C84) *L. Li*, T.J. Walsh, and M.L. Littman: Towards a unified theory of state abstraction for MDPs. In *the 9th International Symposium on Artificial Intelligence and Mathematics (AI&Math)*, 2006.

(C85) *L. Li*, M.L. Littman: Lazy approximation for solving continuous finite-horizon MDPs. In *the 20th National Conference on Artificial Intelligence (AAAI)*, 2005.

(C86) *L. Li*, V. Bulitko, and R. Greiner: Batch reinforcement learning with state importance (extended abstract). In *the 15th European Conference on Machine Learning (ECML)*, LNCS 3201, 2004.

(C87) V. Bulitko, *L. Li*, R. Greiner, and I. Levner: Lookahead pathologies for single agent search (poster paper). In *the 18th International Joint Conference on Artificial Intelligence (IJCAI)*, 2003.

(C88) I. Levner, V. Bulitko, *L. Li*, G. Lee, and R. Greiner: Towards automated creation of image interpretation systems. In *the 16th Australian Joint Conference on Artificial Intelligence*, LNCS 2903, 2003.

(C89) *L. Li*, V. Bulitko, R. Greiner, and I. Levner: Improving an adaptive image interpretation system by leveraging. In *the 8th Australian and New Zealand Intelligent Information System Conference*, 2003.

Surveys, Books and Book Chapters

(B1) J. Gao, M. Galley, and *L. Li*: Neural approaches to Conversational AI. Foundations and Trends in Information Retrieval, 13(2–3):127–298, 2019. ISBN 978-1-68083-552-6.

(B2) K. Hofmann, *L. Li*, and F. Radlinski: Online Evaluation for Information Retrieval. Foundations and Trends in Information Retrieval, 10(1):1–107, 2016. ISBN 978-1-68083-163-4.

(B3) *L. Li*: Sample complexity bounds of exploration. In Marco Wiering and Martijn van Otterlo, editors, *Reinforcement Learning: State of the Art*, Springer Verlag, 2012. ISBN 978-3642276446.

(B4) M. Shao, *L. Li*, Z. Zheng, and C. He: Practical Programming in XML. *Tsinghua University Press*, Beijing, China, December, 2002. ISBN 7-900643-85-0.

Theses

(T1) *L. Li*: A unifying framework for computational reinforcement learning theory. *Doctoral dissertation*, Department of Computer Science, Rutgers University, New Brunswick, NJ, USA, May, 2009.

(T2) *L. Li*: Focus of attention in reinforcement learning. *MSc thesis*, Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada, July, 2004.

(T3) *L. Li*: Design and implementation of an agent communication module based on KQML. *Bachelor degree thesis*, Department of Computer Science and Technology, Tsinghua University, Beijing, China, June, 2002.

Other

(O1) Y. Li, A. Geramifard, *L. Li*, Cs. Szepesvári, and T. Wang: Guest editorial: special issue on reinforcement learning for real life. Machine Learning, 110(9):2291–2293, 2021.

(O2) Y. Liu and *L. Li*: A map of bandits for E-commerce. KDD workshop on multi-armed bandits and reinforcement learning (MARBLE), 2021.

(O3) X. Li, Z.C. Lipton, B. Dhingra, *L. Li*, J. Gao, Y.-N. Chen: A user simulator for task-completion dialogues. MSR technical report, December 2016.

(O4) E. Brunskill and *L. Li*: The online discovery problem and its application to lifelong reinforcement learning. CoRR abs/1506.03379, June 2015.

(O5) D. Yankov, P. Berkhin, and *L. Li*: Evaluation of explore-exploit policies in multi-result ranking systems. Microsoft Journal on Applied Research, volume 3, pages 54–60, 2015. Also available as Microsoft Research Technical Report #MSR-TR-2015-34, May 2015.

(O6) Z. Qin, V. Petricek, N. Karampatziakis, *L. Li*, and J. Langford: Efficient online bootstrapping for large scale learning. *NIPS Workshop on Big Data*, December, 2013. Also available as Microsoft Research Technical Report #MSR-TR-2013-132.

(O7) *L. Li* and O. Chapelle: Regret bounds for Thompson sampling (Open Problems). In the *Twenty-Fifth Annual Conference on Learning Theory (COLT)*, 2012

(O8) *L. Li* and M.L. Littman: Prioritized sweeping converges to the optimal value function. Technical report DCS-TR-631, Department of Computer Science, Rutgers University, May 2008.

(O9) A.L. Strehl, *L. Li*, and M.L. Littman: PAC reinforcement learning bounds for RTDP and Rand-RTDP. *AAAI technical report WS-06-11*, pages 50-56, July 2006.

(O10) *L. Li* and M.L. Littman: Lazy approximation: A new approach for solving continuous finite-horizon MDPs. Technical report DCS-TR-577, Department of Computer Science, Rutgers University, May 2005.

(O11) *L. Li*, V. Bulitko, and R. Greiner: Focus of attention in sequential decision making. AAAI technical report WS-04-08, pages 43-48, July 2004.