

Shangtong Zhang

St Catherine's College, Manor Road, Oxford, United Kingdom, OX1 3UJ
shangtong.zhang@cs.ox.ac.uk <https://shangtongzhang.github.io>

Research Interest	The goal of my research is to solve sequential decision making problems in a scalable and reliable way. Currently, I focus on off-policy and offline reinforcement learning as solution methods.	
Education	University of Oxford , United Kingdom Doctor of Philosophy in Computer Science Advisor: Prof. Shimon Whiteson	Oct. 2018 - July. 2022
	University of Alberta , Canada Master of Science in Computer Science, Advisor: Prof. Richard S. Sutton	Sept. 2016 - Aug. 2018
	Fudan University , China Bachelor of Science in Computing Science Advisor: Prof. Xiaoqing Zheng and Prof. Wenqiang Zhang	Sept. 2012 - Jun. 2016
Research Internships	Microsoft Research Montreal , Canada Collaboration: Remi Tachet des Combes, Romain Laroche, and Harm van Seijen	Jun. 2021 - Sept. 2021
	DeepMind London , United Kingdom Collaboration: AlphaStar team (Michael Mathieu, Oriol Vinyals, etc) Collaboration: Adam White and Hado van Hasselt	Feb. 2021 - Jun. 2021
	DeepDrive , Edmonton, Canada Collaboration: Hengshuai Yao	Sept. 2020 - Dec. 2020
	Microsoft Research Montreal , Canada Collaboration: Remi Tachet des Combes, Romain Laroche, and Harm van Seijen	Jun. 2020 - Aug. 2020
	Noah's Ark Lab, Huawei , Edmonton, Canada Collaboration: Hengshuai Yao	May. 2018 - Aug. 2018
Publications	<ol style="list-style-type: none"><u>Truncated Emphatic Temporal Difference Methods for Prediction and Control</u> Shangtong Zhang, Shimon Whiteson. Journal of Machine Learning Research (JMLR), 2022.<u>On the Chattering of SARSA with Linear Function Approximation</u> Shangtong Zhang, Remi Tachet des Combes, Romain Laroche. arXiv:2202.06828, 2022.<u>A Deeper Look at Discounting Mismatch in Actor-Critic Algorithms</u> Shangtong Zhang, Romain Laroche, Harm van Seijen, Shimon Whiteson, Remi Tachet des Combes. International Conference on Autonomous Agents and Multiagent Systems (AAMAS), 2022 Acceptance rate: 26% Oral Presentation	

4. Learning Expected Emphatic Traces for Deep RL
Ray Jiang, **Shangdong Zhang**, Veronica Chelu, Adam White, Hado van Hasselt.
AAAI Conference on Artificial Intelligence (**AAAI**), 2022.
Acceptance rate: 15%.
5. Global Optimality and Finite Sample Analysis of Softmax Off-Policy Actor Critic under State Distribution Mismatch
Shangdong Zhang, Remi Tachet des Combes[‡], Romain Laroche[‡].
arXiv:2111.02997, 2021.
Under review of Journal of Machine Learning Research (JMLR).
6. StarCraft II Unplugged: Large Scale Offline Reinforcement Learning
Michael Mathieu*, Sherjil Ozair*, Srivatsan Srinivasan, Caglar Gulcehre, **Shangdong Zhang**, Ray Jiang, Tom Le Paine, Konrad Zolna, Richard Powell, Julian Schrittwieser, David Choi, Petko Georgiev, Daniel Kenji Toyama, Aja Huang, Roman Ring, Igor Babuschkin, Timo Ewalds, Mahyar Bordbar, Sarah Henderson, Sergio Gomez Colmenarejo, Aaron van den Oord, Wojciech M. Czarnecki, Nando de Freitas, Oriol Vinyals.
Deep RL Workshop at NeurIPS, 2021
7. Breaking the Deadly Triad with a Target Network
Shangdong Zhang, Hengshuai Yao, Shimon Whiteson.
International Conference on Machine Learning (**ICML**), 2021.
Acceptance rate: 21.5%.
8. Average-Reward Off-Policy Policy Evaluation with Function Approximation
Shangdong Zhang*, Yi Wan*, Richard S. Sutton, Shimon Whiteson.
International Conference on Machine Learning (**ICML**), 2021.
Acceptance rate: 21.5%.
9. Mean-Variance Policy Iteration for Risk-Averse Reinforcement Learning
Shangdong Zhang, Bo Liu, Shimon Whiteson.
AAAI Conference on Artificial Intelligence (**AAAI**), 2021.
Acceptance rate: 21.4%.
10. Learning Retrospective Knowledge with Reverse Reinforcement Learning
Shangdong Zhang, Vivek Veeriah, Shimon Whiteson.
Conference on Neural Information Processing Systems (**NeurIPS**), 2020.
Acceptance rate: 20.1%.
11. GradientDICE: Rethinking Generalized Offline Estimation of Stationary Values
Shangdong Zhang, Bo Liu, Shimon Whiteson.
International Conference on Machine Learning (**ICML**), 2020.
Acceptance rate: 21.8%.
12. Provably Convergent Two-Timescale Off-Policy Actor-Critic with Function Approximation
Shangdong Zhang, Bo Liu, Hengshuai Yao, Shimon Whiteson.
International Conference on Machine Learning (**ICML**), 2020.
Acceptance rate: 21.8%.
13. Deep Residual Reinforcement Learning
Shangdong Zhang, Wendelin Boehmer, Shimon Whiteson.
International Conference on Autonomous Agents and Multiagent Systems (**AAMAS**), 2020.
Acceptance rate: 23%.
Best Paper Award.
14. Mega-Reward: Achieving Human-Level Play without Extrinsic Rewards
Yuhang Song, Jianyi Wang, Thomas Lukasiewicz, Zhenghua Xu, **Shangdong Zhang**, Andrzej Wojcicki, Mai Xu

- AAAI Conference on Artificial Intelligence (**AAAI**), 2020.
Acceptance rate: 20.6%.
15. DAC: The Double Actor-Critic Architecture for Learning Options
Shangtong Zhang, Shimon Whiteson.
Conference on Neural Information Processing Systems (**NeurIPS**), 2019.
Acceptance rate: 21.2%.
 16. Generalized Off-Policy Actor-Critic
Shangtong Zhang, Wendelin Boehmer, Shimon Whiteson.
Conference on Neural Information Processing Systems (**NeurIPS**), 2019.
Acceptance rate: 21.2%.
 17. Distributional Reinforcement Learning for Efficient Exploration
Borislav Mavrin, **Shangtong Zhang**, Hengshuai Yao, Linglong Kong,
Kaiwen Wu, Yaoliang Yu
International Conference on Machine Learning (**ICML**), 2019.
Acceptance rate: 22.6%.
A short version is accepted as an extended abstract at AAMAS 2019.
 18. ACE: An Actor Ensemble Algorithm for Continuous Control with Tree Search
Shangtong Zhang, Hao Chen, Hengshuai Yao.
AAAI Conference on Artificial Intelligence (**AAAI**), 2019.
Acceptance rate: 16.2%.
 19. QUOTA: The Quantile Option Architecture for Reinforcement Learning
Shangtong Zhang, Borislav Mavrin, Linglong Kong, Bo Liu, Hengshuai Yao.
AAAI Conference on Artificial Intelligence (**AAAI**), 2019.
Acceptance rate: 16.2%.
 20. MLPack 3: A Fast, Flexible Machine Learning Library
Ryan Curtin, Marcus Edel, Mikhail Lozhnikov, Yannis Mentekidis, Sumedh Ghaisas,
Shangtong Zhang
Journal of Open Source Software (**JOSS**), 2018.
 21. Crossprop: Learning Representations by Stochastic Meta-Gradient Descent
in Neural Networks
Vivek Veeriah*, **Shangtong Zhang***, Richard S. Sutton.
European Conference on Machine Learning and Principles and Practice of Knowledge
Discovery in Databases (**ECML-PKDD**), 2017.
Acceptance rate: 27.1%.
 22. A Deeper Look at Experience Replay
Shangtong Zhang, Richard S. Sutton.
Deep RL Symposium at NIPS, 2017.
 23. Comparing Deep Reinforcement Learning and Evolutionary Methods
in Continuous Control
Shangtong Zhang, Osmar R. Zaiane
Deep RL Symposium at NIPS, 2017.
 24. A Demon Control Architecture with Off-Policy Learning and Flexible Behavior
Policy
Shangtong Zhang, Richard S. Sutton.
Hierarchical RL Workshop at NIPS, 2017.
 25. A Deep Neural Network for Modeling Music
Pengjing Zhang, Xiaoqing Zheng, Wenqiang Zhang, Siyan Li, Sheng Qian,
Wenqi He, **Shangtong Zhang**, Ziyuan Wang
International Conference on Multimedia Retrieval (**ICMR**), 2015.
Acceptance rate: 31%.

*: Equal contribution
‡: Equal advising
† : My name does not appear in the ICML proceedings due to a mistake in submission.
See Acknowledgments, arXiv, or AAMAS proceedings for clarification.

Academic Services

Meta Reviewer & Area Chair
ACML 2022

Expert Reviewer
ICML 2021

Reviewer & Program Committee

Transactions on Pattern Analysis and Machine Intelligence (1)
Transaction of Machine Learning Research (2)
Journal of Machine Learning Research (1)
Artificial Intelligence Journal (with green open access) (2)
AISTATS 2022
NeurIPS 2020, 2021, 2022
ICML 2020, 2022
AAAI 2020, 2021, 2022, 2023
ICLR 2021, 2022, 2023
SIGCOMM 2022
Offline Reinforcement Learning Workshop at NeurIPS 2020, 2021
Deep Reinforcement Learning Workshop at NeurIPS 2019, 2020, 2021, 2022
Adaptive and Learning Agents Workshop at AAMAS 2019, 2020
Optimization Foundations for Reinforcement Learning Workshop at NeurIPS 2019
Reinforcement Learning for Real Life Workshop at ICML 2019, 2021, NeurIPS 2022

Honours

<i>EPSRC studentship</i> , University of Oxford	2018 - 2022
<i>AAMAS Student Scholarship</i>	2022
<i>ICLR Outstanding Reviewer</i>	2021
<i>NeurIPS Reviewer Award</i>	2020
<i>ICML Reviewer Award</i>	2020
<i>Light Senior Scholarship</i> , St Catherine's College, University of Oxford	2020
<i>AAMAS Travel Award</i>	2020
<i>AAMAS Best Paper Award</i>	2020
<i>NeurIPS Optimization Foundations for RL Workshop Travel Award</i>	2019
<i>NeurIPS Travel Award</i>	2019
<i>AAAI Travel Award</i>	2019
<i>NIPS Hierarchical RL Workshop Travel Award</i>	2017
<i>Second Class Scholarship</i> , Fudan University	2015
<i>EMC Scholarship</i> , Fudan University	2014

Invited Talks

<i>Breaking the Deadly Triad in Off-Policy Reinforcement Learning</i>	
School of Informatics, University of Edinburgh	2021
School of Computing Science, Simon Fraser University	2022
Department of Electrical & Computer Engineering, University of Waterloo	2022
Department of Computer Science, University of Virginia	2022
<i>Breaking the Deadly Triad in Reinforcement Learning</i>	2021
RL team, DeepMind	

	<i>Breaking the Deadly Triad with a Target Network</i> Microsoft Research Summit	2021
	<i>Off-Policy Evaluation</i> Data Fest 2020, Open Data Science	2020
	<i>Off-Policy Evaluation and Control</i> ByteDance AI Lab, Shanghai	2020
	<i>Off-Policy Actor-Critic Algorithms</i> Latent Logic LTD, Oxford	2019
	<i>Generalized Off-Policy Actor-Critic</i> Noah's Ark Lab, Huawei, Edmonton	2019
	<i>Exploration with Quantile Options</i> Huawei RL Workshop, Edmonton	2018
	<i>Coding Deep RL Papers</i> NIPS MLTrain Workshop, Long Beach	2017
Teaching	<i>University of Oxford</i> , Teaching Assistant AIMS CDT Lectures	Michaelmas 2019
	<i>University of Alberta</i> , Teaching Assistant CMPUT 229 Computer Organization and Architecture	Fall 2016
Code	<i>PyTorch Deep RL</i> A zoo of popular deep RL algorithms in PyTorch with 2.5k stars in Github.	
	<i>Reinforcement Learning: An Introduction</i> Python implementation of the book <i>Reinforcement Learning: An Introduction</i> with 10.6k stars in Github.	
	<i>Google Summer of Code (GSoC) 2017</i> Contributed to MLPack by implementing a deep RL framework.	
	<i>Google Summer of Code (GSoC) 2014</i> Contributed to Xapian by optimizing the post list and the position list.	