

## Shangtong Zhang

---

422 Rice Hall, 85 Engineer's Way, Charlottesville, VA, 22903, United States  
[shangtong@virginia.edu](mailto:shangtong@virginia.edu)    <https://shangtongzhang.github.io>

<b>Research Interest</b>	The goal of my research is to solve sequential decision making problems in a scalable and reliable way. Currently, I focus on reinforcement learning as a solution method.	
<b>Academic Employments</b>	<b>Assistant Professor</b> Department of Computer Science University of Virginia, United States	Aug. 2022 - Present
<b>Education</b>	<b>University of Oxford</b> , United Kingdom Doctor of Philosophy in Computer Science Advisor: Prof. Shimon Whiteson	Oct. 2018 - July. 2022
	<b>University of Alberta</b> , Canada Master of Science in Computer Science, Advisor: Prof. Richard S. Sutton	Sept. 2016 - Aug. 2018
	<b>Fudan University</b> , China Bachelor of Science in Computing Science	Sept. 2012 - Jun. 2016
<b>Publications</b>	<ol style="list-style-type: none"><li>1. <u>Improving Monte Carlo Evaluation with Offline Data</u> Shuze Liu<sup>[advisee]</sup>, <b>Shangtong Zhang</b> arXiv:2301.13734, 2023.</li><li>2. <u>On the Convergence of SARSA with Linear Function Approximation</u> <b>Shangtong Zhang</b>, Remi Tachet des Combes, Romain Laroche. International Conference on Machine Learning (<b>ICML</b>), 2023. Acceptance rate: 28%</li><li>3. <u>Global Optimality and Finite Sample Analysis of Softmax Off-Policy Actor Critic under State Distribution Mismatch</u> <b>Shangtong Zhang</b>, Remi Tachet des Combes<sup>†</sup>, Romain Laroche<sup>‡</sup>. Journal of Machine Learning Research (<b>JMLR</b>), 2022.</li><li>4. <u>Truncated Emphatic Temporal Difference Methods for Prediction and Control</u> <b>Shangtong Zhang</b>, Shimon Whiteson. Journal of Machine Learning Research (<b>JMLR</b>), 2022.</li><li>5. <u>A Deeper Look at Discounting Mismatch in Actor-Critic Algorithms</u> <b>Shangtong Zhang</b>, Romain Laroche, Harm van Seijen, Shimon Whiteson, Remi Tachet des Combes. International Conference on Autonomous Agents and Multiagent Systems (<b>AAMAS</b>), 2022. Acceptance rate: 26%</li><li>6. <u>Learning Expected Emphatic Traces for Deep RL</u> Ray Jiang, <b>Shangtong Zhang</b>, Veronica Chelu, Adam White, Hado van Hasselt. AAAI Conference on Artificial Intelligence (<b>AAAI</b>), 2022. Acceptance rate: 15%.</li></ol>	

7. 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
8. 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%.
9. 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%.
10. 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%.
11. 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%.
12. 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%.
13. 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%.
14. 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.**
15. 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%.
16. DAC: The Double Actor-Critic Architecture for Learning Options  
**Shangdong Zhang**, Shimon Whiteson.  
Conference on Neural Information Processing Systems (**NeurIPS**), 2019.  
Acceptance rate: 21.2%.
17. Generalized Off-Policy Actor-Critic  
**Shangdong Zhang**, Wendelin Boehmer, Shimon Whiteson.

- Conference on Neural Information Processing Systems (**NeurIPS**), 2019.  
Acceptance rate: 21.2%.
18. Distributional Reinforcement Learning for Efficient Exploration  
Borislav Mavrin, **Shangtong Zhang**<sup>†</sup>, 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.
  19. 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%.
  20. 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%.
  21. 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.
  22. 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%.
  23. A Deeper Look at Experience Replay  
**Shangtong Zhang**, Richard S. Sutton.  
**Deep RL Symposium at NIPS**, 2017.
  24. Comparing Deep Reinforcement Learning and Evolutionary Methods  
in Continuous Control  
**Shangtong Zhang**, Osmar R. Zaiane  
**Deep RL Symposium at NIPS**, 2017.
  25. A Demon Control Architecture with Off-Policy Learning and Flexible Behavior  
Policy  
**Shangtong Zhang**, Richard S. Sutton.  
**Hierarchical RL Workshop at NIPS**, 2017.
  26. 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.

Services	<b>Meta Reviewer &amp; Area Chair</b>		
	ACML 2022, 2023		
	<b>Expert Reviewer</b>		
	ICML 2021		
	<b>Reviewer &amp; Program Committee</b>		
	Transactions on Pattern Analysis and Machine Intelligence (1)		
	Transaction of Machine Learning Research (2)		
	Journal of Machine Learning Research (3)		
	Artificial Intelligence Journal (2)		
	Transactions on Intelligent Systems and Technology (2)		
	IJCAI 2023		
	AISTATS 2022		
	NeurIPS 2020, 2021, 2022, 2023		
	ICML 2020, 2022, 2023		
	AAAI 2020, 2021, 2022, 2023		
	ICLR 2021, 2022, 2023		
	SIGCOMM 2022		
	Offline Reinforcement Learning Workshop at NeurIPS 2020, 2021, 2022		
	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		
	Reinforcement Learning for Real Life Workshop at NeurIPS 2022		
	<b>Departmental Services</b>		
	Graduate Admission Committee at CS of UVA, 2022 - 2023 AY		
Honours	<i>Runner-Up for the IFAAMAS Victor Lesser Dissertation Award</i>		2022
	<i>Alf Weaver Junior Faculty Fellowship, University of Virginia</i>		2022 - 2027
	<i>EPSRC Studentship, University of Oxford</i>		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, St Catherine's College, University of Oxford</i>		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>EMC Scholarship, Fudan University</i>		2014
Supervision	<b>PhD Students</b>		
	Shuze Liu (2022 - )		
	<b>Master Students</b>		
	Zhengkun Xiao (2022 - )		
	<b>Undergraduates</b>		

Ja-Zhua Cheng (2022 - )

### Selected Research Assistants

Jiuqi Wang (2022 - )

Xiaochi Qian (2022 - )

### PhD Committees

Ingy ElSayed-Aly (Advised by Prof. Lu Feng at UVA)

Sudhir Shenoy (Advised by Prof. Afsaneh Doryab at UVA)

Chuanhao Li (Advised by Prof. Hongning Wang at UVA)

*Qualification:*

Zeyu Mu (Advised by Prof. Brian Park at UVA)

Dane Williamson (Advised by Prof. Yangfeng Ji at UVA)

Matthew Landers (Advised by Prof. Afsaneh Doryab at UVA)

Ethan Harrison Blaser (Advised by Prof. Hongning Wang at UVA)

Invited Talks	<i>Offline Reinforcement Learning: Current and Future</i>	2023
	AAAI New Faculty Highlight Program	
	<i>Breaking the Deadly Triad in Off-Policy Reinforcement Learning</i>	
	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
	School of Informatics, University of Edinburgh	2021
	<i>Breaking the Deadly Triad in Reinforcement Learning</i>	2021
	RL team, DeepMind, hosted by Hado van Hasselt	
	<i>Breaking the Deadly Triad with a Target Network</i>	2021
	Microsoft Research Summit	
	<i>Off-Policy Evaluation</i>	2020
	Data Fest 2020, Open Data Science	
	<i>Off-Policy Evaluation and Control</i>	2020
	ByteDance AI Lab, Shanghai	
Teaching	<i>Off-Policy Actor-Critic Algorithms</i>	2019
	Latent Logic LTD, Oxford	
	<i>Generalized Off-Policy Actor-Critic</i>	2019
	Noah's Ark Lab, Huawei, Edmonton	
	<i>Exploration with Quantile Options</i>	2018
	Huawei RL Workshop, Edmonton	
	<i>Coding Deep RL Papers</i>	2017
	NIPS MLTrain Workshop, Long Beach	
Teaching	<i>University of Virginia, Instructor</i>	
	CS6501: Topics in Reinforcement Learning	Fall 2022

*University of Oxford*, Teaching Assistant  
AIMS CDT Lectures

Michaelmas 2019

*University of Alberta*, Teaching Assistant  
CMPUT 229 Computer Organization and Architecture

Fall 2016

## Research Internships

**Microsoft Research Montreal**, Canada Jun. 2021 - Sept. 2021  
Collaboration: Remi Tachet des Combes, Romain Laroche, and Harm van Seijen

**DeepMind London**, United Kingdom Feb. 2021 - Jun. 2021  
Collaboration: AlphaStar team (Michael Mathieu, Oriol Vinyals, etc)  
Collaboration: Adam White and Hado van Hasselt

**DeepDrive**, Edmonton, Canada Sept. 2020 - Dec. 2020  
Collaboration: Hengshuai Yao

**Microsoft Research Montreal**, Canada Jun. 2020 - Aug. 2020  
Collaboration: Remi Tachet des Combes, Romain Laroche, and Harm van Seijen

**Noah's Ark Lab, Huawei**, Edmonton, Canada May. 2018 - Aug. 2018  
Collaboration: Hengshuai Yao

## Code

*PyTorch Deep RL*  
A zoo of popular deep RL algorithms in PyTorch with **2.8k stars** in Github.

*Reinforcement Learning: An Introduction*  
Python implementation of the book *Reinforcement Learning: An Introduction*  
with **11.7k stars** in Github.

*Google Summer of Code (GSoC) 2017*  
Contributed to MLPack by implementing a deep RL framework.

*Google Summer of Code (GSoC) 2014*  
Contributed to Xapian by optimizing the post list and the position list.