

## Shangtong Zhang

Department of Computer Science  
School of Engineering and Applied Science  
University of Virginia  
Charlottesville, VA, 22903

Rice Hall 422  
85 Engineer's Way  
Charlottesville, VA, 22903  
[shangtong@virginia.edu](mailto:shangtong@virginia.edu)

### RESEARCH INTEREST

The goal of my research is to solve sequential decision making problems in a scalable and reliable way. Currently, I focus on Reinforcement Learning (RL) as a solution method. In particular, I work on stochastic approximations for RL, theories and algorithms of RL, and applications by RL.

### ACADEMIC EMPLOYMENTS

<b>Assistant Professor</b> Department of Computer Science University of Virginia, VA, United States	Aug 2022 - Present
<b>Research Scientist Interns</b> Microsoft Research Montreal DeepMind London Microsoft Research Montreal	Jun 2021 - Sep 2021 Feb 2021 - Jun 2021 Jun 2020 - Aug 2020

### EDUCATION

<b>Doctor of Philosophy</b> , Computer Science University of Oxford, Oxford, United Kingdom Advisor: Prof. Shimon Whiteson	Oct 2018 - Jul 2022
<b>Master of Science</b> , Computer Science University of Alberta, Edmonton, Canada Advisor: Prof. Richard S. Sutton	Sep 2016 - Jul 2018
<b>Bachelor of Science</b> , Computing Science Fudan University, Shanghai, China	Sep 2012 - Jul 2016

### PUBLICATIONS

Advisees of SZ are underlined; \* indicates equal contribution; <sup>†</sup> indicates equal supervision.

#### Preprints

- (P1) *CRASH: Challenging Reinforcement-Learning Based Adversarial Scenarios For Safety Hardening.*  
Amar Kulkarni, **Shangtong Zhang**, Madhur Behl.  
arXiv:2411.16996, 2024.

- (P2) [\*Almost Sure Convergence Rates and Concentration of Stochastic Approximation and Reinforcement Learning with Markovian Noise.\*](#)  
Xiaochi Qian\*, Xinyu Liu\*, Zixuan Xie\*, **Shangdong Zhang**.  
arXiv:2411.13711, 2024.
- (P3) [\*Almost Sure Convergence of Average Reward Temporal Difference Learning.\*](#)  
Ethan Blaser, **Shangdong Zhang**.  
arXiv:2409.19546, 2024.
- (P4) [\*Almost Sure Convergence of Linear Temporal Difference Learning with Arbitrary Features.\*](#)  
Jiuqi Wang, **Shangdong Zhang**.  
arXiv:2409.12135, 2024.
- (P5) [\*StarCraft II Unplugged: Large Scale Offline Reinforcement Learning.\*](#)  
Michael Mathieu\*, Sherjil Ozair\*, Srivatsan Srinivasan\*, Caglar Gulcehre\*, **Shangdong Zhang\***, Ray Jiang\*, Tom Le Paine\*, Richard Powell, Konrad Zolna, 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.  
arXiv:2308.03526, 2023

#### Invited Articles

- (I1) [\*A New Challenge in Policy Evaluation.\*](#)  
**Shangdong Zhang**.  
AAAI Conference on Artificial Intelligence (**AAAI**), 2023  
New Faculty Highlights Program.

#### Refereed Journals

- (J1) [\*The ODE Method for Stochastic Approximation and Reinforcement Learning with Markovian Noise.\*](#)  
Shuze Liu, Shuhang Chen, **Shangdong Zhang**.  
Journal of Machine Learning Research (**JMLR**), 2025.
- (J2) [\*Global Optimality and Finite Sample Analysis of Softmax Off-Policy Actor Critic under State Distribution Mismatch.\*](#)  
**Shangdong Zhang**, Remi Tachet des Combes<sup>†</sup>, Romain Laroche<sup>‡</sup>.  
Journal of Machine Learning Research (**JMLR**), 2022.
- (J3) [\*Truncated Emphatic Temporal Difference Methods for Prediction and Control.\*](#)  
**Shangdong Zhang**, Shimon Whiteson.  
Journal of Machine Learning Research (**JMLR**), 2022.
- (J4) [\*MLPack 3: A Fast, Flexible Machine Learning Library.\*](#)  
Ryan Curtin, Marcus Edel, Mikhail Lozhnikov, Yannis Mentekidis, Sumedh Ghaisas, **Shangdong Zhang**  
Journal of Open Source Software (**JOSS**), 2018.

#### Refereed Conference Papers

- (C1) [\*Transformers Learn Temporal Difference Methods for In-Context Reinforcement Learning.\*](#)  
Jiuqi Wang\*, Ethan Blaser\*, Hadi Daneshmand, **Shangdong Zhang**.  
International Conference on Learning Representations (**ICLR**), 2025.  
Acceptance rate: 32.08%.  
**QuantCo Spotlight Award** at the ICML Workshop on In-Context Learning, 2024.
- (C2) [\*Revisiting a Design Choice in Gradient Temporal Difference Learning.\*](#)  
Xiaochi Qian, **Shangdong Zhang**.

- International Conference on Learning Representations (**ICLR**), 2025.  
Acceptance rate: 32.08%.
- (C3) *Efficient Policy Evaluation with Safety Constraint for Reinforcement Learning*.  
Claire Chen<sup>\*</sup>, Shuze Liu<sup>\*</sup>, **Shangdong Zhang**.  
International Conference on Learning Representations (**ICLR**), 2025.  
Acceptance rate: 32.08%.
- (C4) *Doubly Optimal Policy Evaluation for Reinforcement Learning*.  
Shuze Liu, Claire Chen, **Shangdong Zhang**.  
International Conference on Learning Representations (**ICLR**), 2025.  
Acceptance rate: 32.08%.
- (C5) *Efficient Multi-Policy Evaluation for Reinforcement Learning*.  
Shuze Liu, Claire Chen, **Shangdong Zhang**.  
AAAI Conference on Artificial Intelligence (**AAAI**), 2025.  
Acceptance rate: 23.4%. **Oral presentation (4.6%)**.
- (C6) *Efficient Policy Evaluation with Offline Data Informed Behavior Policy Design*.  
Shuze Liu, **Shangdong Zhang**.  
International Conference on Machine Learning (**ICML**), 2024.  
Acceptance rate: 27.5%
- (C7) *On the Convergence of SARSA with Linear Function Approximation*.  
**Shangdong Zhang**, Remi Tachet des Combes, Romain Laroche.  
International Conference on Machine Learning (**ICML**), 2023.  
Acceptance rate: 28%
- (C8) *A Deeper Look at Discounting Mismatch in Actor-Critic Algorithms*.  
**Shangdong 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**.
- (C9) *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%.
- (C10) *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%.
- (C11) *Average-Reward Off-Policy Policy Evaluation with Function Approximation*.  
**Shangdong Zhang**<sup>\*</sup>, Yi Wan<sup>\*</sup>, Richard S. Sutton, Shimon Whiteson.  
International Conference on Machine Learning (**ICML**), 2021.  
Acceptance rate: 21.5%.
- (C12) *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%.
- (C13) *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%.
- (C14) *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%.

- (C15) *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%.
- (C16) *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**.
- (C17) *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%.
- (C18) *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%.
- (C19) *Generalized Off-Policy Actor-Critic.*  
**Shangdong Zhang**, Wendelin Boehmer, Shimon Whiteson.  
 Conference on Neural Information Processing Systems (**NeurIPS**), 2019.  
 Acceptance rate: 21.2%.
- (C20) *Distributional Reinforcement Learning for Efficient Exploration.*  
 Borislav Mavrin, **Shangdong Zhang**, Hengshuai Yao, Linglong Kong, Kaiwen Wu, Yaoliang Yu  
 International Conference on Machine Learning (**ICML**), 2019.  
 Acceptance rate: 22.6%.
- (C21) *ACE: An Actor Ensemble Algorithm for Continuous Control with Tree Search.*  
**Shangdong Zhang**, Hao Chen, Hengshuai Yao.  
 AAAI Conference on Artificial Intelligence (**AAAI**), 2019.  
 Acceptance rate: 16.2%. **Spotlight presentation**.
- (C22) *QUOTA: The Quantile Option Architecture for Reinforcement Learning.*  
**Shangdong Zhang**, Borislav Mavrin, Linglong Kong, Bo Liu, Hengshuai Yao.  
 AAAI Conference on Artificial Intelligence (**AAAI**), 2019.  
 Acceptance rate: 16.2%. **Oral presentation**.
- (C23) *Crossprop: Learning Representations by Stochastic Meta-Gradient Descent in Neural Networks.*  
 Vivek Veeriah\*, **Shangdong 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%.
- (C24) *A Deep Neural Network for Modeling Music.*  
 Pengjing Zhang, Xiaoqing Zheng, Wenqiang Zhang, Siyan Li, Sheng Qian,  
 Wenqi He, **Shangdong Zhang**, Ziyuan Wang  
 International Conference on Multimedia Retrieval (**ICMR**), 2015.  
 Acceptance rate: 31%.

#### Refereed Workshop Papers (Non-Archival)

- (W1) *A Deeper Look at Experience Replay.*  
**Shangdong Zhang**, Richard S. Sutton.  
 Deep RL Symposium at NIPS, 2017.

- (W2) [Comparing Deep Reinforcement Learning and Evolutionary Methods in Continuous Control.](#)  
**Shangdong Zhang**, Osmar R. Zaiane  
 Deep RL Symposium at NIPS, 2017
- (W3) [A Demon Control Architecture with Off-Policy Learning and Flexible Behavior Policy.](#)  
**Shangdong Zhang**, Richard S. Sutton.  
 Hierarchical RL Workshop at NIPS, 2017.

## GRANTS

[Enhancing the Security of Large Language Models Against Persuasion-Based Jailbreak Attacks in Multi-Turn Dialogues](#)  
 CCI Coastal Virginia Node, **Co-PI**, Total \$60,000, My Share \$20,000 2025 - 2026  
[RAMPART: Reinforcement Against Malicious Penetration by Adversaries in Realistic Topologies.](#)  
 DARPA HR001123S0002, **Co-PI**, Total \$2,150,000, My Share \$77,000. 2023 - 2027  
[SLES: CRASH: Challenging Reinforcement-Learning Based Adversarial Scenarios for Safety Hardening.](#)  
 NSF 2331904, **Co-PI**, Total \$800,000, My Share \$400,000. 2023 - 2026  
[III: Small: Moving Offline Learning to Rank Online, from Theory to Practice.](#)  
 NSF 2128019, **PI**, Total \$500,000, My Share \$500,000. 2021 - 2025

## HONORS

QuantCo Spotlight Award at the ICML In-Context Learning Workshop, 2024  
 Rising Stars in AI by KAUST AI Initiative, 2024  
 AAAI New Faculty Highlights, 2023  
 IFAAMAS Victor Lesser Dissertation Award (Runner-Up), 2022  
 Alf Weaver Junior Faculty Fellowship, UVA, 2022 - 2025  
 ICLR Outstanding Reviewer, 2021  
 NeurIPS Reviewer Award, 2020  
 ICML Reviewer Award, 2020  
 AAMAS Best Paper Award, 2020  
 Light Senior Scholarship, St Catherine's College, University of Oxford, 2020  
 EPSRC Studentship, University of Oxford, 2018 - 2022  
 EMC Scholarship, Fudan University, 2014

## SERVICES

### Organizers

CPS Rising Star Workshop 2024, Co-Chair

### Panelists

National Science Foundation (NSF), 2024  
 Virginia's Commonwealth Cyber Initiative (CCI), 2024

### Meta Meta Reviewer

RL Conference 2025 (Senior Area Chair)

### Meta Reviewer

ICML 2025 (Area Chair)

ICLR 2024, 2025 (Area Chair)  
 L4DC 2025 (Program Committee)  
 AAMAS 2025 (Senior Program Committee)  
 RL Conference 2024 (Area Chair)  
 AISTATS 2024, 2025 (Area Chair)  
 ACML 2022, 2023, 2024 (Area Chair)

### Reviewer

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, 2021, 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

### Conference Session Chair

AAAI 2023, “Reinforcement Learning Theory & Algorithms”

## SUPERVISION

### Doctrnal Students

Shuze Liu	2022 - Now
Ethan Blaser, NSF Graduate Research Fellowship	2023 - Now
Jiuqi Wang	2023 - Now
Xinyu Liu	2024 - Now
Zixuan Xie, UVA Provost’s Fellowship, UVA Engineering Distinguished Fellowship	2024 - Now
Amir Moeini	2024 - Now

### Alumni with Theses or Publications

Kefan Song, Master, UVA	2023 - 2024
Licheng Luo, Master, UVA	2023 - 2024
Steve Zhou, Undergraduate, UVA, DMP thesis	2023 - 2024
Xiaochi Qian, Undergraduate, University of Oxford, Research Intern	2022 - 2024

### PhD Committees

Ingy ElSayed-Aly  
 Sudhir Shenoy  
 Chuanhao Li

## INVITED TALKS

<b>Understanding the Training and Inference of Reinforcement Learning</b> Tsinghua University, hosted by Prof. Hongning Wang	June 2024
<b>On the Cheating of Offline Reinforcement Learning</b> KAUST Rising Stars in AI Symposium	Feb 2024
<b>Offline Reinforcement Learning: Current and Future</b> AAAI New Faculty Highlight Program	Feb 2023
<b>Breaking the Deadly Triad in Off-Policy Reinforcement Learning</b> Department of Computer Science, University of Virginia	Mar 2022
School of Computing Science, Simon Fraser University	Feb 2022
Department of Electrical & Computer Engineering, University of Waterloo	Feb 2022
School of Informatics, University of Edinburgh	Oct 2021
<b>Breaking the Deadly Triad with a Target Network</b> Microsoft Research Summit	Oct 2021
<b>Breaking the Deadly Triad in Reinforcement Learning</b> RL team, DeepMind, hosted by Dr. Hado van Hasselt	Sep 2021
<b>Off-Policy Evaluation</b> Data Fest 2020, Open Data Science	Oct 2020
<b>Off-Policy Evaluation and Control</b> ByteDance AI Lab, Shanghai	Oct 2020
<b>Coding Deep RL Papers</b> NIPS MLTrain Workshop, Long Beach	Dec 2019
<b>Off-Policy Actor-Critic Algorithms</b> Latent Logic LTD, Oxford	Apr 2019

## TEACHING

<b>University of Virginia</b>	
CS 4501: Reinforcement Learning	Fall 2024
CS 6316: Machine Learning	Spring 2024
CS 4501: Optimization	Fall 2023
CS 6501: Topics in Reinforcement Learning	Fall 2022

## OPEN SOURCE CODE

<b>GitHub Repo: PyTorch Deep RL</b> A zoo of popular deep RL algorithms in PyTorch with <b>3k stars</b> .	
<b>GitHub Repo: Reinforcement Learning: An Introduction</b> Python implementation of the book <i>Reinforcement Learning: An Introduction</i> with <b>13.8k stars</b> .	
<b>Google Summer of Code (GSoC)</b>	
MLPack	2017
The Xapian Project	2014