

## 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 Huawei Noah's Ark Lab Edmonton	Jun 2021 - Sep 2021 Feb 2021 - Jun 2021 Jun 2020 - Aug 2020 May 2018 - Aug 2018

### 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) *Direct Gradient Temporal Difference Learning*.  
Xiaochi Qian, **Shangtong Zhang**.  
arXiv:2308.01170, Aug 2023. Under review of Journal of Machine Learning Research.

- (P2) [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, Aug 2023. DeepMind technical report.
- (P3) [Improving Monte Carlo Evaluation with Offline Data.](#)  
Shuze Liu, **Shangdong Zhang**.  
 arXiv:2301.13734, Jan 2023.

## Invited Articles

- (I1) [A New Challenge in Policy Evaluation.](#)  
**Shangdong Zhang**.  
*Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, vol. 37, no. 13, pp. 15465-  
 15465, Feb 2023. New Faculty Highlights Program.

## Refereed Journals

- (J1) [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)*, vol. 23, no. 343, pp. 1-91, Oct 2022.
- (J2) [Truncated Emphatic Temporal Difference Methods for Prediction and Control](#)  
**Shangdong Zhang**, Shimon Whiteson.  
*Journal of Machine Learning Research (JMLR)*, vol. 23, no. 153, pp. 1-59, May 2022.
- (J3) [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)*, vol. 3, no. 26, pp. 726, Jun 2018.

## Refereed Conference Papers

- (C1) [On the Convergence of SARSA with Linear Function Approximation.](#)  
**Shangdong Zhang**, Remi Tachet des Combes, Romain Laroche.  
*Proceedings of the International Conference on Machine Learning (ICML) in Proceedings of Machine  
 Learning Research*, vol. 202, pp. 41613-41646, Jul 2023. Acceptance rate: 28%
- (C2) [A Deeper Look at Discounting Mismatch in Actor-Critic Algorithms.](#)  
**Shangdong Zhang**, Romain Laroche, Harm van Seijen, Shimon Whiteson, Remi Tachet des Combes.  
*Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*,  
 pp. 1491, May 2022. Acceptance rate: 26%
- (C3) [Learning Expected Emphatic Traces for Deep RL.](#)  
 Ray Jiang, **Shangdong Zhang**, Veronica Chelu, Adam White, Hado van Hasselt.  
*Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, vol. 36, no. 6, pp. 7015-7023,  
 Feb 2022. Acceptance rate: 15%.
- (C4) [Breaking the Deadly Triad with a Target Network.](#)  
**Shangdong Zhang**, Hengshuai Yao, Shimon Whiteson.  
*Proceedings of the International Conference on Machine Learning (ICML) in Proceedings of Machine  
 Learning Research*, vol. 139, pp. 12621-12631, Jul 2021. Acceptance rate: 21.5%.
- (C5) [Average-Reward Off-Policy Policy Evaluation with Function Approximation.](#)  
**Shangdong Zhang\***, Yi Wan\*, Richard S. Sutton, Shimon Whiteson.

- Proceedings of the International Conference on Machine Learning (ICML) in Proceedings of Machine Learning Research*, vol. 139, pp. 12578-12588, Jul 2021.. Acceptance rate: 21.5%.
- (C6) [Mean-Variance Policy Iteration for Risk-Averse Reinforcement Learning](#).  
**Shangdong Zhang**, Bo Liu, Shimon Whiteson.  
*Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, vol. 35, no. 12, pp. 10905-10913, Feb 2021. Acceptance rate: 21.4%.
- (C7) [Learning Retrospective Knowledge with Reverse Reinforcement Learning](#).  
**Shangdong Zhang**, Vivek Veeriah, Shimon Whiteson.  
*Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 19976-19987, Dec 2020. Acceptance rate: 20.1%.
- (C8) [GradientDICE: Rethinking Generalized Offline Estimation of Stationary Values](#).  
**Shangdong Zhang**, Bo Liu, Shimon Whiteson.  
*Proceedings of the International Conference on Machine Learning (ICML) in Proceedings of Machine Learning Research*, vol. 119, pp. 11194-11203, Jul 2020. Acceptance rate: 21.8%.
- (C9) [Provably Convergent Two-Timescale Off-Policy Actor-Critic with Function Approximation](#).  
**Shangdong Zhang**, Bo Liu, Hengshuai Yao, Shimon Whiteson.  
*Proceedings of the International Conference on Machine Learning (ICML) in Proceedings of Machine Learning Research*, vol. 119, pp. 11204-11213, Jul 2020. Acceptance rate: 21.8%.
- (C10) [Deep Residual Reinforcement Learning](#).  
**Shangdong Zhang**, Wendelin Boehmer, Shimon Whiteson.  
*Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pp. 1611–1619, May 2020. Acceptance rate: 23%. **Best Paper Award**.
- (C11) [Mega-Reward: Achieving Human-Level Play without Extrinsic Rewards](#).  
Yuhang Song, Jianyi Wang, Thomas Lukasiewicz, Zhenghua Xu, **Shangdong Zhang**, Andrzej Wojcicki, Mai Xu  
*Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, vol. 34, no. 4, pp. 5826-5833, Feb 2020. Acceptance rate: 20.6%.
- (C12) [DAC: The Double Actor-Critic Architecture for Learning Options](#).  
**Shangdong Zhang**, Shimon Whiteson.  
*Advances in Neural Information Processing Systems (NeurIPS)*, vol. 32, pp. 2012-2022, Dec 2019. Acceptance rate: 21.2%.
- (C13) [Generalized Off-Policy Actor-Critic](#).  
**Shangdong Zhang**, Wendelin Boehmer, Shimon Whiteson.  
*Advances in Neural Information Processing Systems (NeurIPS)*, vol. 32, pp. 2001-2011, Dec 2019. Acceptance rate: 21.2%.
- (C14) [Distributional Reinforcement Learning for Efficient Exploration](#).  
Borislav Mavrin, **Shangdong Zhang**, Hengshuai Yao, Linglong Kong, Kaiwen Wu, Yaoliang Yu  
*Proceedings of the International Conference on Machine Learning (ICML) in Proceedings of Machine Learning Research*, vol. 2019, pp. 4424-4434, Jul 2019. Acceptance rate: 22.6%.
- (C15) [ACE: An Actor Ensemble Algorithm for Continuous Control with Tree Search](#).  
**Shangdong Zhang**, Hao Chen, Hengshuai Yao.  
*Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, vol. 33, no. 1, pp. 5789-5796, Feb 2019. Acceptance rate: 16.2%.
- (C16) [QUOTA: The Quantile Option Architecture for Reinforcement Learning](#).  
**Shangdong Zhang**, Borislav Mavrin, Linglong Kong, Bo Liu, Hengshuai Yao.  
*Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, vol. 33, no. 01, pp. 5797-5804, Feb 2019. Acceptance rate: 16.2%.
- (C17) [Crossprop: Learning Representations by Stochastic Meta-Gradient Descent in Neural Networks](#).  
Vivek Veeriah\*, **Shangdong Zhang\***, Richard S. Sutton.  
*Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge*

*Discovery in Databases (ECML-PKDD) in Lecture Notes in Computer Science*, vol. 10534, pp. 445-459, Sep 2017. Acceptance rate: 27.1%.

- (C18) [A Deep Neural Network for Modeling Music](#).  
Pengjing Zhang, Xiaoqing Zheng, Wenqiang Zhang, Siyan Li, Sheng Qian,  
Wenqi He, **Shangdong Zhang**, Ziyuan Wang  
*Proceedings of the International Conference on Multimedia Retrieval (ICMR)*, pp. 379-386, Jun 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, Dec 2017.
- (W2) [Comparing Deep Reinforcement Learning and Evolutionary Methods in Continuous Control](#).  
**Shangdong Zhang**, Osmar R. Zaiane  
Deep RL Symposium at NIPS, Dec 2017.
- (W3) [A Demon Control Architecture with Off-Policy Learning and Flexible Behavior Policy](#).  
**Shangdong Zhang**, Richard S. Sutton.  
Hierarchical RL Workshop at NIPS, Dec 2017.

#### FUNDING

[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 \$400,000. 2023 - 2024

#### HONORS

KAUST Rising Stars in AI<sup>1</sup>, 2024  
AAAI New Faculty Highlights, 2023  
IFAAMAS Victor Lesser Dissertation Award (Runner-Up), 2022  
Alf Weaver Junior Faculty Fellowship, UVA, 2022  
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  
EMC Scholarship, Fudan University, 2014

#### SERVICES

##### Organizers

CPS Rising Star Workshop 2024, Co-Chair

---

<sup>1</sup>Organized by Juergen Schmidhuber

### Meta Reviewer & Area Chair

ICLR 2024  
AISTATS 2024  
ACML 2022, 2023

### Reviewer & Program Committee

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”

### Departmental Services

Faculty Search Committee, UVA CS	2023 - 2024
Graduate Admission Committee, UVA CS	2022 - 2024

## SUPERVISION

### Doctral Students

Ethan Blaser, NSF GRFP	2023 - Now
Jiuqi Wang	2023 - Now
Shuze Liu	2022 - Now

### Master of Science Students

Kefan Song	2023 - Now
Licheng Luo	2023 - Now
Zhengkun Xiao. Then PhD student at University of Florida	2022 - 2023

### Undergraduate Researchers

Xi (Cici) Wang	2023 - Now
Steve Zhou, BA CS Distinguished Major Program	2023 - Now
Pawan Jayakumar	2023 - Now
Ja-Zhua Cheng	2022 - Now

### Research Interns

Jiuqi Wang, University of Alberta, next stop: PhD student at UVA	2022 - 2023
--	-------------

Xiaochi (Joe) Qian, University of Oxford

2022 - Now

### PhD Committees

Sudhir Shenoy

Chuanhao Li

Kun Yang (Proposal)

Zeyu Mu (Proposal)

Ingy ElSayed-Aly (Proposal)

Matthew Landers (Qualification)

Amar Kulkarni (Qualification)

### INVITED TALKS

#### Offline Reinforcement Learning: Current and Future

AAAI New Faculty Highlight Program

Feb 2023

#### Breaking the Deadly Triad in Off-Policy Reinforcement Learning

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

#### Breaking the Deadly Triad with a Target Network

Microsoft Research Summit

Oct 2021

#### Breaking the Deadly Triad in Reinforcement Learning

RL team, DeepMind, hosted by Hado van Hasselt

Sep 2021

#### Off-Policy Evaluation

Data Fest 2020, Open Data Science

Oct 2020

#### Off-Policy Evaluation and Control

ByteDance AI Lab, Shanghai

Oct 2020

#### Coding Deep RL Papers

NIPS MLTrain Workshop, Long Beach

Dec 2019

#### Off-Policy Actor-Critic Algorithms

Latent Logic LTD, Oxford

Apr 2019

### TEACHING

#### University of Virginia

CS 6316: Machine Learning

Spring 2024

CS 4501: Optimization

Fall 2023

CS 6501: Topics in Reinforcement Learning

Fall 2022

### OPEN SOURCE CODE

#### GitHub Repo: PyTorch Deep RL

A zoo of popular deep RL algorithms in PyTorch with **3k stars**.

#### GitHub Repo: Reinforcement Learning: An Introduction

Python implementation of the book *Reinforcement Learning: An Introduction* with **13.8k stars**.

**Google Summer of Code (GSoC)**

MLPack

2017

The Xapian Project

2014