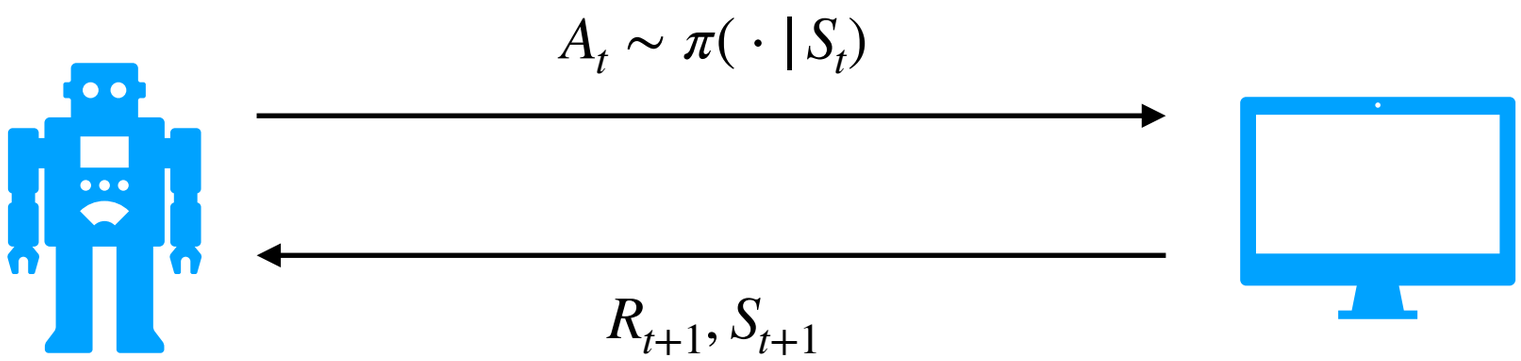
On the Cheating of Offline RL

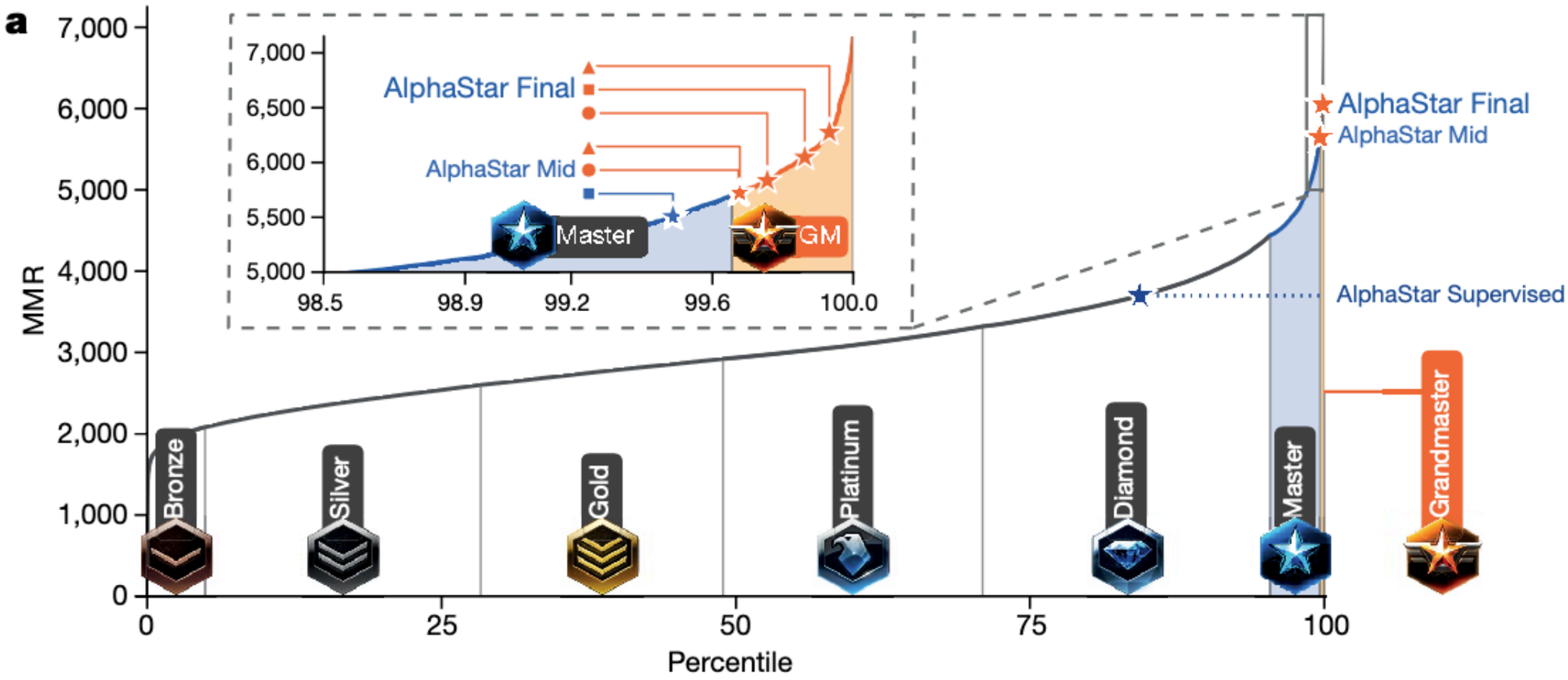
Shangtong Zhang, Assistant Professor

Department of Computer Science
University of Virginia
https://shangtongzhang.github.io/

Canonical RL relies on agent-env interaction



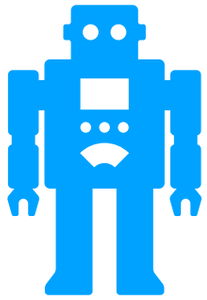
Case study: AlphaStar



(Vinyals et al. 2019)

Trillions of interactions with SCII simulator!

Online interaction can be slow

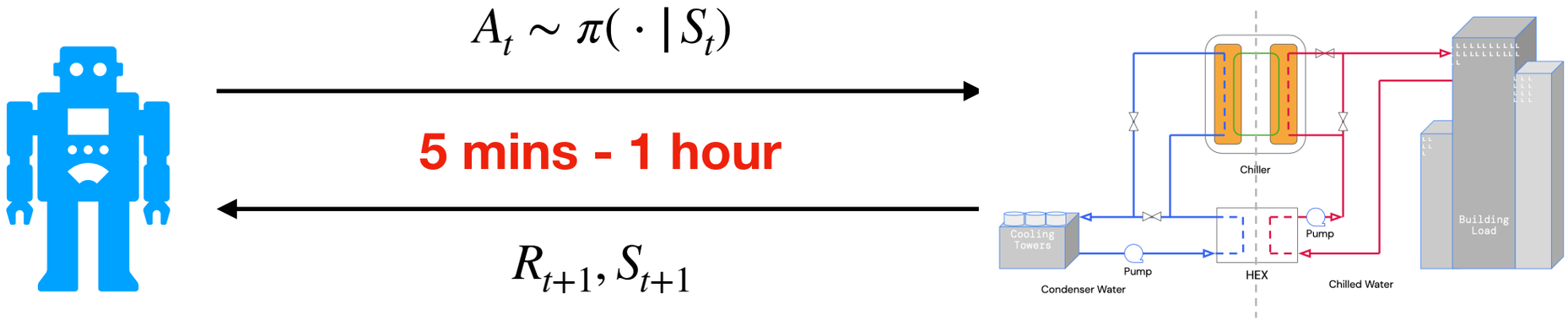


$A_t \sim \pi(\cdot \mid S_t)$

R_{t+1}, S_{t+1}

DMV

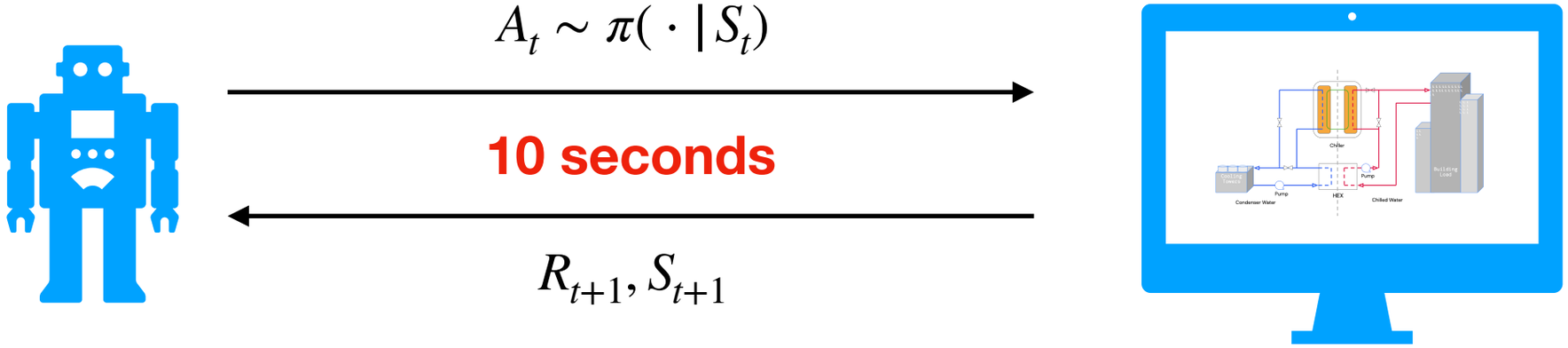
Case study: industrial cooling system



(Chervonyi et al. 2022)

1M training steps is nothing in RL

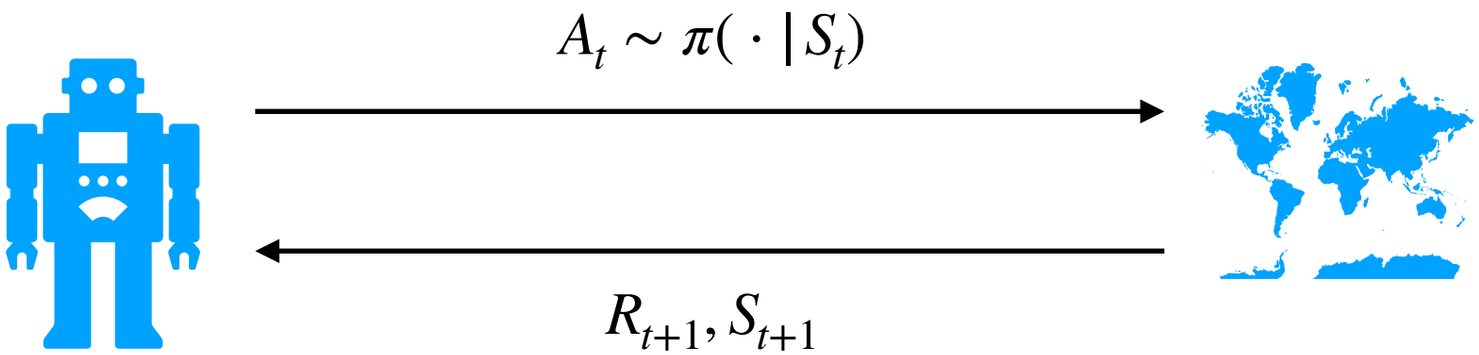
Case study: industrial cooling system



(Chervonyi et al. 2022)

$10 \times 10^6 \div 3600 \div 24 \approx 116$ **days**

Online interaction can be dangerous

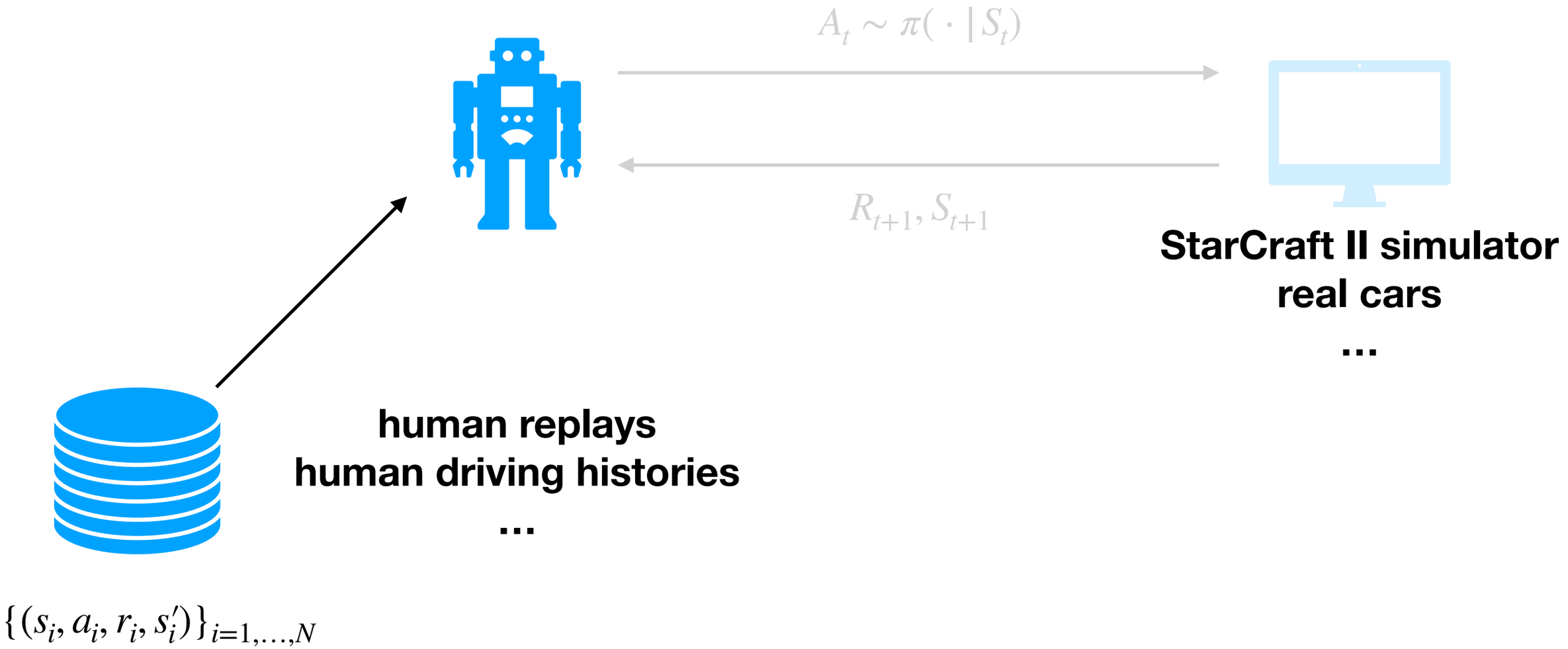


TECH NEWS

Self-driving Uber car that hit and killed woman did not recognize that pedestrians jaywalk

The automated car lacked "the capability to classify an object as a pedestrian unless that object was near a crosswalk," an NTSB report said.

Offline RL uses previously logged data



Case study: Offline AlphaStar

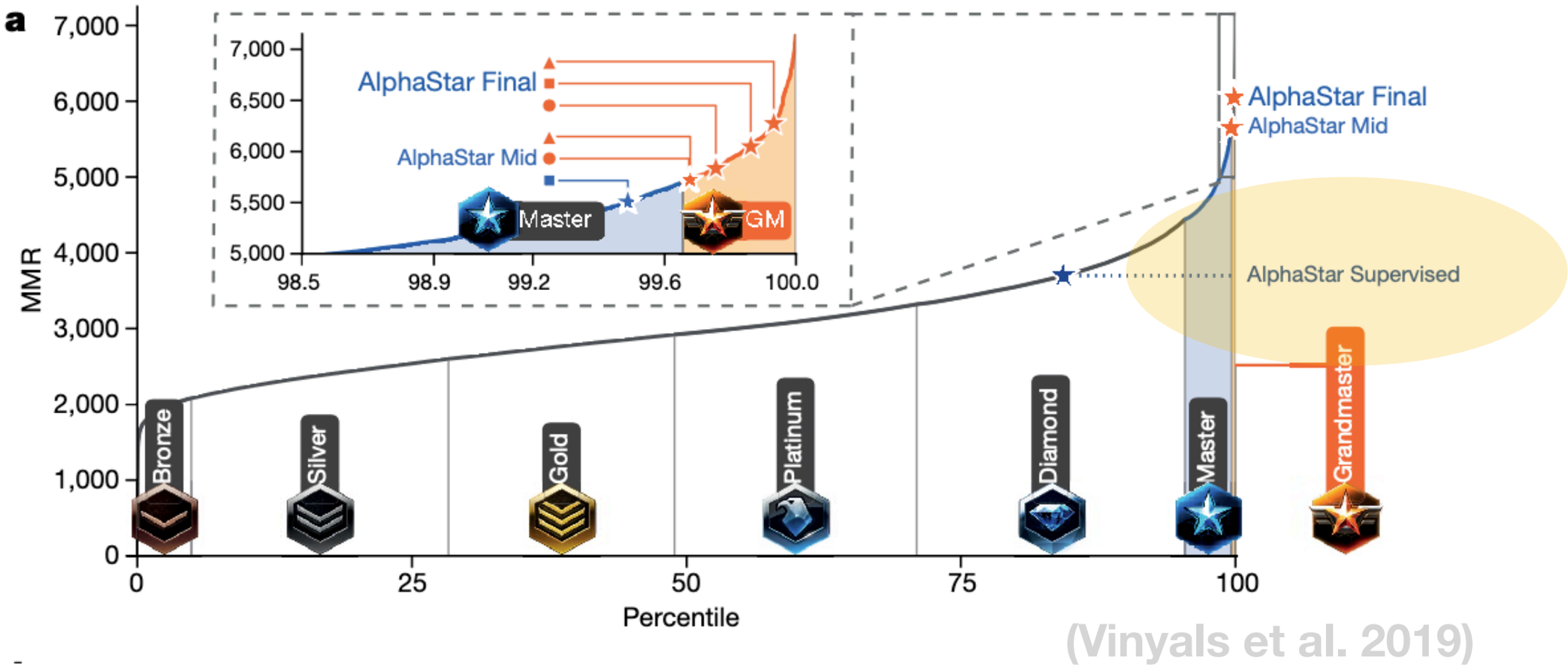
DeepMind

2023-8-8

AlphaStar Unplugged: Large-Scale Offline Reinforcement Learning

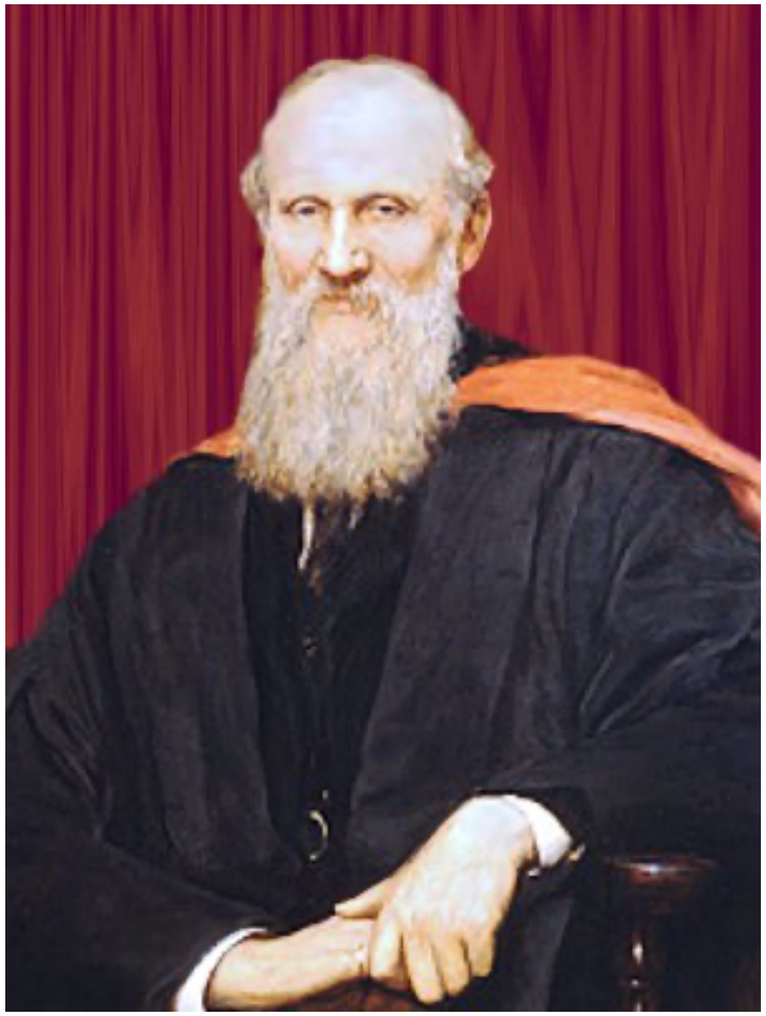
Michaël Mathieu[*,1], Sherjil Ozair[*,1], Srivatsan Srinivasan[*,1], Caglar Gulcehre[*,1], Shangtong Zhang[*,2], Ray Jiang[*,1], Tom Le Paine[*,1], Richard Powell[1], Konrad Żołna[1], Julian Schrittwieser[1], David Choi[1], Petko Georgiev[1], Daniel Toyama[1], Aja Huang[1], Roman Ring[1], Igor Babuschkin[1], Timo Ewalds[1], Mahyar Bordbar[1], Sarah Henderson[1], Sergio Gómez Colmenarejo[1], Aäron van den Oord[1], Wojciech Marian Czarnecki[1], Nando de Freitas[1] and Oriol Vinyals[1]

Case study: Offline AlphaStar



(Vinyals et al. 2019)

Offline AlphaStar has more than 90% win-rate against AlphaStar Supervised.



William Thomson, Lord Kelvin
1824 - 1907

Only two small clouds remained on the horizon of knowledge in ~~physics~~ offline RL.

Kyunghyun Cho
@kchonyc

how do people tune hyperparameters in offline reinforcement learning???

3:26 PM · Jun 23, 2023 · **105.2K** Views

Kyunghyun Cho

New York University, Genentech
Verified email at nyu.edu - Homepage

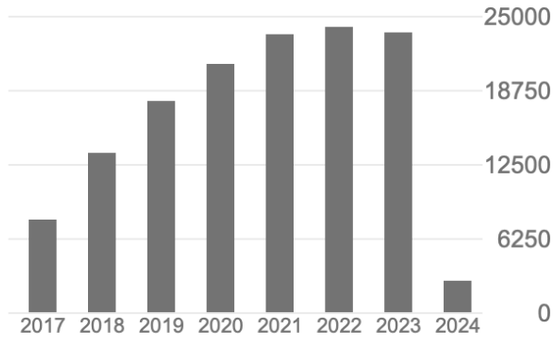
Machine Learning Deep Learning

FOLLOW

TITLE	CITED BY	YEAR
Neural machine translation by jointly learning to align and translate D Bahdanau, K Cho, Y Bengio ICLR 2015	33050	2014
Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation K Cho, B van Merrienboer, C Gulcehre, F Bougares, H Schwenk, ... Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)	28511	2014
Empirical evaluation of gated recurrent neural networks on sequence modeling J Chung, C Gulcehre, KH Cho, Y Bengio arXiv preprint arXiv:1412.3555	15084	2014

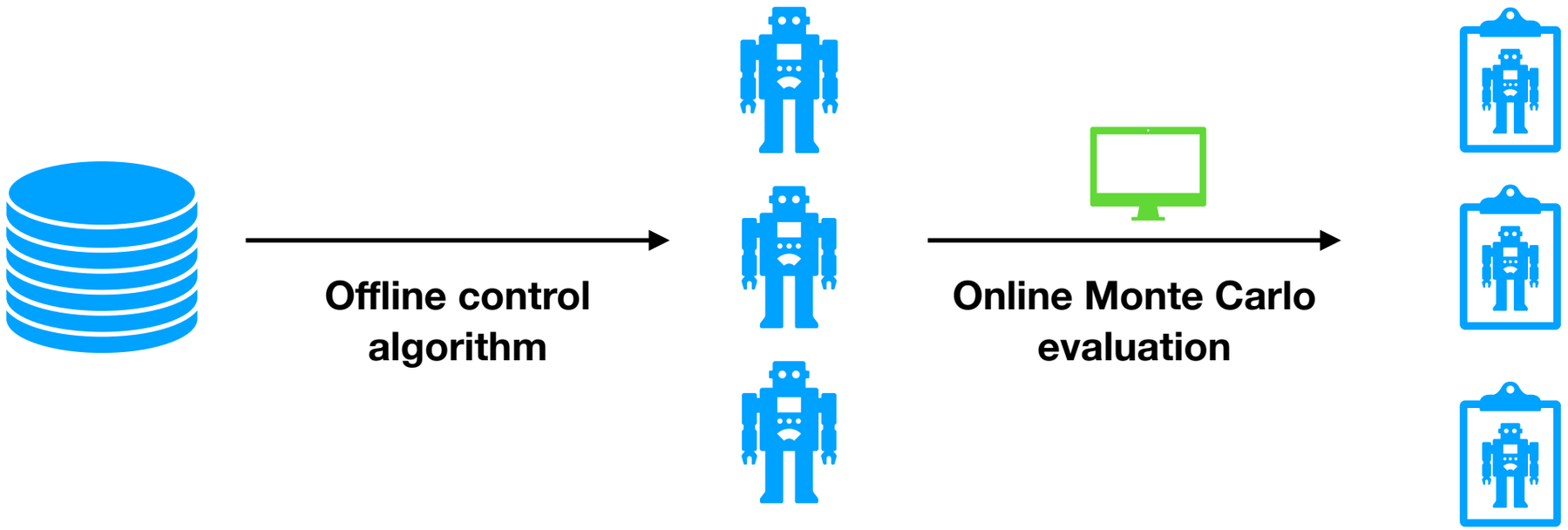
Cited by VIEW ALL

	All	Since 2019
Citations	141989	113051
h-index	98	90
i10-index	234	220

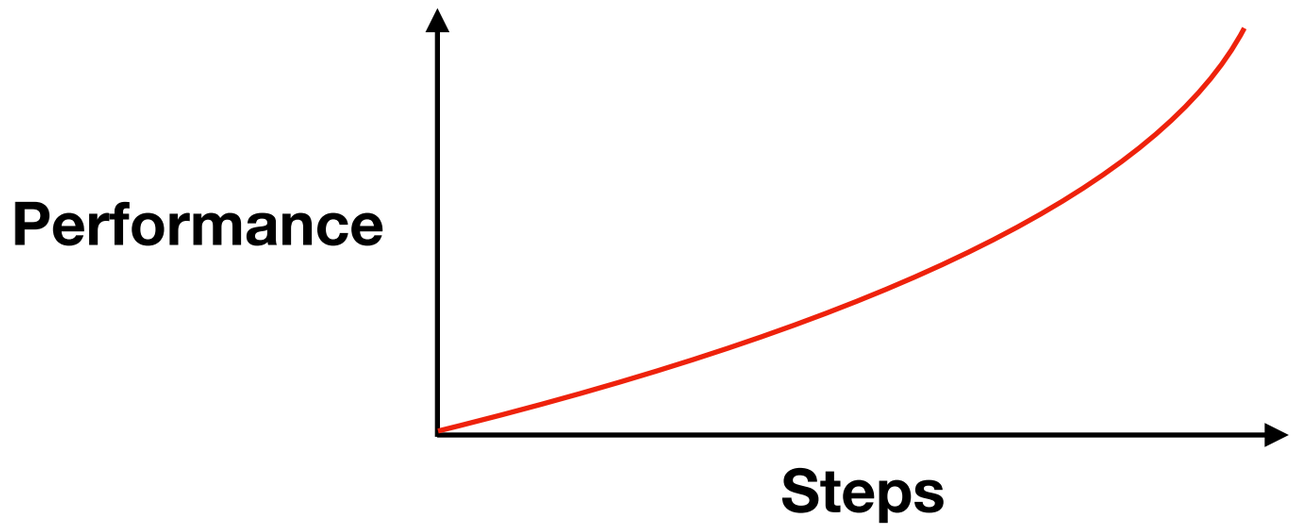


Public access VIEW ALL

Offline RL uses simulator for model selection

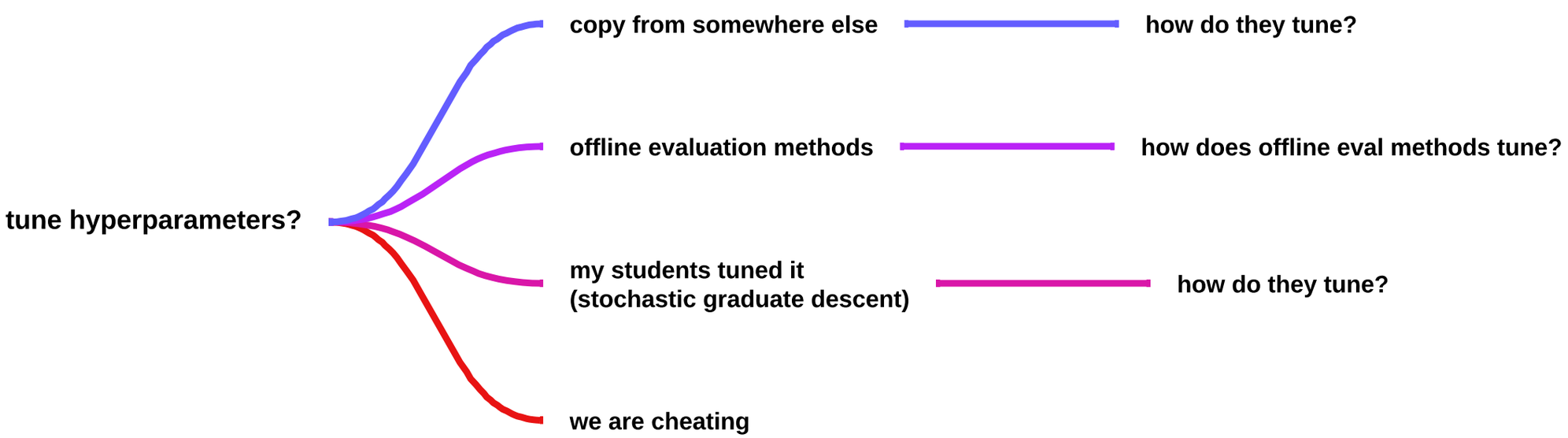


Monte Carlo dominates RL evaluation



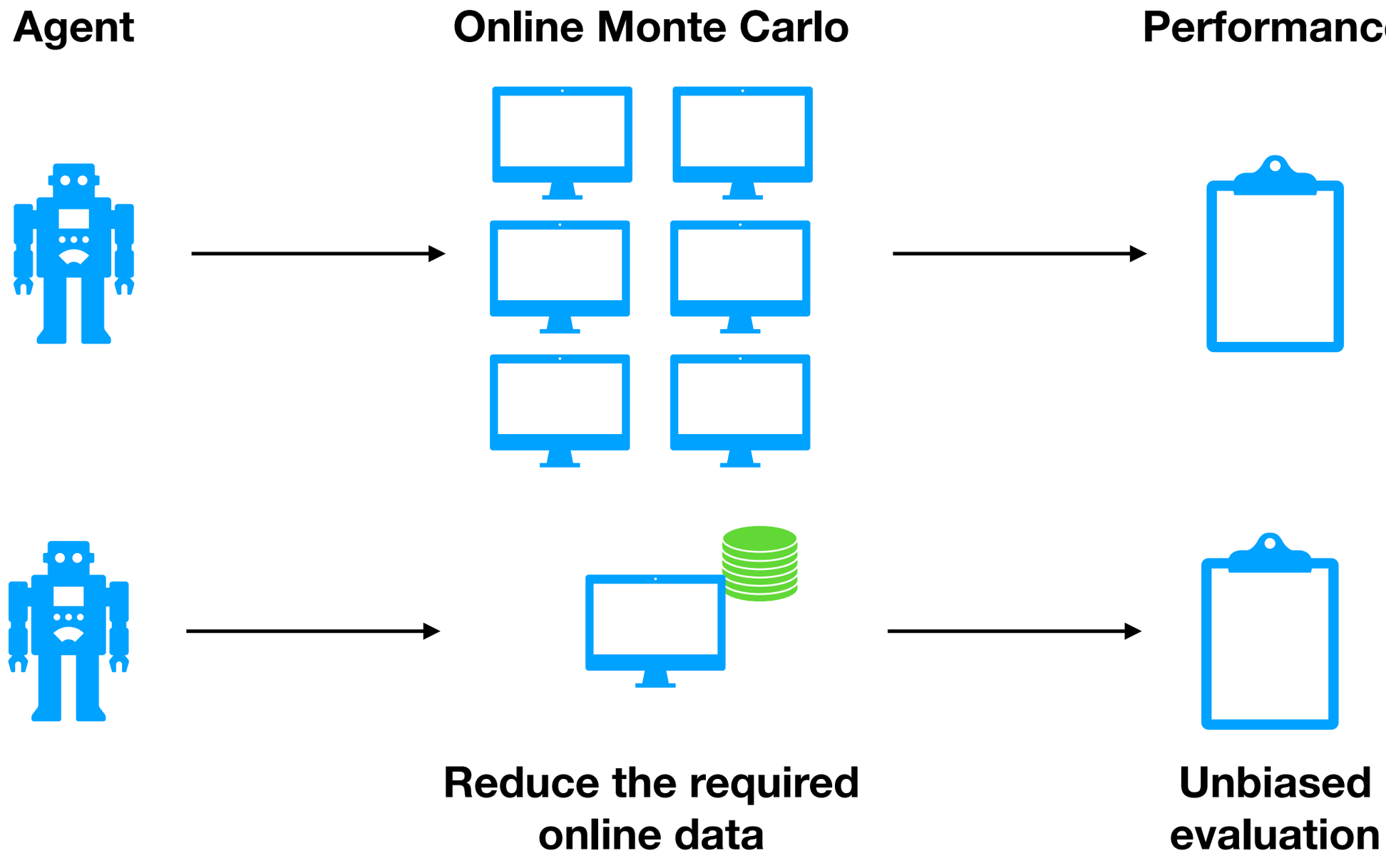
99% of such curves in RL papers are generated by online Monte Carlo

Guideline for attending offline RL talks / posters



Simulator!!!

Our approach: admit that we have to use Monte Carlo but try to use Monte Carlo smartly



Improving Monte Carlo Evaluation with Offline Data.
Shuze Liu, Shangtong Zhang. arXiv:2301.13734, 2023.

Thanks & Questions