

# Structural Causal Model Overhead and Convergence Gains in CausalMixFT Fine-Tuning

Assignee Research

June 9, 2026

## Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: What is the computational overhead of fitting Structural Causal Models during CausalMixFT fine-tuning relative to the training time reduction gained through faster convergence on large-scale tabular. 15 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.1/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Causal-Paced Deep Reinforcement Learning. Research question: What is the computational overhead of fitting Structural Causal Models during CausalMixFT fine-tuning relative to the training time reduction gained through faster convergence on large-scale tabular datasets?.

## 2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.1/10.

## 3 Results

12 papers retrieved. 15 claims extracted; 0 independently verified. Quality review score: 3.1/10.

## 4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

## 5 Extracted Claims

Claim	Verified	Confidence
CP-DRL was evaluated on the Point Mass (PM) and Bipedal Walker (BW) benchmark environments.	×	0.10
Proximal Policy Optimization (PPO) was used as the student algorithm for the Point Mass environment.	×	0.05
Soft Actor-Critic (SAC) was used as the student algorithm for the Bipedal Walker environment.	×	0.02
The baselines evaluated include CURROT, SPRL, GoalGAN, ALPGMM, ACL, PLR, and VDS.	×	0.00
Experiments for the Point Mass environment were conducted with 10 random seeds.	×	0.02
Experiments for Bipedal Walker trivial tasks were conducted with 5 random seeds.	×	0.08
Experiments for Bipedal Walker infeasible tasks were conducted with 3 random seeds.	×	0.06
In the Point Mass environment, the target context distribution is bimodal, corresponding to two gate positions on opposi	×	0.01
Each training epoch in the Point Mass environment consists of 4,096 rollouts.	×	0.02
All methods in the Point Mass environment were trained for 200 epochs.	×	0.05
CP-DRL reached a cumulative discounted return of $6.17 \pm 0.08$ at epoch 195 in the Point Mass environment.	×	0.07
CURROT achieved a cumulative discounted return of $5.6 \pm 0.34$ at epoch 195 in the Point Mass environment.	×	0.01
CP-DRL outperformed CURROT by approximately 10.2% in cumulative discounted return in the Point Mass environment.	×	0.07
CP-DRL maintained consistently low standard errors during training in the Point Mass environment.	×	0.08
CURROT exhibited increasing variance over time during training in the Point Mass environment.	×	0.03

## References

- <http://arxiv.org/abs/2507.02910v1>
- <http://arxiv.org/abs/2110.06500v2>
- <http://arxiv.org/abs/1911.07420v1>