Neuro-Symbolic Reinforcement Learning for Cognitive Digital Twins: A Framework for Lifelong Intelligence Amplification

Patrick O. Akinwumi, Meihua Qian, Clemson University, USA

IEEE WCCI 2026 Maastricht

Introduction

Artificial intelligence systems are typically optimized for immediate reward maximization or short-term task performance. While reinforcement learning (RL) has achieved major advances in sequential decision-making, conventional RL often struggles under delayed transfer conditions, hierarchical dependencies, and long-horizon adaptation [1]. Human cognition, however, develops through gradual capability accumulation, prerequisite acquisition, transfer learning, and continuous adaptation.

Recent work in curiosity-driven RL improves exploratory behavior [2], while neuro-symbolic AI integrates neural learning with symbolic reasoning to improve compositional structure and interpretability [3]. However, existing approaches remain primarily optimized for local reward or exploration objectives rather than long-horizon cognitive growth.

Research Question

Can intelligence amplification systems optimize cognitive growth itself rather than immediate task reward?

Aim and Objectives

Aim

Develop a neuro-symbolic reinforcement learning framework that optimizes long-horizon cognitive growth.

Objectives

- Introduce Cognitive Growth Rate (CGR).
- Develop Cognitive Digital Twins (CDTs).
- Integrate RL with symbolic prerequisite reasoning.
- Evaluate delayed transfer emergence.
- Analyze structure-exploration trade-offs.

Exiting Work and Limitations

Existing Approach	Limitation
Reward-based RL	Myopic optimization under delayed rewards
Curiosity RL	No explicit transfer objective
Neuro-symbolic AI	Symbolic constraints may reduce exploration
Knowledge tracing	Predictive rather than optimization-driven

Methods

Proposed Framework

Cognitive Digital Twin (CDT)

The learner is modeled as a latent cognitive state:

$$C_t \in [0, 1]^K$$

where each dimension represents concept mastery within a structured prerequisite graph G.

The framework integrates:

- latent neural representations,
- symbolic knowledge graphs,
- reinforcement learning policies,
- intelligence-based optimization.

The reinforcement learning policy is defined as:

$$\pi_\omega(a_{t'} \mid C_t G)$$

5. Cognitive Growth Rate (CGR)

Instead of optimizing immediate reward, the proposed framework maximizes capability growth over time:

$$r_t^{CGR} = I(C_{t+1}) - I(C_t)$$

Optimization objective:

$$J_{CGR}(\pi) = \mathbb{E}\left[\sum_t \gamma^t \left(I(C_{t+1}) - I(C_t)\right)\right]$$

6 .Intelligence Functional

$$I(C) = \alpha_1 Transfer + \alpha_2 Coherence + \alpha_3 Balance + \alpha_4 Mastery - \alpha_5 Overspecialization$$

This formulation captures transfer capability, structural coherence, balanced progression, and long-horizon capability growth.

Cognitive Digital Twin for Lifelong Intelligence Amplification (CDT-LIA) Architecture

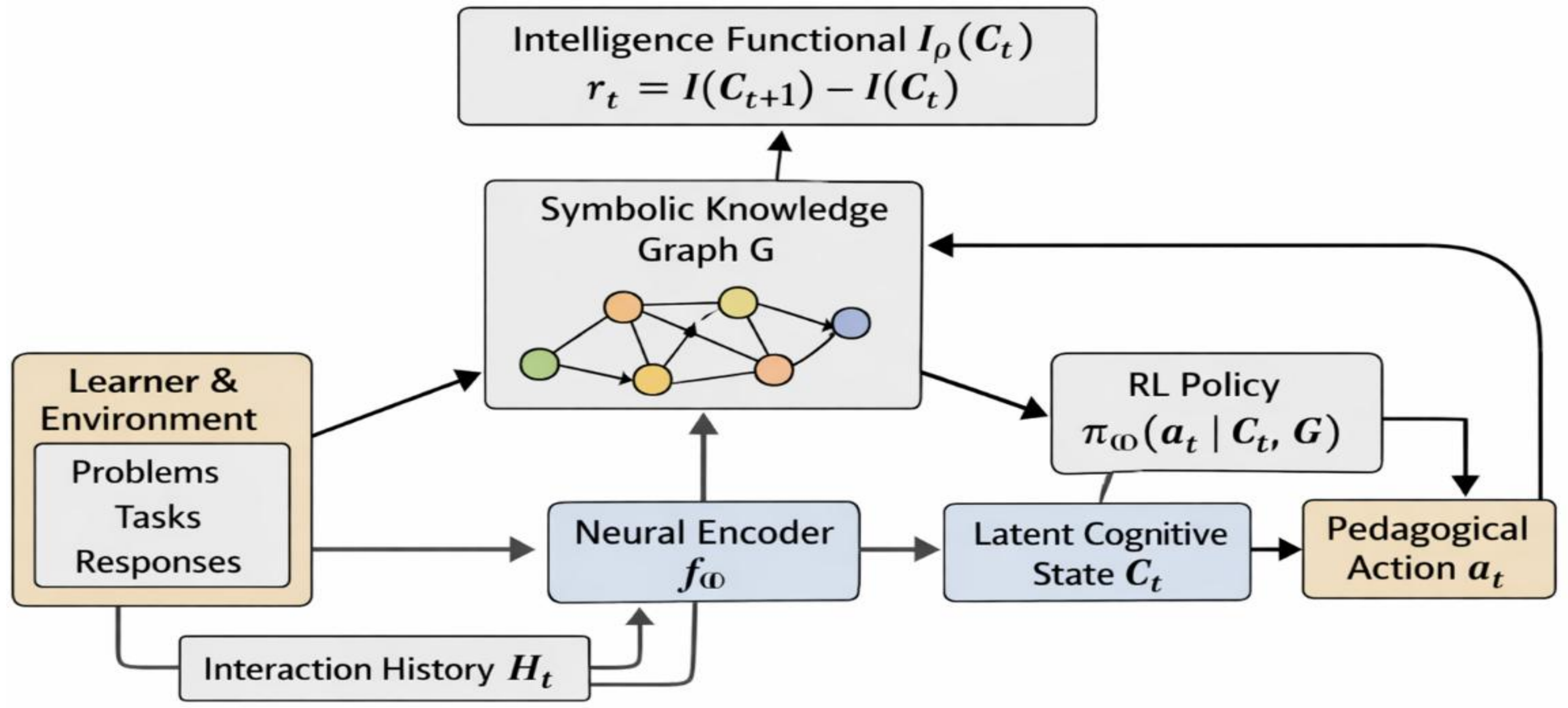

Fig. 1. CDT-LIA architecture: a neuro-symbolic reinforcement learning agent maintainis a latent cognitive state C_t, reasons over a knowledge graph G, and selects pedagogical actions to maximize Cognitive Growth Rate.

Caption: CDT-LIA integrates latent cognitive representations, symbolic prerequisite graphs, and reinforcement learning policies to optimize Cognitive Growth Rate (CGR).

8. Neuro-Symbolic RL Architecture

The policy is defined as:

$$\pi_\omega(a_{t'} \mid C_t G)$$

where C_t is the latent cognitive state and $G$is the symbolic prerequisite graph.

The framework combines:

Neural components: latent encoder and actor–critic policy.

Symbolic components: prerequisite graph and structural reasoning constraints.

Experimental Benchmark

- The benchmark environment includes:
- hierarchical prerequisite dependencies,
- competing learning pathways,
- delayed transfer emergence,
- cross-path interference,
- composite multi-concept tasks.

Methods Compared

- Random
- Greedy Immediate Gain
- Actor-Critic (performance reward)
- Actor-Critic (curiosity reward)
- Actor-Critic (CGR no-symbolic)
- Actor-Critic (CGR + symbolic)

Evaluation Metrics

- Transfer emergence
- Structural coherence
- Path balance
- Area Under Growth Curve (AUGC)

Results

Delayed Transfer Emergence

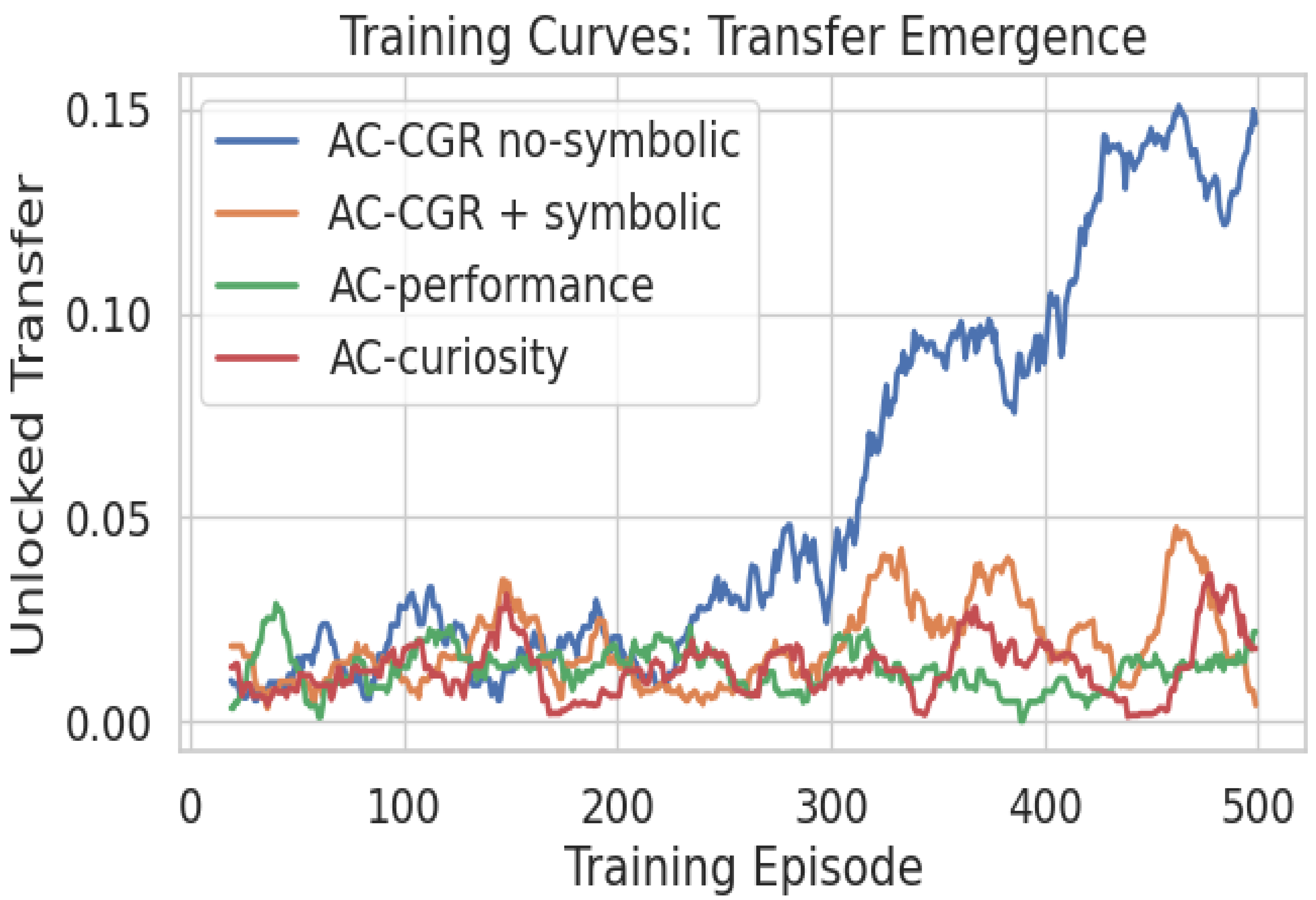

Fig 2: CGR-based optimization produces delayed long-horizon transfer emergence substantially exceeding conventional reward-based reinforcement learning objectives.

AUGC Training Dynamics

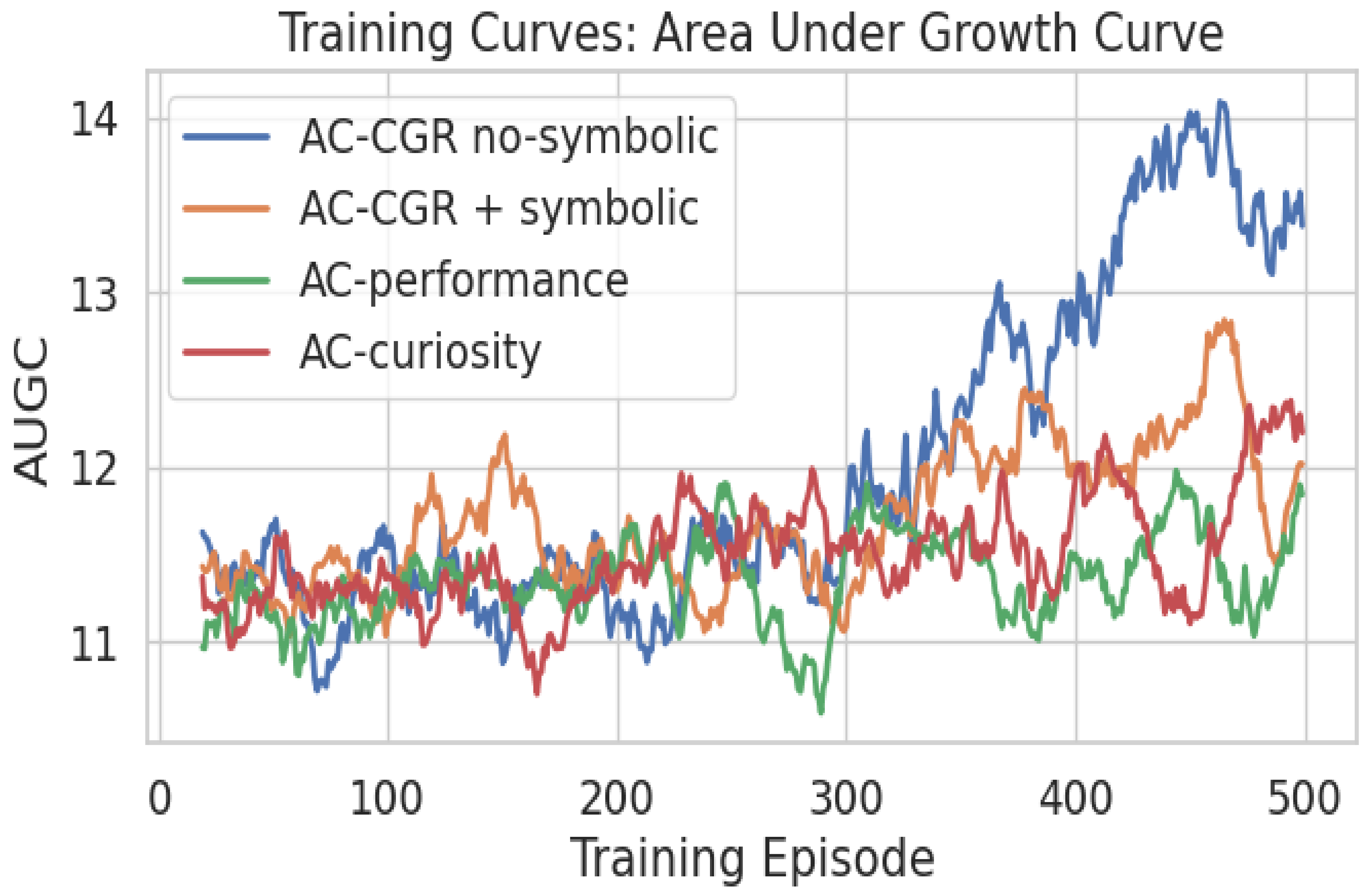

Fig 3: CGR policies sustain larger long-horizon capability accumulation throughout training.

Structural Metrics

Method	Coherence	Path Balance	AUGC
Random	0.110	0.859	11.182
AC-performance	0.149	0.853	11.455
AC-curiosity	0.208	0.837	11.797
AC-CGR + symbolic	0.166	0.932	12.667
AC-CGR no-symbolic	**0.317**	0.736	**13.782**

Transfer Performance

Method	Transfer
Random	0.012
AC-performance	0.017
AC-curiosity	0.025
AC-CGR + symbolic	0.025
AC-CGR no-symbolic	**0.150**
Greedy immediate gain	0.279

Key Findings

- Standard RL largely fails under delayed transfer conditions.
- CGR induces emergent long-horizon transfer behavior.
- Symbolic structure improves coherence and path balance.
- Excessive symbolic constraints may reduce exploratory transfer acquisition.
- Intelligence amplification may require optimizing developmental growth trajectories rather than immediate reward.

Discussion and Implications

The results suggest that optimizing cognitive growth produces learning dynamics fundamentally different from immediate reward maximization. CGR induces delayed transfer emergence and sustained long-horizon capability growth, while symbolic structure introduces a trade-off between coherence and exploratory learning. These findings have implications for intelligent tutoring systems, adaptive curricula, lifelong learning systems, and human-AI cognitive augmentation. However, the study is limited by its synthetic benchmark environment and manually designed intelligence functional; future work will explore real educational datasets, adaptive symbolic reasoning, stronger RL baselines, and dynamic knowledge graphs.

Intelligence may be better optimized as a developmental growth trajectory rather than static task performance.

Optimizing cognitive growth objectives produces qualitatively different long-horizon learning dynamics.

References

[1] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[2] D. Pathak, P. Agrawal, A. A. Efros, and T. Darrell, "Curiosity-driven exploration by self-supervised prediction," 2017: PMLR, pp. 2778-2787.

[3] D. Yu, B. Yang, D. Liu, H. Wang, and S. Pan, "A survey on neural-symbolic learning systems," *Neural Networks*, vol. 166, pp. 105-126, 2023.