Project Stargate Mk IV ("Hunter")

An Integral-Hybrid Cognitive Architecture for Sample-Efficient Visual Locomotion

Status: Exploratory Research Prototype
Author: Mubasher Ikram
Date: December 2025

Abstract

This paper presents **Stargate Mk IV**, an exploratory integral-hybrid cognitive architecture designed to study fast behavioral emergence in embodied agents. The system combines representation learning, dual-timescale world modeling, active inference–style planning, and a simple homeostatic drive mechanism. We evaluate the architecture in a visually guided continuous-control locomotion environment using raw pixel observations.

In a single exploratory experiment, the agent exhibits rapid emergence of forceful movement behaviors under a velocity-focused reward. The policy does not converge to a stable gait and frequently destabilizes, indicating strong exploration pressure and incomplete control learning. These results are **not benchmark-level** and are presented as early evidence of fast behavioral discovery rather than solved locomotion.

---

1. Motivation

Most modern reinforcement learning systems achieve locomotion by relying on large-scale training, dense rewards, and extensive hyperparameter tuning. While effective, these approaches are often sample-inefficient and brittle when transferred beyond narrowly defined tasks.

Biological organisms, by contrast, rapidly discover meaningful behaviors from limited experience, often before those behaviors become stable or optimal. Early movement is noisy, unstable, and exploratory, yet informative.

The goal of Stargate Mk IV is **not** to immediately solve locomotion, but to investigate whether integrating biologically inspired components can accelerate the *emergence* of non-trivial behavior from pixels.

---

2. Architectural Overview

Stargate Mk IV is organized into four interacting subsystems:

2.1 Perceptual Core (VQ-VAE)

A Vector-Quantized Variational Autoencoder compresses raw pixel observations into discrete latent representations. This provides: - Dimensionality reduction - Noise tolerance - A discrete symbolic-like latent space for downstream modeling

2.2 Dual-Process World Model

The system maintains two complementary internal models: - **Fast Model:** Updates rapidly, prioritizing recent observations and short-term adaptation. - **Slow Model:** Updates conservatively, capturing longer-term structure and reducing catastrophic drift.

This separation allows the agent to adapt quickly without fully overwriting prior internal structure.

2.3 Cognitive Layer (Active Inference–Inspired Planner)

Action selection is framed as minimizing expected future surprise (expected free energy) under the learned world model. Rather than directly optimizing reward, the planner balances: - Predicted outcomes - Goal-directed preferences - Uncertainty reduction

This component operates on latent representations rather than raw pixels.

2.4 Autonomic Layer (Homeostatic Drive)

A simple internal energy variable decays over time and is replenished through movement-related reward. This mechanism: - Forces continuous action - Discourages inactivity - Encourages exploration even in the absence of stable control

---

3. Experimental Setup

3.1 Environment

- Continuous-control quadruped locomotion task
- Observations: RGB pixel input
- Actions: Continuous joint torques
- Episode termination disabled to allow recovery attempts

3.2 Reward Structure

- Primary reward emphasizes forward velocity
- No explicit penalties for instability or energy usage
- No curriculum learning or shaping

3.3 Training Conditions

- Single random seed

- Single training run
- No hyperparameter sweep
- No baseline agent trained in parallel
- Hardware: Single NVIDIA T4 GPU (Google Colab free tier)

---

4. Observed Behavior

During training, the agent rapidly discovers strategies that generate strong forward acceleration. Early in learning, this results in aggressive movement patterns that frequently destabilize the body.

Notable characteristics: - Rapid onset of high-energy motion - Frequent flips and falls following acceleration - Persistent attempts to recover after falling - High variance reward signal with intermittent spikes

The agent does **not** learn a stable gait in this experiment. Instability dominates behavior, particularly as reward incentives favor speed over balance.

---

5. Results and Interpretation

The primary signal observed is **speed of behavioral discovery**, not control quality.

Within a relatively small number of environment interactions, the agent transitions from near-random motion to forceful, directed movement. While this movement is unstable, it reflects non-trivial coordination emerging from pixel input.

These results should be interpreted as: - Evidence of strong exploration pressure - Early-stage motor pattern formation - A prototype validation of architectural integration

They should **not** be interpreted as: - Solved locomotion - Stable control - Superiority over established RL baselines

---

6. Limitations

This work has significant limitations: - Single-seed experiment - No comparison against PPO, SAC, or other baselines - No stability regularization or recovery shaping - No evaluation of generalization - No statistical significance claims

The current system is unsuitable for deployment or performance benchmarking.

---

7. Future Work

Key next steps include: - Multi-seed evaluation - Inclusion of explicit stability and energy penalties - Curriculum learning for recovery behaviors - Direct baseline comparisons - Transfer to additional morphologies

---

8. Conclusion

Stargate Mk IV represents an exploratory step toward biologically inspired cognitive architectures for embodied agents. While the current results are unstable and incomplete, they suggest that integrating perception, world modeling, planning, and homeostatic drives may accelerate the early emergence of meaningful behavior from raw sensory input.

This work is intended as a research prototype and architectural study, not a finished system.

Appendix A: Ethical and Scope Statement

This project was developed independently as an educational and research exercise. No claims are made regarding safety, deployment readiness, or commercial viability in its current form.