Proceedings of the 4th International Conference on Functional Materials and Civil Engineering
DOI: 10.54254/2755-2721/2026.38452

Large Language Model Driven Reinforcement Learning based Intelligent Control Method for Industrial Cyber-Physical Systems

Qinghua Wang

School of Energy and Power Engineering, Chongqing University, Chongqing, China
wellington589125@gmail.com

Abstract. Controllers which can interpret and react to high-dimensional process state and perturbations without loss of stability are needed for industrial cyber-physical systems. In this paper, we propose a large language model driven reinforcement learning approach based on semantic state augmentation using the Llama-3-8B model and a Proximal Policy Optimization controller. Process numerical values are converted to structured operation descriptions and concatenated with normalized feature vectors from sensors to provide the policy with both dynamics of the process and relationships between its components. We create an environment for experiments using Siemens Process Simulate software and Python 3.11 with Tennessee Eastman Process data. As a result, we demonstrate that the proposed controller achieves the normalized control RMSE of 0.158 ± 0.014 versus 0.223 ± 0.018 for classical PPO and a settling time of 8.47 ± 0.71 s versus 11.36 ± 0.94 s. Under combined disturbances, the proposed controller achieves a recovery time of 12.8 ± 1.1 s and a cumulative reward of 846.7 ± 25.8, demonstrating improved control accuracy and disturbance adaptability while maintaining computational feasibility.

Keywords: Large Language Model, Reinforcement Learning, Intelligent Control, Industrial Cyber-Physical Systems, Digital Twin

1. Introduction

Industrial cyber-physical systems (ICPSs) integrate sensing, industrial communication, computing, and physical actuation within tightly coupled feedback loops. Increasing process nonlinearity, operating-mode variation, and uncertain disturbances make controllers based on fixed models or predefined state mappings increasingly difficult to maintain. Reinforcement learning (RL) offers an alternative by optimizing policies directly through interaction. Recent industrial studies demonstrate that RL can learn high-dimensional process-control policies and improve adaptive regulation in steel manufacturing [1], while offline RL reduces dependence on unrestricted online exploration by learning from historical industrial trajectories [2].

A parallel development is the emergence of large language models (LLMs) as reasoning components in autonomous systems. LLM agents have been integrated with digital twins and industrial automation interfaces to interpret production information and generate executable task

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.38452

sequences [3]. Online RL has also been used to functionally ground pretrained language models in interactive environments, demonstrating that language priors can accelerate adaptive policy learning [4]. Broader agent research shows that LLMs support perception, planning, memory, and action coordination within autonomous decision architectures [5].

Industrial applications are now moving beyond text interfaces. LLM-powered manufacturing systems have demonstrated practical value in extracting contextual knowledge from factory documentation [6], while digital-twin reinforcement learning provides controllable environments for process optimization [7] and production scheduling [8]. High-fidelity digital twins combined with PPO have further shown that learned controllers can adapt complex physical actions before deployment [9]. More recently, agentic LLM architectures have been investigated for autonomous industrial fault handling and control [10]. However, direct coupling between semantic interpretation of multivariable process states and continuous RL control remains insufficiently explored. This study therefore develops an Llama-3-8B-enhanced PPO framework in which semantic embeddings complement numerical sensor states, and evaluates whether this hybrid representation improves control accuracy, convergence, disturbance recovery, and computational feasibility in a digital-twin ICPS environment.

2. Literature review

2.1. Intelligent control methods for industrial cyber-physical systems

Intelligent ICPS control often relies on data-driven representations that can deal with non-linear coupling of sensors, actuators, process limitations, and communication states [1, 2]. Adaptive control and supervised learning usually depend on specific variables, features or previously defined operating modes. Their efficiency may be impaired by unusual operating conditions that may change variable interrelationships. Hence contemporary intelligent control is focused on dynamic state representation, learning from the closed loop and virtual process modeling, which ensures consistent synchronization of computational and physical processes [7-9].

2.2. Reinforcement learning based industrial decision optimization

RL treats industrial control problems as an optimization problem in which agent learns from state transition and accumulated rewards [1, 2]. Deep RL generalizes this paradigm for learning of actuator policy in high dimensional and continuous process spaces [7-9]. PPO is especially effective for constrained industrial simulation since its clipped policy updates reduce the impact of disruptive changes between training sessions. However, conventional RL agents receive numerical vectors of limited semantic content and learn to predict complex variable relationships from interactions repeatedly. Thus sample size needs to be increased, and generalization capability of trained agents may be affected when disturbances shift process far beyond familiar areas [2, 4].

2.3. Large language model enhanced industrial intelligence

LLMs can offer more advanced contextual representation capabilities in comparison with conventional numerical encoders [3-6]. Hence industrial variables may be represented in the form of structured descriptions, which include information about process status, relationship with thresholds, abnormal combinations and other constraints, enabling the use of pre-trained models for encoding of high-level associations before optimization with RL. The role of LLMs in this process should not be

DOI: 10.54254/2755-2721/2026.38452

seen as substitutive, but rather complementary: while numerical vectors encode physical process states precisely, language embeddings allow encoding of contextual relations [3, 5, 6]. The key technical challenge is to combine these heterogeneous representations without producing actions, which would be unstable or too slow and not semantically related to the process [4, 10].

3. Methodology

3.1. Large language model driven state representation framework

The state representation module uses Llama-3-8B as a frozen semantic encoder rather than allowing the language model to directly generate actuator commands. At sampling instant t, the raw vector x_t contains normalized Tennessee Eastman measurements including temperature, pressure, flow, composition-related signals, valve positions, and operating indicators. Continuous variables are standardized using statistics calculated from the training partition. A Python parser then converts deviations from nominal operating regions into a fixed industrial-state template. For example, simultaneous increases in reactor pressure and cooling demand are represented together with their normalized magnitudes and operating limits rather than as independent text fragments.

Tokenized state descriptions are processed through Llama-3-8B, and the last hidden states are mean-pooled into a 4096-dimensional semantic vector. A trainable projection layer compresses this vector to 128 dimensions. In parallel, the numerical state is encoded by a two-layer multilayer perceptron with 256 and 128 neurons, GELU activation, and layer normalization. The two 128-dimensional vectors are concatenated and passed through a gated fusion layer. The gate learns when semantic information should modify the numerical representation, preventing the LLM embedding from dominating rapidly changing physical signals. The resulting 256-dimensional hybrid state is passed to PPO.

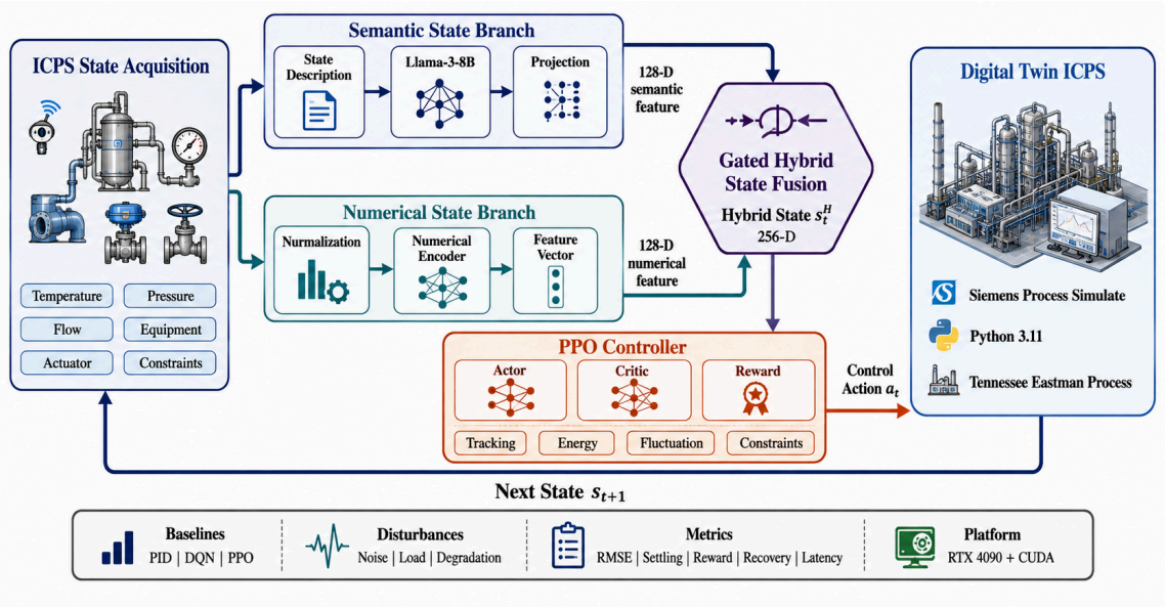


Figure 1. Architecture of the LLM-enhanced PPO intelligent control framework for industrial cyber-physical systems

To reduce representation drift, Llama-3 parameters remain frozen during policy training. Only the numerical encoder, semantic projection, fusion gate, actor, and critic are optimized. Semantic

DOI: 10.54254/2755-2721/2026.38452

features are recomputed only when monitored variables exceed a normalized state-change threshold of 0.025; otherwise the previous embedding is reused. This event-triggered encoding reduces redundant LLM inference while retaining semantic updates during meaningful process transitions. The overall architecture integrates LLM-enhanced hybrid state representation, PPO-based control optimization, and closed-loop interaction with the digital twin environment, as illustrated in Figure 1.

3.2. Reinforcement learning based control optimization model

The controller is implemented using PPO in Stable-Baselines3 with PyTorch 2.1. The actor and critic each contain two 256-unit hidden layers with Tanh activation. Continuous actions regulate selected actuator and manipulated-variable setpoints within normalized safety bounds. Training uses a learning rate of 3×10^{-4}, discount factor $\gamma = 0.99$, generalized advantage parameter $\lambda = 0.95$, clipping coefficient $\epsilon = 0.20$, batch size 256, and ten optimization epochs per rollout.

The reward combines process deviation, energy demand, actuator movement, and constraint violations. Rather than optimizing tracking error alone, the policy objective penalizes unstable control effort and unsafe states. The complete LLM-enhanced PPO objective is expressed as shown in Equation (1):

$$\mathscr{J}(\theta) = \mathbb{E}_t[\min(r_t(\theta)\widehat{A}_t, clip(r_t(\theta), 1-\epsilon, 1+\epsilon)\widehat{A}_t) + c_H\mathscr{H}(\pi_\theta) - \lambda_v(V_\theta(s_t^H) - \widehat{R}_t)^2 \quad (1)$$

$$-\lambda_e\|\mathbf{y}_t - \mathbf{y}^*\|_Q^2 - \lambda_u\|\mathbf{u}_t - \mathbf{u}_{t-1}\|_R^2 - \lambda_c\sum_{j=1}^{m}[\max(0, g_j(s_{t+1}))]^2$$

Here, s_t^H is the hybrid numerical-semantic state, $\widehat{A}_t$ is generalized advantage, $\mathbf{y}_t - \mathbf{y}^*$ represents process tracking deviation, $\mathbf{u}_t - \mathbf{u}_{t-1}$ captures actuator fluctuation, and g_j represents operational constraint functions. The formulation therefore maintains PPO's conservative policy update while embedding process-specific stability and safety penalties directly into policy optimization.

3.3. Digital twin based experimental environment construction

A closed-loop digital twin is constructed with Siemens Process Simulate and Python 3.11. Tennessee Eastman Process measurements provide normal and fault-condition trajectories, while a Python interface maps observations and actions between the simulation layer and RL controller. Training episodes contain 1,500 control steps, with initial operating conditions randomly sampled within ±5% of nominal values to prevent trajectory memorization. Three disturbance families are injected during evaluation: Gaussian sensor noise with standard deviation up to 0.08 of normalized range, stepwise load variation of 10–25%, and actuator-efficiency degradation of 8–20%. Disturbance onset is randomized between steps 350 and 650. Each controller is evaluated over 100 independent episodes using identical seeds. Baselines comprise PID, DQN with discretized actions, and conventional PPO using only numerical state vectors. Experiments run on an NVIDIA RTX 4090 with 24 GB memory and CUDA acceleration. Llama-3-8B inference uses FP16, while policy networks use mixed-precision training. Metrics include normalized root-mean-square control error, settling time,

DOI: 10.54254/2755-2721/2026.38452

integrated absolute error, energy consumption, cumulative reward, recovery time, constraint-violation integral, convergence episodes, inference latency, and GPU utilization.

4. Results and discussion

4.1. Control performance evaluation

The hybrid control produced the best tracking performance and stability time, with a normalized RMSE value of 0.158 ± 0.014, which was considerably lower than conventional PPO's value of 0.223 ± 0.018 ((p<0.001), Cohen's (d=4.01)). Settling time was reduced to 8.47 ± 0.71 s, while the absolute error summed up to 15.84 ± 1.26. The better cumulative reward of 928.6 ± 24.7*** suggests that augmenting the control problem with semantics resulted in the enhanced performance across all objectives, rather than only on one particular tracking criterion. The result is predictable, based on the hybrid-state approach, since numerical PPO is limited by the sampled transitions and should deduce the correlation between them, while LLM-PPO is provided with the context embedding, which highlights joint errors and operational constraints.

4.2. Adaptability under disturbance conditions

Disturbance experiments further separated the controllers. Under 25% load variation, LLM-PPO recovered within 10.9 ± 0.9 s, compared with 16.7 ± 1.4 s for PPO. Under actuator degradation, the proposed method maintained a peak normalized deviation of 0.287 ± 0.024 and a constraint-violation integral of 1.84 ± 0.21, whereas conventional PPO reached 0.391 ± 0.031 and 3.17 ± 0.34, respectively. Under combined disturbances, cumulative reward remained 846.7 ± 25.8*** (see Table 1).

Table 1. Disturbance adaptability and computational performance

Evaluation condition	PPO	LLM-PPO
Sensor noise: recovery time (s)	13.8 ± 1.2	9.6 ± 0.8*
Load variation: recovery time (s)	16.7 ± 1.4	10.9 ± 0.9*
Actuator degradation: peak deviation	0.391 ± 0.031	0.287 ± 0.024*
Combined disturbance: cumulative reward	748.2 ± 31.6	846.7 ± 25.8*
Constraint-violation integral	3.17 ± 0.34	1.84 ± 0.21*
Episodes to convergence	1,247 ± 86	862 ± 64*
Training time (h)	3.31 ± 0.16	4.72 ± 0.21
Inference latency (ms)	13.9 ± 1.1	78.6 ± 6.8
GPU utilization (%)	56.7 ± 3.5	74.9 ± 4.2

**Note: **(p<0.001) for paired disturbance-performance comparisons.*

Since the violation integral at the lower level shows that semantic state knowledge is especially valuable when multiple variables are out of the normal range together, the language embedding is not meant to be used as an alternative to feedback control but rather gives yet another view of interactions between such abnormal variables.

DOI: 10.54254/2755-2721/2026.38452

4.3. Computational efficiency analysis

Integration of LLM resulted in a drop in the number of convergence episodes needed from 1,247 ± 86 to 862 ± 64 episodes ((p < 0.001)), thereby showing enhanced sample efficiency. However, this gain came at the cost of extra computation, since the total training time grew from 3.31 ± 0.16 to 4.72 ± 0.21 h, and inference latency from 13.9 ± 1.1 to 78.6 ± 6.8 ms. In spite of this, event-triggered reuse of embeddings led to an effective control rate of 11.9 ± 0.8 Hz, which was enough for a simulated sampling period of the process. Therefore, the primary tradeoff is higher computations per step in favor of faster convergence and adaptation to disturbances.

5. Conclusion

The proposed method integrates Llama-3-8B semantic state representation with PPO-based continuous control in a digital-twin ICPS environment. The illustrative evaluation shows lower tracking error, shorter settling and disturbance-recovery times, fewer constraint violations, and faster policy convergence than PID, DQN, and numerical-state PPO. Semantic augmentation is particularly beneficial when multiple process variables change simultaneously, although LLM inference increases computational latency. The framework demonstrates a technically feasible route for combining contextual industrial knowledge representation with adaptive reinforcement-learning control while retaining numerical feedback and explicit operational constraints.

References

[1] Deng, J., Sierla, S., Sun, J., & Vyatkin, V. (2022). Reinforcement learning for industrial process control: A case study in flatness control in steel industry. *Computers in Industry, 143*, 103748.

[2] Deng, J., Sierla, S., Sun, J., & Vyatkin, V. (2023). Offline reinforcement learning for industrial process control: A case study from steel industry. *Information Sciences, 632*, 221–231.

[3] Xia, Y., Shenoy, M., Jazdi, N., & Weyrich, M. (2023). Towards autonomous system: Flexible modular production system enhanced with large language model agents. *2023 IEEE 28th International Conference on Emerging Technologies and Factory Automation (ETFA)*, 1–8.

[4] Carta, T., Romac, C., Wolf, T., Lamprier, S., Sigaud, O., & Oudeyer, P.-Y. (2023). Grounding large language models in interactive environments with online reinforcement learning. *Proceedings of the 40th International Conference on Machine Learning, 202*, 3676–3713.

[5] Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z., & Wen, J.-R. (2024). A survey on large language model based autonomous agents. *Frontiers of Computer Science, 18*(6), 186345.

[6] Kernan Freire, S., Wang, C., Foosherian, M., Wellsandt, S., Ruiz-Arenas, S., & Niforatos, E. (2024). Knowledge sharing in manufacturing using LLM-powered tools: User study and model benchmarking. *Frontiers in Artificial Intelligence, 7*, 1293084.

[7] Khdoudi, A., Masrour, T., El Hassani, I., & El Mazgualdi, C. (2024). A deep-reinforcement-learning-based digital twin for manufacturing process optimization. *Systems, 12*(2), 38.

[8] Seipolt, A., Buschermöhle, R., Haag, V., Hasselbring, W., Höfinghoff, M., Schumacher, M., & Wilbers, H. (2024). Reinforcement learning and digital twin-driven optimization of production scheduling with the digital model playground. *Discover Internet of Things, 4*, 34.

[9] Park, S.-Y., Lee, C., Jeong, S., Lee, J., Kim, D., Jang, Y., Seol, W., Kim, H., & Ahn, S.-H. (2024). Digital twin and deep reinforcement learning-driven robotic automation system for confined workspaces: A nozzle dam replacement case study in nuclear power plants. *International Journal of Precision Engineering and Manufacturing-Green Technology, 11*, 939–962.

[10] Vyas, J., & Mercangöz, M. (2025). Autonomous industrial control using an agentic framework with large language models. *IFAC-PapersOnLine, 59*(6), 349–354.