PPI: A Co-Evolution Framework for Parallel Physical Intelligence

Xinyuan Zhang, Jiaxi Liu, Siheng Yi, and Hui Zhang

***Abstract*—Physical artificial intelligence (AI) is becoming an important paradigm for embodied intelligence, where agents must reason, imagine, and act under real-world physical constraints. However, current physical reasoning model (PRM) and physical generation model (PGM) may still produce physically inconsistent judgments, scenes, and motions, exposing the lack of a verifiable truth anchor for reliable physical intelligence. Existing simulation-based methods can provide local physical supervision, but their limited coverage makes them insufficient for open and evolving physical environments. To address this bottleneck, this letter proposes parallel physical intelligence (PPI), a truth-anchored co-evolution framework based on parallel intelligence and the artificial systems, computational experiments, and parallel execution (ACP) methodology. PPI introduces the physical fidelity anchor (PFA) to calibrate both PRM and PGM, while enabling PRM and PGM to co-evolve through mutual data augmentation and physical plausibility feedback. During parallel execution, embodied agents and disembodied agents interact through a virtual-real closed loop, where real-world failures and novel physical phenomena update the PFA and trigger subsequent co-evolution. This provides a scalable and adaptive pathway for physically grounded embodied intelligence in complex environments.**

I. Introduction

Physical artificial intelligence (AI) has emerged as a next-generation AI paradigm that grounds intelligence in physical systems [1, 2]. Unlike conventional AI models mainly optimized for symbolic or semantic tasks, physical AI requires the integration of physical perception, causal reasoning, and embodied interaction under real-world constraints [3]. Such capability is essential for autonomous driving [4], unmanned aerial vehicle (UAV) navigation [5], robotic manipulation [6], and surgical intervention [7], where agents must infer physical properties, predict dynamics, and act safely in uncertain environments. Nevertheless, current AI systems often generate semantically plausible but physically inconsistent results, making physical grounding a key bottleneck for reliable embodied intelligence.

Current efforts toward physical intelligence mainly proceed through two model-centric directions: multimodal large language model (MLLM)-based physical reasoning and diffusion-based physical generation. MLLM-based physical reasoning has shown preliminary physical commonsense, but it is not explicitly grounded in mechanical laws and therefore remains unreliable in robust physical perception, forecasting, and causal reasoning [8, 9]. Diffusion-based physical generation can synthesize realistic scenes and motions, but it remains difficult to guarantee the physical consistency required for reliable reasoning and embodied decision-making [10, 11]. Physics simulators provide controllable physical ground truth, yet their coverage is constrained by predefined object models, interaction rules, and parameter ranges. However, these components are often studied as separate modules rather than organized into a unified physical intelligence framework. The absence of a unified truth anchor and closed-loop collaboration mechanism makes it difficult to cross-calibrate reasoning, generation, and simulation, thereby constraining the further evolution of physical intelligence.

Addressing this limitation requires a system-level organization of physical reasoning, physical generation, and physical simulation, rather than the independent enhancement of each component. Parallel intelligence (PI), together with the artificial systems, computational experiments, and parallel execution (ACP) methodology, provides a suitable theoretical basis for such organization by integrating virtual modeling, computational exploration, and real-world feedback into a closed loop [12]. Following this principle, we propose parallel physical intelligence (PPI), a truth-anchored co-evolution framework in which the artificial system is formulated as a physical fidelity anchor (PFA). The PFA is not intended to fully replicate the physical world as a digital twin; instead, it functions as a task-schema-based physical data engine that generates executable physical cases, produces physical and observational rollouts, and provides verified labels and diagnoses for key interaction laws and constraints. These PFA-grounded cases calibrate the physical reasoning model (PRM) and the physical generation model (PGM) at the model-training level, rather than merely correcting outputs after inference. After calibration, PRM and PGM further co-evolve in computational experiments: PGM expands the physical hypothesis space by generating diverse and counterfactual rollout candidates, while PRM evaluates physical plausibility, causal continuity, and constraint satisfaction to provide diagnostic feed-

Manuscript received: 11 May 2026; revised: 21 May 2026; accepted: 25 May 2026. *(Corresponding author: Hui Zhang.)*

Citation: X. Zhang, J. Liu, S. Yi, and H. Zhang, PPI: A co-evolution framework for parallel physical intelligence, *Int. J. Intell. Control Syst.*, 2026, 31(2), 225–232.

Xinyuan Zhang is with School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: zhangxinyuan23@mails.ucas.ac.cn).

Jiaxi Liu is with School of Social Sciences, University of Manchester, Manchester M13 9PL, UK (e-mail: jiaxi.liu-3@postgrad.manchester.ac.uk).

Siheng Yi is with Decentralized Science Center of Parallel Intelligence, Óbuda University, Budapest H-1034, Hungary (e-mail: sihengyi539@gmail.com).

Hui Zhang is with School of Computer Science and Technology, Beijing Jiaotong University, Beijing 100044, China (e-mail: huizhang1@bjtu.edu.cn).

Digital Object Identifier 10.62678/IJICS202606.10372

back for generation. High-impact, uncertain, or disputed cases are routed back to the PFA for verification, thereby preserving a verifiable truth anchor. During parallel execution, PPI serves as a physical intelligence layer over existing planners or controllers, where PGM provides future imagination, PRM provides physical judgment, and real-world failures or novel physical phenomena update the PFA to trigger the next round of co-evolution. Accordingly, PPI provides a scalable and adaptive pathway for physically grounded embodied intelligence in open physical environments.

The main contributions of this letter are as follows:

(1) We propose PPI, a PRM-PGM-PFA architecture that formulates PFA as a physical data engine for truth-anchored physical intelligence.

(2) We design a PFA-driven calibration and PRM-PGM co-evolution mechanism for jointly improving physical reasoning and physical generation.

(3) We establish a virtual-real closed-loop deployment paradigm that updates PFA through embodied feedback for continual physical intelligence evolution.

II. Preliminary

A. Physical Reasoning Models

Physical reasoning models, often built upon or coupled with vision-language models (VLMs)/MLLMs, aim to infer task-relevant physical factors from visual or multimodal observations. These factors can be organized into static and dynamic categories. Static physical reasoning concerns relatively time-invariant properties and constraints, including object-level physical properties [13], affordance structures [14, 15], geometric-spatial relations [16, 17], and embodiment-conditioned feasibility [18, 19]. Dynamic physical reasoning, in contrast, concerns temporally evolving physical processes, including motion dynamics [20, 21], collision outcomes [22], quantitative kinematic variables [23], continuum or interaction process evolution [24, 25], and causal reasoning [26].

To mitigate these limitations, existing studies introduce physical grounding into PRMs through data construction, inference-time augmentation, and representation design. Physically annotated training resources have been developed to improve object-property understanding, spatial reasoning, affordance recognition, and embodiment-conditioned feasibility estimation [13, 14, 17, 19]. External grounding signals have also been incorporated during reasoning or planning, including specialized visual experts, skill-level value functions, three-dimensional (3D) value maps, and simulator-in-the-loop parameter estimation for constraining physical reasoning and action selection [15, 18]. In addition, representation-level methods embed physical priors into the perception-reasoning interface, such as reachability maps, spatial-temporal cues, and motion-aware dynamics [19–21].

B. Physical Generative Models

Physical generative models, often built upon diffusion-based video generation models or world models, aim to synthesize or predict task-relevant physical scene evolution from visual, linguistic, or action-conditioned inputs. Representative large-scale models such as Sora, Genie, and related video and world foundation models demonstrate the potential of generative modeling for dynamic physical environments [27, 28]. However, recent benchmarks indicate that current video and world generative models still struggle with physical commonsense [29, 30], physical-law consistency in contact dynamics and conservation principles [10, 31], and instruction-conditioned or action-conditioned consistency [31], even when their outputs are visually plausible.

Recent efforts have improved the physical reliability of visual generative world models through several complementary strategies. Some methods adapt pretrained video generators with physics-aware post-training or embodied-domain fine-tuning [32, 33]. Others modify the generative representation or conditioning interface by introducing action prediction, physical latent variables, velocity masks, or force-control signals for object-level dynamics and interactions [34–37]. A third line learns generative physics modules that predict physically grounded trajectories before video synthesis [38]. In parallel, simulation-centered pipelines combine perception, material or structure estimation, physical simulation, and rendering to produce physics-compliant video or four-dimensional (4D) content [39–41].

Taken together, PRMs and PGMs reveal a shared bottleneck: both require physically grounded supervision, but in different forms. PRMs need verified physical labels and causal outcomes for reliable reasoning, whereas PGMs need physically consistent trajectories and constraint signals for reliable generation. Existing methods improve either reasoning or generation through annotations, external grounding, architectural priors, or simulation-aided refinement, but these signals are often fragmented and not organized into a unified calibration loop. Therefore, the key challenge is not merely to impose mechanical laws on individual models, but to establish a unified physical data lifecycle that continuously generates, verifies, filters, reuses, and updates physically valid reasoning and generation cases across training, inference, and deployment.

C. Physical Simulation as a Truth Anchor

Physical simulation provides controllable environments and executable supervision for anchoring physical reasoning and generation. Classical engines such as MuJoCo [42], Isaac Lab [43], and Brax [44] support rigid-body dynamics, contact modeling, actuator simulation, and scalable policy training, while differentiable simulators such as DiffTaichi [45] and ChainQueen [46] further allow gradients to be propagated through physical dynamics. Recent studies show that explicit physical constraints and simulation feedback can improve local reasoning and world modeling: PhysVLM [19] incorporates robotic reachability constraints into embodied visual reasoning, LLMPhy [25] couples LLM program synthesis with simulator-in-the-loop estimation of latent physical parameters for zero-shot multi-body reasoning, and physics-informed world models [47] or physics-integrated representations [39] introduce differentiable dynamics or Newtonian physical simulation to improve dynamic prediction and generative motion modeling. However, these efforts remain task-specific and fragmented, while simulators themselves are

constrained by fixed interfaces, predefined physical models, and manually designed parameterizations, limiting their generalization to unmodeled physical effects and open-ended real-world phenomena. This motivates the PFA perspective adopted in the proposed framework, where simulation is not regarded as a complete replacement for the physical world, but as a physical fidelity anchor that provides hard physical supervision for jointly calibrating PRMs and PGMs within its valid domain.

D. PI and the ACP Method

Parallel intelligence, grounded in the ACP method of artificial systems, computational experiments, and parallel execution, provides a methodological foundation for organizing complex intelligent systems through virtual-real feedback and co-evolution [12, 48]. In ACP, artificial systems abstract task-relevant counterparts of actual systems, computational experiments explore and optimize candidate models or policies in controllable virtual spaces, and parallel execution links virtual and real systems through continuous feedback and iterative deployment [49]. Compared with digital twins that emphasize strict one-to-one replication, PI focuses on key mappings, dynamic feedback, and virtual-real co-evolution.

Recent PI-based frameworks instantiate ACP for different purposes. Parallel deep foundation model (PDFM) supports continuous foundation-model evolution through front-end and shadow models [50]; PRALLM applies PI to public-private-personal knowledge integration through a parallel retrieval-augmented generation (RAG)-agent-LLM architecture [51]; and PiVLA reconstructs VLA training and deployment through artificial scenarios, disembodied agents, real scenarios, and embodied agents [52]. The distinction of PPI lies in the role assigned to the artificial system. Rather than using the artificial system mainly for model lifecycle evolution, knowledge integration, or VLA policy exploration and deployment, PPI constructs it as a physical fidelity anchor, making task-relevant physical validity the central criterion of virtual-real co-evolution. In this sense, PPI complements existing PI frameworks by providing a physical grounding layer for reliable embodied intelligence.

III. Framework

The overall architecture of the proposed PPI framework is shown in Fig. 1. PPI organizes physical reasoning, physical world generation, and physical simulation into a truth-anchored virtual-real co-evolution process. In the artificial physical space, the PFA is instantiated as a schema-guided physical data engine that generates physical source data for model-level calibration. The generated source data are converted into reasoning supervision for PRM and rollout supervision for PGM. In computational experiments, PRM and PGM are further refined through a bidirectional training-data exchange loop, where PGM provides generated rollouts and hard cases for PRM, while PRM provides plausibility judgments, diagnoses, and preference signals for PGM. During parallel execution, embodied agents deploy the current physical intelligence layer, and real-world logs, failures, and novel physical effects are fed back to update the PFA.

Following the ACP methodology, PPI instantiates parallel intelligence through three coupled processes. The artificial system is constructed as a PFA-based artificial physical system, which does not aim to fully replicate the real world but preserves task-relevant physical laws, interaction constraints, dynamic properties, and evaluation rules. Within

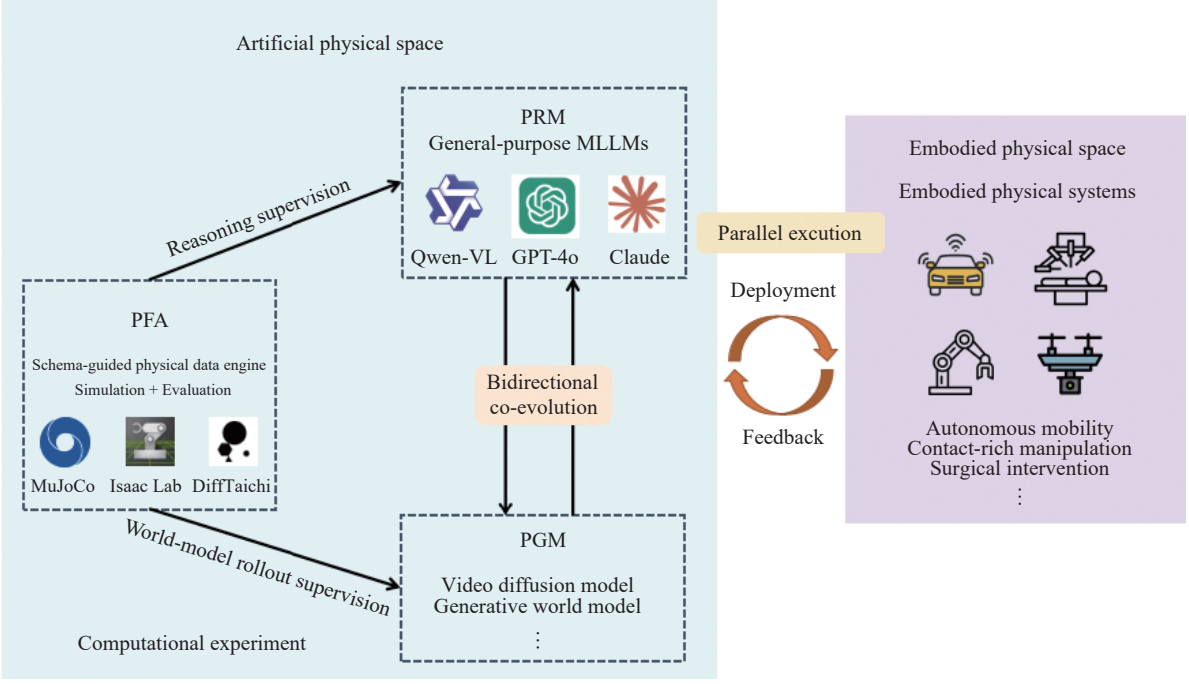


Fig. 1 Overall architecture of the proposed PPI framework. The PFA serves as a schema-guided physical data engine in the artificial physical space, providing reasoning supervision for PRM and world-model rollout supervision for PGM. PRM and PGM are refined through bidirectional co-evolution, while deployment feedback from embodied physical systems updates the artificial physical space through the ACP loop.

this artificial physical system, computational experiments are performed by disembodied agents: PFA-generated cases first calibrate PRM and PGM, and then PRM and PGM co-evolve through bidirectional data augmentation, plausibility evaluation, and hard-case verification.

Parallel execution connects the artificial physical system with embodied physical systems. During deployment, PPI serves as a physical intelligence layer over existing planners or controllers, where PGM provides action-conditioned future imagination and PRM provides physical feasibility and risk judgment. Real-world failures, rare events, and prediction-execution discrepancies are fed back to update the PFA, thereby triggering the next round of calibration and co-evolution. In this way, PPI realizes a closed-loop parallel physical intelligence process across artificial construction, computational experimentation, and embodied execution.

A. PFA as a Physical Data Engine

PFA starts from a task schema space $\Omega_{\mathcal{T}}$. Each schema specifies the task goal, embodiment setting, state and parameter space, action space, constraint and evaluation rules, and the corresponding simulator-renderer backend. Concrete physical cases are generated by sampling

$$\mathcal{T}_i \sim \Omega_{\mathcal{T}}, \quad x_{0,i} \sim P_X(\cdot|\mathcal{T}_i), \quad a_i \sim P_A(\cdot|\mathcal{T}_i, x_{0,i}),$$

where $\mathcal{T}_i$ is a concrete task setting, $x_{0,i}$ is the initial physical condition, and a_i is a candidate action. The initial-condition generator P_X can be implemented by task templates, domain randomization, real-log replay, or hard-case resampling, while the action generator P_A can be implemented by action primitives, planners, controllers, scripted policies, real executed actions, or hard-case actions.

For each generated triplet, PFA constructs an executable physical case and rolls it out through a task-specific simulator-renderer backend

$$\tau_i = (\tau_i^{\text{phy}}, \tau_i^{\text{obs}}) = \text{SimRender}_{\mathcal{T}_i}(x_{0,i}, a_i),$$

where τ_i^{phy} denotes the physical trajectory, including states, forces, contacts, and event records, while τ_i^{obs} denotes the visual or multimodal observation trajectory, such as rendered frames, depth maps, segmentation masks, or sensor sequences. A physical evaluator then checks the physical trajectory against the constraints in $\mathcal{T}_i$

$$(y_i, d_i) = \text{Eval}_{\mathcal{T}_i}(\tau_i^{\text{phy}}),$$

where y_i is the constraint or outcome label and d_i is the failure diagnosis.

B. PFA-Driven Physical Calibration

Let $\mathcal{D}_{\text{PFA}}$ denote the PFA-grounded source dataset, including executable physical cases, verified rollouts, outcome labels, and failure diagnoses. Based on $\mathcal{D}_{\text{PFA}}$, PFA-driven calibration is performed at the model-training level rather than through post-hoc output correction. The PFA source cases are not tied to a specific optimization algorithm. Instead, they provide physically grounded data that can be reformatted according to the training interface of PRM or PGM.

For PRM, the source cases are converted into physical reasoning data

$$\mathcal{D}_{\text{PRM}} = \{(q_i^{\text{phys}}, r_i^{\text{phys}})\}_{i=1}^{N},$$

where q_i^{phys} is a reasoning query constructed from the task setting, initial condition, candidate action, and physical rollout, and r_i^{phys} contains the corresponding outcome judgment, constraint diagnosis, and causal explanation. This data view supports model-level training of PRM through supervised or instruction tuning, physical representation injection, or reward-based post-training. These interfaces are consistent with recent efforts on physical visual question answering (VQA) and kinematic instruction tuning [14, 53], explicit physical representation encoders [19, 54], and physical AI supervised fine-tuning (SFT)/reinforcement learning (RL) [55]. Thus, the role of PFA is to supply verified physical reasoning supervision, allowing PRM to internalize task-relevant physical constraints rather than depending on external checking.

For PGM, the same source cases are used from a world-model perspective. Given $(\mathcal{T}_i, x_{0,i}, a_i)$, PGM learns to model the conditional rollout distribution

$$p_\theta(\tau_i^{\text{phy}}, \tau_i^{\text{obs}} \mid \mathcal{T}_i, x_{0,i}, a_i),$$

where τ_i^{phy} represents the physical rollout and τ_i^{obs} represents the corresponding visual or multimodal observation rollout. The physical rollout provides dynamics-level supervision, while the observation rollout ensures that the generated future remains visually and sensorially consistent with the underlying physical evolution. This formulation is compatible with generative physics training on simulator trajectories [56], physics-aware video or world-model fine-tuning [32, 57], and post-training with physical preferences or verifiable rewards [58–60]. Therefore, PFA calibrates PGM not by filtering generated outputs alone, but by providing physical-observational rollout data for training a physically grounded world model.

C. Bidirectional Co-Evolution of PRM and PGM

After PFA-driven calibration, PRM and PGM are further refined through a bidirectional training-data exchange loop. This loop does not replace the PFA. Instead, it extends the finite coverage of PFA-generated cases by allowing PGM to construct imagination-based reasoning cases for PRM and allowing PRM to construct judgment-based training signals for PGM. In this process, PFA remains the verifiable anchor that prevents the two models from reinforcing physically inconsistent assumptions.

In the PGM-to-PRM direction, the generative world model augments PRM training by producing diverse and counterfactual physical futures. Given a task setting, an initial condition, and a candidate action, PGM can synthesize multiple possible rollouts under different appearances, viewpoints, disturbances, object states, or temporal variations

$$\hat{\tau}_{i,k} = (\hat{\tau}_{i,k}^{\text{phy}}, \hat{\tau}_{i,k}^{\text{obs}}) \sim \text{PGM}(\mathcal{T}_i, x_{0,i}, a_i, z_k),$$

where $\hat{\tau}_{i,k}^{\text{phy}}$ denotes the generated physical rollout and $\hat{\tau}_{i,k}^{\text{obs}}$

denotes the corresponding observation rollout. These generated futures are not treated as physical ground truth. Instead, they are used as candidate reasoning cases for comparison, uncertainty mining, and hard-case construction. For example, PGM can generate counterfactual futures that differ in action, temporal order, contact mode, or environmental condition, thereby exposing PRM to rare and ambiguous physical cases that are difficult to exhaustively collect from the real world or enumerate within the current PFA.

In the PRM-to-PGM direction, the physical reasoning model turns generated futures into physically meaningful training signals for PGM refinement. Given a PGM-generated rollout, PRM evaluates its physical plausibility, causal continuity, action-outcome consistency, and constraint satisfaction

$$(\hat{y}_{i,k}, \hat{d}_{i,k}, \hat{r}_{i,k}) = \mathrm{PRM}(\mathcal{T}_i, x_{0,i}, a_i, \hat{\tau}_{i,k}),$$

where $\hat{y}_{i,k}$ denotes the judged physical outcome, $\hat{d}_{i,k}$ denotes the diagnosed failure mode, and $\hat{r}_{i,k}$ denotes the physical plausibility score. These reasoning-level outputs can be organized as soft supervision for PGM training, including positive and negative rollouts, preference rankings, reward shaping signals, and hard-case prompts. In this way, PRM guides PGM from visually plausible generation toward physically consistent and observation-consistent world generation.

The two directions form a reciprocal model-training process. PGM increases the diversity and difficulty of PRM's reasoning cases, while PRM converts generated futures into diagnostic and preference-based feedback for PGM. To avoid self-confirming hallucinations, high-impact, uncertain, or repeatedly disputed cases are selected for PFA-based verification. Verified hard cases can then be used to update the PFA schema space, initial-condition generator, action generator, or evaluation rules. Therefore, PRM-PGM co-evolution scales physical exploration through training-data augmentation and feedback construction, while PFA preserves the truth anchor of the overall PPI framework.

D. Deployment and Closed-Loop Evolution

The PFA-calibrated PRM and PGM are connected to embodied execution through parallel deployment. Let m index the virtual-real evolution round, and let $R_{\phi^{(m)}}$ and $G_{\theta^{(m)}}$ be the deployed PRM and PGM calibrated on $\mathcal{D}_{\mathrm{PFA}}^{(m)}$. PPI does not replace the existing action interface or low-level controller. Instead, it serves as a physical intelligence layer, where PGM provides future imagination, PRM provides physical judgment, and PFA maintains truth-anchored feedback for continual evolution. The detailed workflow of PPI, including PFA-driven calibration, PRM-PGM co-evolution, and closed-loop deployment, is summarized in Fig. 2.

At execution step t, the embodied agent obtains a task-relevant state estimate $\hat{x}_t$, and the existing planner or controller proposes candidate actions $\mathcal{A}_t = \{a_{t,k}\}_{k=1}^{K}$. For each candidate, PGM generates an action-conditioned physical future hypothesis

$$\hat{\tau}_{t,k} = (\hat{\tau}_{t,k}^{\mathrm{phy}}, \hat{\tau}_{t,k}^{\mathrm{obs}}) = G_{\theta^{(m)}}(\mathcal{T}_t^r, \hat{x}_t, a_{t,k}),$$

and PRM evaluates it as

$$(\hat{y}_{t,k}, \hat{d}_{t,k}, q_{t,k}) = R_{\phi^{(m)}}(\mathcal{T}_t^r, \hat{x}_t, a_{t,k}, \hat{\tau}_{t,k}),$$

where $q_{t,k}$ summarizes feasibility, risk, causal consistency, and task-constraint satisfaction. The selected action is obtained by

$$k^* = \arg\max_k U(q_{t,k}, \hat{y}_{t,k}), \quad \text{s.t., } \rho(q_{t,k}) \leqslant \epsilon,$$

and is executed through the low-level interface to produce a real trace

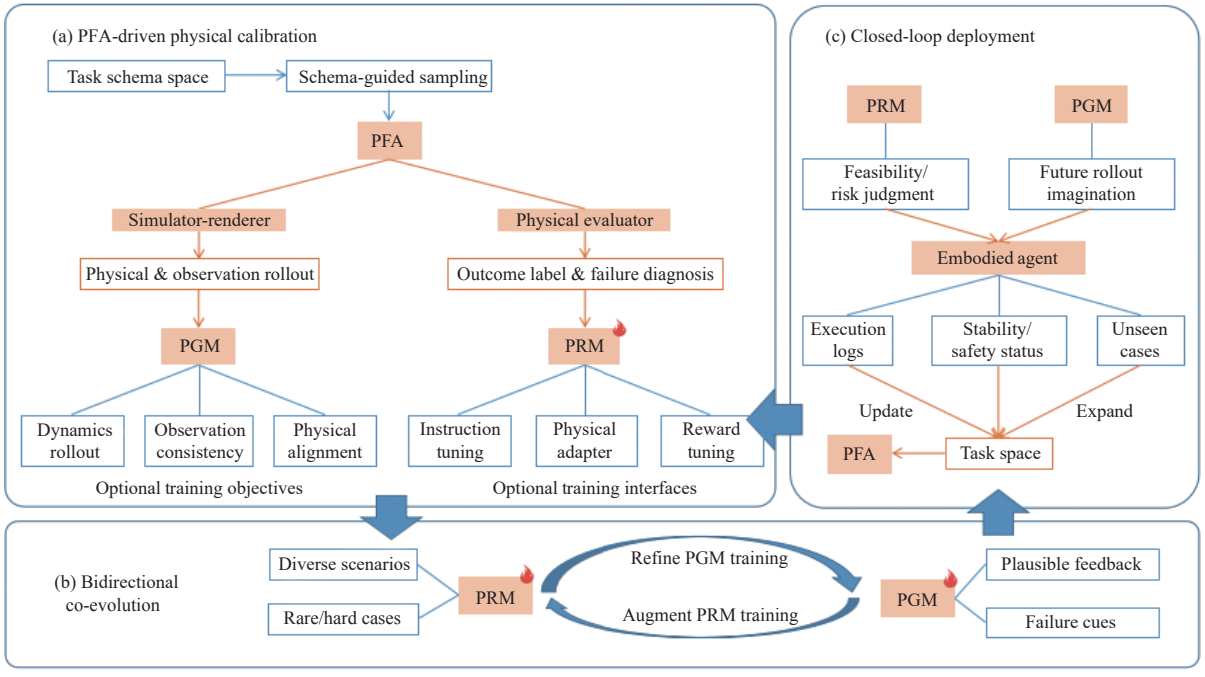


Fig. 2 Detailed workflow of the proposed PPI framework. (a) PFA-driven physical calibration for rollout generation, outcome diagnosis, and model-level training. (b) Bidirectional co-evolution through PGM-based case augmentation and PRM-based feedback refinement. (c) Closed-loop deployment with real-world feedback for task-space and PFA update.

$$\tau_t^r = \mathrm{Exec}_{\mathcal{E}}(\mathcal{T}_t^r, \hat{x}_t, a_{t,k^*}).$$

After execution, a monitor extracts the real outcome and diagnosis

$$(y_t^r, d_t^r) = \mathrm{Mon}_{\mathcal{T}_t^r}(\tau_t^r).$$

Failure cases, rare events, near-violations, high-uncertainty samples, and prediction-execution discrepancies are selected as hard feedback $\mathcal{H}^{(m)}$. Instead of directly updating PRM or PGM with raw experience, PPI routes such feedback back to the PFA data engine

$$\Psi^{(m+1)} = \mathrm{Update}_{\mathrm{PFA}}(\Psi^{(m)}, \mathcal{H}^{(m)}),$$

where $\Psi = (\Omega_{\mathcal{T}}, P_X, P_A, \mathrm{SimRender}, \mathrm{Eval})$. The updated PFA generates new executable physical cases $C_{\mathrm{new}}^{(m+1)}$, and the PFA dataset is updated by

$$\mathcal{D}_{\mathrm{PFA}}^{(m+1)} = \mathcal{D}_{\mathrm{PFA}}^{(m)} \cup C_{\mathrm{new}}^{(m+1)}.$$

New PRM and PGM training views are then derived from $\mathcal{D}_{\mathrm{PFA}}^{(m+1)}$, triggering the next PRM-PGM co-evolution round.

In this way, the real side executes tasks with the currently validated physical intelligence layer, while the artificial side asynchronously converts real-world failures and corner cases into new PFA-grounded calibration data. This separation also improves computational scalability: online deployment only requires the front-end PRM-PGM layer for action-conditioned imagination, physical judgment, and execution support, whereas the heavier PFA expansion, hard-case verification, and PRM-PGM recalibration are performed asynchronously in the artificial physical space. Therefore, as tasks, agents, and feedback cases increase, PPI scales through back-end physical data expansion and computational experiments without increasing the real-time burden of embodied execution or requiring complete world replication and exhaustive real-world trials.

IV. Application

PPI can be instantiated in different embodied domains by specifying a domain-dependent task schema Ω_T, including task goals, physical states, candidate actions, constraints, evaluation rules, and simulator-renderer backends. The PFA then converts these schemas into executable physical cases for PRM and PGM calibration, computational experiments support PRM-PGM co-evolution, and parallel execution feeds real failures or prediction-execution discrepancies back to the PFA. The following examples illustrate three representative physical bottlenecks: constrained motion in open environments, contact-rich object interaction, and safety-critical intervention under deformable dynamics. They also highlight the practical advantages of PPI: physical validity is explicitly anchored by the PFA, risky trial-and-error is shifted to parallelizable virtual experiments, and real-world failures are converted into PFA-updating cases for continual adaptation.

A. PPI for Autonomous Mobility and Navigation

In autonomous mobility and navigation, Ω_T^{mob} can describe road or airspace topology, vehicle or UAV dynamics, terrain conditions, obstacle distributions, weather disturbances, energy limits, and safety constraints. The PFA is instantiated as an artificial mobility system that generates executable state-transition and constraint-evaluation cases for route planning and motion prediction. Based on these cases, PRM is calibrated to judge reachability, stability, collision risk, energy feasibility, and causal consistency, while PGM is calibrated to generate PFA-grounded candidate route futures under different disturbances, occlusions, or traffic interactions. During parallel execution, embodied vehicles or UAVs use the current physical intelligence layer for candidate-action evaluation, whereas deviations, near misses, and unmodeled aerodynamic or ground-contact effects are routed back to the PFA. In this way, PPI supports robust navigation by combining virtual stress testing with real-world feedback, while reducing dependence on risky exhaustive trials.

B. PPI for Contact-Rich Robotic Manipulation

For contact-rich manipulation, Ω_T^{man} can encode object geometry, material properties, surface friction, robot kinematics, end-effector type, manipulation goals, contact constraints, force limits, and task success criteria. The PFA is instantiated as an artificial contact system that produces physically grounded interaction trajectories, force/contact outcomes, slippage events, deformation states, and constraint evaluations. These cases calibrate PRM for stability judgment, affordance reasoning, failure diagnosis, and causal explanation, while providing PGM with PFA-grounded references for generating dynamically plausible contact evolution. In computational experiments, PGM expands object poses, contact modes, tool configurations, perturbations, and failure-like futures, and PRM evaluates their physical plausibility and returns diagnostic feedback. During real execution, unexpected slippage, jamming, breakage, or excessive force is absorbed into the PFA for the next co-evolution round, supporting safer and more adaptive handling of long-tail contact conditions.

C. PPI for Safety-Critical Surgical Intervention

Surgical intervention is a conservative instantiation of PPI, because small physical deviations may cause anatomical boundary violation, tissue trauma, instrument interference, or procedural failure. In this domain, Ω_T^{surg} can specify anatomical geometry, patient-specific tissue properties, instrument kinematics, procedural stages, forbidden regions, contact-force limits, deformation constraints, and safety rules. The PFA is instantiated as an artificial intervention system that provides physically grounded reference cases for instrument-tissue interaction, deformation response, and boundary-risk evaluation. PGM generates counterfactual tool paths, tissue-response sequences, occlusion changes, and patient-specific rare cases, while PRM evaluates contact feasibility, penetration risk, collision risk, boundary violation, causal continuity, and procedural constraint satisfaction. During parallel execution, PPI should be used as a physical intelligence layer for decision support and risk gating rather than as an autonomous replacement for surgical control. Intraoperative observations, near-failure events, and unexpected tissue responses are routed back to the PFA, ensuring that subsequent PRM-PGM evolution remains grounded in updated anatomical and physical constraints.

V. CONCLUSIONS

This letter proposes PPI, a truth-anchored parallel framework for organizing physical reasoning, physical generation, and physical simulation toward reliable embodied intelligence. By introducing PFA and coupling PRM with PGM under PFA-driven calibration, PPI enables bidirectional co-evolution in computational experiments and closes the ACP loop through real-world feedback that updates the PFA for subsequent training.

Future work will focus on the practical deployment and experimental validation of PPI in autonomous mobility, contact-rich manipulation, and safety-critical surgical intervention. By instantiating domain-specific PFAs and connecting PRM-PGM modules with real embodied control interfaces, we will evaluate the closed-loop virtual-real evolution of PPI under real execution feedback and further optimize its physical validity, safety, runtime efficiency, and scalability.

REFERENCES

[1] A. Miriyev and M. Kovač, Skills for physical artificial intelligence, *Nat. Mach. Intell.*, 2020, 2(11), 658–660.

[2] Y. Li, Z. Li, Y. Duan, and A.-B. Spulber, Physical artificial intelligence (PAI): The next-generation artificial intelligence, *Front. Inform. Technol. Electron. Eng.*, 2023, 24(8), 1231–1238.

[3] E.-H. Wu, Y.-Q. Liu, T.-C. Xu, L.-X. Ren, Y.-M. Qin, M.-Y. Wei, X.-W. He, D.-Y. Yuan, W.-C. Hou, Z.-W. Ma et al., Physical AI: Evolution, progress, challenges, and prospects, *J. Comput. Sci. Technol.*, 2026, 41(1), 271–288.

[4] Y. Wang, J. He, L. Fan, H. Li, Y. Chen, and Z. Zhang, Driving into the future: Multiview visual forecasting and planning with world model for autonomous driving, in *Proc. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 2024, 14749–14759.

[5] E. Kaufmann, L. Bauersfeld, A. Loquercio, M. Müller, V. Koltun, and D. Scaramuzza, Champion-level drone racing using deep reinforcement learning, *Nature*, 2023, 620(7976), 982–987.

[6] C. Chi, Z. Xu, S. Feng, E. Cousineau, Y. Du, B. Burchfiel, R. Tedrake, and S. Song, Diffusion policy: Visuomotor policy learning via action diffusion, *Int. J. Robot. Res.*, 2025, 44(10&11), 1684–1704.

[7] J. Xu, B. Li, B. Lu, Y.-H. Liu, Q. Dou, and P.-A. Heng, SurROL: An open-source reinforcement learning centered and dVRK compatible platform for surgical robot learning, in *Proc. 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Prague, Czech Republic, 2021, 1821–1828.

[8] S. Qiu, S. Guo, Z.-Y. Song, Y. Sun, Z. Cai, J. Wei, T. Luo, Y. Yin, H. Zhang, Y. Hu et al., PHYBench: Holistic evaluation of physical perception and reasoning in large language models, arXiv preprint arXiv: 2504.16074v2, 2025.

[9] F. Zhou, J. Huang, J. Li, D. Ramanan, and H. Shi, PAI-Bench: A comprehensive benchmark for physical AI, arXiv preprint arXiv: 2512.01989, 2025.

[10] X. Guo, J. Huo, Z. Shi, Z. Song, J. Zhang, and J. Zhao, T2VPhysBench: A first-principles benchmark for physical consistency in text-to-video generation, arXiv preprint arXiv: 2505.00337, 2025.

[11] J. Gu, X. Liu, Y. Zeng, A. Nagarajan, F. Zhu, D. Hong, Y. Fan, Q. Yan, K. Zhou, M.-Y. Liu et al., "PhyWorldBench": A comprehensive evaluation of physical realism in text-to-video models, arXiv preprint arXiv: 2507.13428, 2025.

[12] F.-Y. Wang, Parallel system methods for management and control of complex systems, *Control Decis.*, 2004, 19(5), 485–489,514.

[13] J. Gao, B. Sarkar, F. Xia, T. Xiao, J. Wu, B. Ichter, A. Majumdar, and D. Sadigh, Physically grounded vision-language models for robotic manipulation, arXiv preprint arXiv: 2309.02561, 2023.

[14] S. Huang, I. Ponomarenko, Z. Jiang, X. Li, X. Hu, P. Gao, H. Li, and H. Dong, ManipVQA: Injecting robotic affordance and physically grounded information into multi-modal large language models, in *Proc. 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Abu Dhabi, United Arab Emirates, 2024, 7580–7587.

[15] W. Huang, C. Wang, R. Zhang, Y. Li, J. Wu, and F.-F. Li, VoxPoser: Composable 3D value maps for robotic manipulation with language models, in *Proc. 7th Conference on Robot Learning*, Atlanta, GA, USA, 2023, 540–562.

[16] B. Chen, Z. Xu, S. Kirmani, B. Ichter, D. Sadigh, L. Guibas, and F. Xia, SpatialVLM: Endowing vision-language models with spatial reasoning capabilities, in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 2024, 14455–14465.

[17] C. H. Song, V. Blukis, J. Tremblay, S. Tyree, Y. Su, and S. Birchfield, RoboSpatial: Teaching spatial understanding to 2D and 3D vision-language models for robotics, in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, TN, USA, 2025, 15768–15780.

[18] B. Ichter, A. Brohan, Y. Chebotar, C. Finn, K. Hausman, A. Herzog, D. Ho, J. Ibarz, A. Irpan, E. Jang et al., Do as I can, not as I say: Grounding language in robotic affordances, in *Proc. Conference on Robot Learning*, Auckland, New Zealand, 2023, 287–318.

[19] W. Zhou, M. Tao, C. Zhao, H. Guo, H. Dong, M. Tang, and J. Wang, PhysVLM: Enabling visual language models to understand robotic physical reachability, in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, TN, USA, 2025, 6940–6949.

[20] X. Wu, Z. Li, J. Jin, G. Shi, K. V. Gouthaman, V. Raj, N. Sinha, J. Chen, F. Du, and D. Manocha, MASS: Motion-aware spatial-temporal grounding for physics reasoning and comprehension in vision-language models, arXiv preprint arXiv: 2511.18373, 2025.

[21] Y.-W. Zhan, X. Wang, H. Chen, T. Feng, W. Feng, R. Wang, G. Li, Q. Li, and W. Zhu, PhyVLLM: Physics-guided video language model with motion-appearance disentanglement, arXiv preprint arXiv: 2512.04532, 2025.

[22] G. Sreekumar and V. N. Boddeti, InPhyRe discovers: Large multimodal models struggle in inductive physical reasoning, arXiv preprint arXiv: 2509.12263, 2026.

[23] P. Li, T. Xiang, E. Mao, S. Wei, X. Chen, A. Masood, F.-F. Li, and E. Adeli, QUANTIPHY: A quantitative benchmark evaluating physical reasoning abilities of vision-language models, in Proc. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Denver, CO, USA, 2026, 33174–33184.

[24] Z. Zheng, X. Yan, Z. Chen, J. Wang, Q. Z. E. Lim, J. B. Tenenbaum, and C. Gan, ContPhy: Continuum physical concept learning and reasoning from videos, in *Proc. 41st International Conference on Machine Learning*, Vienna, Austria, 2024, 61526–61558.

[25] A. Cherian, R. Corcodel, S. Jain, and D. Romeres, LLMPhy: Parameter-identifiable physical reasoning combining large language models and physics engines, arXiv preprint arXiv: 2411.08027, 2026.

[26] A. Foss, C. Evans, S. Mitts, K. Sinha, A. Rizvi, and J. T. Kao, CausalVQA: A physically grounded causal reasoning benchmark for video models, arXiv preprint arXiv: 2506.09943, 2025.

[27] OpenAI, Video generation models as world simulators [Online], https://openai.com/index/video-generation-models-as-world-simulators/, 5 May 2026.

[28] J. Bruce, M. D. Dennis, A. Edwards, J. Parker-Holder, Y. Shi, E. Hughes, M. Lai, A. Mavalankar, R. Steigerwald, C. Apps et al., Genie: Generative interactive environments, in *Proc. 41st International Conference on Machine Learning*, Vienna, Austria, 2024, 4603–4623.

[29] H. Bansal, Z. Lin, T. Xie, Z. Zong, M. Yarom, Y. Bitton, C. Jiang, Y. Sun, K.-W. Chang, and A. Grover, VideoPhy: Evaluating physical commonsense for video generation, in *Proc. 13th International Conference on Learning Representations*, Singapore, Singapore, 2025.

[30] F. Meng, J. Liao, X. Tan, Q. Lu, W. Shao, K. Zhang, Y. Cheng, D. Li, and P. Luo, Towards world simulator: Crafting physical commonsense-based benchmark for video generation, in *Proc. 42nd International Conference on Machine Learning*, Vancouver, Canada, 2025, 1759.

[31] D. Li, Y. Fang, Y. Chen, S. Yang, S. Cao, J. Wong, M. Luo, X. Wang, H. Yin, J. E. Gonzalez et al., WorldModelBench: Judging video generation models as world models, arXiv preprint arXiv: 2502.20694, 2025.

[32] K. Zhang, C. Xiao, J. Xu, Y. Mei, and V. M. Patel, Think before you

diffuse: Infusing physical rules into video diffusion, arXiv preprint arXiv: 2505.21653, 2025.

[33] H. Yue, S. Huang, Y. Liao, S. Chen, P. Zhou, L. Chen, M. Yao, and G. Ren, EWMBench: Evaluating scene, motion, and semantic quality in embodied world models, arXiv preprint arXiv: 2505.09694, 2025.

[34] C. Zhu, R. Yu, S. Feng, B. Burchfiel, P. Shah, and A. Gupta, Unified world models: Coupling video and action diffusion for pretraining on large robotic datasets, arXiv preprint arXiv: 2504.02792, 2025.

[35] Z. Cao, Z. Chen, L. Pan, and Z. Liu, PhysX-3D: Physical-grounded 3D asset generation, in *Proc. 39th Conference on Neural Information Processing Systems*, San Diego, CA, USA, 2025.

[36] D. Romero, A. Bermudez, V. Iablochnikov, H. Li, F. Pizzati, and I. Laptev, Learning to generate rigid body interactions with video diffusion models, arXiv preprint arXiv: 2510.02284, 2025.

[37] N. Gillman, C. Herrmann, M. Freeman, D. Aggarwal, E. Luo, D. Sun, and C. Sun, Force prompting: Video generation models can learn and generalize physics-based control signals, in *Proc. 39th Conference on Neural Information Processing Systems*, San Diego, CA, USA, 2025.

[38] C. Wang, C. Chen, Y. Huang, Z. Dou, Y. Liu, J. Gu, and L. Liu, PhysCtrl: Generative physics for controllable and physics-grounded video generation, in *Proc. 39th Conference on Neural Information Processing Systems*, San Diego, CA, USA, 2025.

[39] T. Xie, Z. Zong, Y. Qiu, X. Li, Y. Feng, Y. Yang, and C. Jiang, PhysGaussian: Physics-integrated 3D Gaussians for generative dynamics, in *Proc. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 2024, 4389–4398.

[40] S. Liu, Z. Ren, S. Gupta, and S. Wang, PhysGen: Rigid-body physics-grounded image-to-video generation, in *Proc. 18th European Conference on Computer Vision*, Milan, Italy, 2024, 360–378.

[41] J. Lin, Z. Wang, D. Xu, S. Jiang, Y. Gong, and M. Jiang, Phys4DGen: Physics-compliant 4D generation with multi-material composition perception, in *Proc. 33rd ACM International Conference on Multimedia*, Dublin, Ireland, 2025, 10398–10407.

[42] E. Todorov, T. Erez, and Y. Tassa, MuJoCo: A physics engine for model-based control, in *Proc. 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vilamoura-Algarve, Portugal, 2012, 5026–5033.

[43] NVIDIA, Isaac Lab: A GPU-accelerated simulation framework for multi-modal robot learning, arXiv preprint arXiv: 2511.04831, 2025.

[44] C. D. Freeman, E. Frey, A. Raichuk, S. Girgin, I. Mordatch, and O. Bachem, Brax—A differentiable physics engine for large scale rigid body simulation, arXiv preprint arXiv: 2106.13281, 2021.

[45] Y. Hu, L. Anderson, T.-M. Li, Q. Sun, N. Carr, J. Ragan-Kelley, and F. Durand, DiffTaichi: Differentiable programming for physical simulation, arXiv preprint arXiv: 1910.00935, 2019.

[46] Y. Hu, J. Liu, A. Spielberg, J. B. Tenenbaum, W. T. Freeman, J. Wu, D. Rus, and W. Matusik, ChainQueen: A real-time differentiable physical simulator for soft robotics, in *Proc. International Conference on Robotics and Automation*, Montreal, Canada, 2019, 6265–6271.

[47] W. Li, H. Zhao, Z. Yu, Y. Du, Q. Zou, R. Hu, and K. Xu, PIN-WM: Learning physics-informed world models for non-prehensile manipulation, arXiv preprint arXiv: 2504.16693, 2025.

[48] X. Wang, J. Yang, Y. Liu, Y. Wang, F.-Y. Wang, M. Kang, Y. Tian, I. Rudas, L. Li, M. P. Fanti et al., Parallel intelligence in three decades: A historical review and future perspective on ACP and cyber-physical-social systems, *Artif. Intell. Rev.*, 2024, 57(9), 255.

[49] F.-Y. Wang, Toward a paradigm shift in social computing: The ACP approach, *IEEE Intell. Syst.*, 2007, 22(5), 65–67.

[50] Y. Tian, F. Lin, C. Wang, J. Ge, Z. Luo, and F.-Y. Wang, Parallel deep foundation model: A co-evolution framework for analogical imagination and embodied cognition of parallel intelligence, *Int. J. Intell. Control Syst.*, 2025, 30(1), 83–90.

[51] B. Sun, Y. Liu, C. Wang, Z. Luo, and Y. Tian, PRALLM: A parallel RAG-agent-LLM framework for unified public-private knowledge integration, *Int. J. Intell. Control Syst.*, 2025, 30(3), 260–266.

[52] F. Lin, T. Zhang, J. Huang, Q. Ni, J. Ge, and Y. Tian, PiVLA: Vision-language-action based on parallel intelligence, *Int. J. Intell. Control Syst.*, 2025, 30(3), 253–259.

[53] D. Ko, S. Kim, Y. Suh, V. K. B. G, M. Yoon, M. Chandraker, and H. J. Kim, ST-VLM: Kinematic instruction tuning for spatio-temporal reasoning in vision-language models, arXiv preprint arXiv: 2503.19355, 2025.

[54] E. Zhu, Y. Liu, Z. Zhang, X. Li, J. Zhou, X. Yu, M. Huang, and H. Wang, MAPS: Advancing multi-modal reasoning in expert-level physical science, in *Proc. 13th International Conference on Learning Representations*, Singapore, Singapore, 2025.

[55] A. Azzolini, J. Bai, H. Brandon, J. Cao, P. Chattopadhyay, H. Chen, J. Chu, Y. Cui, J. Diamond, Y. Ding et al., Cosmos-Reason1: From physical common sense to embodied reasoning, arXiv preprint arXiv: 2503.15558, 2025.

[56] C. Wang, C. Chen, Y. Huang, Z. Dou, Y. Liu, J. Gu, and L. Liu, PhysCtrl: Generative physics for controllable and physics-grounded video generation, arXiv preprint arXiv: 2509.20358, 2025.

[57] S. Zhou, H. Wang, H. Cheng, J. Li, D. Wang, J. Jiang, Y. Jin, J. Huang, S. Mao, S. Liu et al., PhysInOne: Visual physics learning and reasoning in one suite, arXiv preprint arXiv: 2604.09415, 2026.

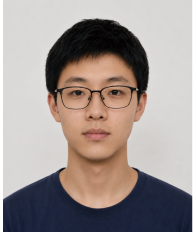
[58] Y. Cai, K. Li, M. Jia, J. Wang, J. Sun, F. Liang, W. Chen, F. Juefei-Xu, C. Wang, A. Thabet et al., PhyGDPO: Physics-aware groupwise direct preference optimization for physically consistent text-to-video generation, arXiv preprint arXiv: 2512.24551, 2025.

[59] M.-Q. Le, Y. Zhu, V. Kalogeiton, and D. Samaras, What about gravity in video generation? Post-training Newton's laws with verifiable rewards, arXiv preprint arXiv: 2512.00425, 2025.

[60] R. Wang, Q. Liu, Y. Deng, G. Liu, Z. Liu, and K. Jia, EVA: Aligning video world models with executable robot actions via inverse dynamics rewards, arXiv preprint arXiv: 2603.17808, 2026.

Xinyuan Zhang is currently pursuing the BS degree in artificial intelligence at University of Chinese Academy of Sciences, Beijing, China. Her research interests include robotic perception, spatial intelligence, and digital twins.

Jiaxi Liu received the BA degree in sociology from Tulane University, New Orleans, LA, USA, in 2021. She is currently pursuing the MS degree in social network analysis at University of Manchester, Manchester, UK. Her research interests include social network analysis and the application of complex systems in sociology, with a particular focus on structural composition, dynamic patterns, and societal impacts within social operations, as well as corresponding countermeasures informed by complex systems approaches.



Siheng Yi received the MS degree from Óbuda University, Hungary, in 2025. He is currently pursuing the PhD degree in computer science at Óbuda University, Budapest, Hungary. His research interests include machine learning, model interpretability, and trustworthy AI.

Hui Zhang received the PhD degree in control theory and control engineering from University of Chinese Academy of Sciences, Beijing, China, in 2020. She is currently an associate professor at School of Computer Science and Technology, Beijing Jiaotong University, China. Her research interests include multi-agent collaboration, parallel vision, embodied intelligence, and intelligent transportation systems.