HERACLES: BRIDGING PRECISE TRACKING AND GENERATIVE SYNTHESIS FOR GENERAL HUMANOID CONTROL

Anonymous authors
Paper under double-blind review

ABSTRACT

Achieving general-purpose humanoid control requires a delicate balance between the precise execution of commanded motions and the flexible, anthropomorphic realignment of state and intent. Current general controllers predominantly formulate motion control as a rigid reference-tracking problem. This formulation implicitly assumes that the reference is always reachable from the current state; once this assumption breaks, these trackers produce myopic, single-step corrections and exhibit brittle, non-anthropomorphic failure modes, lacking the generative adaptability inherent to human motor control. To overcome this limitation, we propose Heracles, a novel state-conditioned generative middleware that bridges precise motion tracking and generative synthesis. Rather than relying on rigid tracking paradigms or explicit mode-switching between a tracker and a dedicated recovery policy, Heracles operates as an intermediary layer between high-level reference motions and low-level physics trackers. By conditioning on the robot's real-time state, the flow-matching model implicitly adapts its behavior: it approximates an identity map when the state closely aligns with the reference, preserving zero-shot tracking fidelity. Conversely, when encountering significant state deviations, it seamlessly transitions into a generative synthesizer, producing an anthropomorphic whole-body maneuver back toward the reference. Our framework demonstrates that integrating generative priors into the control loop not only significantly enhances robustness against extreme perturbations but also elevates humanoid control from a rigid tracking paradigm to a generative general-purpose architecture. Videos and demos are available at https://heracles-humanoid-control-submission.github.io/.

1 INTRODUCTION

General-purpose humanoid control demands architectures that execute precise, goal-directed motions while maintaining the flexible, anthropomorphic adaptability to bridge any divergence between the robot's own state and its reference motion. Humans seamlessly blend exact task execution with intuitive, generative transitions whenever such a divergence arises. Yet the prevailing paradigm in robotics formulates this as reference-driven tracking: deep RL controllers (Yin et al., 2026b; Luo et al., 2026; Wang et al., 2026b; He et al., 2025a) minimize kinematic deviation from reference trajectories, achieving impressive zero-shot execution of diverse MoCap skills under nominal conditions. However, this rigid tracking objective implicitly assumes that the reference remains reachable from the robot's current state. Once the robot's state lies far from the reference manifold, trackers tend to produce myopic, single-step corrective actions that lack the coordinated multi-step reasoning required to return to it, precipitating non-anthropomorphic failures. Dedicated recovery policies (He et al., 2025b; Chen et al., 2025a; Xu et al., 2025b) handle only fall recovery, and integrating them with a general tracker requires an explicit mode switch. Conversely, entirely dropping the tracking objective in favor of behavioral foundation models (Li et al., 2026a) yields more natural, human-like movements but fundamentally sacrifices the spatial and temporal precision required for strict task execution. A significant gap remains: bridging the exactitude of precise trackers with the generative adaptability of reference motion.

To overcome this fundamental limitation, we propose Heracles, a state-conditioned generative middleware that bridges precise motion tracking and generative synthesis (Fig. 1). Rather than

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

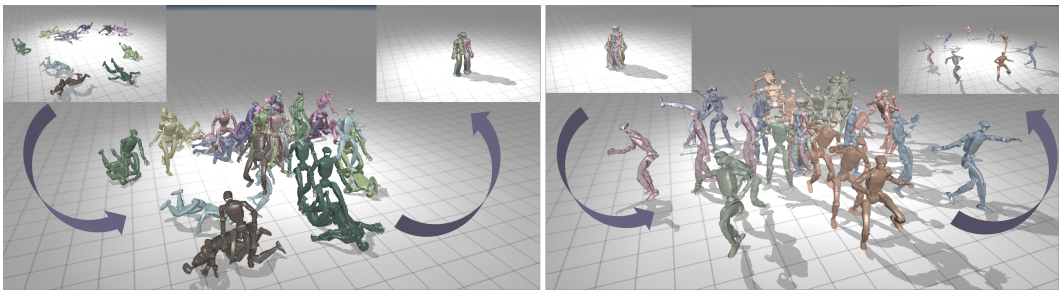

Figure 1: **Heracles synthesizes anthropomorphic transitions from any state to any reference.** (Left) Many states, one reference: a ring of randomly sampled fallen configurations, all commanded by a single standing reference, each resolved by its own transition to standing. (Right) One state, many references: an identical standing pose in different orientations, each commanded by a distinct reference at a distance (high kick, ball kick, run, jump) and reached through a natural transition. In both cases the tracker executes only the trajectory synthesized by the middleware, without fall detector, mode switch, or recovery-specific reward.

engineering a monolithic end-to-end controller or relying on explicit mode-switching, Heracles operates as an intermediary layer between high-level motion commands and a low-level physics tracker, injecting generative priors into the control loop without disrupting high-frequency execution.

The core innovation lies in its implicit, state-driven adaptability. Conditioned on the robot's real-time physical state, the model synthesizes the transition toward the reference motion. In nominal tracking the state lies close to the reference and the synthesized transition is simply the next segment of the reference motion. The model therefore approximates an identity map and preserves tracking fidelity. Under large deviations, the model seamlessly transitions into a generative synthesizer, producing natural transitions back to the reference for the tracker to execute. This design unifies nominal tracking and generative synthesis within a single framework, eliminating the need for explicit heuristics or separate recovery policies.

In summary, we make three contributions: **(1)** We introduce Heracles, a state-conditioned generative middleware that bridges the precision of motion tracking with the flexibility of generative synthesis, decoupling high-level intent generation from low-level physical execution. **(2)** We improve the underlying physics tracker with finite scalar quantization and adaptive motion sampling, retaining high-fidelity tracking while seamlessly integrating with the generative middleware. **(3)** We deploy Heracles on physical humanoid robots, demonstrating emergent, human-like whole-body maneuvers and robust generative adaptability under severe deviations from the reference motion.

2 Related Work

General Humanoid Motion Control. Mimic-based tracking has emerged as the dominant paradigm for general-purpose humanoid control. Pioneered by DeepMimic (Peng et al., 2018), which demonstrated that physics-based characters can be trained via deep RL to closely imitate reference motion clips, this paradigm has since been scaled considerably: HOVER (He et al., 2025a) consolidates navigation and loco-manipulation into a unified controller via multi-mode distillation. Subsequent work trains single policies over increasingly diverse repertoires: GMT (Chen et al., 2025b) unifies heterogeneous distributions via adaptive sampling and mixture-of-experts, AMS (Pan et al., 2025c) fuses agile MoCap with physically constrained balance motions, and BeyondMimic (Liao et al., 2026) distills a compact agile tracker into a unified diffusion policy for zero-shot task composition via classifier guidance. Concurrently, SONIC (Luo et al., 2026) scales tracking along capacity, data (100M+ frames), and compute while introducing a universal token space for multimodal interfaces, and OmniXtreme (Wang et al., 2026b) breaks the representation bottleneck by decoupling skill learning from physical refinement via flow-matching policies. The paradigm extends to language-driven control (Sun et al., 2024; Li et al., 2026b), parkour (Zhuang et al., 2026; Zhu et al., 2026), interaction (Yang et al., 2025; Lin et al., 2026b; Zhang et al., 2026a), and athletic skills (Chen et al., 2026; Wang et al., 2025; Yu et al., 2025; Zhang et al., 2026c), with simulated Olympic sports (Luo et al., 2024) serving as a benchmark. Despite this breadth, tracking-first controllers remain anchored

to references; once the robot's state diverges far from the reference, they produce myopic corrective actions rather than a coordinated transition back to it. Complementarily, unsupervised RL and Behavioral Foundation Models (Yuan et al., 2025; Zeng et al., 2025; Li et al., 2026a) exhibit natural physical compliance but mapping their unconstrained latent spaces to high-precision tracking remains challenging. The unification of tracking fidelity and generative robustness remains unresolved.

Fall Recovery. Tracking failure in deployment typically leaves the robot in a fallen configuration, the most representative and consequential divergence between the robot's state and its reference motion. Dedicated recovery policies target the fallen state: HumanUP (He et al., 2025b) discovers a get-up motion in simulation and imitates a slowed, smoothed version of it, HiFAR (Chen et al., 2025a) acquires high-dynamics fall recovery through a multi-stage curriculum, and FIRM (Xu et al., 2025b) unifies fall prevention, impact mitigation, and recovery in one policy learned from a few demonstrations. These policies nonetheless recover from diverse initial configurations toward one designated posture; a transition between any state and any command lies beyond their scope. Pairing them with a general tracker would accordingly require an explicit mode switch. PHC (Luo et al., 2023) places the mode switch inside a single imitation controller, changing the goal from full-body imitation to a root point-goal once the root deviates from the reference by more than 0.5 m.

Motion Generation. Synthesizing diverse human motion has progressed from transformer baselines (MDM, Tevet et al., 2022) through multi-task language models (MotionGPT, Jiang et al., 2023; Wang et al., 2026a), million-scale (Fan et al., 2025) and multimodal (Li et al., 2025) autoregressive generation, to billion-parameter flow-matching transformers (Wen et al., 2025), supported by large-scale datasets (Lin et al., 2023; Zhang et al., 2025a). Despite achieving remarkable diversity, these models operate in an open-loop kinematic domain, ignoring embodied constraints such as joint torque limits and contact dynamics.

Hybrid Tracking-Synthesis Architectures. Within the physics-based controller itself, ASE (Peng et al., 2022) constructs continuous latent spaces for anthropomorphic downstream control, VMP (Serifi et al., 2024) distills multipurpose motion priors for tracking and perturbation resilience, AMOR (Alegre et al., 2025) trains weight-conditioned policies spanning the Pareto front of reward trade-offs, ADD (Zhang et al., 2025b) replaces hand-designed rewards with an adversarial differential discriminator, a whole-body locomotion controller (Zhang et al., 2024) imitates human references through an adversarial motion prior, and RoboStriker (Yin et al., 2026a) decouples strategic reasoning from physical execution via hierarchical latent manifolds. Within the kinematic domain, COMET (Lee et al., 2026) employs a conditional VAE with reference-guided feedback, while MotionStreamer (Xiao et al., 2025) streams motion from an autoregressive model in a causal latent space and DartControl (Zhao et al., 2025) enforces sequential synthesis under spatial constraints. Through a physics tracker, CLoSD (Tevet et al., 2025) closes the loop between a text- and goal-conditioned diffusion planner and the simulated character, PARC (Xu et al., 2025a) alternates generation and tracking to expand a terrain-traversal dataset, and on hardware a whole-body locomotion pipeline (Zhang et al., 2026b) tracks references generated online from a heading command and a terrain scan. However, in these hybrid paradigms the generative target follows the task command rather than the divergence between the robot's physical state and its reference motion. A mechanism bridging precise tracking and generative synthesis whenever such a divergence arises remains absent.

3 METHOD

3.1 SYSTEM OVERVIEW AND PROBLEM FORMULATION

General humanoid control requires executing desired motion commands while producing coordinated whole-body maneuvers back toward the reference motion under severe state deviations. We formulate this dual objective—high-fidelity motion tracking and generative synthesis—within a hierarchical control architecture. The proposed framework, Heracles, intrinsically decouples high-level intent generation from high-frequency physical execution (Fig. 2) via two components: a state-conditioned generative middleware and a general-purpose physics tracker.

In standard tracking paradigms (Peng et al., 2018; Chen et al., 2025b; Luo et al., 2026), a policy minimizes the kinematic deviation between state $\mathbf{p}_t$ and reference $\mathbf{m}_t$. When the state remains close to the reference manifold, directly tracking $\mathbf{m}_t$ yields optimal zero-shot performance. However, once the state lies far from the reference manifold, a policy that strictly adheres to $\mathbf{m}_t$ produces rigid,

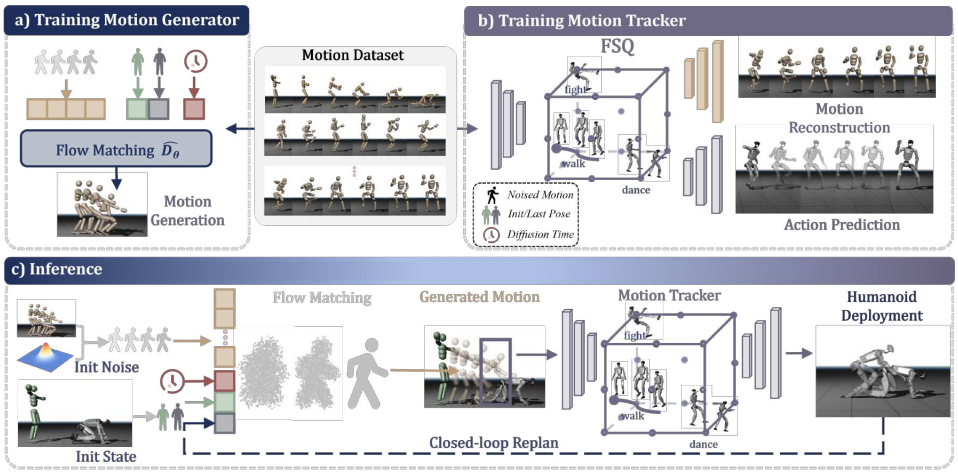

Figure 2: **Overview of the Heracles framework. (a)** A flow matching model $\hat{D}_\theta$ learns to synthesize feasible keyframe trajectories conditioned on the current state. **(b)** Reference motions are quantized into discrete tokens via FSQ, shared by reconstruction and action prediction heads. **(c)** At inference, the middleware generates trajectories through closed-loop replanning for the motion tracker to execute.

myopic corrective actions that lack the coordinated multi-step reasoning required to return to the reference, precipitating catastrophic failure.

To resolve this limitation, we formulate the core problem as learning an intermediary mapping that dynamically modulates the reference command based on real-time physical feasibility. The generative middleware functions as this state-conditioned trajectory synthesizer. Operating at a lower planning frequency, it observes the current state $\mathbf{p}_t$ and the original reference $\mathbf{m}_t$, predicting a short-horizon, kinematically continuous keyframe trajectory:

$$\boldsymbol{\tau}_t = f_\theta(\mathbf{p}_t, \mathbf{m}_t) \in \mathbb{R}^{K \times D}, \tag{1}$$

where K denotes the number of keyframes and D the state dimension. This mapping unifies nominal tracking and generative synthesis: when the robot operates near the reference manifold, f_θ approximates an identity transformation to preserve high-fidelity tracking; conversely, under large state deviations, f_θ synthesizes anthropomorphic transition trajectories that guide the robot back toward the reference manifold.

The K synthesized keyframes $\boldsymbol{\tau}_t$, uniformly spaced over the planning horizon Δt, define the reference trajectory from which the low-level physics tracker draws per-step commands at its control frequency. We model this tracking process as a discounted MDP $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, r, \gamma)$, with observation:

$$\mathbf{o}_t = \{\mathbf{p}_t,\ \mathbf{m}'_t,\ \mathbf{z}_d\}, \tag{2}$$

where $\mathbf{m}'_t$ is the modulated reference sampled from $\boldsymbol{\tau}_t$ and $\mathbf{z}_d$ is a discrete motion embedding (Sec. 3.3). The policy maximizes:

$$J(\pi) = \mathbb{E}_{\tau \sim \pi}\left[\sum_{t=0}^{T-1} \gamma^t r_t\right], \tag{3}$$

where r_t represents a task reward composed of tracking precision and physical regularization terms. Deployment follows a receding-horizon replanning loop with execution interval N_{exec}: every N_{exec} control steps, the middleware updates $\boldsymbol{\tau}_t$ from the latest proprioceptive observation, while the tracking policy executes the resulting dense trajectory at the fundamental control frequency, forming a seamless, closed-loop tracking-generation architecture.

3.2 State-Conditioned Generative Middleware via Flow Matching

We formulate the trajectory generation process as a conditional flow matching problem over a geometrically constrained residual space, bridging exact motion tracking and generative synthesis without relying on explicit mode-switching heuristics.

Geometric Residual Parameterization. Directly predicting absolute state coordinates wastes model capacity on approximating the identity mapping during near-nominal execution. Instead, we predict residual trajectories relative to the current state. Let $\boldsymbol{\beta}_t$ denote the static baseline anchored at the current proprioceptive state:

$$\boldsymbol{\beta}_{t,k} = \mathbf{p}_t, \qquad \forall\, k \in \{0, \ldots, K-1\}. \tag{4}$$

The middleware predicts only the residual deviation $\mathbf{r}_t$, recovering the final trajectory as:

$$\boldsymbol{\tau}_t = \boldsymbol{\beta}_t + \mathbf{r}_t. \tag{5}$$

When $\mathbf{p}_t \approx \mathbf{m}_t$, the residual directly recovers the reference trajectory, reducing the middleware to an identity map. Under severe deviations, the model leverages the gap between $\mathbf{p}_t$ and $\mathbf{m}_t$ in its conditioning input to synthesize anthropomorphic trajectories that depart from the reference infeasible from the current state. The reference $\mathbf{m}_t$ enters only as a conditioning input and never as part of the target. The model therefore always predicts the motion increment from the current state over the next Δt; the magnitude of this target stays bounded however far $\mathbf{m}_t$ lies from $\mathbf{p}_t$.

Continuous Conditional Flow Matching. To synthesize the complex, multimodal distributions of these residuals, we employ continuous flow matching (Lipman et al., 2023). Let $\mathbf{x}_0$ denote the normalized ground-truth residual data and $\mathbf{x}_1 \sim \mathcal{N}(0, \mathbf{I})$ the prior noise. We define a probability path via linear interpolation:

$$\mathbf{x}_t = (1-t)\mathbf{x}_0 + t\mathbf{x}_1, \qquad t \in [0, 1]. \tag{6}$$

The model is trained to regress the continuous vector field by minimizing the velocity matching objective:

$$\mathcal{L}_{\text{vel}} = \mathbb{E}_{t,\mathbf{x}_0,\mathbf{x}_1}\left[\left\|\hat{\mathbf{v}}(\mathbf{x}_t, t, \mathbf{c}_t) - (\mathbf{x}_1 - \mathbf{x}_0)\right\|_2^2\right], \tag{7}$$

where $\mathbf{c}_t = [\mathbf{p}_t, \mathbf{m}_t]$ serves as the state-conditioning vector. During inference, anthropomorphic trajectories are sampled by integrating the learned velocity field $\hat{\mathbf{v}}$ from $t = 1$ to $t = 0$ via a minimal number of Euler steps.

Architecture and Kinematic Continuity. The velocity field is parameterized by an AdaLN-modulated Transformer (Peebles & Xie, 2023), where the conditioning vector $\mathbf{c}_t$ and the flow timestep embedding are injected into each block via adaptive shift-scale-gate modulation. To guarantee kinematic continuity between the real-time physical state and the synthesized trajectory, we pin the first residual token via an inpainting constraint during integration:

$$\mathbf{x}_t[0] = (1 - t_{\text{next}})\mathbf{r}_0 + t_{\text{next}}\boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I}), \tag{8}$$

where $\mathbf{r}_0$ is the zero-residual anchor corresponding to the initial state.

Receding-Horizon Planning with Directional Warm Start. By holding the temporal window Δt and execution interval N_{exec} constant regardless of the state-command gap, the middleware confines each prediction to a bounded local horizon; long-range transitions toward the reference consequently emerge through autoregressive replanning across successive cycles rather than single-pass trajectory optimization.

To accelerate convergence during inference, we replace the standard Gaussian initialization with a directional prior that linearly interpolates toward the target:

$$\mathbf{r}_k^{\text{init}} = \frac{k}{K-1}(\mathbf{m}_t - \mathbf{p}_t), \qquad k \in \{0, \ldots, K-1\}, \tag{9}$$

and begin ODE integration from a partially noised version of this prior at $t_{\text{start}} < 1$:

$$\mathbf{x}_{t_{\text{start}}} = (1 - t_{\text{start}})\,\text{normalize}(\mathbf{r}^{\text{init}}) + t_{\text{start}}\,\boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I}), \tag{10}$$

where t_{start} controls the noise-to-prior ratio. Analogous to SDEdit (Meng et al., 2022), this initialization supplies a coarse directional estimate that the learned velocity field progressively refines into a human-like trajectory, reducing the required ODE steps relative to pure-noise sampling. The recovered K keyframes define the reference trajectory over Δt, from which the tracker draws commands at its control frequency, closing the middleware–tracker loop.

3.3 General-Purpose Physics Tracker

3.3.1 Motion Tracking Formulation

Observation and Action Spaces. The observation $\mathbf{o}_t = \{\mathbf{p}_t, \mathbf{m}_t, \mathbf{z}_d\}$ comprises proprioceptive state, reference kinematics, and a discrete motion embedding. The proprioceptive state is

$$\mathbf{p}_t = [\mathbf{g}_t^{\text{proj}}, \boldsymbol{\omega}_t, \mathbf{q}_t - \mathbf{q}_0, \dot{\mathbf{q}}_t, \mathbf{a}_{t-1}], \tag{11}$$

comprising the projected gravity vector, root angular velocity, joint positions relative to the default pose $\mathbf{q}_0$, joint velocities, and the previous action. The reference $\mathbf{m}_t = [\mathbf{v}_t^{\text{ref}}, \boldsymbol{\omega}_t^{\text{ref}}, \mathbf{e}_t^{\text{root}}, \mathbf{q}_t^{\text{ref}}]$ provides target root linear and angular velocities, a 6D root orientation error, and target joint positions. The tracker is trained as a standalone motion tracker on dataset references and accepts middleware output $\mathbf{m}'_t$ through the same input at deployment, without retraining or modification. The policy outputs target joint positions $\mathbf{a}_t \in \mathbb{R}^{29}$, executed by per-joint PD controllers.

Rewards and Domain Randomization. The reward combines exponential tracking terms and regularization penalties, with no task- or recovery-specific shaping. Domain randomization over friction, CoM offsets, and external pushes supports sim-to-real transfer. Reward weights and randomization ranges are in Tab. 3.

3.3.2 General Motion Tracking Policy

A motion encoder maps the 10-frame reference window $\mathbf{M}_{t:t+9}$ to a continuous latent $\mathbf{z}_c$, which an iFSQ quantizer discretizes into the token $\mathbf{z}_d$. A reconstruction decoder recovers $\hat{\mathbf{M}}_{t:t+9}$ to shape the representation through

$$\mathcal{L}_{\text{rec}} = \frac{1}{10} \sum_{k=0}^{9} \|\hat{\mathbf{m}}_{t+k} - \mathbf{m}_{t+k}\|_2^2, \tag{12}$$

while an action decoder combines $\mathbf{z}_d$ with a 10-step proprioceptive history $\mathbf{P}_{t-9:t}$ to produce $\mathbf{a}_t$.

Discrete Motion Tokenization. Quantization maps reference segments, including noisy or out-of-distribution inputs, to a fixed codebook (Van Den Oord et al., 2017). Discrete codes help separate distinct behaviors, avoiding the collapse or poor sampling associated with strongly or weakly regularized continuous latents (Zhu et al., 2023). We adopt iFSQ (Lin et al., 2026a), a variant of Finite Scalar Quantization (FSQ) (Mentzer et al., 2024) that retains its fixed grid while improving grid-bin utilization. Each channel of the d-dimensional continuous latent is bounded to $[-1, 1]$ and quantized into $L = 2^b + 1$ uniformly spaced levels:

$$z_{d,j} = \text{round}\left(\frac{L-1}{2}\left(f(z_{c,j}) + 1\right)\right), \qquad j = 1, \ldots, d. \tag{13}$$

iFSQ replaces the $\tanh$ bound in FSQ with

$$f(x) = 2.0\,\sigma(1.6x) - 1, \tag{14}$$

encouraging approximately uniform bin utilization for Gaussian latents. The straight-through estimator passes gradients through rounding, and the resulting integer vector forms the token $\mathbf{z}_d$.

Adaptive Motion Sampling. To balance locomotion and agile skills, we adopt a difficulty-aware curriculum inspired by ZEST (Sleiman et al., 2026). Per-bin difficulty estimates are updated via exponential moving average of tracking scores, kernel-smoothed over neighboring temporal regions, and mixed with a uniform baseline to form the sampling distribution. Details are in App. A.

3.4 Training Paradigm for Unified Tracking and Generation

Dataset Construction. Training tuples are generated via receding-horizon sampling. Segment lengths ℓ are drawn from a log-uniform distribution $\ell \sim \exp(\mathcal{U}(\log \ell_{\min}, \log \ell_{\max}))$ with $\ell_{\min} = H$ (the planning horizon). From each segment, K uniformly spaced keyframes are extracted covering only the first H frames (corresponding to $\Delta t = H/\text{fps}$ seconds), regardless of the total segment length. The conditioning vector $\mathbf{c}_t = [\tilde{\mathbf{p}}_t, \mathbf{m}_t]$ pairs the start state with the reference command at the segment endpoint. The residual supervision is computed against the static baseline $\boldsymbol{\beta}_{t,k} = \tilde{\mathbf{p}}_t$

(Eq. (5)), so each training target encodes the motion increment over the next Δt seconds from the current state. This design ensures that (i) the model never predicts trajectories exceeding a fixed temporal horizon, bounding residual magnitude regardless of the state-command gap; (ii) successive starting points within the same motion yield overlapping local windows, naturally covering every phase of long motion sequences; and (iii) eliminating near-zero-length segments ($\ell < H$) removes the trivial zero-residual bias that otherwise dominates under MSE training.

Noisy-State Augmentation. Deployment introduces a systematic discrepancy between noisy physical state estimation and the clean reference commands available during training. To close this gap, we apply asymmetric start-state perturbations: only the initial proprioceptive state is corrupted with channel-wise Gaussian noise,

$$\tilde{\mathbf{p}}_t = \mathbf{p}_t + \boldsymbol{\epsilon}, \tag{15}$$

while the target reference remains clean. This asymmetric augmentation mirrors the deployment scenario where the robot's proprioceptive state reflects both sensor noise and accumulated tracking errors, improving robustness to real-world state-command discrepancies.

Kinematics-Aware Loss Weighting. Beyond the primary velocity matching objective (Eq. (7)), we introduce a kinematics-aware weighting scheme motivated by the observation that identical joint-space errors can induce vastly different body-space deviations depending on the current kinematic configuration. We weight each state dimension q_d by a pose-dependent Jacobian magnitude:

$$w_d(\mathbf{q}) = \sum_b \left\| \frac{\partial \mathbf{p}_b}{\partial q_d} \right\|_2^2 \approx \sum_b \left\| \frac{\mathbf{p}_b(\mathbf{q} + \delta e_d) - \mathbf{p}_b(\mathbf{q} - \delta e_d)}{2\delta} \right\|_2^2, \tag{16}$$

where $\mathbf{p}_b$ is the Cartesian position of body link b. This approximates the diagonal of $\mathbf{J}^\top \mathbf{J}$ evaluated at each training pose and keyframe, producing a weight tensor of shape (N, K, D) that captures pose-dependent lever arm geometry. Weights are normalized to unit mean per dimension, clamped to prevent zero-gradient dimensions, and precomputed via differentiable forward kinematics during dataset construction.

State Representation. We evaluate two state parameterizations: a 38D configuration (29 joint positions + 3D root position + 6D root orientation) and a 35D variant omitting root position. The 38D formulation enables the middleware to synthesize root translational commands, allowing autonomous correction of global positional drift. The 35D variant delegates root translation entirely to the reference motion, eliminating dependence on external positioning for localization-free deployment with onboard proprioception and IMU alone. All quantitative results use the 38D configuration unless otherwise noted.

4 Experiments

4.1 Implementation Details

All experiments use the Unitree G1 humanoid (1.32 m, ∼35 kg, 29 DoFs). Training runs in IsaacLab (Mittal et al., 2025) with 16,384 parallel environments on a single A100 GPU (physics at 200 Hz, policy at 50 Hz). The tracker is trained end-to-end via PPO; the flow matching generator is trained offline, predicting K=8 keyframes spanning a 0.2 s horizon and replanning at 25 Hz via 5 Euler steps. We provide **38D** and **35D** (localization-free) state variants. For real-world deployment, both networks are accelerated with TensorRT on the G1's onboard NVIDIA Jetson Orin NX; the 5-step ODE solve completes in ∼20 ms, comfortably within the 40 ms replan budget at 25 Hz, requiring no offboard computation. The training corpus is entirely public (App. B); evaluation uses MuJoCo on 1,000 unseen sequences (up to 20 s each), comprising 700 daily-activity sequences retargeted from AMASS, 150 fall-and-recovery sequences from a proprietary MoCap collection held out for evaluation, and 150 discretized references, each holding a standing keyframe that switches abruptly to a fallen keyframe and later to a risen keyframe, with pelvis displacements of 0.37 to 0.90 m between successive keyframes. Full training details, reward functions, motion dataset, and architecture specifications are in App. C.

4.2 Comparisons

We evaluate nine configurations varying policy architecture, tokenizer, observation design, and generative planning. All configurations trained in this work share the same simulator, robot, reward,

and training budget. The baselines are an **MLP** actor–critic, a TokenHSI-style **Transformer** (Pan et al., 2025b), a **VAE** with a continuous latent representation, and a **VQ-VAE** tokenizer (Sun et al., 2024), all with BeyondMimic observations (Liao et al., 2026). **SONIC** (Luo et al., 2026) is evaluated zero-shot from its officially released checkpoint with its original observation design. Our variants isolate the remaining design choices. **iFSQ$_{\text{BM}}$** pairs the iFSQ tokenizer with BeyondMimic observations. **iFSQ** follows the observation design of Sec. 3.3; **iFSQ$_{+\text{H}}$** adds explicit height features to iFSQ. **Heracles** adds the flow matching middleware to iFSQ. We report Completion Rate (CR; the fraction of frames with height error <0.3 m and orientation error <1.2 rad, Eq. (20)) and joint, root height, orientation, velocity errors, all defined in App. D.

Table 1: **Quantitative comparison on 1,000 held-out motions**, overall and per subset. Blue = best per metric within a subset. Colored %: teal = better, red = worse vs. MLP. Changes (%): $100(x-b)/b$, with b the same-subset MLP baseline.

Set	Method	CR (%) ↑	Joint Err (rad) ↓	Height Err (m) ↓	Ori Err (rad) ↓	LinVel Err (m/s) ↓
Overall (1,000)	MLP	84.1	1.065	0.112	0.338	0.216
	Transformer	82.1 (−2.4%)	1.080 (−1.4%)	0.124 (−10.7%)	0.344 (−1.8%)	0.097 (+55.1%)
	VAE	78.5 (−6.7%)	1.441 (−35.3%)	0.133 (−18.8%)	0.471 (−39.3%)	0.226 (−4.6%)
	VQ-VAE	91.9 (+9.3%)	1.883 (−76.8%)	0.072 (+35.7%)	0.214 (+36.7%)	0.230 (−6.5%)
	SONIC	84.6 (+0.6%)	1.610 (−51.2%)	0.109 (+2.7%)	0.306 (+9.5%)	0.198 (+8.3%)
	iFSQ$_{\text{BM}}$	86.0 (+2.3%)	1.339 (−25.7%)	0.096 (+14.3%)	0.330 (+2.4%)	0.235 (−8.8%)
	iFSQ$_{+\text{H}}$	85.3 (+1.4%)	1.458 (−36.9%)	0.094 (+16.1%)	0.362 (−7.1%)	0.174 (+19.4%)
	iFSQ	85.2 (+1.3%)	1.109 (−4.1%)	0.105 (+6.3%)	0.380 (−12.4%)	0.165 (+23.6%)
	Heracles	94.0 (+11.8%)	0.969 (+9.0%)	0.047 (+58.0%)	0.145 (+57.1%)	0.180 (+16.7%)
Daily (700)	MLP	95.0	0.717	0.048	0.184	0.203
	Transformer	94.3 (−0.7%)	0.726 (−1.3%)	0.049 (−2.1%)	0.181 (+1.6%)	0.086 (+57.6%)
	VAE	92.3 (−2.8%)	0.918 (−28.0%)	0.049 (−2.1%)	0.294 (−59.8%)	0.228 (−12.3%)
	VQ-VAE	97.4 (+2.5%)	1.674 (−133.5%)	0.038 (+20.8%)	0.162 (+12.0%)	0.228 (−12.3%)
	SONIC	95.7 (+0.7%)	1.089 (−51.9%)	0.040 (+16.7%)	0.176 (+4.3%)	0.178 (+12.3%)
	iFSQ$_{\text{BM}}$	96.2 (+1.3%)	0.945 (−31.8%)	0.037 (+22.9%)	0.175 (+4.9%)	0.222 (−9.4%)
	iFSQ$_{+\text{H}}$	96.3 (+1.4%)	0.994 (−38.6%)	0.035 (+27.1%)	0.169 (+8.2%)	0.151 (+25.6%)
	iFSQ	96.3 (+1.4%)	0.802 (−11.9%)	0.039 (+18.8%)	0.185 (−0.5%)	0.143 (+29.6%)
	Heracles	93.7 (−1.4%)	0.895 (−24.8%)	0.050 (−4.2%)	0.142 (+22.8%)	0.152 (+25.1%)
Recovery (150)	MLP	88.6	1.385	0.095	0.282	0.214
	Transformer	86.5 (−2.4%)	0.998 (+27.9%)	0.095 (0.0%)	0.279 (+1.1%)	0.090 (+57.9%)
	VAE	68.7 (−22.5%)	2.358 (−70.3%)	0.202 (−112.6%)	0.623 (−120.9%)	0.202 (+5.6%)
	VQ-VAE	90.8 (+2.5%)	2.424 (−75.0%)	0.094 (+1.1%)	0.272 (+3.5%)	0.231 (−7.9%)
	SONIC	88.8 (+0.2%)	1.727 (−24.7%)	0.087 (+8.4%)	0.267 (+5.3%)	0.209 (+2.3%)
	iFSQ$_{\text{BM}}$	84.8 (−4.3%)	1.855 (−33.9%)	0.113 (−18.9%)	0.382 (−35.5%)	0.223 (−4.2%)
	iFSQ$_{+\text{H}}$	90.7 (+2.4%)	1.826 (−31.8%)	0.073 (+23.2%)	0.311 (−10.3%)	0.183 (+14.5%)
	iFSQ	92.2 (+4.1%)	1.441 (−4.0%)	0.068 (+28.4%)	0.323 (−14.5%)	0.175 (+18.2%)
	Heracles	97.8 (+10.4%)	1.100 (+20.6%)	0.025 (+73.7%)	0.143 (+49.3%)	0.173 (+19.2%)
Discrete (150)	MLP	28.6	2.370	0.425	1.111	0.282
	Transformer	20.6 (−28.0%)	2.814 (−18.7%)	0.502 (−18.1%)	1.171 (−5.4%)	0.158 (+44.0%)
	VAE	23.4 (−18.2%)	2.966 (−25.1%)	0.457 (−7.5%)	1.145 (−3.1%)	0.240 (+14.9%)
	VQ-VAE	67.1 (+134.6%)	2.320 (+2.1%)	0.207 (+51.3%)	0.400 (+64.0%)	0.236 (+16.3%)
	SONIC	28.4 (−0.7%)	3.924 (−65.6%)	0.454 (−6.8%)	0.954 (+14.1%)	0.274 (+2.8%)
	iFSQ$_{\text{BM}}$	39.4 (+37.8%)	2.661 (−12.3%)	0.355 (+16.5%)	0.998 (+10.2%)	0.308 (−9.2%)
	iFSQ$_{+\text{H}}$	28.9 (+1.0%)	3.255 (−37.3%)	0.393 (+7.5%)	1.315 (−18.4%)	0.273 (+3.2%)
	iFSQ	26.3 (−8.0%)	2.208 (+6.8%)	0.451 (−6.1%)	1.344 (−21.0%)	0.258 (+8.5%)
	Heracles	92.0 (+221.7%)	1.181 (+50.2%)	0.055 (+87.1%)	0.159 (+85.7%)	0.314 (−11.3%)

Results. Tab. 1 reports quantitative performance. Heracles attains the highest CR overall (94.0%).

(i) Nominal tracking: On daily motions every configuration is near saturation; Heracles tracks with orientation error among the lowest and joint error within the trackers' range, consistent with the middleware acting as a near-identity map under nominal conditions.

(ii) Continuous fall-and-recovery: Adding the middleware raises the CR of iFSQ from 92.2% to 97.8%, the highest of all configurations. At the contact changes of a get-up motion, a few steps of accumulated tracking error displace the robot's state from the reference; the middleware then replaces the reference infeasible from the current state with a synthesized transition toward it.

(iii) Discretized references: Each discretized reference holds one keyframe and switches abruptly to the next, with no transition between them, so every switch leaves the reference far from the current state. On this subset Heracles outperforms every tracker by a wide margin. All standalone trackers except VQ-VAE achieve CR below 40%, failing catastrophically at each switch; VQ-VAE reaches 67.1% CR but retains large joint (2.320 rad) and orientation (0.400 rad) errors, indicating limited tracking fidelity. **Heracles achieves 92.0% CR**, exceeding VQ-VAE by 24.9 percentage points and more than doubling the CR of every other tracker, with the lowest joint error (1.181 rad) and orientation error (0.159 rad). On dense fall-and-recovery references every configuration reaches 68.7% to 97.8% CR; the divide arises at these switches. Heracles alone bridges this gap, a capability absent in every other configuration in Tab. 1.

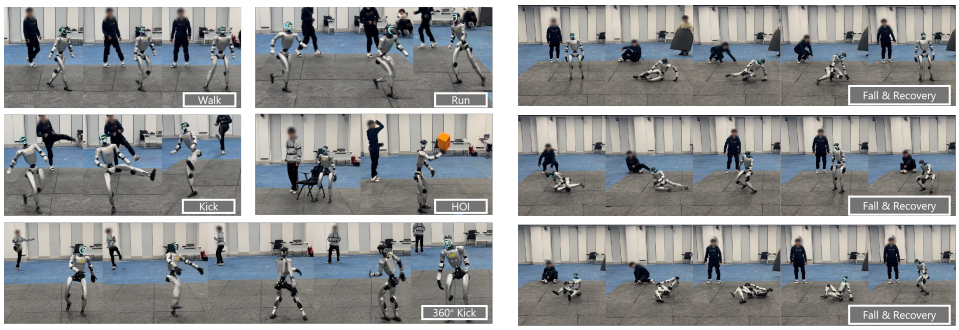

Figure 3: **Real-world deployment.** (Left) Motion tracking from everyday locomotion (walk, run) to dynamic skills (kick, 360° kick) and human-object interaction. (Right) Continuous fall-and-recovery tracking of a live performer. Our model tracks the reference throughout and also covers the deviations from the reference that arise on hardware, without task-specific engineering.

(iv) Tracker and tokenizer: The standalone iFSQ tracker with the proposed observations exceeds SONIC in CR on daily and fall-and-recovery sequences, at lower joint and velocity error. Among VAE, VQ-VAE, and iFSQ$_{\text{BM}}$ (identical except the latent), the continuous VAE trails iFSQ$_{\text{BM}}$ in CR and joint error (78.5% versus 86.0%; 1.441 versus 1.339 rad), and finite scalar quantization reduces joint error relative to VQ-VAE from 1.883 to 1.339 rad (-29%), while VQ-VAE attains the higher completion at the largest joint error of any tracker. Detailed per-component analysis and the learned codebook visualization are in App. E.

Further analysis. To complement the quantitative comparison in Tab. 1, we provide qualitative visual comparisons in MuJoCo (App. F). Heracles maintains accurate tracking on an out-of-distribution martial arts sequence and completes a challenging lie-to-stand transition, while several baselines fail or only partially track these motions. We further analyze how Heracles adapts to deviations from the reference (App. G), evaluate middleware transfer to the official SONIC checkpoint (App. H.1), and assess the human-likeness of synthesized recoveries through comparison with HumanUP (App. H.2).

4.3 Ablation Studies

To isolate each design choice within the generative middleware, we ablate the flow matching middleware while keeping the iFSQ tracker fixed (Tab. 7). Removing noisy-state augmentation causes the most severe degradation (CR: -11.3%), followed by kinematics-aware weighting (-5.6%) and the directional warm start (-4.2%). The 35D localization-free variant retains 92.4% CR overall and 86.7% CR on discretized references (Tab. 8), above every standalone tracker in Tab. 1.

4.4 Real-World Deployment

We deploy Heracles on the physical Unitree G1 (Fig. 3), demonstrating zero-shot sim-to-real transfer across everyday locomotion, dynamic kicks, and human-object interaction. For fall recovery, the robot tracks a live performer in real time across three lie-to-stand trials from supine, lateral, and prone configurations, rising toward different orientations. When disturbances or accumulated tracking error leave the reference infeasible from the current state, the middleware synthesizes the transition back toward the reference and the robot still rises to standing. Heracles therefore unifies nominal tracking and generative synthesis without task-specific engineering.

5 Conclusion

We presented Heracles, a state-conditioned generative middleware that resolves the longstanding dichotomy between strict kinematic tracking and generative synthesis in general-purpose humanoid control. By embedding continuous flow matching within a closed-loop hierarchical architecture, the framework dynamically bridges high-fidelity intent execution with anthropomorphic adaptability. Without relying on explicit mode-switching heuristics or separate recovery modules, Heracles intrinsically preserves zero-shot tracking fidelity under nominal conditions while seamlessly synthesizing human-like whole-body maneuvers under severe deviations from the reference motion. This unified paradigm liberates embodied control from rigid, reference-bound execution, establishing a scalable foundation for deploying agile and resilient general-purpose humanoids in complex physical environments.

AI USE STATEMENT

In this work, generative AI tools were used only to polish the wording of the manuscript text. They were not used for any task with required disclosure: not for generating data, developing the framework or method, formulating or proving claims, designing or running experiments, implementing methods, processing or analyzing data, or interpreting results. We have reviewed all AI-assisted text and take responsibility for the final content of this work.

REPRODUCIBILITY STATEMENT

The formulation, the middleware, and the tracker are specified in Sec. 3 with all objectives stated explicitly. App. C gives the simulation setup, reward function, domain-randomization ranges (Tab. 3), and all hyperparameters (Tab. 4); App. B lists the training corpus, which is entirely public; and App. D specifies every compared system and the evaluation protocol. Code, pre-trained models, and the sequence lists of the public part of the benchmark will be released upon acceptance.

REFERENCES

Lucas N Alegre, Agon Serifi, Ruben Grandia, David Müller, Espen Knoop, and Moritz Bächer. Amor: Adaptive character control through multi-objective reinforcement learning. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers*, pp. 1–11, 2025.

Penghui Chen, Yushi Wang, Changsheng Luo, Wenhan Cai, and Mingguo Zhao. Hifar: Multi-stage curriculum learning for high-dynamics humanoid fall recovery. In *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2908–2915. IEEE, 2025a.

Yeke Chen, Shihao Dong, Xiaoyu Ji, Jingkai Sun, Zeren Luo, Liu Zhao, Jiahui Zhang, Wanyue Li, Ji Ma, Bowen Xu, et al. Learning human-like badminton skills for humanoid robots. *arXiv preprint arXiv:2602.08370*, 2026.

Zixuan Chen, Mazeyu Ji, Xuxin Cheng, Xuanbin Peng, Xue Bin Peng, and Xiaolong Wang. Gmt: General motion tracking for humanoid whole-body control. *arXiv preprint arXiv:2506.14770*, 2025b.

Ke Fan, Shunlin Lu, Minyue Dai, Runyi Yu, Lixing Xiao, Zhiyang Dou, Junting Dong, Lizhuang Ma, and Jingbo Wang. Go to zero: Towards zero-shot motion generation with million-scale data. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 13336–13348. IEEE, 2025.

Félix G Harvey, Mike Yurick, Derek Nowrouzezahrai, and Christopher Pal. Robust motion in-betweening. *ACM Transactions on Graphics (TOG)*, 39(4):60–1, 2020.

Tairan He, Wenli Xiao, Toru Lin, Zhengyi Luo, Zhenjia Xu, Zhenyu Jiang, Jan Kautz, Changliu Liu, Guanya Shi, Xiaolong Wang, et al. Hover: Versatile neural whole-body controller for humanoid robots. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 9989–9996. IEEE, 2025a.

Xialin He, Runpei Dong, Zixuan Chen, and Saurabh Gupta. Learning getting-up policies for real-world humanoid robots. *arXiv preprint arXiv:2502.12152*, 2025b.

Inwoo Hwang, Jian Wang, Bing Zhou, et al. Snapmogen: Human motion generation from expressive texts. *Advances in Neural Information Processing Systems*, 38:99939–99955, 2026.

Biao Jiang, Xin Chen, Wen Liu, Jingyi Yu, Gang Yu, and Tao Chen. Motiongpt: Human motion as a foreign language. *Advances in Neural Information Processing Systems*, 36:20067–20079, 2023.

Eunjong Lee, Eunhee Kim, Sanghoon Hong, Eunho Jung, and Jihoon Kim. Controllable long-term motion generation with extended joint targets. In *2026 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 5164–5173. IEEE, 2026.

Yitang Li, Zhengyi Luo, Tonghe Zhang, Cunxi Dai, Anssi Kanervisto, Andrea Tirinzoni, Haoyang Weng, Kris Kitani, Mateusz Guzek, Ahmed Touati, et al. Bfm-zero: A promptable behavioral foundation model for humanoid control using unsupervised reinforcement learning. In *International Conference on Learning Representations*, volume 2026, pp. 79697–79725, 2026a.

Zhe Li, Weihao Yuan, Weichao Shen, Siyu Zhu, Zilong Dong, and Chang Xu. Omnimotion: Multi-modal motion generation with continuous masked autoregression. *arXiv preprint arXiv:2510.14954*, 2025.

Zhe Li, Yangyang Wei, Boan Zhu, Yibo Peng, Tao Huang, Pengwei Wang, Zhongyuan Wang, Cheng Chi, Chang Xu, and Shanghang Zhang. From language to locomotion: Retargeting-free humanoid control via motion latent guidance. In *International Conference on Learning Representations*, volume 2026, pp. 129292–129319, 2026b.

Qiayuan Liao, Takara E Truong, Xiaoyu Huang, Yuman Gao, Guy Tevet, Koushil Sreenath, and C Karen Liu. Beyondmimic: From motion tracking to versatile humanoid control via guided diffusion. *Science Robotics*, 11(117):eadx8924, 2026.

Bin Lin, Zongjian Li, Yuwei Niu, Kaixiong Gong, Yunyang Ge, Yunlong Lin, Mingzhe Zheng, JianWei Zhang, Miles Yang, Zhao Zhong, et al. ifsq: Improving fsq for image generation with 1 line of code. *arXiv preprint arXiv:2601.17124*, 2026a.

Jing Lin, Ailing Zeng, Shunlin Lu, Yuanhao Cai, Ruimao Zhang, Haoqian Wang, and Lei Zhang. Motion-x: A large-scale 3d expressive whole-body human motion dataset. *Advances in Neural Information Processing Systems*, 36:25268–25280, 2023.

Yutang Lin, Jieming Cui, Yixuan Li, Baoxiong Jia, Yixin Zhu, and Siyuan Huang. Lessmimic: Long-horizon humanoid interaction with unified distance field representations. *arXiv preprint arXiv:2602.21723*, 2026b.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *International Conference on Learning Representations (ICLR)*, 2023.

Zhengyi Luo, Jinkun Cao, Kris Kitani, Weipeng Xu, et al. Perpetual humanoid control for real-time simulated avatars. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10895–10904, 2023.

Zhengyi Luo, Jiashun Wang, Kangni Liu, Haotian Zhang, Chen Tessler, Jingbo Wang, Ye Yuan, Jinkun Cao, Zihui Lin, Fengyi Wang, et al. Smplolympics: Sports environments for physically simulated humanoids. *arXiv preprint arXiv:2407.00187*, 2024.

Zhengyi Luo, Ye Yuan, Tingwu Wang, Chenran Li, Fernando Castañeda, Sirui Chen, Zi-Ang Cao, Jiefeng Li, David Minor, Qingwei Ben, et al. Sonic: Supersizing motion tracking for natural humanoid whole-body control. *Science Robotics*, 11(117):eaed4592, 2026.

Naureen Mahmood, Nima Ghorbani, Nikolaus F Troje, Gerard Pons-Moll, and Michael J Black. Amass: Archive of motion capture as surface shapes. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 5442–5451, 2019.

Ian Mason, Sebastian Starke, and Taku Komura. Real-time style modelling of human locomotion via feature-wise transformations and local motion phases. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 5(1):1–18, 2022.

Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. SDEdit: Guided image synthesis and editing with stochastic differential equations. In *International Conference on Learning Representations (ICLR)*, 2022.

Fabian Mentzer, David Minnen, Eirikur Agustsson, and Michael Tschannen. Finite scalar quantization: VQ-VAE made simple. In *International Conference on Learning Representations (ICLR)*, 2024. arXiv:2309.15505.

Mayank Mittal, Pascal Roth, James Tigue, Antoine Richard, Octi Zhang, Peter Du, Antonio Serrano-Munoz, Xinjie Yao, René Zurbrügg, Nikita Rudin, et al. Isaac lab: A gpu-accelerated simulation framework for multi-modal robot learning. *arXiv preprint arXiv:2511.04831*, 2025.

Chaoyi Pan, Changhao Wang, Haozhi Qi, Zixi Liu, Homanga Bharadhwaj, Akash Sharma, Tingfan Wu, Guanya Shi, Jitendra Malik, and Francois Hogan. Spider: Scalable physics-informed dexterous retargeting. *arXiv preprint arXiv:2511.09484*, 2025a.

Liang Pan, Zeshi Yang, Zhiyang Dou, Wenjia Wang, Buzhen Huang, Bo Dai, Taku Komura, and Jingbo Wang. Tokenhsi: Unified synthesis of physical human-scene interactions through task tokenization. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5379–5391. IEEE, 2025b.

Yixuan Pan, Ruoyi Qiao, Li Chen, Kashyap Chitta, Liang Pan, Haoguang Mai, Qingwen Bu, Hao Zhao, Cunyuan Zheng, Ping Luo, et al. Agility meets stability: Versatile humanoid control with heterogeneous data. *arXiv preprint arXiv:2511.17373*, 2025c.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4172–4182. IEEE, 2023.

Xue Bin Peng, Pieter Abbeel, Sergey Levine, and Michiel Van de Panne. Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM Transactions On Graphics (TOG)*, 37(4):1–14, 2018.

Xue Bin Peng, Yunrong Guo, Lina Halper, Sergey Levine, and Sanja Fidler. Ase: Large-scale reusable adversarial skill embeddings for physically simulated characters. *ACM Transactions On Graphics (TOG)*, 41(4):1–17, 2022.

Agon Serifi, Ruben Grandia, Espen Knoop, Markus Gross, and Moritz Bächer. Vmp: Versatile motion priors for robustly tracking motion on physical characters. In *Computer graphics forum*, volume 43, pp. e15175. Wiley Online Library, 2024.

Jean Pierre Sleiman, He Li, Alphonsus Adu-Bredu, Robin Deits, Arun Kumar, Kevin Bergamin, Mohak Bhardwaj, Scott Biddlestone, Nicola Burger, Matthew A Estrada, et al. Zest: Zero-shot embodied skill transfer for athletic robot control. *Science Robotics*, 11(117):eaec7695, 2026.

Jingkai Sun, Qiang Zhang, Yiqun Duan, Xiaoyang Jiang, Chong Cheng, and Renjing Xu. Prompt, plan, perform: Llm-based humanoid control via quantized imitation learning. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 16236–16242. IEEE, 2024.

Guy Tevet, Sigal Raab, Brian Gordon, Yonatan Shafir, Daniel Cohen-Or, and Amit H Bermano. Human motion diffusion model. *arXiv preprint arXiv:2209.14916*, 2022.

Guy Tevet, Sigal Raab, Setareh Cohan, Daniele Reda, Zhengyi Luo, Xue Bin Peng, Amit Bermano, and Michiel Van de Panne. Closd: Closing the loop between simulation and diffusion for multi-task character control. In *International Conference on Learning Representations*, volume 2025, pp. 46506–46520, 2025.

Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. *Advances in neural information processing systems*, 30, 2017.

Yinhuai Wang, Qihan Zhao, Runyi Yu, Hok Wai Tsui, Ailing Zeng, Jing Lin, Zhengyi Luo, Jiwen Yu, Xiu Li, Qifeng Chen, et al. Skillmimic: Learning basketball interaction skills from demonstrations. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17540–17549. IEEE, 2025.

Yuan Wang, Di Huang, Yaqi Zhang, Wanli Ouyang, Jile Jiao, Xuetao Feng, Dan Xu, and Shixiang Tang. Motiongpt-2: A general-purpose motion-language model for motion generation and understanding. *IEEE Transactions on Circuits and Systems for Video Technology*, 2026a.

Yunshen Wang, Shaohang Zhu, Peiyuan Zhi, Yuhan Li, Jiaxin Li, Yong-Lu Li, Yuchen Xiao, Xingxing Wang, Baoxiong Jia, and Siyuan Huang. Omnixtreme: Breaking the generality barrier in high-dynamic humanoid control. *arXiv preprint arXiv:2602.23843*, 2026b.

Yuxin Wen, Qing Shuai, Di Kang, Jing Li, Cheng Wen, Yue Qian, Ningxin Jiao, Changhai Chen, Weijie Chen, Yiran Wang, et al. Hy-motion 1.0: Scaling flow matching models for text-to-motion generation. *arXiv preprint arXiv:2512.23464*, 2025.

Lixing Xiao, Shunlin Lu, Huaijin Pi, Ke Fan, Liang Pan, Yueer Zhou, Ziyong Feng, Xiaowei Zhou, Sida Peng, and Jingbo Wang. Motionstreamer: Streaming motion generation via diffusion-based autoregressive model in causal latent space. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10086–10096. IEEE, 2025.

Michael Xu, Yi Shi, KangKang Yin, and Xue Bin Peng. Parc: Physics-based augmentation with reinforcement learning for character controllers. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers*, pp. 1–11, 2025a.

Zhengjie Xu, Ye Li, Kwan-yee Lin, and Stella X Yu. Unified humanoid fall-safety policy from a few demonstrations. *arXiv preprint arXiv:2511.07407*, 2025b.

Lujie Yang, Xiaoyu Huang, Zhen Wu, Angjoo Kanazawa, Pieter Abbeel, Carmelo Sferrazza, C Karen Liu, Rocky Duan, and Guanya Shi. Omniretarget: Interaction-preserving data generation for humanoid whole-body loco-manipulation and scene interaction. *arXiv preprint arXiv:2509.26633*, 2025.

Kangning Yin, Zhe Cao, Wentao Dong, Weishuai Zeng, Tianyi Zhang, Qiang Zhang, Jingbo Wang, Jiangmiao Pang, Ming Zhou, and Weinan Zhang. Robostriker: Hierarchical decision-making for autonomous humanoid boxing. *arXiv preprint arXiv:2601.22517*, 2026a.

Kangning Yin, Weishuai Zeng, Ke Fan, Minyue Dai, Zirui Wang, Qiang Zhang, Zheng Tian, Jingbo Wang, Jiangmiao Pang, and Weinan Zhang. Unitracker: Learning universal whole-body motion tracker for humanoid robots. *IEEE Robotics and Automation Letters*, 2026b.

Runyi Yu, Yinhuai Wang, Qihan Zhao, Hok Wai Tsui, Jingbo Wang, Ping Tan, and Qifeng Chen. Skillmimic-v2: Learning robust and generalizable interaction skills from sparse and noisy demonstrations. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers*, pp. 1–11, 2025.

Mingqi Yuan, Tao Yu, Wenqi Ge, Xiuyong Yao, Dapeng Li, Huijiang Wang, Jiayu Chen, Bo Li, Wei Zhang, Wenjun Zeng, et al. A survey of behavior foundation model: Next-generation whole-body control system of humanoid robots. *IEEE transactions on pattern analysis and machine intelligence*, 2025.

Weishuai Zeng, Shunlin Lu, Kangning Yin, Xiaojie Niu, Minyue Dai, Jingbo Wang, and Jiangmiao Pang. Behavior foundation model for humanoid robots. *arXiv preprint arXiv:2509.13780*, 2025.

Qiang Zhang, Peter Cui, David Yan, Jingkai Sun, Yiqun Duan, Gang Han, Wen Zhao, Weining Zhang, Yijie Guo, Arthur Zhang, et al. Whole-body humanoid robot locomotion with human reference. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 11225–11231. IEEE, 2024.

Qiang Zhang, Jiahao Ma, Peiran Liu, Shuai Shi, Zeran Su, Zifan Wang, Jingkai Sun, Wei Cui, Jialin Yu, Gang Han, et al. Meshmimic: Geometry-aware humanoid motion learning through 3d scene reconstruction. *arXiv preprint arXiv:2602.15733*, 2026a.

Yuhong Zhang, Jing Lin, Ailing Zeng, Guanlin Wu, Shunlin Lu, Yurong Fu, Yuanhao Cai, Ruimao Zhang, Haoqian Wang, and Lei Zhang. Motion-x++: A large-scale multimodal 3d whole-body human motion dataset. *arXiv preprint arXiv:2501.05098*, 2025a.

Zewei Zhang, Kehan Wen, Michael Xu, Junzhe He, Chenhao Li, Takahiro Miki, Clemens Schwarke, Chong Zhang, Xue Bin Peng, and Marco Hutter. Learning whole-body humanoid locomotion via motion generation and motion tracking. *IEEE Robotics and Automation Letters*, 2026b.

Zhikai Zhang, Haofei Lu, Yunrui Lian, Ziqing Chen, Yun Liu, Chenghuai Lin, Han Xue, Zicheng Zeng, Zekun Qi, Shaolin Zheng, et al. Learning athletic humanoid tennis skills from imperfect human motion data. *arXiv preprint arXiv:2603.12686*, 2026c.

Ziyu Zhang, Sergey Bashkirov, Dun Yang, Yi Shi, Michael Taylor, and Xue Bin Peng. Physics-based motion imitation with adversarial differential discriminators. In *Proceedings of the SIGGRAPH Asia 2025 Conference Papers*, pp. 1–12, 2025b.

Kaifeng Zhao, Gen Li, and Siyu Tang. Dartcontrol: A diffusion-based autoregressive motion model for real-time text-driven motion control. In *International Conference on Learning Representations*, volume 2025, pp. 23569–23592, 2025.

Qingxu Zhu, He Zhang, Mengting Lan, and Lei Han. Neural categorical priors for physics-based character control. *ACM Transactions on Graphics (TOG)*, 42(6):1–16, 2023.

Shaoting Zhu, Baijun Ye, Jiaxuan Wang, Jiakang Chen, Ziwen Zhuang, Linzhan Mou, Runhan Huang, and Hang Zhao. Ttt-parkour: Rapid test-time training for perceptive robot parkour. *arXiv preprint arXiv:2602.02331*, 2026.

Ziwen Zhuang, Shaoting Zhu, Mengjie Zhao, and Hang Zhao. Deep whole-body parkour. *arXiv preprint arXiv:2601.07701*, 2026.

A Adaptive Motion Sampling

Training a unified policy over a large, heterogeneous motion corpus introduces severe optimization imbalance: uniform sampling overfits to trivial locomotion while underperforming on dynamic, agile skills. Inspired by the bin-level adaptive curriculum in ZEST (Sleiman et al., 2026), we design a continuous temporal-bin variant with smoothed difficulty propagation.

We partition the concatenated motion corpus into B uniformly spaced temporal bins. At each episode termination, the per-step tracking score $\rho \in [0, 1]$ is averaged over the episode to yield a difficulty estimate $1 - \bar{\rho}$, which is accumulated into the corresponding bin j via exponential moving average:

$$F_j \leftarrow \alpha\,(1 - \bar{\rho}) + (1 - \alpha)\,F_j, \tag{17}$$

where F_j denotes the difficulty of bin j and α controls the update rate. To prevent isolated hard bins from dominating the sampling distribution and to propagate difficulty to temporally adjacent regions, we apply a 1D kernel smoothing operation followed by an outlier cap:

$$\tilde{F}_j = \left(\mathcal{K} * \hat{F}\right)_j, \quad \hat{F}_j = \min\left(F_j,\, c \cdot \bar{F}\right), \tag{18}$$

where $\mathcal{K}$ is an exponentially decaying kernel and c bounds the maximum bin weight relative to the global mean $\bar{F}$. The final sampling distribution mixes the smoothed difficulty with a uniform baseline to ensure exploration:

$$P_j = \eta\,\tilde{F}_j + \frac{1 - \eta}{B}, \tag{19}$$

where η controls the balance between difficulty-driven and uniform sampling. This mechanism continuously steers the training distribution toward challenging temporal regions of the motion manifold, while the kernel smoothing ensures that difficulty information propagates to neighboring segments, preventing abrupt sampling discontinuities.

B Training Corpus

The training corpus (Tab. 2) is assembled from public sources only: SnapMoGen (Hwang et al., 2026), 100STYLE (Mason et al., 2022), LAFAN1 (Harvey et al., 2020), and AMASS (Mahmood et al., 2019). SnapMoGen clips were filtered for physical plausibility through the SPIDER pipeline (Pan et al., 2025a). AMASS jumping clips come from Transitions (39), EyesJapan (35), KIT (29), SOMA (3), and SFU (2); crawling clips from KIT (29), Transitions (7), ACCAD (6), and SFU (1). Fall-and-recovery training content consists of six LAFAN1 clips (0.63% of frames), identical for every method we trained. No proprietary motion capture is used for training; the proprietary recovery collection appears only in the evaluation benchmark.

Table 2: **Training corpus composition** (1,268 clips, 41.5 h at 50 fps). †: filtered via SPIDER.

Source	Clips	Clip %	Frames	Frame %	Duration
SnapMoGen†	871	68.7%	4,975,347	66.6%	27.6 h
100STYLE	212	16.7%	2,037,235	27.3%	11.3 h
LAFAN1	34	2.7%	360,216	4.8%	2.0 h
of which fall-and-get-up	6	0.47%	46,729	0.63%	15.6 min
AMASS jumping†	108	8.5%	76,038	1.0%	0.4 h
AMASS crawling	43	3.4%	17,842	0.2%	0.1 h
Total	1,268		7,466,678		41.5 h

C Training Configuration

Simulation Environment. All experiments are conducted on the Unitree G1 humanoid platform, a full-size bipedal robot standing approximately 1.32 m tall with a total mass of roughly 35 kg. The robot features 29 actuated degrees of freedom spanning the torso, two 7-DoF arms, and two 6-DoF legs, all driven by proprietary electric actuators. Training is carried out in IsaacLab (Mittal

et al., 2025), a GPU-accelerated simulator built on NVIDIA Isaac Sim, where we instantiate 16,384 parallel environments on a single NVIDIA A100 (80 GB) GPU. The physics simulation runs at a 5 ms timestep (200 Hz) on flat ground with randomized friction coefficients (Tab. 3b), while the control policy queries observations and emits actions at 50 Hz (every 4 simulation substeps). Each action is converted to joint torques by a per-joint PD controller executing at the full 200 Hz rate. For evaluation, all policies are tested in the MuJoCo physics engine on a held-out set of 1,000 unseen motion sequences (Sec. 4.1), comprising 700 daily-activity sequences retargeted from AMASS, 150 fall-and-recovery sequences from a proprietary motion-capture collection held out for evaluation, and 150 discretized references in which continuous reference trajectories are replaced with piecewise-constant signals consisting of static poses separated by abrupt transitions, thereby removing all smooth interpolation and testing the policy's ability to track discontinuous commands. Each sequence is rolled out for its full duration (up to 20 s).

Tracker Training. The general-purpose physics tracker is trained via Proximal Policy Optimization (PPO) within an asymmetric actor–critic framework, following established practices in prior whole-body tracking systems (Liao et al., 2026; Luo et al., 2026). The iFSQ encoder, reconstruction decoder, and action decoder are trained jointly end-to-end with the PPO objective, where the reconstruction loss (Eq. (12)) and the RL policy gradient share the same encoder–quantizer pathway. The complete reward function and domain randomization ranges are specified in Tab. 3; all remaining hyperparameters are listed in Tab. 4.

Table 3: **Training configuration for the general-purpose physics tracker.** Left: reward function with exponential tracking terms and regularization penalties. Right: domain randomization ranges.

(a) Reward function. Tracking: $r_i = w_i \exp\!\left(-\|e_i\|^2/\sigma_i^2\right)$.

Tracked Quantity	w	σ
Root Position	0.1	0.30
Root Orientation	0.5	0.40
Root Linear Vel.	1.0	0.50
Root Angular Vel.	1.0	1.00
Rel. Body Position	1.0	0.30
Rel. Body Orientation	1.0	0.40
Body Linear Vel.	1.0	1.00
Body Angular Vel.	1.0	$\sqrt{\pi}$
Regularization	w	**Penalty**
Action Rate	-0.1	$\|a_t - a_{t\text{-}1}\|^2$
Joint Limit	-10	$\sum \max(0,\ q - q_{\lim})$
Undesired Contacts	-0.1	$\sum \mathbb{1}(F_c > 1)$

(b) Domain randomization ranges.

Parameter	Range
Command-level perturbations	
Target Joint Pos. (rad)	$\pm\,0.01$
Target Ang. Vel. (rad/s)	$\pm\,0.2$
Target Root Rot. (rad)	$\pm\,0.05$
Default Joint Pos. (rad)	$\pm\,0.01$
Physical properties	
CoM Offset (m)	x: $\pm\,0.05$; y, z: $\pm\,0.1$
Static Friction	$[0.3,\ 1.6]$
Dynamic Friction	$[0.3,\ 1.2]$
External disturbances	
Push Frequency (s)	$1.0 - 3.0$
Push Lin. Vel. (m/s)	xy: $\pm\,0.5$; z: $\pm\,0.2$
Push Ang. Vel. (rad/s)	RP: $\pm\,0.52$; Y: $\pm\,0.78$

Flow Matching Middleware Training. The state-conditioned flow matching middleware is trained offline on paired trajectory tuples extracted from the motion corpus following the dataset construction procedure described in Sec. 3.4. The velocity field $\hat{\mathbf{v}}$ is parameterized by an AdaLN-modulated Transformer (Peebles & Xie, 2023) that predicts $K{=}8$ uniformly spaced keyframes over a fixed planning horizon of $\Delta t{=}0.2$ s. We train two state representation variants: a **38D** configuration encoding 29 joint positions, 3D root position, and 6D root orientation, and a **35D** variant that omits root position entirely. The 38D model retains full awareness of global positioning, enabling the middleware to synthesize root translational commands and correct global positional drift. The 35D model relies solely on joint encoders and an IMU for root orientation, making it directly deployable on hardware without external localization; global position tracking is delegated to the reference motion source. We optimize the velocity matching objective (Eq. (7)) jointly with the kinematics-aware loss weighting (Eq. (16)). Noisy-state augmentation (Eq. (15)) is applied throughout training with channel-wise Gaussian noise with reduced magnitudes for root pose and orientation channels. During deployment, trajectory samples are generated via 5 Euler integration steps from $t{=}0.9$ to $t{=}0$, initialized with the directional warm start (Eq. (10)) at $t_{\text{start}}{=}0.9$. The generator replans every $N_{\text{exec}}{=}2$ control steps (0.04 s), yielding a closed-loop replanning frequency of 25 Hz. Full hyperparameters for both the tracker and the flow matching middleware are listed in Tab. 4.

Table 4: **Training hyperparameters for the physics tracker and flow matching middleware.**

Generator Hyperparameter	Value
Attention blocks / heads / dim	6 / 4 / 512
Conditioning injection	AdaLN ($\mathbf{c}_t$=[$\mathbf{p}_t$, $\mathbf{m}_t$] + timestep)
Keyframes K / horizon Δt	8 / 0.2 s
State dimension D	38 / 35
Optimizer	AdamW (β_1=0.9, β_2=0.999)
Learning rate	1×10^{-4} (cosine decay)
Weight decay	10^{-4}
Batch size	256
Training epochs	4,000
Parameters	22.9 M
Inference ODE steps	5 (Euler, t: 0.9→0)
Warm-start t_{start}	0.9
Replan interval N_{exec}	2 steps (0.04 s, 25 Hz)

Tracker Hyperparameter	Value
Parallel environments	16,384
Rollout horizon	24 steps
Discount factor γ	0.99
GAE λ	0.95
PPO epochs / mini-batches	5 / 4
Clipping ratio ϵ	0.2
Actor learning rate	2×10^{-3}
Critic learning rate	1×10^{-3}
PPO KL target	0.01
Entropy coefficient	0.005
Gradient clip norm	1.0
Total training iterations	∼100,000

D Compared Methods, Evaluation Protocol, and Architectures

Compared Methods. All policies trained in this work receive a 10-frame proprioceptive history and a 10-frame future motion reference window (frames t through t+9). We organize the evaluated methods into two groups.

Architecture and external baselines. **MLP** employs a standard MLP actor–critic with BeyondMimic-style (BM) observations (Liao et al., 2026). **Transformer** follows the TokenHSI architecture (Pan et al., 2025b), encoding proprioceptive and motion inputs into per-modality tokens that are processed by a multi-head Transformer with a learnable aggregation token before an MLP action head.

VQ-VAE replaces the iFSQ quantizer with a VQ-VAE tokenizer (Sun et al., 2024) while retaining the same encoder–decoder policy structure. **VAE** retains this structure but replaces discrete quantization with a 512-dimensional diagonal-Gaussian latent distribution. The encoder predicts the latent mean and variance; the action decoder uses the posterior mean together with proprioceptive observations, while the reconstruction decoder uses a sampled latent to reconstruct the motion observations. Training combines PPO with motion reconstruction and KL regularization toward a standard-normal prior. **SONIC** (Luo et al., 2026) serves as an external baseline, evaluated zero-shot using its officially released checkpoint and original observation design.

Heracles variants. **iFSQ$_{\text{BM}}$** pairs the iFSQ tokenizer and encoder–decoder policy with BeyondMimic observations, isolating the tokenizer contribution under a standard observation design. **iFSQ$_{\text{+H}}$** combines the iFSQ-based policy with the proposed observation design (Sec. 3.3) including explicit height features. **iFSQ** uses the proposed observation design without height, representing the best standalone tracker configuration. **Heracles** augments the iFSQ tracker with the state-conditioned flow matching middleware (Eq. (1)), constituting the full proposed system.

The component configuration of each variant is summarized in Tab. 5.

Evaluation Protocol. All methods are evaluated on the same held-out set of 1,000 motion sequences unseen during training, spanning locomotion, dance, sports, daily activities, continuous fall-and-recovery, and discretized references. Each sequence is rolled out for its full duration (up to 20 s). We report five metrics. (i) *Completion Rate (CR)* is the fraction of reference frames for which the policy maintains a root height error below 0.3 m and a root orientation error below 1.2 rad,

$$\mathrm{CR} = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\big[\,|h_t - h_t^{\mathrm{ref}}| < 0.3\,\mathrm{m} \;\wedge\; \theta_t < 1.2\,\mathrm{rad}\,\big], \tag{20}$$

where T is the number of reference frames, h_t and h_t^{ref} the robot and reference root heights, and θ_t the root orientation error excluding yaw; episodes are not terminated at a failed frame, so later frames count again once the thresholds are met. (ii) *Joint Position Error* is the L_2 norm of joint-position deviations; (iii) *Root Height Error* the absolute height deviation in the world frame; (iv) *Root Orientation Error* the orientation error excluding yaw; and (v) *Root Linear Velocity Error* the velocity error in the body frame. All metrics are computed against the original reference motion for every method, including Heracles, whose rewritten reference does not enter the scoring.

Detailed network architecture specifications are provided in Tab. 6.

Table 5: **Component configuration of all evaluated methods.** Obs. column: BM = BeyondMimic-style (Liao et al., 2026); † = proposed (Sec. 3.3); +H = with height features. ✓/✗ indicates presence or absence of the middleware.

Method	Obs.	Tokenizer	Policy	Middleware
MLP	BM	–	MLP	✗
Transformer	BM	–	Trans.	✗
VAE	BM	Gaussian	Enc-Dec	✗
VQ-VAE	BM	VQ	Enc-Dec	✗
SONIC	SONIC	–	SONIC	✗
iFSQ$_{\mathrm{BM}}$	BM	iFSQ	Enc-Dec	✗
iFSQ$_{+\mathrm{H}}$	†+H	iFSQ	Enc-Dec	✗
iFSQ	†	iFSQ	Enc-Dec	✗
Heracles	†	iFSQ	Enc-Dec	✓

Table 6: **Network architecture specifications for all evaluated methods.** $[\cdot]$ denotes MLP hidden-layer widths. iFSQ$_*$ covers all iFSQ variants.

Method	Component	Configuration
MLP	Actor	$[4096, 2048, 1024, 512, 256]$ MLP
	Critic	$[4096, 2048, 1024, 512, 256]$ MLP
Transformer	Tokenizer	$[512, 512]$ MLP $\times$ 2 modalities $\rightarrow$ 3 tokens (512-dim)
	Backbone	3-layer, 4-head, 512-dim Transformer
	Action head	$[2048, 1024, 256]$ MLP
	Critic	$[3072, 1536, 768, 512]$ MLP
VAE	Latent	512-dim continuous diagonal Gaussian
	Policy	Enc-Dec; posterior mean to action head
	Regularizer	KL to standard normal prior
VQ-VAE	Quantizer	Codebook $\lvert\mathcal{C}\rvert$=10,240, dim=512
	Policy	Enc-Dec (identical to iFSQ)
SONIC	–	Official release (Luo et al., 2026); stride-5 reference sampling
iFSQ$_*$	Policy	iFSQ Enc-Dec (Sec. 3.3)
Heracles	Tracker	iFSQ Enc-Dec (Sec. 3.3)
	Middleware	6-layer, 4-head, 512-dim AdaLN Transformer (Peebles & Xie, 2023)

E Ablation Details and Token Usage

The ablation results are summarized in Tab. 7. Here we provide detailed per-component analysis.

Table 7: **Middleware ablation** on the 1,000-motion benchmark (overall), same iFSQ tracker. Percentages are relative changes from Heracles (full), computed from unrounded values.

Variant	CR (%) ↑	Joint (rad) ↓	Height (m) ↓	Ori (rad) ↓	LinVel (m/s) ↓
Heracles (full)	94.0	0.969	0.047	0.145	0.180
w/o directional warm start	90.1 (−4.2%)	1.097 (+13.2%)	0.069 (+46.2%)	0.214 (+48.0%)	0.186 (+3.5%)
w/o noisy-state augmentation	83.4 (−11.3%)	1.408 (+45.3%)	0.107 (+125.6%)	0.312 (+115.9%)	0.181 (+1.1%)
w/o kinematics-aware weighting	88.8 (−5.6%)	1.213 (+25.2%)	0.075 (+58.6%)	0.250 (+72.6%)	0.195 (+8.6%)
Local 35-D (no root position)	92.4 (−1.7%)	1.044 (+7.8%)	0.059 (+24.7%)	0.171 (+18.3%)	0.185 (+2.8%)

Table 8: **Global (38-D) vs. local (35-D) middleware on the discretized-reference subset** (150 references). The best standalone tracker on this subset reaches 67.1% CR (Tab. 1).

Variant	CR (%) ↑	Joint (rad) ↓	Height (m) ↓	Ori (rad) ↓	LinVel (m/s) ↓
Heracles (38-D)	92.0	1.181	0.055	0.159	0.314
Heracles (35-D)	86.7	1.331	0.086	0.221	0.237

Directional Warm Start. Replacing the directional motion prior (Eq. (10)) with pure Gaussian initialization degrades all metrics: completion drops from 94.0% to 90.1% (−4.2%) and joint error increases by 13.2%. The warm start seeds the ODE solver with a coarse linear interpolation toward the target, enabling the learned velocity field to focus its refinement budget on naturalness rather than gross direction estimation. Without this prior, the generator must expend additional integration steps to discover the heading toward the reference.

Noisy-State Augmentation. Removing the asymmetric noise injection (Eq. (15)) during training produces the most severe degradation in completion rate among all ablation variants, with CR falling to 83.4% (−11.3%) and joint error increasing by 45.3%. Height error more than doubles (+125.6%), and orientation error increases by 115.9%. Without noise augmentation, the generator overfits to clean state inputs; at deployment, accumulated tracking drift and sensor noise push the conditioning state away from the training distribution, causing catastrophic failure on out-of-distribution motions. The augmented variant bridges this train–deploy distribution gap, enabling robust trajectory generation even from noisy proprioceptive readings.

Kinematics-Aware Loss Weighting. Removing the Jacobian-based weighting (Eq. (16)) produces the second-largest degradation in completion rate, with CR falling to 88.8% (−5.6%). Height error increases by 58.6% and orientation error by 72.6%. The severity of this ablation indicates that pose-dependent lever-arm geometry is critical for robust tracking: a unit-radian shoulder error in an extended-arm configuration induces far larger Cartesian displacement than the same error with arms at rest, and the weighting scheme enables the generator to prioritize these geometrically sensitive configurations.

State Representation: 38D vs. 35D. The 35D variant, which omits root position from the state representation, achieves 92.4% CR (−1.7%) overall with comparable joint error (1.044 vs. 0.969 rad) and linear velocity error (0.185 vs. 0.180 m/s). The gap widens on discretized references (Tab. 8), where 35D CR drops to 86.7% (−5.7% vs. 38D), above the best standalone tracker in Tab. 1. This is expected: synthesizing the transition to a distant keyframe benefits from explicit root position awareness, which enables the middleware to plan global translational corrections. Nevertheless, the 35D configuration retains strong overall performance, validating its suitability for localization-free deployment where only onboard proprioception and IMU are available.

Learned Discrete Representation. Beyond component-level ablations, we examine whether the iFSQ tokenizer acquires a semantically structured codebook after training on the full motion corpus. Fig. 4 visualizes the discrete code activations projected into a three-dimensional embedding space, with each point representing a quantized motion token colored by its source motion category. The visualization reveals clearly separable clusters for walking, running, jumping, martial arts, dance, parkour, crawling, balance, and fall recovery, even though the quantizer receives no explicit category labels during training. Notably, semantically related skills occupy neighboring regions. The walking and running clusters lie adjacent, while crawling and balance form a separate group, indicating that the iFSQ codebook captures meaningful kinematic similarity structure. This emergent organization

confirms that the finite scalar quantization not only compresses high-frequency motion signals into compact tokens but also distills a structured motion taxonomy that enables the downstream action decoder and flow matching middleware to reason over semantically coherent motion abstractions.

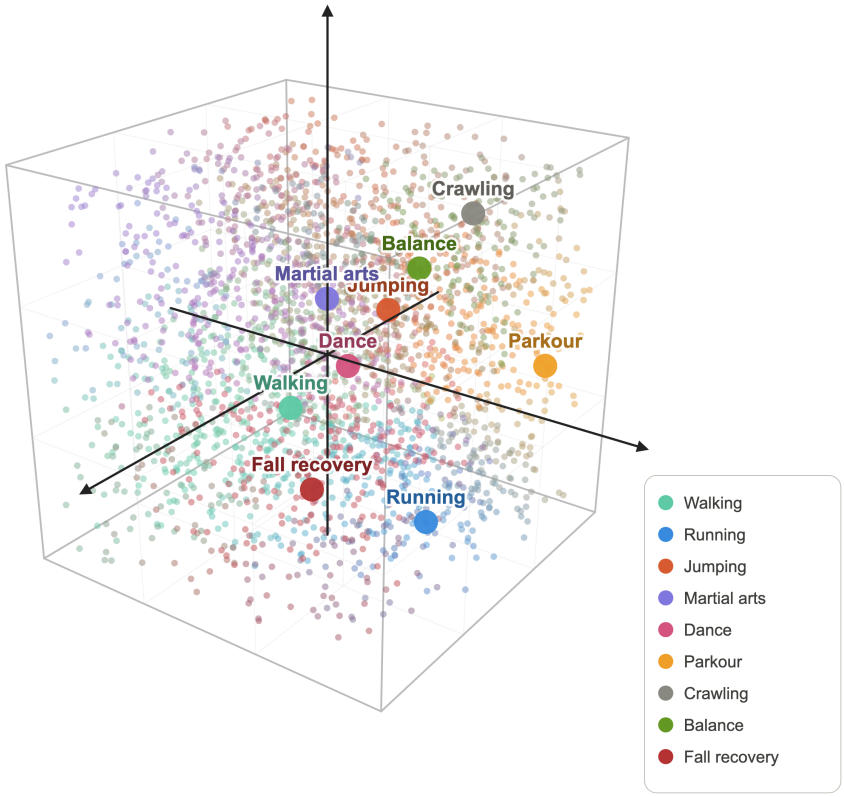

Figure 4: **Emergent semantic clustering of iFSQ token usage.** Each point is a quantized motion token of the trained tracker projected to 3D via PCA; colors denote motion categories. Without category labels during training, token usage organizes into clusters corresponding to distinct motor skills.

F Additional Qualitative Results

We provide qualitative visualizations that complement the quantitative evaluations in the main text. Fig. 5 presents frame-by-frame rollouts on an out-of-distribution martial arts sequence, while Fig. 6 shows a challenging lie-to-stand transition. In both cases, reactive baselines (MLP, VAE, Transformer, SONIC) collapse early, whereas Heracles maintains accurate tracking throughout. These results visually confirm the quantitative advantages reported in Tab. 1. Real-world deployment results—zero-shot sim-to-real transfer and omnidirectional fall recovery—are presented in the main text (Fig. 3).

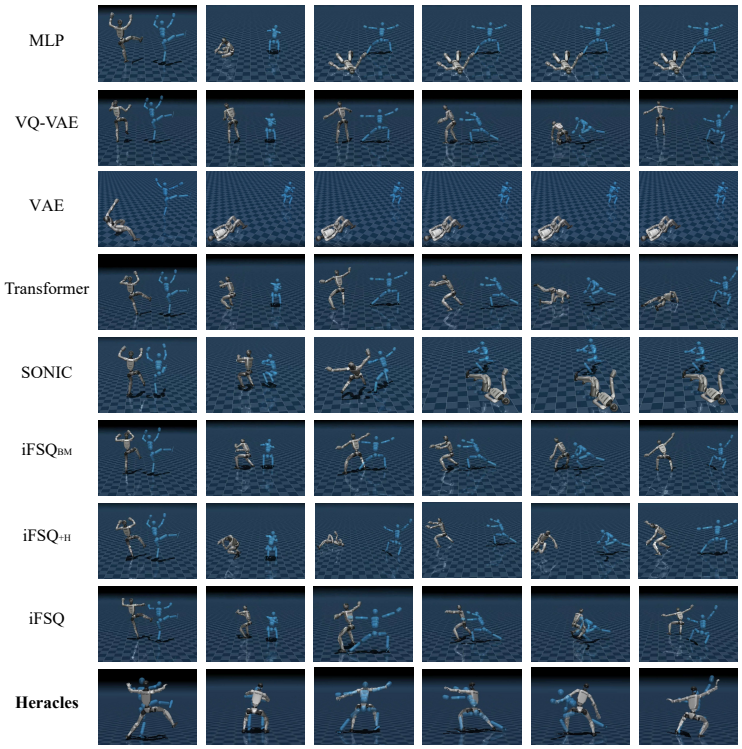

Figure 5: **Qualitative sim-to-sim comparison on an OOD martial arts sequence.** MLP, VAE, Transformer, and SONIC collapse early; VQ-VAE barely tracks. iFSQ_{BM} and iFSQ survive without falling, while $\text{iFSQ}_{+\text{H}}$ falls but recovers. **Heracles** tracks most accurately throughout.

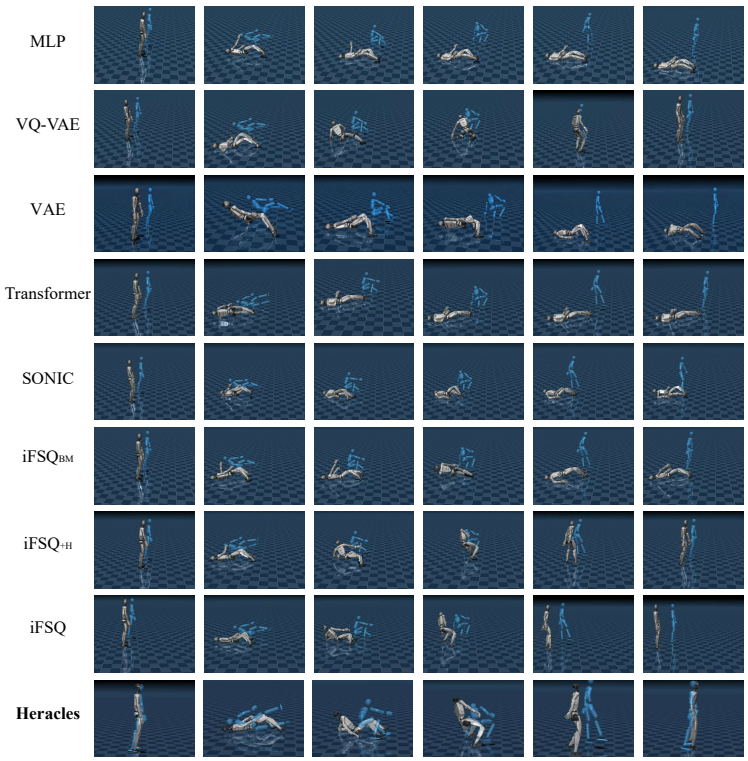

Figure 6: **Qualitative sim-to-sim comparison on an OOD lie-to-stand sequence.** MLP, VAE, Transformer, SONIC, and iFSQ_{BM} fail; VQ-VAE, $\text{iFSQ}_{\text{+H}}$, and iFSQ partially track. **Heracles** (Ours) completes the full transition and most accurately tracks root position.

G Behavioral Analysis

Beyond quantitative metrics, we examine the qualitative character of whole-body behaviors to understand *how* the generative middleware transforms the robot's response to severe deviations from the reference motion. This analysis reveals a fundamental distinction between a standalone tracker and the same tracker augmented with the middleware.

Tracking-Only Failure Modes. Once the robot's state lies far from the reference, standalone trackers, including all baselines and the iFSQ variant, exhibit a characteristic failure pattern. Upon being displaced from the reference trajectory, the tracker computes a single-step corrective action that minimizes the instantaneous state-reference error. This myopic strategy produces rigid, jerky corrective torques that lack the temporal coordination required for a multi-step return to the reference. In the most severe cases, the tracker's insistence on returning to the exact reference pose forces physically infeasible joint configurations, accelerating rather than preventing tracking failure. Even when the tracker avoids catastrophic failure, its corrective motions appear distinctly non-anthropomorphic: abrupt whole-body stiffening, unnatural arm postures, and a conspicuous absence of the compensatory stepping strategies that characterize human motor control.

Generative Whole-Body Maneuvers. Heracles produces qualitatively different whole-body behaviors. Once the robot's state lies far from the reference manifold, its generative middleware synthesizes a short-horizon anthropomorphic transition toward the reference. This manifests as emergent human-like strategies such as compensatory stepping to widen the base of support, coordinated arm counter-motions to redistribute angular momentum, and gradual torso realignment before resuming the original motion. From a fallen configuration the same mechanism yields rolling, propping, and kneeling before a rise to standing (Fig. 3). Crucially, these behaviors are not hand-designed or reward-engineered; they emerge naturally from the flow matching model's learned distribution over human motion transitions, conditioned on the robot's real-time state.

Anthropomorphic Behaviors Without Reward Engineering. The distinction has an architectural origin. A step-wise tracking reward penalizes the tracking error at every step, and the discount factor weights the earliest steps most heavily, so a discounted-return objective favors reducing the error as early as possible. Multi-phase transitions such as rolling, propping, and kneeling pass through intermediate configurations in which the tracking error temporarily increases, and the objective itself therefore offers little incentive to produce them.

The middleware obtains these behaviors from a single regression objective. It is trained by conditional regression on human motion capture, minimizing reconstruction error against recorded transitions instead of maximizing a return. No signal therefore favors faster completion, and no discriminator, style reward, or auxiliary loss enters its training. Its outputs reflect the speed, coordination, and contact sequencing of the human motion corpus. Conditioned on the robot's real-time state at every replanning step, the middleware synthesizes a trajectory tailored to the current configuration; different fallen configurations therefore produce different recovery motions (Fig. 3). The anthropomorphic, multi-phase behaviors observed in practice thus arise from the regression objective itself.

We further evaluate this generalization capability on omnidirectional fall recovery tasks in both simulation and the real world. As shown in Fig. 6, Heracles completes the full lie-to-stand transition in MuJoCo while all baseline methods fail or only partially succeed. Fig. 3 presents the corresponding real-world results: across three trials, the robot is initialized in distinct fallen configurations—supine, lateral, and prone postures. In all cases, the robot successfully executes a complete lie-to-stand recovery by following the reference motion, progressively transitioning through intermediate support phases. Critically, the recovery directions vary across trials, with the robot rising toward different orientations relative to its initial fallen heading, demonstrating true omnidirectional recovery rather than a memorized fixed-direction strategy.

H Middleware Transferability and Human-Likeness

Tab. 9 compares the three classes of system along three capabilities. General continuous motion tracking denotes following a dense reference trajectory; SONIC and Heracles both track such references. Dedicated recovery policies accept no reference motion, and tracking of both kinds therefore lies outside their scope. Discrete (keyframe) reference tracking denotes following a

reference given as keyframes alone, with no trajectory between them; Heracles synthesizes the motion between keyframes through the middleware, whereas for a tracker each keyframe is a reference infeasible from the current state (Tab. 1). Autonomous fall recovery denotes rising from a fallen configuration without a get-up reference; dedicated recovery policies restore standing from diverse fallen configurations, and Heracles does so with a single standing frame as its only reference (Tab. 10). Given the same standing frame, SONIC does not rise (Tab. 10); given a continuous get-up trajectory, it tracks the trajectory as any other continuous reference (Tab. 1).

Table 9: **Capabilities of dedicated recovery policies, a general tracker, and Heracles.**

Capability	HumanUP	HiFAR	FIRM	SONIC	Heracles
General continuous motion tracking	✗	✗	✗	✓	✓
Discrete (keyframe) reference tracking	✗	✗	✗	✗	✓
Autonomous fall recovery	✓	✓	✓	✗	✓

H.1 Transferability of the Middleware

The middleware communicates with a tracker solely through the reference-motion input and can therefore be attached to an external tracker. To isolate its contribution from that of our own tracker, we attach it to the official SONIC checkpoint and evaluate SONIC, SONIC with the middleware, iFSQ, and Heracles. SONIC's weights, observations, and gains remain untouched, and the middleware is attached to SONIC exactly as to our own tracker. Any difference between SONIC with and without the middleware therefore arises from the middleware alone. The benchmark consists of 40 fallen configurations with headings spaced uniformly over 360° to test omnidirectional recovery. In each configuration the robot, tipped along its heading into a braced ground posture, receives a static standing pose as reference for 20 s. A configuration counts as recovered when the pelvis remains above 0.66 m for 1 s (Tab. 10).

Table 10: **Middleware transfer on 40 fallen configurations.** Errors are whole-episode means against the standing reference.

System	**Success ↑**	**Height Err (m) ↓**	**Joint Err (rad) ↓**	**Ori Err (rad) ↓**
SONIC	0/40 (0%)	0.623 ± 0.055	4.503 ± 0.463	2.052 ± 0.527
SONIC + middleware	18/40 (45.0%)	0.398 ± 0.089	3.592 ± 0.250	1.502 ± 0.187
iFSQ	0/40 (0%)	0.567 ± 0.025	2.591 ± 0.315	1.623 ± 0.180
Heracles (iFSQ + middleware)	37/40 (92.5%)	0.275 ± 0.057	3.278 ± 0.425	1.142 ± 0.169

Given the standing reference alone, neither tracker recovers (Tab. 10). With the middleware attached, SONIC recovers in 18 of 40 configurations and every error decreases; iFSQ with the same middleware recovers in 37 of 40. The remaining gap reflects the mismatch between the middleware's output and SONIC's reference interface rather than tracker capability. The middleware emits a dense 0.2 s transition trajectory that fills iFSQ's ten-frame command window and only a fraction of SONIC's longer and sparser reference window. Both interfaces are kept unmodified, since the transfer is meant to test the middleware itself rather than an interface tuned to SONIC.

H.2 Human-Likeness

To assess the human-likeness of the synthesized recoveries, we compare them with the get-up motions discovered by HumanUP (He et al., 2025b) through reinforcement learning exploration in simulation. Heracles is evaluated over 100 trials from fallen configurations with uniformly sampled fall directions. The only command in every trial is a single static standing pose, so the entire rise is synthesized by the middleware. Neither Heracles nor HumanUP receives a get-up reference. We quantify human-likeness against human motion capture with three sets of 100 get-up motions: *MoCap*, human rise segments retargeted to the G1; *Heracles*, recoveries synthesized by the middleware under the static standing command; and *HumanUP*, get-up motions from a policy retrained with its official code and unmodified configuration. NN-DTW is the dynamic-time-warping distance to the nearest MoCap rise segment, NN cosine the DTW-aligned cosine similarity of per-joint-centered pose vectors, and peak

contact force the maximum summed external contact force over the recovery in body weights (BW); the MoCap column serves as a leave-one-out anchor.

Table 11: **Human-likeness of get-up motions** (100 per column; MoCap is a leave-one-out anchor, percentages relative to HumanUP).

Metric	MoCap (LOO)	Heracles	HumanUP
NN-DTW ↓	0.58 ± 0.24	1.12 ± 0.09 (−45%)	2.02 ± 0.03
NN cosine ↑	0.74 ± 0.15	0.60 ± 0.06 (+40%)	0.43 ± 0.02
Peak contact force (BW) ↓	–	1.47 ± 0.41 (−84%)	9.40 ± 6.84

Heracles is markedly closer to the human motion distribution on both similarity metrics (Tab. 11), reaching 81% of the MoCap leave-one-out cosine similarity versus 58% for HumanUP, with 84% lower peak contact force. HumanUP's recoveries carry a high peak contact force, whereas Heracles synthesizes the transition suited to each fallen configuration (rolling, arm support, gradual torso realignment) from the learned motion prior alone (App. G).

I Limitations and Future Work

While the motion prior already supports diverse recovery behaviors, its coverage is naturally bounded by the training corpus. As large-scale video and multimodal motion datasets become increasingly available, scaling beyond MoCap-only priors may further enrich the repertoire of emergent skills and extend the framework to broader deployment scenarios.