GigaBrain-WBC-0.5: A Behavior World Model for Robust Whole-Body Control with Environment Interaction

Ziyang Cheng[1,2] Tianshu Tang[1,2] Jinxin Lan[1,2] Xinze Chen[2] Yuhan Gong[2]
Zhichao Liu[2] Changzhong Wu[2] Yahao Mao[2,3] Zongyan Deng[2,4] Mingxuan Ma[2,4]
Huasen Xi[2,5,6] Yilong Liu[1,2] Yutong Wu[2] Xiaofeng Wang[2] Yang Wang[2]
Yun Ye[2] Guan Huang[2] Xiaojie Jin[4] Zheng Zhu[2,#] Jiwen Lu[1,#]

[1]Tsinghua University, [2]GigaAI, [3]University of Shanghai for Science and Technology
[4]Beijing Jiaotong University, [5]Institute of Automation, Chinese Academy of Sciences
[6]University of Chinese Academy of Sciences

[#]Corresponding authors.

Project page: this link

Abstract

Whole-body motion tracking policies turn a humanoid into a robust control interface: the teleoperator—or an upstream model—only supplies a coarse movement intent, while the low-level policy keeps the robot balanced and physically feasible. Existing trackers deliver this interface only on flat ground: trained in empty scenes, they never learn how contact with terrain and objects reshapes their dynamics, and they attempt to teach the policy to balance under any command by continually enlarging the reference-motion corpus, which stops working once feasible behaviors become environment-dependent. We present GigaBrain-WBC-0.5, the first *Behavior World Model* (BWM) for humanoid whole-body control. Rather than a purely reactive tracker, we train a causal Transformer to jointly predict its next action, next state, and the distribution over its next latent behavior command, so the network that acts also models how the environment shapes what it can do next. An automatic terrain-annotation pipeline recovers full 3D contact geometry from retargeted motion, enabling terrain annotation at the scale of existing motion datasets. The predicted distribution is reused at deployment to detect implausible commands online and retract them onto learned behaviors, so the robot attempts tasks in a "best-effort" manner. The result is a unified policy that takes real-time command, interacts with environment, and stays robust to implausible commands, falls, and disturbances. GigaBrain-WBC-0.5 achieves the highest success rate across all four regimes among three large-scale tracker baselines: 81.3% on terrain interaction ($4.3\times$ the strongest baseline), 83.1% under implausible commands, and 99.3% fall recovery ($16.8\times$ the strongest baseline). Hardware trials show robust interaction under missing supports and disturbances; the Unitree G1 checkpoint transfers to the Maker L01 robot with simple fine-tuning.

1 Introduction

Humanoid robots are valuable because they can use hands and feet together in environments built for people with the potential to cross complex terrains. Realizing that value requires a low-level controller that accepts *whole-body* commands online. Large-scale motion trackers [5, 8, 21, 23, 30, 52] have shown this is achievable: imitating hundreds of thousands of retargeted human clips under a single objective yields one policy that walks, runs, crouches, dances and gets up. Such a tracker is not merely a demonstration of agility—it is an *interface*. Because the policy absorbs the burden of balance and physical feasibility, whatever produces the command—a human in a VR rig [14, 44], a kinematic planner, or a vision-language-action model [3]—only has to express a rough movement

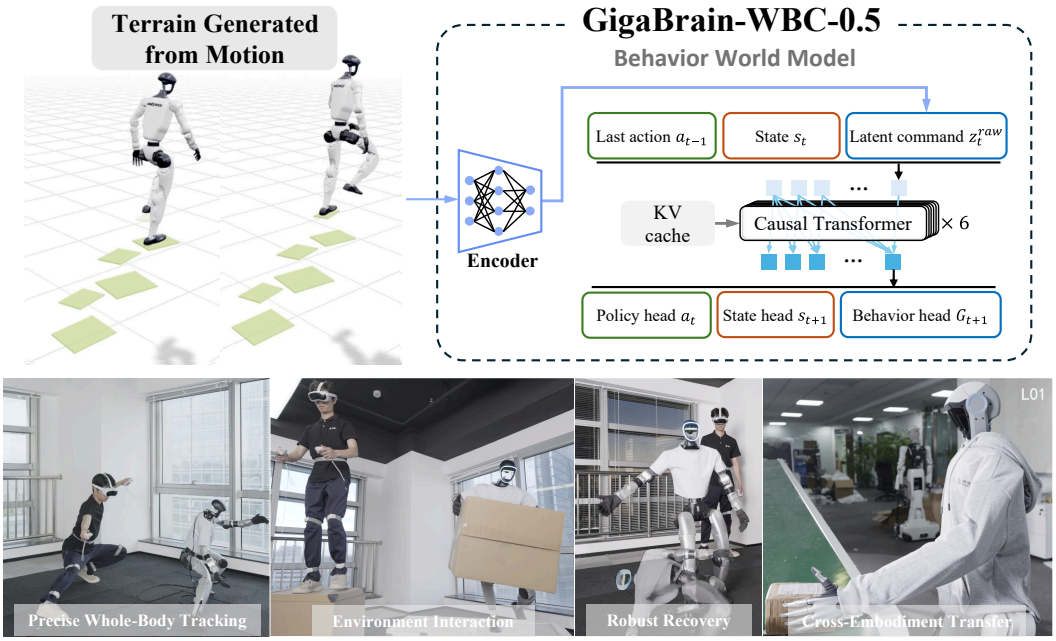


Figure 1. GigaBrain-WBC-0.5: the first Behavior World Model for humanoid whole-body control. GigaBrain-WBC-0.5 couples automatic motion–terrain annotation with a causal Behavior World Model. Recovering full 3D contact geometry from ordinary retargeted motions scales spatially grounded interaction training beyond flat-ground corpora, enabling the policy to exploit environmental contacts. Jointly predicting the action, next proprioceptive state and next-command distribution lets the same network act while modeling what can happen next, and turns its predicted distribution into an online criterion for retracting implausible commands to best-effort behaviors. The resulting live-command controller combines precise tracking, object-and-terrain interaction, fall recovery without a separate controller, and transfer from Unitree G1 to Maker L01 robot through fine-tuning.

intent rather than a high-frequency, dynamically consistent control signal. That is what makes these controllers usable both for collecting whole-body demonstrations [45] and as the execution layer beneath a learned high-level policy [23].

These interfaces, however, live on flat ground. Almost all such controllers are trained and evaluated in an empty scene with a flat floor, so the robot never steps onto anything, sits on anything, or carries anything heavy. The policy is never asked to notice that its own dynamics have changed—that ground reaction force now arrives 18 cm higher, or that a payload has displaced its center of mass—and never learns to exploit such a change. The obstacle is data: an environment-interactive tracker needs terrain and object geometry *paired* with consistent reference motions at scale, and such data is scarce. Chen et al. [7] attack exactly this bottleneck by reconstructing the scene from the motion in hindsight, the same overall direction we take. Their terrain, however, is a 2.5D elevation map, which cannot express the spatial geometry whole-body interaction requires—a chair seat with clearance underneath, a table edge, a handrail—their corpus is 7.5 hours of specialized interaction data, and their policy is driven by per-link contact labels, a command channel a teleoperator does not have.

A second, less obvious difficulty is that the recipe used to make flat-ground trackers robust does not transfer. Which motions are feasible and human-like depends on the dynamics constraints the environment imposes. On flat ground, the field's answer to command robustness has been to enlarge the reference distribution until the policy stays balanced under almost any command [5, 23, 30]. Once the environment varies, that answer is unavailable: a wide lunge that is feasible on flat ground may be infeasible on a narrow step, and a motion of leaning back into a chair is nonsense on flat ground. The feasible set becomes *conditional* on the environment, and RL policies are in any case known to behave unpredictably once a command departs from what they were trained on—a risk terrain sharply amplifies, since no amount of unconditional data coverage covers a conditional set. What is needed is

an online mechanism that recognizes such commands and responds with a best-effort motion rather than an emergency stop or a fallback to standing still, both of which abort the operator's task and, on a staircase, are *more* dangerous than continuing. Cheng et al. [9] show that an infinite-horizon version of this problem for whole-body loco-manipulation reduces to a cheap test applied at every control step; we build a filter around a similar idea, driven directly by our world model's own predictions.

This requirement has a second face. A command can be impossible, but the body can also end up somewhere the command does not assume—on the ground. Standing up is a capability in its own right [17]; what decides whether the operator's task survives is where it lives. Handing control to a specialist get-up controller interrupts the task exactly as the fall did, so we train fallen initializations into the tracker itself and let recovery be one of its behaviors. Staying driveable when the command is infeasible and staying driveable when the body is down are complementary, and it is the two together, in one policy, that we mean by *robust* whole-body control.

We present GigaBrain-WBC-0.5, to our knowledge the first *Behavior World Model* (BWM) for humanoid whole-body control: a controller trained to model its own future behavior rather than only to reproduce an action. The policy is a causal Transformer that consumes latent behavior commands from a reference encoder and emits, at every control step, the action, a prediction of its own next proprioceptive state, and a mixture distribution over the next latent behavior command. Training the controller this way changes what it must represent: predicting its own next state requires internalizing the contact dynamics currently acting on it, and predicting the next command distribution requires modeling which behaviors the environment admits—quantities flat-ground trackers never need and never acquire. These are precisely the two capabilities we are after, environment interaction and extreme robustness, and they come from one model rather than from bolted-on modules. Our contributions are:

- A behavior world model that predicts its own next state and next command distribution alongside the action, yielding a single causal policy that interacts with terrain and objects while remaining driveable by a live operator.
- An automatic spatial terrain-annotation pipeline that recovers the geometry a motion *must* have been performed on directly from the retargeted trajectory. Unlike elevation maps the output is genuine 3D geometry—chairs, tables, boxes, stair treads—which lets us assemble a terrain-paired motion corpus at the scale of existing motion datasets.
- An online filter, driven by the world model's own predicted distribution, that recognizes out-of-distribution commands and projects them onto the closest behavior the policy can execute rather than rejecting them, producing a continuous best-effort motion with no separate classifier or fallback controller. The filter is stateless and closed-form, exposing a single runtime-tunable safety radius that trades off precision and robustness as the task demands.

2 Related Work

2.1 Whole-body motion tracking and behavior foundation models

Whole-body control (WBC) is the general problem of coordinating a humanoid's entire body toward a commanded objective; its dominant instantiation, and the one we adopt, is whole-body motion tracking, in which that objective is a reference motion. Physics-based motion imitation [28, 29] became the standard recipe for it because motion capture provides dense per-frame supervision without per-task reward engineering, and the field has scaled it aggressively. SONIC [23] trains on 100M frames with 128 GPUs and maps heterogeneous inputs into a shared quantized token space; HoloMotion-1 [5] scales a video-derived hybrid corpus with a sparse Mixture-of-Experts Transformer and KV-cache inference; Humanoid-GPT [30] distills a library of RL motion experts into a GPT-style causal Transformer trained on 2B frames. Complementary lines of work push agility on dynamic skills [13, 15], disturbance rejection [52], versatility through diffusion or masked inpainting [21, 33], universality across morphologies and command modes [16, 37, 42], and the fidelity–diversity trade-off [36]; teleoperation systems [1, 14, 44, 45] turn such trackers into data-collection interfaces. Behavior foundation models [43] generalize the same pre-training idea beyond tracking, prompting one policy across control modes through masked distillation over a behavior corpus [46] or a shared

	Diverse whole-body tracking	Whole-body teleoperation	Terrain interaction	Object interaction	Robust to OOD	Robust to falls
GMT [8]	✓	✗	✗	✗	✗	✗
TWIST [44]	✓	✓	✗	✗	✗	✗
SONIC [23]	✓	✓	✗	✗	✗	✗
HoloMotion-1 [5]	✓	✓	✗	✗	✗	✗
Humanoid-GPT [30]	✓	✓	✗	✓	✗	✗
PHP [39]	✗	✗	✓	✗	✗	✗
GALLANT [2]	✗	✗	✓	✗	✗	✗
OmniRetarget [41]	✗	✗	✓	✓	✗	✗
SceneBot [7]	✓	✗	✓	✓	✗	✗
CMP [9]	✓	✓	✗	✓	✓	✗
BFM-Zero [20]	✓	✓	✗	✗	✗	✓
GigaBrain-WBC-0.5 (ours)	✓	✓	✓	✓	✓	✓

Table 1. Capability comparison. The teleoperation column requires a causal, low-latency interface driven only by signals an operator can produce online—not per-link contact prompts, or privileged state. The last column requires that a fall be recoverable by the tracking policy itself: the robot stands back up and resumes following the reference, with no separate recovery controller and no human reset.

latent over motions, goals and rewards [20], though so far at a smaller data scale.

Two limitations recur across this literature. First, these models are trained on flat ground in an empty scene [5, 8, 20, 23, 30, 46], so their behavior distribution is free-space motion—expressive, but never in contact with anything but the floor. Second, generalization remains narrower than the ambition: Wang et al. [36] document a "generality barrier" in which tracking fidelity collapses as the motion library grows, Tessler et al. [33] and Luo et al. [22] report the same tension between coverage and precision in the physics-based character setting, and the one tracker that does reach scene interaction [7] does so from a 7.5-hour specialized corpus. GigaBrain-WBC-0.5 keeps the scalable causal-tracking recipe but changes what the policy is asked to represent, from reproducing a reference to modeling what the current environment permits.

2.2 Terrain- and object-interactive whole-body control

A second line of work targets the environment directly, from classical footstep planning with contact-aware trajectory optimization and MPC [11, 19] to learned perceptive locomotion over stairs, slopes and 3D-constrained terrain [2, 48, 50, 54]. Parkour-level agility has been reached by chaining human skills through motion matching [39] and by reconstructing coupled motion–terrain data from monocular video [49], while scene-aware retargeting and interaction-preserving data generation [38, 41, 47, 53] enable tracking of large-object manipulation, and terrain-aware kinematic generators can be co-trained with a tracker [40, 51].

We do not position GigaBrain-WBC-0.5 against this literature, for two reasons. Terrain-traversal policies of this kind are typically specialists trained for one skill family, and—as surveyed by Chen et al. [7]—are commanded through a low-dimensional root velocity or navigation interface [2, 50, 54]. Such an interface states where the base should go but not what the arms, torso and hands should do, so these policies cannot serve as the general low-level layer of a whole-body manipulation stack. Methods that do track a full whole-body reference over scene geometry [18, 38, 41, 53] instead rely on paired scene assets and are demonstrated per trajectory rather than as general controllers. The closest prior work is SceneBot [7], which unifies free-space motion, terrain and object interaction in a single tracking policy by conditioning on per-link contact labels and reconstructing 2.5D elevation terrain in hindsight. GigaBrain-WBC-0.5 differs in three respects: our annotation recovers full spatial geometry rather than a height field, our command channel is a reference window an operator can produce online, and our policy is trained as a world model rather than conditioned on externally supplied contact prompts.

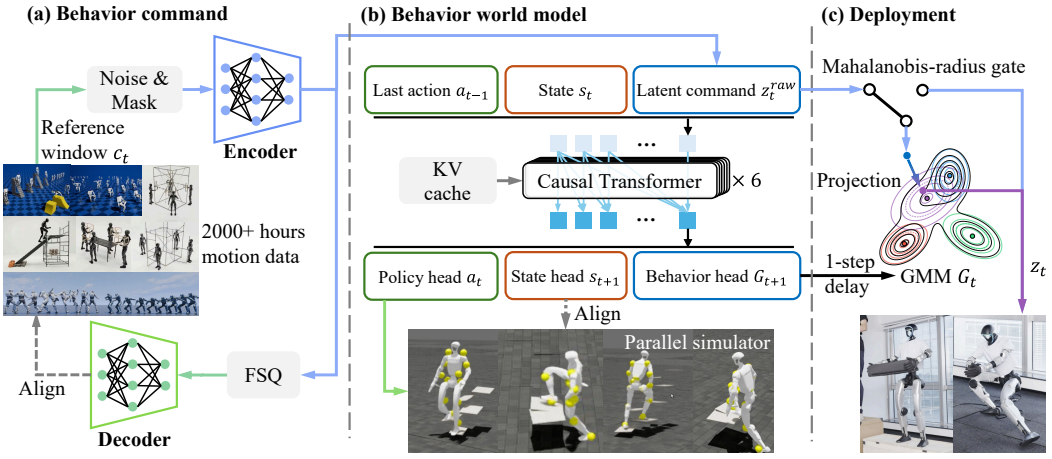


Figure 2. GigaBrain-WBC-0.5 overview. (a) The reference window c_t is perturbed with noise and, for the robot-relative root translation, intermittently masked, then mapped by an MLP encoder to the latent behavior command z_t^{raw}; a finite scalar quantizer and a decoder reconstruct the clean reference for a training-only cycle-consistency loss (dashed). **(b)** $[s_t, a_{t-1}, z_t]$ is processed by a 6-layer causal Transformer with a KV cache; three heads emit the action a_t, the next proprioceptive state s_{t+1} (aligned against parallel-simulator rollouts), and the Gaussian Mixture Model (GMM) G_{t+1} over the next latent command. **(c)** At deployment, a stateless Mahalanobis test checks z_t^{raw} against G_t—the mixture emitted one step earlier—and radially retracts out-of-distribution commands onto the safety ellipsoid of the selected component before committing z_t. The real-robot frames at bottom right illustrate commands produced by this deployment path on hardware.

2.3 Robustness to out-of-distribution commands and to falls

Robustness in whole-body tracking has mostly been pursued by widening the training distribution, through domain randomization, adversarial pushes and hindsight perturbation [6, 15, 52]. Cheng et al. [9] instead filter commands online, reducing the underlying infinite-horizon constraint to a test applied at every control step; they explore both a lightweight filter that flags commands lying outside the training distribution and a more elaborate one that estimates, through an auxiliary learned module, which commands the policy has actually mastered, and report that the latter is more effective but adds system complexity. Latent-space constraints have also been imposed by tying behavior to a learned motion prior [22, 33], which restricts what can be commanded but without an online admissibility test. GigaBrain-WBC-0.5 adopts the same single-step, distribution-based idea without the auxiliary module: the mixture already predicted by our world model gives a distribution over the next latent command for free, so an out-of-distribution command can be identified and projected back toward it with no additional machinery.

An infeasible command is one face of robustness; a body already on the ground is the other, and it is usually handled by a dedicated skill. HoST [17] learns standing up across diverse ground postures with a multi-critic curriculum and deploys it on the same G1 platform we use, while SafeFall [25] addresses the preceding moment, predicting an unavoidable fall and shaping the impact to protect the hardware. Where that capability lives decides whether the operator's task survives the fall: handing control to a specialist get-up controller interrupts the task exactly as the fall did. BFM-Zero [20] shows recovery need not be a separate mode—its unsupervised latent policy recovers from being pulled to the ground and resumes tracking—and we take the same position, training fallen initializations directly into the tracker (Section 3.5) so that standing up is one of its behaviors rather than an exit from it. Remaining driveable when the command is impossible and remaining driveable when the body is not where the command assumes are complementary halves of the same requirement, and it is the two together, in one policy, that we mean by robust whole-body control.

3 Method

3.1 Problem setting

We control a Unitree G1 humanoid with 29 actuated degrees of freedom at 50 Hz. At each step t the policy receives a proprioceptive observation $\boldsymbol{s}_t \in \mathbb{R}^{67}$ (projected gravity, base angular velocity, torso linear acceleration, joint positions and velocities), the previous action $\boldsymbol{a}_{t-1} \in \mathbb{R}^{29}$, and a reference window $\boldsymbol{c}_t$ over the next 10 frames; it outputs PD joint targets $\boldsymbol{a}_t \in \mathbb{R}^{29}$. The reference window is produced online by whichever source drives the robot—VR teleoperation, a kinematic planner, or offline replay—so the policy must remain driveable from whatever that source can supply. Figure 2 summarizes the resulting architecture.

3.2 Behavior world model

The reference window is assembled from the motion the command source provides. For each of 10 future frames at 0.02 s spacing we take (i) the 29 reference joint positions, (ii) the 6-D rotation of the reference root relative to the *current robot* root, (iii) the reference root translation relative to the reference root one frame earlier, (iv) the reference root translation relative to the current robot root, and (v) the world down direction expressed in the reference root frame, giving $\boldsymbol{c}_t \in \mathbb{R}^{440}$; block (v) supplies absolute gravity alignment without global localization. Each block is perturbed with additive uniform noise, and an MLP encoder E produces a continuous latent behavior command $\boldsymbol{z}_t^{\text{raw}} = E(\boldsymbol{c}_t + \boldsymbol{\epsilon}_t) \in \mathbb{R}^{64}$, represented as two 32-dimensional tokens.

How that latent space is shaped governs everything downstream, so we route reconstruction through a quantizer rather than through the policy network. A finite scalar quantizer [26] and a small kinematic MLP decoder are attached to the encoder during training only: the decoder maps the *quantized* latent back to the clean, unmasked 440-d window, and that reconstruction is re-encoded and aligned with the original continuous latent. The policy itself always receives the unquantized latent. This routing matters in both directions. SONIC [23] quantizes on the path to its action decoder as well, which we found leaves fine motion poorly resolved; conversely, dropping the quantizer and reconstructing directly from a Transformer output head—the other natural choice—removes all pressure on the latent, and we found it shapes the space markedly less well. Confining quantization to the auxiliary branch makes it a pure regularizer: with 32 levels per dimension, a latent that folds too much information into too few dimensions becomes indistinguishable after rounding and fails to reconstruct, so the encoder must spread information across dimensions and keep the space loosely packed. Control resolution stays continuous, while the latent acquires the smooth, well-spread structure that the mixture of Section 3.4 needs in order to describe it.

Among those input blocks, (iv) needs care. Terrain interaction is intolerant of global drift: if the robot's root wanders from where the reference expects it to be, the geometry under its feet stops matching the motion, and errors accumulate over long-horizon tasks such as climbing a staircase. The usual remedy is to cut long motions into short clips and reset the robot back onto the reference frequently, which bounds drift by construction. That remedy is closed to us: six stacked layers of 32-frame local attention give an effective receptive field of 187 frames, and episodes short enough to bound drift by resetting would never fill it, so the policy would never learn to exploit the history it can see. We keep episodes long and control drift through the input instead. Block (iv) is exactly the signal that corrects drift, but on hardware the robot's true root position is generally unavailable, so we cannot simply supply it at all times. During training it is visible about a third of the time and zeroed otherwise, resampled per environment every 2–5 s; the policy therefore learns to run without it, and to re-anchor itself to the reference whenever it briefly reappears. At deployment the block is held at zero throughout.

The masked latent then enters the core network. The per-frame input $\boldsymbol{e}_t = [\boldsymbol{s}_t, \boldsymbol{a}_{t-1}, \boldsymbol{z}_t] \in \mathbb{R}^{160}$ is projected to 256 dimensions and processed by a 6-layer causal Transformer [35] with 4 heads and rotary position embeddings [32], restricted to a 32-frame window by a mask that simultaneously enforces causality, the window bound and episode boundaries. Rollout, evaluation and deployment all run single-step with a per-layer KV cache, so the marginal cost of the long context is one frame of

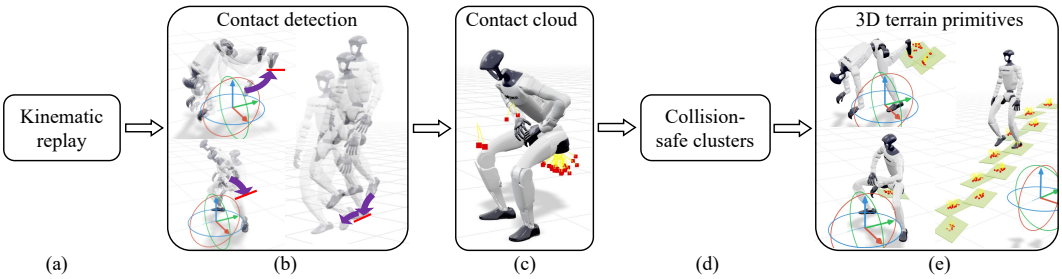


Figure 3. From motion to spatial terrain. **(a)** A retargeted motion is replayed kinematically. **(b)** Contact detection flags frames where a sampled surface point's normal and tangential velocity fall below threshold (arrows show incoming velocity at the moment of contact). **(c)** Detected contacts accumulate into a 3D point cloud. **(d)** Points are grouped into collision-safe clusters after the whole-body penetration filter. **(e)** Each cluster is fitted with proper primitives, giving the 3D terrain used for training—genuine spatial geometry rather than a height field, so chairs, tables, boxes and stair treads are all representable.

computation. Three linear heads read the final hidden state:

$$\boldsymbol{a}_t \in \mathbb{R}^{29}, \qquad \hat{\boldsymbol{s}}_{t+1} \in \mathbb{R}^{67}, \qquad \mathcal{G}_t = \{\pi_k, \boldsymbol{\mu}_k, \log \boldsymbol{\sigma}_k\}_{k=1}^{K} \in \mathbb{R}^{516}, \tag{1}$$

with $K = 4$ diagonal Gaussian components over the next latent command $\boldsymbol{z}_{t+1} \in \mathbb{R}^{64}$. The first head is the policy; the other two make the network a world model. Predicting $\boldsymbol{s}_{t+1}$ forces it to represent the contact dynamics currently acting on the body—exactly what changes when the robot steps onto a box or picks up a load. Predicting $p(\boldsymbol{z}_{t+1} \mid \text{history}_t)$ forces it to represent which behaviors the situation admits, since the reachable set of next commands on a staircase is not the reachable set on flat ground. The mixture is deliberately multi-modal: distinct behavior modes—walking across speeds and headings, take-offs across styles—occupy separate regions of the latent space, and collapsing them into a single Gaussian would smear those regions together, so the single admissible region Section 3.4 projects onto would span behaviors the current situation does not admit.

The full objective combines the PPO loss with the four auxiliary terms:

$$\mathcal{L} = \mathcal{L}_{\text{PPO}} + 0.01\,\mathcal{L}_{\text{recon}} + 1.0\,\mathcal{L}_{\text{cycle}} + 0.01\,\|\hat{\boldsymbol{s}}_{t+1} - \boldsymbol{s}_{t+1}\|^2 - 0.005\,\log \mathcal{G}_t(\boldsymbol{z}_{t+1}), \tag{2}$$

with both next-step terms masked across episode boundaries so no spurious transition is supervised.

3.3 Automatic spatial terrain annotation

That masking presupposes the policy trains on terrain at all, and ordinary motion capture carries no scene. We resolve this in hindsight: a motion captured on a staircase contains, in its own kinematics, evidence of where the staircase was. Our pipeline (Figure 3) turns that evidence into simulator-ready geometry automatically, which is what allows a terrain-paired corpus to be built at the scale of existing motion datasets rather than at the scale of scene-capture sessions.

We first uniformly sample points on the collision geometry of the contact-relevant links—by default the two ankle links, optionally extended to hands, knees, elbows, pelvis and the back of the torso—and replay the retargeted trajectory kinematically in MuJoCo [34] to obtain each sample's world position and outward normal per frame. After Savitzky–Golay smoothing, a per-sample state machine marks a contact when normal speed falls below 0.2 m/s and tangential speed below 0.3 m/s, provided at least 2 of the preceding 10 frames show normal deceleration above 1 m/s^2; this deceleration signature is what distinguishes a support event from a limb that merely happens to be momentarily slow. A contact ends when the sample drifts more than 5 cm or begins to slide, and is discarded below 10 frames.

Detected contacts are then turned into geometry. Points below 10 cm or with downward-facing normals are discarded as ordinary floor or non-supporting surfaces; the rest is deduplicated on a 1 mm voxel grid. Every surviving point is then tested against the *whole* robot over the *whole* trajectory: it is kept only if a support placed there, offset outward along its normal by at least 5 cm, never intersects

any body at any frame. This penetration filter is what prevents the reconstruction from inventing geometry the robot would have walked through.

Points are then grouped into planar supports by a pairwise affinity in which two points connect only if their normals agree within 40° and their mutual offset along the normal is below $10\,\text{cm}$, so DBSCAN separates stair treads at different heights even when laterally adjacent. Each cluster is fitted with a $1\,\text{cm}$-thick oriented box whose in-plane axes come from a PCA of the tangential projection and whose normal offset is aligned to the lowest contact point, so shallow contacts do not become deep ones; clusters whose bounding box is much emptier than the area their points occupy are recursively split along the principal axis, decomposing L-shaped and annular supports into several tight rectangles. Each primitive is finally expanded laterally by a penetration-safe binary search, giving the robot margin around the footprint it originally used.

Primitives are spawned per environment as static rigid bodies in Isaac Lab [27], co-located with that environment's reference motion. Because primitive count and size differ per motion, terrain is rebuilt only at motion-resample boundaries and motion sampling is frozen within an epoch, so reference and geometry can never disagree. Terrain categories are read from the labelled corpus, and each is sampled at a configurable rate so terrain and flat-ground motions mix at a controlled ratio of 0.2.

3.4 Filtering out-of-distribution commands

RL policies are known to behave unpredictably once an input departs from the distribution they were trained on, and Section 1 argues why a fixed, environment-independent training distribution cannot be enlarged away on terrain. At deployment the mixture $\mathcal{G}_{t-1}$ emitted at the previous step is a prediction, made before seeing $\boldsymbol{c}_t$, of which behavior commands are consistent with training-time experience in the robot's current situation. We treat it as a live, one-step approximation of the training distribution and check incoming commands against it, following the single-step reduction of Cheng et al. [9]. The ordering is strict, $\boldsymbol{c}_t \to E \to \boldsymbol{z}_t^{\text{raw}} \xrightarrow{\mathcal{G}_{t-1}} \boldsymbol{z}_t^{\star} \to \pi \to (\boldsymbol{a}_t, \hat{\boldsymbol{s}}_{t+1}, \mathcal{G}_t)$: filtering the current latent with the mixture produced at the current step would leak the command into its own admissibility test. The first control step has no predecessor mixture and is passed through unfiltered, and so is the first step after any event that discards $\mathcal{G}_{t-1}$ rather than carrying it forward—a commanded behavior switch, a temporal reset, or a hand-off to or from direct PD control.

The natural first choice—accept $\boldsymbol{z}$ iff its log-likelihood under $\mathcal{G}_{t-1}$ exceeds a fixed threshold—is unusable, because a conditional mixture can be flat enough that *no* $\boldsymbol{z}$ attains the threshold and the admissible set is empty. We instead define, per component, an in-distribution region via the squared Mahalanobis radius $M_k^2(\boldsymbol{z}) = \sum_{i=1}^{D}((z_i - \mu_{k,i})/\sigma_{k,i})^2$, $D = 64$, with $\sigma_{k,i} = \exp(\log \sigma_{k,i})$ and $\log \sigma_{k,i}$ clamped to $[-3, 10]$ before use so a degenerate head cannot collapse or explode the region it defines, and the set $\mathcal{Z}_k = \{\boldsymbol{z} : M_k^2(\boldsymbol{z}) \le R_{\text{safe}}^2\}$, which always contains $\boldsymbol{\mu}_k$ and is therefore never empty.

Since each component covers one behavior mode, we first ask which mode the incoming command most plausibly belongs to and then check and, if needed, project against that component alone. The responsibility head is

$$k^\star = \arg\max_k \Big[\log(\pi_k + 10^{-10}) \; - \; \frac{1}{2D} \sum_{i=1}^{D} \big[((z_i - \mu_{k,i})/\sigma_{k,i})^2 + 2\log\sigma_{k,i} + \log 2\pi\big] \Big], \quad (3)$$

i.e. a MAP-*style* rule in which the Gaussian log-density is averaged over the $D = 64$ latent dimensions while the mixing log-weight is left at its native scale, so the prior over behavior modes weighs more heavily in the choice than dimension-consistent scaling would give it. This deliberately biases selection toward modes the world model considers likely in the current situation, rather than toward whichever mode happens to have the loosest covariance. Letting an arbitrary component serve instead of $k^\star$ allows an unrelated mode to sit close to an out-of-distribution command and silently bypass correction, which we found to be a major failure mode.

Once the head is fixed, the test is a single per-frame inequality: the command is out-of-distribution iff $M_{k^\star}^2(\boldsymbol{z}_t^{\text{raw}}) > R_{\text{safe}}^2$. When it fires, we do not replace the command by the component mean—a

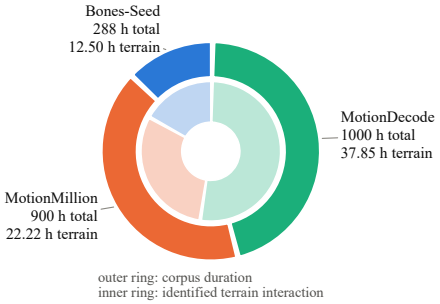
mean is the center of a conditional distribution, not a motion to execute—but retract it radially toward $\boldsymbol{\mu}_{k^\star}$, just far enough to land on the boundary of $\mathcal{Z}_{k^\star}$:

$$\boldsymbol{z}_t^\star = \boldsymbol{\mu}_{k^\star} + \sqrt{R_{\text{safe}}^2 \,/\, M_{k^\star}^2(\boldsymbol{z}_t^{\text{raw}})}\,\big(\boldsymbol{z}_t^{\text{raw}} - \boldsymbol{\mu}_{k^\star}\big), \tag{4}$$

which by construction satisfies $M_{k^\star}^2(\boldsymbol{z}_t^\star) = R_{\text{safe}}^2$ exactly. This is not the Euclidean-nearest point on the ellipsoid surface, which would require solving for a Lagrange multiplier every frame; it is a closed-form rescaling along the ray from the mode's center through the raw command. The retracted command still points from the trusted mode toward what the operator asked for, so the robot keeps attempting the task instead of freezing.

The filter is memoryless: each frame re-selects $k^\star$ against the freshly emitted $\mathcal{G}_{t-1}$ and re-evaluates the inequality on its own, so the admissible region tracks the conditional mixture as the situation evolves and the whole procedure reduces to two closed-form expressions. It is $O(1)$ per step and costs well under 1 ms against a 20 ms control period, negligible in the deployment budget. It also leaves a single tunable knob, the safety radius R_{safe}, which an operator can move at runtime, without retraining, to trade fidelity for robustness along the frontier of Figure 6.

3.5 Training



We draw on three large retargeted human-motion corpora: Bones-Seed [4] (288 h), MotionMillion [12] (900 h) and MotionDecode [10] (1000 h). Within them we identify 12.50 h, 22.22 h and 37.85 h respectively as containing spatial terrain interaction, each clip labelled with the terrain type it involves. This identified subset is cleaned and filtered—removing retargeting artifacts and the annotation failures caught by the audit of Section 4.3—before being mixed with flat-ground motion at a controlled rate for training, so that free-space tracking quality is preserved while contact-rich behavior is learned.

Figure 4. Source corpora and identified terrain coverage. The outer ring is the total duration of each corpus, the inner ring the duration identified as terrain interaction; the two are independent pies, so the inner ring reads as the terrain composition rather than a sliver of the total. Corpus identity is carried by color across the rings.

We optimize with PPO [31] in Isaac Lab [27] over the corpus mix of Figure 4, under a sequence-level update: each update recomputes the current rollout segment with gradients while attaching the detached per-layer KV prefix saved at the rollout start, so the attention context seen during the update is exactly the context the policy acted under. Rollouts use 4096 parallel environments on flat ground, reduced to 512 once terrain geometry and fallen-state initialization are enabled, since spawning heterogeneous static primitives per environment raises the broad-phase collision cost. Two ingredients target robustness. A fraction of flat-ground episodes start from synthetic fallen poses under a curriculum over trunk inclination up to fully prone, with a validated whole-body collision proxy placing the robot at ground contact, and precise tracking rewards are gated by a smooth target-relative recovery gate so the policy is not penalized while it is still standing up. Persistent random external forces at the wrists and torso emulate payloads and contact loads.

Transferring to a second embodiment reuses this recipe unchanged. Retargeting the corpus to a Maker L01 humanoid and fine-tuning the G1 checkpoint on a single 8-GPU node recovers whole-body tracking on the new body quickly, while training the same architecture on L01 from scratch converges slowly: the behavior world model appears to carry structure about how a humanoid interacts with its environment that is not specific to one set of link lengths. Part of the gallery in Section 4.5 is recorded on L01.

Every tracking term maps an error $\boldsymbol{e}$ to

$$r = \exp\big(-\|\boldsymbol{e}\|^2/\sigma^2\big), \tag{5}$$

Reward term	Weight	σ
Anchor position	1.0	0.30
Anchor orientation	0.5	0.40
Relative body position	1.0	0.30
Relative body orientation	1.0	0.40
Body linear velocity	1.0	1.00
Body angular velocity	1.0	3.14
Local head/wrist points	2.0	0.10
Local wrist orientation	2.0	0.30
Local feet position	1.0	0.15
Action rate	$-1e{-1}$	–
Joint limit	-10.0	–
Undesired contacts	$-1e{-1}$	–
Anti-shake angular vel.	$-5e{-3}$	–
Ankle joint acceleration	$-2.5e{-6}$	–

Randomized quantity	Range	Resample
Static friction	$[0.3, 1.6]$	startup
Dynamic friction	$[0.3, 1.2]$	startup
Restitution	$[0.0, 0.5]$	startup
Wrist/torso mass	$\times[0.8, 2.5]$	startup
Torso CoM x	± 0.025 m	startup
Torso CoM y, z	± 0.05 m	startup
Joint default offset	± 0.01 rad	startup
Push linear x, y	± 0.5 m/s	4–6 s
Push linear z	± 0.2 m/s	4–6 s
Push angular r, p	± 0.52 rad/s	4–6 s
Push angular yaw	± 0.78 rad/s	4–6 s
Wrist/torso force	± 5 N	2–5 s
Wrist/torso torque	± 0.3 N m	2–5 s
Root-translation mask	masked w.p. $2/3$	2–5 s
Reference-window noise	± 0.05	each step

Table 2. Reward terms (left) and domain randomization (right). Tracking terms use the Gaussian kernel of Equation (5); the remaining terms are plain penalties. Startup quantities are drawn once per environment at scene construction, the rest at the stated interval.

	Standard			**Terrain**			**OOD**		**Fall**	
Method	MPKPE↓	RootVel↓	SR↑	MPKPE↓	RootPos↓	SR↑	MPKPE↓	SR↑	SR† ↑	Jerk↓
SONIC [23]	82.3	189.6	94.1	331.2	294.7	15.3	327.6	50.0	5.9	1295.5
HoloMotion-1 [5]	109.4	**121.3**	89.0	330.0	280.7	18.7	248.7	67.7	0.7	2000.0
Humanoid-GPT [30]	90.9	205.8	91.9	283.3	326.7	14.0	208.0	70.6	2.9	3598.1
GigaBrain-WBC-0.5 (ours)	**76.6**	211.1	**96.3**	**93.3**	**100.7**	**81.3**	**158.0**	**83.1**	**99.3**	**1050.6**

Table 3. Sim-to-sim comparison in MuJoCo. Standard: AMASS test split. **Terrain**: terrain-interaction motions with annotated geometry loaded. **OOD**: physically implausible motions from MotionMillion. **Fall**: Standard split initialized from fallen poses. MPKPE (mean per-keypoint position error) and RootPos in mm, RootVel in mm/s, Jerk in rad/s^3, SR in %. Each split is scored on the metrics that decide it (Section 4.1): OOD asks first whether the robot stays up, Fall whether it gets up and how violently. †Fall SR uses the recovery criterion—falling is permitted and an episode succeeds if the robot returns to tracking—whereas every other SR treats a fall as failure. GigaBrain-WBC-0.5 runs at its deployment setting $R_{\text{safe}} = 3$ throughout.

so weights and kernel widths σ fully specify it; Table 2 lists both, together with the randomization ranges. Anchor terms act on the root, relative-body terms on all tracked links in the root frame, and the local terms on the head proxy, wrists and feet expressed in the root frame—the same points a VR operator provides. The critic, discarded at deployment, is a separate MLP over 1756-d privileged observations: multi-frame reference targets, root position and rotation differences, body positions and orientations, and 10 frames of proprioception and action history.

4 Experiments

We evaluate three claims. First, that training the tracker as a behavior world model buys competence on terrain and survival under commands it was never trained for while also yielding a highly capable flat-ground policy (Section 4.2). Second, that the automatic terrain annotation is accurate enough to train on, and that the policy it produces interacts with real environments rather than merely tolerating them (Section 4.3). Third, that the out-of-distribution filter converts commands the policy was not trained for into survivable behavior at a precision cost that is both acceptable and—since the safety radius is tunable at runtime—chosen rather than fixed (Section 4.4). This release reports the core comparisons; a more extensive evaluation, including real-robot latency measurements and component ablations, will be added in the future.

4.1 Experimental setup

Benchmarks and metrics. We evaluate on four held-out sets, each targeting a distinct failure mode. **Standard** is the AMASS [24] test split and measures ordinary flat-ground tracking; it contains 136 clips. **OOD** is a test subset of MotionMillion [12] containing physically implausible references—

self-penetration, retargeting artifacts, motions no humanoid of this morphology can execute—and measures what a controller does when the command falls outside what it was trained on; it contains 136 clips. **Terrain** is a test subset of Bones-Seed [4] whose motions traverse or interact with terrain, evaluated with the annotated geometry loaded; it contains 150 clips. **Fall** reuses the 136 Standard clips but initializes the robot from fallen poses.

SR uses a deviation-from-reference criterion under which a fall always counts as a failure. The rollout is not cut short when that happens: every clip is played to its end and the error metrics are accumulated over the full reference, so a motion the policy cannot handle cannot also shrink its own error by terminating early. **Fall** is the one exception to the failure criterion: there the robot starts on the ground, and counting any time spent down as a failure would score every episode as failed before it begins, so being down is permitted and an episode counts as a success once the robot has stood up and returned to tracking the reference. Fall SR is therefore a recovery rate and is not comparable to the other three columns. **MPKPE** is the mean per-keypoint position error in mm over the tracked links, **RootPos** and **RootVel** the root position and linear-velocity tracking errors in mm and mm/s, and **Jerk** the mean magnitude of joint jerk in rad/s^3.

Not every metric decides every split, and reporting all of them everywhere would obscure which number matters. The root is scored by *velocity* on **Standard** but by *position* on **Terrain**, for the same reason the root-translation block of Section 3.2 is masked: on flat ground the root position carries no scene meaning and only the velocity profile matters, whereas on terrain an absolute drift of even a few centimeters puts the feet against geometry the reference did not anticipate. On **OOD** the reference is by construction infeasible, so tracking it accurately is not the goal and a root-velocity error against an impossible trajectory is not interpretable; what matters is whether the robot survives a command it should never have been given, and secondarily how far the executed motion still drifts from the intent, so we report SR and MPKPE only. On **Fall** the question is whether the policy gets back up at all and how violently it does so—hardware does not tolerate the torque spikes that an aggressive stand-up produces—so we report the recovery rate together with joint jerk, the quantity that decides whether a recovery is deployable rather than merely successful in simulation.

Baselines and protocol. We compare against SONIC [23], HoloMotion-1 [5] and Humanoid-GPT [30], the three strongest publicly described large-scale whole-body trackers. All policies, including ours, are evaluated sim-to-sim in MuJoCo [34] under the same failure criterion and the same full-length rollouts, so no method benefits from the simulator it was trained in. No policy is tested with a signal it could not obtain on hardware: the reference-root translation block of Section 3.2—the robot's true root position, which the real G1 does not measure—is masked throughout evaluation and appears only during training, with probability $1/3$. Since these baselines use different corpora and retargeting pipelines, the comparison is evidence about generalization and about the effect of environment-aware training, not a data-matched ablation. GigaBrain-WBC-0.5 is evaluated exactly as it is deployed, with the filter of Section 3.4 active at $R_{\text{safe}} = 3$: the filter is part of the controller, not a wrapper around it, so we report the system as shipped rather than a filter-free variant we would never run.

4.2 Motion tracking across regimes

Table 3 reports all four policies on the four splits of Section 4.1. On **Standard**, training as a behavior world model costs nothing in free space: at 76.6 mm MPKPE and 96.3% SR our policy is the most accurate and the most reliable of the four on exactly the flat-ground regime the baselines are specialized for, ahead of SONIC (82.3 mm, 94.1%). The one metric we do not lead is root linear velocity (211.1 vs. HoloMotion-1's 121.3 mm/s), which we attribute to the intermittent root-translation masking of Section 3.2: a policy that must run without absolute root feedback two-thirds of the time re-anchors itself in small corrections rather than matching the reference's velocity profile continuously. That is the price of being driveable from a command source that cannot supply global position, and it does not show up in keypoint accuracy.

The **Terrain** split is where environment-aware training separates from it. All three baselines collapse to 14–19% SR with MPKPE between 283 and 331 mm—having never experienced non-floor contact,

Terrain category	# sampled	Correct (%)	Dominant failure mode
Stairs / steps	50	92	Fast footfall obscures the deceleration signature
Chair / seat	50	94	Torso sway misplaces the backrest contact
Box / platform	50	98	Sparse contacts fail to form a primitive
Other	50	84	Hand-retargeting noise hides the velocity drop
Overall	200	92	—

Table 4. Human audit of automatic terrain annotation. A motion counts as correct only if every support it relies on is reconstructed at the correct position and orientation and no spurious primitive is introduced.

they treat a step or a seat as a disturbance to reject rather than a surface to load, and the reference diverges from the executed motion within a few frames. Our policy holds 81.3% SR at 93.3 mm, a $4.3\times$ improvement in survival over the strongest baseline and a $3.0\times$ reduction in tracking error. The more telling comparison is internal: moving from flat ground to terrain raises our MPKPE by 22% ($76.6 \rightarrow 93.3$ mm), while it raises the baselines' by a factor of three to four. Terrain is a regime for our policy and an out-of-distribution event for theirs.

The last two splits test what happens when the command or the initial state is one no policy was trained for. On **OOD**, where every reference is physically impossible to execute, GigaBrain-WBC-0.5 survives 83.1% of clips against 50.0–70.6% for the baselines, and its executed motion stays closest to the intent (158.0 mm vs. 208.0 mm for the best baseline)—the filter is not discarding the command, it is retracting it to the nearest thing the policy can do. On **Fall** the gap is categorical rather than incremental: the baselines recover from 0.7 to 5.9% of fallen initializations while ours recovers from 99.3%. Their jerk figures (1295–3598 rad/s^3) measure thrashing on the ground rather than a costly recovery; ours is both the most successful and the smoothest at 1050.6 rad/s^3, which is what makes the behavior something we are willing to run on hardware.

4.3 Environment interaction

The pipeline of Section 3.3 is unsupervised, so its output is audited before it is used as training data. That audit has to be human: the failures an automatic check can catch—above all a primitive intersecting the body—are exactly the ones the penetration filter already rules out, so it cannot serve as its own test. What is left is judgement: reading the motion for the intent behind it, deciding whether the contacts that intent implies are physically sound, and asking whether motion and terrain together look like something a person would actually do.

Retargeting quality varies across our source corpora, and Bones-Seed [4] is the cleanest of the three—only a minority of its hand motions retarget poorly—so an audit there isolates errors introduced by the annotation pipeline itself rather than by upstream retargeting noise. We ground the audit there: for each of the four categories in Table 4 we sampled 50 annotated motions at random and had a human judge whether the reconstructed primitives are consistent with the motion—whether every support the robot uses exists at the right position and orientation without intersecting the body, so that the motion looks physically possible under the annotated scene.

All four categories annotate reliably, at 92% overall. Boxes and platforms are the easiest at 98%: stepping onto a raised surface produces a long, sharply decelerating contact, and the rare failure is a touch too brief to accumulate the points a primitive needs. Stair treads (92%) and seats (94%) fail because the feet or torso move fast enough to blur the deceleration signature, so the tread or backrest is sometimes missing or displaced. *Other* is weakest at 84%, mainly for a reason external to the pipeline: it is dominated by hand-supported motions, where retargeting error on the wrists perturbs precisely the velocity signal contact detection reads. All failure kinds stay conservative, because the penetration filter can only delete candidate geometry and never add it: an error leaves a support absent or misplaced, never an obstacle the robot would collide with. The audited failures are in any case removed from the corpus before training (Section 3.5).

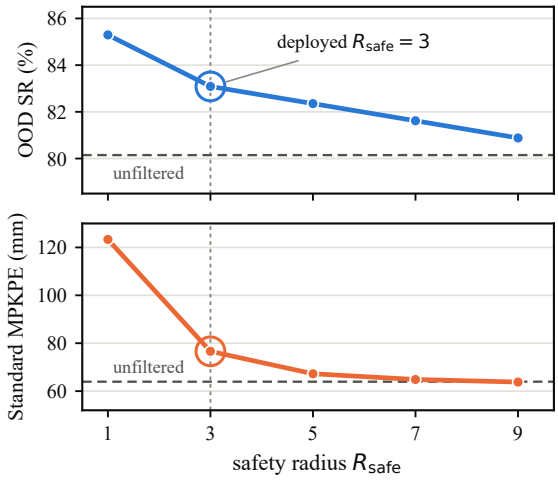
Figure 5. Environment interaction on hardware. Matched hardware comparisons in (a)–(d), with SONIC [23] on the left and GigaBrain-WBC-0.5 on the right, driven simultaneously by the same operator command. **(a)** Ours loads the box as a seat while SONIC remains in a half-squat. **(b)** Ours steps onto the platform and stabilizes; SONIC fails to climb it. **(c)** Ours lifts and holds the box while SONIC loses a stable grasp. **(d)** Ours completes a kneel-to-stand transition to lift a 2 kg fire extinguisher while SONIC falls during the rise. **(e)** Ours carries a large case while walking forward onto the platform and down to the floor.

The Terrain column of Table 3 establishes that training on this annotated geometry pays off in simulation; Figure 5 shows what the same policy does on hardware, where the scene is real and nothing about it is pre-supplied to the controller. To assess the advantage over a strong whole-body tracker under controlled inputs, panels (a)–(d) compare GigaBrain-WBC-0.5 with SONIC using the same live commands from one operator. The contrast is diagnostic: when contact shifts from flat ground to a seat or step, SONIC behaves like a tracker trained for free-space motion and fails to establish load-bearing support, whereas GigaBrain-WBC-0.5 uses the resulting contact to complete the intended motion. The box-lifting and fire-extinguisher trials further show that this advantage extends to maintaining sustained contact forces while balancing under changing upper-body loads. Most strikingly, Figure 5(e) couples both sources of difficulty: while carrying a long black case, the robot walks forward onto the platform and continues down to the floor without dropping the object. These hardware results provide qualitative evidence that terrain-paired training produces a general interaction capability rather than a collection of scripted, scene-specific skills.

4.4 Robustness

The OOD column of Table 3 reports the deployment setting $R_{\text{safe}} = 3$; Figure 6 shows the frontier that setting was chosen from. Because the filter is stateless and R_{safe} is read at runtime rather than fixed during training, this curve is not a design-time choice made once: an operator can move along it on the running robot with a single scalar, filtering harder when the environment is unforgiving and backing off when precision matters more.

Figure 7 shows the qualitative counterpart. The hardware trials expose two complementary requirements for keeping a tracker driveable: adapting before a failure and recovering after one. Before a failure, when an expected environmental support is absent, the commanded motion cannot be realized as specified; in Figure 7(a, b), the policy instead produces a stable best-effort response without committing to nonexistent contact. After a failure, panel (c) shows the same tracker standing



A filter that never intervenes is useless, and one that always intervenes destroys precision. Sweeping the single knob R_{safe} of Section 3.4 traces the achievable frontier between survival under commands the policy was not trained for and fidelity under commands it was. The deployed radius $R_{\text{safe}} = 3$ sits at the knee: tightening it from the unfiltered limit (dashed) buys 2.9 points of OOD survival for 12.7 mm of Standard MPKPE, and the remainder of the sweep down to $R_{\text{safe}} = 1$ trades a further 2.2 points against 46.7 mm.

Figure 6. Robustness versus precision as a function of the safety radius. Standard-split MPKPE (in-distribution fidelity) and OOD success rate against the single filter parameter R_{safe}. Small R_{safe} shrinks the admissible ellipsoid and projects more aggressively, buying OOD survival at the cost of in-distribution tracking; large R_{safe} approaches the unfiltered policy (dashed). The circled radius $R_{\text{safe}} = 3$ is the operating point reported in Table 3.

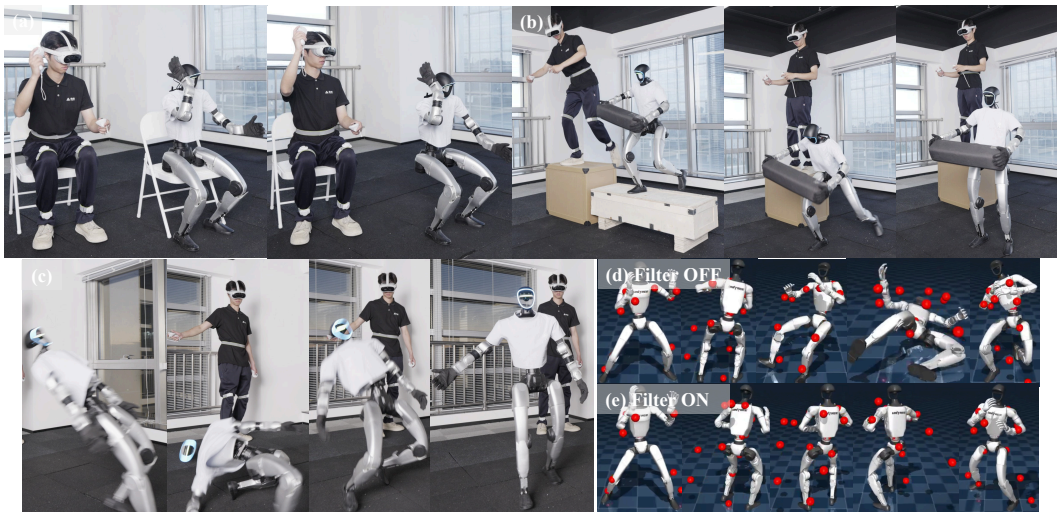


Figure 7. Robustness to unsafe commands and to physical disturbances. (a, b) Removing the chair or platform invalidates the original command, yet the policy settles into a stable best-effort behavior instead of a emergency stop, in which case the carried object would be dropped. **(c)** The same tracking policy recovers from a kick-induced fall without a separate recovery controller. **(d, e)** Under the same self-interfering and challenging spinning-kick reference in simulation, disabling the safety filter causes the robot to attempt the kick and fall (though subsequently recover); enabling it preserves the commanded turn, suppresses the difficult kick and keeps the robot upright.

up and resuming control without handing the task to a dedicated recovery mode. The simulation comparison examines another before-failure mechanism, this time in command space: for the same highly dynamic, self-interfering spinning-kick reference, the unfiltered command causes a fall that the policy later recovers from, whereas the filtered command avoids the fall while preserving the commanded turn in a "best-effort" manner. This anticipatory correction comes directly from the world model: the preceding-step mixture defines a conditional Mahalanobis ellipsoid and retracts an out-of-distribution latent to its boundary, thereby filtering unsafe commands that depart from learned conditional behavior.

Figure 8. Whole-body motion capability. Agile motion, household tasks and industrial tasks, all produced by one tracking policy under live whole-body command. Frames are badged with the robot they were recorded on: the controller runs on a Unitree G1 and, after the fine-tuning of Section 3.5, on a Maker L01.

4.5 Whole-body motion capability

Figure 8 collects what the controller does beyond the specific claims tested above. With the same live whole-body command interface, the policy tracks demanding motions including squat jumps, high kicks and low Tai Chi stances. The same tracker also drives household and industrial behaviors under live command, coordinating locomotion with object interaction without switching to a task-specific manipulation stack. After fine-tuning the G1 checkpoint, several household and logistics trials run on Maker L01, showing that the controller is not tied to a single embodiment.

5 Conclusion

We presented GigaBrain-WBC-0.5, a behavior world model for humanoid whole-body control. Its premise is that the two capabilities missing from current whole-body trackers—interacting with the environment, and remaining well-behaved when asked to do something the environment does not permit—are the same capability seen from two sides, and that both follow from making the controller predict its own future rather than only its next action. Around that premise we built an automatic pipeline that recovers spatial contact geometry from ordinary retargeted motion, and a stateless, closed-form single-step projection that turns the model's own predicted command distribution into an

online out-of-distribution filter governed by one scalar radius.

Across sim-to-sim evaluation, the resulting policy achieves the highest success rate in all four regimes among three large-scale tracker baselines. It reaches 81.3% on terrain interaction ($4.3\times$ the strongest baseline) and 83.1% under physically implausible out-of-distribution commands, while recovery from fallen initializations rises from single-digit baseline rates to 99.3%. On hardware, the policy itself tracks agile motions, interacts with real environments and remains robust to missing supports and physical disturbances. When a command is infeasible under current environmental conditions, it settles into a stable best-effort response; in simulation, its future-prediction filter similarly retracts commands that depart from learned behavior before execution. If a fall nevertheless occurs, the same policy stands up and resumes control without a separate recovery controller. Fine-tuning further transfers the G1 checkpoint to Maker L01 robot for household and logistics tasks. By introducing, to our knowledge, the first behavior world model for humanoid whole-body control, this work opens a promising direction toward controllers that combine environment interaction with robust execution.

Several limitations remain. The filter is validated in simulation and its radius must be re-calibrated per checkpoint and platform before being relied upon on hardware; it flags commands that are unlike the policy's training experience rather than reasoning about physical risk directly, so interception of infeasible commands is high-probability rather than certain. The terrain annotation is derived from contact evidence and therefore recovers only the geometry a motion actually touches: surfaces that were present but unused are not reconstructed, and the resulting scenes are supports rather than complete environments. Extending the reconstruction to non-supporting geometry, and connecting the world model's state prediction to a more direct notion of physical risk, are natural next steps.

References

[1] Qingwei Ben, Feiyu Jia, Jia Zeng, Junting Dong, Dahua Lin, and Jiangmiao Pang. HOMIE: Humanoid Loco-Manipulation with Isomorphic Exoskeleton Cockpit. *arXiv preprint arXiv:2502.13013*, 2025.

[2] Qingwei Ben, Botian Xu, Kailin Li, Feiyu Jia, Wentao Zhang, Jingping Wang, Jingbo Wang, Dahua Lin, and Jiangmiao Pang. GALLANT: Voxel Grid-Based Humanoid Locomotion and Local-Navigation across 3D Constrained Terrains. *arXiv preprint arXiv:2511.14625*, 2025.

[3] Johan Bjorck, Fernando Castañeda, Nikita Cherniadev, Xingye Da, Runyu Ding, Linxi Fan, Yu Fang, Dieter Fox, Fengyuan Hu, Spencer Huang, et al. GR00T N1: An Open Foundation Model for Generalist Humanoid Robots. *arXiv preprint arXiv:2503.14734*, 2025.

[4] Bones Studio. Bones-Seed: A large-scale human motion capture dataset, 2026. https://huggingface.co/datasets/bones-studio/seed.

[5] Maiyue Chen, Kaihui Wang, Bo Zhang, Xihan Ma, Zhiyuan Yang, Yi Ren, Qijun Huang, Zihao Zhu, Yucheng Wang, and Zhizhong Su. HoloMotion-1 Technical Report. *arXiv preprint arXiv:2605.15336*, 2026.

[6] Sirui Chen, Zi-ang Cao, Zhengyi Luo, Fernando Castañeda, Chenran Li, Tingwu Wang, Ye Yuan, Linxi Fan, C Karen Liu, Yuke Zhu, et al. CHIP: Adaptive Compliance for Humanoid Control through Hindsight Perturbation. *arXiv preprint arXiv:2512.14689*, 2025.

[7] Sirui Chen, Shibo Zhao, Zhen Wu, Jiaman Li, Guanya Shi, and C. Karen Liu. SceneBot: Contact-Prompted General Humanoid Whole Body Tracking with Scene-Interaction. *arXiv preprint arXiv:2606.27581*, 2026.

[8] 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*, 2025.

[9] Ziyang Cheng, Haoyu Wei, Hang Yin, Xiuwei Xu, Bingyao Yu, Jie Zhou, and Jiwen Lu. CMP: Robust Whole-Body Tracking for Loco-Manipulation via Competence Manifold Projection. *arXiv preprint arXiv:2604.07457*, 2026.

[10] ChingMu. MotionDecode: Chingmu 1000-hour embodied motion dataset, 2026. https://huggingface.co/datasets/CMRobot/MotionDecode.

[11] Robin Deits and Russ Tedrake. Footstep Planning on Uneven Terrain with Mixed-Integer Convex Optimization. In *IEEE-RAS International Conference on Humanoid Robots*, pages 279–286. IEEE, 2014.

[12] 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 *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13336–13348, 2025.

[13] Jinrui Han, Weiji Xie, Jiakun Zheng, Jiyuan Shi, Weinan Zhang, Ting Xiao, and Chenjia Bai. KungfuBot2: Learning Versatile Motion Skills for Humanoid Whole-Body Control. *arXiv preprint arXiv:2509.16638*, 2025.

[14] Tairan He, Zhengyi Luo, Xialin He, Wenli Xiao, Chong Zhang, Weinan Zhang, Kris Kitani, Changliu Liu, and Guanya Shi. OmniH2O: Universal and Dexterous Human-to-Humanoid Whole-Body Teleoperation and Learning. *arXiv preprint arXiv:2406.08858*, 2024.

[15] Tairan He, Jiawei Gao, Wenli Xiao, Yuanhang Zhang, Zi Wang, Jiashun Wang, Zhengyi Luo, Guanqi He, Nikhil Sobanbab, Chaoyi Pan, et al. ASAP: Aligning Simulation and Real-World Physics for Learning Agile Humanoid Whole-Body Skills. *arXiv preprint arXiv:2502.01143*, 2025.

[16] 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)*, pages 9989–9996. IEEE, 2025.

[17] Tao Huang, Junli Ren, Huayi Wang, Zirui Wang, Qingwei Ben, Muning Wen, Xiao Chen, Jianan Li, and Jiangmiao Pang. Learning Humanoid Standing-up Control across Diverse Postures. *arXiv preprint arXiv:2502.08378*, 2025.

[18] Yifeng Jiang, Yuting Ye, Deepak Gopinath, Jungdam Won, Alexander W. Winkler, and C. Karen Liu. Transformer Inertial Poser: Real-Time Human Motion Reconstruction from Sparse IMUs with Simultaneous Terrain Generation. In *SIGGRAPH Asia Conference Papers*, pages 1–9, 2022.

[19] Scott Kuindersma, Robin Deits, Maurice Fallon, Andrés Valenzuela, Hongkai Dai, Frank Permenter, Twan Koolen, Pat Marion, and Russ Tedrake. Optimization-Based Locomotion Planning, Estimation, and Control Design for the Atlas Humanoid Robot. *Autonomous Robots*, 40(3):429–455, 2016.

[20] Yitang Li, Zhengyi Luo, Tonghe Zhang, Cunxi Dai, Anssi Kanervisto, Andrea Tirinzoni, Haoyang Weng, Kris Kitani, Mateusz Guzek, Ahmed Touati, Alessandro Lazaric, Matteo Pirotta, and Guanya Shi. BFM-Zero: A Promptable Behavioral Foundation Model for Humanoid Control Using Unsupervised Reinforcement Learning. *arXiv preprint arXiv:2511.04131*, 2025.

[21] Qiayuan Liao, Takara E Truong, Xiaoyu Huang, Guy Tevet, Koushil Sreenath, and C Karen Liu. BeyondMimic: From Motion Tracking to Versatile Humanoid Control via Guided Diffusion. *arXiv preprint arXiv:2508.08241*, 2025.

[22] Zhengyi Luo, Jinkun Cao, Josh Merel, Alexander Winkler, Jing Huang, Kris Kitani, and Weipeng Xu. Universal Humanoid Motion Representations for Physics-Based Control. *arXiv preprint arXiv:2310.04582*, 2023.

[23] Zhengyi Luo, Ye Yuan, Tingwu Wang, Chenran Li, Sirui Chen, Fernando Castañeda, Zi-Ang Cao, Jiefeng Li, David Minor, Qingwei Ben, et al. SONIC: Supersizing Motion Tracking for Natural Humanoid Whole-Body Control. *arXiv preprint arXiv:2511.07820*, 2025.

[24] 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 (ICCV)*, 2019.

[25] Ziyu Meng, Tengyu Liu, Le Ma, Yingying Wu, Ran Song, Wei Zhang, and Siyuan Huang. SafeFall: Learning Protective Control for Humanoid Robots. *arXiv preprint arXiv:2511.18509*, 2025.

[26] Fabian Mentzer, David Minnen, Eirikur Agustsson, and Michael Tschannen. Finite Scalar Quantization: VQ-VAE Made Simple. *arXiv preprint arXiv:2309.15505*, 2023.

[27] Mayank Mittal, Pascal Roth, James Tigue, Antoine Richard, Octi Zhang, Peter Du, Nikita Rudin, et al. Isaac Lab: A GPU-Accelerated Simulation Framework for Multi-Modal Robot Learning, 2025.

[28] 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*, 37(4):1–14, 2018.

[29] Xue Bin Peng, Ze Ma, Pieter Abbeel, Sergey Levine, and Angjoo Kanazawa. AMP: Adversarial Motion Priors for Stylized Physics-Based Character Control. *ACM Transactions on Graphics*, 40(4), 2021.

[30] Zekun Qi, Xuchuan Chen, Dairu Liu, Chenghuai Lin, Yunrui Lian, Sikai Liang, Zhikai Zhang, Yu Guan, Jilong Wang, Wenyao Zhang, Xinqiang Yu, He Wang, and Li Yi. Humanoid-GPT: Scaling Data and Structure for Zero-Shot Motion Tracking. *arXiv preprint arXiv:2606.03985*, 2026.

[31] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal Policy Optimization Algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[32] Jianlin Su, Murtadha Ahmed, Yu Lu, Shengfeng Pan, Wen Bo, and Yunfeng Liu. RoFormer: Enhanced Transformer with Rotary Position Embedding. *Neurocomputing*, 568:127063, 2024.

[33] Chen Tessler, Yunrong Guo, Ofir Nabati, Gal Chechik, and Xue Bin Peng. MaskedMimic: Unified Physics-Based Character Control Through Masked Motion Inpainting. *ACM Transactions on Graphics*, 43(6):1–21, 2024.

[34] Emanuel Todorov, Tom Erez, and Yuval Tassa. MuJoCo: A Physics Engine for Model-Based Control. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5026–5033. IEEE, 2012.

[35] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention Is All You Need. *Advances in Neural Information Processing Systems*, 30, 2017.

[36] 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*, 2026.

[37] Yuxuan Wang, Ming Yang, Weishuai Zeng, Yu Zhang, Xinrun Xu, Haobin Jiang, Ziluo Ding, and Zongqing Lu. From Experts to a Generalist: Toward General Whole-Body Control for Humanoid Robots. *arXiv preprint arXiv:2506.12779*, 2025.

[38] Haoyang Weng, Yitang Li, Nikhil Sobanbabu, Zihan Wang, Zhengyi Luo, Tairan He, Deva Ramanan, and Guanya Shi. HDMI: Learning Interactive Humanoid Whole-Body Control from Human Videos. *arXiv preprint arXiv:2509.16757*, 2025.

[39] Zhen Wu, Xiaoyu Huang, Lujie Yang, Yuanhang Zhang, Koushil Sreenath, Xi Chen, Pieter Abbeel, Rocky Duan, Angjoo Kanazawa, Carmelo Sferrazza, et al. Perceptive Humanoid Parkour: Chaining Dynamic Human Skills via Motion Matching. *arXiv preprint arXiv:2602.15827*, 2026.

[40] Michael Xu, Yi Shi, KangKang Yin, and Xue Bin Peng. PARC: Physics-Based Augmentation with Reinforcement Learning for Character Controllers. In *SIGGRAPH Conference Papers*, pages 1–11, 2025.

[41] 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.

[42] 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. *arXiv preprint arXiv:2507.07356*, 2025.

[43] Mingqi Yuan, Tao Yu, Wenqi Ge, Xiuyong Yao, Huijiang Wang, Jiayu Chen, Bo Li, Wei Zhang, Wenjun Zeng, Hua Chen, and Xin Jin. A Survey of Behavior Foundation Model: Next-Generation Whole-Body Control System of Humanoid Robots. *arXiv preprint arXiv:2506.20487*, 2025.

[44] Yanjie Ze, Zixuan Chen, João Pedro Araújo, Zi ang Cao, Xue Bin Peng, Jiajun Wu, and C. Karen Liu. TWIST: Teleoperated Whole-Body Imitation System, 2025.

[45] Yanjie Ze, Siheng Zhao, Weizhuo Wang, Angjoo Kanazawa, Rocky Duan, Pieter Abbeel, Guanya Shi, Jiajun Wu, and C Karen Liu. TWIST2: Scalable, Portable, and Holistic Humanoid Data Collection System. *arXiv preprint arXiv:2511.02832*, 2025.

[46] 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.

[47] Chong Zhang, Wenli Xiao, Tairan He, and Guanya Shi. WoCoCo: Learning Whole-Body Humanoid Control with Sequential Contacts. *arXiv preprint arXiv:2406.06005*, 2024.

[48] Chong Zhang, Victor Klemm, Fan Yang, and Marco Hutter. AME-2: Agile and Generalized Legged Locomotion via Attention-Based Neural Map Encoding. *arXiv preprint arXiv:2601.08485*, 2026.

[49] Qiang Zhang, Jiahao Ma, Peiran Liu, Shuai Shi, Zeran Su, Zifan Wang, Jingkai Sun, Wei Cui, Jialin Yu, Gang Han, Wen Zhao, Pihai Sun, Kangning Yin, Jiaxu Wang, Jiahang Cao, Lingfeng Zhang, Hao Cheng, Xiaoshuai Hao, Yiding Ji, Junwei Liang, Jian Tang, Renjing Xu, and Yijie Guo. MeshMimic: Geometry-Aware Humanoid Motion Learning through 3D Scene Reconstruction. *arXiv preprint arXiv:2602.15733*, 2026.

[50] Yuanhang Zhang, Younggyo Seo, Juyue Chen, Yifu Yuan, Koushil Sreenath, Pieter Abbeel, Carmelo Sferrazza, Karen Liu, Rocky Duan, and Guanya Shi. RPL: Learning Robust Humanoid Perceptive Locomotion on Challenging Terrains. *arXiv preprint arXiv:2602.03002*, 2026.

[51] 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. *arXiv preprint arXiv:2604.17335*, 2026.

[52] Zhikai Zhang, Jun Guo, Chao Chen, Jilong Wang, Chenghuai Lin, Yunrui Lian, Han Xue, Zhenrong Wang, Maoqi Liu, Huaping Liu, et al. Track Any Motions under Any Disturbances. *arXiv preprint arXiv:2509.13833*, 2025.

[53] Siheng Zhao, Yanjie Ze, Yue Wang, C Karen Liu, Pieter Abbeel, Guanya Shi, and Rocky Duan. ResMimic: From General Motion Tracking to Humanoid Whole-Body Loco-Manipulation via Residual Learning. *arXiv preprint arXiv:2510.05070*, 2025.

[54] Shaoting Zhu, Ziwen Zhuang, Mengjie Zhao, Kun-Ying Lee, and Hang Zhao. Hiking in the Wild: A Scalable Perceptive Parkour Framework for Humanoids. *arXiv preprint arXiv:2601.07718*, 2026.