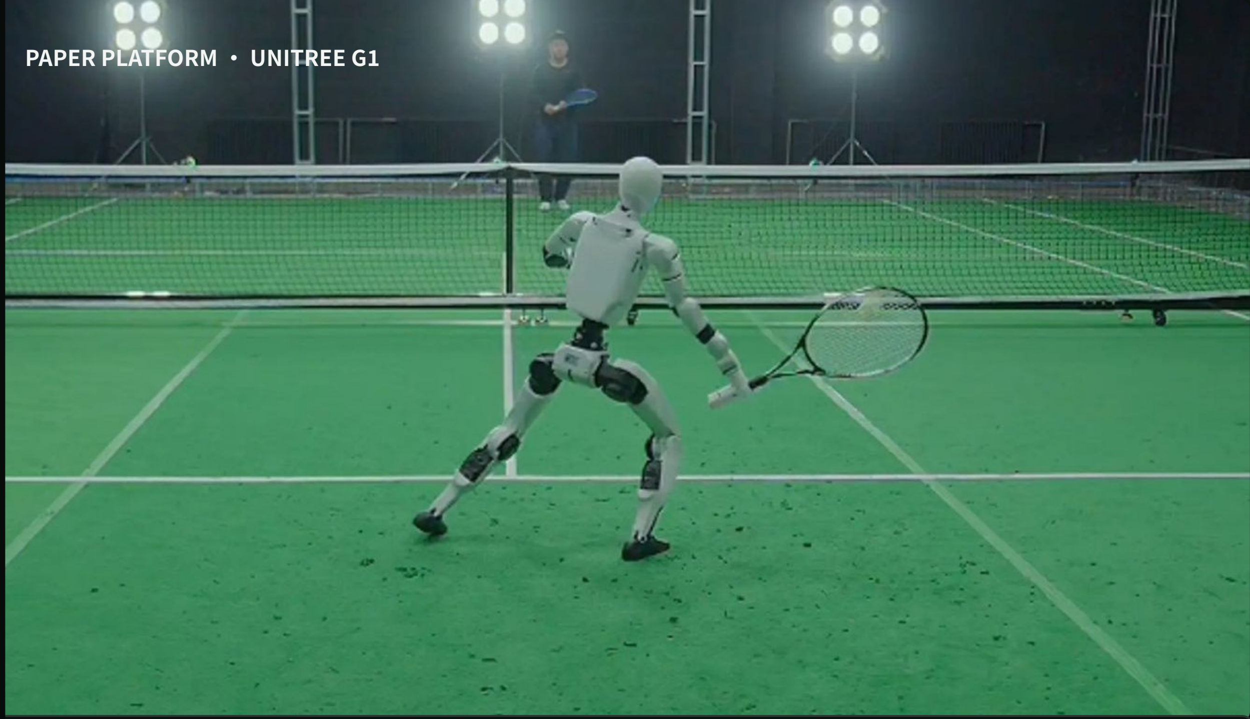
How Galbot taught a humanoid to play tennis

From a Unitree G1 research deployment to a Galbot ET1 arena demo:
what LATENT reveals—and does not reveal—about cross-embodiment whole-body learning

TECHNICAL BRIEF

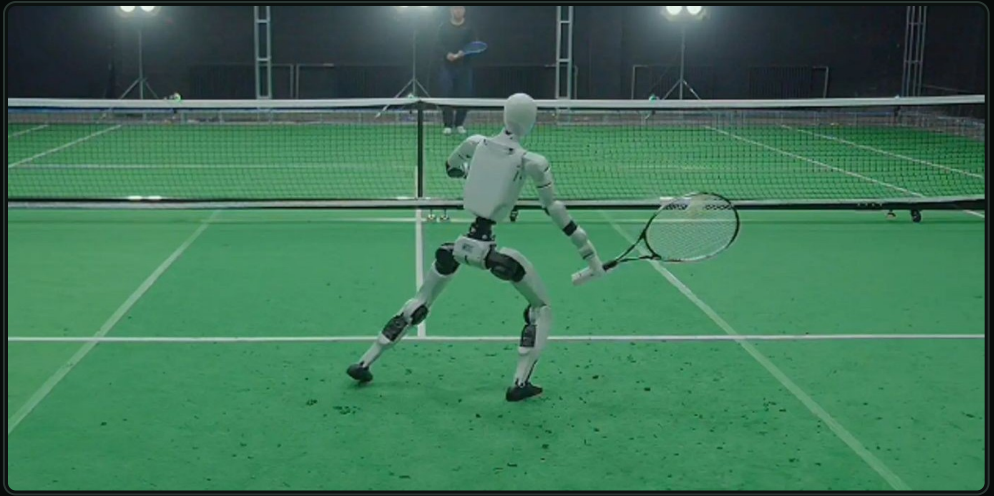
ONE SYSTEM / TWO BODIES



01 / EMBODIMENT UPDATE

First correction: LATENT has appeared on two different bodies

MAR 2026 • PAPER

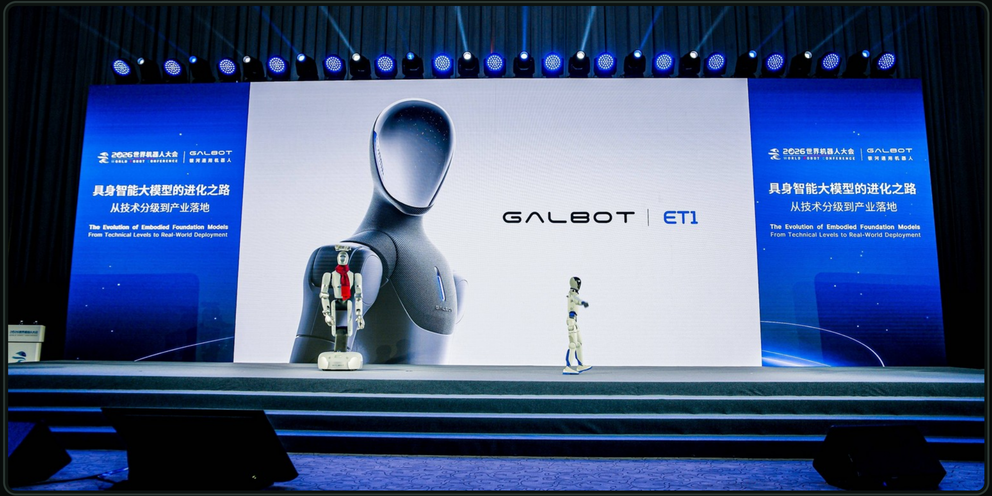


UNITREE G1

- Platform used for every LATENT paper experiment
- Commercial 29-DoF compact humanoid
- External optical mocap + modified racket end-effector

LATENT

AUG 2026 • PRODUCT



GALBOT ET1 • GALAXY STAR

- Galbot’ s first in-house biped
- Public tennis showcase on 22 August used LATENT
- Transfer method and control stack remain undisclosed

The same system name on two bodies is evidence of portability—not proof that identical weights transferred unchanged.

Sources: LATENT paper; Unitree; Xinhua; Galbot event coverage

The paper proved LATENT on a 29-DoF Unitree G1

INPUT	Global ball and robot state; external optical mocap in the real deployment
PRIOR	Forehand, backhand, shuffle and crossover motion fragments
REPRESENTATION	Motion tracker → online distillation → correctable latent action space
POLICY	High-level PPO composes skills and corrects latent and right-wrist actions
OUTPUT	50 Hz whole-body joint targets; a PD controller produces actuator torques

5 h
motion fragments from 5 amateurs

50 Hz
high- and low-level control

90.9%
real-world forehand return SR

77.8%
real-world backhand return SR

Source: LATENT paper, Sections 3–5 and Appendix B

Tennis collapses five robotics problems into a few hundred milliseconds

PERCEPTION	→	PREDICTION	→	LOCOMOTION	→	CONTACT	→	RECOVERY
Fast small target position / velocity / spin	→	Flight and bounce strike point / timing	→	Large lateral motion reach + balance	→	Racket pose / speed millisecond window	→	Reposition prepare the next shot

Table tennis pushes perception and strike timing to the limit. Tennis adds a larger workspace, whole-body coordination and a much stronger need for motion priors.

Sources: LATENT introduction; HITTER introduction

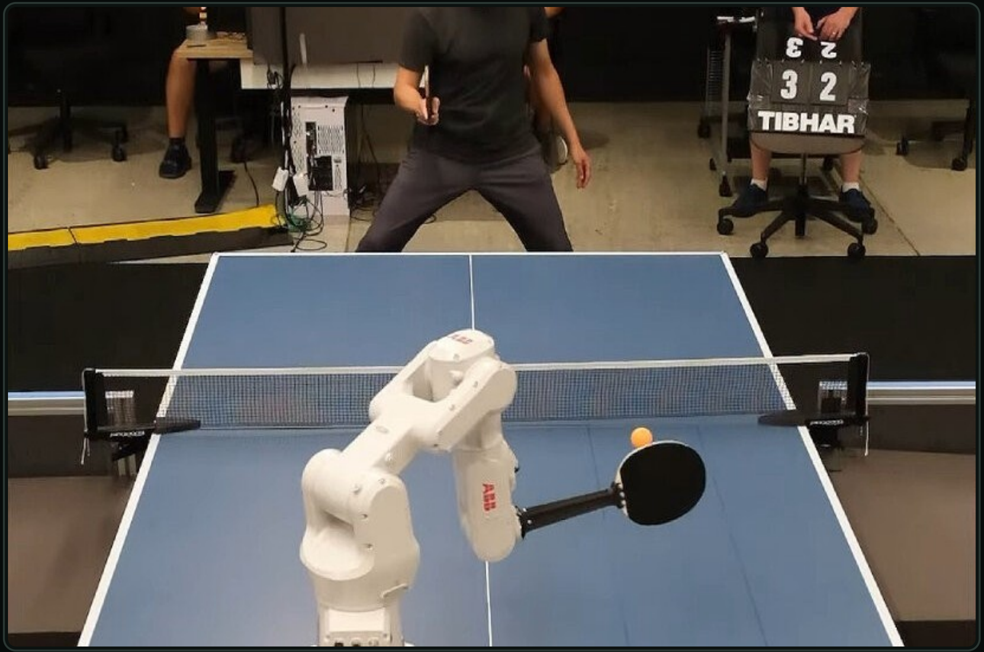
Ball-playing robots keep changing the hierarchy, not removing it

2010–13	2024	2025	MAR 2026	APR 2026	AUG 2026
MOTION PRIMITIVES	**DEEPMIND**	**HITTER**	**LATENT**	**SONY ACE**	**GALBOT ET1**
Demonstration templates explicit trajectories	Skill library high-level selection	Physics planner RL whole-body	Human prior latent skill	Event vision model-free RL	In-house body product showcase

The field did not converge on “no structure.” As interaction gets faster, systems become more explicit about what belongs to physics, policy, and motor priors.

Sources: DeepMind; HITTER; LATENT; Sony AI; Galbot ET1 releases

DeepMind separated stroke execution from match-level skill selection



SPECIALIZED ARM + LINEAR RAIL

It removes biped balance to isolate striking and match strategy.

LOW LEVEL

Multiple specialized controllers

Forehand, backhand, target locations and speeds are represented as separate skills.

HIGH LEVEL

Select a skill in real time and adapt to unseen opponents

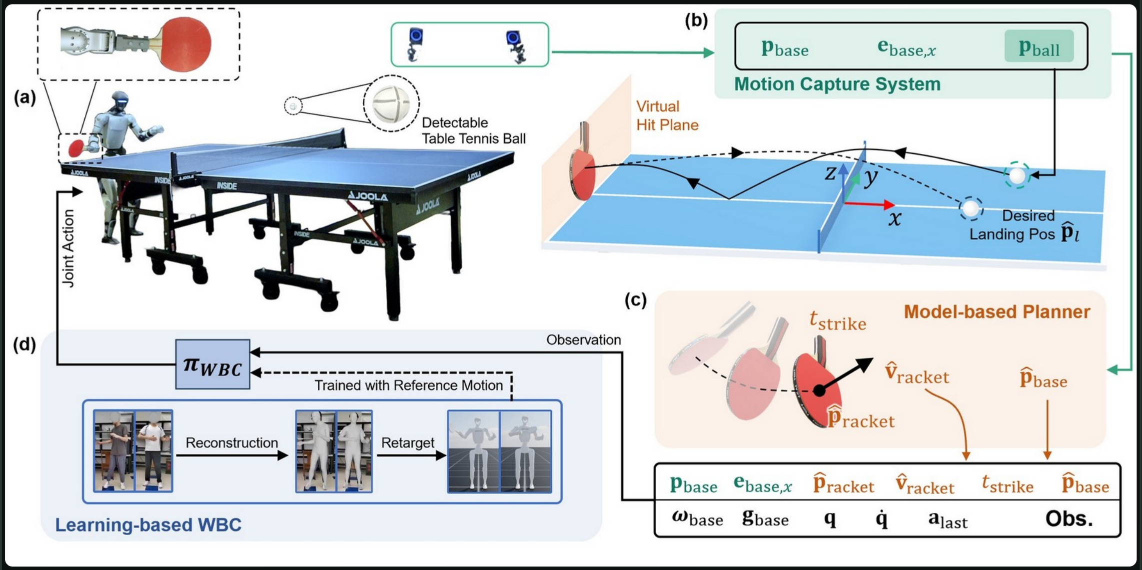
13 / 29

Human matches won; 100% vs beginners and 55% vs intermediate players.

Intelligence sits mainly in skill selection, not one network generating every joint action.

Source: Google DeepMind, Achieving Human Level Competitive Robot Table Tennis

HITTER lets physics choose the strike; RL makes the body execute it



01 **STATE ESTIMATION**

Nine OptiTrack cameras at 360 Hz; smooth position and estimate velocity.

02 **TRAJECTORY PREDICTION**

Explicit gravity, air drag and table-bounce dynamics.

03 **STRIKE PLANNING**

Compute racket position, velocity and timing from the target landing point.

04 **WHOLE-BODY EXECUTION**

PPO outputs 29 joint targets; a PD controller converts them to torques.

106 maximum consecutive shots with a human

Source: HITTER arXiv 2508.21043

07 / MODEL-BASED PLANNER

HITTER’ s lesson is what does not need to be learned

FLIGHT

$$a = -k \| v \| v + g$$

air drag + gravity

→

BOUNCE

$$v^{+} = C \cdot v^{-}$$

horizontal / vertical restitution

→

RACKET

incoming → desired outgoing

solve for strike velocity and direction

0.5 s before strike: position error falls within the 7.5 cm racket-radius threshold

0.3 s before strike: timing error below 20 ms—one 50 Hz control cycle

Philosophy: the planner provides a stable, low-dimensional command interface; RL solves 29-DoF whole-body coordination.

Source: HITTER, Sections IV–VI

LATENT learns the language of human movement before learning to return the ball

HUMAN FRAGMENTS		MOTION TRACKER		ONLINE DISTILLATION		HIGH-LEVEL RL		WHOLE BODY
Forehand / backhand shuffle / crossover	→	Retarget human motion into executable skills	→	Compress tracking skills into one latent space	→	Compose skills from ball state and output task corrections	→	Stable, natural whole-body tennis motion

≈ 5 h
human motion

3 × 5 m
capture area

8 GPU
parallel PPO

50 Hz
control

The high-level policy no longer searches 29 joints directly; it searches a low-dimensional manifold that already contains footwork, hip rotation and stroke structure.

Sources: LATENT paper; official GitHub pipeline

Imperfect motion is enough—if it supplies structure, not answers

TRADITIONAL TARGET	LATENT'S CUT	WHAT IT BUYS
Complete match motion	**Capture primitive fragments only**	Much lower collection cost
Complete ball + human trajectories	**Allow incomplete fragments**	Data scales more easily
Precise wrist + racket pose	**Learn style and coordination first**	Preserve a human-like low-level prior
Large space, long capture sessions	**Let RL recover task precision**	High level can still optimize for the ball

Key assumption: human data only needs to supply the structure of primitive skills; the downstream policy can recover task-specific precision.

Source: LATENT, data collection and method

The prior must guide the policy without locking it down

MOTION PRIOR		TASK CORRECTION		EXECUTABLE SKILL
Natural motion whole-body coordination footwork and stroke structure	→	Condition on the incoming ball correct latent + right wrist satisfy the current strike	→	Human-like motion strike meets the target dynamic stability

Imitation alone can look human and still miss the ball. Task RL alone can exploit the latent space and produce jittery, unnatural motion.

Source: LATENT method and ablations

The body changed—and so did the unresolved perception question

UNITREE G1 • DOCUMENTED SYSTEM

50+
motion-capture cameras

2048^2
camera resolution

120 Hz
mocap update rate

19 × 15 m
capture area

≈ ¥350k
three-week venue + system rental

The venue supplies global ball and robot pose.

GALBOT ET1 • PUBLIC SHOWCASE

PUBLIC DESCRIPTION

Event coverage says ET1 used binocular vision for the ball and LATENT for autonomous tennis.

Undisclosed: any remaining external cameras, the ball-state estimator, end-to-end latency, control rate, and which policies or interfaces changed from G1 to ET1.

Sources: LATENT GitHub deployment details; Galbot ET1 event coverage

Four ball-playing systems are solving four different problems

SYSTEM	PRIMARY PROBLEM	HIGH LEVEL	LOW LEVEL	EXTERNALIZED
DEEPMIND 2024	Competitive play / opponent adaptation	**skill selection**	learned low-level skills	biped / whole body
HITTER 2025	High-speed humanoid striking	**physics planner**	RL whole-body controller	external mocap / simplified spin
LATENT 2026	Whole-body tennis composition	**latent skill policy**	motion prior + RL	paper: external mocap
SONY ACE 2026	Elite speed and spin	**model-free rally policy**	optimizer + high-speed execution	specialized hardware cost

None is purely end-to-end. The structured middle layer is a skill, physics target, latent prior or trajectory optimizer.

Sources: DeepMind; HITTER; LATENT; Sony AI / Nature

Two bodies suggest portability; they do not prove plug-and-play transfer

WHAT IT SUGGESTS

- **The tennis-skill representation can be redeployed to bodies with different scale and dynamics.**
- Galbot is moving a paper capability into in-house hardware and a product narrative.
- The body becomes a test of model reuse, not just experimental hardware.

WHAT IT DOES NOT PROVE

- G1 and ET1 share identical policy weights.
- The 29-DoF action space maps to ET1 without redesign.
- No retargeting, fine-tuning or low-level controller retraining was required.
- ET1’ s vision loop has paper-level reproducible evidence.

The next evidence should document G1 → ET1 retargeting, control interfaces, training and perception.

RoboAtlas analysis based on disclosed system boundaries

The next milestone is evidence across four interfaces—not a faster highlight reel

01	**ONBOARD PERCEPTION**	Can onboard cameras estimate ball position, velocity and spin under speed, occlusion and motion blur?
02	**TRANSFER DISCLOSURE**	What changed in retargeting, action space, low-level control and training from G1 to ET1?
03	**GAME STRATEGY**	Move from return/rally to serves, spin, placement, opponent modelling and genuine match decisions.
04	**SYSTEM COST**	Outside a high-spec venue, can perception, compute, joints and reliability sustain the same loop?

ET1 changes the question from “does the paper work?” to “does the capability survive a new body, sensing stack and engineering system?”

RoboAtlas analysis; source boundaries documented in prior slides

More robotics technology briefs—continuously updated at ROBOATLAS.AI

RoboAtlas is a global robotics intelligence platform that organizes companies, robots, embodied AI models, components and technical routes into structured, source-linked profiles.

roboatlas.ai

Research, product updates and collaboration

FOLLOW ROBOATLAS

WEBSITE

roboatlas.ai

LINKEDIN

linkedin.com/company/roboatlas

X

x.com/RoboAtlasbot

ROBOTS • MODELS • COMPANIES • COMPONENTS