

● SIM-TO-REAL · HUMANOID ROBOTICS · UNITREE G1

Sim-to-Real Training für Humanoid-Roboter.

Verfahren, Herausforderungen und robuste Trainingsarchitekturen am Beispiel des **Unitree G1**.

23–43 FREIHEITSGRADE UNITREE G1	≈ 50^{min} 1 JAHR ROBOTERTRAINING HOVER, 1× GPU	100k+ BEWEGTE TOTES DIGIT @ GX0	9 BRÜCKEN-BAUSTEINE DER HYBRIDEN PIPELINE
--	---	--	--

HERAUSGEBER

wetando Management GmbH
Dr. Martin Schunk · Leipzig /
Berlin

ZIELGRUPPE

Forschung · Industrie · KMU
Ingenieurinnen und Ingenieure
der Robotik

KONTAKT

hallo@robotskills.de
www.robotskills.de

ISBN – INTERNAL PREPRINT
RS-WP-01 / 2026.05

IMPRESSUM & DISCLAIMER

Über dieses Whitepaper.

HERAUSGEBER

wetando Management GmbH

Mädler-Passage (Treppe B)

Grimmaische Straße 2–4

04109 Leipzig

Vertreten durch Dr. Martin Schunk

BETRIEBSSTÄTTE BERLIN

Schönhauser Allee 163

10435 Berlin

hallo@robotskills.dewww.robotskills.de

ZIELGRUPPE

Forschungsinstitute, Hochschulen, Robotik-
Ingenieurinnen und -Ingenieure, Industrie-
unternehmen, Innovations- und Transferzentren,
KI- und robotik-affine KMU.

STAND

20. Mai 2026

Whitepaper-Reihe robotskills

Band 01

Vollständiges Impressum unter robotikcheck.de/impressum. Dieses Whitepaper fasst den Stand der wissenschaftlichen und industriellen Praxis zum Stand Mai 2026 zusammen. Es ersetzt keine technische, rechtliche oder förderrechtliche Beratung im konkreten Anwendungsfall.

Achtzehn Kapitel.

Von der wissenschaftlichen Verortung des Sim-to-Real-Problems über eine technische Tiefenanalyse bis zur kritischen Bewertung industrieller Architekturen — und konkreter Handlungsempfehlungen.

01	Executive Summary	06
02	Einführung	08
03	Stand der Humanoid-Robotik 2025/2026	10
04	Das Sim-to-Real-Problem	14
05	Reinforcement Learning für Humanoids	18
06	Simulationsplattformen	21
07	Trainingsverfahren im Vergleich	24
08	Sim-to-Sim als kritische Zwischenstufe	27
09	unitree_rl_mjlab-Referenzstack	30
10	Teacher-Student-Distillation	33
11	Hierarchical Policies und modulare Skill-Systeme	36
12	Industrial Humanoid Robotics	39
13	Warum Haushaltsrobotik schwieriger ist	42
14	Sicherheits- und Runtime-Aspekte	44
15	Zukunft: Foundation Policies und World Models	46
16	Kritische Bewertung aktueller Ansätze	48
17	Fazit und Handlungsempfehlungen	50
18	Literaturverzeichnis	53

Executive Summary.

Humanoide Roboter erleben 2025/2026 einen technischen Durchbruch, der weniger durch neue Mechanik als durch eine konvergierende Lerninfrastruktur getragen wird.

GPU-beschleunigte Physik (Isaac Lab, MuJoCo XLA, Newton), preiswerte Forschungsplattformen wie der Unitree G1 (23 bis 43 Freiheitsgrade, 1,32 m Körperhöhe, ca. 35 kg, Einstiegspreis um 16.000 USD) und eine ausgereifte Pipeline aus Reinforcement Learning (PPO), Domain Randomization, Teacher-Student-Distillation und Sim-to-Sim-Validierung treffen aufeinander. Die zentrale Frage moderner Humanoid-Robotik ist nicht mehr „*können wir laufen*“, sondern „*können wir eine Policy aus der Simulation robust auf reale Hardware übertragen*“ — das Sim-to-Real-Problem.

Dieses Whitepaper analysiert die aktuell führenden Verfahren, kontextualisiert sie wissenschaftlich, bewertet sie kritisch und entwickelt eine klare Position:

THESE

Die robusteste praktisch verfügbare Architektur für Humanoid-Training ist **keine reine End-to-End-AGI, keine reine klassische Robotik, keine reine Motion-Imitation und keine reine Teleoperation**. Es ist eine hybride Pipeline aus massivem RL, Domain Randomization, Teacher-Student-Distillation, Sim-to-Sim-Validierung, hierarchischen modularen Skill-Policies und einem deterministischen Safety- und PD-Layer auf der Hardware.

Genau diese Architektur bildet auch der offene Referenzstack `unitree_rl_mjlab` ab – ergänzt um drei Brücken-Mechanismen, die wir in diesem Whitepaper systematisch herleiten.

Konkret argumentieren wir, dass

- Sim-to-Sim die *unterschätzte Lebensversicherung* jedes Sim-to-Real-Projekts ist,
- Runtime-Gleichheit zwischen Trainings- und Deployment-PD-Loop kritischer ist als ausgefeiltes Reward-Engineering,
- Teacher-Student-Distillation aktuell den State of the Art für die Lücke zwischen privilegierten Simulationsbeobachtungen und realer Sensorik darstellt (Lee et al., 2020; Isaac Lab Documentation, 2025),
- modulare Skill-Stacks (HOVER, ASE, Helix) realistischer skalieren als eine monolithische General-Policy,
- Humanoide kurzfristig in Industrie und Logistik nutzbar werden, im Haushalt aber noch viele Jahre brauchen.

Einführung.

2.1 Warum dieses Whitepaper

In den letzten 24 Monaten hat sich die Diskussion über humanoide Roboter dramatisch verschoben. Wo 2022/2023 vor allem spektakuläre Marketing-Demos dominierten, dominieren 2025/2026 reproduzierbare, peer-reviewte Sim-to-Real-Pipelines. Gleichzeitig sind die wesentlichen Komponenten — physikalisch akkurate GPU-Simulatoren, eine offene Foundation-Policy für Humanoide (Isaac GROOT N1), ein bezahlbarer Forschungsroboter (Unitree G1) und Open-Source-Trainingsframeworks (MJLab, Isaac Lab, MuJoCo Playground) — erstmals frei zugänglich.

01

Demokratisierung

Einzelne Entwickler können mit dem offenen Stack `unitree_rl_mjlab` (Unitree Robotics, 2025) innerhalb weniger Tage eine PPO-trainierte Lauf-Policy für einen realen Unitree G1 erzeugen — ohne institutionellen Cluster.

02

Überfrachtung

Gleichzeitig dominieren VLAs, Foundation Models und World Models die öffentliche Wahrnehmung und erzeugen Erwartungen, die ein realer Sim-to-Real-Stack heute nicht einlöst.

2.2 Vorgehen

- 01** **Wissenschaftliche Verortung** des Sim-to-Real-Problems mit Quellen aus *Science Robotics*, ICRA, CoRL, RSS und arXiv von 2017 bis 2026.
- 02** **Technische Tiefen-Analyse** der relevanten Verfahren (PPO, Domain Randomization, SysID, Teacher-Student, Multi-Sim, AMP/ASE/DeepMimic, Hierarchical Policies, VLAs, World Models).
- 03** **Kritische Bewertung** der Architekturen entlang industriell relevanter Achsen: Robustheit, Datenbedarf, Compute-Bedarf, Safety, Skalierbarkeit, Hardware-Risiko.

Alle Quellen sind im Literaturverzeichnis (Kapitel 18) im APA-Stil mit URL angegeben.



ABB. A Realer Unitree G1 im wetando-Trainingssetup. Skill-Aufzeichnung am physischen Roboter ist eine zentrale Datenquelle, die Sim-to-Real-Pipelines ergänzt — wir kommen in Kapitel 12 ausführlich darauf zurück.

Stand der Humanoid-Robotik 2025 / 2026.

3.1 Hardware-Landschaft

Die heutige humanoide Hardware ist heterogen, lässt sich aber grob in drei Klassen einteilen:

KLASSE	BEISPIELE	HÖHE / GEWICHT	ZIELMARKT	PREISINDIKATION
Forschung & Education	Unitree G1 (EDU), Booster T1, Fourier GR-1	1,2–1,7 m / 30–55 kg	Hochschulen, Startups	16k – 80k USD
Industrial-Pilots	Aptronik Apollo, Agility Digit V4, Boston Dynamics Electric Atlas, 1X NEO Beta	1,5–1,9 m / 65–95 kg	Logistik, Fertigung	RaaS · hohe sechsstellig
Verbraucher / Home	1X NEO, Figure 03 (geplant)	~1,7 m / 30–55 kg	Haushalt (perspektivisch)	20.000 USD · 499 USD/Mt.

Quellen u. a.: Unitree Robotics (2026), Aptronik & Mercedes-Benz (2024), Agility Robotics (2024, 2025), Boston Dynamics (2024, 2025), 1X Technologies (2025, 2026), Humanoids Daily (2026).

3.2 Unitree G1 als Referenzplattform

Wir wählen den **Unitree G1** als durchgängiges Bezugssystem dieses Whitepapers, weil er fünf Kriterien erfüllt, die ihn zur faktischen Referenzplattform für offene Sim-to-Real-Forschung machen:

<p>Verfügbarkeit</p> <p>Kommerziell ab ca. 16.000 USD (Basismodell) bis 73.900 USD (Top-EDU-Konfiguration mit Dex3-1-Hand), bestellbar als Einzelgerät.</p>	<p>Offene Tooling-Kette</p> <p>URDF / MJCF, unitree_rl_mjlab mit PPO, Sim-to-Sim via unitree_mujoco, Sim-to-Real über Ethernet (192.168.123.222, Gamepad).</p>
<p>Sinnvolle Mechanik</p> <p>23 DOF Basis, bis 43 DOF EDU mit Dex3-1-Hand; Knie-Drehmoment 120 Nm; Armnutzlast bis 3 kg; ≈ 2 m/s Gang.</p>	<p>Nähe zur Industrie</p> <p>Größe und Aktuator-Klasse vergleichbar mit Aptronik Apollo oder Agility Digit — Methoden grundsätzlich übertragbar.</p>
<p>Aktive Forschungs-Community</p> <p>Zahlreiche peer-reviewte Arbeiten 2025/2026 nutzen den G1: He et al. 2025a (HOVER), Lei et al. 2025 (PolySim), He et al. 2025b (VIRAL).</p>	



ABB. B Standard-Aufbau für Sim-to-Real-Tests: G1 in Hebevorrichtung mit angeflanschem Gewicht-Set. Das Tethering ist zwingend, solange neue Policies noch nicht in mehreren Sim-to-Sim-Iterationen validiert sind (Kapitel 8, 14).

3.3 Software-Landschaft

Drei Stacks dominieren 2026 die offene Forschung:

<p>01 · NVIDIA</p> <p>Isaac Lab</p> <p>Mittal et al., 2025</p> <p>Nachfolger von Isaac Gym mit der neuen Newton-Physik-Engine (gemeinsam mit Google DeepMind und Disney Research entwickelt). Bis zu 4.096 parallele Environments auf einer GPU.</p>	<p>02 · GOOGLE</p> <p>MuJoCo Playground</p> <p>Zakka et al., 2025</p> <p>MuJoCo XLA + JAX, eingebaute Madrona-Render-Batch, Zero-Shot-Sim-to-Real demonstriert. <i>Outstanding Demo Paper Award</i> RSS 2025.</p>	<p>03 · UNITREE</p> <p>MJLab</p> <p>Zakka et al., 2026</p> <p>Leichtgewichtiger Stack: Isaac-Lab-Manager-API + MuJoCo Warp. Pragmatische Alternative für Forschungs-PCs.</p> <p><code>unitree_rl_mjlab</code> baut darauf auf.</p>
--	--	---

Zusätzlich existieren Foundation-Modelle: **NVIDIA Isaac GROOT N1** (erstes offenes Humanoid-Foundation-Modell, Dual-System-Architektur), **Helix** (Figure AI, System-2 7–9 Hz / System-1 200 Hz, Onboard-GPU), **OpenVLA** (7B, Open Source, Open X-Embodiment), **1X World Model** (video-prädiktiv).

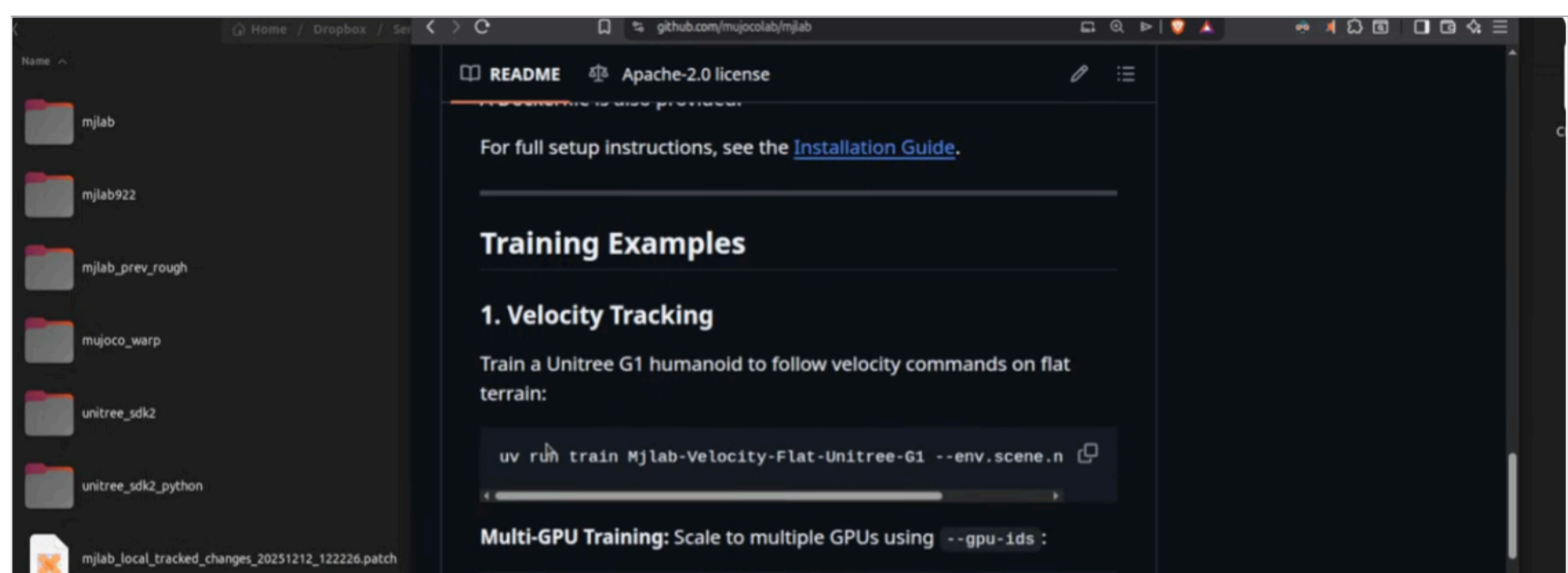


ABB. C Ausschnitt aus dem offenen mjlab-Repository (Zakka et al., 2026). Trainings-Templates für Velocity-Tracking auf dem G1 sind aus einem einzigen Kommando reproduzierbar.

3.4 Warum gerade jetzt der Durchbruch?

01

02

GPU-Physik. Über 100k Steps/s pro GPU senkt die Wall-Clock-Kosten eines RL-Trainings um Größenordnungen. HOVER: 1 Jahr Training \approx 50 Minuten realer Zeit.

Reife der RL-Algorithmen. PPO ist seit fast einem Jahrzehnt Workhorse; Reward- und Curriculum-Patterns sind in `legged_gym`, `rs1_rl`, MJLab und Isaac Lab kanonisiert.

03

Foundation-Priors. VLAs lernen hohe Semantik aus Internet-Daten, statt sie im RL-Loop zu rekonstruieren.

04

Demokratisierte Hardware. Mit dem G1 unter 20k USD wird die Plattform erstmals für Einzelne erschwinglich.

Das Sim-to-Real-Problem.

4.1 Definition

Sim-to-Real bezeichnet das Problem, eine Policy, die in einem Simulator trainiert wurde, ohne signifikanten Performance-Verlust auf reale Hardware zu übertragen. Salvato et al. (2021) definieren die *Reality Gap* als „Mismatch von simulierter und realer Steuerungs-Performance, verursacht durch eine ungenaue Repräsentation der realen Umgebung in der Simulation“.



ABB. 1 Die Reality Gap und die drei wichtigsten Brücken — Domain Randomization, System Identification und Sim-to-Sim.

4.2 Quellen des Reality Gaps

Die Reality Gap zerfällt analytisch in fünf Domänen, die jeweils eigene Gegenmaßnahmen erfordern:

DOMÄNE	TYPISCHE DISKREPANZEN	GEGENMASSNAHMEN
Dynamik	Massen, Trägheitstensoren, Reibung, Kontaktphysik, Gelenkspiel, Kabelzug	Domain Randomization, SysID, Actuator Networks
Aktuatorik	Torque/Strom-Kennlinie, Reaktionszeit, Thermik, PD-Verstärkungen	Actuator Networks (Hwangbo et al., 2019), Torque-Limit-Modeling
Sensorik	IMU-Drift, Encoder-Quantisierung, Belichtung, Sensorlatenz	Visual Domain Randomization (VIRAL), Noise-Injection
Latenz	Steuerfrequenz, Kommunikations-Jitter (EtherCAT, CAN, Ethernet)	Latenz-Randomisierung, Frame-Stacking
Beobachtungs-Mismatch	Privilegierte Sim-Größen (Basisgeschw., Bodenkontakt, Terrain-Höhe) nicht real messbar	Teacher-Student-Distillation, Asymmetric Actor-Critic

4.3 Warum Humanoide besonders schwierig sind

Sim-to-Real ist bei Humanoiden härter als bei Manipulatoren oder Quadrupeden, aus mehreren strukturellen Gründen:

- **Underactuated Dynamics.** Ein bipedaler Roboter ist im Stand bereits ein inverses Pendel; jeder Modellierungsfehler propagiert direkt in den Schwerpunktverlauf (ZMP / Capture-Point).
- **Hohe Dimensionalität.** 23–43 Aktionen und 50–100 Beobachtungen je Step. Der Raum, in dem Domain Randomization kompensieren muss, wächst exponentiell.
- **Catastrophic Failures.** Ein Sturz beschädigt das System dauerhaft — ein Quadruped kann meistens einfach wieder aufstehen.
- **Loco-Manipulation-Kopplung.** Sobald der Roboter etwas trägt, ändert sich die Dynamik des Gangs nicht-linear; eine reine Manipulations-Policy ist nicht ausreichend.
- **Sicherheits- und Ethikrahmen.** Reale Tests in der Nähe von Menschen unterliegen ISO 10218, ISO 15066 und perspektivisch dem EU AI Act.

§ 4.4 KONSEQUENZ

Die heutige beste Praxis ist **kein einzelnes Wundermittel**, sondern eine defensive Pipeline mit mehreren redundanten Brücken. Genau diese Pipeline beschreibt der Rest des Whitepapers.

Reinforcement Learning für Humanoids.

5.1 Grundbegriffe

Ein Reinforcement-Learning-Problem wird typischerweise als Markow-Entscheidungsprozess (MDP) modelliert: Zustand s , Aktion a , Übergang $p(s'|s,a)$, Belohnung $r(s,a)$, Diskontierungsfaktor γ . Eine Policy $\pi(a|s)$ wird so optimiert, dass der erwartete diskontierte Reward maximiert wird.

ZUSTAND S_T

Proprioception (q, \dot{q}), IMU (ω, \hat{g}), Torques τ , plus Kommando c_t (z.B. Soll-Basisgeschwindigkeit v_x^*, v_y^*, ω_z^*) und optional Beobachtungshistorie.

AKTION A_T

Sollpositionen für einen unterlagerten PD-Regler (Position-Target-Mode) oder direkte Torques. Position-Target-Mode ist in der Praxis Standard, weil er robuster gegen Lernfehler ist.

REWARD R_T

Gewichtete Summe aus Velocity-Tracking, Action-Smoothness, Joint-Limits, Air-Time, Foot-Slip, Energy etc. (vgl. [legged_gym](#) - Templates, ETH/RSL).

5.2 PPO — der Workhorse

Proximal Policy Optimization (Schulman et al., 2017) ist seit 2017 der praktische Standard:

- **Trust-Region-light.** PPO klippt das Verhältnis $\pi_\theta(a|s)/\pi_{\theta_{old}}(a|s)$ auf $[1-\epsilon, 1+\epsilon]$ und vermeidet die teure KL-Constraint-Optimierung von TRPO.
- **On-policy mit Massiv-Parallelisierung.** Mit 4.096–16.384 parallelen Environments amortisiert sich der On-Policy-Charakter.
- **Robust gegen Hyperparameter.** In der Praxis sind nur Lernrate, Clip-Range, Entropie-Koeffizient und Number-of-Epochs wirklich kritisch.

Praktisch alle in diesem Whitepaper diskutierten Trainings-Stacks (Isaac Lab, MJLab, MuJoCo Playground, [unitree_rl_mjlab](#), HOVER, VIRAL) verwenden PPO als Innerschleife.

5.3 Off-Policy- und modellbasierte Alternativen

SAC, TD3

Off-policy und theoretisch sample-effizienter, in der Praxis aber empfindlich gegenüber Reward-Skalierung und schwieriger zu massiv-parallelisieren. Sie dominieren in Real-Robot-Settings mit wenig Daten — nicht in Sim-to-Real-Pipelines mit Milliarden Schritten.

Dreamer V3

Das aktuell ausgereifteste world-model-basierte RL-Verfahren (Hafner et al., 2023), mit fixen Hyperparametern über sehr verschiedene Domänen hinweg robust. Für reines Locomotion-Training auf GPUs überflüssig, für sample-effiziente Echt-Hardware-Anwendungen (1X World Model, Optimus „Neural World Simulator“) zunehmend relevant.

5.4 Reward-Engineering vs. Reference-Tracking

Reines Reward-Engineering erzeugt häufig „RL-typische“ Gangbilder, die mechanisch effizient, aber unnatürlich aussehen und nahe Singularitäten operieren. Drei Pfade umgehen das:

- 01 **DeepMimic** (Peng et al., 2018a) führt einen Imitationsterm gegen eine Referenzbewegung ein — funktioniert sehr gut für simulierte Charaktere, ist aber MoCap-abhängig.

- 02 **AMP** (Peng et al., 2021) ersetzt diesen handgeschriebenen Imitationsterm durch einen adversarial trainierten Diskriminator und erreicht naturalistische Bewegung — mit nachweislichem Sim-to-Real-Transfer auf Quadrupeden.

- 03 **ASE** (Peng et al., 2022) erweitert AMP zu einem unsupervised gelernten Skill-Embedding — nutzbar als generischer „Bewegungs-Prior“ für Downstream-Tasks.

Diese drei sind die Grundlagen, auf denen Motion-Imitation-Komponenten in modernen Humanoid-Pipelines (HOVER, OmniH2O) aufbauen.

Simulationsplattformen.

6.1 Anforderungen an einen Sim-to-Real-tauglichen Simulator

Damit ein Simulator als ernsthafter Sim-to-Real-Anker dienen kann, muss er fünf Kriterien gleichzeitig erfüllen:

<p>01</p> <p>Numerisch stabile Kontaktphysik bei großer Anzahl gleichzeitiger Kontakte.</p>	<p>02</p> <p>Hohe Parallelität auf GPU — mindestens 10^3 Environments pro GPU.</p>	<p>03</p> <p>Determinismus für Reproduzierbarkeit und Debugging.</p>	<p>04</p> <p>Korrekte Aktuator-Modellierung, inkl. Torque-Sättigung und Reflected-Inertia.</p>	<p>05</p> <p>Saubere Exportierbarkeit des Modells (URDF/MJCF/USD) und der Policy (ONNX, TorchScript).</p>
--	--	---	---	--

6.2 Plattformvergleich

PLATTFORM	BACKEND	PARALLELITÄT	STÄRKE	SCHWÄCHE
MuJoCo Todorov et al., 2012	CPU/GPU (MJX/Warp)	hoch	Goldstandard Kontaktphysik, große Community	Vanilla single-threaded; ohne MJX/Warp limitiert
Isaac Sim / Isaac Lab Mittal et al., 2025	PhysX 5 → Newton	sehr hoch	Photoreal. Rendering, ROS2, USD-Workflow	komplexer Stack, hoher VRAM-Bedarf
Newton NVIDIA et al., 2025	Warp / GPU	sehr hoch, differentiable	flexible Solver, DeepMind/Disney	jung, API im Fluss (Beta 2025)
MJLab Zakka et al., 2026	MuJoCo Warp	hoch	Isaac-API + MuJoCo, minimale Deps	Beta, Feature-Lücken zu Isaac Lab
MuJoCo Playground Zakka et al., 2025	MJX (JAX)	hoch	End-to-end GPU, Madrona, Colab-tauglich	weniger Sensor-Modelle, kleinere User-Base
Isaac Gym Makoviychuk et al., 2021	PhysX	sehr hoch	Forschungs-Klassiker	deprecated, geht in Isaac Lab auf
Brax Freeman et al., 2021	JAX	sehr hoch	rein-JAX-Workflow	Kontaktphysik weniger ausgereift
PyBullet / Gazebo	CPU	niedrig	Standard für ROS-Welt	nicht GPU-skalierbar

6.3 Bedeutung der Simulator-Bias

Lei et al. (2025) zeigen in **PolySim**, dass jede Physik-Engine eine spezifische *inductive bias* hat — Annahmen über Solver, Kontaktmodell und Constraint-Handling, die in der Simulation unsichtbar bleiben, im Real-Transfer aber als

Performanzeinbruch erscheinen. Ihre Antwort ist Multi-Sim-Training: Eine Policy wird parallel über mehrere heterogene Simulatoren trainiert, was die Sim-zu-Sim-Performance um **52,8 %** auf MuJoCo verbessert und zero-shot auf einem realen Unitree G1 erfolgreich deployt wurde.

Praktische Konsequenz. Den Trainingssimulator nicht als „Realität in einer Box“ verstehen, sondern als eine Hypothese über die Realität. Wer nur in einem Simulator trainiert, optimiert auch nur gegen dessen Annahmen.

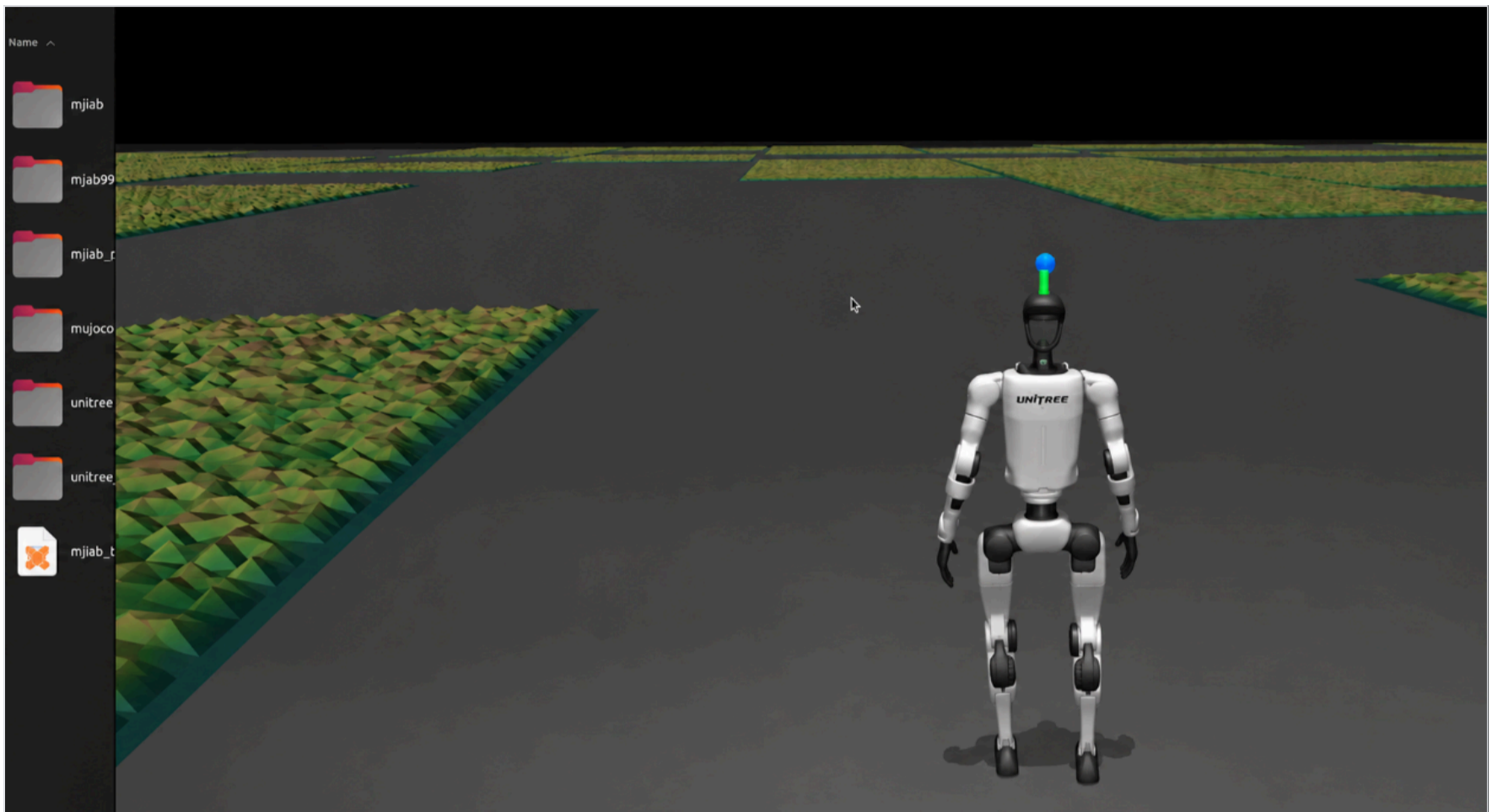


ABB. D Trainings-Simulation eines G1 im offenen mjab-Stack. Wer nur in dieser Umgebung trainiert, optimiert auch nur gegen deren spezifische Annahmen — ein klassischer Fall von Simulator-Bias, dem mit Multi-Sim-Strategien (Kapitel 6 / 8) begegnet wird.

Trainingsverfahren im Vergleich.

In diesem Kapitel ordnen wir die wesentlichen Trainingsparadigmen ein und vergleichen sie entlang einer pragmatischen Matrix.

7.1 Übersicht

- › **Klassische Robotik (modellbasiert).** ZMP, MPC, inverse Dynamik, QP-WBC — analytisch fundiert, deterministisch, aber rigide.
- › **Reinforcement Learning (PPO, SAC, Dreamer).** Lernt Steuerung direkt, robust unter Randomization, datenhungrig.
- › **Imitation Learning / Behavior Cloning.** Aus Demonstrationen (MoCap, Teleop), datenarm pro Skill, aber begrenzt.
- › **Motion-Imitation-Pipelines** (DeepMimic, AMP, ASE, OmniH2O). Hybrid aus RL und Imitation, naturalistisch, MoCap-abhängig.
- › **Teleoperation.** Mensch im Loop; zentral für Daten-Bootstrap, ungeeignet als Endzustand.
- › **Foundation Models / VLAs** (GROOT N1, OpenVLA, Helix). Internet-skalig, gute Semantik, schwach in Long-Horizon-LoCo-Manip.
- › **World Models** (Dreamer V3, 1X, Tesla „Neural World Simulator“). Sample-effizient, aber neu und mit eigener Reality Gap.
- › **Hybride Architekturen.** Kombinieren mehrere obige Bausteine über hierarchische Strukturen.

7.2 Vergleichsmatrix

VERFAHREN	SIM-TO-REAL	SKALIERBARKEIT	INDUSTRIELLE EIGNUNG	DATENBEDARF	COMPUTE	SAFETY
Klassische Robotik / WBC	★★★★	★	★★★★	-	gering	★★★★
Pure RL (PPO + DR)	★★	★★★★	★★	mittel	hoch	★
Imitation Learning / BC	★	★	★	mittel (real)	gering	★
Motion Imitation (AMP/ASE)	★★	★★	★★	mittel (MoCap)	hoch	★
Teleoperation	n/a	★	Bootstrap	hoch (real)	gering	★
Foundation / VLA	★★	★★★★	★	sehr hoch	sehr hoch	★
World Models	★	★★	★	mittel	mittel-hoch	★
Hybrid (RL + DR + T/S + WBC + VLA)	★★★★	★★★★	★★★★	mittel-hoch	hoch	★★★★

Legende: ★★★★ sehr gut ·★★ gut · ★ ausreichend.

7.3 Lessons Learned

- 01 **Keine einzelne Familie genügt.** Reine VLAs erreichen heute nicht die Robustheit klassischer Steuerung; reine RL-Policies verstehen keine Sprache; reine BC-Policies generalisieren schlecht.
- 02 **Die robusten Stacks sind hybride Stacks.** Helix ist explizit System-1 / System-2; GROOT N1 ebenfalls; HOVER ist Multi-Mode-Distillation.
- 03 **Sim-to-Real-Robustheit kommt aus klassischen Bausteinen.** Domain Randomization, SysID, PD-Loop-Treue, Sim-to-Sim. VLAs und World Models adressieren ein anderes Problem.
- 04 **Industriell deployt: Safety mitdenken.** Eine Architektur ohne deterministische Safety-Schicht ist im EU-Industrienumfeld nicht zulassungsfähig (Kapitel 14).

Sim-to-Sim als kritische Zwischenstufe.

8.1 Was Sim-to-Sim leistet

Sim-to-Sim ist die Praxis, eine Policy vor dem Real-Deployment in einem **zweiten, andersartig implementierten Simulator** zu evaluieren. Im Stack von `unitree_rl_mjlab` entspricht das dem `play`-Schritt im `unitree_mujoco`-Vorschau-Simulator: Die Policy wurde z. B. in MJLab / MuJoCo Warp trainiert und wird in „klassischem“ MuJoCo erneut evaluiert. Isaac Lab geht analog vor: Newton-trainierte Policies werden „first verified through sim-to-sim transfer before real robot deployment“ (Isaac Lab Documentation, 2025).

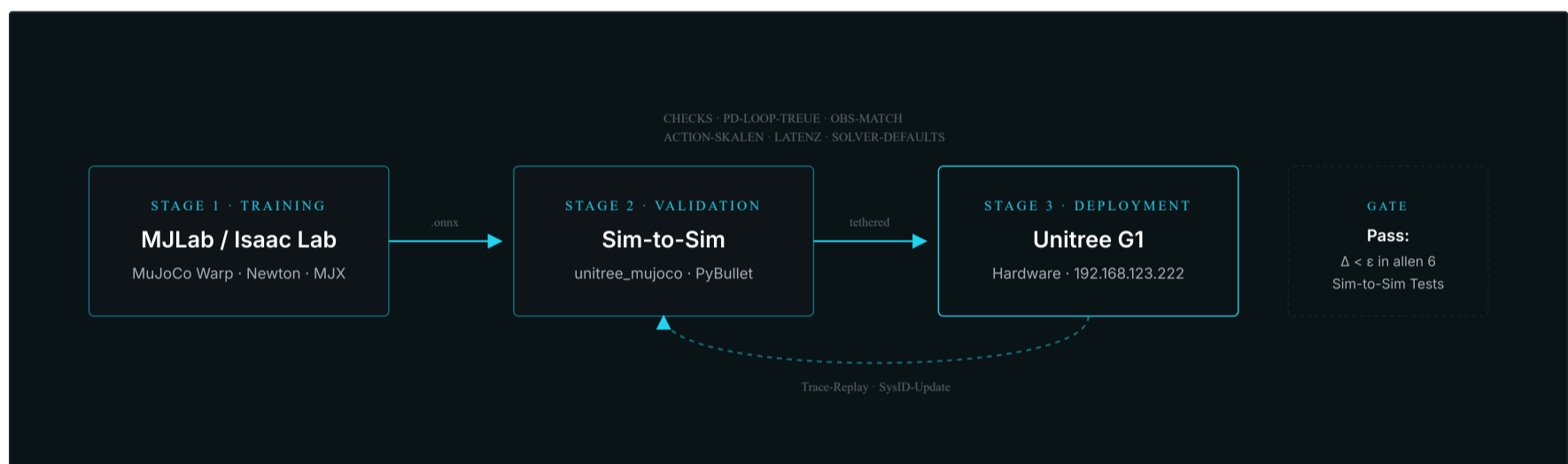


ABB. 3 Sim-to-Sim sichert die Konsistenz aller Schnittstellen zwischen Training und Deployment — bevor reale Hardware Risiko trägt.

8.2 Warum das so wichtig ist

Sim-to-Sim deckt eine Klasse von Fehlern auf, die in der Trainings-Simulation prinzipiell unsichtbar sind, im Real-Robot aber Hardware-Schäden verursachen:

- **Observation-Layout-Drift.** Reihenfolge, Skalen, Normalisierung der Beobachtungen weichen zwischen Training und Deployment ab.
- **Implizite vs. explizite PD-Controller.** Wenn der Trainings-Sim einen PD-Regler implizit im Solver hat, der Deployment-Sim aber einen expliziten, ändert sich die effektive Aktuator-Dynamik massiv.
- **Action-Skalierung.** Tanh-Squashing-Faktoren, Position-Offsets, Default-Posen werden in einer Codebasis abgesetzt, in der anderen vergessen.
- **Steuerfrequenz.** Wenn Trainings-Sim mit 100 Hz, Deployment aber mit 50 Hz läuft, kollabiert die Policy in der Hardware.
- **Solver-Annahmen.** Maximale Iterationen, Soft-Constraint-Stiffness, Friction-Cone — jede Engine hat eigene Defaults.
- **Numerische Latenzen.** Ein zusätzlicher 1-Step-Delay zwischen Beobachtung und Aktion ändert die effektive Phasenlage drastisch.

In allen genannten Fällen läuft die Policy in einem Simulator perfekt — und versagt im anderen reproduzierbar.

8.3 Kanonische Sim-to-Sim-Checkliste

Vor jedem Hardware-Test empfehlen wir folgende Sim-to-Sim-Pipeline:

- ✓ 01 **Identisches Robotermodell** (URDF/MJCF) in beiden Simulatoren laden — Prüfung über `mujoco-validate` und Visual-Diff.
- ✓ 02 **PD-Loop-Treue**. Trainings-PD-Gains 1:1 in den Deployment-Sim übernehmen und in einer Roh-Trace (Sprungantwort) vergleichen.
- ✓ 03 **Observation-Match-Test**. Policy-frei generierte Trajektorie in beiden Simulatoren ausführen und Observation-Vektoren bitweise vergleichen (max. Abweichung $< \epsilon$).
- ✓ 04 **Action-Skalentest**. Identische deterministische Policy (`tanh(0)`) in beiden Sims für 1 s laufen lassen.
- ✓ 05 **Latenz-Stress**. Künstlicher 1- und 2-Step-Delay einfügen; Policy muss robust bleiben.
- ✓ 06 **Worst-Case-DR im Eval-Sim**. Reibung 0,3–1,2 · Masse $\pm 20\%$ · Latency 10–30 ms.
- ✓ 07 **Failure-Replay**. Bekannte gefährliche Trajektorien (kniende Pose, asymmetrischer Sprung) im Eval-Sim erzwingen.

Erst nach grünem Lauf dieser Pipeline geht die Policy auf reale Hardware — idealerweise zunächst *tethered*, an einer Hebebühne und mit Kill-Switch.

Analyse des unitree_rl_mjlab- Referenzstacks.

9.1 Der Referenz-Workflow in 7 Schritten

Der offene Stack `unitree_rl_mjlab` (Unitree Robotics, 2025) ist der aktuell vollständigste herstellergetragene Open-Source-Referenz-Workflow für ein End-to-End-Sim-to-Real-Pipeline auf einem realen Unitree G1. Ein typischer Lauf durch den Stack umfasst sieben Schritte:

01 MJCF laden. Offizielles Unitree-G1-Modell in MJLab (MuJoCo Warp).	02 PPO-Training (<code>rsl_rl</code>) mit mitgelieferten Velocity-Tracking-Reward-Templates.
03 Domain Randomization für Massen, Reibung und Steuerlatenz.	04 Evaluation der Policy in <code>unitree_mujoco</code> — Sim-to-Sim.
05 Export nach ONNX.	06 Übertragung auf einen realen G1 über Ethernet (192.168.123.222, Gamepad im Debug-Mode).
07 Tethered Tests im Innenraum, danach semianspruchsvoller Untergrund.	

Auf den ersten Blick wirkt das wie ein simples Hobby-Skript. Tatsächlich ist es ein nahezu vollständiger State-of-the-Art-Sim-to-Real-Workflow — ohne die Buzzwords, mit allen wesentlichen Bausteinen.

9.2 Warum dieser Stack so robust ist

Wir bewerten den Referenzstack entlang der in Kapitel 4 definierten Reality-Gap-Domänen:

DOMÄNE	ANTWORT IM <code>UNITREE_RL_MJLAB</code> -STACK
Dynamik	Domain Randomization über Massen, Reibung, Inertien (Standard in MJLab-Templates)
Aktuatorik	Identische PD-Loop in Trainings- und Deployment-Sim (MuJoCo Warp vs. MuJoCo), 1:1-Gain-Übertragung
Sensorik	Reduzierte Observation (IMU + Encoder + Kommando), keine privilegierten Größen → kein Beobachtungs-Mismatch
Latenz	Steuerfrequenz und Action-Repeat zwischen Sim und Real abgeglichen
Beobachtungs-Mismatch	Architektur ohne Privileged Information; Teacher-Student-Distillation separat für komplexere Tasks vorgesehen

Genau diese fünf Punkte sind die Trockenarbeit, die in akademischen Papers oft als „*implementation detail*“ untergeht — und im realen Sim-to-Real die Differenz zwischen einer Policy, die geht, und einer Policy, die fällt, ausmacht.

9.3 Runtime-Gleichheit als eigentlicher Erfolgsfaktor

- `unitree_rl_mjlab` und `unitree_mujoco` teilen den MuJoCo-Solver. Die Wahrscheinlichkeit, dass der Sim-to-Sim-Sprung verzerrt, ist daher gering.
- Der Roboter selbst läuft einen PD-Loop, der vom Hersteller spezifiziert ist — und der MJCF-Modell-Default ist genau darauf abgestimmt.
- Der Action-Space (Sollpositionen für jedes Gelenk) ist deckungsgleich.

Andere Stacks (z. B. Isaac Lab → Real-Robot) verlieren hier Performance, weil PhysX-Aktuator-Defaults und MuJoCo-Aktuator-Defaults unterschiedlich sind. Der MJLab-Ansatz minimiert genau diese Friction.

9.4 Grenzen des Referenz-Workflows

<p>Grenze 01</p> <p>Skill-Skalierung fehlt</p> <p>Walk + Velocity-Tracking ist demonstriert; Manipulation, Pick & Place oder Loco-Manip-Kopplung sind nicht abgedeckt.</p>	<p>Grenze 02</p> <p>Keine privilegierte Information</p> <p>Im 1-Skill-Setup eine Stärke, im Multi-Skill-Stack eine Einschränkung — komplexere Reward-Strukturen profitieren von Teacher-Student-Distillation (Kapitel 10).</p>
<p>Grenze 03</p> <p>Keine Vision</p> <p>Sensor-Input ist propriozeptiv; visuelle Tasks brauchen separate Pipelines wie VIRAL (He et al., 2025b).</p>	<p>Grenze 04</p> <p>Keine industrielle Safety</p> <p>Für reale Industrieanwendungen fehlt der deterministische Safety-Layer (Kapitel 14).</p>

KERNBOTSCHAFT

Der `unitree_rl_mjlab`-Referenzstack ist nicht magisch, sondern **konservativ-pragmatisch**. Er enthält genau die Bausteine, die in mehr als acht Jahren Sim-to-Real-Forschung als robust herauskristallisiert wurden — und lässt jeden Baustein weg, der noch experimentell ist. Das ist die richtige Default-Architektur für einen Einstieg in Humanoid-RL.

Teacher-Student-Distillation und moderne Pipelines.

10.1 Privileged Learning und der Ursprung

Der zentrale moderne Sim-to-Real-Paradigma stammt aus der Quadruped-Forschung der ETH Zürich: Lee et al. (2020) führten in *Science Robotics* das Schema **privileged learning + student distillation** ein. Eine Teacher-Policy wird mit allen verfügbaren simulierten Beobachtungen trainiert — inklusive Terrain-Höhenkarte, Bodenkontaktindikatoren, Roboter-Schwerpunktgeschwindigkeit — und erreicht so eine Performance, die ohne diese Größen unerreichbar wäre. Ein Student-Netz lernt dann per Behavior Cloning, mit nur proprioceptiver Information (TCN / RNN über die letzte Sensorhistorie) dieselben Aktionen zu produzieren.

Das Schema wurde seitdem auf Humanoide übertragen (Zhang et al., 2025 / Distillation-PPO; He et al., 2025a / HOVER; Isaac Lab Documentation, 2025) und ist heute der dokumentierte Standardpfad.

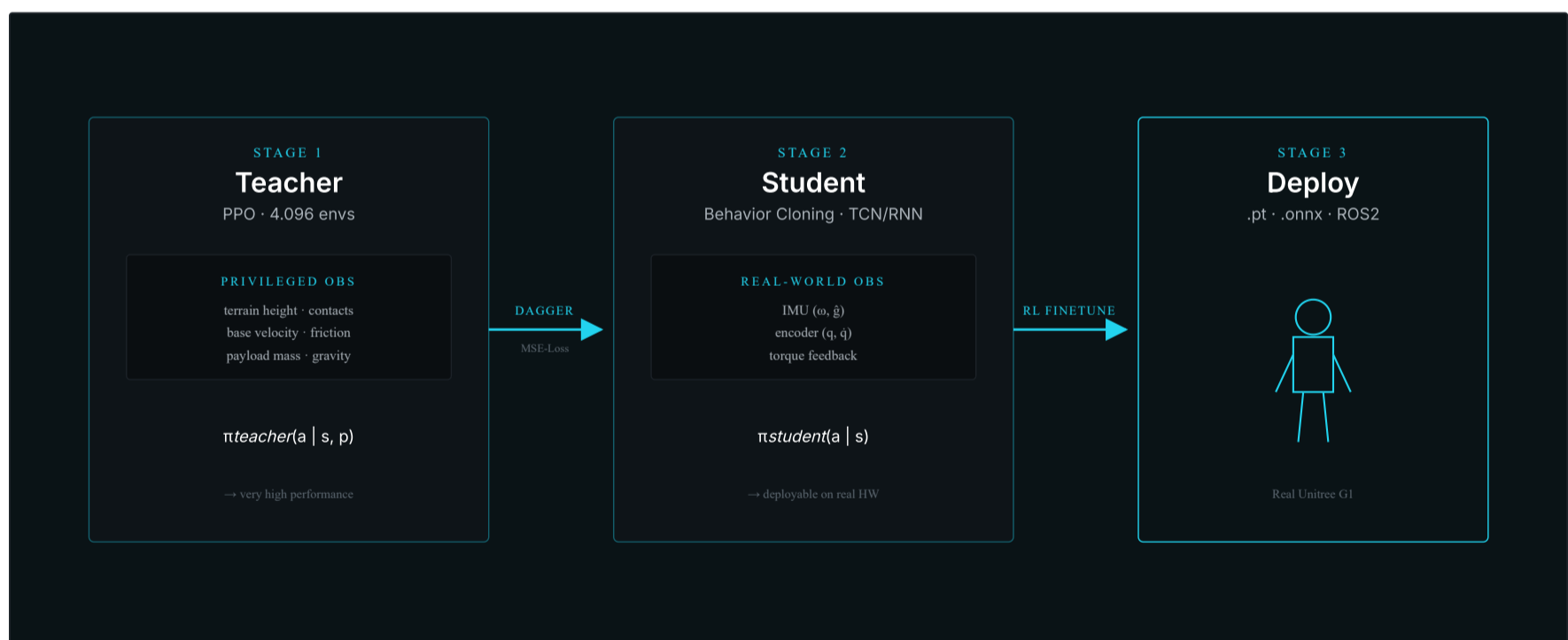


ABB. 4 Teacher-Student-Distillation — die heute robusteste Methode, um die Lücke zwischen privilegierter Sim-Beobachtung und realer Sensorik zu überbrücken.

10.2 Der Isaac-Lab-Workflow im Detail

Die offizielle Isaac-Lab-Dokumentation (2025) beschreibt den Workflow am konkreten Beispiel [Isaac-Velocity-Flat-G1-v1](#) in drei Stufen:

- 01 **Stage 1 — Teacher Training.** PPO mit 4.096 parallelen Environments, alle simulierten Beobachtungen verfügbar (inkl. Basislineargeschwindigkeit).
- 02 **Stage 2 — Student Distillation.** Behavior Cloning mit MSE-Loss zwischen Teacher- und Student-Aktionen; Student sieht nur reale Sensorik: IMU, Encoder, Torque-Feedback.

03 **Stage 3 — RL-Finetuning.** Der distillierte Student wird mit reduziertem Beobachtungsraum noch einmal via RL nachoptimiert, um über das reine Imitieren des Teachers hinaus an Robustheit zu gewinnen.

Wichtige Designdetails:

- Der Student bekommt nicht die exakte Roh-IMU, sondern bereits gefilterte Größen, um die Distillation gegen Sensorrauschen zu stabilisieren.
- Die Privileged-Information wird strikt nur im Teacher-Training und nicht im Student-Loop genutzt — sonst entsteht ein „Leak“ und Sim-to-Real bricht zusammen.
- Die exportierten Policies werden in `.pt` und `.onnx` für deployment-Flexibilität ausgegeben.

10.3 HOVER — universeller Whole-Body-Controller

He et al. (2025a) verallgemeinern Teacher-Student für humanoide Whole-Body-Control mit **HOVER**:

- **Trainings-Setup:** Isaac mit massiv paralleler GPU-Physik — „1 year robot training in 50 minutes wall clock on a single GPU“.
- **Oracle:** Eine Policy lernt sehr breit definierte Modi (kinematic position tracking, local joint angle tracking, root tracking) auf privilegierter Information.
- **Student:** Ein einzelnes Netz lernt via DAgger einen Multi-Mode-fähigen Whole-Body-Controller, der zero-shot auf reale Humanoide transferiert.

10.4 Multi-Simulator-Distillation: PolySim und VIRAL

PolySim · Lei et al., 2025

Multi-Sim Dynamics Randomization

Policy parallel über mehrere heterogene Simulatoren trainiert. **52,8 % Verbesserung** in MuJoCo-Sim-to-Sim-Evaluation und Zero-Shot-Deployment auf realem G1.

VIRAL · He et al., 2025b

Skaliertes Teacher-Student bis 64 GPUs

Aggressive visuelle Domain Randomization + Real-to-Sim-Alignment. Bis zu **54 ununterbrochene Loco-Manipulation-Zyklen** auf einem Unitree G1 ohne Real-World-Finetuning.

Konsequenz für den Standard-Stack. Wer mehr will als ein einzelnes Lauf-Skill, sollte Teacher-Student als Default-Pfad einplanen — nicht weil es elegant ist, sondern weil es die einzige skalierbare Antwort auf die Beobachtungs-Mismatch-Frage ist.

Hierarchical Policies und modulare Skill-Systeme.

11.1 Warum eine einzige große Policy keine gute Idee ist

Es gibt eine intuitive Versuchung, Humanoid-Steuerung als „eine einzige, sehr große End-to-End-Policy“ zu sehen, die alles tut — laufen, greifen, sortieren, sprechen. Vier Gründe sprechen dagegen:

- 01 **Frequenz-Asymmetrie.** Sprachverstehen und Planung brauchen 1–10 Hz, Skill-Auswahl 10–50 Hz, Whole-Body-Control 200–500 Hz, Aktuator-PD 1–2 kHz. Eine einzige Policy kann diese vier Zeitskalen nicht effizient bedienen.
- 02 **Trainings-Effizienz.** Pro Skill optimiertes Reward-Engineering ist um Größenordnungen sample-effizienter als ein einziges, globales Reward.
- 03 **Wartbarkeit.** Ein neuer Skill darf nicht die Performance bestehender Skills zerstören (Catastrophic-Forgetting-Risiko bei monolithischen Policies).
- 04 **Sicherheit.** Eine deterministische Untergrenze (PD-Loop, Torque-Limits, Reflexe) ist nur in einem geschichteten Stack realisierbar.

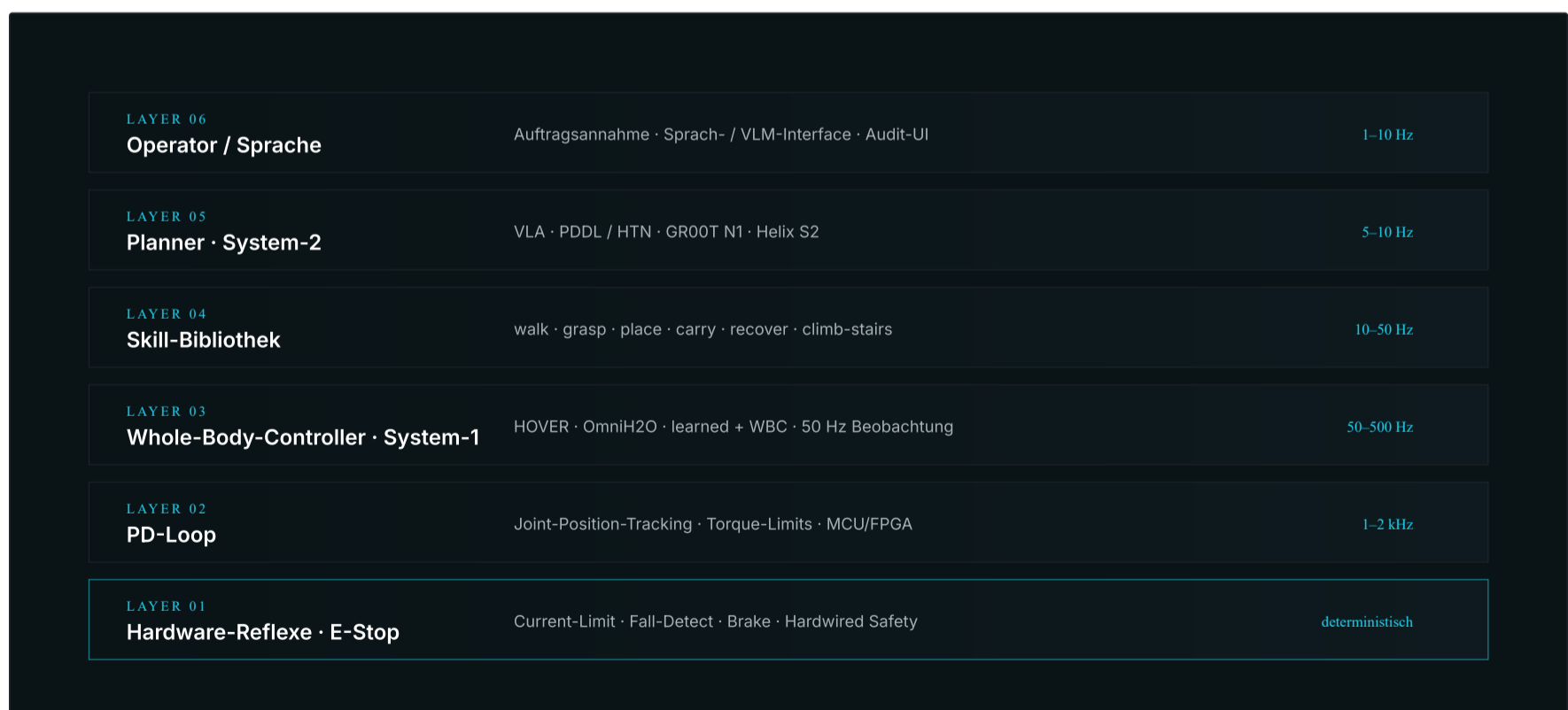


ABB. 5-6 Hierarchische Policies trennen Sprache & Planung, Skill-Auswahl, Whole-Body-Control und Aktuator-Loop entlang ihrer natürlichen Frequenzen. Diese Trennung spiegelt sich 1:1 in Helix (Figure AI, 2025) und GROOT N1 (NVIDIA Research, 2025) wider.

11.2 Skill-Bibliothek und Auswahl

Locomotion

Manipulation

walk · sidestep · turn · stand-up · climb-stairs (Siekmann et al., 2021 / Cassie Blind Stair Traversal)

reach · grasp · place · push · hand-over

Loco-Manip

carry-and-walk · tray-balance · two-handed lift

Recovery

fall-prevent · soft-landing · recover

Die Auswahl, welcher Skill aktiviert wird, übernimmt ein höherwertiger Planner — entweder symbolisch (PDDL / HTN) oder generativ (VLM / VLA, OpenVLA / GROOT N1).

11.3 Distillation als Brücke zwischen Skills

Ein wichtiger Bonus modularer Skill-Systeme: Sie sind ideal mit Teacher-Student-Distillation kompatibel. Jeder einzelne Skill kann mit Privileged Information trainiert und anschließend in einen sensorisch sparsamen Student überführt werden — und am Ende übernimmt ein Whole-Body-Controller (z. B. HOVER) die multi-mode-Distillation aller Skills in eine einzelne, deploybare Policy.

Industrial Humanoid Robotics.

12.1 Anwendungsklassen

KLASSE	BEISPIEL	CYCLE TIME	REIFEGRAD 2026
Sortieren / Sequenzieren	Pakete nach Größe, Werkstücke nach Form	5–15 s	Frühe Pilots
Pick & Place / Materialhandling	Tote-Movement (Agility Digit @ GXO)	10–30 s	Kommerziell · > 100k Totes
Montage / Assembly	Lineside-Delivery, Kit-Vorbereitung (Apptronik @ Mercedes-Benz)	30–120 s	Pilot
Qualitätsprüfung	Visuelle Inspektion mit Multi-Posen-Erfassung	5–30 s	Forschung
Mobile Manipulation	Tray- / Werkzeug-Transport durch Werkhalle	minutenskalig	Forschung / Frühe Pilots
MRK-Kollaboration	Werker reicht Bauteil, Humanoid stabilisiert	-	Forschung

Quellen u. a.: GXO Logistics & Agility Robotics (2024, 2025), Apptronik & Mercedes-Benz (2024), Boston Dynamics & RAI Institute (2025), Humanoids Daily (2026).

12.2 Architektur eines industriellen Humanoid-Stacks

Aus den realen Pilots der letzten 24 Monate (Apollo @ Mercedes-Benz, Digit @ GXO, Atlas-Forschung, Optimus @ Tesla) lässt sich eine wiederkehrende Architektur ableiten:

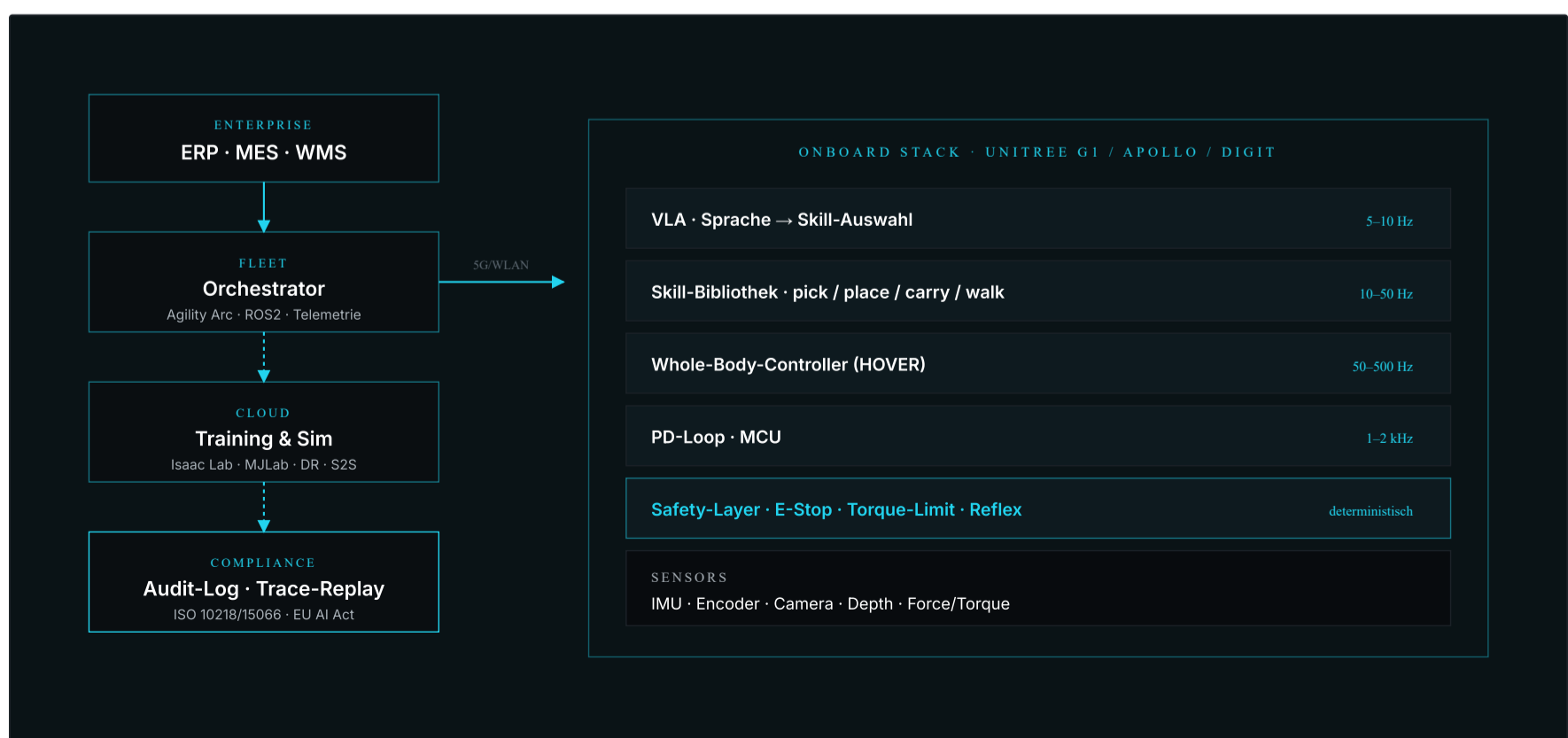


ABB. 7 Industrielle Humanoid-Mobile-Manipulation-Architektur — ERP / MES, Fleet-Orchestrator, Cloud-Training, Onboard-Stack, Audit-Log.

Die zwei nicht-verhandelbaren Bausteine sind:

- 01 **Ein Fleet-Orchestrator** (Agility Arc, ROS2 + Orchestrator-Layer, proprietäre Pendants). Er kapselt Auftrags-Backlog, Task-Dispatch und Telemetrie.
- 02 **Ein Audit-Log / Safety-Track**, der für jede Aktion die ausgewählte Skill-Policy, die Eingabe-Observation und die Ausgabe-Aktion zeitgestempelt protokolliert. Notwendig für ISO 10218, 15066 und EU AI Act.

12.3 Warum Sim-to-Real industriell besonders attraktiv ist

Industrielle Umgebungen sind paradoxerweise *einfacher* für humanoide Roboter als ein Wohnzimmer:

- **Strukturierte Welt.** Stellplätze, Markierungen, Lichtverhältnisse weitgehend kontrolliert.
- **Wiederholbare Tasks.** Ein Skill, einmal trainiert, läuft millionenfach unverändert.
- **Vorhersehbare Sicherheits-Zonen.** Keine Kleinkinder, keine Haustiere, keine improvisierten Möbel.
- **Wirtschaftliche Skalierung.** Ein Roboter, der eine 8-Stunden-Schicht ersetzt, amortisiert sich mit RaaS-Modellen (GXO ↔ Agility) in 12–24 Monaten.

Aus Sim-to-Real-Sicht. Trainings-Domain möglichst eng auf die spezifische Industrie-Umgebung zuschneiden, Domain Randomization eng halten, Foundation-Modelle nur dort einsetzen, wo der Skill-Raum tatsächlich breit ist.

12.4 Konkrete Belege aus realen Pilots

GXO · AGILITY DIGIT

Mehrjähriges RaaS-Abkommen, mehr als **100.000 bewegte Totes** in einer einzigen Standort-Installation.

MERCEDES-BENZ · APTRONIK APOLLO

Pilot am *Mercedes-Benz Digital Factory Campus Berlin* zur autonomen Bereitstellung von Montagekits inkl. Komponenten-Inspektion.

FIGURE AI · HELIX

Helix-02-Roboter halten in einem öffentlichen Livestream eine **8-Stunden-Schicht autonomer Paket-Sortierung** durch (Figure AI, 2026).

BOSTON DYNAMICS · ATLAS

Partnerschaft mit dem RAI Institute für RL-basierte Whole-Body-Manipulation und dynamic running.

TESLA · OPTIMUS

Pivot von MoCap-Teleop-Daten auf Camera-Rig + *Neural World Simulator* (Humanoids Daily, 2026).

Diese Pilots haben *keine* der Magie-VLA-Träume aus dem Marketing — aber sie haben *alle* die Bausteine, die wir in diesem Whitepaper diskutieren: massives Sim-Training, Sim-to-Sim-Validierung, Distillation, hierarchische Policies, Safety-Layer.

12.5 Die Rolle von Skill-Recording am realen Roboter

Eine wichtige industrielle Ergänzung zum reinen Sim-Training ist die **Aufzeichnung am Arbeitsplatz**: Der eigentliche Skill wird nicht in einer abstrakten Simulationsumgebung definiert, sondern dort, wo er später läuft. Die wetando-Praxis (siehe Abb. A und B) zeigt, dass folgende Kette pragmatisch funktioniert:

- 01 **Reale Skill-Demonstration** durch einen Mitarbeiter
- 02 **Konvertierung in MoCap-/CSV-Referenz** (analog zu [unitree_rl_mjlab](#) motion imitation)
- 03 **RL-Finetuning** in MJLab / Isaac Lab

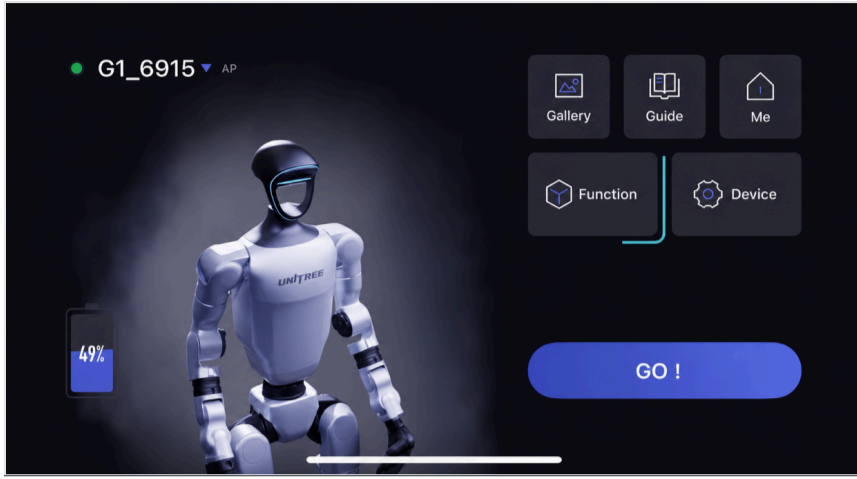


ABB . E Smartphone-Oberfläche zur Skill- und Funktions-Steuerung eines realen G1.

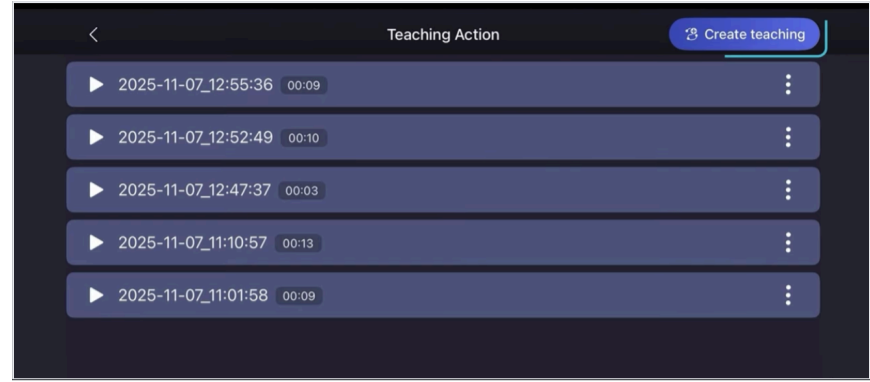


ABB . F Aufgezeichnete Skill-Demonstrationen aus einer Teaching-Session.

Warum Haushaltsrobotik deutlich schwieriger ist.

Haushaltsrobotik wirkt visuell einfacher (kein Schweißroboter, kein Stahlwerk), ist aber technisch dramatisch härter.

13.1 Die fünf strukturellen Hindernisse

- 01 **Unstrukturierte Welt.** Möbel werden verschoben, Spielzeug liegt am Boden, Lichtverhältnisse ändern sich täglich.
- 02 **Long-Tail-Objekte.** Millionen verschiedener Tassen, Geschirrtücher, Verpackungen — jede mit eigener Greifgeometrie.
- 03 **Sicherheit mit Kleinkindern und Haustieren.** Eine Bag-of-Reflexes-Logik wie in der Fabrik ist nicht ausreichend.
- 04 **Soziale Erwartungen.** Geschwindigkeit, Geräuschpegel, „Höflichkeit“ sind harte Anforderungen, die im Industrieumfeld irrelevant sind.
- 05 **Wirtschaftliches Pricing.** 20.000 USD pro Gerät sind für viele Haushalte prohibitiv — während sechsstellige RaaS-Verträge in der Logistik trivial finanzierbar sind.

13.2 Was VLAs heute schon können

Mit Helix (Figure AI, 2025) und GROOT N1 (NVIDIA Research, 2025) ist 2025/2026 bewiesen worden, dass VLAs Tausende von Haushaltsobjekten zero-shot greifen und in Sprache umsetzen können. Helix-02 demonstriert sogar mehrere Stunden autonomen Schichtbetrieb auf einer Sortierstraße (Figure AI, 2026). Das ist beachtlich.

WAS DIESE DEMOS NICHT ZEIGEN

- 24/7-Betrieb über Wochen ohne menschliche Korrektur,
- robuste Fehlerbehandlung bei unerwarteten Objekten (z. B. zerbrochenes Glas),
- regulatorisch zugelassener Betrieb in einem Privathaushalt mit Kindern.

13.3 Konsequenz

Die *erste echte kommerzielle Massen Anwendung* humanoider Roboter wird Industrie und Logistik sein — nicht der Haushalt. Diese Reihenfolge ändert sich nach unserer Einschätzung mindestens bis 2030 nicht.

Gestützt durch die tatsächliche Verteilung kommerzieller Pilots 2025/2026: zwei Drittel sind industrielle Deployments, ein Drittel sind Haushalts-Forschungsdemos.

Sicherheits- und Runtime-Aspekte.

14.1 Der Runtime-Stack

Ein deployter humanoider Roboter im industriellen Einsatz hat eine geschichtete Runtime, in der jede Schicht ihre eigene Frequenz, ihre eigene Sprache und ihr eigenes Failure-Behavior hat:

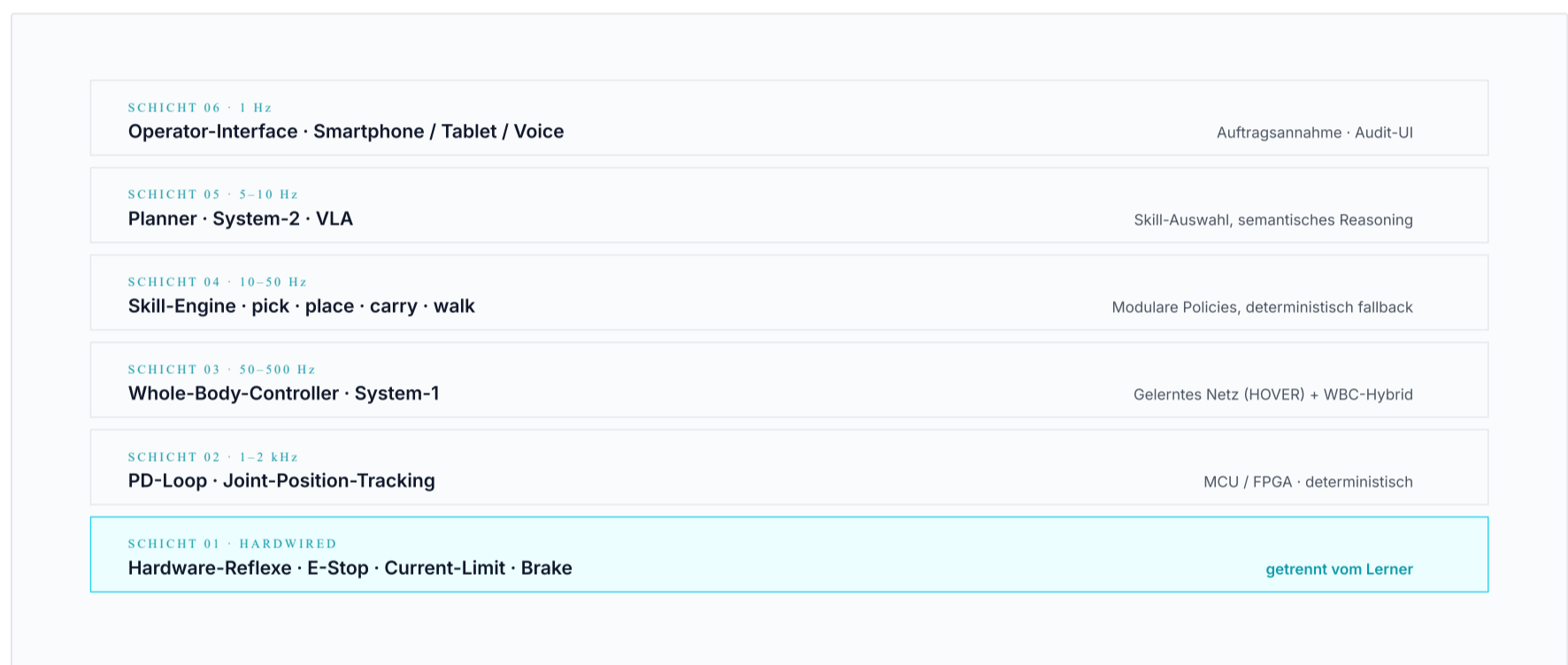


ABB. 9 Runtime-Stack vom Operator-Interface bis zu den Hardware-Reflexen. Wir schlagen vor, diese sechs Schichten als verbindliche Konvention für jeden ernst gemeinten industriellen Stack zu etablieren.

14.2 Drei Safety-Prinzipien

01

Determinismus auf Hardware-Nähe

PD-Loop, Torque-Limits und E-Stop dürfen *nicht* in einer gelernten Policy liegen. Sie sind in MCU/FPGA implementiert, getrennt vom Lerner.

02

Fail-Safe-Fallback

Bei Verlust der ML-Policy (Latenz > Schwelle, NaN-Aktion, Sensor-Drop) muss der Whole-Body-Controller in eine analytische Default-Pose („sitzen“, „bracing“) gehen.

03

Audit-Trail per Trace-Replay

Jede ausgelieferte Aktion, jede Observation und jede aktive Policy-Version wird in einem zeitlich annotierten Log gespeichert, das hardware-extern repliziert wird.

14.3 Regulatorischer Rahmen

EN ISO
10218-1/-2

Industrieroboter, klassisch.

ISO/TS
15066

Collaborative Robots; speed-
and-separation, power-
and-force.

ISO/IEC
22989

KI-Grundbegriffe.

EU
AI Act

Mind. „High-Risk“ für viele
industrielle Humanoid-
Anwendungen, mit
dokumentationspflichtigen
Trainings- und
Validierungsschritten.

Praktische Konsequenz. Ein VLA, das auf Web-Daten trainiert wurde, ist im EU-Kontext nicht trivial einsetzbar. Die Wahrscheinlichkeit, dass die ersten industriellen Massen-Deployments deshalb *hybride* Architekturen sind — mit deterministischen Safety-Layern, geprüften Skill-Policies und VLAs ausschließlich für die *Auftragsannahme*, nicht für die *Steuerung* — ist hoch.

Zukunft: Foundation Policies und World Models.

15.1 Foundation Policies

Foundation Policies versuchen, das, was große Sprachmodelle für Text erreicht haben, für Steuerung zu replizieren: ein einziges, sehr großes Modell, das *generalisiert*.

NVIDIA · 2025

Isaac GROOT N1

Erstes offenes Humanoid-Foundation-Modell mit System-1 / System-2-Architektur, trainiert auf realen Trajektorien, Human-Videos und synthetischen Daten.

FIGURE AI · 2025

Helix

Geschlossenes VLA, bestes industriell demonstriertes Beispiel für hochfrequente System-1-Steuerung (200 Hz) mit semantischer System-2-Schicht (7–9 Hz).

STANFORD · 2024

OpenVLA

7B-Parameter, Open Source, trainiert auf Open X-Embodiment.

GOOGLE & CO. · 2024

RT-X

22 Roboter-Embodiments, 527 Skills, 160.266 Tasks; demonstriert positiver Transfer über Embodiments hinweg.

Die Frage 2026 ist nicht mehr „funktionieren diese Modelle“, sondern „sind sie reaktiv genug, sicher genug, validierbar genug für Industrie“.

15.2 World Models

World Models lernen einen Surrogat-Simulator aus realen Sensordaten und planen in dessen Latent-Space:

- **Dreamer V3** (Hafner et al., 2023) generalisiert über mehr als 150 Domänen mit fixen Hyperparametern; erstes Verfahren, das *Minecraft Diamond* ohne Demonstrationen löst.
- **1X World Model** (1X Technologies, 2026) sagt zukünftiges Video aus Beobachtungs- und Aktionssequenzen vorher und ist explizit für *Evaluation* von Home-Robotern gedacht.
- **Tesla „Neural World Simulator“** (Humanoids Daily, 2026) nutzt die FSD-Trainings-Infrastruktur, jetzt auch für Optimus.

World Models sind die wahrscheinliche zweite große Welle nach VLAs — mit einer eigenen Reality Gap (Modell-Bias) und einer eigenen Methodik (Multi-Sim-Konsistenz, Differential-Physics, vgl. Newton).

15.3 Was bleibt vom klassischen Sim-to-Real?

Domain Randomization , weil reale Hardware altert.	Teacher-Student , weil privilegierte Beobachtungen niemals vollständig auf reale Sensoren reduzierbar sind.	Sim-to-Sim , weil jede Engine eigene Annahmen hat.	PD-Loop-Treue , weil die Aktuatoren sich nicht ändern lassen.	Safety-Layer , weil regulatorische Anforderungen unabhängig vom KI-Trend bleiben.
--	---	--	---	---

Das ist die *konservative* These dieses Whitepapers: Foundation Policies und World Models **ergänzen** den Stack, sie ersetzen ihn nicht.

Kritische Bewertung aktueller Ansätze.

16.1 End-to-End-AGI-Versprechen

Mehrere Marktakteure suggerieren, eine einzige, sehr große End-to-End-Policy könne sämtliche humanoiden Skills abdecken („Solve robotics end-to-end“). Wir halten das aus drei Gründen für unrealistisch:

- › **Frequenzmismatch.** Ein 7B-Modell auf einer Onboard-GPU kann nicht 1 kHz steuern; ein 1-kHz-PD-Loop kann nicht semantisch reasonen.
- › **Failure-Modi.** Eine monolithische Policy hat keine deterministische Untergrenze; bei Sensor-Drop oder Latenz-Spike gibt es keinen sauberen Fallback.
- › **Regulatorische Validierung.** Eine monolithische Policy ist als Single-Use „High-Risk-AI“ nach EU AI Act schwer zu zertifizieren.

16.2 Reine klassische Robotik

- › **Schlechte Generalisation** über Terrain, Lasten, Beobachtungsrauschen.
- › **Sehr hoher Engineering-Aufwand** pro Skill.
- › **Schwer kombinierbar** mit modernen VLAs (kein direkter Sprach-Input).

16.3 Reine Motion-Imitation

DeepMimic / AMP / OmniH2O sind exzellent für *Bewegungsqualität*, aber:

- › **MoCap-Abhängigkeit.** Für viele industrielle Tasks gibt es keine sinnvollen MoCap-Daten.
- › **Embodiment-Mismatch.** Menschliche Bewegungen sind nicht 1:1 auf ein 23-DOF-System übertragbar.
- › **Begrenzte Generalisation** über die Trainings-Bibliothek hinaus.

16.4 Reine Teleoperation

Teleoperation (HumanPlus, OmniH2O) ist als *Datenquelle* und *Backup-Modus* wertvoll, aber kein Endzustand: Sie löst nicht das Skalierungsproblem (1 Mensch ↔ 1 Roboter).

16.5 Die hybride Architektur als Konsens

Die in den letzten zwei Jahren erfolgreichsten Stacks (Helix, GROOT N1, HOVER, VIRAL, PolySim, `unitree_rl_mjlab`, Apollo, Digit) sind alle *hybride* Stacks. Sie kombinieren:

massives RL (PPO)

Sim-to-Sim

Sim-to-Real mit DR

Teacher-Student

Hierarchical Policies

System Identification

Runtime-Konsistenz

Safety-Layer

modulare Skill-Policies

KONSENS 2025/2026

Genau diese Kombination — nicht ein einzelner Bestandteil — ist die Antwort auf das Sim-to-Real-Problem.

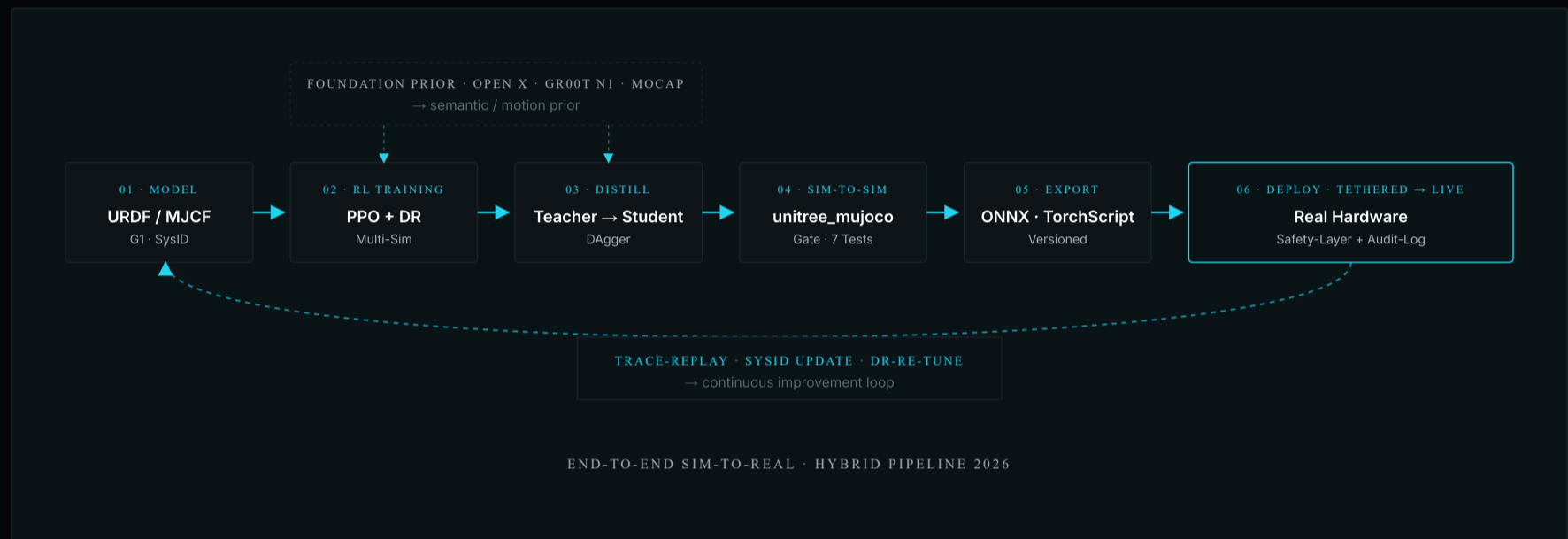


ABB. 2 · 8 Die hybride End-to-End-Pipeline — Robotermodell, massives RL mit Multi-Sim, Distillation, Sim-to-Sim, Export, Deployment, Trace-Replay-Feedback und Foundation-Priors.

Fazit und Handlungsempfehlungen.

17.1 Fazit

Humanoide Roboter sind 2025/2026 an einem Punkt, an dem sie **industriell nutzbar** werden. Sie sind es nicht trotz der konservativen Bausteine — sie sind es *wegen* dieser Bausteine.

Die hybride Architektur, die wir in diesem Whitepaper beschrieben haben, ist nicht spektakulär. Sie ist robust. Der offene Referenzstack [unitree_rl_mjlab](#) auf einem Unitree G1 illustriert diese Logik pars pro toto: weder magisch noch innovativ, aber *vollständig*. Genau diese Vollständigkeit ist das, was Sim-to-Real-Projekte heute auszeichnet, die funktionieren.

Foundation Policies (Helix, GROOT N1, OpenVLA) und World Models (Dreamer V3, 1X World Model) ergänzen den Stack um Semantik und Sample-Effizienz — sie ersetzen aber keine der klassischen Brücken, weil die zugrunde liegenden Probleme (Beobachtungs-Mismatch, Solver-Bias, Aktuator-Reality-Gap, regulatorische Validierbarkeit) struktureller Natur sind.

17.2 Handlungsempfehlungen

A · FORSCHUNG & HOCHSCHULEN

- Beginnen Sie mit [unitree_rl_mjlab](#) oder MuJoCo Playground auf einer Single-GPU-Workstation. Sie reproduzieren damit in wenigen Tagen den State of the Art für 1-Skill-Locomotion.
- Investieren Sie in eine *Sim-to-Sim-Pipeline*, bevor Sie in Hardware investieren. 90 % der Sim-to-Real-Probleme sind in Sim-to-Sim sichtbar.
- Setzen Sie auf offene Foundation-Modelle (OpenVLA, GROOT N1) statt auf proprietäre Stacks.

B · INDUSTRIEUNTERNEHMEN

- Setzen Sie nicht auf *eine* Plattform, sondern auf *eine Architektur*. Plattformen werden sich konsolidieren; die hybride Architektur ist stabil.
- Pilot-Use-Case eng definieren: ein wiederholbarer Task, in einer kontrollierten Umgebung, mit klarer Cycle-Time- und Safety-Anforderung.
- Verlangen Sie von Anbietern *Trace-Replay-Logs* und *Sim-to-Sim-Reports*, nicht nur YouTube-Demos.

C · INNOVATIONS- UND TRANSFERZENTREN

- Bauen Sie *eine* Referenz-Workbench (z. B. Unitree G1 + Isaac Lab + MJLab + [unitree_rl_mjlab](#)) auf, die mehrere KMU teilen können.
- Vermitteln Sie an KMU, dass Humanoid-Robotik in der Industrie heute *nicht magisch* ist, sondern aus dem hier beschriebenen Engineering-Workflow stammt.

D · KMU MIT KONKRETEM ANWENDUNGSFALL

- Beginnen Sie mit Robots-as-a-Service-Verträgen (Agility, Aptronik), bevor Sie eigene Hardware kaufen.
- Definieren Sie den Mehrwert über Cycle Time und 24/7-Verfügbarkeit — nicht über Faszination.
- Verlangen Sie eine Safety-Strategie inkl. ISO 10218 / 15066 / EU AI Act — auch wenn der Anbieter sie nicht ungefragt liefert.

17.3 Schluss

SCHLUSSBILD

Das Sim-to-Real-Problem ist nicht „gelöst“. Es ist *beherrschbar geworden*. Genau in diesem Übergang — von „unbeherrschbar“ zu „beherrschbar“ — entstehen die wirtschaftlich nutzbaren Anwendungen.

Wer die hier beschriebenen Bausteine systematisch anwendet, kann die nächsten 24 Monate humanoider Industrie-Robotik aktiv mitgestalten, statt nur darauf zu reagieren.

AUSBLICK

Band 02 der robotskills-Whitepaper-Reihe

„Skill-Recording am Arbeitsplatz — wie KMU eigene Humanoid-Skills wirtschaftlich aufbauen.“

Erscheint im Sommer 2026.

Literaturverzeichnis.

Alle Quellen im APA-Stil (7. Auflage). URLs entsprechen dem Stand 20. Mai 2026.

01. 1X Technologies. (2026). *1X World Model*. 1x.tech/discover/1x-world-model
02. Agility Robotics. (2024). *GXO signs industry-first multi-year agreement with Agility Robotics*. agilityrobotics.com
03. Agility Robotics. (2025, November). *Digit moves over 100k totes in commercial deployment*. agilityrobotics.com
04. Aptronik & Mercedes-Benz. (2024). *Aptronik and Mercedes-Benz enter commercial agreement*. prnewswire.com
05. Boston Dynamics & RAI Institute. (2025). *Boston Dynamics and the Robotics and AI Institute partner to advance humanoid robots through reinforcement learning*. bostondynamics.com
06. Figure AI. (2025). *Helix: A vision-language-action model for generalist humanoid control*. figure.ai/news/helix
07. Figure AI. (2026). *Figure AI's Helix-02 robots complete full 8-hour autonomous shifts*. techtimes.com
08. Freeman, C. D., Frey, E., Raichuk, A., Girgin, S., Mordatch, I., & Bachem, O. (2021). *Brax — A differentiable physics engine for large scale rigid body simulation*. arXiv. arxiv.org/abs/2106.13281
09. Hafner, D., Pasukonis, J., Ba, J., & Lillicrap, T. (2023). *Mastering diverse domains through world models*. arXiv. arxiv.org/abs/2301.04104
10. He, T., Xiao, W., Luo, Z., Wang, Z., Lin, T., Yuan, Y., Shi, G., Liu, C., Fan, L., & Zhu, Y. (2025a). *HOVER: Versatile neural whole-body controller for humanoid robots*. arXiv. arxiv.org/abs/2410.21229
11. He, T., Wang, Z., Xue, H., Ben, Q., Luo, Z., Xiao, W., Yuan, Y., Da, X., Castañeda, F., Sastry, S., Liu, C., Shi, G., Fan, L., & Zhu, Y. (2025b). *VIRAL: Visual sim-to-real at scale for humanoid loco-manipulation*. arXiv. arxiv.org/abs/2511.15200
12. Humanoids Daily. (2026). *Tesla AI Chief details unified „World Simulator“ for FSD and Optimus*. humanoidsdaily.com
13. Hwangbo, J., Lee, J., Dosovitskiy, A., Bellicoso, D., Tsounis, V., Koltun, V., & Hutter, M. (2019). Learning agile and dynamic motor skills for legged robots. *Science Robotics*, 4(26), eaau5872. science.org
14. Isaac Lab Documentation. (2025). *Sim-to-real policy transfer (Newton physics integration)*. NVIDIA / Isaac Sim. isaac-sim.github.io
15. Kim, M. J., Pertsch, K., Karamcheti, S., Xiao, T., Balakrishna, A., Nair, S., Rafailov, R., Foster, E., Lam, G., Sanketi, P., Vuong, Q., Kollar, T., Burchfiel, B., Tedrake, R., Sadigh, D., Levine, S., Liang, P., & Finn, C. (2024). *OpenVLA: An open-source vision-language-action model*. arXiv. arxiv.org/abs/2406.09246
16. Kinsley, H. (sentdex). (2025). *Training a Unitree G1 to walk with reinforcement learning*. YouTube. youtube.com
17. Lee, J., Hwangbo, J., Wellhausen, L., Koltun, V., & Hutter, M. (2020). Learning quadrupedal locomotion over challenging terrain. *Science Robotics*, 5(47), eabc5986. science.org
18. Lei, Z., Zhou, Z., Yin, S., Chen, Y., Xu, Q., Li, W., Wang, Y., Tang, B., Jing, W., & Chen, S. (2025). *PolySim: Bridging the sim-to-real gap for humanoid control via multi-simulator dynamics randomization*. arXiv. arxiv.org/abs/2510.01708
19. Makoviychuk, V., Wawrzyniak, L., Guo, Y., Lu, M., Storey, K., Macklin, M., Hoeller, D., Rudin, N., Allshire, A., Handa, A., & State, G. (2021). *Isaac Gym: High performance GPU-based physics simulation for robot learning*. arXiv. arxiv.org/abs/2108.10470
20. Mercedes-Benz Group. (2026). *Mercedes-Benz Digital Factory Campus (MBDFC) in Berlin — Humanoid robots in production*. group.mercedes-benz.com
21. Mittal, M., Yu, C., Yu, Q., Liu, J., Rudin, N., Hoeller, D., Yuan, J. L., et al. (2025). *Isaac Lab: A GPU-accelerated simulation framework for multi-modal robot learning*. arXiv. arxiv.org/abs/2511.04831
22. NVIDIA. (2025). *NVIDIA launches Newton physics engine and GROOT AI at CoRL 2025*. The Robot Report. therobotreport.com
23. NVIDIA Research. (2025). *NVIDIA Isaac GROOT N1: An open foundation model for humanoid robots*. research.nvidia.com
24. Open X-Embodiment Collaboration. (2024). *Open X-Embodiment: Robotic learning datasets and RT-X models*. arXiv / ICRA. arxiv.org/abs/2310.08864
25. Peng, X. B., Abbeel, P., Levine, S., & van de Panne, M. (2018a). DeepMimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM Transactions on Graphics*, 37(4), 143. arxiv.org/abs/1804.02717
26. Peng, X. B., Andrychowicz, M., Zaremba, W., & Abbeel, P. (2018b). Sim-to-real transfer of robotic control with dynamics randomization. *IEEE ICRA*. arxiv.org/abs/1710.06537
27. Peng, X. B., Ma, Z., Abbeel, P., Levine, S., & Kanazawa, A. (2021). AMP: Adversarial motion priors for stylized physics-based character control. *ACM Transactions on Graphics*, 40(4), 144. arxiv.org/abs/2104.02180

28. Peng, X. B., Guo, Y., Halper, L., Levine, S., & Fidler, S. (2022). ASE: Large-scale reusable adversarial skill embeddings for physically simulated characters. *ACM Transactions on Graphics*, 41(4), Art. 94. doi.org/10.1145/3528223.3530110
29. Robotics & Automation News. (2025, 24. November). *Agility Robotics' Digit humanoid passes 100,000-tote milestone*. roboticsandautomationnews.com
30. Salvato, E., Fenu, G., Medvet, E., & Pellegrino, F. A. (2021). Crossing the reality gap. *IEEE Access*, 9, 153171–153187. ieeexplore.ieee.org
31. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). *Proximal policy optimization algorithms*. arXiv. arxiv.org/abs/1707.06347
32. Siekmann, J., Green, K., Warila, J., Fern, A., & Hurst, J. (2021). *Blind bipedal stair traversal via sim-to-real reinforcement learning*. RSS XVII. arxiv.org/abs/2105.08328
33. Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W., & Abbeel, P. (2017). Domain randomization. *IEEE/RSJ IROS*. arxiv.org/abs/1703.06907
34. Todorov, E., Erez, T., & Tassa, Y. (2012). MuJoCo: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems* (S. 5026–5033). IEEE. doi.org/10.1109/IROS.2012.6386109
35. Unitree Robotics. (2025). *unitree_rl_mjlab* [Software]. GitHub. github.com/unitreerobotics
36. Unitree Robotics. (2026). *G1 humanoid robot — product page and specifications*. unitree.com/g1
37. Yu, W., Kumar, V. C. V., Turk, G., & Liu, C. K. (2019). *Sim-to-real transfer for biped locomotion*. arXiv / IROS. arxiv.org/abs/1903.01390
38. Zakka, K., Tabanpour, B., Liao, Q., Haiderbhai, M., Holt, S., Luo, J. Y., Allshire, A., Frey, E., Sreenath, K., Kahrs, L. A., Sferrazza, C., Tassa, Y., & Abbeel, P. (2025). *MuJoCo Playground*. arXiv. arxiv.org/abs/2502.08844
39. Zakka, K., Liao, Q., Yi, B., Le Lay, L., Sreenath, K., & Abbeel, P. (2026). *mjlab: A lightweight framework for GPU-accelerated robot learning* (arXiv:2601.22074). arxiv.org/abs/2601.22074
40. Zhang, Q., et al. (2025). *Distillation-PPO: A novel two-stage reinforcement learning framework for humanoid robot perceptive locomotion* (arXiv:2503.08299). arxiv.org/abs/2503.08299

Was sich für Sie wirklich lohnt.

Sie haben jetzt die Architektur gesehen, die heute funktioniert. Die nächste Frage ist die ökonomische: **wo in Ihrem Unternehmen rechnet sich ein humanoider Roboter tatsächlich?** Genau dafür haben wir den **Robotik-Check** gebaut — eine methodische, herstellerunabhängige Beratungsleistung mit belastbarer Amortisationsrechnung.

Phase 01	Phase 02	Phase 03	Phase 04
Vorgespräch & Zieldefinition Unternehmenskontext, Erwartungen, Rahmen — und ob ein Check für Sie überhaupt sinnvoll ist.	Einblick & Realitätscheck Überblick über aktuelle Humanoide und Trainingsmethoden — mit ehrlicher Einordnung echter Fähigkeiten gegenüber Marketing-Videos.	Use-Case-Identifikation & Priorisierung Methodische Erfassung Ihrer Einsatzfelder und Bewertung nach Wertschöpfungspotenzial. Das Kernstück.	Auswertung & Roadmap Priorisierte Use-Case-Liste, Amortisationsrechnung und Trainingsvorschlag — auf Wunsch mit Demo am realen Unitree G1.
Remote · 60 Min.	Vor Ort · oder remote	Workshop · vor Ort	Bericht + Übergabe

BEWERTUNGSLOGIK

- A · **Wertschöpfungspotenzial** — Entlastung, Durchsatz, Qualität

- B · **Technische Machbarkeit** — realistisch mit heutiger Hardware umsetzbar?

- C · **Integrationsaufwand** — Einrichtung, Safety, Anlernen im laufenden Betrieb

- D · **Amortisation** — ab wann trägt sich die Investition

WAS SIE ERHALTEN

- Robotik-Check-Report** — priorisierte Use-Case-Liste, Bewertung je Einsatzfeld, Amortisationsrechnung, Handlungsempfehlung.

- Beratungsbericht** — strukturiert nach Anforderungen öffentlicher Förderprogramme, einreichbar als Verwendungsnachweis.

AB

4.500 €

ZZGL. MWST · ENDPREIS JE NACH UMFANG

HÄUFIG FÖRDERFÄHIG

Als externe Beratungsleistung über mehrere öffentliche Förderprogramme anteilig bezuschussbar — **Zuschüsse bis 80 %** möglich (BAFA, INQA-Coaching, Länderprogramme). Welches Programm zu Ihnen passt, prüfen wir vor dem ersten Termin.

[Erstgespräch buchen →](#)
[robotikcheck.de ↗](https://robotikcheck.de)
[hallo@robotskills.de ↗](mailto:hallo@robotskills.de)

